

Carl von Ossietzky  
Universität Oldenburg

Studiengang Hörtechnik und Audiologie  
Master of Science (M.Sc.)

## Masterarbeit

Titel:

**Analyse der menschlichen und  
algorithmischen Fähigkeit zur  
Schätzung der relativen Sprecheraus-  
richtung**

vorgelegt von:  
Sven Franz (Mat.Nr: 1031147)

Betreuernder Gutachter: Herr Prof. Dr.-Ing. Jörg Bitzer

Zweiter Gutachter: Herr Prof. Dr. ir. Simon Doclo

Oldenburg, 30. März 2012

# **Analyse der menschlichen und algorithmischen Fähigkeit zur Schätzung der relativen Sprecherausrichtung**

Masterarbeit

von

**SVEN FRANZ**

Mat.Nr: 1031147

Datum: 30. März 2012

Betreuer und Erstgutachter: Herr Prof. Dr. Ing. Jörg Bitzer

Zweiter Gutachter: Herr Prof. Dr. ir. Simon Doclo

Carl von Ossietzky Universität Oldenburg  
Fakultät V - Mathematik und Naturwissenschaften  
Medizinische Physik  
Carl-von-Ossietzky-Str. 9 - 11  
26129 Oldenburg

und

Institut für Hörtechnik und Audiologie  
Jade Hochschule Oldenburg  
Fachbereich Bauwesen und Geoinformation (B+G)  
Ofener Str. 16  
D-26121 Oldenburg

## Zusammenfassung:

Diese Arbeit befasst sich mit der Wahrnehmung und Unterscheidungsfähigkeit der relativen Sprecherausrichtung von normalhörenden Menschen und von Signalverarbeitungsalgorithmen. Die Ziel dieser Arbeit war die Entwicklung von Signalverarbeitungsalgorithmen für binaural gekoppelte Hörgeräte, die in der Lage sind, Hauptsprecher in einer Multi-Sprecher Situation zu bestimmen und so eine Richtungsschätzung mit einer zusätzlichen Information für eine verbesserte Ausrichtung von Richtmikrofonen zu unterstützen. Ein sinnvolles Attribut für einen Hauptsprecher stellt die Sprecherausrichtung relativ zum Empfänger dar, da sich diese bei einer Konversation meist anschauen. Zusätzlich stellte sich die Frage, inwieweit normalhörende Probanden in der Lage sind, unterschiedliche Sprecherausrichtungen zu detektieren. Diese Fähigkeit wurde in unterschiedlichen Hörversuchen betrachtet und ausgewertet.

Die Wahrnehmung und Unterscheidungsfähigkeit der Sprecherausrichtung wurde bisher wenig untersucht. Entsprechende Signalverarbeitungsalgorithmen existieren bereits, basieren aber meist auf großen Mikrofonarrays oder räumlich verteilten Mikrofonen und sind daher ungeeignet für Hörgeräte. Für die Entwicklung neuer Ansätze war es erforderlich, die Einflüsse der relativen Sprecherausrichtung auf die empfangenen Sensorsignale zu untersuchen, z.B. binaurale Eigenschaften und der Einfluss auf den Direktschall und die frühen Reflexionen.

Drei unterschiedliche Algorithmen werden vorgestellt und evaluiert. Der erste Algorithmus nutzt die durch die Kopfabschattung entstehenden Klangverfärbungen des Nutzsignals. Die anderen Algorithmen nutzen das Verhältnis zwischen Direktschallanteile und frühen Reflexionen mit Hilfe des globalen Kohärenzfelds bzw. dem LPC-Restsignal. Es zeigt sich, dass die Sprecherausrichtung mithilfe des LPC-Restsignals am besten unterschieden werden kann.

## Abstract:

This thesis addresses the ability to distinguish and discriminate the facing angle of another speaker for normal hearing subjects and signal processing algorithms. The motivation is the development of algorithms for binaural hearing aids which are able to determine the main speaker in multiple-speakers situations. This additional information should lead to an improvement of the time delay estimation. One relevant property of a main speaker is his facing angle relative to the receiver. The facing angle discrimination of normal hearing subjects was evaluated in experiments.

So far, few investigations cover this topic. Existing signal processing algorithms usually employ arrays with a large number of microphones or spatial distributions not suitable for hearing aids. New approaches were developed by evaluation of the influence of the facing angle on the received sensor signals. In order to do so, binaural cues and the influence of direct sound and early reflections were investigated. Three different algorithms are presented and evaluated. The first algorithm makes use of differences in the frequency domain resulting from the head shadow effect. The other algorithms use the ratio between direct sound and early reflections by means of the global coherence field or the LPC-residual respectively. It is shown that the best discrimination of different facing angles can be achieved using the LPC-residual.

**Keywords:** relative Sprecherausrichtung, Hauptsprecherdetektion, Richtungsschätzung, binaurale Hörgeräte, Cocktail-Party Effekt, Mikrofonarrays, lineare Prädiktion, globales Kohärenzfeld

# Inhaltsverzeichnis

<b>Einleitung</b>	<b>1</b>
<b>Problembeschreibung und Lösungsansätze für die Haupt sprecherde tektion</b>	<b>3</b>
<b>Schwerpunkte und Gliederung dieser Arbeit</b>	<b>6</b>
<b>1 Grundlagen zur Raumakustik, der psychoakustischen Wahrnehmung von Quellen im Raum und der verwendeten Signalverarbeitung</b>	<b>8</b>
1.1 Grundlagen der Raumakustik . . . . .	8
1.2 Räumliches Hören: interaurale Pegel- und Laufzeitdifferenzen, Kopf-übertragungsfunktion . . . . .	10
1.2.1 Binaurale Effekte: interaurale Pegel- und Laufzeitdifferenzen . . . . .	11
1.2.2 Monaurale Effekte: Cone of Confusion und der Effekt der Kopfübertragungsfunktion . . . . .	12
1.3 Signalmodell mehrkanaliger Mikrofone . . . . .	13
1.4 Lineare Prädiktion . . . . .	15
1.5 Algorithmen zur Richtungsschätzung und zur Schätzung der Sprecherausrichtung . . . . .	16
1.5.1 Schätzung der Sprecherposition mittels allgemeiner Kreuzkorrelation mit Phasentransformation . . . . .	17
1.5.2 Schätzung der Sprecherposition unter Verwendung des globalen Kohärenzfeldes . . . . .	19
1.5.3 Schätzung der Sprecherausrichtung unter Verwendung des <i>oriented global coherence field</i> . . . . .	20
<b>2 Wahrnehmung der relativen Sprecherausrichtung bei normalhörenden Probanden und Voruntersuchungen für die implementierten Algorithmen</b>	<b>22</b>
2.1 Begriffsdefinition <i>relative Sprecherausrichtung</i> . . . . .	23
2.2 Evaluation der Sprecherausrichtung in der Literatur . . . . .	23

2.2.1	Untersuchung der menschlichen Fähigkeit zur Bestimmung der Specherausrichtung im reflexionsarmen Raum bei stationärer und rotierender Sprecherausrichtung (nach Kato et al.) . . . . .	24
2.2.2	Untersuchung der wahrnehmbaren Specherausrichtung bei stationärer und rotierender Sprecherausrichtung (nach Neuhoff) . . . . .	26
2.2.3	Untersuchung zur minimal wahrnehmbaren Quellausrichtung (nach Neuhoff et al.) . . . . .	29
2.3	Richtungsabhängige Abstrahlcharakteristik des menschlichen Kopfes und Vergleich mit der Abstrahlcharakteristik der Lautsprecherbox Fostex 6301B . . . . .	31
2.4	Einfluss der Sprecherausrichtung auf die Raumimpulsantwort	35
2.5	Akustische Parameter zur Schätzung der Sprecherausrichtung	41
<b>3</b>	<b>Analyse der menschlichen Fähigkeit zur Schätzung der Sprecherausrichtung mit normalhörenden Probanden</b>	<b>45</b>
3.1	Messaufbau und erstellte Audiodatenbank . . . . .	47
3.2	Experimente und Durchführung . . . . .	52
3.2.1	Ja/Nein Experimente . . . . .	52
3.2.1.1	Experimente 1 und 9: Unterscheidung der Sprecherausrichtung im reflexionsarmen Raum ( $0^\circ$ vs. $180^\circ$ ) . . . . .	53
3.2.1.2	Experimente 2 und 10: Unterscheidung der Sprecherausrichtung im Büro ( $0^\circ$ vs. $180^\circ$ ) .	53
3.2.1.3	Experimente 3 und 11: Unterscheidung der Sprecherausrichtung im reflexionsarmen Raum ( $0^\circ$ vs. $90^\circ$ ) . . . . .	54
3.2.1.4	Experimente 4 und 12: Unterscheidung der Sprecherausrichtung im Büro ( $0^\circ$ vs. $90^\circ$ ) .	54
3.2.2	Zuordnungsexperimente . . . . .	54
3.2.2.1	Experiment 5: Monotische Schätzung der Sprecherausrichtung mit Sprachsignalen . . . .	55
3.2.2.2	Experiment 6: Monotische Schätzung der Sprecherausrichtung mit LPC-Residual-Signalen .	56
3.2.2.3	Experiment 7: Dichotische Schätzung der Sprecherausrichtung mit Sprachsignalen . . . .	56
3.2.2.4	Experiment 8: Dichotische Schätzung der Sprecherausrichtung mit LPC-Residual-Signalen .	56
3.3	Versuchsergebnisse: Schätzung der Sprecherausrichtung mit normalhörenden Probanden . . . . .	57
3.3.1	Ja/Nein Experimente . . . . .	57

3.3.1.1	Experiment 1 und 9: Unterscheidung der Sprecherausrichtung im reflexionsarmen Raum (0 ° vs. 180 °) . . . . .	58
3.3.1.2	Experimente 2 und 10: Unterscheidung der Sprecherausrichtung im Büro (0 ° vs. 180 °) . .	59
3.3.1.3	Experimente 3 und 11: Unterscheidung der Sprecherausrichtung im reflexionsarmen Raum (0 ° vs. 90 °) . . . . .	60
3.3.1.4	Experimente 4 und 12: Unterscheidung der Sprecherausrichtung im Büro (0 ° vs. 90 °) . .	61
3.3.2	Zuordnungsexperimente . . . . .	62
3.3.2.1	Experiment 5: Monotische Schätzung der Sprecherausrichtung mit Sprachsignalen . . . . .	64
3.3.2.2	Experiment 6: Monotische Schätzung der Sprecherausrichtung mit LPC-Residual-Signalen .	66
3.3.2.3	Experiment 7: Dichotische Schätzung der Sprecherausrichtung mit Sprachsignalen und mit synthetischen Signalen . . . . .	68
3.3.2.4	Experiment 8: Dichotische Schätzung der Sprecherausrichtung mit LPC-Residual-Signalen .	72
3.3.3	Erkenntnisse aus den Messergebnissen der durchgeführten Experimente zur Bewertung der Sprecherausrichtung mit normalhörenden Probanden . . . . .	73
<b>4</b>	<b>Analyse der algorithmischen Fähigkeit zur Schätzung der Sprecherausrichtung</b>	<b>76</b>
4.1	Entwickelte Signalverarbeitungsalgorithmen zur Schätzung der Sprecherausrichtung . . . . .	76
4.1.1	Algorithmus zur Schätzung der Sprecherausrichtung unter Verwendung der mittlere Steigung des Frequenzgangs . . . . .	77
4.1.2	Algorithmus zur Schätzung der Sprecherausrichtung unter Verwendung des globalen Kohärenzfeldes . . . .	79
4.1.2.1	Laufzeitausgleich zur Verwendung des globalen Kohärenzfeldes mit binauralen Hörgeräten	79
4.1.2.2	Datenreduktion des globalen Kohärenzfeldes für die Schätzung der Sprecherausrichtung mit binauralen Hörgeräten . . . . .	82
4.1.2.3	Algorithmus zur Schätzung der Sprecherausrichtung unter Verwendung des modifizierten globalen Kohärenzfeldes . . . . .	83
4.1.3	Algorithmus zur Schätzung der Sprecherausrichtung unter Verwendung der linearen Prädiktion . . . . .	86

4.2	Versuchsdurchführung: Schätzung der Sprecherausrichtung mit Signalverarbeitungsalgorithmen . . . . .	89
4.2.1	Bewertung der berechneten Maße für die Ja/Nein Experimente . . . . .	90
4.2.2	Bewertung der berechneten Maße für die Zuordnungs- experimente . . . . .	91
4.3	Versuchsergebnisse: Schätzung der Sprecherausrichtung mit Signalverarbeitungsalgorithmen . . . . .	92
4.3.1	Ja/Nein Experimente . . . . .	92
4.3.1.1	Experiment 1: Unterscheidung der Kopfaus- richtung im reflexionsarmen Raum (0 ° vs. 180 °) . . . . .	92
4.3.1.2	Experiment 2: Unterscheidung der Kopfaus- richtung im Büro (0 ° vs. 180 °) . . . . .	94
4.3.1.3	Experiment 3: Unterscheidung der Kopfaus- richtung im reflexionsarmen Raum (0 ° vs. 90 °) . . . . .	94
4.3.1.4	Experiment 4: Unterscheidung der Kopfaus- richtung im Büro (0 ° vs. 90 °) . . . . .	96
4.3.2	Zuordnungsexperimente . . . . .	97
4.3.3	Erkenntnisse aus den Messergebnissen der durchgeführten Experimente zur algorithmischen Bewertung der Spre- cherausrichtung . . . . .	100
<b>5</b>	<b>Zusammenfassung, Erkenntnisse und Ausblick</b>	<b>102</b>
5.1	Zusammenfassung . . . . .	102
5.2	Erkenntnisse . . . . .	104
5.3	Ausblick . . . . .	104
<b>A</b>	<b>Anhang</b>	<b>106</b>
A.1	Grafische Benutzer Oberflächen der durchgeführten psychoaku- stische Experimente zur Schätzung der Sprecherausrichtung mit normalhörenden Probanden . . . . .	106
A.2	Ergebnisse der Hörtests mit normalhörenden Probanden . . . . .	108
A.2.1	Experiment 5: Monotische Schätzung der Kopfaus- richtung mit Sprachsignalen . . . . .	108
A.2.2	Experiment 6: Monotische Schätzung der Kopfaus- richtung mit LPC-Residual-Signalen . . . . .	110
A.2.3	Experiment 7: Dichotische Schätzung der Kopfaus- richtung mit Sprachsignalen . . . . .	111
A.2.4	Experiment 8: Dichotische Schätzung der Kopfaus- richtung mit LPC-Residual-Signalen . . . . .	114

A.3	Ergebnisse der Algorithmen-Evaluation zur Schätzung der Sprecherausrichtung unter Verwendung der mittlere Steigung des Frequenzgangs . . . . .	115
A.4	Ergebnisse der Algorithmen-Evaluation zur Schätzung der Sprecherausrichtung unter Verwendung des globalen Kohärenzfeldes . . . . .	116
A.5	Ergebnisse der Algorithmen-Evaluation zur Schätzung der Sprecherausrichtung unter Verwendung der linearen Prädiktion (Kurtosis $P^K$ ) . . . . .	117
A.5.1	Experiment 1: Unterscheidung der Kopfausrichtung im reflexionsarmen Raum ( $0^\circ$ vs. $180^\circ$ ) . . . . .	118
A.5.2	Experiment 2: Unterscheidung der Kopfausrichtung im Büro ( $0^\circ$ vs. $180^\circ$ ) . . . . .	118
A.5.3	Experiment 3: Unterscheidung der Kopfausrichtung im reflexionsarmen Raum ( $0^\circ$ vs. $90^\circ$ ) . . . . .	119
A.5.4	Experiment 4: Unterscheidung der Kopfausrichtung im Büro ( $0^\circ$ vs. $90^\circ$ ) . . . . .	119
A.5.5	Experiment 7: Dichotische Schätzung der Kopfausrichtung mit Sprachsignalen . . . . .	120
A.6	Ergebnisse der Algorithmen-Evaluation zur Schätzung der Sprecherausrichtung unter Verwendung der linearen Prädiktion (Crest Faktor $P^C$ ) . . . . .	121
A.6.1	Experiment 1: Unterscheidung der Kopfausrichtung im reflexionsarmen Raum ( $0^\circ$ vs. $180^\circ$ ) . . . . .	122
A.6.2	Experiment 2: Unterscheidung der Kopfausrichtung im Büro ( $0^\circ$ vs. $180^\circ$ ) . . . . .	122
A.6.3	Experiment 3: Unterscheidung der Kopfausrichtung im reflexionsarmen Raum ( $0^\circ$ vs. $90^\circ$ ) . . . . .	123
A.6.4	Experiment 4: Unterscheidung der Kopfausrichtung im Büro ( $0^\circ$ vs. $90^\circ$ ) . . . . .	123
A.6.5	Experiment 7: Dichotische Schätzung der Kopfausrichtung mit Sprachsignalen . . . . .	124
A.7	Ergebnisse der Algorithmen-Evaluation zur Schätzung der Sprecherausrichtung unter Verwendung der linearen Prädiktion (Peakigkeit $P^P$ ) . . . . .	125
<b>B</b>	<b>Abkürzungen</b>	<b>126</b>
	<b>Literaturverzeichnis</b>	<b>132</b>
	<b>Danksagung</b>	<b>132</b>

# Abbildungsverzeichnis

1	Schematische Skizzen zur Sprachverständlichkeitsverbesserung	4
2	Schematische Darstellung der verwendeten Hörgeräteempfangsapparatur Siemens <i>Triano 3</i> . . . . .	6
1.1	Darstellung einer künstlich erzeugten Raumimpulsantwort .	9
1.2	Qualitative Beschreibung der Spiegelquellenmethode . . . . .	10
1.3	Schematische Darstellung zur der interauralen Laufzeitdifferenz . . . . .	11
1.4	Schematische Darstellung zum Effekt des <i>Cone Of Confusion</i> .	13
1.5	Signalmodell zur Funktionsbeschreibung von mehrkanaligen Mikrofon-Systemen . . . . .	14
1.6	Beispiel eines zweidimensionalen globalen Kohärenzfeldes .	20
1.7	Grafische Darstellung zur Schätzung der Sprecherausrichtung unter Verwendung eines zweidimensionalen <i>oriented global coherence fields</i> . . . . .	21
2.1	Schematische Darstellung der Sprecherausrichtung $\varphi$ . . . . .	23
2.2	Durchschnittliche Fehlerraten zur Schätzung der Sprecherausrichtung eines männlichen Sprechers . . . . .	26
2.3	Ergebnisse zur Untersuchung der wahrnehmbaren Sprecherausrichtung bei stationärer und rotierender Quelle . . . . .	28
2.4	Richtdiagramme der Schallabstrahlcharakteristik der Lautsprecherbox <i>Radio Shack Optimus XTS 40</i> . . . . .	29
2.5	Versuchsergebnisse der Bestimmung der minimal wahrnehmbaren Quellenausrichtung . . . . .	30
2.6	Richtungsabhängige Abstrahlcharakteristiken in Terzbändern von 20 männlichen Sprechern und der Lautsprecherbox <i>Fostex 6301B</i> . . . . .	32
2.7	Vergleich der Abstrahlcharakteristik und der Sprachspektren eines männlichen Sprechers für mittlere, laute und leise Sprache	33
2.8	Zweidimensionale Simulation der Schallausbreitung im Freifeld in Abhängigkeit der Sprecherausrichtung $\varphi$ bei einer Frequenz $f = 1000$ Hz . . . . .	37

2.9 Zweidimensionale Simulation der Schallausbreitung im Freifeld in Abhangigkeit der Sprecherausrichtung $\varphi$ bei einer Frequenz $f = 8000$ Hz . . . . .	37
2.10 Zweidimensionale Simulation der Schallausbreitung in einem halligen Raum in Abhangigkeit der Sprecherausrichtung $\varphi$ bei einer Frequenz $f = 1000$ Hz . . . . .	39
2.11 Zweidimensionale Simulation der Schallausbreitung in einem halligen Raum in Abhangigkeit der Sprecherausrichtung $\varphi$ bei einer Frequenz $f = 8000$ Hz . . . . .	39
2.12 Darstellung der ersten 10 ms der gemessenen Raumimpulsantworten in einem Bororaum in Abhangigkeit der Sprecherausrichtungen $\varphi$ . . . . .	40
3.1 Schematische Skizze des Messaufbaus zur Aufnahme des verwendeten Audiomaterials . . . . .	47
3.2 Mastabsgerechte Skizze des genutzten reflexionsarmen Raumes . . . . .	48
3.3 Mastabsgerechte Skizze des genutzten Bororaums . . . . .	49
3.4 Darstellungen der Ausrichtungen von Sprecher/Lautsprecherbox und Kunstkopf wahrend den Aufnahmen, welche fur die Durchführung der Experimente und der Evaluation der entwickelten Algorithmen verwendet wurden. . . . .	50
3.5 Beispiele fur Verwechslungsmatrizen zur besseren Deutung der im Anhang dargestellten Ergebnisse. Auf der x-Achse befinden sich die wahren Sprecherausrichtungen, auf der y-Achse die von den Probanden oder Algorithmen zugeordneten Ausrichtungen. . . . .	64
3.6 Ergebnisse Experimente 5: Monotische Schatzung der Sprecherausrichtung mit Sprachsignalen von normalhorenden Probanden (1) . . . . .	65
3.7 Ergebnisse Experimente 5: Monotische Schatzung der Sprecherausrichtung mit Sprachsignalen von normalhorenden Probanden (2) . . . . .	66
3.8 Ergebnisse Experimente 6: Monotische Schatzung der Sprecherausrichtung mit LPC-Residual-Signalen von normalhorenden Probanden (1) . . . . .	67
3.9 Ergebnisse Experimente 6: Monotische Schatzung der Sprecherausrichtung mit LPC-Residual-Signalen von normalhorenden Probanden (2) . . . . .	68
3.10 Ergebnisse Experimente 7: Dichotische Schatzung der Sprecherausrichtung mit Sprachsignalen von normalhorenden Probanden (1) . . . . .	69
3.11 Ergebnisse Experimente 7: Dichotische Schatzung der Sprecherausrichtung mit Sprachsignalen von normalhorenden Probanden (2) . . . . .	70

3.12	Ergebnisse Experiment 7: Dichotische Schätzung der Sprecherausrichtung mit synthetischen Signalen von normalhörenden Probanden . . . . .	72
3.13	Ergebnisse Experimente 8: Dichotische Schätzung der Sprecherausrichtung mit LPC-Residual-Signalen von normalhörenden Probanden (1) . . . . .	73
3.14	Ergebnisse Experimente 8: Dichotische Schätzung der Sprecherausrichtung mit LPC-Residual-Signalen von normalhörenden Probanden (2) . . . . .	74
4.1	Blockschaltbild zur Veranschaulichung des Algorithmus zur Schätzung der Sprecherausrichtung mittels gemittelten Sprachspektrums . . . . .	78
4.2	Beispielhafte Darstellungen des globalen Kohärenzfeldes . . . . .	80
4.3	Beispielhafte Darstellungen des modifizierten globalen Kohärenzfeldes . . . . .	80
4.4	Darstellung der Geometrie für die Berechnung der Laufzeit an einem idealen Kugelkopfmodell . . . . .	82
4.5	Veranschaulichung der Datenreduktion des mGCF und Resultat der Mittlung des modifizierten GCF . . . . .	83
4.6	Blockschaltbild zur Veranschaulichung des Algorithmus zu Schätzung der Sprecherausrichtung mittels modifizierten globalen Kohärenzfeld . . . . .	84
4.7	Beispiel für die Berechnung des modifizierten globalen Kohärenzfeldes für unterschiedliche Sprecherausrichtungen . . . . .	85
4.8	Zeitsignal des Vokals /a/ eines männlichen Sprechers in Abhängigkeit von der Sprecherausrichtung $\varphi_s$ . . . . .	87
4.9	LPC-Residual-Signal des Vokals /a/ eines männlichen Sprecher in Abhängigkeit von der Sprecherausrichtung $\varphi_s$ . . . . .	88
4.10	Blockschaltbild zur Veranschaulichung des Algorithmus zur Schätzung der Sprecherausrichtung mittels linearer Prädiktion	89
4.11	Ergebnisse Experiment 1: Blinde Schätzung der Kopfausrichtung im reflexionsarmen Raum ( $0^\circ$ vs. $180^\circ$ ) . . . . .	91
4.12	Ergebnisse Experiment 7: Dichotische Schätzung der Kopfausrichtung mit Sprachsignalen unter Verwendung des gemittelten Sprachspektrums . . . . .	98
4.13	Ergebnisse Experiment 7: Dichotische Schätzung der Kopfausrichtung mit Sprachsignalen unter Verwendung des modifizierten globalen Kohärenzfeldes . . . . .	99
4.14	Ergebnisse Experiment 7: Dichotische Schätzung der Kopfausrichtung mit Sprachsignalen unter Verwendung der linearen Prädiktion (Peakigkeit) . . . . .	100
A.1	Screenshot der Oberfläche eines Ja/Nein Experiments . . . . .	107
A.2	Screenshot der Oberfläche eines Zuordnungsexperimentes . .	107

A.3 Verwechslungsmatrizen Experimente 5: Monotische Schätzung der Sprecherausrichtung mit Sprachsignalen von normalhörenden Probanden im reflexionsarmen Raum in Abhängigkeit des Quellsignals . . . . .	108
A.4 Verwechslungsmatrizen Experimente 5: Monotische Schätzung der Sprecherausrichtung mit Sprachsignalen von normalhörenden Probanden im halligen Büror Raum . . . . .	109
A.5 Verwechslungsmatrizen Experimente 5: Monotische Schätzung der Sprecherausrichtung mit Sprachsignalen von normalhörenden Probanden in Abhängigkeit von den Räumen und einer möglichen Hochpassfilterung . . . . .	109
A.6 Verwechslungsmatrizen Experimente 6: Monotische Schätzung der Sprecherausrichtung mit LPC-Residual-Signalen von normalhörenden Probanden im reflexionsarmen Raum in Abhängigkeit des Quellsignals . . . . .	110
A.7 Verwechslungsmatrizen Experimente 6: Monotische Schätzung der Sprecherausrichtung mit LPC-Residual-Signalen von normalhörenden Probanden im halligen Büror Raum in Abhängigkeit des Quellsignals . . . . .	110
A.8 Verwechslungsmatrizen Experimente 6: Monotische Schätzung der Sprecherausrichtung mit LPC-Residual-Signalen von normalhörenden Probanden in Abhängigkeit von den Räumen und einer möglichen Hochpassfilterung . . . . .	110
A.9 Verwechslungsmatrizen Experimente 7: Dichotische Schätzung der Sprecherausrichtung von normalhörenden Probanden im reflexionsarmen Raum in Abhängigkeit des Quellsignals . . .	111
A.10 Verwechslungsmatrizen Experimente 7: Dichotische Schätzung der Sprecherausrichtung von normalhörenden Probanden im halligen Büror Raum in Abhängigkeit des Quellsignals . . . .	112
A.11 Verwechslungsmatrizen Experimente 7: Dichotische Schätzung der Sprecherausrichtung von normalhörenden Probanden von normalhörenden Probanden in Abhängigkeit von den Räumen und einer möglichen Hochpassfilterung . . . . .	113
A.12 Ergebnisse Experimente 8: Dichotische Schätzung der Sprecherausrichtung mit LPC-Residual-Signalen von normalhörenden Probanden im reflexionsarmen Raum in Abhängigkeit des Quellsignals . . . . .	114
A.13 Ergebnisse Experimente 8: Dichotische Schätzung der Sprecherausrichtung mit LPC-Residual-Signalen von normalhörenden Probanden im halligen Büror Raum in Abhängigkeit des Quellsignals . . . . .	114
A.14 Ergebnisse Experimente 8: Dichotische Schätzung der Sprecherausrichtung mit LPC-Residual-Signalen von normalhörenden Probanden von normalhörenden Probanden in Abhängigkeit von den Räumen und einer möglichen Hochpassfilterung . . .	114

A.15 Verwechslungsmatrizen Experimente 7: Dichotische Schätzung der Sprecherausrichtung unter Verwendung der mittlere Steigung des Frequenzgangs in Abhängigkeit von den Räumen. . .	115
A.16 Verwechslungsmatrizen Experimente 7: Dichotische Schätzung der Sprecherausrichtung unter Verwendung des globalen Kohärenzfeldes in Abhängigkeit von den Räumen. . . . .	116
A.17 Verwechslungsmatrizen Experimente 7: Dichotische Schätzung der Sprecherausrichtung unter der linearen Prädiktion (Kurtosis $P^K$ ) in Abhängigkeit von den Räumen. . . . .	120
A.18 Verwechslungsmatrizen Experimente 7: Dichotische Schätzung der Sprecherausrichtung unter der linearen Prädiktion (Crest Faktor $P^C$ ) in Abhängigkeit von den Räumen. . . . .	124
A.19 Verwechslungsmatrizen Experimente 7: Dichotische Schätzung der Sprecherausrichtung unter der linearen Prädiktion (Peakigkeit $P^P$ ) in Abhängigkeit von den Räumen. . . . .	125

# Tabellenverzeichnis

2.1	Messergebnisse für Experimente zur Schätzung der Wahrnehmung unterschiedlicher Ausrichtungen eines männlichen Sprechers und normalhörenden Probanden, durchgeführt in einem reflexionsarmen Raum. . . . .	25
3.1	Ergebnisse Experimente 1 und 9: Unterscheidung der Sprecherausrichtung im reflexionsarmen Raum von normalhörenden Probanden . . . . .	59
3.2	Ergebnisse Experimente 2 und 10: Unterscheidung der Sprecherausrichtung im Büro von normalhörenden Probanden . .	60
3.3	Ergebnisse Experimente 3 und 11: Unterscheidung der Sprecherausrichtung im reflexionsarmen Raum von normalhörenden Probanden . . . . .	61
3.4	Ergebnisse Experimente 4 und 12: Unterscheidung der Sprecherausrichtung im Büro von normalhörenden Probanden . .	62
4.1	Ergebnisse Experiment 1: Unterscheidung der Sprecherausrichtung im reflexionsarmen Raum von Signalverarbeitungsalgorithmen . . . . .	93
4.2	Ergebnisse Experiment 2: Unterscheidung der Sprecherausrichtung im Büro von Signalverarbeitungsalgorithmen . . .	95
4.3	Ergebnisse Experiment 3: Unterscheidung der Sprecherausrichtung im reflexionsarmen Raum von Signalverarbeitungsalgorithmen . . . . .	96
4.4	Ergebnisse Experiment 4: Unterscheidung der Sprecherausrichtung im Büro von Signalverarbeitungsalgorithmen . . .	97
A.1	Ergebnisse Experiment 1: Unterscheidung der Sprecherausrichtung im reflexionsarmen Raum unter Verwendung der linearen Prädiktion (Kurtosis $P^K$ ) . . . . .	118
A.2	Ergebnisse Experiment 2: Unterscheidung der Sprecherausrichtung im Büro unter Verwendung der linearen Prädiktion (Kurtosis $P^K$ ) . . . . .	118
A.3	Ergebnisse Experiment 3: Unterscheidung der Sprecherausrichtung im reflexionsarmen Raum unter Verwendung der linearen Prädiktion (Kurtosis $P^K$ ) . . . . .	119

A.4 Ergebnisse Experiment 4: Unterscheidung der Sprecherausrichtung im Büro unter Verwendung der linearen Prädiktion (Kurtosis $P^K$ ) . . . . .	119
A.5 Ergebnisse Experiment 1: Unterscheidung der Sprecherausrichtung im reflexionsarmen Raum unter Verwendung der linearen Prädiktion (Crest Faktor $P^C$ ) . . . . .	122
A.6 Ergebnisse Experiment 2: Unterscheidung der Sprecherausrichtung im Büro unter Verwendung der linearen Prädiktion (Crest Faktor $P^C$ ) . . . . .	122
A.7 Ergebnisse Experiment 3: Unterscheidung der Sprecherausrichtung im reflexionsarmen Raum unter Verwendung der linearen Prädiktion (Crest Faktor $P^C$ ) . . . . .	123
A.8 Ergebnisse Experiment 4: Unterscheidung der Sprecherausrichtung im Büro unter Verwendung der linearen Prädiktion (Crest Faktor $P^C$ ) . . . . .	123

---

# Einleitung

In geräuschvollen Umgebungen besitzen Normalhörende die Fähigkeit, ihre Aufmerksamkeit gezielt auf einzelne Nutzquellen zu fokussieren. Das Gehör ist dabei in der Lage eine Geräuschunterdrückung gegenüber den Umweltgeräuschen von 9 bis 15 dB zu erreichen [Sla92], selbst wenn weitere Sprecher als Störquellen aktiv sind. Diese Fähigkeit wird als *Cocktailparty-Effekt* oder *selektives Hören* bezeichnet und ist eine Leistung des binauralen Hörens. Zusätzlich kann das Gehör die Aufmerksamkeit in Bruchteilen einer Sekunde auf andere Gespräche im Raum und wieder zurück lenken.

Die Verarbeitung des Gehörs ist bis heute nicht komplett verstanden und es existieren keine ausreichenden Modelle, die diese Fähigkeit beschreiben.

Schwerhörige und Hörgeräteträger bemängeln vor allem die Sprachverständlichkeit in geräuschvollen Umgebungen. Der Effekt des selektiven Hörens verringert sich mit steigendem Hörverlust und kann auch durch Hörgeräte nicht vollständig kompensiert werden.

Um eine Verbesserung der Sprachverständlichkeit von Schwerhörigen in geräuschvollen Umgebungen zu erreichen, existieren eine Vielzahl von Signalverarbeitungsalgorithmen. So reduzieren einkanalige Geräuschreduktionsalgorithmen mit mehreren Frequenzbändern deren Verstärkung, wenn enthaltende Störgeräusche detektiert werden. Frequenzbänder mit Sprachanteilen werden unmodifiziert übertragen. Diese Algorithmen helfen vor allem, wenn das Störgeräusch sprachunähnlich ist und in Frequenzbändern außerhalb des Hauptsprachbereichs dominant ist, wie z.B. tieffrequente Fahrzeuggeräusche oder Lüfterbrummen. In einer Cocktailparty-Situation, in der Sprache sowohl Nutz- als auch Störquellen darstellt (*Multitalker-Situation*), führen diese Algorithmen nur bedingt zu einer Verbesserung der Sprachverständlichkeit [BC06].

Verbesserungen der Sprachverständlichkeit einer Nutzquelle in einer Multitalker-Situation ermöglicht der Einsatz von Raum-Zeit-Filtern (*Richtmikrofone, Beamformern*) bei mehrkanaligen Hörgeräten. Mehrkanalige Hörgeräte besitzen zwei oder mehr Mikrofone und können Signale einer

---

Vorzugsrichtung unverändert wiedergeben, während Signale aus anderen Richtungen gedämpft werden. Somit werden Sprecher außerhalb der Vorzugsrichtung (*Hauptkeule*) gedämpft. Durch die Wahl einer Vorzugsrichtung, welche meist der Blickrichtung des Hörgeräteträgers entspricht, verbessert sich das Verhältnis zwischen dem gegenüberstehenden Nutzsprecher zu den umliegenden Störquellen sowohl bei Sprache als auch bei Geräuschquellen in geräuschvoller Umgebung.

Von Nachteil ist bei diesem Verfahren, dass sich Gesprächspartner häufig nicht frontal anschauen, die Kopfausrichtung zueinander leicht abgewandt ist. So kann die Nutzquelle außerhalb der gewünschten, frontalen Vorzugsrichtung liegen und wird daher ebenfalls gedämpft. Dem entgegenwirkend können zum einen breitere Hauptkeulen verwendet werden, die wiederum mehr unerwünschte Quellen weniger dämpfen und die Sprachverständlichkeit somit erneut reduzieren. Zum anderen können Verfahren zur Richtungsschätzung (engl. *time delay estimation, TDE*) angewandt werden, welche die Position einer räumlich stationären Quelle ermitteln und so adaptiv die Vorzugsrichtung anpassen. Ein bekanntes und gängiges Verfahren ist die *allgemeine Kreuzkorrelation mit Phasentransformation* (engl. *Generalized Cross Correlation with Phase Transform, GCC-PHAT*) [BOS05]. Eine Begrenzung der Ausrichtung auf den Blickbereich in Kombination mit einer ausreichend schmalen Hauptkeule gleicht so die zuvor beschriebenen Nachteile aus. In [Roh08] wird eine solche Kombination aus adaptiven Beamformern mit adaptiver Ausrichtung der Vorzugsrichtung vorgestellt. Um die Schätzung der Positionen auf die vorhandenen Sprecher einzuschränken ist der Einsatz von Sprachpausen-Detektoren (engl. *Speech Activity Detector, VAD*) sinnvoll und verbessert die Schätzungen.

Aus der Richtungsschätzung ergibt sich die Richtung der dominantesten räumlich stationären Quelle. Durch Einschränkung des möglichen Winkelbereichs auf das Blickfeld des Sprechers wird die Richtkeule nur auf frontale Quellen ausgerichtet. Es kann aber keine Aussage getroffen werden, ob diese Quelle die tatsächlich gewünschte Nutzquelle des Hörgeräteträgers ist. Ebenso ist der Hörgeräteträger nach wie vor aus anderen Richtungen, außerhalb der Vorzugsrichtung, schlecht ansprechbar.

Eine Lösung dieser Problematik ergibt sich, wenn die Hörgeräte in der Lage sind, akustische Eigenschaften aus den Mikrofonsignalen zu extrahieren, die für die Bestimmung eines möglichen Hauptsprechers relevant sind. So ist es denkbar, die Vorzugsrichtung auf den Sprecher zu richten, der im Sichtfeld des Hörgeräteträgers steht und zusätzlich in Richtung des Trägers blickt. Dieser Sprecher hat für den Hörgeräteträger in lärmvoller Umgebung

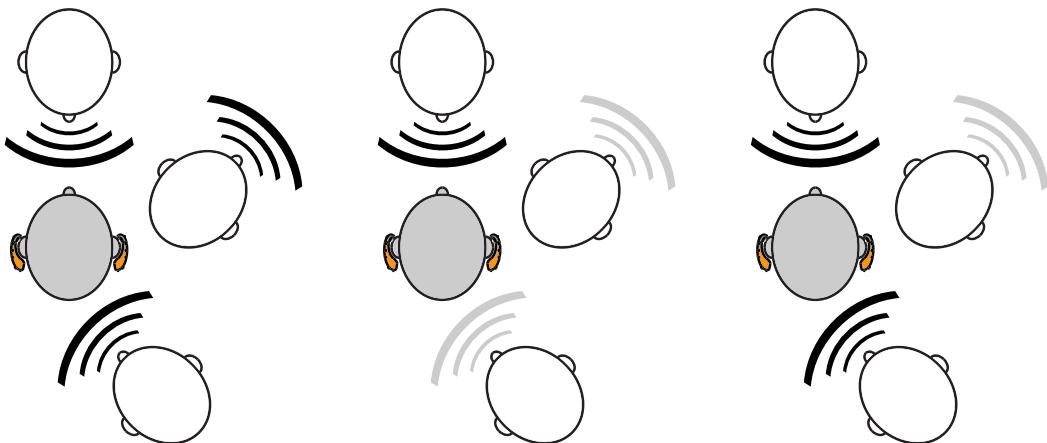
---

wahrscheinlich größere Relevanz als ein Sprecher, der seinen Blick abgewandt hat. Ebenso können Sprecher detektiert werden, die außerhalb des Sichtfeldes liegen, aber in Richtung des Trägers schauen und in ansprechen. In dieser Situation könnte das Hörgerät ebenfalls eine Richtkeule in diese Richtung ausrichten oder ein entsprechendes akustisches Signal geben, so dass der Betroffene erkennt, dass er angesprochen wird.

## **Problembeschreibung und Lösungansätze für die Haupt sprecherdetektion**

In einer lärmvollen Umgebung oder in einer Multisprechersituation werden die Quellen von konventionellen Hörgeräten mit omnidirektionalen Mikrofonen richtungsunabhängig übertragen. Somit wird ein frontal Sprecher genauso von den Hörgeräten übertragen, wie ein seitlicher Sprecher. Aufgrund des verschlechterten Cocktailparty-Effekts eines Schwerhörigen, verschlechtert sich in diesen Situationen die Sprachverständlichkeit. Abbildung 1(a) zeigt beispielhaft eine Multisprechersituation mit einem Hörgeräteträger im Zentrum und drei Sprechern um diesen herum. Der frontale Sprecher blickt in Richtung des Hörgeräteträgers und stellt den direkten Gesprächspartner dar. Der vom Hörer abgewandte Sprecher kann als Störquelle angesehen werden, da er nicht in Richtung des Hörgeräteträgers blickt. Weiterhin wird angenommen, dass der dritte Sprecher außerhalb des Gesichtsfeldes den Hörgeräteträger anspricht, um seine Aufmerksamkeit zu erhalten. Die Abbildung soll verdeutlichen, dass alle Quellen gleichberechtigt von den Hörgeräten aufgenommen und verstärkt werden.

Konventionelle Hörgeräte mit direkionalen Mikrofonen bieten die Möglichkeit, Signale aus einer Vorzugsrichtung unverändert zu übertragen, während Signale aus den anderen Richtungen gedämpft werden. Meist wird die Vorzugsrichtung in Blickrichtung des Hörgeräteträgers gewählt. Dies würde den Signal-Rauschabstand zwischen einer frontalen Quelle und den räumlich getrennten Quellen verbessern. Schematisch ist eine solche Situation in Abbildung 1(b) dargestellt. Während der frontale Sprecher besser verstanden werden kann, werden die Signale der anderen Sprecher gedämpft. Allerdings ist der Hörer durch die Dämpfung des Schalls außerhalb der Vorzugsrichtung deutlich schlechter ansprechbar. In diesem Beispiel könnte der rückseitige Sprecher nur schwer die Aufmerksamkeit des Hörers erlangen. Weiterhin muss der frontale Sprecher direkt in der Vorzugsrichtung



- (a) Multisprechersituation mit omnidirektionalen Hörgerätemikrofonen. Die Quellen werden von den Hörgeräten ohne Richtungsfilterung übertragen. Durch das eingeschränkte selektive Hören des Hörgeräteträgers fällt es schwer den Gesprächen der zugewandten Sprechern zu folgen.
- (b) Multisprechersituation mit direktonalen Hörgerätemikrofonen. Die frontale Quelle steht in der Vorzugsrichtung der Richtmikrofone und kann vom Hörgeräteträger besser verstanden werden, während die räumlich getrennten Quellen gedämpft werden. Beim eingeschränkten selektiven Hören verspricht dies eine Verbesserung der Sprachverständlichkeit des frontalen Sprechers. Der Hörgeräteträger ist aus anderen Richtungen als der Vorzugsrichtung vermindert ansprechbar.
- (c) Multisprechersituation mit direktonalen Hörgerätemikrofonen und Hauptsprecherdetektion. Die Hörgeräte klassifizieren die zugewandten Sprecher als Hauptsprecher, so dass Richtkeulen in die entsprechenden Richtungen ausgerichtet werden können. Abgewandte Sprecher werden gedämpft, was eine Verbesserung der Sprachverständlichkeit verspricht. Zusätzlich ist der Sprecher aus anderen Richtungen ansprechbar.

**Abb. 1:** Schematische Skizzen zur Sprachverständlchkeitsverbesserung in einer Multisprechersituation mit unterschiedlichen Hörgerätemikrofonen und Hauptsprecherdetektion.

der Hörgeräte stehen. Leichte Abweichungen oder Kopfbewegungen des Hörgeräteträgers können bereits zu einer Dämpfung dieser Quelle führen. Häufig kann zudem beobachtet werden, dass ein Hörer sein gegenüber nicht direkt frontal anschaut. Der Kopf ist leicht vom Sprecher abgewandt, während die Augen in die Richtung des Sprechers blicken. Verfahren zur Richtungsschätzung der Quellen können hier die Vorzugsrichtung adaptiv an die jeweilige Situation anpassen. Die Hörgeräte würden stets eine dominante Quelle wählen und die Richtkeule der Mikrofone entsprechend ausrichten. Problematisch kann hier die Auswahl der Quellen sein. Es ist denkbar, dass der abgewandte Sprecher als Hauptsprecher detektiert wird, wenn dieser z.B. sehr laut spricht. Die Hörgeräte würden so eine ungewollte Quelle verstärken und die gewünschten Nutzquellen dämpfen.

---

Eine Schätzung der Relevanz für die Hörer könnte die Situation für den Hörgeräteträger verbessern. Die Hörgeräte müssten, zusätzlich zur relativen Position der Quellen, akustische Eigenschaften der Quellen extrahieren, welche diese als relevante Quellen und somit als Hauptsprecher für den Hörer klassifizieren. Abbildung 1(c) zeigt schematisch das Resultat eines solchen Verfahrens. Der frontale Sprecher und der rückseitige Sprecher werden als Hauptsprecher detektiert und in Kombination mit einer Richtungsschätzung können die Richtkeulen auf diese Quellen ausgerichtet werden. Zusätzlich wird der abgewandte Sprecher als Hauptsprecher verworfen und gedämpft. Der Hörgeräteträger nimmt die gewünschte, frontale Quelle wahr und ist in dieser Situation von hinten ansprechbar. Zum aktuellen Zeitpunkt sind keine Verfahren für die Hauptsprecherdetektion mit Hörgeräten bekannt.

Wie an den Abbildungen 1(a) bis 1(c) beispielhaft gezeigt, erscheint die Schätzung der Eigenschaften zur Detektion eines Hauptsprechers mit Hörgeräten eine sinnvolle Erweiterung für die Richtungsschätzung. Die folgenden Eigenschaften sind mögliche Ansätze zur Lösung der beschriebenen Problematik.

- **Schätzung des Abstands zwischen Hörer und Sprecher:** Je näher der Sprecher am Hörgeräteträger steht, umso wahrscheinlicher ist es, dass es sich um einen Hauptsprecher handelt. Es ist zu vermuten, dass ein Sprecher im direkten Gespräch selten weit entfernt steht.
- **Schätzung der Blickrichtung des Sprechers relativ zum Hörgeräteträger:** Je direkter der Sprecher in Richtung des Hörgeräteträgers blickt, umso wahrscheinlicher ist es, dass dieser den Träger anspricht. Abgewandte Sprecher können so als irrelevante Quellen angesehen werden.
- **Detektion der Nennung des Namens des Hörgeräteträgers:** Für den Fall, dass der Hörgeräteträger nicht auf den Sprecher reagiert, da dieser z. B. außerhalb des Blickfeldes des Trägers steht, wird der Sprecher möglicherweise den Namen des Hörgeräteträgers aussprechen, um seine Aufmerksamkeit zu erhalten. Eine Detektion des Namens kann somit auf eine gewünschte Kontaktaufnahme mit dem Hörgeräteträger hindeuten.

Zur Erkennung des Namens müssten Spracherkenner in den Hörgeräten den Namen in unterschiedlichen Ausprägungen, z. B. Vor-, Nach- und Spitznamen, detektieren. Aufgrund der begrenzten Rechenkapazität erscheint eine Spracherkennung beim heutigen Stand der Technik nicht umsetzbar. Es ist

---

zu erwarten, dass die Kombination der vorgestellten Eigenschaften die Erkennungsleistung möglicher Algorithmen verbessert.

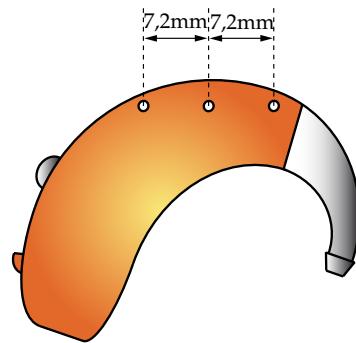
Eine sinnvolle umsetzbare Lösung erscheint die Kombination aus der Abstands- und Blickrichtungsschätzung. Da die Lösungen für beide Einzelprobleme nicht trivial sind, beschränkt sich die vorliegende Arbeit mit der Schätzung der relativen Sprecherausrichtung bezogen auf den Hörgeräteträger.

## Schwerpunkte und Gliederung dieser Arbeit

Im Rahmen dieser Arbeit wurden Verfahren zur Bestimmung von einem oder mehreren Hauptsprechern untersucht und entwickelt. Die vorgestellten Algorithmen sind in der Programmiersprache *MATLAB* implementiert und evaluiert.

Bei der Entwicklung der Algorithmen wurden zwei mehrkanalige Hörgeräte vorausgesetzt, wobei eine binaurale Verschaltung der Hörgeräte als gegeben angenommen wird. Im konkreten Fall der Implementation und Evaluierung standen zwei Siemens *Triano 3* Hörgerätedummies zur Verfügung, mit jeweils 3 Mikrofonen, angeordnet in einer Reihe (*line array*) im Abstand von jeweils 7,2 mm. Abbildung 2 zeigt eine schematische Darstellung der verwendeten Geräte. Die Dummies bestehen aus dem Hörgerätegehäuse, den Mikrofonen und einem Hörer. Die Mikrofonsignale stehen über einem externen Mikrofonvorverstärker und einer Soundkarte zur Verarbeitung zur Verfügung.

Die binaurale Verschaltung von Hörgeräten ermöglicht es, ein oder mehrere Mikrofonsignale des kontralateralen Hörgerätes bei den Hörgerätealgorithmen mit einzubeziehen. Im Fall von Richtmikrofonen verspricht dies eine bessere Geräuschunterdrückung, da die Algorithmen durch den höheren Abstand (Kopfdurchmesser) mehr Informationen des vorliegenden Geräuschfeldes erhalten. Die binaurale Verschaltung ist derzeit in Ansätzen in kommerziellen Hörgeräten vorhanden. Ein kompletter



**Abb. 2:** Schematische Darstellung der für die Experimente und Entwicklung verwendete Hörgeräteempfangsapparatur Siemens *Triano 3*. Gerät besitzt drei Hörgerätemikrofone mit seitlicher Schalleintrittsöffnung im Abstand von 7,2 mm.

---

Austausch der Mikrofonsignale von einem Ohr zum Gegenohr (in beiden Richtungen) findet aufgrund der hohen Übertragungsrate und des hohen Stromverbrauchs derzeit nicht statt. Es ist zu erwarten, dass eine komplette binaurale Verschaltung in künftigen Hörgerätegenerationen zur Verfügung steht. Bei der Implementation und Evaluation wird die binaurale Verschaltung durch die Verarbeitung aufgenommener Mikrofonsignale im Rechner ermöglicht.

Eine Literaturrecherche hat gezeigt, dass eine Vielzahl von Untersuchungen hinsichtlich der Lokalisationsleistung des auditorischen Systems durchgeführt wurde, z.B. [Bla97], [CG91], [Mil58] und [Mil60]. Ebenso existieren einige Untersuchungen zur Abstrahlcharakteristik der menschlichen Sprache und die Abschattungswirkung des menschlichen Kopfes, z.B. [DF39], [MP78], [McK86] und [CW02]. Weniger untersucht ist bisher die Fähigkeit des auditorischen Systems in Hinblick auf die Fähigkeit, die Kopfausrichtung eines Sprechers zu detektieren. Aus diesem Grund wurden parallel zur Algorithmenentwicklung Hörversuche durchgeführt, die diese Fähigkeit bei normalhörenden Probanden untersuchen. Für diesen Test wurden mehrere Sprecher aus unterschiedlichen Richtungen und Kopfausrichtungen mit den zur Verfügung stehenden Hörgerätemikrofonen aufgenommen, von normalhörenden Probanden bewertet und ausgewertet. Die somit zur Verfügung stehenden Aufnahmen dienten weiterhin als Testsignale für die durchgeführte Evaluation der entwickelten Algorithmen.

Die Arbeit ist wie folgt gegliedert:

- **Kapitel 1:** Beschreibung der Grundlagen zur Raumakustik, der psychoakustischen Wahrnehmung von Quellen im Raum und der verwendeten Signalverarbeitung.
- **Kapitel 2:** Wahrnehmung der relativen Sprecherausrichtung bei normalhörenden Probanden und Voruntersuchungen für die implementierten Algorithmen anhand von Beispielen aus der Literatur und Simulationen.
- **Kapitel 3:** Analyse der menschlichen Fähigkeit zur Schätzung der Sprecherausrichtung mit normalhörenden Probanden mithilfe von Hörversuchen.
- **Kapitel 4:** Analyse der algorithmischen Fähigkeit zur Schätzung der relativen Sprecherausrichtung von drei entwickelten und implementierten mehrkanaligen Hörgerätealgorithmen.
- **Kapitel 5:** Zusammenfassung, Erkenntnisse und Ausblick.

---

# Kapitel 1

## Grundlagen zur Raumakustik, der psychoakustischen Wahrnehmung von Quellen im Raum und der verwendeten Signalverarbeitung

In diesem Kapitel werden die Grundlagen der Raumakustik anhand der Raumimpulsantwort beschrieben und ein Modell zur Beschreibung dieser Impulsantworten vorgestellt. Darauf folgen psychoakustische Modelle, die für das selektive Hören und der akustischen Quellortung relevant sind. Diese Modelle werden im folgenden Kapitel genutzt um die akustisch wahrnehmbaren Effekte in Abhängigkeit von der Sprecherausrichtung zu untersuchen und zu beschreiben. Anschließend folgen die Grundlagen der verwendeten Signalverarbeitung die für die entwickelten Hörgerätealgorithmen in Kapitel 4 von Bedeutung sind. Unter anderem werden hier gängige Verfahren zu Richtungs- und Sprecherausrichtungsschätzung beschrieben.

### 1.1 Grundlagen der Raumakustik

„Die Raumimpulsantwort ist die akustische Reaktion eines Raumes auf einen sehr kurzen Schallimpuls“ [Bar08] und enthält alle Informationen der Schallausbreitung zwischen einer Quelle und einem räumlich getrennten Empfängerort. Die Raumimpulsantwort (RIA) kann in drei Bereiche eingeteilt werden:

- **Direktschall:** Der Direktschallanteil ergibt sich aus der direkten, ungestörten Schallausbreitung zwischen der Quelle und dem Empfänger.

# Kapitel 1. Grundlagen zur Raumakustik, der psychoakustischen Wahrnehmung von Quellen im Raum und der verwendeten Signalverarbeitung

---

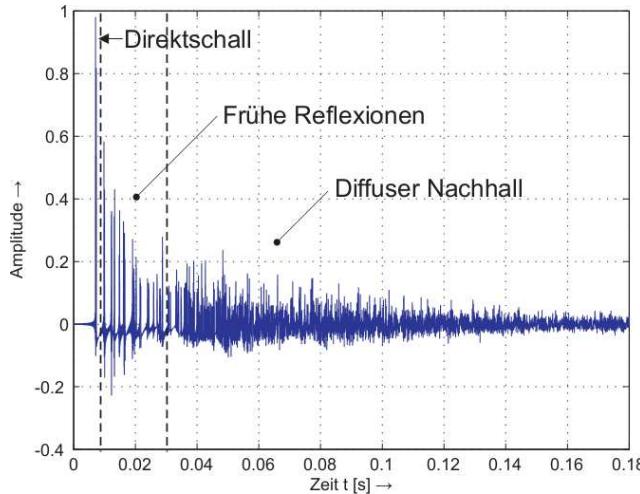


Abb. 1.1: Darstellung einer künstlich erzeugten Raumimpulsantwort (aus [Bar08])

Er trifft als erste Antwort beim Empfänger ein und kann als größter Peak direkt am Beginn der Impulsantwort identifiziert werden.

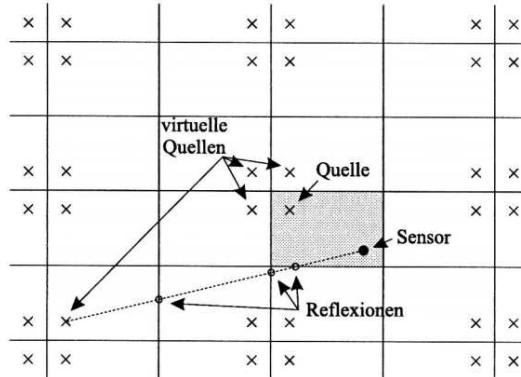
- **frühe Reflexionen:** Die frühen Reflexionen entstehen durch die Schallausbreitung von der Quelle über Wänden und Gegenständen zum Empfängerort. Hier wird die Welle in Abhängigkeit vom Material der Hindernisse gedämpft und reflektiert. Aufgrund der längeren zurückgelegten Wegstrecke erreichen die frühen Reflexionen den Empfänger zeitverzögert. Sie können aber noch deutlich als einzelne Peaks in der RIA detektiert werden (vgl. [Bit02]).
- **Diffuser Nachhall:** Der diffuse Nachhall entsteht durch Reflexionen höherer Ordnung. Die Schallwellen werden immer wieder von den Wände und Gegenstände reflektiert und so „nimmt die Reflexionsdichte mit der Zeit zu, gleichzeitig nimmt die Energie der einzelnen Impulse ab“ [Bit02]. Einzelimpulse sind im diffusen Nachhall der RIA nicht mehr deutlich erkennbar.

In Abbildung 1.1 ist eine künstlich erstellte Raumimpulsantwort zur Veranschaulichung dargestellt. Die drei vorgestellten Bereiche sind markiert und deutlich detektierbar.

Neben der messtechnischen Bestimmung der RIA, z. B. mit einer Impulsquelle, Sinussweeps oder weißem Rauschen, existieren drei Ansätze, die Schallausbreitung in Räumen zu beschreiben: ein wellentheoretisches Modell auf Grundlage der Wellengleichung (z.B. in [Mar95]), ein geometrisches Modell basierend auf den Reflexionen an Flächen und ein statistisches Modell (vgl. [Bit02]). Im Folgenden werden die Grundlagen des geometrischen Modells kurz erläutert.

# Kapitel 1. Grundlagen zur Raumakustik, der psychoakustischen Wahrnehmung von Quellen im Raum und der verwendeten Signalverarbeitung

---

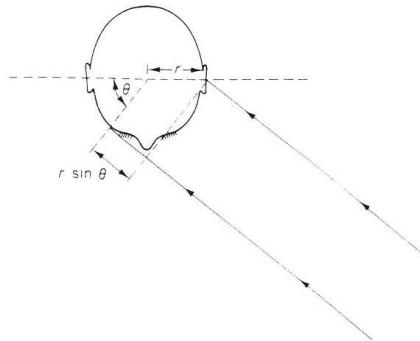


**Abb. 1.2:** Qualitative Beschreibung der Spiegelquellenmethode nach [AB76] (aus [Bit02]).

Beim geometrischen Modell nach [AB76], auch als Spiegelquellenmethode bezeichnet, wird der Raum mit den vorhandenen Quellen an allen Raumbegrenzungen dreidimensional gespiegelt. Je nach gewünschter Ordnung der Reflexionen muss diese Spiegelung entsprechend häufig fortgesetzt werden. Zusätzlich sind die Reflexionskoeffizienten der Wände für die gespiegelten Quellen zu berücksichtigen. Werden nun die Raumbegrenzungen außer acht gelassen, verhält sich die Schallausbreitung einer gespiegelten Quelle aufgrund der geometrischen Reflexion von Schallwellen an Begrenzungen vergleichbar zu den tatsächlichen Reflexionen an den Wänden. Zur Bestimmung der Raumimpulsantwort müssen die Signale der einzelnen Quellen unter Berücksichtigung der Laufzeiten am Empfängerort und der Schallabsorption in der Luft additiv überlagert werden. In Abbildung 1.2 ist schematisch das Vorgehen zur Bestimmung der RIA mithilfe der Spiegelquellenmethode dargestellt.

## 1.2 Räumliches Hören: interaurale Pegel- und Laufzeitdifferenzen, Kopfübertragungsfunktion

Die Fähigkeit des auditorischen Systems, die Position einer Quelle zu bestimmen, beruht vorwiegend auf binauralen Effekten, welche in der von Lord Rayleigh (1907) publizierten *Dublex Theorie* beschrieben werden. Zusätzlich nutzt das Gehör monaurale Effekte, vorwiegend um Mehrdeutigkeiten beim räumlichen Hören aufzulösen. In Abschnitt 2.2 wird betrachtet, in wieweit die Wahrnehmung der Sprecherausrichtung ebenfalls durch die im Folgenden näher erläuterten monauralen und binauralen Effekte beschrieben werden kann.



**Abb. 1.3:** Schematische Darstellung zur der interauralen Laufzeitdifferenz für eine Quelle im Winkel  $\theta$ . Bei einer frontalen Quelle sind keine Laufzeitunterschiede festzustellen. Bei einer seitlich versetzten Quelle erreichen die Schallwellen das zugewandte Ohr früher als das abgewandte Ohr. (aus [Moo82])

### 1.2.1 Binaurale Effekte: interaurale Pegel- und Laufzeitdifferenzen

Es wird angenommen, dass sich eine omnidirektionale Quelle frontal vor einem menschlichen Hörer befindet. In einem reflexionsfreien Raum kann dann davon ausgegangen werden, dass die ausgesendeten Schallwellen mit identischem Pegel und zum gleichen Zeitpunkt die Ohren erreichen. Für den Fall, dass die Quelle seitlich ausgelenkt wird, ergeben sich unterschiedliche Wegstrecken zu den Ohren. Da sich die Schallwellen mit konstanter Geschwindigkeit ausbreiten, erreicht das Signal das der Quelle zugewandte Ohr früher. Beim abgewandten Ohr trifft die Schallwelle später ein. Dieser Effekt wird als interaurale Laufzeitdifferenz (engl. *interaural time difference, ITD*) bezeichnet und ist in Abbildung 1.3 grafisch dargestellt. Die ITD  $\Delta T$  kann durch die winkelabhängige Entfernungsdifferenz, die der Schall von der Quelle zum jeweiligen Ohr benötigt, berechnet werden:

$$\Delta T = \frac{r}{c} \cdot (\sin(\theta) + \theta), \quad (1.1)$$

mit dem individuellen Kopfradius  $r$  und der Schallgeschwindigkeit  $c$ . [Moo03]

Alternativ kann die interaurale Phasendifferenz IPD  $\Delta\varphi$  in Abhängigkeit von der Frequenz  $f$  berechnet werden

$$\Delta\varphi = \Delta T \cdot f \cdot 2\pi. \quad (1.2)$$

Für höhere Frequenzen zeigt sich, dass die ITD bzw. die IPD zu höheren Frequenzen Mehrdeutigkeiten aufweisen. So ergibt sich aus einem 200 Hz Sinuston mit einer ITD von 500  $\mu\text{s}$  eine äquivalente IPD von 36 °. Oberhalb einer

## **Kapitel 1. Grundlagen zur Raumakustik, der psychoakustischen Wahrnehmung von Quellen im Raum und der verwendeten Signalverarbeitung**

---

Frequenz von 1500 Hz sind die Phasendifferenzen hochgradig mehrdeutig und können vom auditorischen System nicht mehr sinnvoll ausgewertet werden. [Moo03]

Der zweite wesentliche Effekt für die Lokalisation einer Schallquelle ist die interaurale Pegeldifferenz (engl. *interaural level difference, ILD*). Diese entsteht durch die wellenlängenabhängigen Beugungs- und Abschattungeigenschaften des menschlichen Kopfes. Tiefe Frequenzen mit großen Wellenlängen im Vergleich zum Kopfdurchmesser werden am Kopf gebeugt, so dass eine kleine bis gar keine Abschattung für tiefe Frequenzen auftritt. Hohe Frequenzen weisen eine kleine Wellenlänge auf und führen zu einer großen Abschatten durch den Kopf. Weiterhin ist die Abschattung abhängig vom Abstand der Quellen zum Kopf. Bei einer entfernten Quelle tritt die Abschattung nur bei hohen Frequenzen auf, während eine nahe liegende Quelle auch bei tiefen Frequenzen zu einer ILD führt. (aus [Moo82])

In der Duplex Theorie werden die Effekte ILD und ITD zusammengefasst. Die ILD wirkt vor allem bei hohen Frequenzen während die ITD bei tiefen Frequenzen die größten Effekte zeigt. Die Kombination dieser Effekte nutzt das auditorische System für die Lokalisation einer Quelle im Raum.

### **1.2.2 Monaurale Effekte: Cone of Confusion und der Effekt der Kopfübertragungsfunktion**

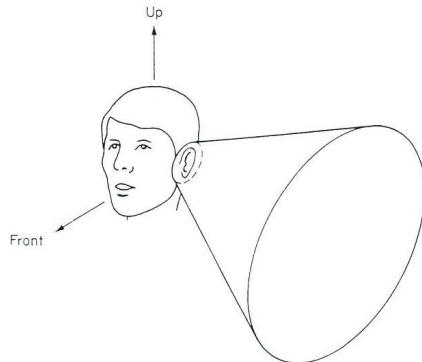
Die ITDs und ILDs sind nicht für alle Winkel eindeutig. Unter der Vereinfachung, dass der Kopf einer Kugel entspricht, würde eine Quelle vorne rechts zu den gleichen interauralen Laufzeit- und Pegelunterschieden führen, wie eine Quelle hinten rechts. Die gleichen Unterschiede ergeben sich so theoretisch für alle Quellen, die sich auf der Oberfläche eines Kegels befinden. Dieser Kegel wird als *Cone of Confusion* bezeichnet und ist in Abbildung 1.4 schematisch dargestellt.

Trotz der Vereinfachung ist ersichtlich, dass beim menschlichen Kopf beliebig viele Quellpositionen zu identischen ITDs und ILDs führen können. Am anschaulichsten ist dies bei einem Azimuth von Null Grad. Unabhängig von der Elevation, müssten die Signale links und rechts identisch und nicht differenzierbar sein. Allerdings zeigen eine Vielzahl von Untersuchungen, z.B. [Bla97], dass eine Unterscheidbarkeit der Elevation möglich ist.

Möglich wird dies erst durch die individuellen Kopfübertragungsfunktionen (engl. *head related transfer functions, HRTF*) oder Außenohrübertragungsfunktionen. Diese beschreiben die komplexe Filterfunktion des menschlichen Kopfes und werden vornehmlich durch

## Kapitel 1. Grundlagen zur Raumakustik, der psychoakustischen Wahrnehmung von Quellen im Raum und der verwendeten Signalverarbeitung

---



**Abb. 1.4:** Schematische Darstellung zum Effekt des *Cone Of Confusion*. Allein durch die ILDs und ITDs würde jede Quelle auf der Kugeloberfläche theoretisch zu identischen interauralen Laufzeit- und Pegeldifferenzen führen, so dass u.a. die Lokalisationsauflösung der Normalhörenden in Abhängigkeit der Elevation nicht hinreichend erklärt werden kann. (aus [Moo82])

die Form des Kopfes, des Außenohres und dem Oberkörper bestimmt. Die HRTF ist richtungsabhängig. Das menschliche auditorische System ist in der Lage die richtungsabhängigen Unterschiede zum Auflösen der Mehrdeutigkeiten auszunutzen. Weiterhin führen auch Kopfbewegungen zum Auflösen der Mehrdeutigkeiten des *Cone of Confusion*.

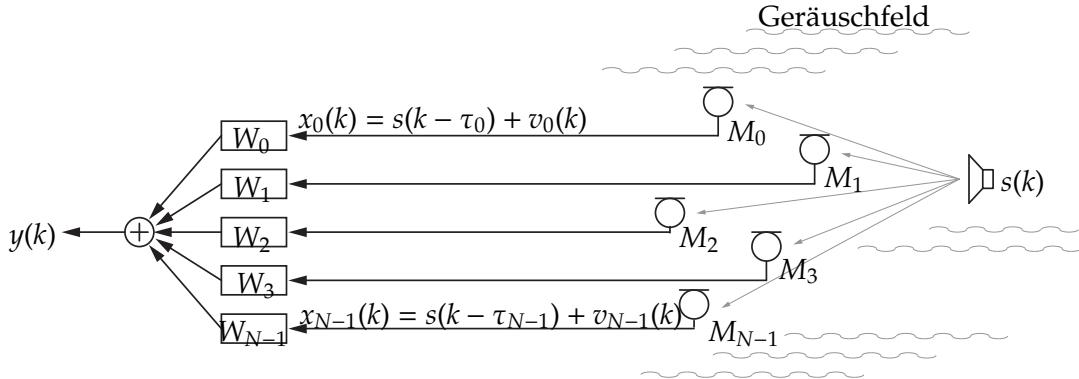
### 1.3 Signalmodell mehrkanaliger Mikrofone

Das Signalmodell zur Funktionsbeschreibung von Multimikrofon-Systemen ist in Abbildung 1.5 dargestellt. Es wird angenommen, dass sich eine Nutzschallquelle  $s(k)$  im Fernfeld einer mehrkanaligen Empfangsapparatur befindet, so dass die Schallwellen der Quelle die Mikrofone als ebene Wellen erreichen. Zusätzlich zum Nutzsignal wird ein beliebig räumlich verteiltes Geräuschfeld  $v(k)$  angenommen, welches sich additiv mit dem Nutzsignal überlagert. Die Unkorreliertheit der Signale wird vorausgesetzt [BMS01].

Eine Sensorgruppe (engl. *Array*) in diesem Geräuschfeld tastet das resultierende Schallfeld an diskreten Raumpunkten ab. Durch die räumliche Anordnung der Sensoren entsprechen die diskreten Eingangssignale jedes einzelnen Mikrofons einer zeitversetzten und gedämpften Version des gesendeten Nutzsignals  $s(k)$ . Für den Eingang des  $i$ -ten Mikrofons ergibt sich  $x_i(k) = a_i s(k - \tau_i)$  mit der Zeitverzögerung  $\tau$  und für den Eingang der gesam-

# Kapitel 1. Grundlagen zur Raumakustik, der psychoakustischen Wahrnehmung von Quellen im Raum und der verwendeten Signalverarbeitung

---



**Abb. 1.5:** Signalmodell zur Funktionsbeschreibung von mehrkanaligen Mikrofon-Systemen. Die Mikrofone nehmen das Schallfeld, bestehend aus der additiven Überlagerung eines beliebigen Geräuschenfeldes  $v(k)$  und dem Signal einer im Fernfeld befindlichen Schallquelle  $s(k)$ , auf. Die Mikrofone tasten das resultierende Schallfeld an diskreten Raumpunkten ab. Nach [BMS01].

ten Sensorgruppe mit  $N$  Mikrofonen im Zeitbereich

$$\begin{pmatrix} x_0(k) \\ x_1(k) \\ \vdots \\ \vdots \\ x_{N-1}(k) \end{pmatrix} = \begin{pmatrix} a_0s(k - \tau_0) \\ a_1s(k - \tau_1) \\ \vdots \\ \vdots \\ a_{N-1}s(k - \tau_{N-1}) \end{pmatrix} + \begin{pmatrix} v_0(k) \\ v_1(k) \\ \vdots \\ \vdots \\ v_{N-1}(k) \end{pmatrix}$$

$$\mathbf{x} = \mathbf{a} \cdot \mathbf{s}(k - \tau) + \mathbf{v}(k). \quad (1.3)$$

Mit Hilfe der Fouriertransformation kann das diskrete Eingangssignal im Zeitbereich aus Gleichung 1.3 ohne Verlust der Gültigkeit in den Frequenzbereich überführt werden.

$$\mathbf{X}(e^{j\Omega}) = \mathbf{S}(e^{j\Omega})\mathbf{d} + \mathbf{V}(e^{j\Omega}). \quad (1.4)$$

Der Vektor  $\mathbf{d}$  beschreibt die frequenzabhängige und zeitverzögerte Verstärkung der Kanäle in Abhängigkeit von der gewählten Array-Geometrie und der Position der Quellen:

$$\mathbf{d}^T = [a_0 e^{-j\frac{\Omega f_s}{c} \tau_0}, a_1 e^{-j\frac{\Omega f_s}{c} \tau_1}, \dots, a_{N-1} e^{-j\frac{\Omega f_s}{c} \tau_{N-1}}]. \quad (1.5)$$

Der Ausgang der Array-Struktur berechnet sich durch die Summierung der

## Kapitel 1. Grundlagen zur Raumakustik, der psychoakustischen Wahrnehmung von Quellen im Raum und der verwendeten Signalverarbeitung

---

gewichteten Eingangskanäle

$$Y(e^{j\Omega}) = \sum_{n=0}^{N-1} W_n^*(e^{j\Omega}) X_n(e^{j\Omega}) = \mathbf{W}^H(e^{j\Omega}) \mathbf{X}(e^{j\Omega}), \quad (1.6)$$

mit den frequenzabhängigen Koeffizienten  $W_n^*(e^{j\Omega})$  des  $n$ -ten Mikrofons. Das zeitdiskrete Ausgangssignal des Raum-Zeit-Filters ergibt sich aus der inversen Fouriertransformation von  $Y(e^{j\Omega})$  [BMS01]. (gesamter Abschnitt entnommen aus [Fra08])

## 1.4 Lineare Prädiktion

Die Lineare Prädiktion (*LP*) beschreibt, wie ein beliebiges Sample  $s[k]$  eines Signales  $s$  aus einer Linearkombination der vorangegangenen  $p$  Samples geschätzt werden kann, wobei  $p$  der Ordnung der Prädiktion entspricht [GMvdP<sup>+</sup>09] [Mak75]. Die Schätzung  $\hat{s}[k]$  des  $k$ -ten Samples wird bestimmt durch

$$\hat{s}[k] = - \sum_{n=1}^p a_k s[k-n] \quad (1.7)$$

mit den LP Koeffizienten  $a_k$ . Das Fehlersignal  $e[k]$  ergibt sich bei dieser Voraussage aus der Differenz des Signals  $s[k]$  und der Schätzung  $\hat{s}[k]$

$$e[k] = s[k] - \hat{s}[k] = s[k] + \sum_{n=1}^p a_k s[k-n]. \quad (1.8)$$

Entsprechend des Quelle-Filter-Modells (basierend auf [Fra70]) kann das Fehlersignal als das Anregungssignal bei der Spracherzeugung interpretiert werden.

Typischerweise wird beim LP-Verfahren das zeitdiskrete Signal in kurze, überlappende Blöcke zerlegt. Die Blocklänge liegt meist im Größenbereich von etwa 20 ms, da in diesem Bereich von einer Stationarität der Sprache ausgegangen werden kann. Obwohl die Komponenten der Spracherzeugung sowohl zeit- als auch frequenzabhängig sind, können einzelne Teilespekte, wie z.B. die Abstrahlung des Signals als zeitinvariant und die Amplitude als frequenzinvariant angenommen werden [VHH98]. Die optimalen LP-Filterkoeffizienten  $a_k$  ergeben sich durch die Minimierung des mittleren quadratischen Fehlers innerhalb eines Blocks. Die LP-Filterkoeffizienten

## Kapitel 1. Grundlagen zur Raumakustik, der psychoakustischen Wahrnehmung von Quellen im Raum und der verwendeten Signalverarbeitung

---

beschreiben das inverse Filter  $A(z)$

$$A(z) = 1 + \sum_{n=1}^p a_n z^{-n}. \quad (1.9)$$

Für ein aufgenommenes Sprachsignal ist die folgende Betrachtung möglich und wurde aus [GMvdP<sup>+</sup>09] entnommen. Bei der Spracherzeugung wird nach dem Quelle-Filter-Modell (basierend auf [Fra70]) ein Glottissignal  $G(z)$  erzeugt. Dieses Signal besteht in erster Näherung aus einer gewichteten Summe von Impulskämmen mit variabler Grundfrequenz für stimmhafte Laute (Vokale) und Rauschen für stimmlose Laute. Dieses Signal wird durch den Mund-, Rachen- und Nasenraum mit der Übertragungsfunktion  $T(z)$  gefiltert. Die Abstrahlung und Übertragung vom Mund zum Empfänger kann mit der Übertragungsfunktion  $H(z)$  beschrieben werden. Ein mit einem Mikrofon im Raum aufgenommenes Sprachsignal  $S(z)$  ergibt sich so aus

$$S(z) = H(z)T(z)G(z). \quad (1.10)$$

Die Übertragungsfunktion des Vokaltraktes ist annähernd minimalphasig, die Raumübertragungsfunktion hingegen meist nicht. Zudem ist die All-Pole Übertragungsfunktion der LP-Komponenten  $1/A(z)$  dazu geeignet eine minimalphasige Übertragungsfunktion wie die des Vokaltrakts  $T(z)$  nachzubilden. Das Fehlersignal  $E(z)$  ergibt sich so annähernd aus

$$E(z) = A(z)S(z) = A(z)H(z)T(z)G(z) \quad (1.11)$$

$$\approx A(z)H(z)\frac{1}{A(z)}G(z) \quad (1.12)$$

$$\approx H(z)G(z). \quad (1.13)$$

Hier zeigt sich, dass das Fehlersignal  $E(z)$  vom Einfluss des Vokaltraktfilters befreit wurde und in erster Näherung aus dem mit der RIA gefilterten Glottissignal besteht.

## 1.5 Algorithmen zur Richtungsschätzung und zur Schätzung der Sprecherausrichtung

Bisherige Verfahren zur Schätzung der Sprecherausrichtung basieren meist auf mehreren im Raum verteilten Mikrofon-Arrays oder Arrays mit einer Vielzahl von Mikrofonen und/oder großen Mikrofonabständen. Das in [CFSC<sup>+</sup>08] vorgeschlagene Verfahren basiert auf dem Verhältnis zwischen

# Kapitel 1. Grundlagen zur Raumakustik, der psychoakustischen Wahrnehmung von Quellen im Raum und der verwendeten Signalverarbeitung

---

den Leistungen eines tiefen und eines hohen Frequenzbandes und wird dort in Kombination mit einem bildverarbeitenden Verfahren zur Schätzung der Sprecherausrichtung eingesetzt. In [NYN09] wird hingegen ein Verfahren zur Schätzung der Sprecherausrichtung vorgestellt, welches auf einem künstlichen neuronalen Netz aufbaut.

Ein Standardverfahren zur Bestimmung der Sprecherausrichtung in Räumen ist das in [BOS05] vorgestellte *oriented Global Coherence Field (oGCF)*. Bei diesem Verfahren wird mithilfe der *allgemeinen Kreuzkorrelation mit Phasentransformation* (engl. *generalized cross-correlation phase transform, GCC-PHAT*) und dem *globalen Kohärenzfeld* (engl. *global coherence field, GCF*) die Hauptstrahlrichtung einer Quelle im Raum geschätzt. Um das oGCF zu beschreiben werden im Folgenden die zugrundeliegenden Verfahren GCC-PHAT und GCF vorgestellt.

## 1.5.1 Schätzung der Sprecherposition mittels allgemeiner Kreuzkorrelation mit Phasentransformation

Ein gängiges Verfahren zur akustischen Schätzung der Sprecherposition basiert auf der allgemeinen Kreuzkorrelation mit Phasentransformation zweier Mikrofonsignale  $x_i$  und  $x_l$ , welche in [KC76] und [DHB01] ausführlich beschrieben wird. Die folgenden Zusammenfassung basiert auf [Roe06].

Die Standard-Korrelationsfunktion  $r_{x_i x_l}$  ist definiert als der Erwartungswert  $E\{\cdot\}$  „aus dem Produkt der zeitlich um den Verschiebungskoeffizienten  $\kappa$  verschobenen Zufallsvariablen“ [Roe06]

$$r_{x_i x_l}(\kappa) = E\{x_i(k)x_l(k + \kappa)\}. \quad (1.14)$$

Als Schätzung der Korrelationsfunktion  $\hat{r}_{x_i x_l}(\kappa)$  aus zwei begrenzten Vektoren der Länge  $N$  ergibt sich

$$\hat{r}_{x_i x_l}(\kappa) = \frac{1}{N} \sum_{k=0}^{N-1} x_i(k)x_l(k + \kappa) \begin{cases} i = l, AKF : r_{x_i x_i} \\ i \neq l, KKF : r_{x_i x_l} \end{cases} \quad (1.15)$$

mit der Autokorrelation (AKF)  $r_{x_i x_i}$  und der Kreuzkorrelation (KKF)  $r_{x_i x_l}$ .

Die allgemeine Kreuzkorrelation  $R_{x_i x_l}(\kappa)$  ergibt sich aus der Korrelation von

# Kapitel 1. Grundlagen zur Raumakustik, der psychoakustischen Wahrnehmung von Quellen im Raum und der verwendeten Signalverarbeitung

---

zwei gefilterten Signalen

$$R_{x_i x_l}(\kappa) = \text{IDFT} \{(G_i(n)X_i(n))(G_l(n)X_l(n))^*\} \quad (1.16)$$

$$= \text{IDFT} \{G_i(n)G_i(n)^* \text{E}\{X_i(n)X_l(n)^*\}\} \quad (1.17)$$

$$= \text{IDFT} \{\Psi_{il}(n)\phi_{X_i X_l}(n)\} \quad (1.18)$$

mit den Filtern  $G_i(n)$  und  $G_l(n)$ , der inversen diskreten Fouriertransformation  $\text{IDFT}\{\cdot\}$ , der frequenzabhängigen Filterfunktion  $\Psi_{il}(n) = G_i(n)G_i(n)^*$  und dem Kreuzleistungsdichtespektrum (KLDS)  $\phi_{X_i X_l}(n)$ .

Zur Schätzung der internen Laufzeitverzögerung  $\hat{\tau}$  zwischen den Mikrofonsignalen muss das Maximum der allgemeinen Kreuzkorrelation bestimmt werden, welches sich bildet, wenn die Mikrofonsignale durch die Verschiebung gleich sind

$$\hat{\tau} = \arg \max_k R_{x_i x_l}[k]. \quad (1.19)$$

Bei der allgemeinen Kreuzkorrelation mit Phasentransformation wird das Gewichtungsfilter so gewählt, „dass die Beträge der KLDS  $\phi_{X_i X_l}(n)$  zu eins werden“, [Roe06]

$$\Psi_{X_i X_l}^{PHAT}(n) = \frac{1}{|\phi_{X_i X_l}|}. \quad (1.20)$$

Bei diesem Verfahren, welches als pre-wighting bezeichnet wird, wird das Betragspektrum eliminiert, da es für eine Laufzeitschätzung nicht erforderlich ist. Die allgemeine Kreuzkorrelation mit Phasentransformation weist anschließend ein deutliches Maximum bei der gesuchten relativen Verzögerung zwischen den Mikrofonsignalen auf. Mithilfe dieser Verzögerung kann die relative Richtung der Quellposition in Bezug auf die Mikrofonpositionen bestimmt werden.

Die Einschränkung der allgemeinen Kreuzkorrelation mit Phasentransformation besteht darin, dass sie nur in der Lage ist, die relative Richtung der Quellposition in Bezug auf die Mikrofonpositionen bestimmen zu können. Die relative Position (Richtung und Entfernung) zu den Mikrofonen oder die absolute Position in einem Raum ist mit dem beschriebenen *GCC-PHAT* Verfahren so nicht möglich.

### 1.5.2 Schätzung der Sprecherposition unter Verwendung des globalen Kohärenzfeldes

In [SS04] wird dazu ein großes Mikrofon-Array mit 512 an den Raumwänden angebrachten und gleichverteilten Mikrofonen vorgestellt, um die Sprecherposition und -ausrichtung im Raum zu schätzen. Im Gegensatz dazu, wird in [BOS05] das globale Kohärenzfeld vorgestellt und untersucht, welches die Mikrofonanzahl auf eine praktische Anzahl reduziert. Die Autoren verwenden fünf im Raum verteilte T-förmige Mikrofonarrays, bestehend aus vier Mikrofonen. Das globale Kohärenzfeld basiert auf der *cross-power spectrum phase (CSP)* zwischen zwei Mikrofonsignalen  $X_i$  und  $X_l$ . Das CSP  $C_{il}^{CSP}$  wird bestimmt nach

$$C_{il}^{CSP}(\kappa) = IDFT \left\{ \frac{X_i(n)X_l(n)^*}{|X_i(n)X_l(n)^*|} \right\}, \quad (1.21)$$

mit dem Verschiebungsindex der Korrelationsfunktionen  $\kappa$ . Das CSP entspricht der allgemeinen Kreuzkorrelation mit Phasentransformation. Die relative Laufzeitverzögerung des Quellsignals zwischen den Sensoren  $i$  und  $k$  kann ebenfalls aus der Position des Maximums des CSP bestimmt werden.

Angenommen wird nun ein Ensemble  $\Omega$ , welches  $Q$  Mikrofonpaare enthält, und die theoretische Laufzeitverzögerung  $\delta_{ik}(x, y, z)$  für das Mikrofonpaar  $(i, k)$  für die Quelle an der dreidimensionalen Position  $(x, y, z)$ . So ergibt sich das globale Kohärenzfeld  $GCF_\Omega$  mit dem CSP zum diskreten Zeitindex  $k$  nach

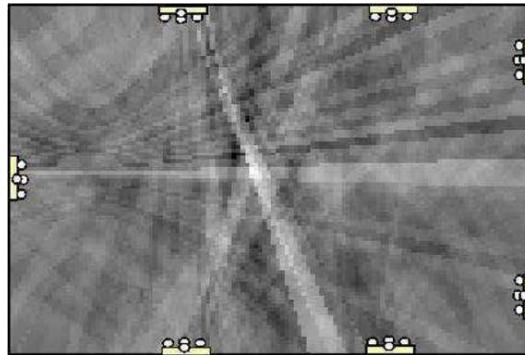
$$GCF_\Omega(k, x, y, z) = \frac{1}{Q} \sum_{(i,l) \in \Omega} C_{il}^{CSP}(k, \delta_{ik}(x, y, z)) \quad (1.22)$$

Die theoretische Laufzeitverzögerung  $\delta_{ik}(x, y, z)$  wird bestimmt, indem der betrachtete Raum in ein Raster möglicher Sprecherpositionen in einer frei wählbaren Auflösung aufgeteilt wird. Für jeden Punkt des Rasters kann die theoretische Laufzeitverzögerung zu den Mikrofonen bestimmt werden

$$\delta_{ik}(x, y, z) = \frac{d_l - d_i}{c} f_s, \quad (1.23)$$

mit den Abständen  $d_l$  und  $d_i$  vom Rasterpunkt an der Position  $(x, y, z)$ , der Schallgeschwindigkeit  $c$  und der Abtastfrequenz  $f_s$  [Roe06].

Das GCF bestimmt so die Wahrscheinlichkeit, dass sich eine Quelle an einem bestimmten Punkt des Rasters befindet. Am wahrscheinlichsten Rasterpunkt weiß das GCF ein Maximum auf. In Abbildung 1.6 ist grafisch



**Abb. 1.6:** Beispiel eines zweidimensionalen globalen Kohärenzfeldes unter Verwendung der fünf links positionierten Mikrofonarrays. Je heller das GCF, desto wahrscheinlicher ist es, dass sich an der entsprechenden Position eine Quelle befindet. (aus [BOS05])

ein GCF dargestellt.

### 1.5.3 Schätzung der Sprecherausrichtung unter Verwendung des *oriented global coherence field*

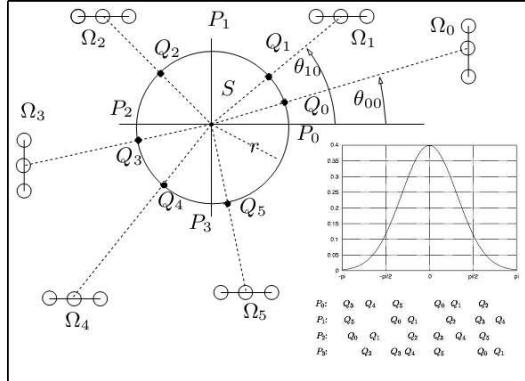
Neben der Möglichkeit die Sprecherposition mithilfe des GCF zu schätzen, wird in [BOS05] darauf aufbauend das *oriented global coherence field* vorgestellt. Das oGCF erlaubt eine Bewertung, ob eine direktive Quelle an einer bestimmten Position in eine der betrachteten Richtungen ausgerichtet ist.

Angenommen, eine Quelle befindet sich im Raum an der Position  $S = (x_s, y_s, z_s)$ , dann kann die Position der Quelle mithilfe des GCF geschätzt werden. Beim oGCF wird nun ein Kreis um  $C$  mit dem Radius  $r$  um die Quellposition  $S$  angenommen. Auf dem Kreis  $C$  können  $N$  äquidistante Punkte  $P_j$  gewählt werden, welche als mögliche Richtungen der Quelle angenommen werden. Im nächsten Schritt werden  $l$  Schnittpunkte  $Q_l$  auf dem Kreis  $C$  bestimmt, die sich durch eine gerade Verbindung zwischen der Position  $S$  und den verwendeten Mircofonarray  $\Omega_l$  ergeben. [BOS05] In Abbildung 1.7 wird das Verfahren zur Bestimmung der erforderlichen Punkte  $P_j$  und  $Q_l$  grafisch dargestellt.

Die Wahrscheinlichkeit, dass die Quelle in Richtung  $P_j$  ausgerichtet ist, wird durch das Bewertungsmaß  $o_j$  mit  $j = 0 \dots N-1$  bestimmt. „Für jede Blickrichtung  $P_j$  wird die Summe über alle  $GCF_{\Omega_l}(Q_l)$  gebildet. Um die Sensorarrays in der gedachten Blickrichtung stärker zu bewerten als die hinter dem Sprecher, erfolgt eine gaußsche Gewichtung aller Arrays.“ [Roe06]. Das oGCF  $O_j$

# Kapitel 1. Grundlagen zur Raumakustik, der psychoakustischen Wahrnehmung von Quellen im Raum und der verwendeten Signalverarbeitung

---



**Abb. 1.7:** Grafische Darstellung zur Schätzung der Sprecherausrichtung unter Verwendung eines zweidimensionalen *oriented global coherence fields* mit sechs Mikrofonarrays (aus [BOS05]).

an der Position  $S$  ergibt sich aus

$$O_j(S) = \sum_{l=0}^{L-1} GCF_{\Omega_l}(Q_l) \omega(\theta_{lj}), \quad (1.24)$$

mit der Gewichtungsfunktion

$$\omega(\theta) = \frac{1}{\sqrt{2\pi}} e^{\frac{\theta^2}{2}} \quad (1.25)$$

und dem Winkel  $\theta_{lj} \in [-\pi, \pi]$  zwischen der Linie durch  $S$  und  $P_j$  und der Linie von  $S$  nach  $Q_l$ . [BOS05]

Die geschätzte Sprecherausrichtung  $\hat{P}_j$  ergibt sich anschließend aus dem Maximum des oGCF

$$\hat{P}_j = \max\{O_j(S)\}. \quad [Roe06] \quad (1.26)$$

---

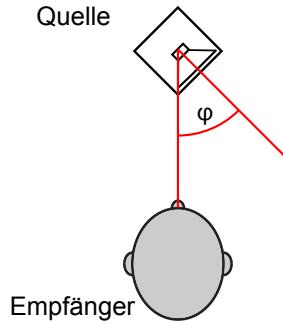
## Kapitel 2

# Wahrnehmung der relativen Sprecherausrichtung bei normalhörenden Probanden und Voruntersuchungen für die implementierten Algorithmen

In diesem Kapitel werden die Grundlagen zur relativen Sprecherausrichtung vorgestellt. Eine Vielzahl von Veröffentlichungen behandeln die Wahrnehmung des räumlichen Hörens von normalhörenden und schwerhörigen Menschen. Typischerweise wird hier die räumliche Position oder der Einfluss von Nachhall auf die Wahrnehmung betrachtet. Weitauß weniger untersucht wurde bisher die auditorische Wahrnehmung und Unterscheidbarkeit der Ausrichtung von direktiven Quellen. Neben einer Definition des Begriffes *relative Sprecherausrichtung* werden die Ergebnisse von drei veröffentlichten Studien zur Untersuchung der menschlichen Wahrnehmung der relativen Sprecherausrichtung und deren Ergebnisse vorgestellt. Darauf folgt ein Vergleich der Abstrahlcharakteristik vom menschlichen Kopf und der Lautsprecherbox Fostex 6301B, da beide Quellen für die durchgeführten Experimente in den Kapiteln 2 und 3 verwendet wurden. Abschließend folgt eine Betrachtung des Einflusses der relativen Sprecherausrichtung auf wahrnehmbare akustische Eigenschaften und eine Zusammenfassung der bisherigen Erkenntnisse.

## Kapitel 2. Wahrnehmung der relativen Sprecherausrichtung bei normalhörenden Probanden und Voruntersuchungen für die implementierten Algorithmen

---



**Abb. 2.1:** Schematische Darstellung der Sprecherausrichtung  $\varphi$ , welche als der Winkel zwischen der Hauptabstrahlrichtung einer Quelle und der direkten Verbindung zwischen Quelle und Empfänger definiert ist. (nach [Neu03])

### 2.1 Begriffsdefinition *relative Sprecherausrichtung*

Im Allgemeinen ist die relative Quellenausrichtung (engl. *facing angle*) als der Winkel  $\varphi$  zwischen der Hauptabstrahlrichtung einer Quelle und der direkten Verbindung zwischen Quelle und Empfänger definiert. Abbildung 2.1 stellt diesen Zusammenhang grafisch dar. Die relative Quellenausrichtung ist für alle direktiven Quellen bestimmbar, im Gegensatz zu omnidirektionalen Quellen, die keine Richtwirkung aufweisen [Neu03]. Aus der Definition und den frequenzabhängigen Abstrahlcharakteristiken aller Quellen lässt sich so eine frequenzabhängige auditorisch wahrnehmbare Quellenausrichtung ableiten, da im Realen keine optimal direktionalen Quellen existieren und so immer eine frequenzabhängige Richtwirkung aufweisen. Folgend wird die relative Quellenausrichtung auf den menschlichen Kopf oder Quellen mit kopfähnlicher Abstrahlcharakteristik betrachtet und somit durchgängig der Begriff (*relative*) *Sprecherausrichtung* verwendet.

### 2.2 Evaluation der Sprecherausrichtung in der Literatur

Bisher wurden nur wenige Ergebnisse zur Wahrnehmung und Unterscheidbarkeit der Sprecherausrichtung von normalhörenden Probanden veröffentlicht. Untersuchungen zur Wahrnehmung und Unterscheidbarkeit der Sprecherausrichtung mit Schwerhörigen sind zum Zeitpunkt dieser Arbeit nicht bekannt. Die im Folgenden vorgestellten Versuche, Experimente und Ergebnisse unterscheiden sich vor allem in der Art des Raumes und der Wahl der verwendeten Quelle.

## **Kapitel 2. Wahrnehmung der relativen Sprecherausrichtung bei normalhörenden Probanden und Voruntersuchungen für die implementierten Algorithmen**

---

### **2.2.1 Untersuchung der menschlichen Fähigkeit zur Bestimmung der Sprecherausrichtung im reflexionsarmen Raum bei stationärer und rotierender Sprecherausrichtung (nach Kato et al.)**

In [KTNM10] wurde die Fähigkeit zur Bestimmung der Sprecherausrichtung in einem reflexionsarmen Raum (BxTxH: 4,8 m x 5,4 m x 4 m) bei stationärer und rotierender Sprecherausrichtung mit 12 normalhörenden Probanden (7 weibliche, 5 männliche) untersucht. Als Quelle diente ein männlicher Sprecher, welcher auf einem Drehstuhl saß und jeweils einen von vier Sätzen sprach. Den Probanden wurden die Augen verbunden und sie wurden frontal auf einem Stuhl in einem Abstand von 1,2 m und 2,4 m positioniert. Mit diesem Versuchsaufbau bestand die Aufgabe der Probanden darin, die Sprecherrichtung zu schätzen, wobei diese in verschiedenen Versuchsreihen variierte.

In der ersten Versuchsreihe blicke der Sprecher während er sprach konstant in eine Richtung. Zufällig verteilt wurden pro Proband alle Azimuths zwischen -135 ° und 180 ° in 45 ° Schritten getestet. Der Proband sollte anschließend angeben, in welche Richtung der Sprecher geblickt hat. Während der zweiten Versuchsreihe begann der Sprecher bei einer der möglichen Richtungen und drehte sich während er sprach um 45 ° nach rechts oder links. Dort angekommen, verstummte er. Hier sollte der Proband bestimmen, in welcher Richtung die Drehung erfolgte. Im Unterschied zur zweiten Versuchsreihe erfolgte in der Dritten die Drehung um 45 ° während einer kurzen Sprechpause. Hier sollte überprüft werden, ob die Probanden anhand der Differenz zwischen den zwei Signalen die Richtung der Drehung bestimmen können.

Die letzten zwei Messreihen befassten sich mit der Wahrnehmung der Sprecherausrichtung bei unterschiedlichem Elevationswinkel. Der Ablauf der vierten Messreihe entsprach der Ersten, allerdings unter Verwendung der Elevation -45 °, 0 ° und 45 °. Der Proband gab anschließend die geschätzte Blickrichtung an. Während er sprach, musste der Sprecher im fünften Versuch seinen Kopf um 45 ° nach oben oder unten bewegen. Auch hier sollte der Proband angeben, in welche Richtung die Kopfbewegung durchgeführt wurde.

In Tabelle 2.1 sind die Messergebnisse der durchgeführten Versuchsreihen dargestellt. Die Messergebnisse für die getesteten Distanzen von 1,2 m und 2,4 m und der unterschiedlichen Sätze wurden zusammengefasst, da sich dabei keine signifikanten Unterschiede zeigten. Die große Diskrepanz zwi-

## Kapitel 2. Wahrnehmung der relativen Sprecherausrichtung bei normalhörenden Probanden und Voruntersuchungen für die implementierten Algorithmen

Aufgabenstellung	Gesamtanzahl der Antworten	Korrekte Antworten	Durchschnittliche Fehlerrate
fixer Azimuthwinkel	768	57 % (12,5 %)	23,5 ° (90 °)
horizontale Drehung während der Sprache	768	85 % (50 %)	
horizontale Drehung während Sprachpause	768	86 % (50 %)	
fixer Elevationswinkel	864	74 % (33 %)	12,9 ° (40 °)
vertikale Drehung während der Sprache	576	80 % (50 %)	

**Tab. 2.1:** Messergebnisse für Experimente zur Schätzung der Wahrnehmung unterschiedlicher Ausrichtungen eines männlichen Sprechers und normalhörenden Probanden, durchgeführt in einem reflexionsarmen Raum. Getestet wurden unterschiedliche statische als auch dynamische Azimuth- und Elevationswinkel. Die Werte in den Klammern geben die Ratewahrscheinlichkeit an. (aus [KTNM10])

schen dem fixen Azimuth und dem fixen Elevationswinkel ergeben sich vermutlich aus der geringeren Anzahl an unterschiedlichen Möglichkeiten bei den Elevationswinkeln. Zur Vergleichbarkeit werden diese Ergebnisse um die Ratewahrscheinlichkeit bereinigt. So ergibt sich für die fixen Azimuths  $0,57 \cdot \frac{7}{8} = 0,499$  und für die fixen Elevationswinkel  $0,74 \cdot \frac{2}{3} = 0,493$ . Somit ist kein Unterschied in diesen Ergebnissen erkennbar. Dies überrascht, da die Änderung des Azimuth sowohl Einfluss auf monaurale als auch auf binaurale akustische Eigenschaften haben sollte. Mögliche monaurale Eigenschaften sind in diesem Zusammenhang z. B. spektrale Unterschiede zwischen den unterschiedlichen Winkeln. Die binauralen Eigenschaften sind in den winkelabhängigen interauralen Laufzeit- und Pegeldifferenzen zu suchen. Bei unterschiedlichem Elevationswinkel können nur monaurale Effekte auftreten. Änderungen von binauralen Eigenschaften sollten theoretisch nicht auftreten.

Weiterhin ist anzumerken, dass bei den Versuchen mit unterschiedlichen Azimuths der Mund als Rotationsmittelpunkt gewählt wurde, so dass die Signalabstrahlung immer von einem fixen Punkt im Raum ausging. Im Gegensatz dazu wurde bei der Änderung der Elevationswinkel der Kopf vom Nacken aus auf und ab geneigt. Dies führt zu winkelabhängigen Quellpositionen und kann das Ergebnis der Elevationswinkel hin zu besseren Erkennungsleistungen verfälschen.

Bei den Versuchen mit Rotation des Sprechers zeigt sich, dass die durchschnittlichen Erkennungsraten zwischen 80 % und 86 % liegen. Die Darbie-

## Kapitel 2. Wahrnehmung der relativen Sprecherausrichtung bei normalhörenden Probanden und Voruntersuchungen für die implementierten Algorithmen

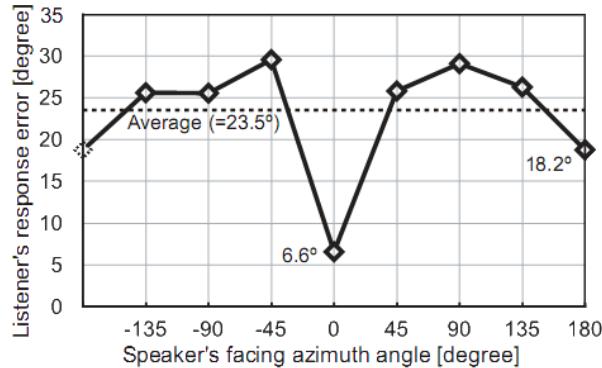


Fig. 3. Listeners' accuracy as a function of speaker's facing azimuth.

**Abb. 2.2:** Durchschnittliche Fehlerraten zur Schätzung der Sprecherausrichtung eines männlichen Sprechers von normalhörenden Probanden in einem reflexionsarmen Raum. (aus [KTNM10])

tung mit Sprechpausen zeigt dabei das beste Ergebnis. Dies überrascht, da zu erwarten gewesen wäre, dass die Probanden bei der Sprache während der Kopfbewegung von mehr akustischen Informationen während der kontinuierlichen Bewegung profitieren können.

In Abbildung 2.2 sind die durchschnittlichen Fehler der Probanden für die Versuchsreihe mit fixem Azimuth über der tatsächlichen Sprecherausrichtung dargestellt. Die mittlere Fehlerrate beträgt  $23,5^\circ$ . Bei einer Sprecherausrichtung von  $0^\circ$  können die Probanden mit einem mittleren Fehler von  $6,6^\circ$  die Sprecherausrichtung am sichersten bestimmen. Die Fehlerraten steigen mit über  $25^\circ$  bereits bei abweichenden Winkeln von  $\pm 45^\circ$  stark an. Für eine Sprecherausrichtung von  $180^\circ$  ergibt sich ein lokales Minimum der Fehlerrate von  $18,2^\circ$ . Eine mögliche Erklärung für die Minima bei  $0^\circ$  und  $180^\circ$  könnte sich daraus ergeben, dass für beide Fälle die Unterschiede zwischen den Signalen maximal sind und daher von den Probanden leichter bestimmt werden können. Die veröffentlichten Daten enthalten keine weiteren Streuungsmaße, so dass eine weitere Interpretation der Fehlerraten schwer fällt. So ist z.B. nicht ersichtlich, ob die Abweichungen möglicherweise nur durch rechts-links Verwechslungen entstanden sind.

### 2.2.2 Untersuchung der wahrnehmbaren Sprecherausrichtung bei stationärer und rotierender Sprecherausrichtung (nach Neuhoff)

Im Gegensatz zu den zuvor beschriebenen Experimenten in Abschnitt 2.2.1 wurden in [Neu03] Versuche zur Bestimmung der Sprecherausrichtung mit 13 (in Experiment 1) bzw. 20 (in Experiment 2) normalhörenden Proban-

## Kapitel 2. Wahrnehmung der relativen Sprecherausrichtung bei normalhörenden Probanden und Voruntersuchungen für die implementierten Algorithmen

---

den (jeweils zwischen 18 und 25 Jahren) in einem halligen Raum (BxTxH: 2,74 m x 3,66 m x 2,44 m) mit einer mittleren Nachhallzeit von  $T_{60} = 0.45$  s durchgeführt. Als Signalquelle diente eine Lautsprecherbox *Radio Shack Optimus XTS 40* (BxHxT: 12,5 cm x 12,5 cm x 11,4 cm), welche um ihren Gehäusemittelpunkt drehbar montiert wurde. Als Nutzsignal dienten eine viersekündige Aufnahme eines männlichen Sprechers, welche in einem reflexionsarmen Raum bei einem Pegel von  $\approx 65$  dB-A in einem Meter Abstand aufgenommen und über den Lautsprecher wiedergegeben wurde.

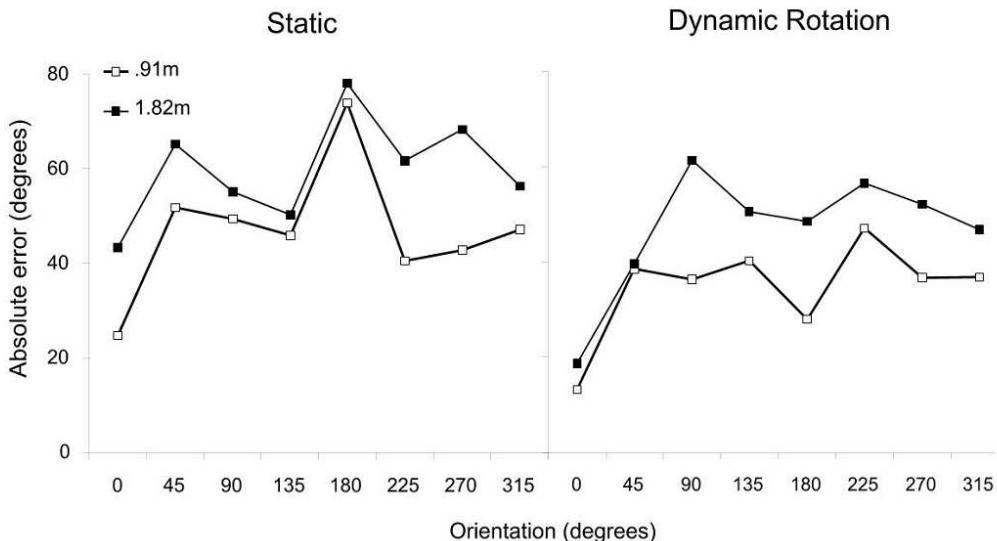
Den Probanden wurden die Augen verbunden und sie wurden in einem Abstand von 0,91 m bzw. 1,82 m frontal vor der Quelle positioniert. Im ersten Experiment nahm der Lautsprecher lautlos eine der acht Ausrichtungen zwischen  $-180^\circ$  und  $180^\circ$  in  $45^\circ$  Schritten ein. Mit einem zweiten, vor den Probanden positionierten Lautsprecher, sollten diese dann die geschätzte Ausrichtung der Quelle in entsprechenden  $45^\circ$  Schritten einstellen. Beim zweiten Experiment drehte sich die Quelle während der Darbietung um  $180^\circ$  und die Probanden hatten die geschätzte finale Ausrichtung an ihrem Lautsprecher einzustellen. Der Lautsprecher drehte sich in den ersten zwei Sekunden des Sprachstimulus um  $90^\circ/\text{s}$  um in der Endausrichtung die verbleibenden 2 Sekunden des Signals abzuspielen. Bei diesem Experiment wurde erwartet, dass die sich ändernden akustischen Eigenschaften der Quelle eine bessere Schätzung der Quellposition ermöglichen. Weiterhin wurden die besseren Ergebnisse bei geringerem Abstand und bei einer Sprecherausrichtung von  $0^\circ$  vermutet.

Im linken Diagramm der Abbildung 2.3 sind die Ergebnisse des ersten Experiments dargestellt. Es ist die gemittelte absolute Abweichung der Probandenschätzung über die wahre Sprecherausrichtung in Abhängigkeit von der Distanz aufgetragen. Es zeigt sich, dass in dem leicht halligen Raum die Erkennungsleitung in den getesteten Situationen signifikant von der Entfernung der Quelle zur Versuchsperson abhängt. Wie bereits bei den Experimenten aus Abschnitt 2.2.1 ergeben sich die kleinsten absoluten Abweichungen für eine Sprecherausrichtung von  $0^\circ$ . Die Abweichungen betragen hier ca.  $22^\circ$  für einen Abstand von 0,91 m und  $42^\circ$  für 1,82 m. Abweichend von Kato et al. zeigen die Daten von Neuhoff bei  $180^\circ$  kein lokales Minimum, sondern mit ca.  $70^\circ$  bis  $80^\circ$  die höchsten absoluten Abweichungen.

Das rechte Diagramm der Abbildung 2.3 zeigt die Ergebnisse des zweiten Experiments. Hier sind die gemittelten absoluten Abweichungen der Probandenschätzung über die wahre finale Sprecherausrichtung in Abhängigkeit von der Distanz aufgetragen. Wie zuvor im ersten Experiment unterscheiden sich die Ergebnisse zwischen den getesteten Distanzen signifikant. Bei

## Kapitel 2. Wahrnehmung der relativen Sprecherausrichtung bei normalhörenden Probanden und Voruntersuchungen für die implementierten Algorithmen

---



**Abb. 2.3:** Ergebnisse zur Untersuchung der wahrnehmbaren Sprecherausrichtung bei stationärer und rotierender Quelle (aus [Neu03])

einer Distanz von 0,91 m führt die Sprecherausrichtung ungleich  $0^\circ$  zu einer Abweichung von etwa  $40^\circ$  mit einem lokalem Minimum von etwa  $30^\circ$  bei  $180^\circ$ . Bei 1,82 m verläuft die mittlere Abweichung bei diesen Winkeln im Mittel bei ca.  $55^\circ$ . Auffallend ist die Sprecherausrichtung von  $0^\circ$ . Hier zeigt sich eine deutlich geringere absolute Abweichung von unter  $20^\circ$ . Neuhoff stellt fest, dass sich die Ergebnisse von den Sprecherausrichtungen ungleich  $0^\circ$  nicht signifikant unterscheiden.

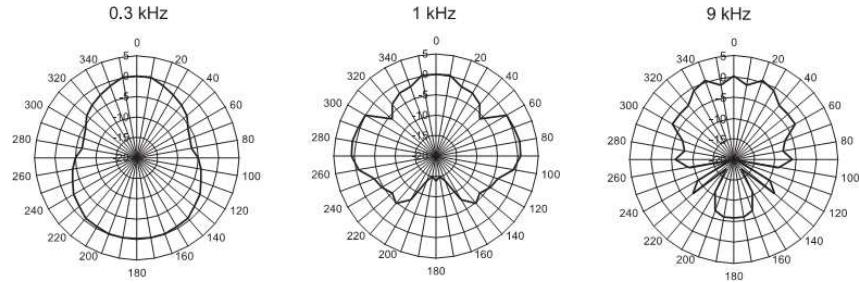
Beim zweiten Experiment wurde die Drehbewegung sowohl im Uhrzeigersinn als auch entgegengesetzt durchgeführt. Dabei stellte sich interessanterweise heraus, dass die Probanden die finale Ausrichtung signifikant besser schätzen konnten, wenn der Lautsprecher sich durch die  $0^\circ$  Ausrichtung bewegte.

Im Vergleich zum ersten Experiment mit statischer Lautsprecherausrichtung zeigt sich, dass die gemittelten absoluten Abweichungen im zweiten Experiment signifikant besser waren. Die Probanden konnten mit den zusätzlichen Informationen durch die Drehbewegung die Lautsprecherausrichtung besser schätzen als ohne.

Die Testergebnisse sind nicht direkt mit einem menschlichen Sprecher vergleichbar, da sich die richtungsabhängige Abstrahlcharakteristik der verwendeten Lautsprecherbox deutlich von der eines menschlichen Kopfes unterscheidet. In Abbildung 2.4 sind Richtdiagramme der Schallabstrahlcharakteristik der verwendeten Lautsprecherbox *Radio Shack Optimus XTS 40* für die Frequenzen 300 Hz, 1 kHz und 9 kHz abgebildet. Bei einer Frequenz von 300 Hz ist bereits eine richtungsabhängige Abstrahlung mit einem

## Kapitel 2. Wahrnehmung der relativen Sprecherausrichtung bei normalhörenden Probanden und Voruntersuchungen für die implementierten Algorithmen

---



**Abb. 2.4:** Richtdiagramme der Schallabstrahlcharakteristik der von Neuhoff verwendeten Lautsprecherbox *Radio Shack Optimus XTS 40*. (aus [Neu03])

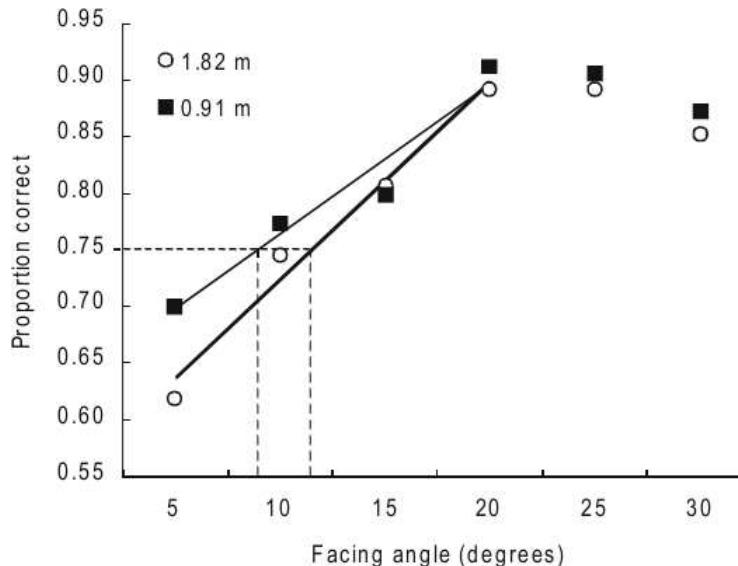
Minimum bei etwa  $\pm 70^\circ$  von -7 dB erkennbar. Diese Charakteristik wird mit steigender Frequenz ausgeprägter, so dass sich bei 1 kHz bereits eine nierenähnliche Richtcharakteristik mit einer Dämpfung von etwa -15 dB bei von  $180^\circ$  abzeichnet. Bei 9kHz sind bei Richtungen zwischen  $180^\circ$  und  $270^\circ$  weitere Nebenkeulen ausgeprägt. Richtdiagramme der Abstrahlcharakteristik eines menschlichen Kopfes sind in Abbildung 2.6 dargestellt. Diese weisen keine Nullstellen und Nebenkeulen auf und fallen für nahezu alle gemessenen Frequenzen mit steigenden Winkel annähernd monoton aus.

Bei den durchgeführten Versuchen ist anzumerken, dass als Rotationsachse der Mittelpunkt der Lautsprecherbox gewählt wurde. Somit liegt die Position der Signalerzeugung (Lautsprechermembran) außerhalb der Rotationsachse. Bei der gegebenen Abmessung des Lautsprechers wandert die Position der Quelle beim Drehen um einen Abstand von 0,91 m zum Probanden um etwa  $\pm 3,5^\circ$  von der Mittelachse zwischen Lautsprecher und Proband weg. Bei einem Abstand von 1,82 m wandert die Quelle um etwa  $\pm 1,8^\circ$ . Der minimale auditiv wahrnehmbare Winkel (engl. *minimum audible angle, MMA*) einer Quelle liegt bei einer frontalen Quelle frequenzabhängig zwischen 1 und 3 dB [Mil58]. Somit ist es bei den durchgeführten Versuchen möglich, dass die Probanden die sich ändernde Position der Quelle als eine zusätzliche auditive Eigenschaft wahrgenommen haben. Die Bewertung der Sprecherausrichtung könnte somit verfälscht sein. Um diesen Effekt auszuschließen ist zu beachten, dass sich der verwendete Lautsprecher um den Ursprung der Signalerzeugung dreht.

### 2.2.3 Untersuchung zur minimal wahrnehmbaren Quellausrichtung (nach Neuhoff et al.)

In [NRV01] wurden die Ergebnisse der Bestimmung der minimal wahrnehmbaren Quellausrichtung (engl. *minimum audible facing angle, MAFA*) mit 15 normalhörenden Probanden (zwischen 18 und 20 Jahren) un-

## Kapitel 2. Wahrnehmung der relativen Sprecherausrichtung bei normalhörenden Probanden und Voruntersuchungen für die implementierten Algorithmen



**Abb. 2.5:** Versuchsergebnisse der Bestimmung der minimal wahrnehmbaren Quellenausrichtung. Aufgetragen sind die gemittelten korrekten Antworten in Prozent über der getesteten Quellenausrichtung in Abhängigkeit von der Distanz zwischen Quelle und Proband. (aus [NRV01])

tersucht und veröffentlicht. Der MAFA ist definiert als die Quellenausrichtung, welche die normalhörenden Probanden zu 75 % korrekt bestimmen können. Wie bereits beschrieben, ist die Schätzung der Quellenausrichtung sowohl von der Distanz zwischen Quelle und Empfänger als auch von der Abstrahlcharakteristik der Quelle abhängig.

Das Experiment erfolgte in einem leicht halligen Raum ( $B \times T \times H: 2,74 \text{ m} \times 3,66 \text{ m} \times 2,43 \text{ m}$ ) mit einer Nachhallzeit von  $T_{60} = 0.15 \text{ s}$ . Ein zwischen 220 Hz und 10 kHz bandpassgefiltertes weißes Rauschen wurde über einen horizontal frei drehbaren *Radio Shack Optimus XTS 40* Lautsprecher abgespielt. Das Rauschen wurde in fünf jeweils 100 ms langen Impulsen mit einer Lückendauer von 100 ms bei 70 dB-A dargeboten. Die Probanden saßen 0,91 m bzw. 1,82 m vom Lautsprecher entfernt. Pro Testdurchlauf spielte der Lautsprecher das Testsignal bei einer Ausrichtung von  $0^\circ$  ab. Anschließend wurde der Lautsprecher auf einen der Winkel zwischen  $\pm 30^\circ$  in  $5^\circ$  Schritten eingestellt und das Testsignal erneut wiedergegeben. Die Probanden sollten anschließend angeben, in welche Richtung der Lautsprecher gedreht wurde.

Abbildung 2.5 zeigt die gemittelten korrekten Antworten in Prozent über den wahren Lautsprecherausrichtungen. Als minimal wahrnehmbare Quellenausrichtung wurde die Schwelle gewählt, bei der die Probanden die Drehrichtung des Lautsprechers zu 75 % korrekt bestimmen konnten. Für einen Abstand von 0,91 m zwischen Quelle und Proband ergibt sich der MAFA zu  $7,5^\circ$ . Für eine Distanz von 1,82 m beträgt der MAFA  $12,5^\circ$ .

Es wird vermutet, dass der MAFA direkt mit der resultierenden ILD korre-

## **Kapitel 2. Wahrnehmung der relativen Sprecherausrichtung bei normalhörenden Probanden und Voruntersuchungen für die implementierten Algorithmen**

---

liert. Die theoretische ILD kann durch die Entfernung zwischen Quelle und Sender und der bekannten Richtcharakteristik des Lautsprechers bestimmt werden. Der kleinste wahrnehmbare interaurale Pegelunterschied liegt für die meisten Frequenzen zwischen 0,5 dB und 1 dB [NRV01], [Mil60]. Die Kombination dieser Informationen führt zu einem theoretischen MAFA von  $6,8^\circ$  für einen Abstand von 0,91 m und von  $13,5^\circ$  für einen Abstand von 1,82 m. Diese Werte decken sich mit den Messergebnissen. Somit erscheint die ILD als die primäre Eigenschaft zur Schätzung der Sprecherausrichtung.

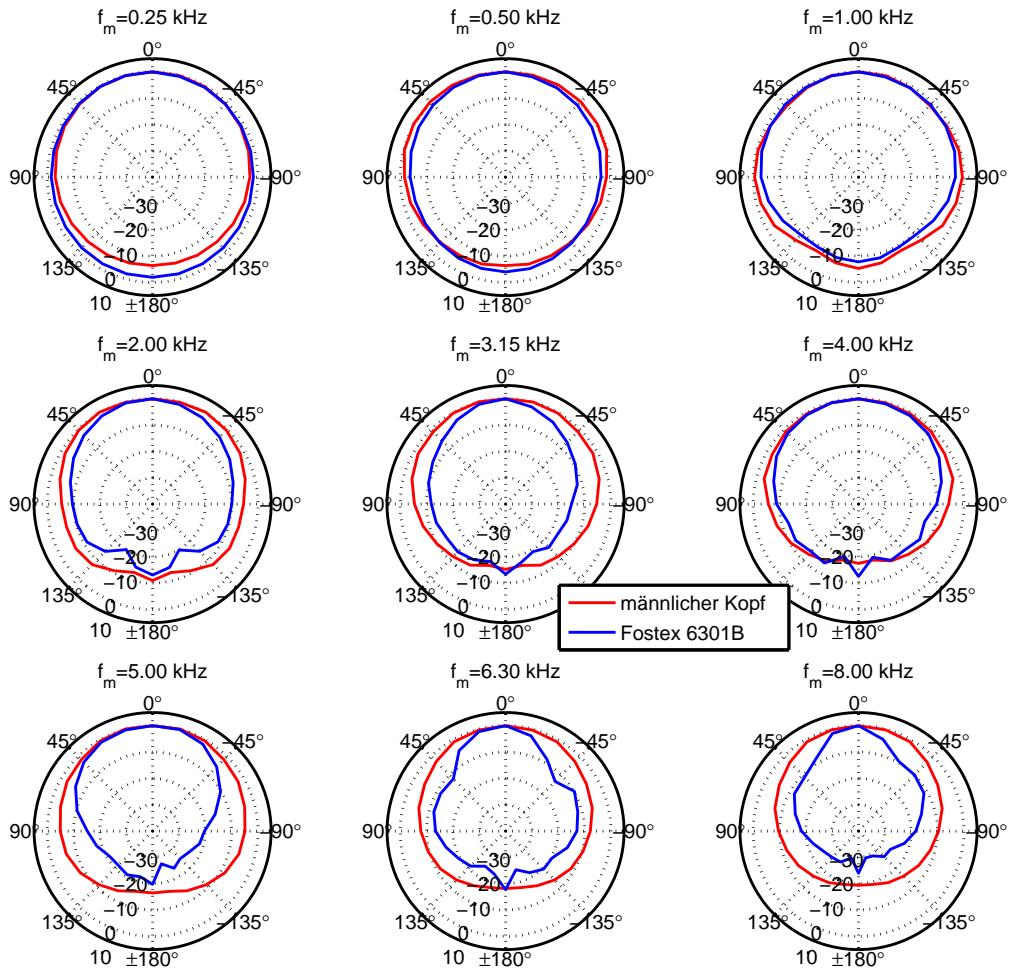
Wie bereits in Abschnitt 2.2.2 angemerkt, wurde auch bei diesen Versuchen der Drehpunkt des Lautsprechers in der Mitte der Box gewählt. Somit ist ebenfalls ein Einfluss der versetzten Quellposition als wahrnehmbare Positionsänderung anzunehmen.

### **2.3 Richtungsabhängige Abstrahlcharakteristik des menschlichen Kopfes und Vergleich mit der Abstrahlcharakteristik der Lautsprecherbox Fostex 6301B**

Für die Schätzung der Sprecherausrichtung erscheint die richtungsabhängige Abstrahlcharakteristik der Quelle eine wichtige Eigenschaft zu sein. Die Messung der Abstrahlcharakteristik des menschlichen Kopfes ist mit einem hohen Messaufwand verbunden und benötigt eine geeignete, aufwendige Messapparatur. Bisher existieren nur wenige Veröffentlichungen, aus der detaillierte Messdaten entnommen und verwendet werden können. Eine sehr frühe Untersuchung ist [DF39] aus dem Jahr 1939 bei der die Kopfabstrahlcharakteristik von einer Versuchsperson bestimmt wurde.

Deutlich umfangreicher sind die Daten in [CW02]. Hier wurden die Kopfabstrahlcharakteristiken von 40 Versuchspersonen (20 männliche, 20 weibliche) bei unterschiedlichen Sprachen (Englisch und Französisch) aufgezeichnet und können der Veröffentlichung als gemittelte Daten in Terzbändern entnommen werden. In einem reflexionsarmen Raum wurde Sprache mit einem dreh- und schwenkbaren Mikrofonarray für unterschiedliche Azimuths und Elevationen in einem Abstand von 1 m aufgezeichnet und ausgewertet. Die Sprecher sollten pro Aufnahme einen beliebigen 40 s langen Text sprechen, wobei es nicht erforderlich war, dass der Text immer identisch war. Aufgezeichnet wurden die Daten für die Azimuths zwischen  $0^\circ$  und  $180^\circ$  mit einer Auflösung von  $15^\circ$  sowie 9 Elevationen zwischen  $0^\circ$

## Kapitel 2. Wahrnehmung der relativen Sprecherausrichtung bei normalhörenden Probanden und Voruntersuchungen für die implementierten Algorithmen

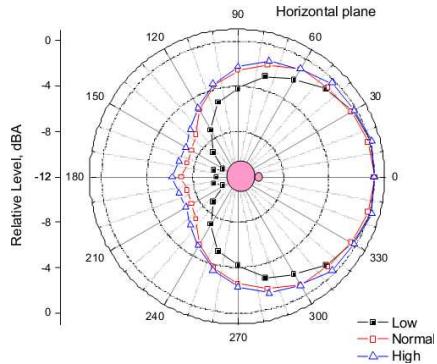


**Abb. 2.6:** Richtungsabhängige Abstrahlcharakteristiken in Terzbändern mit einer Richtungsauflösung von  $15^\circ$  in der Horizontalebene: Vergleich zwischen den gemittelten Abstrahlcharakteristiken von 20 männlichen Sprechern (nach [CW02]) und der hochkant aufgestellten Lautsprecherbox Fostex 6301B.

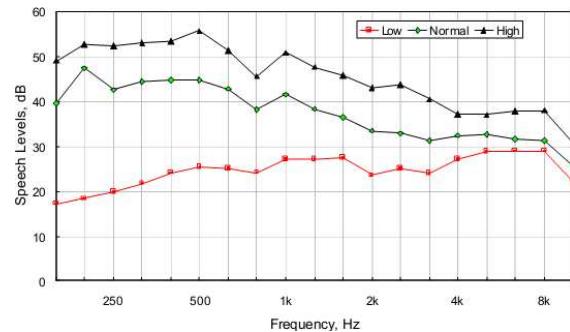
und  $347^\circ$ . Im folgenden wird nur auf die Abstrahlcharakteristik bei einer Elevation von  $0^\circ$  eingegangen. Für die folgenden Betrachtungen ist diese Einschränkung ausreichend, da in erster Näherung davon ausgegangen wird, dass Sprecher und Hörer sich auf gleicher Höhe gegenüberstehen. Bei den Auswertungen zeigte sich u. a., dass sich für die Abstrahlcharakteristiken zwischen männlichen und weiblichen Sprechern keine signifikanten Unterschiede ergaben.

Abbildung 2.6 zeigt die Polardiagramme der Abstrahlcharakteristik der gemittelten männlichen Sprecher für 9 Terzbänder von 250 Hz bis 8 kHz. Die Richtdiagramme des Kopfes zeigen keine ausgeprägten Nebenkeulen und verlaufen relativ gleichmäßig. Bei einigen Frequenzen können kleine Einbrüche der Richtwirkung bei ca.  $170^\circ$  und  $190^\circ$  erkannt werden, die durch destruktive Interferenzen entstehen. Bei  $180^\circ$  addieren sich die um den Kopf gebeugten Schallwellen konstruktiv, so dass sich hier eine leichte Nebenkeu-

## Kapitel 2. Wahrnehmung der relativen Sprecherausrichtung bei normalhörenden Probanden und Voruntersuchungen für die implementierten Algorithmen



(a) Abstrahlcharakteristik des mittleren A-gewichteten Schalldruckpegels eines männlichen Sprechers für mittlere, laute und leise Sprache. (aus [CW02])



(b) Sprachspektren für mittlere, laute und leise Sprache eines männlichen Sprechers. (aus [CW02])

**Abb. 2.7:** Vergleich der Abstrahlcharakteristik und der Sprachspektren eines männlichen Sprechers für mittlere, laute und leise Sprache. Für leise Sprache weicht die Abstrahlcharakteristik signifikant von denen der mittleren und lauten Sprache ab. Die Spektren für mittlere und laute Sprache verlaufen annähernd parallel und sind deutlich tiefenbetonter als das Spektrum für leise Sprache.

le ausbildet. Bis 1 kHz beträgt die Dämpfung nach hinten ca. 8 dB und steigt zu höheren Frequenzen bis auf -20 dB an.

Zusätzlich wird in [CW02] gezeigt, dass ein Zusammenhang zwischen der Lautstärke des Sprechers und der Abstrahlcharakteristik bestehen kann. Abbildung 2.7(a) zeigt die Richtcharakteristik als mittleren A-gewichteten Schalldruckpegel eines männlichen Sprechers. Für laute und mittlere Sprache ergeben sich keine signifikanten Unterschiede, außer bei 215°, welches als Messfehler interpretiert wurde. Für leise Sprache weicht die Richtcharakteristik signifikant ab. In Abbildung 2.7(b) sind die entsprechenden Spektren in Abhängigkeit von den betrachteten Sprachpegeln dargestellt. Auch hier weicht das Spektrum für leise Sprache signifikant von den anderen Pegeln ab und weist im Gegensatz zu den anderen Sprachpegeln eine Hochpasscharakteristik auf. Die Spektren für laute und leise Sprache verlaufen annähernd parallel. Aufgrund des Lombard-Effekts [EL11] wäre hier zu erwarten gewesen, dass die Spektren für laute Sprache hochfrequenter werden, denn Lombard hat gezeigt, dass Menschen die Tonhöhe bei lauterer Sprache anheben.

Der beobachtete Effekt der pegelabhängigen Richtcharakteristik kann hier mithilfe der durchgeführten A-Bewertung in Kombination mit dem tief-frequenten Spektrum bei mittlerer und lauter Sprache erklärt werden. Die A-Bewertung gewichtet tiefe Frequenzen weniger, so dass deren Einfluss auf die Richtcharakteristik sinkt. Die mittleren Frequenzen zwischen 1

## **Kapitel 2. Wahrnehmung der relativen Sprecherausrichtung bei normalhörenden Probanden und Voruntersuchungen für die implementierten Algorithmen**

---

und 4 kHz fließen dementsprechend stärker bei der Richtcharakteristik ein. Während der Kopf bei tiefen Frequenzen omnidirektionaler ist als bei hohen Frequenzen, ist das dargestellte, direktionale Richtdiagramm für leise Sprachpegel nachvollziehbar. Als weitere mögliche Erklärung ist es denkbar, dass bei höheren Pegeln der Kopf beim Sprecher stärker zu Schwingungen angeregt wird und er somit als Quelle wirkt und auch nach hinten mehr Schall abstrahlt. Zusätzlich verändert sich die Artikulation in Abhängigkeit der Sprachlautstärke und könnte Einflüsse auf die Richtcharakteristik haben. Diese Untersuchung wurde allerdings nur für einen männlichen Sprecher durchgeführt und kann somit nur als Indiz für den Einfluss der Sprachlautstärke auf die Abstrahlcharakteristik gewertet werden.

Für die durchgeführten Hörversuche in Kapitel 3, aber auch für die Evaluation der entwickelten Signalverarbeitungsalgorithmen in Kapitel 4, wurden natürliche Sprecher als Quelle verwendet. Um den Einfluss der Sprecherausrichtung auf synthetische Signale, wie z. B. Rauschen und harmonische Tonkomplexe, zu betrachten, wurden diese Signale von einer Lautsprecherbox Fostex 6301B abgespielt und aufgezeichnet. Der Einsatz dieses Lautsprechers bei den Versuchen ist legitim, sofern die Richtcharakteristik mit der des menschlichen Kopfes vergleichbar ist.

Zur Bestimmung der Abstrahlcharakteristik der Lautsprecherbox Fostex 6301B wurden Aufnahmen in einem reflexionsarmen Raum (siehe Abbildung 3.2) durchgeführt und ausgewertet. Die Lautsprecherbox wurde mit einem Stativ auf einem Drehteller in einer Höhe von 1,2 m aufgebaut. Dabei wurde darauf geachtet, dass sich die Drehachse des Lautsprechers auf Höhe der Lautsprechermembran befindet. Somit dreht sich der Lautsprecher annähernd um den Bereich der Signalerzeugung. Dies sollte Einflüsse auf die Messung durch eine geänderte Position der Quelle reduzieren. Als Quellsignale wurden Sinussweeps mit quadratisch ansteigender Phase von 100 Hz bis 10 kHz bei einer Abtastrate von  $f_s = 44100$  Hz verwendet und von einem *Behringer ECM 8000* Mikrofon in einem Abstand von 1,5 m aufgezeichnet.

Neben der Abstrahlcharakteristik des menschlichen Kopfes enthalten die Richtdiagramme in Abbildung 2.6 zusätzlich die gemessenen Abstrahlcharakteristiken der verwendeten Lautsprecherbox. Für Frequenzen bis 1 kHz zeigen sich nur kleine Abweichungen zwischen den Richtdiagrammen des Lautsprechers und des mittleren männlichen Kopfes. Zu höheren Frequenzen weist die Abstrahlcharakteristik deutliche Einbrüche durch destruktive

## **Kapitel 2. Wahrnehmung der relativen Sprecherausrichtung bei normalhörenden Probanden und Voruntersuchungen für die implementierten Algorithmen**

---

Interferenzen auf. Die Abweichung zwischen den Richtdiagrammen beträgt in diesen Frequenzen ca. 5-8 dB. Ab 5 kHz sind die Kurven deutlich unsymmetrischer. Diese Abweichungen sind vermutlich auf Messfehler und äußere Einflüsse zurückzuführen. Die Abbildungen zeigen, dass die Lautsprecherbox vor allem für tiefe Frequenzen bis hinein in den Hauptsprachbereich, in erster Näherung eine zum menschlichen Kopf vergleichbare Abstrahlcharakteristik aufweist. Zu hohen Frequenzen ab 5 kHz wird der Lautsprecher direktiver. Gesamt betrachtet erscheint die Lautsprecherbox Fostex 6301B für Aufnahmen und Tests zur Untersuchung der Sprecherausrichtung unter Verwendung von synthetischen Signalen ausreichend zu sein.

## **2.4 Einfluss der Sprecherausrichtung auf die Raumimpulsantwort**

Um das Verhalten und die Einflüsse unterschiedlicher Sprecherausrichtungen zu untersuchen, wurde die entsprechende Schallausbreitung in einem zweidimensionalen Geräuschfeld simuliert. Hierfür wurde die Software *Lambda 2.0* [RB09] verwendet, welche die 2D Schallausbreitung mittels der *Transmission-Line-Matrix-Methode, TLM* [JB71] berechnet.

In der Simulation wurden die Köpfe von Sprecher und Hörer vereinfacht als Kreise mit einem Durchmesser  $d = 0,18$  m und einem Reflexionsfaktor der Kopfberandung von  $r = 0,1$  modelliert. Als Quellsignal dienten Sinussignale mit der Frequenz  $f = 1000$  Hz und  $f = 8000$  Hz um den frequenzabhängigen Einfluss abzuschätzen. Bei den Simulationen wurde das jeweilige Schallfeld in Abhängigkeit von der Sprecherausrichtung von  $0^\circ$  bis  $180^\circ$  in  $45^\circ$  Schritten berechnet. Die Drehung der Sprecherpositon erfolgte so, dass die Quelle sich immer an einem fixen Punkt im Abstand von 0,75 m zum Kopfmittelpunkt des Hörers befand.

Um den Einfluss der Sprecherausrichtung unter unterschiedlichen Bedingungen zu betrachten, wurden folgende einfache Szenarien geschaffen:

- **Schallausbreitung im Freifeld:** Die Schallausbreitung erfolgt ohne Berandung im freien Schallfeld. In diesem Szenario sollen Einflüsse der Sprecherausrichtung auf die ITD, ILD betrachtet werden. Ebenso sollen Effekte der direktonalen Quelle ersichtlich werden.
- **Schallausbreitung im halligen Raum:** Der simulierte Raum (LxB: 1,5 m x 1,5 m) wurde an drei Seiten mit einem Reflexionsfaktor von  $r = 1$  berandet, während der Raum hinter dem Sprecher unberandet war.

## Kapitel 2. Wahrnehmung der relativen Sprecherausrichtung bei normalhörenden Probanden und Voruntersuchungen für die implementierten Algorithmen

---

Diese Vereinfachung wurde gewählt um die Komplexität des resultierenden Schallfeldes zu reduzieren. Der Abstand der Quelle zur hinteren Wand beträgt 0,375 m und zu den Seitenwänden jeweils 0,75 m. In diesem Szenario sollen Einflüsse der Sprecherausrichtung auf die resultierenden Reflexionen erkennbar sein.

Obwohl die Simulationen stark vereinfachten Modellen entsprechen und die Dimension des halligen Raumes für reale Situationen unterdimensioniert ist, ergeben sich aus ihnen einige Erkenntnisse, die für die Entwicklung der Algorithmen zur Schätzung der Sprecherausrichtung relevant sein können. Im Folgenden werden die einzelnen Szenarien betrachtet und die Ergebnisse und Erkenntnisse beschrieben.

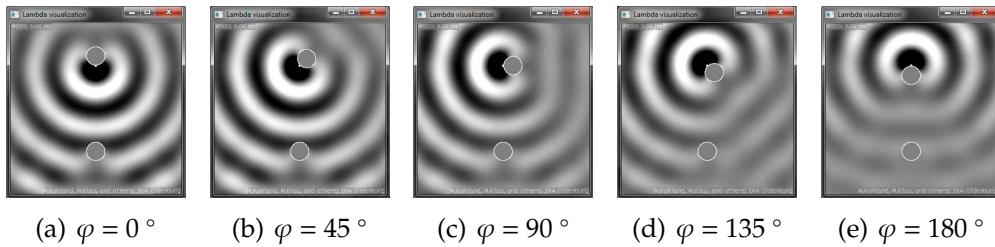
### Schallausbreitung im Freifeld

Die Abbildungen 2.8(a) bis 2.8(e) zeigen eine Momentaufnahme der Schallausbreitung im Freifeld nach  $\approx 8,06$  ms für die Frequenz  $f = 1000$  Hz. Die sich ausbreitenden Schallwellen sind farblich kodiert, wobei helle Bereiche Maxima und dunkle Minima der momentanen Amplitude entsprechen. Der Kontrast zwischen Minima und Maxima wird mit zunehmenden Abstand zur Quelle immer geringer, wodurch die entfernungsabhängige Dämpfung der Amplituden erkennbar ist. Die Berandungen der kreisförmigen Köpfe sind als weiße Linien dargestellt. Die folgenden Betrachtungen erfolgen unter der Annahme, dass der Hörer direkt in Richtung der Quelle blickt und sich die Ohren bei  $\pm 90^\circ$  befinden.

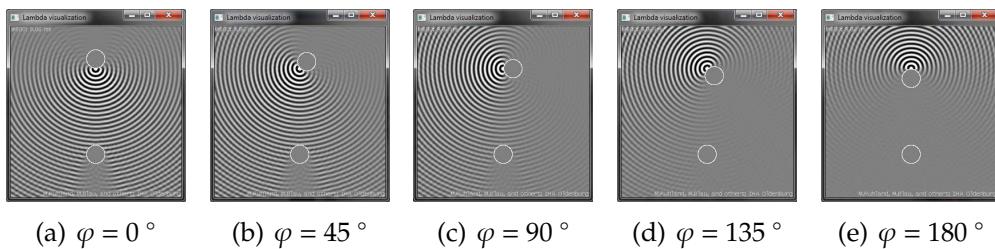
In den Abbildungen ist die direktionale Abstrahlung des Sprecherkopfes mit einer Hauptkeule in Blickrichtung deutlich erkennbar. Diese Richtcharakteristik entspricht in erster Näherung der gemessenen Richtcharakteristik aus Abbildung 2.6 für die betrachtete Frequenz. Hinter dem Hörerkopf zeichnet sich ein deutlicher Schallschatten ab, welcher bei seitlichem Schalleinfall für die interauralen Pegeldifferenzen verantwortlich ist (vgl. Abschnitt 1.2.1). Für die Sprecherausrichtungen  $0^\circ$  und  $180^\circ$  ergeben sich erwartungsgemäß keine interauralen Laufzeit- und Pegelunterschiede an den Ohren des Hörers. Die Schallwellen erreichen aufgrund der Symmetrie der simulierten Situation die Ohren mit identischem Pegel und identischer Laufzeit. Im Gegensatz zu den Beobachtungen von Neuhoff (siehe Abschnitt 2.2.3) sind für die Sprecherausrichtungen von  $45^\circ$  bis  $135^\circ$  keine bis sehr geringe Laufzeitunterschiede erkennbar. Ebenso zeichnen sich nur kleine interaurale Pegelunterschiede ab. Die deutlichsten interauralen Unterschiede ergeben sich für eine Sprecherausrichtung von  $135^\circ$ . Erkennbar an den leicht versetzten Maxima rechts und links neben dem Hörer.

Die Abbildungen 2.9(a) bis 2.9(e) zeigen eine Momentaufnahme der Schall-

## Kapitel 2. Wahrnehmung der relativen Sprecherausrichtung bei normalhörenden Probanden und Voruntersuchungen für die implementierten Algorithmen



**Abb. 2.8:** Zweidimensionale Simulation der Schallausbreitung im Freifeld in Abhängigkeit der Sprecherausrichtung  $\varphi$  bei einer Frequenz  $f = 1000$  Hz. Im Raum befinden sich zwei kreisrunde Trennflächen als Simulation des Sprechers (oben) und des Empfängers (unten).



**Abb. 2.9:** Zweidimensionale Simulation der Schallausbreitung im Freifeld in Abhängigkeit der Sprecherausrichtung  $\varphi$  bei einer Frequenz  $f = 8000$  Hz. Im Raum befinden sich zwei kreisrunde Trennflächen als Simulation des Sprechers (oben) und des Empfängers (unten).

ausbreitung im Freifeld nach  $\approx 8,06$  ms für die Frequenz  $f = 8000$  Hz. Die Richtcharakteristik entspricht in erster Näherung der gemessenen Richtcharakteristik aus Abbildung 2.6 für die betrachtete Frequenz. Auch hier ergeben sich erwartungsgemäß keine interauralen Pegel- und Laufzeitunterschiede für die Sprecherausrichtungen  $0^\circ$  und  $180^\circ$ . Für die Sprecherausrichtungen von  $45^\circ$  bis  $135^\circ$  sind geringe Laufzeitunterschiede erkennbar, wobei diese für  $135^\circ$  am stärksten ausgeprägt sind. Interaurale Pegeldifferenzen rechts und links vom Hörerkopf sind hauptsächlich für die Sprecherausrichtungen von  $90^\circ$  und  $135^\circ$  erkennbar.

Bei der Betrachtung der Simulationen über alle Sprecherausrichtungen zeigt sich für beide Frequenzen, dass die Amplitude der ankommenden Schallwelle mit steigender Sprecherausrichtung sinkt. Die Dämpfung ist abhängig von der Frequenz. Je höher die Frequenz und die Sprecherausrichtungen desto höher ist die Dämpfung. Dieser Zusammenhang ergibt sich direkt aus der frequenzabhängigen Richtcharakteristik des menschlichen Kopfes, welcher eine Tiefpasscharakteristik aufweist (vgl. Abschnitt 2.3). Die interauralen Pegeldifferenzen können ebenfalls mit der Richtcharakteristik der Quelle erklärt werden. Bezogen auf die Quellposition ergeben sich unterschiedliche Winkel zu den Positionen der Ohren des Hörers und somit unterschiedliche Pegel.

## **Kapitel 2. Wahrnehmung der relativen Sprecherausrichtung bei normalhörenden Probanden und Voruntersuchungen für die implementierten Algorithmen**

---

Die geringen Laufzeitunterschiede können durch Beugungseffekte am Kopf der Quelle beschrieben werden. Speziell in den Abbildung 2.8(d) und 2.9(d) für die Sprecherausrichtung von  $135^\circ$  ist ersichtlich, dass die Schallwellen zum rechten Ohr des Hörers einen etwas längeren Weg um den Kopf des Sprechers haben als zum linken Ohr. Diese Laufzeitunterschiede sind aber deutlich geringer als von Neuhoff beschrieben. Auch wenn dieser eine Lautsprecherbox mit anderer Richtcharakteristik in seinen Experimenten verwendet hat, scheint es, als resultiert die Laufzeitdifferenz aus der veränderten Quellposition beim Drehen des Lautsprechers um seine Mittelachse (vgl. Abschnitt 2.2.3).

### **Schallausbreitung im halligen Raum**

Die Abbildungen 2.10(a) bis 2.10(e) zeigen eine Momentaufnahme der Schallausbreitung im halligen Raum nach  $\approx 5,12$  ms für die Frequenz  $f = 1000$  Hz. Zum gewählten Zeitpunkt erreicht die erste seitliche Reflexion bei einer Sprecherausrichtung von  $0^\circ$  den Kopf des Hörers. Anders als im Freifeld ergeben sich auf Grund der reflektierenden Wände komplexere Wellenmuster. Beim Vergleich zwischen der Ausbreitung der Schallwellen im Freifeld und der Ausbreitung des Direktschalls im halligen Raum zeigen sich keine Unterschiede.

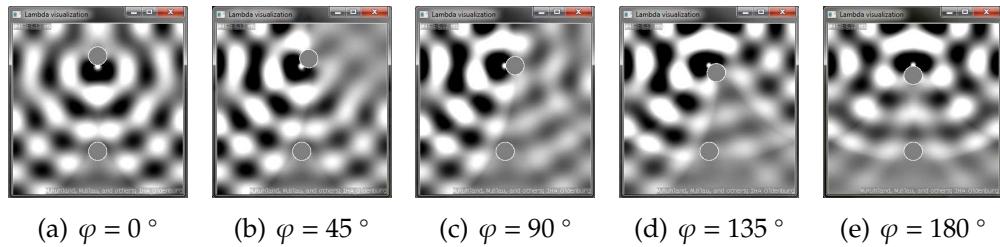
Aufgrund der Reflexionen im halligen Raum sind interaurale Unterschiede des Schalldrucks deutlich erkennbar. Für die Sprecherausrichtungen von  $45^\circ$  bis  $135^\circ$  ist der Schalldruck auf der Ohrseite, die dem Mund der Quelle am nächsten ist, sichtbar höher als auf der Gegenseite. Bei  $0^\circ$  und  $180^\circ$  ergeben sich so erwartungsgemäß keine Schalldruckunterschiede. Zusätzlich ist zu erkennen, dass der Pegel an jeder Ohrseite in Abhängigkeit vom Betrag der Sprecherausrichtung sinkt.

Weiterhin sind zwischen den ersten frühen Reflexionen Laufzeitunterschiede erkennbar. Zuerst trifft die Reflexion von der Wand, in die der Sprecher blickt, am näher gelegenen Empfängerohr ein. Die reflektierte Welle der gegenüberliegenden Wand trifft deutlich später am näher gelegenen Ohr ein, da diese Welle auf dem Weg zur reflektierenden Wand um den Kopf herumgebeugt wird. Beide betrachteten Reflexionen erzeugen aufgrund des schrägen Einfallswinkel zusätzlich einen interauralen Laufzeitunterschied.

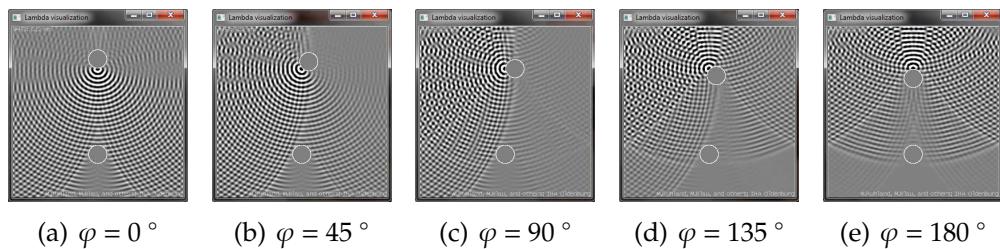
Die Abbildungen 2.11(a) bis 2.11(e) zeigen eine Momentaufnahme der Schallausbreitung im halligen Raum nach ca.  $5,12$  ms für die Frequenz  $f = 8000$  Hz. Zum gewählten Zeitpunkt erreicht die erste seitliche Reflexion bei einer Sprecherausrichtung von  $0^\circ$  den Kopf des Hörers. Die Richtcharakteristik entspricht in erster Näherung der gemessenen Richtcharakteristik aus Abbildung 2.6 für die betrachtete Frequenz. Wie zu erwarten, bestätigen

## Kapitel 2. Wahrnehmung der relativen Sprecherausrichtung bei normalhörenden Probanden und Voruntersuchungen für die implementierten Algorithmen

---



**Abb. 2.10:** Zweidimensionale Simulation der Schallausbreitung in einem halligen Raum in Abhängigkeit der Sprecherausrichtung  $\varphi$  bei einer Frequenz  $f = 1000$  Hz. Im Raum befinden sich zwei kreisrunde Trennflächen als Simulation des Sprechers (oben) und des Empfängers (unten).

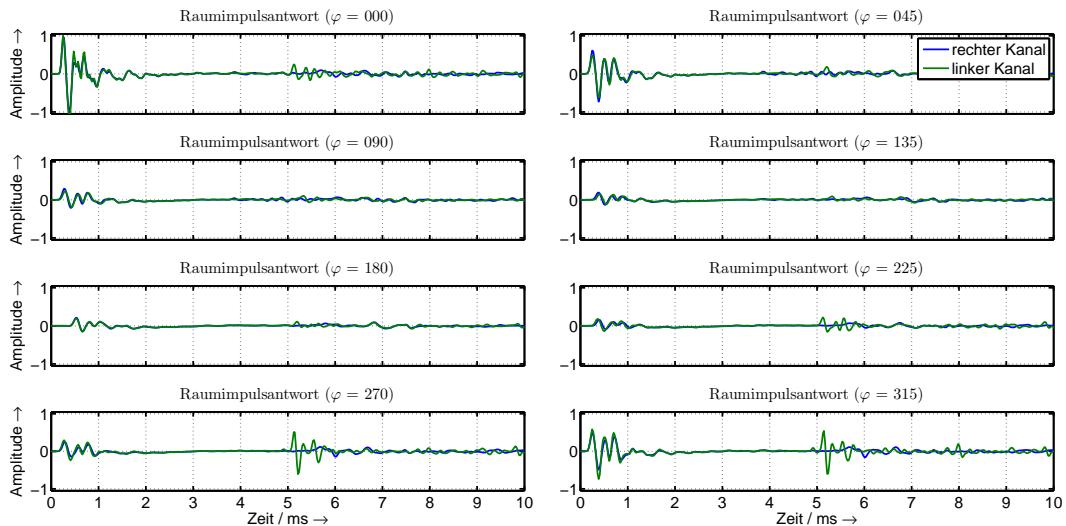


**Abb. 2.11:** Zweidimensionale Simulation der Schallausbreitung in einem halligen Raum in Abhängigkeit der Sprecherausrichtung  $\varphi$  bei einer Frequenz  $f = 8000$  Hz. Im Raum befinden sich zwei kreisrunde Trennflächen als Simulation des Sprechers (oben) und des Empfängers (unten).

sich hier die bisherigen Erkenntnisse in Bezug auf die interauralen Laufzeit- und Pegeldifferenzen.

Um die Ergebnisse und Erkenntnisse aus den Simulationen mit einer realen Situation zu vergleichen, wurden Raumimpulsantworten in einem halligen Raum aufgenommen. Für diese Analyse wurde eine Fostex 6301B Lautsprecherbox in einem Abstand von einem Meter vor einem Neumannkunstkopf mit KM 83 spezial Gehörgangsmikrofonen aufgestellt. Quelle und Empfänger wurden frontal ausgerichtet und in einer Höhe von  $\approx 1,7$  m so aufgestellt, dass sich die Quelle in der Höhe der Ohren des Kunstkopfes befanden. Gemessen wurden die Impulsantworten für die Sprecherausrichtung von  $0^\circ$  bis  $315^\circ$  in  $45^\circ$  Schritten. Die Drehachse des Lautsprechers wurde in der Höhe der Lautsprechermembran positioniert, so dass der Lautsprecher möglichst um das Zentrum der Quellerzeugung rotiert. Somit kann sichergestellt werden, dass die Quellposition für alle Messungen räumlich annähernd stationär war. Aufgenommen wurden die Daten in einem Bürraum ( $B \times L \times H: 7,9$  m  $\times$   $3,9$  m  $\times$   $3,15$  m) mit einer mittleren Nachhallzeit von  $0.45$  s. Ein Grundriss des Raumes ist in Abbildung 3.3 dargestellt. Die Position von Quelle und Empfänger kann aus der Abbildung entnommen werden. Anders als in der Simulation wurde hier keine symmetrische Anordnung gewählt.

## Kapitel 2. Wahrnehmung der relativen Sprecherausrichtung bei normalhörenden Probanden und Voruntersuchungen für die implementierten Algorithmen



**Abb. 2.12:** Darstellung der ersten 10 ms der gemessenen Raumimpulsantworten in einem Bürraum in Abhängigkeit der Sprecherausrichtungen  $\varphi$ . Die Direktschallanteile weisen sehr geringe interaurale Laufzeit- und Pegeldifferenzen auf. Die Amplituden und der Frequenzgehalt der Direktschallanteile weisen eine Abhängigkeit von der Sprecherausrichtung auf. Deutliche Unterschiede sind in den ersten frühen Reflexionen zu erkennen.

In Abbildung 2.12 werden die ersten 10 ms der Raumimpulsantworten vom linken und rechten Kunstkopfmikrofon für die gemessenen Sprecherausrichtungen  $\varphi$  gezeigt. Dargestellt sind somit nur der Direktschallanteil und die erste frühe Reflexion, da hier nach den bisherigen Untersuchungen die größten Unterschiede in der Impulsantworten zu erwarten sind.

Wie aus den vorangegangen Untersuchungen zu erwarten war, zeigen sich für alle Sprecherausrichtungen vernachlässigbar kleine interaurale Laufzeit und Pegeldifferenzen beim Direktschallanteil. Die Kurven sind innerhalb der ersten 4 ms nahezu deckungsgleich. Im Gegensatz dazu unterscheiden sich die Direktschallanteile in Abhängigkeit von der Sprecherausrichtung. Mit steigendem Betrag der Sprecherausrichtung sinkt ebenfalls die Amplitude des Direktschalls. Zusätzlich verlieren die Kurven an Feinstruktur, was hier auf geringere Anteile hoher Frequenzen hindeutet. Beide Effekte können durch die Richtcharakteristik der Quelle beschrieben werden. Mit steigendem Winkel sinkt die frequenzabhängige Abstrahlleistung der Quelle. Zusätzlich weiß die Quelle zu höheren Frequenzen eine stärkere Direktivität aus. Beides zusammen führt zu den beschriebenen Effekten.

Deutlich ausrichtungsabhängige Unterschiede der Impulsantworten zeigen sich dagegen bei den ersten frühen Reflexionen. Diese liegen bei etwa 5 ms und ergeben sich durch die nächstliegende Wand. Hierbei handelt es sich um eine harte Wand mit Rauhfaser Tapete an der ein Whiteboard befestigt ist, so dass von einer hartreflektierenden Wand auszugehen ist. Auf der zu

## **Kapitel 2. Wahrnehmung der relativen Sprecherausrichtung bei normalhörenden Probanden und Voruntersuchungen für die implementierten Algorithmen**

---

dieser Wand zugewandten Kopfseite (linker Kanal) ist die erste Reflexion erwartungsgemäß am stärksten ausgeprägt. Unabhängig von der Sprecherausrichtung trifft die erste Reflexion nahezu zeitgleich ein, da sich die Quelle immer am selben Punkt im Raum befindet. Die Amplitude ist wie der Direktschallanteil deutlich von der Ausrichtung der Quelle abhängig. Hier zeigt sich, dass die Amplitude maximal ist, wenn die Quelle direkt in die Richtung der reflektierenden Wand ausgerichtet ist, in diesem Beispiel bei  $\varphi = 315^\circ$ .

Deutliche Unterschiede zeigen sich bei dem Vergleich der Amplituden der ersten Reflexion zwischen den beiden Kopfseiten. Die Amplitude des von der nächstgelegenen Wand abgewandten Ohres ist aufgrund der Abschattung des Kopfes deutlich reduziert. So ist die erste Reflexion hier nur sehr schwach ausgeprägt und aufgrund der Laufzeitdifferenz verzögert.

## **2.5 Akustische Parameter zur Schätzung der Sprecherausrichtung**

Aus den bisherigen Betrachtungen und Untersuchungen können einige mögliche Parameter für die Schätzung der Sprecherausrichtung bestimmt werden. Im Folgenden werden diese und weitere zusammengefasst und die mögliche Umsetzung in Algorithmen für die Verwendung in Hörgeräten betrachtet.

### **Richtungsabhängige Pegeldifferenzen**

Es hat sich gezeigt, dass sich aus unterschiedlichen Sprecherausrichtungen bereits monaural Pegeldifferenzen an den Ohren ergeben. Diese Unterschiede wurden in [KTNM10] ebenfalls festgestellt und weiter untersucht. Hier hat sich im reflexionsarmen Raum gezeigt, dass bei einem Abstand von 1,2 m zwischen Sprecher und Hörer eine maximale Pegeldifferenz von ca. 13 dB ergibt und für einen Abstand von 2,4 m eine Differenz von 10 dB. Diese maximale Differenz ergibt sich zwischen einer Sprecherausrichtung von  $0^\circ$  und  $180^\circ$  und resultiert aus der direktionalen Abstrahlcharakteristik des menschlichen Kopfes. Es ist davon auszugehen, dass diese Differenz bei den Versuchen in Abschnitt 2.2.1 von den normalhörenden Probanden für die vorne-hinten-Unterscheidung ausgenutzt wurde und so zu den guten Ergebnissen bei diesen zwei Sprecherausrichtungen führen.

Für den Einsatz in Hörgeräten ist diese Information alleine nicht ausreichend, da der Pegel einer Quelle abhängig vom Abstand zum Hörer ist und von der Schallleistung der Quelle abhängt. Beide Informationen stehen in

## **Kapitel 2. Wahrnehmung der relativen Sprecherausrichtung bei normalhörenden Probanden und Voruntersuchungen für die implementierten Algorithmen**

---

einem Hörgerät nicht ausreichend genau zur Verfügung.

### **Frequenzabhängige Übertragungsfunktion**

Die in Abbildung 2.6 dargestellten frequenzabhängigen Abstrahlcharakteristiken des menschlichen Kopfes zeigen bei tiefen Frequenzen ein direktionales Verhalten, welches mit steigender Frequenz zunehmend direktiver wird. Daraus lässt sich schließen, dass sich die Klangfarbe des Sprechers mit zunehmenden Betrag der Sprecherausrichtung hin zu tiefen Frequenzen verschiebt. Der Kopf wirkt wie ein richtungsabhängiger Tiefpassfilter. Allerdings ist die menschliche Sprache nichtstationär, so dass sich das Spektrum der Sprache permanent verändert und zudem noch einen sprecherabhängigen Einfluss hat. Vor allem in reflexionsarmen Räumen und in Räumen mit geringer Nachhallzeit kann die Auswertung des mittleren Frequenzabfalls der Sprache ein Indiz für die Sprecherausrichtung sein.

### **Interaurale Laufzeitdifferenz**

Einen nennenswerten Einfluss der Sprecherausrichtung auf die interaurale Laufzeitdifferenz für den Direktschall konnte nicht beobachtet werden. Eine Laufzeitdifferenz ist, wie in Abschnitt 2.4 beschrieben, nur für die Sprecherausrichtungen zwischen  $90^\circ$  und  $180^\circ$  zu erwarten. In [KTNM10] wurde ebenfalls die interaurale Laufzeitdifferenz in einem reflexionsarmen Raum untersucht. Hier zeigte sich, dass die maximalen Laufzeitdifferenzen bei einer Abtastrate  $f_s = 48000$  unterhalb des zeitlichen Auflösungsvermögens von  $20,8 \mu\text{s}$  liegen. Somit scheint die ILD ungeeignet für die Verwendung in Algorithmen für die Schätzung der Sprecherausrichtung zu sein. Allerdings hat sich gezeigt, dass bei den ersten frühen Reflexionen ILDs nachzuweisen sind.

### **Interaurale Pegeldifferenz**

Einflüsse der Sprecherausrichtung auf die interauralen Pegeldifferenzen konnten in den Simulationen aus Abschnitt 2.4 sowohl im reflexionsarmen Raum als auch in einem halligen Raum nachgewiesen werden. Diese Differenzen können mithilfe der richtungsabhängigen Abstrahlcharakteristik des menschlichen Kopfes erklärt werden. Auch diese Eigenschaft wurde in [KTNM10] anhand von gemessenen Impulsantworten im reflexionsarmen Raum bestimmt. Hier zeigt sich, dass die ILD zwei dB oder weniger beträgt und somit in der Größenordnung der kleinsten wahrnehmbaren Pegeldifferenz liegt. Für Hörgeräte erscheint die Verwendung solch geringer Pegeldifferenzen ungeeignet, da die Bestimmung der ILD nur mit sehr gut kalibrierten Hörgerätemikrofonen funktionieren kann. Deutlich wird die Kalibrier-Problematik bereits bei der DIN EN 60:118-7:2005, der Norm für Hörgerätemessboxen und Prüfgeräte. Hier heißt es „Die Schallquelle

## Kapitel 2. Wahrnehmung der relativen Sprecherausrichtung bei normalhörenden Probanden und Voruntersuchungen für die implementierten Algorithmen

---

(für reine Töne) in Verbindung mit einem Regelmikrofon mit bekanntem Druckübertragungsmaß muss in der Lage sein, am Prüfpunkt die benötigten Schalldruckpegel zwischen 50dB und 90dB in einer Schrittweite von wenigstens 5dB im Frequenzbereich von 200Hz bis 2000Hz mit einer Grenzabweichung von +/-1.5dB und im Frequenzbereich von 2000Hz bis 5000Hz mit einer Grenzabweichung von 2.5dB zu erzeugen.“ [IEC05]. Somit wird deutlich, dass die erforderliche Genauigkeit der Hörgerätemikrofone in einer Hörgerätemessbox aufgrund der Toleranz nur schwer kontrollierbar und anpassbar ist.

### Direktschall und frühe Reflexionen

Die bisher größten Unterschiede in Abhängigkeit von der Sprecherausrichtung haben sich im Zusammenhang mit den ersten frühen Reflexionen gezeigt. Die ersten frühen Reflexionen werden direkt von der Kopfabstrahcharakteristik des Kopfes beeinflusst. An Wänden zu denen die Quelle gerichtet ist, weisen die Reflexionen höhere Amplituden auf, als zu den abgewandten Wänden. Bei einem vom Hörer abgewandten Sprecher nimmt zusätzlich die Amplitude des Direktschalls aufgrund des Kopfschattens ab. Zusätzlich zeigten sich Laufzeitunterschiede beim Eintreffen der ersten frühen Reflexionen.

Um diese Erkenntnisse für einen Hörgerätealgorithmus zur absoluten Schätzung der Sprecherausrichtung zu nutzen, müsste a-priori Wissen vorliegen. Um mindestens eine Unterscheidung zwischen einem direkt zum Hörer blickenden Sprecher und einer davon abweichenden Richtung zu ermöglichen, muss das zu erwartende Verhältnis zwischen der Amplitude des Direktschalls und der ersten Reflexion für eine Sprecherausrichtung von  $0^\circ$  bekannt sein oder geschätzt werden. Allerdings sind diese Parameter von der Raumimpulsantwort abhängig, welche von den vorliegenden Raumparametern und der Position der Quelle und des Empfängers im Raum abhängig ist. Die Schätzung der Raumimpulsantwort ist nicht trivial. Verfahren wie z. B. die *Cepstrale Mittelung* (engl. *cepstral mean*) sind aber zu ungenau um die erforderliche Schätzung der Raumimpulsantwort zu erreichen. In [SRHE09] wird dieses Verfahren für die Schätzung der Nachhallzeit  $\tau_{60}$  verwendet, weist aber bereits zu große Abweichungen zur wahren Nachhallzeit auf. Eine direkte Schätzung dieser Parameter zur Bestimmung der Sprecherausrichtung erscheint somit derzeit nicht trivial.

### Quellbreite

Eine weitere mögliche, bisher noch nicht behandelte Eigenschaft zur Schätzung der Sprecherausrichtung ist die Quellbreite. Ein Mund kann in erster Näherung als Punktquelle betrachtet werden. Unter der Annahme,

## **Kapitel 2. Wahrnehmung der relativen Sprecherausrichtung bei normalhörenden Probanden und Voruntersuchungen für die implementierten Algorithmen**

---

dass der Kopf beim Sprechen ebenfalls in Schwingungen versetzt wird, dient er somit als zusätzliche Schallquelle. Der Kopf kann aufgrund seiner Ausmaße in entsprechender Entfernung nicht mehr als Punktquelle betrachtet werden, sondern als eine räumlich ausgedehnte Quelle. Somit müsste die wahrgenommene Quellbreite in Abhängigkeit der Sprecherausrichtung variieren. Denn bei einem zugewandten Sprecher kann davon ausgegangen werden, dass die Punktquelle des Mundes dominanter als die Kopfquelle wahrgenommen wird. Für den Fall, dass der Sprecher abgewandt steht, wird die Punktquelle durch die Abschattung des Kopfs gedämpft. Das Verhältnis zwischen den Quellen müsste sich zu Gunsten der Kopfquelle verändern. Bisher sind keine Untersuchungen des Einflusses der Sprecherausrichtung auf die Quellbreite bekannt und werden im Folgenden nicht weiter untersucht.

Die bisherigen Untersuchungen haben gezeigt, dass die Ausrichtung einer direktiven Quelle große Einflüsse auf die Klangfärbung des Signals hat. Dieser Effekt lässt sich durch die ausrichtungsabhängige Abstrahlcharakteristik der betrachteten Quellen erklären. Zusätzlich wird das Verhältnis zwischen Direktschall und frühen Reflexionen sowie die Laufzeitdifferenz der frühen Reflexionen in halligen Räumen durch unterschiedliche Ausrichtungen einer direktiven Quelle beeinflusst. Diese Eigenschaften werden im Kapitel 4 für die Entwicklung von Algorithmen zur Schätzung der Sprecherausrichtung mit binaural verschalteten Hörgeräten genutzt.

---

## Kapitel 3

# Analyse der menschlichen Fähigkeit zur Schätzung der Sprecherausrichtung mit normalhörenden Probanden

Im Abschnitt 2.2 wurden bereits Ergebnisse der menschlichen Fähigkeit zur Schätzung der Sprecherausrichtung mit normalhörenden Probanden vorgestellt. Dabei wurden zwei unterschiedliche Verfahren in unterschiedlichen akustischen Umfeldern betrachtet. Zum einen erfolgten die Messungen in einem reflexionsarmen Raum (Abbildung 3.2) und zum anderen in einem Raum mit einer Nachhallzeit von  $\approx 0,45$  s (Abbildung 3.3). Im reflexionsarmen Raum wurden die Hörversuche mit natürlichen Sprechern durchgeführt, die sich pro Test in eine der vorgegebenen Richtungen zu drehen hatten. Im halligen Raum wurden die Signale von einem ausgerichteten Lautsprecher wiedergegeben. Den Probanden wurden die Augen verbunden und sie mussten für jede Situation die Ausrichtung der Quelle angeben.

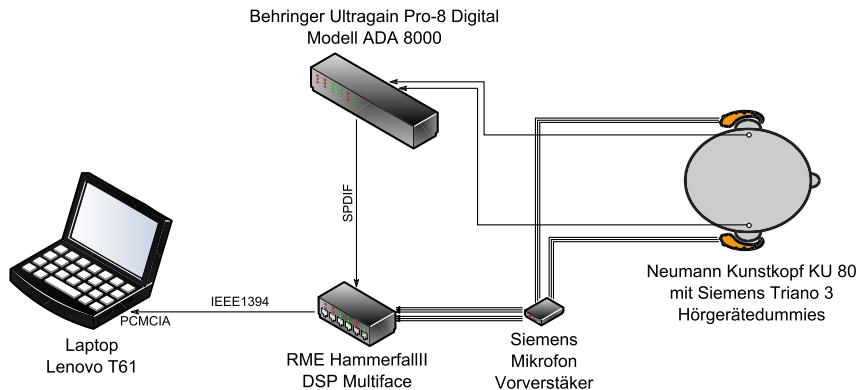
Bereits bei diesen zwei Versuchen wird deutlich, dass die Untersuchungen zur Schätzung der Sprecherausrichtung mit unterschiedlichen Verfahren durchgeführt werden können. Bei genauerer Betrachtung zeigt sich, dass bei jedem Versuchsaufbau und bei der Durchführung eine Vielzahl von äußeren Einflüssen die Ergebnisse verfälschen können. Die folgende Auflistung beschreibt einige mögliche Fehlerquellen, welche durch einen geeigneten Versuchsaufbau zu minimieren sind.

- Geräusche während der Ausrichtung der Quelle müssen vermieden werden. Hierzu zählen z.B. Geräusche der Schuhe beim Drehen eines natürlichen Sprechers oder des Stuhls, sowie die Motorgeräusche eines

möglichen Drehtellers.

- Pausen zwischen zwei Signalen müssen unabhängig von der Winkeländerung der Quelle sein, d. h. eine große Winkeländerung darf nicht wahrnehmbar länger dauern, als eine kleine.
- Die Quellsignale sollten für jeden Test identisch sein. So können natürliche Sprecher das Sprachsignal unbewusst durch unterschiedliche Aussprache oder Betonung modifizieren und dem Probanden zusätzliche Informationen liefern.
- Lautsprecher besitzen eine vom menschlichen Kopf abweichende Richtcharakteristik. Für vergleichbare Ergebnisse, sollte ein Lautsprecher gewählt werden, der eine zum Kopf vergleichbare Richtcharakteristiken aufweist. Sinnvoller kann der Einsatz eines Kunstkopfes mit Mundsimulator sein. Doch auch hier können Abweichungen durch die Schwingungseigenschaften der jeweiligen Quelle entstehen (siehe Abschnitte 2.3 und 2.4).
- Der Drehpunkt der Quelle muss mit dem Zentrum der Signalerzeugung übereinstimmen, um Positionsänderungen der Quelle im Raum zu vermeiden.
- Bewegungen des Probanden oder der Quelle während der Durchführung können zusätzliche akustische Informationen liefern.

Im Gegensatz zu den vorgestellten Experimenten in Abschnitt 2.2 wurden in dieser Arbeit Aufnahmen verwendet, die mit einem Kunstkopf aufgezeichnet wurden. Dadurch ergeben sich einige Vorteile. Die Testsignale sind für alle Probanden identisch und es können zwischen den Signalen keine Pausen entstehen, die durch die Drehung der Quelle hervorgerufen werden. Weiterhin müssen den Probanden die Augen nicht verbunden werden. Ein entscheidender Vorteil für die folgenden Tests ist die Möglichkeit der nachträglichen Modifizierung der Aufnahmen. So ist es z. B. möglich, die Pegel der Signale anzupassen, um die Lautstärke als nutzbare akustische Eigenschaft auszuschließen. Ebenso kann die Tiefpasscharakteristik der Quelle ausgeglichen werden. Auch hier ist zu erwarten, dass den Probanden eine weitere akustische Eigenschaft zur Schätzung der Sprecherausrichtung genommen wird. Die Modifizierung der Signale ist in den folgenden Experimenten sinnvoll, um für die Entwicklung möglicher Algorithmen zur Schätzung der Sprecherausrichtung mittels Hörgeräten mögliche akustische Eigenschaften zu untersuchen, die für eine algorithmische Detektion herangezogen werden können.



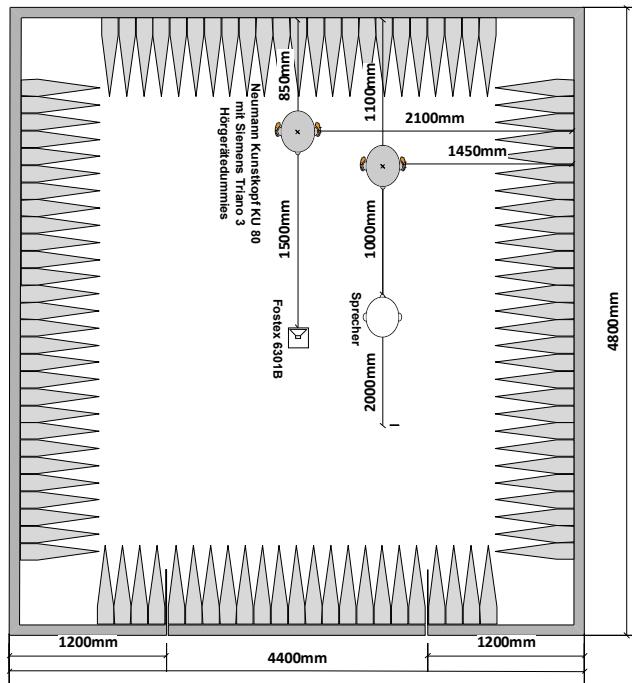
**Abb. 3.1:** Schematische Skizze des Messaufbaus zur Aufnahme des in Hörversuchen und der Algorithmen-evaluation verwendeten Audiomaterials. Die Signale der Hörgerätemikrofone und der Gehörgangsmikrofone am Kunstkopf wurden über das dargestellte Audioequipment aufgenommen und verarbeitet.

Allerdings ergeben sich mit den Kunstkopfaufnamen systematische Fehler. Als entscheidender Nachteil ist zu erwähnen, dass die HRTFs des Kunstkopfes von den HRTFs der Probanden abweichen und somit mit den Aufnahmen kein natürliches räumliches Hören erleben werden. Weiterhin wurden die Hörversuche an einem Rechner mit Kopfhörern in einem beliebigen, ruhigen Raum durchgeführt. Diese Umgebung weicht von den Aufnahmeräumen ab, so dass die entstehende Differenz zwischen den optischen und akustischen Eindrücken beim Hörversuch und einer Hörversuchssituation in einem Büro oder reflexionsarmen Raum die Beurteilung der Probanden beeinflussen kann.

### 3.1 Messaufbau und erstellte Audiodatenbank

Die Testsignale für die durchgeführten Hörversuche und für die Evaluation der entwickelten und implementierten Algorithmen wurden in einem reflexionsarmen Raum ( $L \times B \times H: 4,8 \text{ m} \times 4,4 \text{ m} \times 3,2 \text{ m}$ ) und in einem Bürroraum ( $L \times B \times H: 7,9 \text{ m} \times 3,9 \text{ m} \times 3,15 \text{ m}$ ) mit einer mittleren Nachhallzeit von  $T_{60} \approx 0,45 \text{ s}$  aufgenommen. Als Quellen dienten natürliche Sprecher (ein weiblicher, ein männlicher) und die Lautsprecherbox Fostex 6301B. Aufgenommen wurden die Signale mit dem Neumannkunstkopf KU 80, welcher mit zwei jeweils dreikanaligen Siemens Triano 3 Hörgerätedummies ausgestattet war (siehe Abbildung 2). Zusätzlich wurden die Signale der Gehörgangsmikrofone vom Typ KM 83 *spezial* mit aufgezeichnet, so dass jedes Signal als achtkanalige Aufnahme zur Verfügung stand.

Die Hörgerätemikrofone wurde über einen von der Siemens Audiologischen Technik GmbH gelieferten Mikrofonvorverstärker (ohne Bezeichnung) über

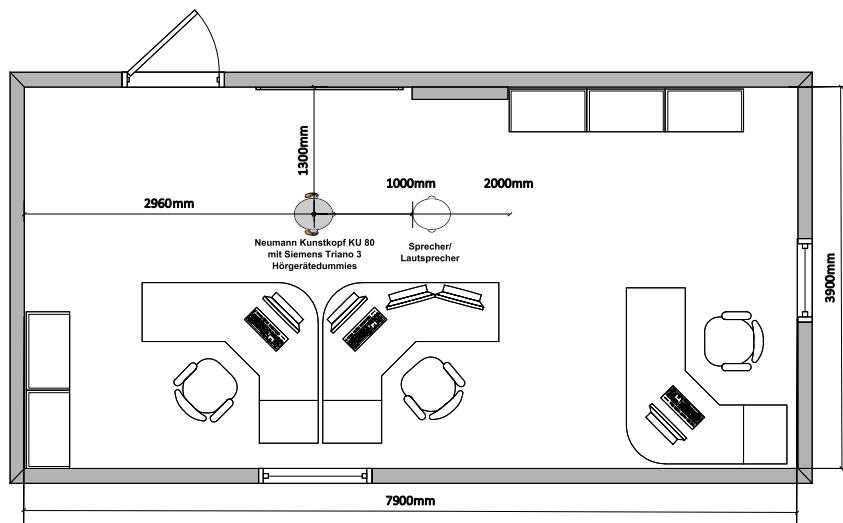


**Abb. 3.2:** Maßstabsgerechte Skizze des für die Aufnahmen genutzten reflexionsarmen Raumes.

eine *RME HammerfallII DSP Multiface* Soundkarte an ein *Lenovo T61 Laptop* per PCMCIA Schnittstelle verbunden. Die Gehörgangsmikrofone wurden über einen *Behringer Ultragain Pro-8 Digital Modell ADA 8000* Vorverstärker übertragen, welcher per TOSLINK mit dem *RME HammerfallII DSP Multiface* verbunden war. Abbildung 3.1 zeigt schematisch den beschriebenen Messaufbau mit den verwendeten Geräten.

Nach den Aufnahmen wurden die gesamten Audiodateien auf fünf Sekunden beschnitten und der Effektivwert (engl. root-mean-square, RMS) angeglichen. Die Effektivwertanpassung ist dadurch motiviert, dass die Probanden die Eigenschaft des Pegelunterschiedes der Audiosignale in Abhängigkeit von der Sprecherausrichtung nicht ausnutzen sollen. Wie in Abschnitt 2.5 beschrieben, kann diese Eigenschaft nur schwer von entsprechenden Algorithmen ausgenutzt werden. Neben der Untersuchung zur Unterscheidung der Sprecherausrichtungsbestimmung von Normalhörenden, war ein Hauptziel dieser Evaluation, mögliche akustische Eigenschaften für die algorithmische Schätzung der Sprecherausrichtung zu betrachten.

Wie bereits im Abschnitt 2 behandelt, haben der menschliche Kopf und der verwendete Lautsprecher eine richtungsabhängige Tiefpasscharakteristik die ebenfalls als akustische Eigenschaft ausgenutzt werden kann. Aus diesem Grund wurde der Frequenzgang einiger Aufnahmen angeglichen, so dass diese Eigenschaft für die Probanden ebenfalls nicht ausgewertet werden kann.



**Abb. 3.3:** Maßstabsgerechte Skizze des für die Aufnahmen genutzten halligen Büroraums mit einer mittleren Nachhallzeit von  $\approx 0,45$  s.

Die durchgeführten Modifizierungen, wie die RMS-Anpassung und der Ausgleich des Frequenzganges, beeinflussen ebenfalls das aufgenommene Sensorrauschen in den Signalen und können die Probanden bei Ihrer Bewertung der Sprecherausrichtung beeinflussen. Daher wurde zu allen Kanälen der aufgenommenen und ggf. modifizierten Signale ein normalverteiltes, unkorreliertes Rauschen mit einem SNR von -30dB addiert, um so ein gleichmäßiges Sensorrauschen zu simulieren.

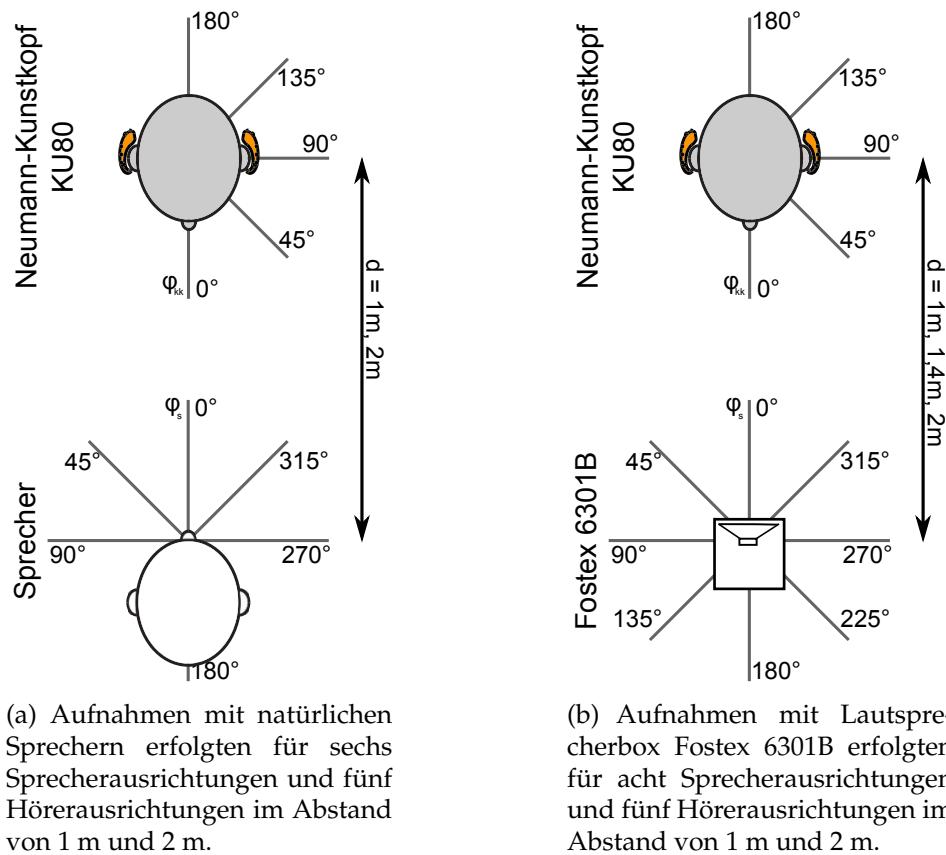
### Nutzsignale mit natürlicher Sprache

Für die Nutzsignale der Experimente mit natürlicher Sprache wurden zwei Sprecher (ein weiblicher, ein männlicher) aufgefordert, jeweils den ersten Satz von August Meißeners Prosafabel „Sonne und Wind“ in der Fassung von Johann Gottfried Herder<sup>1</sup> aufzusagen:

*Einst stritten sich die Sonne und der Wind,  
wer von ihnen beiden der Stärkere sei,  
und man ward einig, derjenige solle dafür gelten,  
der einen Wanderer, den sie eben vor sich sahen,  
am ersten nötigen würde, seinen Mantel abzulegen.*

Pro Aufnahme erfolgte eine Ausrichtung des Sprechers  $\varphi_s$  von  $0^\circ$  (frontal zum Kunstkopf),  $45^\circ$ ,  $90^\circ$ ,  $180^\circ$ ,  $270^\circ$  bis  $315^\circ$  im mathematisch positivem Drehsinn. Nach einem kompletten Durchlauf aller Richtungen wurde die Ausrichtung des Kunstkopfes  $\varphi_{kk}$  jeweils im Bereich von  $0^\circ$  (frontal zum Sprecher) bis  $180^\circ$  in  $45^\circ$ -Schritten im mathematisch positivem Drehsinn erhöht. Zusätzlich zu den jeweiligen Ausrichtungen wurden die Aufnahmen

<sup>1</sup>Quelle: <http://gutenberg.spiegel.de/buch/2017/1> (zuletzt besucht am 23. März 2012)



**Abb. 3.4:** Darstellungen der Ausrichtungen von Sprecher/Lautsprecherbox und Kunstkopf während den Aufnahmen, welche für die Durchführung der Experimente und der Evaluation der entwickelten Algorithmen verwendet wurden.

für die Abstände  $d$  von ein und zwei Metern durchgeführt. Insgesamt ergaben sich pro Raum, Sprecher und Abstand 30 quellenrichtungsabhängige Aufnahmen. Abbildung 3.4(a) zeigt schematisch die beschriebene Positionierung und Ausrichtung von Sprecher und Kunstkopf während der Aufnahmen.

Um eine möglichst räumlich stationäre Quelle pro Aufnahme zu erhalten, wurde im jeweiligen Sprecherabstand eine Positionsstange aufgestellt an der der Sprecher den Mund zu positionieren hatte. So konnten positions- und entfernungsabhängige Einflüsse während der Aufnahmen vermindert werden. Ziel war es, die Quelle (Mund) an einer möglichst konstanten Position zu belassen. Im Gegensatz dazu, hätte eine Drehung um die eigene Achse des Sprechers zu unterschiedlichen Positionen der Quelle und somit zu weiteren möglichen akustischen Einflüssen in den Aufnahmen geführt.

#### Nutzsignale mit Lautsprecherbox Fostex 6301B

Für die Aufnahmen mit, zum Teil synthetischen Lautsprechersignalen, wurde eine Fostex 6301B Lautsprecherbox auf einem Stativ angebracht. Dabei wurde darauf geachtet, dass der Drehpunkt nicht am Stativgewinde im Mit-

telpunkt der Lautsprecherbox lag, sondern in Höhe der Lautsprechermembran. Dieser Drehpunkt sollte eine Rotation möglichst nahe um das akustische Zentrum der Signalerzeugung des Lautsprechers ermöglichen. Eine räumliche Verschiebung des akustischen Zentrums bei der Ausrichtung der Quelle kann so vermindert werden (vgl. Abschnitt 2.2.2 und Abschnitt 2.4).

Pro Aufnahme erfolgte eine Ausrichtung des Lautsprechers  $\varphi_s$  von  $0^\circ$  (frontal zum Kunstkopf) bis  $180^\circ$  in  $45^\circ$ -Schritten im mathematisch positiven Drehsinn. Nach einem kompletten Durchlauf aller Richtungen wurde die Ausrichtung des Kunstkopfes  $\varphi_{kk}$  jeweils im Bereich von  $0^\circ$  (frontal zum Lautsprecher) bis  $180^\circ$  in  $45^\circ$  im mathematisch positiven Drehsinn erhöht. Abbildung 3.4(b) zeigt schematisch die Positionierung und Ausrichtung von Lautsprecherbox und Kunstkopf während der Aufnahmen.

Im Bürorraum wurde diese Prozedur für einen Abstand von einem Meter und zwei Metern zwischen Kunstkopf und Quelle durchgeführt. Im Schallarmenraum wurde eine freie Positionierung von Lautsprecher und Kunstkopf durch die feste Verteilung der im Boden verankerten Stativgewinde verhindert. Hier konnten die Aufnahmen nur im Abstand von  $d = 1.4\text{ m}$  durchgeführt werden. Insgesamt ergaben sich pro Raum, Quellsignal und Abstand 40 quellenrichtungsabhängige Aufnahmen.

Die folgenden Signale wurden für die Aufnahmen generiert und mit dem Lautsprecher abgespielt und aufgenommen:

- Sätze „Britta verleiht elf alte Bilder. Tanja kauft acht nasse Messer.“ aus dem Oldenburger Satztest [Old06].
- Synthetischer Vokal /a/: Dieses Signal wurde durch das Ausschneiden einer Periode aus einem gesprochenem Vokal /a/ eines Satzes des Oldenburger Satztestes extrahiert und periodisch fortgesetzt.
- Weißes Rauschen.
- Rauschen mit einem sprachähnlichem Leistungsdichtespektrum: erzeugt durch eine 1000-fache Überlagerung eines Satzes aus dem Oldenburger Satztest.
- Nullphasiger harmonischer Tonkomplex (HTC) mit fünf aufeinanderfolgenden Komponenten, beginnend ab 150 Hz.

## **3.2 Experimente und Durchführung**

Die Hörversuche wurden mit 21 normalhörenden Probanden (12 weibliche, 9 männliche) im Alter zwischen 20 und 34 Jahren durchgeführt. Für die Dauer der Experimente wurden die Probanden mit einem *Sennheiser HDA 200 Kopfhörer* und einem *Tascam US-144 USB-2.0-Audiointerface* ausgestattet, so dass die Versuche mit einem identischen Audioequipment ausgeführt wurden. Die Probanden wurden darauf hingewiesen, die Experimente in einer ruhigen Umgebung durchzuführen und sich dabei ausreichend Pausen einzuräumen, so dass Ermüdungseffekte reduziert werden können. Die Experimente erfolgten über eine Online-Hörversuchs Applikation, welche in Adobe Flex implementiert und am Institut für Hörtechnik und Audiologie (Oldenburg) entwickelt wurde. Diese Applikation stellt eine Auswahl von Hörversuchsstsetups bereit, z. B. Ja/Nein-Experimente, MUSHRA, Paarvergleiche. Als Audiomaterial können sowohl MP3- als auch WAV-Dateien verwendet werden. Bei den durchgeführten Experimenten wurden ausschließlich WAV-Dateien genutzt. Der Vorteil der verwendeten Online-Applikation besteht darin, dass sie von jedem internetfähigem Rechner mit aktuellem Adobe Flash Plugin ausgeführt werden kann.

Die durchgeführten Hörversuche können in zwei Methoden aufgeteilt werden: in Ja/Nein Experimente und Zuordnungsexperimente. Bei der Durchführung wurden zuerst vier unterschiedliche Ja/Nein Experimente (Experiment 1 bis 4) durchgeführt, anschließend folgten vier Zuordnungsexperimente (Experiment 5 bis 8). Die letzten vier Experimente (9 bis 12) entsprechen einem Re-Test der ersten vier Experimente. Die Testprozeduren der Experimente werden im folgenden erläutert.

### **3.2.1 Ja/Nein Experimente**

Insgesamt wurden acht Ja/Nein-Experimente durchgeführt. Bei allen Experimenten wurde dem Probanden jeweils eine aus dem Testkorpus zufällig auswählte Audiodatei vorgespielt. Die Probanden hatten die Möglichkeit, die Datei so oft wie gewünscht anzuhören und mussten anschließend bewerten, ob die Quelle während der Signalerzeugung in die Richtung des Empfängers blickt oder nicht. Anschließend folgte ein weiteres zufällig ausgewähltes Signal, bis die gesamten Audiodateien bewertet wurden. Abbildung A.1 im Anhang A.1 zeigt einen Screenshot der Online-Applikation mit der Oberfläche für die durchgeführten Ja/Nein-Experimente.

Die ersten vier Experimente verwendeten identisches Quellmaterial, wel-

ches entweder im reflexionsarmen Raum oder im Bürroraum aufgenommen wurden. Als Testsignale wurden bei den Experimenten jeweils natürliche weibliche und männliche Sprache, ein OLSA Satz aus dem Lautsprecher und der OLSA Satz aus dem Lautsprecher mit Frequenzgangsausgleich verwendet. Getestet wurden alle Quellen mit Positionen zwischen  $0^\circ$  und  $180^\circ$  in  $45^\circ$ -Schritten und jeweils zwei unterschiedlichen Sprecherausrichtungen, entweder  $0^\circ$  und  $90^\circ$  oder  $0^\circ$  und  $180^\circ$ . Pro Experiment mussten die Probanden bei 40 unterschiedlichen Audiobeispielen entscheiden, ob der Sprecher in die Richtung des Probanden sprach oder nicht.

Die Experimente 9 bis 12 sind identische Wiederholungen der ersten vier Experimente. So wurde eine Test/Retest Situation geschaffen, um Aussagen auf die Retest-Reliabilität zu treffen und mögliche Lerneffekte bei der Schätzung der Sprecherausrichtung zu erkennen.

### **3.2.1.1 Experimente 1 und 9: Unterscheidung der Sprecherausrichtung im reflexionsarmen Raum ( $0^\circ$ vs. $180^\circ$ )**

Bei den Experimenten 1 und 9 wurden Signale verwendet, die im reflexionsarmen Raum aufgenommen wurden. Dem Probanden wurden Audiodateien vorgespielt, bei denen die Quelle eine Sprecherausrichtung von  $0^\circ$  oder  $180^\circ$  aufwies und somit die maximal mögliche Differenz zwischen den Sprecherausrichtungen betrachtet wurde. Aufgrund der fehlenden Reflexionen und durch die gewählten Sprecherausrichtungen ergeben sich theoretisch für alle Sprecherpositionen keine ausrichtungsabhängigen interauralen Unterschiede (vgl. Abschnitt 2.4). Somit können die Probanden in den betrachteten Situationen ausschließlich Unterschiede in der ausrichtungsabhängigen Klangverfärbung des Quellsignals auswerten. Für die Quellenpositionen  $45^\circ$ ,  $90^\circ$  und  $135^\circ$  ergeben sich ILDs und ITDs, die ausschließlich aus der räumlichen Position der Quelle resultieren.

### **3.2.1.2 Experimente 2 und 10: Unterscheidung der Sprecherausrichtung im Büro ( $0^\circ$ vs. $180^\circ$ )**

Die in den Experimenten 2 und 10 verwendeten Signale wurden im Bürroraum aufgenommen. Als Sprecherausrichtung wurden  $0^\circ$  oder  $180^\circ$  gewählt, die entspricht der maximal möglichen Differenz zwischen den Sprecherausrichtungen. Neben der ausrichtungsabhängigen Klangverfärbung bei  $180^\circ$ , ist zu erwarten, dass die Probanden Unterschiede durch die resultierenden Reflexionen wahrnehmen und somit eine bessere Wahrnehmung der Sprecherausrichtung im Vergleich zum reflexionsarmen

Raum zu erwarten ist (vgl. Abschnitt 2.4). Anders als im reflexionsarmen Raum entstehen hier interaurale Laufzeit- und Pegelunterschiede vornehmlich bei den frühen Reflexionen, da Quelle und Empfänger asymmetrisch im Raum positioniert sind (siehe Abbildung 3.3).

### **3.2.1.3 Experimente 3 und 11: Unterscheidung der Sprecherausrichtung im reflexionsarmen Raum ( $0^\circ$ vs. $90^\circ$ )**

Bei den Experimenten 3 und 11 wurden Signale verwendet, die im reflexionsarmen Raum aufgenommen wurden. Dem Probanden wurden Audiodateien vorgespielt, bei denen die Quelle eine Sprecherausrichtung von  $0^\circ$  oder  $90^\circ$  aufwies. Wie bereits unter Abschnitt 2.4 beschrieben können bei einer Sprecherausrichtung von  $90^\circ$  im reflexionsarmen Raum geringe ausrichtungsabhängige interaurale Laufzeit- und Pegeldifferenzen durch Beugung an der Quelle auftreten und möglicherweise von den Probanden wahrgenommen werden. Wie beim ersten Experiment, ist auch hier zu erwarten, dass die Probanden hauptsächlich die Klangverfärbung des Quellsignals als akustische Eigenschaft zur Schätzung der Sprecherausrichtung nutzen. Allerdings ist diese Klangverfärbung aufgrund der richtungs- und frequenzabhängigen Abstrahlcharakteristik der Quelle für  $90^\circ$  geringer als für  $180^\circ$ .

### **3.2.1.4 Experimente 4 und 12: Unterscheidung der Sprecherausrichtung im Büro ( $0^\circ$ vs. $90^\circ$ )**

Die in den Experimenten 4 und 12 verwendeten Signale wurden im Bürroraum aufgenommen. Als Sprecherausrichtung wurden  $0^\circ$  oder  $90^\circ$  gewählt. Wie bereits beim Experiment 3 sind hier geringe Unterschiede in der Klangverfärbung zwischen den betrachteten Sprecherausrichtung als bei den Sprecherausrichtungen von  $0^\circ$  und  $180^\circ$  zu erwarten. Andererseits sollten die größten interauralen Laufzeit- und Pegelunterschiede aufgrund der asymmetrischen Sprecher- und Quellposition bei den frühen ersten Reflexionen auftreten.

## **3.2.2 Zuordnungsexperimente**

Bei den durchgeführten Zuordnungsexperimenten bestand die Aufgabe der Probanden darin, vorgegebene Audiodateien zu bewerten und der geschätzten Sprecherausrichtung zuzuweisen. Pro zufällig aus dem Testkorpus ausgewählten Testsetup standen dem Probanden sechs (bei natürlichem

Sprecher) bzw. acht (beim Lautsprecher) Audiodateien zu Verfügung, bei denen das identische Quellsignal bei unterschiedlichen Ausrichtung wiedergegeben wurde. Die Probanden konnten die Signale so oft wie gewünscht anhören und mussten dann alle zur Verfügung stehenden Signale einer möglichen Ausrichtung zuordnen. Mehrfachbelegungen waren ausgeschlossen. Nachdem alle Signale einer Gruppe zugeordnet wurden, folgte eine weitere Signalgruppe, bis die gesamten Audiodateien bewertet wurden. Abbildung A.2 im Anhang A.1 zeigt einen Screenshot der Online-Applikation mit der Oberfläche für die durchgeföhrten Zuordnungsexperimente. Im Zentrum sind schematisch die Sprecherpositionen und -ausrichtungen relativ zur Hörerposition abgebildet. An der linken Seite befinden sich die zur Verfügung stehenden, noch nicht zugeordneten Signale, die um die Quelle herum per *drag-and-drop* positioniert werden mussten.

Es wurden vier Zuordnungsexperimente durchgeführt, die sich in der Auswahl und Darbietung der Signale unterschieden. Innerhalb eines Experiments wurden sowohl Aufnahmen aus dem reflexionsarmen Raum wie aus dem Bürroraum verwendet. Die Quellposition betrug immer  $0^\circ$  und der Abstand zwischen Quelle und Empfänger ein Meter, mit Ausnahme von Experiment 7. Hier wurden zusätzlich Situationen getestet, bei denen die Quelle bei  $90^\circ$  oder im Abstand von zwei Metern positioniert war. Zusätzlich wurden bei diesem Experiment neben natürlicher Sprache auch synthetische Signale verwendet. Außer dichotischen Signalen, bei denen die Probanden binaurale Eigenschaften ausnutzen konnten, wurden die Experimente mit den gleichen Signalen aber bei monotonischer Darbietung durchgeführt.

### 3.2.2.1 Experiment 5: Monotische Schätzung der Sprecherausrichtung mit Sprachsignalen

Im Experiment 5 war es die Aufgabe der Probanden, sprecherausrichtungsabhängige monotone Sprachsignale den geschätzten Ausrichtungen zuzuordnen. Durch die monotone Darbietung können keine interauralen Unterschiede auftreten, so dass die Probanden vor allem bei Aufnahmen aus dem reflexionsarmen Raum ausschließlich die richtungsabhängigen Klangverfärbungen zwischen den Signalen nutzen konnten. Im Bürroraum können hingegen die Einflüsse der frühen Reflexionen zur Schätzung der Sprecherausrichtung genutzt werden. Es ist zu erwarten, dass die Probanden aufgrund der fehlenden binauralen Informationen keine Möglichkeit haben, zwischen einer rechten und linken Sprecherausrichtung zu unterscheiden.

**3.2.2.2 Experiment 6: Monotische Schätzung der Sprecherausrichtung mit LPC-Residual-Signalen**

In Abschnitt 4.1.3 wird gezeigt, dass das LPC-Residual-Signal für die Schätzung der Sprecherausrichtung möglicherweise algorithmisch genutzt werden kann. Wie bereits in den Abschnitten 1.4 und 4.1.3 beschrieben, weist das Spektrum des LPC-Residualsignals, abhängig von der Anzahl der verwendeten Filterkoeffizienten, eine möglichst weiße Grobstruktur auf und sollte so geringere Klangverfärbungen aufweisen. Im Zeitsignal sind idealerweise nur das Glottissignal, die Rauschanregung der Spracherzeugung und die Reflexionen enthalten. Um die Einflüsse auf die Schätzung der normalhörenden Probanden zu untersuchen und mit den Ergebnissen der algorithmischen Schätzungen in Kapitel 4 zu vergleichen, wurden für das Experiment 6, sowie für das Experiment 8, die LPC-Residual-Signale der im Experiment 5 verwendeten Sprachdateien berechnet. Die resultierenden Signale wurden den Probanden monotonisch zur Bewertung dargeboten.

**3.2.2.3 Experiment 7: Dichotische Schätzung der Sprecherausrichtung mit Sprachsignalen**

Im Experiment 7 wurden den Probanden Sprachsignale und synthetische Signale dichotisch dargeboten. Für die Signale im reflexionsarmen Raum sind die Ergebnisse mit den Experimenten von Kato et al. (vgl. Abschnitt 2.2.1) und im Büraum mit den Experimenten von Neuhoff (vgl. Abschnitt 2.2.2) zu vergleichen. Die Probanden können sowohl die Unterschiede in der Klangfärbung als auch interaurale Effekte und Einflüsse der frühen Reflexionen bei der Bewertung der Sprecherausrichtung nutzen. Die synthetischen Signale wurden in diesem Experiment verwendet um die bisherigen Erkenntnisse über die Fähigkeit der Normalhörenden genauer zu untersuchen und ggf. für die Entwicklung weiterer Algorithmen auszunutzen.

**3.2.2.4 Experiment 8: Dichotische Schätzung der Sprecherausrichtung mit LPC-Residual-Signalen**

Beim Experiment 8 mussten die Probanden dichotische LPC-Residualsignale bewerten. Vergleichbar mit dem Experiment 6 enthalten diese Signale idealerweise die Glottisimpulse und die Rauschanregung der Spracherzeugung sowie die Reflexionen im Raum bei einer möglichst weißen Grobstruktur im Frequenzspektrum. Zusätzlich bleiben binaurale Informationen in den Signalen enthalten.

### **3.3 Versuchsergebnisse: Schätzung der Sprecherausrichtung mit normalhörenden Probanden**

In den folgenden Abschnitten werden die Ergebnisse und Auswertungen der Experimente zur Schätzung der Sprecherausrichtung mit normalhörenden Probanden vorgestellt und interpretiert. Die Bezeichnungen der verwendeten Audiofiles in den Abbildungen und Tabellen der Ergebnisse sind so gewählt, dass daraus die jeweilige Aufnahmesituation ersichtlich ist. Jede Bezeichnung beginnt mit der Sprecherausrichtung in Grad (*SA000*, ..., *SA360*), gefolgt vom Abstand zwischen Quelle und Empfänger (*1m* oder *2m*). Anschließend folgt das Quellsignal (*männlich*, *weiblich*, *OLSA*, etc). Für den Fall, dass es sich um ein Signal mit einem Ausgleich der quellenabhängigen Tiefpasscharakteristik handelt, wird zusätzlich ein *HP* an die Bezeichnung angehängt. So ergibt sich z.B. für den hochpassgefilterten männlichen Sprecher im Abstand von 1 m und einer Sprecherausrichtung von 135 ° die Bezeichnung *SA135 1m männlich HP*.

#### **3.3.1 Ja/Nein Experimente**

Die Auswertung der durchgeführten Ja/Nein Experimente erfolgt mithilfe der Berechnung und Interpretation der Sensitivität, der Spezifität und der Rate der korrekten Klassifikation. Um diese Größen zu bestimmen, wurden probandenübergreifend für jedes Experiment und jedes verwendete Signal folgende vier Werte bestimmt:

- richtig positiv ( $r_p$ ): der Proband hat angegeben, dass die Quelle direkt zum Empfänger blickt und diese Aussage war korrekt,
- richtig negativ ( $r_n$ ): der Proband hat angegeben, dass die Quelle nicht zum Empfänger blickt und diese Aussage war korrekt,
- falsch positiv ( $f_p$ ): der Proband hat angegeben, dass die Quelle direkt zum Empfänger blickt und diese Aussage war falsch,
- falsch negativ ( $f_n$ ): der Proband hat angegeben, dass die Quelle nicht zum Empfänger blickt und diese Aussage war falsch.

Die Sensitivität wird nach Formel 3.1 bestimmt und gibt an, inwieweit die Probanden in der Lage waren, die Sprecherausrichtung von 0 ° korrekt zu bestimmen. Für den optimalen Fall beträgt die Sensitivität eins und ist null,

## Kapitel 3. Analyse der menschlichen Fähigkeit zur Schätzung der Sprecherausrichtung mit normalhörenden Probanden

---

wenn keine Sprecherausrichtung von  $0^\circ$  erkannt wurden.

$$\text{Sensitivität} = \text{SENS} = \frac{r_p}{r_p + f_n} \quad (3.1)$$

Die Spezifität wird nach Formel 3.2 bestimmt und gibt an, inwieweit die Probanden in der Lage waren, die von  $0^\circ$  abweichende Sprecherausrichtung korrekt zu bestimmen. Für den optimalen Fall beträgt die Spezifität eins und ist null, wenn alle abweichenden Sprecherausrichtungen als  $0^\circ$  bestimmt wurde.

$$\text{Spezifität} = \text{SPEZ} = \frac{r_n}{r_n + f_p} \quad (3.2)$$

Die Korrektklassifikationsrate wird nach Formel 3.3 berechnet und gibt an, inwieweit die Probanden die Sprecherausrichtung korrekt bewertet haben. Sie beträgt eins, wenn die getesteten Sprecherausrichtungen korrekt bestimmt waren und null, wenn alle Bewertungen falsch lagen. Die Ratewahrscheinlichkeit beträgt 50 %.

$$\text{Korrektklassifikationsrate} = \text{KORR} = \frac{r_p + r_n}{r_p + r_n + f_p + f_n} \quad (3.3)$$

### 3.3.1.1 Experiment 1 und 9: Unterscheidung der Sprecherausrichtung im reflexionsarmen Raum ( $0^\circ$ vs. $180^\circ$ )

Die über alle Probanden zusammengefassten Messergebnisse für das erste Experiment sind in Tabelle 3.1(a) in Abhängigkeit von den Testsignalen dargestellt. Die Ergebnisse zeigen, dass die Probanden im Mittel eine Sensitivität von 86 % und eine Spezifität von 84 % erreichen. Insgesamt wurden in 84 % der Fälle erkannt, ob der Sprecher in Richtung des Empfängers blickt oder nicht. Für alle drei Quellen ist zu erkennen, dass die Probanden tendenziell bessere Ergebnisse erzielen, wenn die Quelle seitlich vom Empfänger steht. Die Ergebnisse über die unterschiedlichen Quellen befinden sich in ähnlicher Größenordnung, was durch eine Standardabweichung von maximal 8 % bestätigt wird.

Zum Vergleich wurde bei den OLSA Sätzen ein Ausgleich der Tiefpasscharakteristik der Quelle durchgeführt. Die Messergebnisse liegen für alle drei Ergebnisse im Mittel bei etwa 52 % und entsprechen somit der Ratewahrscheinlichkeit. Auch hier zeigt sich eine geringe Standardabweichung von 5 %, so dass die Ergebnisse über alle Probanden und Konditionen valide sind.

In Tabelle 3.1(b) sind die Ergebnisse des Re-Tests dargestellt. Alle ermittelten

## Kapitel 3. Analyse der menschlichen Fähigkeit zur Schätzung der Sprecherausrichtung mit normalhörenden Probanden

---

**Tab. 3.1:** Ergebnisse Experimente 1 und 9: Unterscheidung der Sprecherausrichtung im reflexionsarmen Raum von normalhörenden Probanden. Tabellen zeigen die Sensitivität, die Spezifität und die Korrektklassifikationsrate für Unterscheidung von den Sprecherausrichtungen von  $0^\circ$  und  $180^\circ$  für unterschiedliche Quellsignale.

(a) Ergebnis Experiment 1				(b) Ergebnis Experiment 9 (Re-Test)			
Signal	SENS	SPEZ	KORR	Signal	SENS	SPEZ	KORR
SA000 1m männlich	0.773	0.800	0.786	SA000 1m männlich	0.667	0.733	0.694
SA045 1m männlich	0.731	0.875	0.786	SA045 1m männlich	0.700	0.750	0.722
SA090 1m männlich	0.941	0.800	0.857	SA090 1m männlich	0.813	0.750	0.778
SA135 1m männlich	0.882	0.760	0.810	SA135 1m männlich	0.737	0.765	0.750
SA180 1m männlich	0.800	0.941	0.857	SA180 1m männlich	0.824	0.789	0.806
SA000 1m weiblich	0.947	0.870	0.905	SA000 1m weiblich	0.941	0.895	0.917
SA045 1m weiblich	0.783	0.842	0.810	SA045 1m weiblich	0.789	0.824	0.806
SA090 1m weiblich	1.000	0.750	0.833	SA090 1m weiblich	1.000	0.783	0.861
SA135 1m weiblich	0.944	0.833	0.881	SA135 1m weiblich	0.933	0.810	0.861
SA180 1m weiblich	0.857	0.857	0.857	SA180 1m weiblich	0.938	0.850	0.889
SA000 1m OLSA	0.792	0.889	0.833	SA000 1m OLSA	0.842	0.882	0.861
SA045 1m OLSA	0.792	0.889	0.833	SA045 1m OLSA	0.762	0.867	0.806
SA090 1m OLSA	0.833	0.750	0.786	SA090 1m OLSA	0.889	0.889	0.889
SA135 1m OLSA	0.938	0.769	0.833	SA135 1m OLSA	1.000	0.857	0.917
SA180 1m OLSA	0.870	0.947	0.905	SA180 1m OLSA	1.000	0.857	0.917
Gesamt $\mu$	0.859	0.838	0.838	Gesamt $\mu$	0.856	0.820	0.831
$\sigma$	0.081	0.065	0.039	$\sigma$	0.111	0.055	0.072
SA000 1m OLSA HP	0.480	0.471	0.476	SA000 1m OLSA HP	0.500	0.500	0.500
SA045 1m OLSA HP	0.517	0.538	0.524	SA045 1m OLSA HP	0.409	0.357	0.389
SA090 1m OLSA HP	0.556	0.600	0.571	SA090 1m OLSA HP	0.471	0.474	0.472
SA135 1m OLSA HP	0.545	0.516	0.524	SA135 1m OLSA HP	0.462	0.478	0.472
SA180 1m OLSA HP	0.533	0.519	0.524	SA180 1m OLSA HP	0.500	0.500	0.500
Gesamt HP $\mu$	0.526	0.529	0.524	Gesamt HP $\mu$	0.468	0.462	0.467
$\sigma$	0.030	0.047	0.034	$\sigma$	0.037	0.060	0.046

Werte liegen in einer vergleichbaren Größenordnung, sind aber tendenziell etwas geringer als beim ersten Testdurchlauf.

### 3.3.1.2 Experimente 2 und 10: Unterscheidung der Sprecherausrichtung im Büro ( $0^\circ$ vs. $180^\circ$ )

Die Ergebnisse für das zweite Experiment sind in der Tabelle 3.2(a) dargestellt. Die Ergebnisse im halligen Büroraum liegen über alle Probanden und Quellen mit einer Sensitivität von 86 %, einer Spezifität von 84 % und einer Korrektklassifikationsrate von 84 % in der gleichen Größenordnung, wie die Ergebnisse im Experiment 1 mit den Signalen aus dem reflexionsarmen Raum. Die Standardabweichung beträgt maximal 9 %.

Anders als im reflexionsarmen Raum können die Probanden die Signale mit Tiefpassausgleich besser klassifizieren. Hier liegen die Ergebnisse mit einer Sensitivität von 84 %, einer Spezifität von 72 % und einer Korrektklassifikationsrate von 76 % deutlich über der Ratewahrscheinlichkeit und in ähnlicher Größenordnung wie die unausgeglichenen Signale.

Der Re-Test in Tabelle 3.2(b) zeigt erneut tendenziell schlechtere aber zum ersten Testdurchlauf dieses Experimentes vergleichbare Ergebnisse.

## Kapitel 3. Analyse der menschlichen Fähigkeit zur Schätzung der Sprecherausrichtung mit normalhörenden Probanden

---

**Tab. 3.2:** Ergebnisse Experimente 2 und 10: Unterscheidung der Sprecherausrichtung im Büro von normalhörenden Probanden. Tabellen zeigen die Sensitivität, die Spezifität und die Korrektklassifikationsrate für Unterscheidung von den Sprecherausrichtungen von  $0^\circ$  und  $180^\circ$  für unterschiedliche Quellsignale.

(a) Ergebnis Experiment 2				(b) Ergebnis Experiment 10 (Re-Test)			
Signal	SENS	SPEZ	KORR	Signal	SENS	SPEZ	KORR
SA000 1m männlich	0.900	0.864	0.881	SA000 1m männlich	0.778	0.750	0.763
SA045 1m männlich	0.909	0.950	0.929	SA045 1m männlich	0.875	0.773	0.816
SA090 1m männlich	0.750	0.833	0.786	SA090 1m männlich	0.933	0.783	0.842
SA135 1m männlich	1.000	0.778	0.857	SA135 1m männlich	0.889	0.850	0.868
SA180 1m männlich	0.875	0.731	0.786	SA180 1m männlich	0.750	0.682	0.711
SA000 1m weiblich	0.905	0.905	0.905	SA000 1m weiblich	0.882	0.810	0.842
SA045 1m weiblich	0.889	0.792	0.833	SA045 1m weiblich	0.783	0.933	0.842
SA090 1m weiblich	0.850	0.818	0.833	SA090 1m weiblich	0.857	0.941	0.895
SA135 1m weiblich	0.882	0.760	0.810	SA135 1m weiblich	0.944	0.900	0.921
SA180 1m weiblich	0.905	0.905	0.905	SA180 1m weiblich	0.824	0.762	0.789
SA000 1m OLSA	0.826	0.895	0.857	SA000 1m OLSA	0.842	0.842	0.842
SA045 1m OLSA	0.909	0.950	0.929	SA045 1m OLSA	1.000	0.950	0.974
SA090 1m OLSA	0.895	0.826	0.857	SA090 1m OLSA	0.895	0.895	0.895
SA135 1m OLSA	0.933	0.741	0.810	SA135 1m OLSA	0.786	0.667	0.711
SA180 1m OLSA	0.938	0.769	0.833	SA180 1m OLSA	1.000	0.731	0.816
Gesamt $\mu$	0.891	0.834	0.854	Gesamt $\mu$	0.869	0.818	0.835
$\sigma$	0.055	0.074	0.048	$\sigma$	0.078	0.093	0.073
SA000 1m OLSA HP	0.944	0.833	0.881	SA000 1m OLSA HP	0.765	0.714	0.737
SA045 1m OLSA HP	0.889	0.792	0.833	SA045 1m OLSA HP	0.857	0.708	0.763
SA090 1m OLSA HP	0.800	0.773	0.786	SA090 1m OLSA HP	0.857	0.708	0.763
SA135 1m OLSA HP	0.833	0.633	0.690	SA135 1m OLSA HP	0.700	0.571	0.605
SA180 1m OLSA HP	0.727	0.581	0.619	SA180 1m OLSA HP	1.000	0.655	0.737
Gesamt HP $\mu$	0.839	0.722	0.762	Gesamt HP $\mu$	0.836	0.672	0.721
$\sigma$	0.083	0.109	0.106	$\sigma$	0.113	0.061	0.066

### 3.3.1.3 Experimente 3 und 11: Unterscheidung der Sprecherausrichtung im reflexionsarmen Raum ( $0^\circ$ vs. $90^\circ$ )

In Tabelle 3.3(a) sind die über alle Probanden zusammengefassten Messergebnisse für das dritte Experiment im reflexionsarmen Raum dargestellt. Im Gegensatz zum ersten Experiment ist hier die Differenz zwischen der gewählten Sprecherausrichtung mit  $90^\circ$  geringer. Für die Signale ohne Ausgleich der Tiefpasscharakteristik sind die Werte mit einer Sensitivität von 80 %, einer Spezifität von 82 % und einer Korrektklassifikationsrate von 79 % geringer als beim ersten Experiment. Zusätzlich steigt die Standardabweichung auf bis zu 12,5 % an. Auch bei diesem Experiment ist die Tendenz zu erkennen, dass die Probanden die Sprecherausrichtung bei seitlicher Quelle besser unterscheiden können.

Für die Signale mit Ausgleich der richtungsabhängigen Tiefpasscharakteristik der Quelle in Tabelle 3.3(b) zeigt sich, dass die Erkennungsleistung der Probanden nahe der Ratewahrscheinlichkeit liegt. Die Re-Test-Ergebnisse befinden sich erneut in der Größenordnung des ersten Testdurchlauf mit einer Tendenz zur schlechteren Erkennungsleistung.

## Kapitel 3. Analyse der menschlichen Fähigkeit zur Schätzung der Sprecherausrichtung mit normalhörenden Probanden

---

**Tab. 3.3:** Ergebnisse Experimente 3 und 11: Unterscheidung der Sprecherausrichtung im reflexionsarmen Raum von normalhörenden Probanden. Tabellen zeigen die Sensitivität, die Spezifität und die Korrektklassifikationsrate für Unterscheidung von den Sprecherausrichtungen von  $0^\circ$  und  $90^\circ$  für unterschiedliche Quellsignale.

(a) Ergebnis Experiment 3				(b) Ergebnis Experiment 11 (Re-Test)			
Signal	SENS	SPEZ	KORR	Signal	SENS	SPEZ	KORR
SA000 1m männlich	0.548	0.636	0.571	SA000 1m männlich	0.519	0.545	0.526
SA045 1m männlich	0.680	0.765	0.714	SA045 1m männlich	0.526	0.526	0.526
SA090 1m männlich	0.714	0.714	0.714	SA090 1m männlich	0.700	0.722	0.711
SA135 1m männlich	0.917	0.667	0.738	SA135 1m männlich	0.733	0.652	0.684
SA180 1m männlich	0.714	0.929	0.786	SA180 1m männlich	0.577	0.667	0.605
SA000 1m weiblich	0.731	0.875	0.786	SA000 1m weiblich	0.727	0.813	0.763
SA045 1m weiblich	0.731	0.875	0.786	SA045 1m weiblich	0.810	0.882	0.842
SA090 1m weiblich	0.696	0.737	0.714	SA090 1m weiblich	0.591	0.625	0.605
SA135 1m weiblich	0.938	0.769	0.833	SA135 1m weiblich	0.875	0.773	0.816
SA180 1m weiblich	0.760	0.882	0.810	SA180 1m weiblich	0.833	0.800	0.816
SA000 1m OLSA	0.870	0.947	0.905	SA000 1m OLSA	0.818	0.938	0.868
SA045 1m OLSA	0.783	0.842	0.810	SA045 1m OLSA	0.810	0.882	0.842
SA090 1m OLSA	0.950	0.909	0.929	SA090 1m OLSA	0.810	0.882	0.842
SA135 1m OLSA	1.000	0.875	0.929	SA135 1m OLSA	1.000	0.864	0.921
SA180 1m OLSA	0.833	0.944	0.881	SA180 1m OLSA	0.938	0.818	0.868
Gesamt $\mu$	0.791	0.824	0.794	Gesamt $\mu$	0.751	0.759	0.749
$\sigma$	0.124	0.102	0.096	$\sigma$	0.146	0.130	0.131
SA000 1m OLSA HP	0.500	0.500	0.500	SA000 1m OLSA HP	0.500	0.500	0.500
SA045 1m OLSA HP	0.526	0.522	0.524	SA045 1m OLSA HP	0.471	0.476	0.474
SA090 1m OLSA HP	0.500	0.500	0.500	SA090 1m OLSA HP	0.444	0.450	0.447
SA135 1m OLSA HP	0.625	0.529	0.548	SA135 1m OLSA HP	0.250	0.471	0.447
SA180 1m OLSA HP	0.556	0.542	0.548	SA180 1m OLSA HP	0.417	0.462	0.447
Gesamt HP $\mu$	0.541	0.519	0.524	Gesamt HP $\mu$	0.416	0.472	0.463
$\sigma$	0.052	0.018	0.024	$\sigma$	0.098	0.019	0.024

### 3.3.1.4 Experimente 4 und 12: Unterscheidung der Sprecherausrichtung im Büro ( $0^\circ$ vs. $90^\circ$ )

Tabelle 3.4(a) beinhaltet die Ergebnisse des vierten Experiments, mit Signalen aus dem halligen Büroraum und Sprecherausrichtung von  $0^\circ$  und  $90^\circ$ . Mit einer Sensitivität von 81 %, einer Spezifität von 78 %, einer Korrektklassifikationsrate von 70 % und einer Standardabweichung von bis zu 11 % für die gemittelten Daten über alle Probanden und Konditionen, liegen die Ergebnisse ebenfalls in der Größenordnung der zuvor betrachteten Ergebnisse. Erneut ist zu erkennen, dass die Probanden leicht bessere Ergebnisse bei seitlich positionierter Quelle erreichen.

Wie zuvor im Experiment 2 zeigt sich in Tabelle 3.4(b), dass die Signale mit einem angepassten Frequenzspektrum im halligen Raum, im Gegensatz zu den vergleichbaren Experimenten im reflexionsarmen Raum, zu deutlich besseren Erkennungsraten führen. Hier liegt die Sensitivität bei 85 %, die Spezifität bei 66 % und die Korrektklassifikationsrate bei 71 %. Erneut liegen die Ergebnisse des Re-Tests in der Größenordnung des ersten Testdurchlaufs mit einer Tendenz zur schlechteren Erkennungsleistung.

## Kapitel 3. Analyse der menschlichen Fähigkeit zur Schätzung der Sprecherausrichtung mit normalhörenden Probanden

---

**Tab. 3.4:** Ergebnisse Experimente 4 und 12: Unterscheidung der Sprecherausrichtung im Büro von normalhörenden Probanden. Tabellen zeigen die Sensitivität, die Spezifität und die Korrektklassifikationsrate für Unterscheidung von den Sprecherausrichtungen von  $0^\circ$  und  $90^\circ$  für unterschiedliche Quellsignale.

(a) Ergebnis Experiment 4				(b) Ergebnis Experiment 12 (Re-Test)			
Signal	SENS	SPEZ	KORR	Signal	SENS	SPEZ	KORR
SA000 1m männlich	0.571	0.643	0.595	SA000 1m männlich	0.583	0.643	0.605
SA045 1m männlich	0.679	0.857	0.738	SA045 1m männlich	0.591	0.625	0.605
SA090 1m männlich	0.818	0.850	0.833	SA090 1m männlich	0.800	0.833	0.816
SA135 1m männlich	0.789	0.739	0.762	SA135 1m männlich	0.750	0.682	0.711
SA180 1m männlich	0.667	0.625	0.643	SA180 1m männlich	0.714	0.625	0.658
SA000 1m weiblich	0.739	0.789	0.762	SA000 1m weiblich	0.800	0.833	0.816
SA045 1m weiblich	0.826	0.895	0.857	SA045 1m weiblich	0.857	0.941	0.895
SA090 1m weiblich	0.870	0.947	0.905	SA090 1m weiblich	0.750	0.778	0.763
SA135 1m weiblich	0.933	0.741	0.810	SA135 1m weiblich	0.750	0.682	0.711
SA180 1m weiblich	0.833	0.750	0.786	SA180 1m weiblich	0.818	0.630	0.684
SA000 1m OLSA	0.900	0.864	0.881	SA000 1m OLSA	0.800	0.833	0.816
SA045 1m OLSA	0.950	0.909	0.929	SA045 1m OLSA	0.833	0.800	0.816
SA090 1m OLSA	0.941	0.800	0.857	SA090 1m OLSA	0.900	0.944	0.921
SA135 1m OLSA	0.824	0.720	0.762	SA135 1m OLSA	0.923	0.720	0.789
SA180 1m OLSA	0.867	0.704	0.762	SA180 1m OLSA	0.846	0.680	0.737
Gesamt $\mu$	0.814	0.789	0.792	Gesamt $\mu$	0.781	0.750	0.756
$\sigma$	0.110	0.097	0.092	$\sigma$	0.097	0.110	0.095
SA000 1m OLSA HP	0.750	0.654	0.690	SA000 1m OLSA HP	0.632	0.632	0.632
SA045 1m OLSA HP	0.800	0.667	0.714	SA045 1m OLSA HP	0.667	0.650	0.658
SA090 1m OLSA HP	0.850	0.818	0.833	SA090 1m OLSA HP	0.789	0.789	0.789
SA135 1m OLSA HP	1.000	0.600	0.667	SA135 1m OLSA HP	0.667	0.552	0.579
SA180 1m OLSA HP	0.800	0.594	0.643	SA180 1m OLSA HP	0.833	0.654	0.711
Gesamt HP $\mu$	0.840	0.666	0.710	Gesamt HP $\mu$	0.718	0.655	0.674
$\sigma$	0.096	0.091	0.074	$\sigma$	0.088	0.086	0.080

### 3.3.2 Zuordnungsexperimente

Die Auswertung und Interpretation der Versuchsergebnisse der Zuordnungsexperimente erfolgt in drei Schritten.

Im ersten Schritt werden für jede Testsituation Verwechslungsmatrizen aus den Probandenbewertungen bestimmt. Dabei wird die von den Probanden geschätzte Ausrichtung des jeweiligen Quellsignals über der wahren Sprecherausrichtung aufgetragen und für alle Probanden aufsummiert. In den Abbildungen 3.5(a) bis 3.5(c) sind beispielhaft entsprechende Verwechslungsmatrizen dargestellt. Auf der x-Achse sind die wahren Ausrichtungen aufgetragen und auf der y-Achse die Geschätzten. Jedes Feld der Matrix wird zur besseren Unterscheidbarkeit entsprechend der relativen Häufigkeit farblich hervorgehoben, wobei rot 100 % und blau 0 % relative Häufigkeit entspricht. Zusätzlich ist in jedem Feld die gerundete, relative Häufigkeit als Zahl eingetragen. Abbildung 3.5(a) zeigt eine Verwechslungsmatrix bei der alle Schätzungen zu 100 % korrekt sind. Allerdings hat sich gezeigt, dass die Bewertungen der Probanden und der Algorithmen zum Teil häufige rechts-links Verwechslungen aufweisen, also z. B. eine Sprecherausrichtung von  $270^\circ$  statt  $90^\circ$  angeben. Abbildung 3.5(b) zeigt eine Verwechslungsmatrix bei der alle Bewertungen komplett rechts-links verwechselt wurden. Die

## Kapitel 3. Analyse der menschlichen Fähigkeit zur Schätzung der Sprecherausrichtung mit normalhörenden Probanden

---

Kombination aus diesen beiden Fällen ist in Abbildung 3.5(c) dargestellt. Hier sind die Bewertungen entweder korrekt angegeben oder von einer rechts-links Verwechslung betroffen.

Aufgrund der unterschiedlichen Testbedingungen bei den Zuordnungsexperimenten ergeben sich eine Vielzahl von Verwechslungsmatrizen. Um die Ergebnisse zusammenzufassen und möglichst überschaubar darzustellen, wurden aus den Verwechslungsmatrizen die mittleren Abweichungen über alle Richtungen als Fehlermaß bestimmt. Die mittlere Abweichung  $a_m$  berechnet sich aus

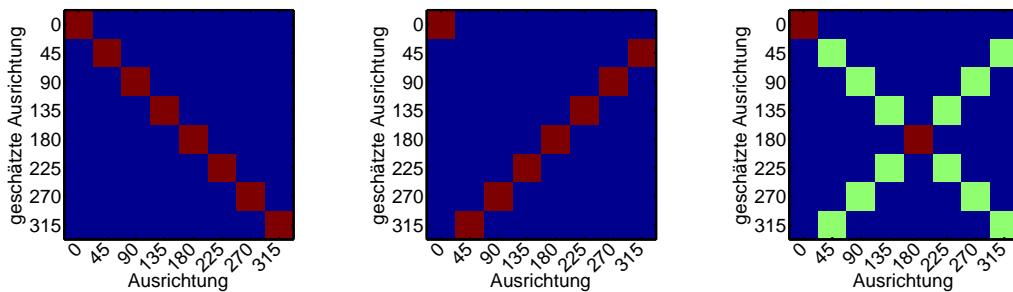
$$a_m = \frac{1}{N} \sum_{n \in [1, N]} |\theta(n) - \theta_0(n)|, \quad (3.4)$$

mit der Anzahl der Bewertungen  $N$ , der geschätzten Sprecherausrichtung  $\theta$  und der wahren Sprecherausrichtung  $\theta_0$ . Um eine Aussage über die Streuung der Bewertungen zu erhalten, wurde die Wurzel der mittleren quadratischen Abweichung  $a_q$  nach

$$a_q = \sqrt{\frac{1}{N} \sum_{n \in [1, N]} (\theta(n) - \theta_0(n))^2} \quad (3.5)$$

bestimmt. Je größer die Streuung der Bewertungen ist, um so größer ist die Differenz zwischen  $a_q$  und  $a_m$ . Bei der Betrachtung der Ergebnisse werden nur die Summen über die Quellsignale in den betrachteten Räumen mit und ohne Ausgleich der richtungsabhängigen Tiefpasscharakteristik betrachtet. Die probandenübergreifenden Einzelergebnisse der verwendeten Quellsignale können im Anhang A.2 nachgeschlagen werden.

Im zweiten Schritt werden  $a_m$  und  $a_q$  für die jeweiligen Raumbedingungen (reflexionsarmer Raum und Bürraum) über das jeweilige Quellsignal aufgetragen. Bei dieser Herangehensweise führen rechts-links Verwechslungen zum Teil zu sehr großen Abweichungen. In einer zusätzlichen Fehlerbetrachtung wurden die Fehlermaße aus den Beträgen der geschätzten Sprecherausrichtung bestimmt, so dass eine reine rechts-links Verwechslung ebenfalls zu einem optimalen Ergebnis führt. Dies erlaubt eine Abschätzung, ob die Probanden in der Lage waren, die Sprecherausrichtung zu unterscheiden und aufsteigend zuzuordnen, aber nicht zwischen einer Ausrichtung nach links oder rechts unterscheiden können. Es ist zu beachten, dass die Fehlermaße mit Verwechslungsausgleich durch dieses Vorgehen immer kleiner oder gleich von den Ergebnissen ohne Beachtung möglicher rechts-links Verwechslungen sind.



- (a) Optimales Ergebnis, alle sprecherausrichtungsabhängigen Signale wurden korrekt zugeordnet, ohne rechts-links Verwechslungen.
- (b) Alle sprecherausrichtungsabhängigen Signale wurden komplett mit rechts-links Verwechslung zugeordnet.
- (c) Alle sprecherausrichtungsabhängigen Signale wurden zur Hälfte korrekt und zu Hälfte mit rechts-links Verwechslung zugeordnet.

**Abb. 3.5:** Beispiele für Verwechslungsmatrizen zur besseren Deutung der im Anhang dargestellten Ergebnisse. Auf der x-Achse befinden sich die wahren Sprecherausrichtungen, auf der y-Achse die von den Probanden oder Algorithmen zugeordneten Ausrichtungen.

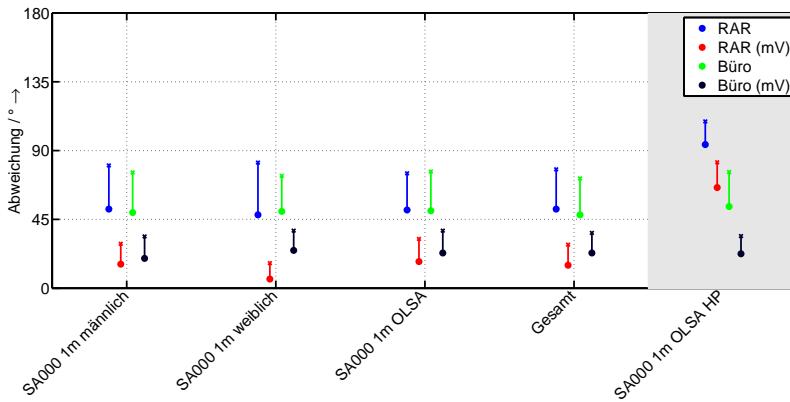
Im dritten Schritt wurden die Bewertungen der unterschiedlichen Quellsignale zusammengefasst und die Fehlermaße raumabhängig über der tatsächlichen Sprecherausrichtung aufgetragen. Diese Darstellung entspricht der Ergebnisdarstellung nach Kato et al. und Neuhoff (siehe Abschnitt 2.2).

### 3.3.2.1 Experiment 5: Monotische Schätzung der Sprecherausrichtung mit Sprachsignalen

Die Verwechslungsmatrizen des fünften Experiments mit normalhörenden Probanden sind im Anhang A.2.1 dargestellt. Die Abbildung 3.6 enthält die Fehlermaße in Abhängigkeit von den Quellsignalen. Die Graphen 3.7(a) bis 3.7(d) enthalten die gemittelten Abweichungen für alle ungefilterten Quellsignale über den wahren Ausrichtungen.

#### Ergebnisse ohne Ausgleich möglicher rechts-links Verwechslungen

Die Verwechslungsmatrizen für die ungefilterten Signale entsprechen in ihrer Form dem Beispiel in Abbildung 3.5(c). Aufgrund der monotonen Darstellung zeichnen sich erwartungsgemäß bei nahezu 50 % der Bewertungen rechts-links Verwechslungen ab, was der Ratewahrscheinlichkeit entspricht. Beim hochpassgefilterten Sprachsignal im reflexionsarmen Raum (Abbildung A.5(b)) zeigt die Verwechslungsmatrix eine gleichmäßige Verteilung der Bewertungen. Die Probanden waren hier nicht in der Lage die Sprecherausrichtungen zu bestimmen. Anders im halligen Bürraum (Abbildung A.5(d)), hier zeigt sich ein zu den ungefilterten Signalen vergleichbares Er-



**Abb. 3.6:** Ergebnisse Experimente 5: Monotische Schätzung der Sprecherausrichtung mit Sprachsignalen von normalhörenden Probanden. Grafik zeigt die mittleren Abweichungen  $a_m$  (●) und die Wurzel der mittleren quadratischen Abweichung  $a_q$  (x) der Probandenbewertung in Abhängigkeit von den Quellsignalen. (blau: Bewertungen der Signale im reflexionsarmen Raum, grün: Bewertungen der Signale im Büroraum, rot: Bewertungen der Signale im reflexionsarmen Raum mit rechts-links Ausgleich, schwarz: Bewertungen der Signale im Büroraum mit rechts-links Ausgleich)

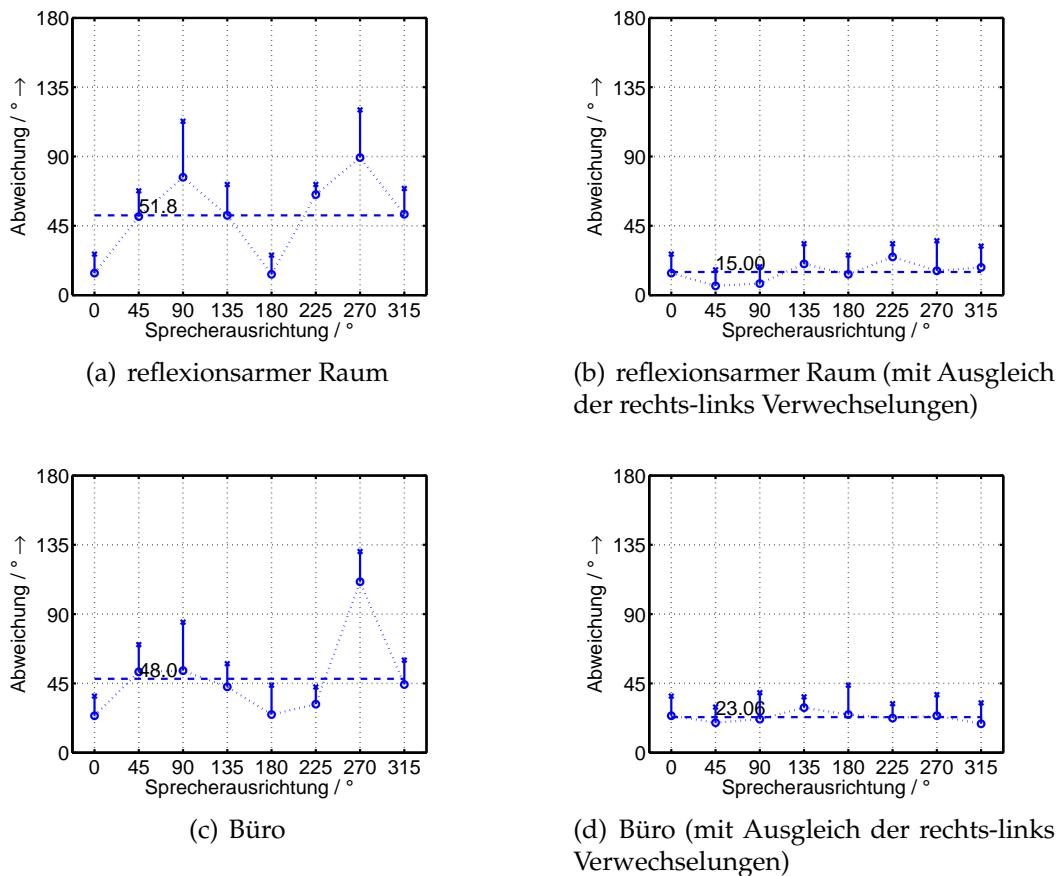
gebnis.

Bei der Darstellung der Fehlermaße in Abhängigkeit von den Quellsignalen in Abbildung 3.6 zeigt sich eine von den Quellsignalen unabhängige Bewertung der Probanden. Die mittleren Abweichungen liegen auf gleichem Niveau und weisen eine ähnliche Streuung auf. Im Mittel zeigen sich nur geringe Unterschiede zwischen den Bewertungen in den untersuchten Räumen.

Die mittlere Abweichung über der wahren Ausrichtung beträgt in beiden Räumen  $\approx 50^\circ$  (Abbildung 3.7(a) und 3.7(c)). Im reflexionsarmen Raum weisen die Schätzungen bei einer Sprecherausrichtung von  $0^\circ$  und  $180^\circ$  ein deutliches Minimum mit einer mittleren Abweichung von weniger als  $20^\circ$  auf, während bei  $90^\circ$  und  $270^\circ$  Abweichungen von etwa  $90^\circ$  auftreten. Die Bewertungen im Büroraum verhalten sich ähnlich. Sie schwanken aber über die wahren Ausrichtung weniger, mit einem Ausreißer bei  $280^\circ$ .

#### Ergebnisse mit Ausgleich möglicher rechts-links Verwechslungen

Die Ergebnisse mit Ausgleich der Verwechslungen zeigen eine mittlere Abweichung von  $15^\circ$  im reflexionsarmen Raum (Abbildung 3.7(b)) und  $23^\circ$  im Büroraum (Abbildung 3.7(d)). Sie sind konsistent über den betrachteten wahren Ausrichtungen. Die Probanden sind daher gut in der Lage die auftretenden akustischen Eigenschaften zu nutzen um die Sprecherausrichtung zu unterscheiden. Aufgrund der monotonen Darbietung ist jedoch keine rechts-links Unterscheidung möglich. Die Schätzungen im reflexionsarmen



**Abb. 3.7:** Ergebnisse Experimente 5: Monotische Schätzung der Sprecherausrichtung mit Sprachsignalen von normalhörenden Probanden. Grafiken zeigen die mittleren Abweichung  $a_m$  (•) und die Wurzel der mittleren quadratischen Abweichung  $a_q$  (x) und Durschschnitt der mittleren Abweichung der Probandenbewertung in Abhängigkeit von den wahren Sprecherausrichtungen.

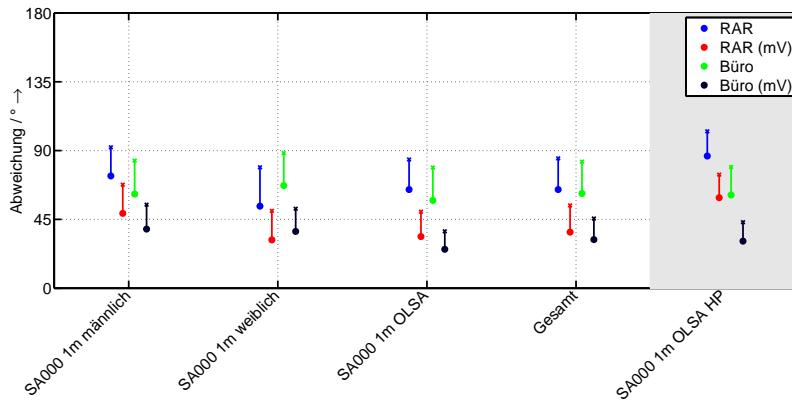
Raum gelingen den Probanden im Vergleich zum Bürraum mit  $23^\circ$  besser. Wie bereits in Abschnitt 2.4 gezeigt, kann in diesem Raum nur die richtungsabhängige Klangverfärbung genutzt werden und im Bürraum zusätzlich weitere akustische Eigenschaften. Diese Tendenz ist unerwartet.

### 3.3.2.2 Experiment 6: Monotische Schätzung der Sprecherausrichtung mit LPC-Residual-Signalen

Die Verwechslungsmatrizen des sechsten Experiments mit normalhörenden Probanden sind im Anhang A.2.2 dargestellt. Die Abbildung 3.8 enthält die Fehlermaße in Abhängigkeit von den Quellsignalen. Die Graphen 3.9(a) bis 3.9(d) enthalten die gemittelten Abweichungen für alle ungefilterten Quellsignale über den wahren Ausrichtungen.

#### Ergebnisse ohne Ausgleich möglicher rechts-links Verwechslungen

Bei diesem Experiment weisen die Verwechslungsmatrizen deutlichere



**Abb. 3.8:** Ergebnisse Experimente 6: Monotische Schätzung der Sprecherausrichtung mit LPC-Residual-Signalen von normalhörenden Probanden. Grafik zeigt die mittleren Abweichungen  $a_m$  (●) und die Wurzel der mittleren quadratischen Abweichung  $a_q$  (x) der Probandenbewertung in Abhängigkeit von den Quellsignalen. (blau: Bewertungen der Signale im reflexionsarmen Raum, grün: Bewertungen der Signale im Büroraum, rot: Bewertungen der Signale im reflexionsarmen Raum mit rechts-links Ausgleich, schwarz: Bewertungen der Signale im Büroraum mit rechts-links Ausgleich)

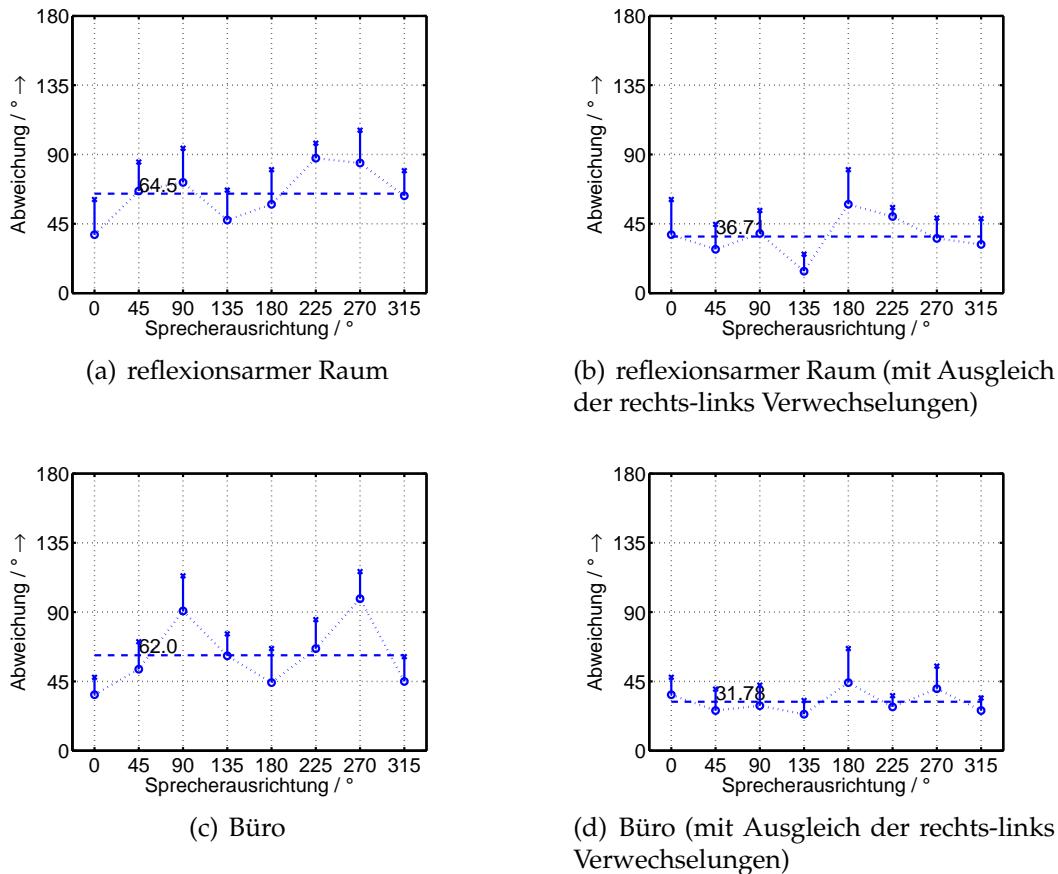
Streuungen der Bewertungen als im fünften Experiment auf. Danach ist auch hier im Mittel zu erkennen, dass die Sprecherausrichtung für die ungefilterten Signale in den beiden Räumen (Abbildung A.8(a) und A.8(c)) und das hochpassgefilterte Signal im Büroraum (Abbildung A.8(d)) unterscheidbar sind. Die Hochpassfilterung des Sprachsignals im reflexionsarmen Raum zeigt aufgrund der Gleichverteilung der Bewertungen, dass die Probanden die Sprecherausrichtung nicht unterscheiden können. Im Büroraum ist es durch die enthaltenen Reflexionen hingegen möglich.

Abbildung 3.8 zeigt, dass die Bewertungen nahezu unabhängig von den verwendeten Quellsignalen sind. Die mittleren Abweichungen liegen auf gleichem Niveau und weisen eine ähnliche Streuung auf. Im Mittel sind keine großen Unterschiede zwischen den Bewertungen in den untersuchten Räumen vorhanden.

Die gemittelten Abweichungen für alle ungefilterten Quellsignale über der wahren Ausrichtung im reflexionsarmen Raum (Abbildung 3.9(a)) und im Büroraum (Abbildung 3.9(c)) betragen  $\approx 63^\circ$ , mit lokalen Minima von  $\approx 45^\circ$  für die Sprecherausrichtungen  $0^\circ$  und  $180^\circ$ .

#### Ergebnisse mit Ausgleich möglicher rechts-links Verwechslungen

Die mittleren Abweichungen mit rechts-links Ausgleich betragen im reflexionsarmen Raum  $37^\circ$  (Abbildung 3.9(b)) und streuen wenig über alle wahren Ausrichtungen, mit Ausnahme von den Ausrichtungen  $135^\circ$  und  $180^\circ$ . Im



**Abb. 3.9:** Ergebnisse Experimente 6: Monotische Schätzung der Sprecherausrichtung mit LPC-Residual-Signalen von normalhörenden Probanden. Grafiken zeigen die mittleren Abweichung  $a_m$  (•) und die Wurzel der mittleren quadratischen Abweichung  $a_q$  (x) und Durchschnitt der mittleren Abweichung der Probandenbewertung in Abhängigkeit von den wahren Sprecherausrichtungen.

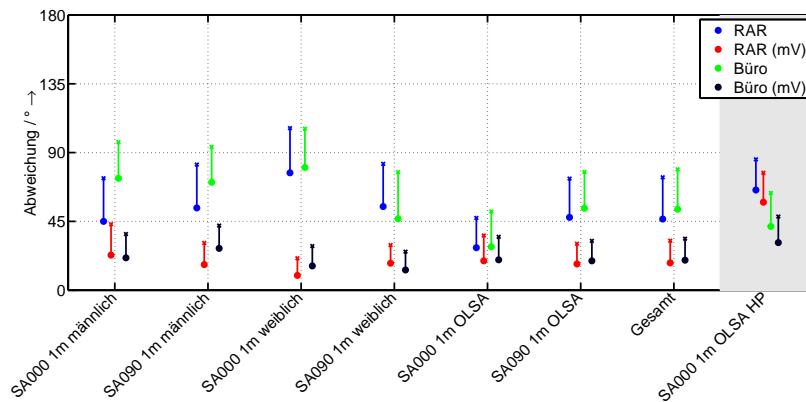
Büroraum liegen die mittleren Abweichungen mit Ausgleich bei  $32^\circ$  (Abbildung 3.9) und weisen eine geringe Streuung über die wahren Ausrichtungen auf.

### 3.3.2.3 Experiment 7: Dichotische Schätzung der Sprecherausrichtung mit Sprachsignalen und mit synthetischen Signalen

In diesem Experiment wurden sowohl Sprache als auch synthetische Signale als Testsignale verwendet. Die Ergebnisse werden im folgenden getrennt vorgestellt. Die Verwechslungsmatrizen des gesamten siebten Experiments mit normalhörenden Probanden unter Verwendung sind im Anhang A.2.3 dargestellt.

**Testsignale: Sprache von natürlichen Sprechern und Lautsprecherbox Fostex6301B**

Die Abbildung 3.10 enthält die Fehlermaße in Abhängigkeit von den Quell-



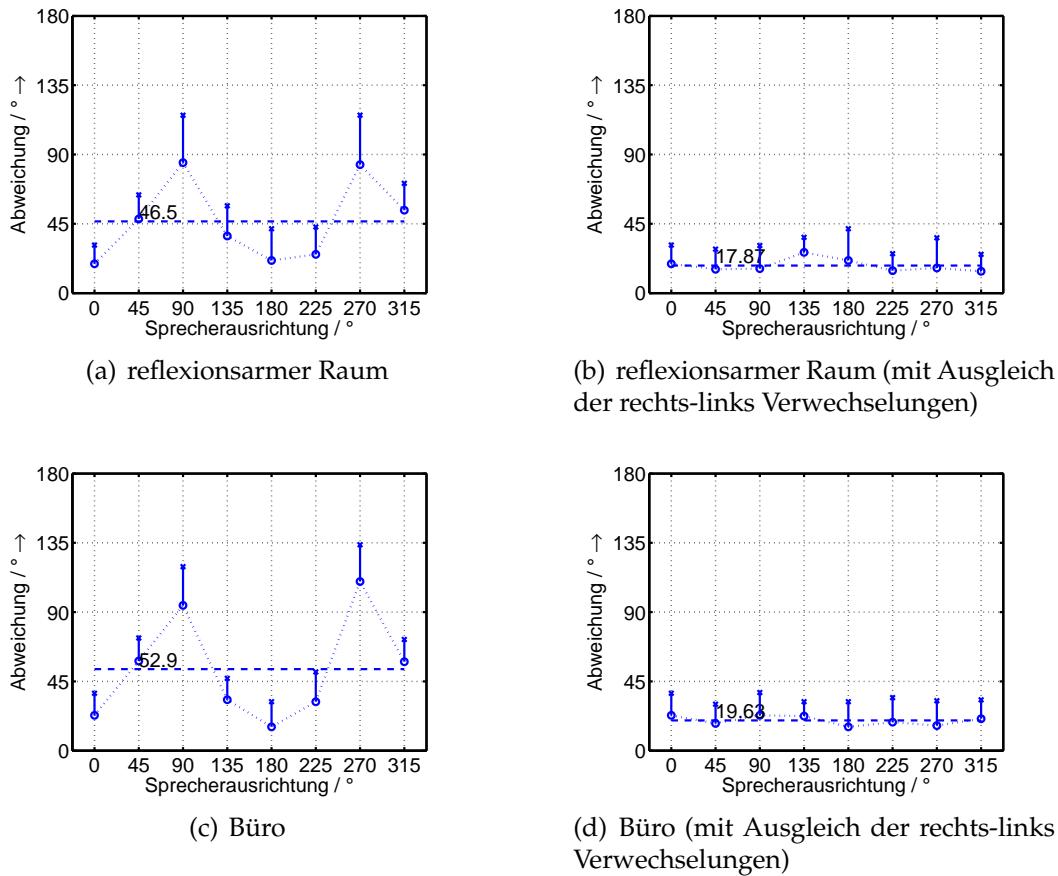
**Abb. 3.10:** Ergebnisse Experimente 7: Dichotische Schätzung der Sprecherausrichtung mit Sprachsignalen von normalhörenden Probanden. Grafik zeigt die mittleren Abweichungen  $a_m$  (•) und die Wurzel der mittleren quadratischen Abweichungen  $a_q$  (x) der Probandenbewertung in Abhängigkeit von den Quellsignalen. (blau: Bewertungen der Signale im reflexionsarmen Raum, grün: Bewertungen der Signale im Büroraum, rot: Bewertungen der Signale im reflexionsarmen Raum mit rechts-links Ausgleich, schwarz: Bewertungen der Signale im Büroraum mit rechts-links Ausgleich)

signalen. Die Graphen 3.11(a) bis 3.11(d) enthalten die gemittelten Abweichungen für alle ungefilterten Sprachsignale über den wahren Ausrichtungen. Die Verwechslungsmatrizen weisen ähnlich Streuungen auf, wie im fünften Experiment, bei dem die Signale monoton dargeboten wurden. Die Probanden waren nicht in der Lage, Signale von links oder rechts trotz der dichotischen Darbietung klar zu unterscheiden.

#### Ergebnisse ohne Ausgleich möglicher rechts-links Verwechslungen

Abbildung 3.10 zeigt, dass in diesem Experiment erstmals deutliche Unterschiede in Abhängigkeit von den verwendeten Quellsignalen auftreten. Die weibliche Sprecherin wird deutlich schlechter den korrekten Ausrichtungen zugewiesen und das OLSA-Sprachsignal deutlich besser, als die restlichen Audiodateien.

Die Ergebnisse in Abbildung 3.11(a) zeigen eine durchschnittliche Abweichung von 47 % im reflexionsarmen Raum. Bei einer Sprecherausrichtung von 0° und 180° werden lokale Minima mit ≈ 20° erreicht. Besonders bei 90° und 270° sind die Abweichungen mit ≈ 90° maximal. Das durchgeführte Testdesign kann mit dem Experiment von Neuhoff verglichen werden (siehe 2.2.1). Es zeigt sich, dass die Ergebnisse von Kato et al. mit einer durchschnittlichen Abweichung von 23,5° deutlich besser ausfallen. Ebenfalls können die Probanden die Sprecherausrichtung von 0° mit einer Abweichung von 6,6° deutlich besser detektieren. Grundsätzlich zeigen beide Ergebnisse,



**Abb. 3.11:** Ergebnisse Experimente 7: Dichotische Schätzung der Sprecherausrichtung mit Sprachsignalen von normalhörenden Probanden. Grafiken zeigen die mittleren Abweichung  $a_m$  (●) und die Wurzel der mittleren quadratischen Abweichung  $a_q$  (x) und Durchschnitt der mittleren Abweichung der Probandenbewertung in Abhängigkeit von den wahren Sprecherausrichtungen.

dass die Probanden die Sprecherausrichtung am besten erkennen, wenn die Quelle direkt auf sie zu- oder abgewandt ist. Bei seitlichen Ausrichtungen sinken die Erkennungsleistungen.

Die Ergebnisse in Abbildung 3.11(c) zeigen eine durchschnittliche Abweichung von 53 % im halligen Büror Raum. Der Kurvenverlauf ist fast identisch zu den Ergebnissen des reflexionsarmen Raums in Abbildung 3.11(a). Das durchgeführte Testdesign kann mit dem statischen Experiment von Neuhoff verglichen werden (siehe 2.2.2). Bei Neuhoff wird für eine Sprecherausrichtung von  $0^\circ$  ebenfalls eine mittlere Abweichung von  $\approx 20^\circ$  erreicht und die Schätzungen steigen mit größeren Ausrichtungen an. Eine große Diskrepanz zeigt sich, bei einer Ausrichtung von  $180^\circ$ . Bei Neuhoff zeigt sich hier die schlechteste Erkennungsleistung und beim Experiment 7 bildet sich ein lokales Minimum. Das Minimum lässt sich vermutlich dadurch erklären, dass bei einem Signal der Sprecherausrichtung von  $180^\circ$  die maximale Un-

terschiede in den Signalen wahrnehmbar sind und die Probanden diese im Vergleich gut der entsprechenden wahren Ausrichtung zuweisen können.

Allerdings muss beachtet werden, dass bei allen durchgeführten Experimenten die mittlere Leistung der Signale ausgeglichen wurde und den Probanden so möglicherweise eine wichtige akustische Eigenschaft genommen wurde.

### **Ergebnisse mit Ausgleich möglicher rechts-links Verwechslungen**

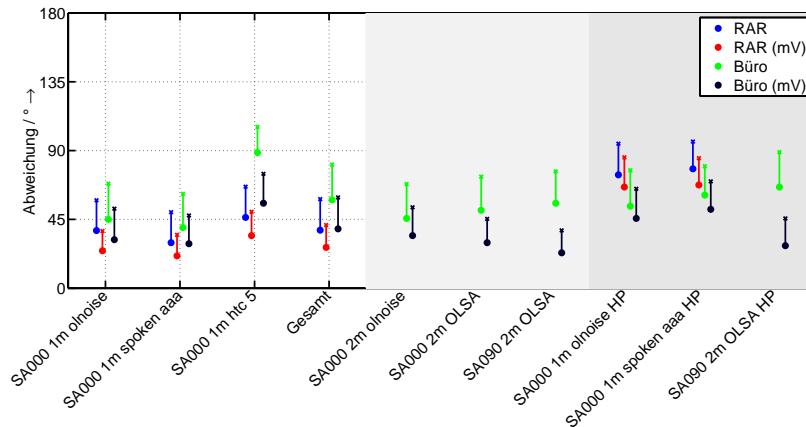
Obwohl in Abbildung 3.8 ohne Ausgleich der rechts-links Verwechslungen zuvor deutliche Unterschiede in Abhängigkeit der Quellsignale sichtbaren waren, reduzieren sich die Unterschiede mit einem entsprechenden Ausgleich. Bei den abweichenden Signalen traten somit entweder deutlich mehr, bzw. deutlich weniger Verwechslungen auf. Insgesamt sind über alle Signale keine großen Differenzen nach dem Ausgleich ersichtlich.

Die zeigt sich auch in den Abbildungen 3.11(b) und 3.11(d). Hier sind weisen alle Bewertungen nach einem Ausgleich der rechts-links Verwechslungen eine konstante durchschnittliche Abweichung von  $\approx 19^\circ$  über alle wahren Ausrichtungen auf. Die Probanden sind somit gut in der Lage, die Sprecherausrichtung zu schätzen. Allerdings sind die Ergebnisse nicht besser als bei der monotischen Darbietung der Signale (vgl. Abbildung 3.7(b) und 3.7(d)). Die Probanden nutzen somit keine oder wenige binauralen Effekte für die Unterscheidung der Sprecherausrichtung aus.

### **Testsignale: synthetische Signale aus Lautsprecherbox Fostex6301B**

Die Abbildung 3.12 enthält die Fehlermaße in Abhängigkeit von den Quellsignalen. Aufgrund der Vielzahl der Ergebnisse werden im folgenden nur die zusammengefassten Ergebnisse über die einzelnen Signale beschrieben.

Neben den unterschiedlichen Signalen mit und ohne Hochpassfilterung wurden zusätzlich abweichende Sprecherpositionen mit  $\theta_0 = 90^\circ$  und Abstände von 2 m im halligen Raum getestet. Das sprechsimulierende Rauschen und der synthetische Vokal /a/ weisen die besten Ergebnisse aller durchgeführten Tests ohne Ausgleich der rechts-links Verwechslungen auf. Dies lässt sich dadurch erklären, dass der HTC nur Frequenzkomponenten bis 750 Hz enthält, bei der der verwendete Lautsprecher noch relativ omnidirektion abstrahlt (vgl. Abschnitt 2.3). Die Ergebnisse für einen Abstand von zwei Meter unterscheiden sich bei den verwendeten Signalen nur geringfügig von den Ergebnissen bei einem Meter. Die hohen Abweichungen für die hochpassgefilterten Signale im reflexionsarmen Raum zeigen erneut, dass die richtungsabhängige Klangverfärbung hier ein wichtiges Maß für die Bewertung der Probanden ist. Im halligen Bürraum verbessern sich die Ergebnisse aufgrund der enthaltenen Reflexionen.



**Abb. 3.12:** Ergebnisse Experiment 7: Dichotische Schätzung der Sprecherausrichtung mit synthetischen Signalen von normalhörenden Probanden. Grafik zeigt die mittleren Abweichung  $a_m$  (●) und die Wurzel der mittleren quadratischen Abweichung  $a_q$  (x) der Probandenbewertung in Abhängigkeit von den Quellsignalen. (blau: Bewertungen der Signale im reflexionsarmen Raum, grün: Bewertungen der Signale im Büroraum, rot: Bewertungen der Signale im reflexionsarmen Raum mit rechts-links Ausgleich, schwarz: Bewertungen der Signale im Büroraum mit rechts-links Ausgleich)

### 3.3.2.4 Experiment 8: Dichotische Schätzung der Sprecherausrichtung mit LPC-Residual-Signalen

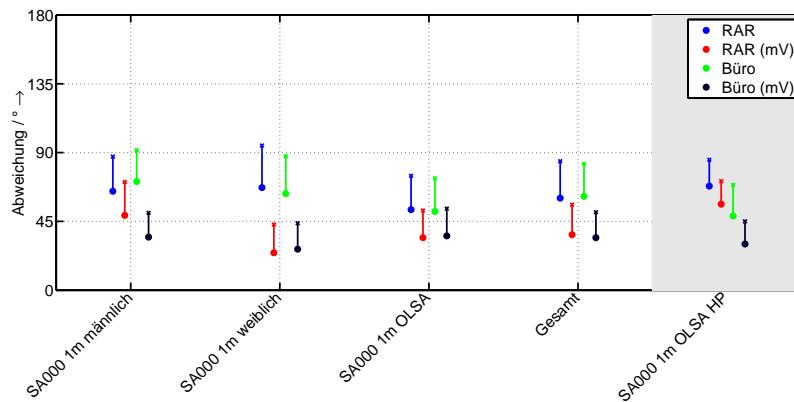
Die Verwechslungsmatrizen des achten Experiments mit normalhörenden Probanden sind im Anhang A.2.4 dargestellt. Die Abbildung 3.13 enthält die Fehlermaße in Abhängigkeit von den Quellsignalen. Die Graphen 3.14(a) bis 3.14(d) enthalten die gemittelten Abweichungen für alle ungefilterten Quellsignale über den wahren Ausrichtungen.

#### Ergebnisse ohne Ausgleich möglicher rechts-links Verwechslungen

Die gemittelten Verwechslungsmatrizen in den Abbildungen A.14(a) bis A.14(c) zeigen mit dem Experiment 6 (Abbildungen A.8(a) bis A.8(c)) vergleichbare Ergebnisse. Gleiches gilt für die Ergebnisse in Abhängigkeit der verwendeten Quellsignale in den Abbildungen 3.8 und 3.13. Obwohl sich beim Vergleich der einzelnen Quellsignale individuelle Unterschiede zeigen, sind keine signifikanten Unterschiede bei der Gesamtbewertung über alle Dateien erkennbar.

Dies gilt auch für die mittleren Abweichungen über die wahren Sprecherausrichtungen (Abbildung 3.14(a) bis 3.14(c)). Sowohl Kurvenverlauf als auch die Mittelwerte über alle wahren Sprecherausrichtungen unterscheiden sich in den betrachteten Räumen nur geringfügig.

#### Ergebnisse mit Ausgleich möglicher rechts-links Verwechslungen



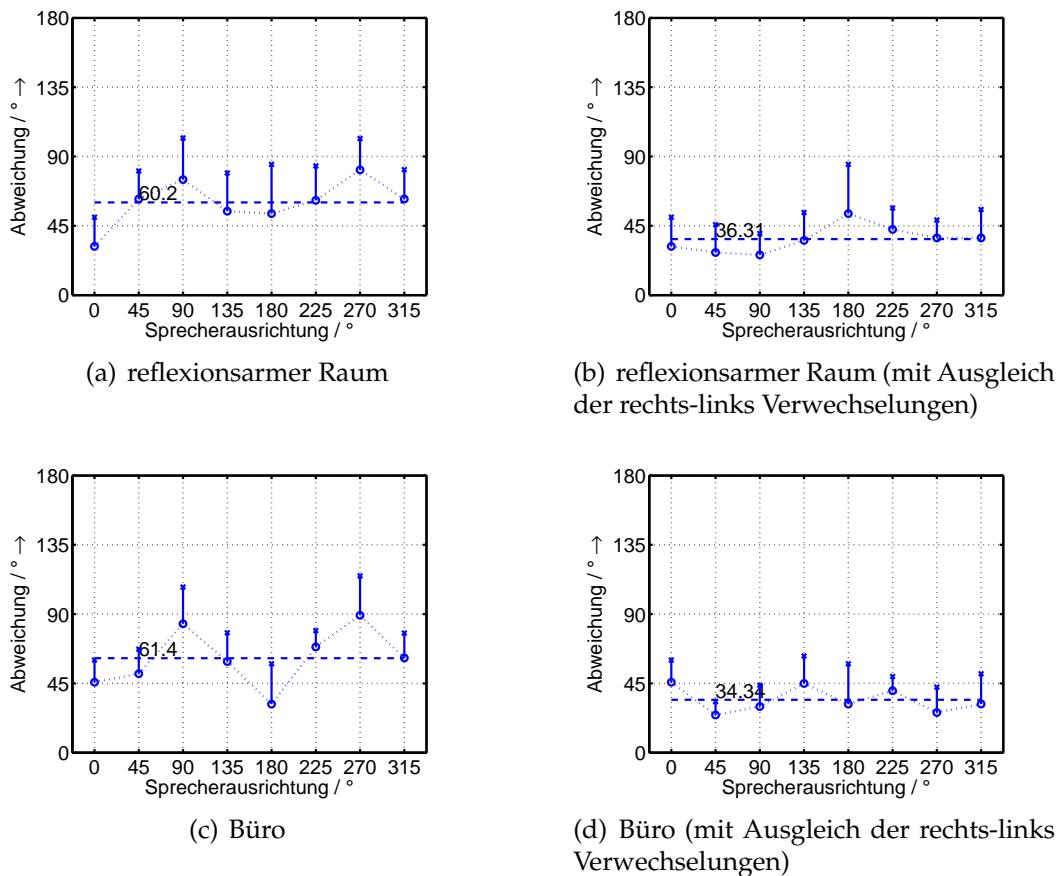
**Abb. 3.13:** Ergebnisse Experimente 8: Dichotische Schätzung der Sprecherausrichtung mit LPC-Residual-Signalen von normalhörenden Probanden. Grafik zeigt die mittleren Abweichung  $a_m$  (●) und die Wurzel der mittleren quadratischen Abweichung  $a_q$  (x) der Probandenbewertung in Abhängigkeit von den Quellsignalen. (blau: Bewertungen der Signale im reflexionsarmen Raum, grün: Bewertungen der Signale im Bürraum, rot: Bewertungen der Signale im reflexionsarmen Raum mit rechts-links Ausgleich, schwarz: Bewertungen der Signale im Bürraum mit rechts-links Ausgleich)

Mit entsprechendem Ausgleich zeigen sich keine signifikanten Unterschiede in der mittleren Abweichung über alle Signale zwischen der monotonischen und der dichotischen Darbietung (vgl. Abbildungen 3.9 und 3.14). Daraus lässt sich schließen, dass die Probanden wie erwartet keine binauralen Effekte ausnutzen können, um die Ausrichtung der Quelle nach rechts oder links zu unterscheiden. Vorteile für die Schätzung des Ausrichtungswinkel zeigen sich durch die dichotische Darbietung der LPC-Residual Signale somit nicht.

### 3.3.3 Erkenntnisse aus den Messergebnissen der durchgeführten Experimente zur Bewertung der Sprecherausrichtung mit normalhörenden Probanden

Bei allen Experimenten konnte gezeigt werden, dass für die Bewertung der Sprecherausrichtung nur die spektralen Verfärbungen in Abhängigkeit von der Ausrichtung genutzt werden können. Im halligen Bürraum führen die zusätzlichen Reflexionen zu weiteren akustischen Eigenschaften, so dass die Unterscheidung der Signale trotz Ausgleich des richtungsabhängigen Frequenzabfalls möglich ist. Dies entspricht den Erwartungen aus Abschnitt 2.4.

Eine Unterscheidung zwischen  $0^\circ$  und  $180^\circ$  gelingt den Probanden mit



**Abb. 3.14:** Ergebnisse Experimente 8: Dichotische Schätzung der Sprecherausrichtung mit LPC-Residual-Signalen von normalhörenden Probanden. Grafiken zeigen die mittleren Abweichung  $a_m$  (•) und die Wurzel der mittleren quadratischen Abweichung  $a_q$  (x) und Durchschnitt der mittleren Abweichung der Probandenbewertung in Abhängigkeit von den wahren Sprecherausrichtungen.

über 83 % Korrektklassifikationsrate sowohl im Bürroraum als auch im reflexionsarmen Raum gut. Die Unterschiede zwischen den Ergebnissen werden bei einer Sprecherausrichtung von  $90^\circ$  gegenüber  $0^\circ$  geringer, eine Unterscheidung ist mit einer Korrektklassifikationsrate von 75 % aber noch immer möglich.

Überraschend ist, dass die Probanden zwischen Test und Re-Test bei den Ja-Nein Experimenten keinen Lerneffekt gezeigt haben. Identische Audiodateien wurden auch in dem Zuordnungsexperiment verwendet. Es wurde erwartet, dass die wiederholte Darbietung der Signale zu einer verbesserten Erkennung beim Re-Test führen.

Nicht erklärt werden kann, warum einige Ergebnisse der Ja-Nein-Experimente bei einer von  $\theta_0 = 0^\circ$  abweichenden Sprecherposition zu zum Teil besseren Ergebnissen führt. Da das Gehör bei seitlichem Schalleinfall häufig unempfindlicher wird, z.B. beim MAA (siehe [Mil58]), wurde

### **Kapitel 3. Analyse der menschlichen Fähigkeit zur Schätzung der Sprecherausrichtung mit normalhörenden Probanden**

---

auch hier eine schlechtere oder gleiche Erkennung der Sprecherausrichtung erwartet. Eine Überprüfung, ob die Ergebnisse statistisch signifikant sind, muss noch erfolgen.

Bei der Verwendung der LPC-Residual-Signale zeigt sich, dass in diesen Signalen ausreichend akustische Eigenschaften zur Bestimmung der Sprecherausrichtung enthalten sind. Sie können somit ohne der Filterung des Vokaltraktes für eine Bewertung der Sprecherausrichtung in Algorithmen verwendet werden.

Für alle getesteten Situationen ist zu erkennen, dass die Probanden nicht in der Lage sind, eine Sprecherausrichtung nach links oder rechts sicher zu unterscheiden. Ein signifikanter Unterschied zwischen der monotischen und dichotischen Darbietung der Signale mit Ausgleich möglicher rechts-links Verwechslungen kann weder für unverarbeitete Sprache noch für die LPC-Residual-Signale erkannt werden. Somit scheinen die Probanden keinen oder nur geringen Nutzen aus binauralen akustischen Informationen in den Signalen für die Bewertung der Sprecherausrichtung zu ziehen.

Es zeigt sich, dass die Bewertungen der Probanden unabhängig von den verwendeten Sprachquellen sind. Beim Vergleich zwischen den natürlichen Sprechern und dem Sprachsignal aus der Lautsprecherbox zeigen sich keine signifikanten Unterschiede. Bessere Ergebnisse zeigen sich bei der Verwendung des sprachsimulierenden Rauschens und des synthetischen Vokals /a/. Diese Signale weisen die besten Bewertungen, selbst mit rechts-links Verwechslungen, auf. Eine mögliche Erklärung ist, dass diese Signale zeitlich stationär sind und dadurch besser von den Probanden verglichen werden können.

---

# Kapitel 4

## Analyse der algorithmischen Fähigkeit zur Schätzung der Sprecherausrichtung

Auf Grundlage der bisherigen Untersuchungen und Erkenntnisse werden in diesem Kapitel mögliche Hörgerätealgorithmen zur Schätzung und Unterscheidung der Sprecherausrichtung unter Verwendung von binaural verschalteten Hörgeräten vorgestellt. Abgeschlossen wird dieses Kapitel mit den Ergebnissen einer Evaluation, bei denen die vorgestellten Hörgerätealgorithmen die gleichen Hörversuche wie die Probanden zu absolvieren hatten.

### 4.1 Entwickelte Signalverarbeitungsalgorithmen zur Schätzung der Sprecherausrichtung

Bisherige Algorithmen zur Schätzung der Sprecherausrichtung, wie sie in Abschnitt 1.5 vorgestellt werden, können bei Hörgeräten nicht angewandt werden. Die bisherigen Algorithmen erwarten entweder ein sehr großes Mikrofonarray oder zumindest mehrere kleine Arrays die im Raum verteilt sind. Von Vorteil ist hier, dass die Position der Mikrofone als bekannt vorausgesetzt werden kann. Anders sieht es bei Hörgeräten aus. Selbst binaural verschaltete, mehrkanalige Hörgeräte weisen einen maximalen Abstand auf, der durch den Kopfdurchmesser bestimmt wird. Zusätzlich kann sich der Hörgeräteträger frei im Raum bewegen, so dass die absolute Position der Mikrofone im Raum unbekannt ist. Es wurde bisher gezeigt, dass die Raumeigenschaften einen großen Einfluss auf die Mikrofonsignale in Abhängigkeit von der Sprecherausrichtung und Position haben. Ebenso verhält es sich mit

der Empfängerposition und -ausrichtung. Eine gemessene oder geschätzte akustische RIA ist immer nur für die jeweilige Mikrofon- und Quellenposition und Quellenausrichtung gültig.

Um die Komplexität der vorgestellten Algorithmen zu reduzieren, wurde im Folgenden davon ausgegangen, dass sich die Quelle auf gleicher Höhe wie die Ohren des Kunstkopfes bzw. der Hörgerätemikrofone befindet. Dies führt u. a. bei der Anwendung des globalen Kohärenzfeldes zu einer zweidimensionalen Betrachtung anstatt einer Dreidimensionalen.

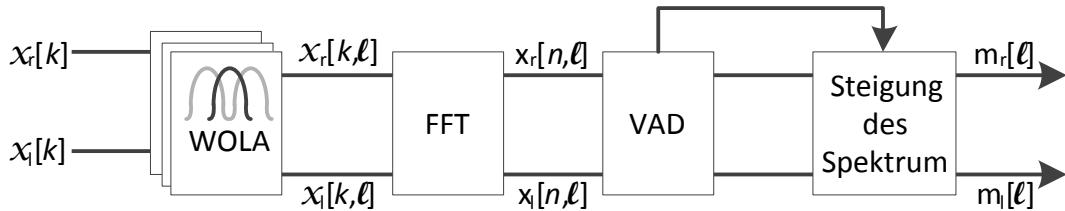
### 4.1.1 Algorithmus zur Schätzung der Sprecherausrichtung unter Verwendung der mittlere Steigung des Frequenzgangs

Wie in Kapitel 2 beschrieben und mithilfe der Probandenversuche in Kapitel 3 bestätigt, ist die Klangverfärbung des Sprachsignals eine wichtige akustische Eigenschaft, die von der Sprecherausrichtung abhängt. Diese Abhängigkeit zeigte sich sowohl im reflexionarmer Raum als auch im halligen Büroraum. Erklären lässt sich dieser Zusammenhang durch die richtungsabhängige Abstrahlcharakteristik des menschlichen Kopfes, die in Abschnitt 2.3 betrachtet wird. Der menschliche Kopf als direktionale Quelle wirkt mit steigender Sprecherausrichtung immer stärker als Tiefpassfilter, da hohe Frequenzen durch die Kopfabschattung mehr und mehr gedämpft werden.

In [CFSC<sup>+</sup>08] und [KTNM10] wird die Tiefpasswirkung des menschlichen Kopfes als eine mögliche akustische Eigenschaft zur Bewertung der Sprecherausrichtung und beispielsweise das Energieverhältnis zwischen einem hohen und tiefen Frequenzband als Indiz für die algorithmischen Bestimmung der Sprecherausrichtung vorgeschlagen.

Aufbauend auf dieser Erkenntnis wurde der in Abbildung 4.1 als Blockschaltbild dargestellte Algorithmus entwickelt und in *MATLAB* implementiert. Der Algorithmus basiert auf zwei diskreten Zeitsignalen  $x_r[k]$  und  $x_l[k]$  von binaural verschalteten Hörgeräten. In der vorliegenden Implementation von den vorderen Mikrofone der zwei Triano 3 Hörgerätedummies.

Die Zeitsignale werden in  $L$  überlappende und gefensterte Blöcke aufgeteilt. Die resultierenden Signale  $x_r[k, \ell]$  und  $x_l[k, \ell]$  des Blockes  $\ell$  werden mittels FFT in die diskreten Kurzzeitspektren  $X_r[n, \ell]$  und  $X_l[n, \ell]$  mit dem Frequenzindex  $n$  transformiert. Pro Block  $\ell$  bestimmt ein VAD ob der jeweilige Block Sprache enthält. In der entwickelten Matlab-Implementation des beschrie-



**Abb. 4.1:** Blockschaltbild zur Veranschaulichung des Algorithmus zur Schätzung der Sprecherausrichtung mittels gemittelten Sprachspektrums. Pro überlappendem Zeitblock wird die Steigung des Frequenzgangs des Leistungsdichtespektrums berechnet. Angenommen wird, dass die Abstrahlcharakteristik der Quelle in Abhängigkeit von der Sprecherausrichtung einen Tiefpasscharakter aufweist und die Steigung des Leistungsdichtespektrums über die Frequenzen bei steigender Sprecherausrichtung im Mittel abnimmt.

benen Algorithmus wurde der VAD nach [MK02] verwendet. Enthält der aktuelle Block  $l$  Sprache, so wird für diesen Block die Steigung  $m$  des logarithmierten Betragsspektrums im Hauptsprachbereich mit Hilfe einer linearen Regression entsprechend der Geradengleichung  $Y[n] = m \cdot X[n] + c$  bestimmt. Für eine stationäre Sprecherausrichtung ergibt sich eine Schätzung der mittleren Steigung des Frequenzverlaufs bei aktiver Sprache als arithmetisches Mittel aus

$$\bar{m}_{r,l} = \mu(m_{r,l}[\ell_s]) \quad (4.1)$$

oder als Quantil

$$\tilde{m}_{r,l} = Q_q(m_{r,l}[\ell_s]) \quad (4.2)$$

mit der Steigung  $m[\ell_s]$  der sprachenthaltenden Blöcke  $\ell_s$ , dem Mittelwert  $\mu(\cdot)$  und dem  $q$ -Quantil  $Q_q(\cdot)$ . Bei der Evaluation des Algorithmus wurden Quantile als Mittelung verwendet. Bei nicht stationären Sprecherausrichtungen bieten sich gleitende Mittelwerte an. Dies ist dadurch begründet, dass Quantile robuster gegenüber Ausreißern sind. Bei einer Paramteroptimierung mittels Monte-Carlo-Simulation wurden die Quantile  $Q_{.1}, Q_{.2}, \dots, Q_{.9}$  zur Verbesserung der Schätzleistung untersucht. Es hat sich gezeigt, dass sich das 0.5-Quantil am besten eignet. Zusätzlich wurden unterschiedliche Blöcklängen untersucht.

Entsprechend der Position des Sprechers muss abschließend entschieden werden, von welcher Kopfseite das Maß zur Bewertung der Sprecherausrichtung herangezogen wird. Sinnvoll ist daher der Einsatz einer zusätzlichen Sprecherpositionsbestimmung wie in Abschnitt 1.5.1 vorgestellt.

## **4.1.2 Algorithmus zur Schätzung der Sprecherausrichtung unter Verwendung des globalen Kohärenzfeldes**

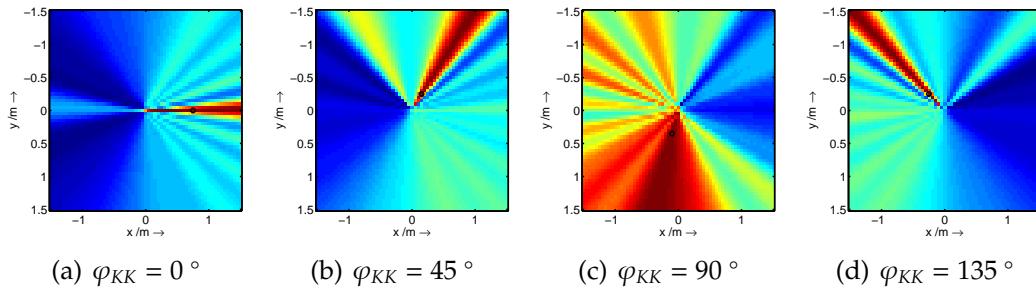
Im Abschnitt 1.5.3 wird beschrieben, dass das oGCF-Verfahren ein mögliches Verfahren für die Schätzung der Sprecherausrichtungen in Räumen ist. Versuche haben gezeigt, dass es nicht für die Verwendung in Hörgeräten geeignet ist, da der maximal mögliche Mikrofonabstand durch den Kopfdurchmesser  $a$  auf etwa 18 cm zu begrenzt ist. Es ist ersichtlich, dass es mit einer solchen Mikrofonanordnung unmöglich ist, mehrere äquidistante Positionen um die Quelle herum zur Analyse der Sprecherausrichtung heranzuziehen.

Allerdings wurde in Abschnitt 1.1 beschrieben, dass die Raumimpulsantwort durch ein Spiegelquellenmodell beschrieben werden kann. Bei diesem Modell „erscheint die Schallquelle an jeder Begrenzungsfläche gespiegelt; eine unendlich ausgedehnte Wand wird ersetzt durch eine spiegelbildlich angeordnete zweite Quelle. Die Intensität der Spiegelquelle ist durch den Absorptionsgrad der Wand bestimmt“ [Gör08]. Die Eigenschaft, dass Reflexionen als Spiegelquellen betrachtet werden können, führte zu der Annahme, dass diese angenommenen Quellen auch im globalen Kohärenzfeld als weitere Quellen identifiziert und für die Bestimmung der Sprecherausrichtung genutzt werden können. Neben einer Hauptquelle aus der Sprecherrichtung müssten weitere Quellen aus den Richtungen der Reflexionen detektierbar sein.

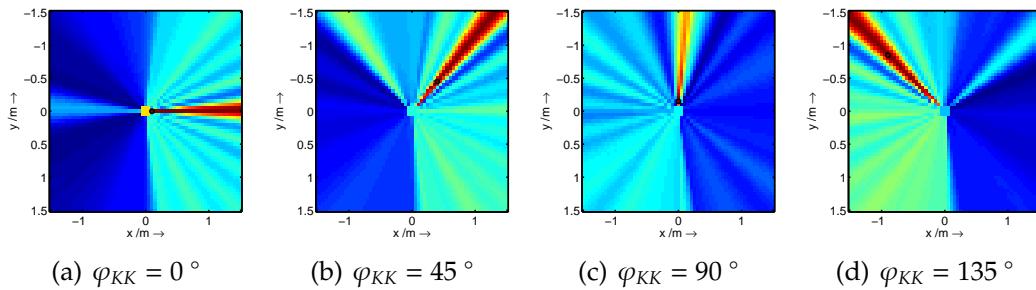
Bevor diese Idee umgesetzt werden konnte, musste das GCF an die akustischen Begebenheiten des Kopfes angepasst werden. Im Folgenden werden die durchgeführten Anpassungen, die Entwicklung eines datenreduzierten GCF und der implementierte Algorithmus zur Schätzung der Sprecherausrichtung beschrieben.

### **4.1.2.1 Laufzeitausgleich zur Verwendung des globalen Kohärenzfeldes mit binauralen Hörgeräten**

Für den Einsatz in binauralen Hörgeräte kann das globale Kohärenzfeld nicht direkt verwendet werden, da die sich ausbreitenden Schallwellen durch die Beugung am Kopf mit zusätzlichen, von der Quellposition abhängigen interauralen Laufzeitunterschieden behaftet sind (vgl. Abschnitt 1.2.1). Verdeutlicht werden die resultierenden Fehler in den Abbildungen 4.2(a) bis 4.2(d). Dargestellt sind die zweidimensionalen globalen Kohärenzfelder der sechs Hörgerätemikrofone des verwendeten Hörgerätedummies am Kunstkopf in der Draufsicht. Die Rasterpunkte be-



**Abb. 4.2:** Beispielhafte Darstellungen des globalen Kohärenzfeldes mit realen Aufnahmen von zwei binauralen Hörgeräten mit jeweils drei Mikrofonen in Abhängigkeit von der Ausrichtung des Empfängers  $\varphi_{KK}$ . Der Empfänger befindet sich im Zentrum der Abbildungen. Die Farbkodierung gibt die Wahrscheinlichkeit an, ob ein Sprecher aus der entsprechenden Richtung spricht. Rot entspricht einer hohen Wahrscheinlichkeit, blau einer geringen. Deutlich erkennbar ist, dass die Schätzung speziell für  $\varphi_{KK} = 90^\circ$  zu starken Abweichungen vom gewünschten Ergebnis führt.



**Abb. 4.3:** Beispielhafte Darstellungen des modifizierten globalen Kohärenzfeldes mit realen Aufnahmen von zwei binauralen Hörgeräten mit jeweils drei Mikrofonen in Abhängigkeit von der Ausrichtung des Empfängers  $\varphi_{KK}$ . Der Empfänger befindet sich im Zentrum der Abbildungen. Die Farbkodierung gibt die Wahrscheinlichkeit an, ob ein Sprecher aus der entsprechenden Richtung spricht. Es ist zu erkennen, dass die durchgeführte kopfbezogene Laufzeitanpassung in allen betrachteten Fällen zum gewünschten Ergebnis führt.

sitzen einen Abstand von 5 cm. Die Aufnahmen erfolgten im halligen Bürroraum mit einer Sprechquelle (Sprecherausrichtung  $\varphi_s = 0^\circ$ ) im Abstand von einem Meter zum Kunstkopf. Das GCF wurde im Mittelpunkt des Kunstkopf zentriert, so das dieser bei  $x = y = 0$  positioniert ist. Pro Aufnahme wurde der Winkel des Kunstkopfes  $\varphi_{KK}$  von  $0^\circ$  bis  $135^\circ$  in  $45^\circ$  Schritten gedreht. Für den Winkel von  $0^\circ$  kann die Richtung der Quelle deutlich anhand des schmalen Maximums des GCFs detektiert werden. Die Maxima bei den Winkeln  $45^\circ$  und  $135^\circ$  sind bereits deutlich breiter, aber die Richtung der Quelle gut zu detektieren. Bei einer Empfängerausrichtung von  $90^\circ$  ist zu erwarten, dass sich das Maximum des GCF in Abbildung 4.2(c) nach oben hin ausbildet. Stattdessen zeigen sich mehrere lokale Maxima rund um die Hörerposition und das Hauptmaximum liegt bei etwa  $-100^\circ$ .

Zur Auflösung der auftretenden Laufzeitfehler müssen die benötigten Verzögerungen aller betrachteten Raumrasterpunkte an die kopfbezogenen Laufzeitverzögerungen angepasst werden. Zur Berechnung dieser Laufzeitverzögerungen kann die Raytracing Formel nach Woodworth und Schlossberg [WS62] zur Schätzung der Laufzeitdifferenzen und der ITD eingesetzt werden [Bla97]. Unter der Annahme eines Kugelkopfes und der Annahme dass die Schallgeschwindigkeit bei freier Schallausbreitung und an der Kugeloberfläche identisch ist, müssen zwei Fälle für die Berechnung der Wegdifferenz  $d$  unterschieden werden. [DM98]

- **Fall 1:** Wenn der Empfängerort sich auf der Halbkugel befindet, die der Quelle zugewandt ist, berechnet sich die Wegstrecke zwischen Quelle und Empfängerort aus

$$d = \sqrt{r^2 + a^2 - 2ra \cos \theta}, \quad (4.3)$$

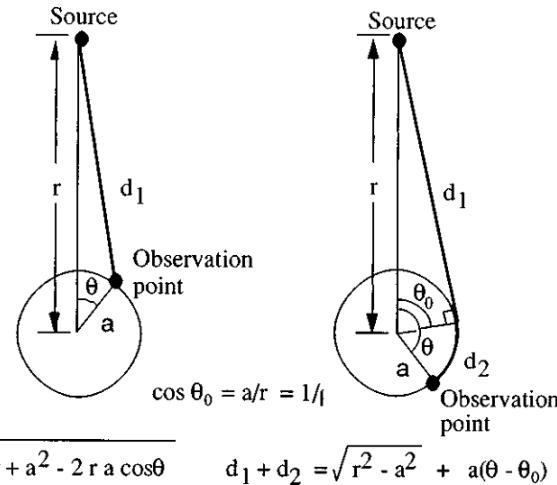
- **Fall 2:** Wenn der Empfängerort sich auf der Halbkugel befindet, die der Quelle abgewandt ist, berechnet sich die Wegstrecke zwischen Quelle und Empfängerort aus

$$d = \sqrt{r^2 - a^2} + a(\theta - \theta_0), \quad (4.4)$$

mit dem Abstand  $r$  zwischen dem Kugelmittelpunkt und dem Ort der Quelle, dem Kugelkopfdurchmesser  $a$ , der Auslenkung des Empfängerortes  $\theta$  und dem Tangentialwinkel  $\theta_0$ . Grafisch ist dieses Modell in Abbildung 4.4 dargestellt.

Die Abbildungen 4.3(a) bis 4.3(d) zeigen die resultierenden globalen Kohärenzfelder  $GCF^{Head}$  unter Verwendung der kopfbezogenen Laufzeiten nach der Woodworth/Schlossberg Formel in Abhängigkeit von der Kopfausrichtung  $\varphi_{KK}$ . Im Vergleich zu den nicht korrigierten GCFs zeigt sich, dass die Maxima deutlich schmäler ausgeprägt sind und im Fall von  $\varphi_{KK} = 45^\circ$  deutlicher in die zu erwartende Richtung zeigt. Bei einer Kopfausrichtung von  $\varphi_{KK} = 90^\circ$  kann nun ebenfalls die Quellenrichtung korrekt bestimmt werden.

Beim GCF sind auftretende interaurale Pegeldifferenzen zu vernachlässigen, da die Betragsspektren der Signale durch die verwendete Phasentransformation verworfen werden (vgl. Abschnitt 1.5.2).



**Abb. 4.4:** Darstellung der Geometrie für die Woodworth/Schlossberg Formel [WS62] zur Berechnung der Laufzeit an einem idealen Kugelkopfmodell. Unter der Annahme einer konstanten Schallgeschwindigkeit  $c$ , ergibt sich die Laufzeit entweder aus der direkten Verbindung  $d_1$  von der Quelle zum Endpunkt oder aus der Summe von  $d_1$  und dem Kreisbogen  $d_2$  (aus [DM98]).

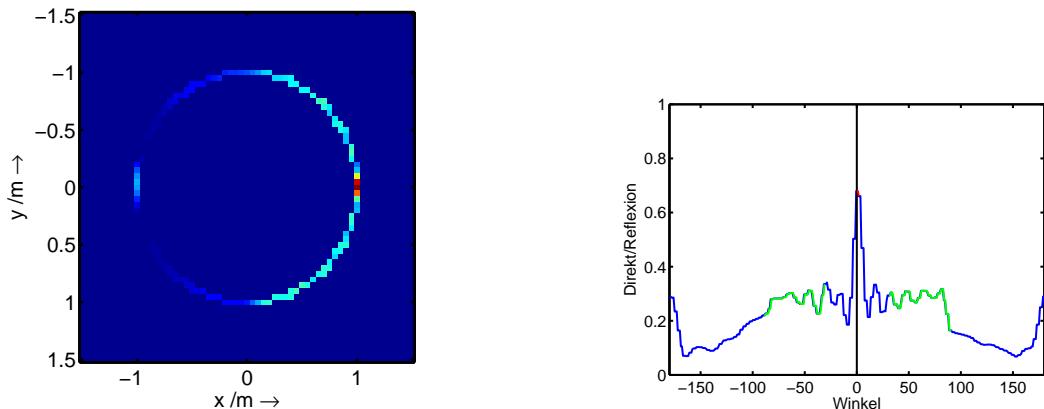
#### 4.1.2.2 Datenreduktion des globalen Kohärenzfeldes für die Schätzung der Sprecherausrichtung mit binauralen Hörgeräten

In den Abbildungen 4.3(a) bis 4.3(d) zeigt sich, dass der geringe Kopfdurchmesser  $a$  relativ klein ist im Vergleich zur betrachteten Fläche von  $1,5 \text{ m} \times 1,5 \text{ m}$  und dem gewählten Abstand zwischen den betrachteten Raumpunkten. Dies führt dazu, dass sich die Struktur des globalen Kohärenzfeldes sternförmig um das Zentrum ausbreitet. Das berechnete GCF ist in der betrachteten Konstellation vornehmlich abhängig vom Azimuth  $\theta$  zwischen Blickrichtung des Empfängers und der Quelle und nicht von der Entfernung zur Quelle. Das resultierende GCF enthält somit nicht erforderliche, redundante Daten, die sich radial um die Empfängerposition ausbreiten. Zur Datenreduktion und Vereinfachung erscheint es ausreichend, die Werte des GCF nur auf einem Kreisbogen im Abstand  $r$  um das Zentrum zu betrachten. Daraus resultiert, bei zweidimensionaler Betrachtung, das modifizierte, kopfbezogene globale Kohärenzfeld  $mGCF^{Head}$  aus

$$mGCF^{Head}(\ell, \theta) = GCF^{Head}\left(\ell, \operatorname{Re}\left\{r \cdot e^{i\theta}\right\}, \operatorname{Im}\left\{r \cdot e^{i\theta}\right\}\right), \quad (4.5)$$

mit dem kopfbezogenen globalen Kohärenzfeld  $GCF^{Head}$ , dem Blockindex  $\ell$ , dem Azimuth  $\theta \in [-\pi, \pi]$  und dem Radius  $r$ .

Abbildung 4.5(a) zeigt ein Beispiel der extrahierten Werte aus einem globalen Kohärenzfeld mit Ausgleich der kopfbezogenen Laufzeitunterschiede als zweidimensionale Darstellung. Für den folgenden Algorithmus werden



(a) Resultat nach der Datenreduktion redundanter Informationen aus dem modifizierten globalen Kohärenzfeld. Erzielt wird die Datenreduktion durch die Extraktion der Daten des GCF auf einem Ring um die Hörerposition.

(b) Modifizierten GCF bei einer Sprecherausrichtung von  $\varphi = 0^\circ$ . Wie zu erwarten, bildet sich ein schmäler, ausgeprägter Peak bei einem Winkel von  $0^\circ$ . Zusätzlich bilden sich durch Reflexionen hervorgerufene gedämpfte Nebenspitzen aus.

**Abb. 4.5:** Darstellungen zur Veranschaulichung der Datenreduktion des modifizierten globalen Kohärenzfeldes und das Resultat der Mittlung des modifizierten GCF einer realen Audiodatei mit Sprachsignalen im halligen Bürraum.

nur die Daten auf dem Ring um die Hörerposition weiterverarbeitet. Das resultierende  $mGCF^{Head}$  ist pro Block nicht mehr von den zwei Dimensionen  $x$  und  $y$  abhängig, sondern nur noch vom Azimuth  $\theta$ . Abbildung 4.5(b) zeigt das  $mGCF^{Head}$  über dem Azimuth  $\theta$  für einen bei  $\theta_0 = 0^\circ$  positioniertem Sprecher und einer Sprecherausrichtung von  $\varphi_s = 0^\circ$ . Wie zu erwarten, bildet sich bei  $\theta = 0$  ein deutliches, schmales Maximum aus. Neben dem Maximum befinden sich Nebenmaxima, die durch die Reflexionen im Raum entstehen.

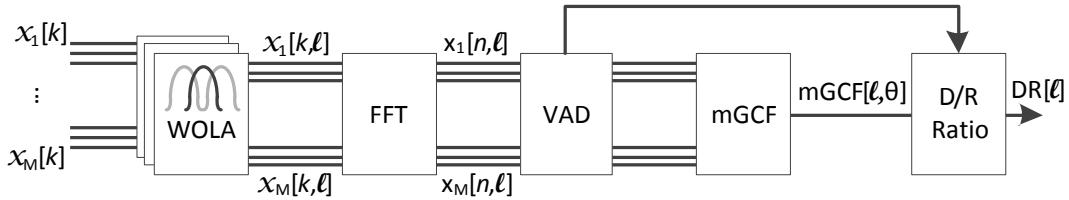
#### 4.1.2.3 Algorithmus zur Schätzung der Sprecherausrichtung unter Verwendung des modifizierten globalen Kohärenzfeldes

Unter Verwendung des  $mGCF^{Head}$  wurde der in Abbildung 4.6 als Blockschaltbild dargestellte Algorithmus entwickelt und in MATLAB implementiert. Der Algorithmus verwendet  $M$  diskrete Zeitsignale  $x_1[k]$  bis  $x_M[k]$  von binaural verschalteten Hörgeräten. In der vorliegenden Implementation wurden die gesamten sechs Mikrofonsignale der zwei Triano 3 Hörgerätedummies verwendet.

Die Zeitsignale der  $M$  Mikrofone werden in  $L$  überlappende und gefensterte Blöcke aufgeteilt. Die resultierenden Signale  $x_1[k, \ell]$  bis  $x_M[k, \ell]$  des Blockes  $\ell$  werden mittels FFT in die diskreten Kurzzeitspektren  $X_1[n, \ell]$  bis  $X_M[n, \ell]$  mit dem Frequenzindex  $n$  transformiert. Pro Block  $\ell$  bestimmt

## Kapitel 4. Analyse der algorithmischen Fähigkeit zur Schätzung der Sprecherausrichtung

---



**Abb. 4.6:** Blockschaltbild zur Veranschaulichung des Algorithmus zu Schätzung der Sprecherausrichtung mittels modifiziertem globalen Kohärenzfeld. Pro überlappenden Zeitblock wird das modifizierte GCF berechnet, die Sprecherposition bestimmt und ein Verhältnis zwischen dem Maximum des mGCF an der Sprecherposition und dem Mittelwert des mGCF der umliegenden Winkel bestimmt. Je höher dieses Verhältnis, um so größer ist der Direktschallanteil und um so wahrscheinlicher ist es, dass der Sprecher in Richtung des Hörers blickt.

ein VAD, ob im jeweiligen Block Sprache enthalten ist. In der entwickelten Matlab-Implementation des beschriebenen Algorithmus wurde der VAD nach [MK02] verwendet. Mit den  $M$  Kurzzeitsprektren wird anschließend  $mGCF^{Head}(\ell, \theta)$  berechnet. Aus dem mGCF wird abschließend für jeden sprachenthaltenden Block die Sprecherposition  $\hat{\theta}_0$  aus dem Argument des mGCF-Maximums und ein Maß für das Verhältnis aus Direkschallanteil und Reflexionen geschätzt. Eine Mögliche Umsetzung dieses Direct/Reverberant Maß ( $D/R Ratio, D$ ) ist das Verhältnis aus dem Maximum und dem Mittelwert aus den umliegenden Azimuthwinkel des mGCF. Angenommen wird dabei wie in Abschnitt 2.4 gezeigt, dass das Verhältnis zwischen Direktschall und den ersten frühen Reflexionen sinkt, wenn die Sprecherausrichtung steigt. Das Direct/Reverberant Maß  $D$  kann nach

$$D(\ell) = \frac{\max \{mGCF^{Head}(\ell, \theta)\}}{\mu \{mGCF^{Head}(\ell, \theta_r)\}} \quad (4.6)$$

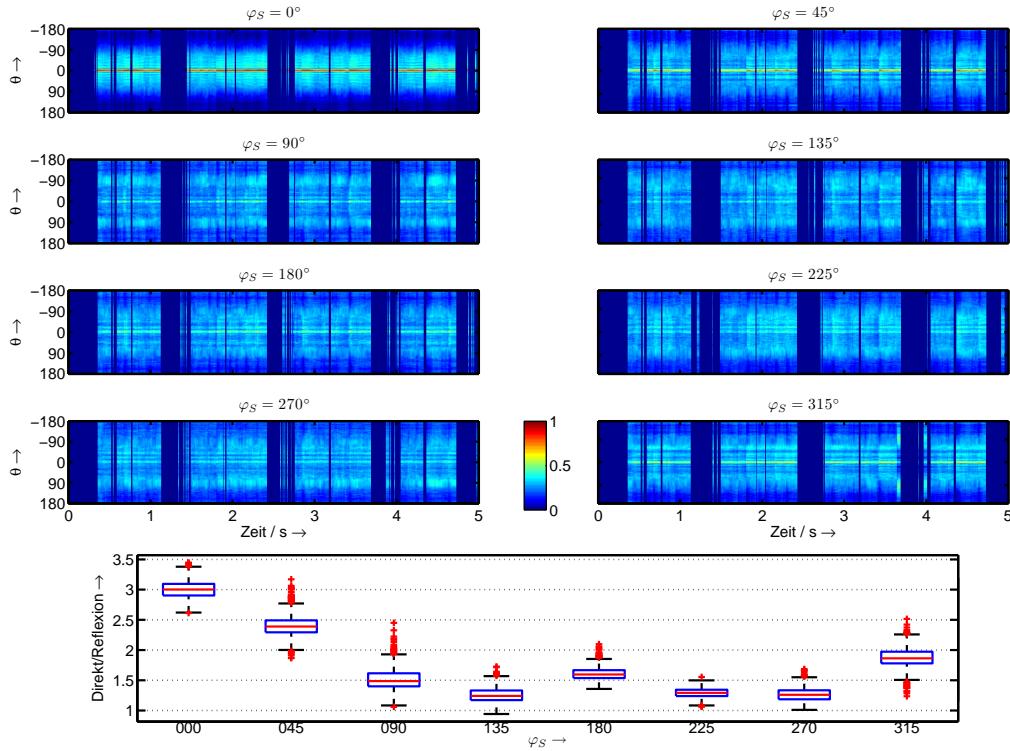
mit  $\theta_r \in [-pi, pi] \setminus \{\hat{\theta}_0\}$  und dem Block  $\ell$  bestimmt werden. Das Intervall  $\theta_r$  sollte nur Azimuths enthalten, in denen Reflexionen zu erwarten sind und kann ggf. weiter eingeschränkt werden. In Abbildung 4.5(b) ist ein solch eingeschränktes Intervall  $\theta_r$  in grün dargestellt.

Für eine stationäre Sprecherausrichtung ergibt sich eine Schätzung des mittleren Direct/Reverberant Maß aktiver Sprache als arithmetisches Mittel aus

$$\bar{D} = \mu(D[l_s]) \quad (4.7)$$

oder als Quantil

$$\tilde{D} = Q_q(D[l_s]) \quad (4.8)$$



**Abb. 4.7:** Beispiel für die Berechnung des modifizierten globalen Kohärenzfeldes für unterschiedliche Sprecherausrichtungen eines männlicher Sprechers. Die oben Abbildungen zeigen das resultierende mGCF ind Abhängigkeit von der Sprecherausrichtung über der Zeit. Der Boxplot zeigt die Verteilung des Direct/Reverberant Maßes  $D$  in Abhängigkeit von der relativen Sprecherausrichtungen.

mit dem Direct/Reverberant Maß  $D[l_s]$  der sprachenthaltenden Blöcke  $l_s$ , dem Mittelwert  $\mu(\cdot)$  und dem q-Quantil  $Q_q(\cdot)$ . Bei nicht stationären Sprecherausrichtungen bieten sich gleitende Mittelwerte an. Ähnlich wie in Abschnitt 4.1.1 wurden Quantile aufgrund der Robustheit verwendet und bei der durchgeföhrten Monte-Carlo-Simulation verglichen. Es hat sich gezeigt, dass sich das 0.5-Quantil am besten eignet.

In Abbildung 4.7 ist das vom Azimuth abhängige Direct/Reverberant Maß über der Zeit für ein Sprachsignal aus einem Lautsprecher im Bürraum bei frontaler Quelle im Abstand von einem Meter zum Kunstkopf für unterschiedliche Quellenausrichtungen dargestellt. Während sich bei einer Quellenausrichtung von  $\varphi_S = 0^\circ$  ein deutliches Maximum für  $\theta = 0^\circ$  abzeichnet, nimmt dieses Maximum bis  $\theta = 180^\circ$  immer mehr ab. In der unteren Abbildung ist ein Boxplot des Direct/Reverberant Maß über die Sprecherausrichtungen abgebildet. Es zeigt sich deutlich, dass das verwendete Maß wie erwartet einen Rückschluss auf die Sprecherausrichtung ermöglicht. Das lokale Maximum bei  $\theta = 180^\circ$  ergibt sich aus den konstruktiven Interferenzen an der Rückseite des Lautsprechers (vgl. Abschnitt 2.3).

### **4.1.3 Algorithmus zur Schätzung der Sprecherausrichtung unter Verwendung der linearen Prädiktion**

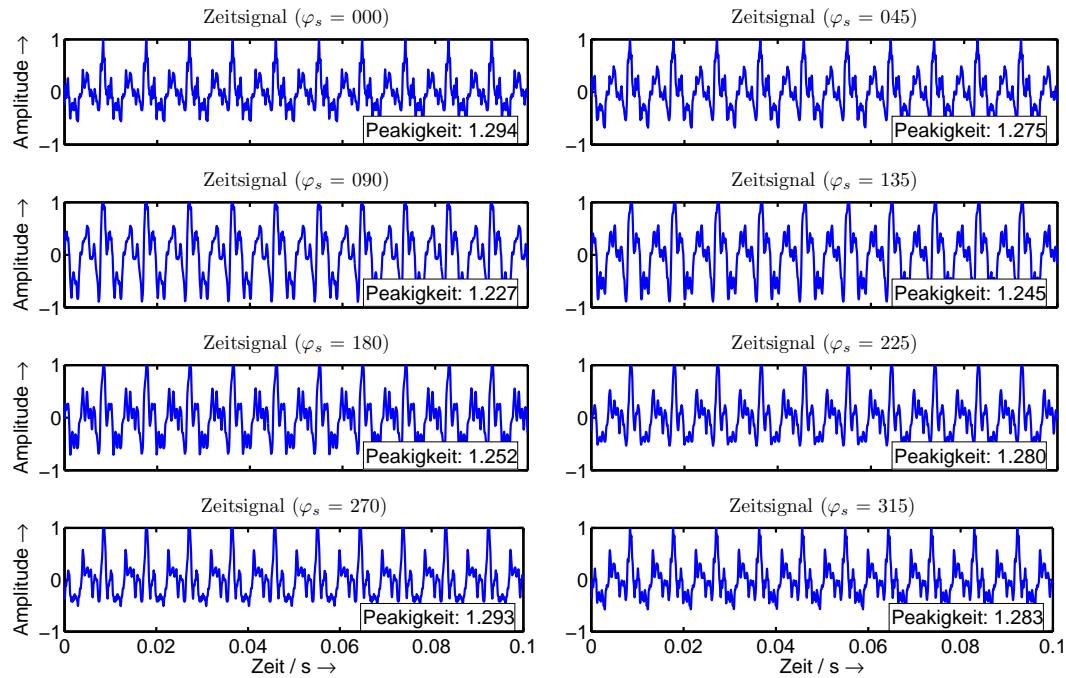
In Abschnitt 1.4 und in [GMvdP<sup>+</sup>09] wurde gezeigt, dass mit Hilfe der linearen Prädiktion ein Mikrofonsignal von der Übertragungsfunktion des Vokaltraktes befreit werden kann. Das resultierende Residual-Signal  $e(k)$  ergibt sich dann annähernd aus dem mit der Raumimpulsantwort gefalteten Glottissignal. Daraus folgt, dass im Residual-Signal bei stimmhafter Anregung die Glottisimpulse als deutliche Spitzen enthalten und die frühen Reflexionen als geringere Spitzen zwischen den Glottisimpulsen erkennbar sind (vgl. [GMvdP<sup>+</sup>09]).

Abbildung 4.8 zeigt beispielhaft aufgenommene und normierte Zeitsignale eines Gehörgangsmikrofons des Neumann Kunstkopfes in Abhängigkeit von der Quellenausrichtung. Als Quellsignal wurde ein periodisch wiederholter Ausschnitt des Vokals /a/ mit der Lautsprecherbox 6301B für die Ausrichtungen von  $0^\circ$  bis  $315^\circ$  in  $45^\circ$  Schritten aufgenommen. Die Aufnahmen erfolgten mit einem Abstand von einem Meter zwischen Quelle und Empfänger im Bürraum, welcher in Abbildung 3.3 dargestellt ist.

In den verarbeiteten Mikrofonsignalen sind aufgrund der Verschmierung des Quellsignals mit der Raumimpulsantwort zwar Unterschiede in den Signalen zu erkennen, eine deutliche Unterscheidung der Wellenform in Abhängigkeit von der Sprecherausrichtung ist augenscheinlich aber nicht möglich.

Im Gegensatz dazu sind in Abbildung 4.9 die LPC-Residual-Signale der entsprechenden Mikrofonsignale abgebildet. Die Signale wurden mit einer Samplingfrequenz von  $f_s = 16$  kHz abgetastet und es wurden 20 LP-Koeffizienten verwendet. In Abschnitt 2.4 wurde für einen halligen Raum beschrieben, dass bei steigender Sprecherausrichtung der Direktschallanteil im Verhältnis zu den ersten frühen Reflexionen sinkt. Dementsprechend müssten bei einer Sprecherausrichtung von  $0^\circ$  der Direktschall der Glottisimpulse im LPC-Residual deutlich als einzelne, periodische Impulse hervortreten. Bei steigender Sprecherausrichtung reduzieren sich die Amplituden der Glottisimpulse und die Anteile der frühen Reflexionen steigen an. Beide Effekte sind in Abbildung 4.9 ersichtlich. Bei den Sprecherausrichtungen von  $315^\circ$ ,  $0^\circ$  und  $45^\circ$  treten die Glottisimpulse noch deutlich hervor, während bei den restlichen Winkeln diese schwerer aus den Anteilen der Reflexionen erkennbar sind.

Der Einfluss der Sprecherausrichtung auf das LPC-Residual-Signal soll mit Hilfe des im Blockschaltbild der Abbildung 4.10 dargestellten Si-



**Abb. 4.8:** Zeitsignal des Vokals /a/ eines männlichen Sprechers in Abhängigkeit von der Sprecherausrichtung  $\varphi_s$ . Obwohl in den Signalen Unterschiede ersichtlich sind, ist es augenscheinlich nicht möglich, aus dem Zeitsignal auf eine Sprecherausrichtung zu schließen.

gnalverarbeitungsalgorithmus eine Schätzung der Sprecherausrichtung ermöglichen. Dazu werden zwei binaurale Eingangssignale  $x_l$  und  $x_r$  in zeitlich überlappende und gefensterte Blöcke aufgeteilt, in der vorliegenden Implementation von den vorderen Mikrofone der zwei Triano 3 Hörgerätedummies. Pro Block  $\ell$  bestimmt ein VAD ob der jeweilige Block Sprache enthält. In der entwickelten MATLAB-Implementation des beschriebenen Algorithmus wurde der VAD nach [MK02] verwendet. Anschließend wird für jeden Kanal das LP-Residual-Signal unter Verwendung der linearen Prädiktion (siehe Abschnitt 1.4) bestimmt. Für die Signalblöcke mit Sprachanteil muss abschließend bestimmt werden, wie sich die Glottisimpulse und die frühen Reflexionen zueinander verhalten.

Möglich wird dies unter Verwendung eines Maßes für die Peakigkeit des jeweiligen Signalblocks. Die drei folgenden Maße wurden gewählt und evaluiert um eine Aussage über das Verhältnis stark ausgeprägter Spitzen im Signal zum Restsignal zu erhalten.

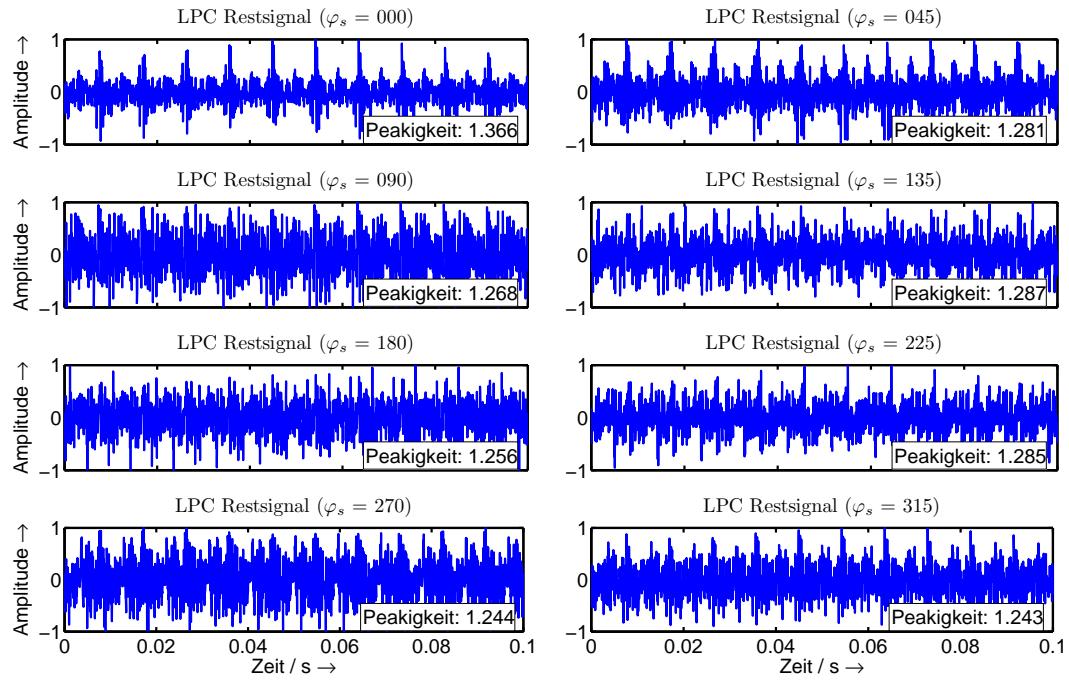
- **Kurtosis:**

$$P^K[\ell] = \text{Kurtosis}[\ell] = \frac{\mu_4(x[k, \ell])}{\sigma^4(x[k, \ell])} \quad (4.9)$$

„Die Kurtosis ist definiert als das Verhältnis zwischen dem vierten

## Kapitel 4. Analyse der algorithmischen Fähigkeit zur Schätzung der Sprecherausrichtung

---



**Abb. 4.9:** LPC-Residual-Signal des Vokals /a/ eines männlichen Sprecher in Abhängigkeit von der Sprecherausrichtung  $\varphi_s$ . Für die Sprecherausrichtung  $\varphi = 315^\circ$  bis  $45^\circ$  sind die Glottisimpulse deutlich zu erkennen, während für die restlichen Signale die Anteile der Reflexionen stärker hervortreten.

zentralen Moment und dem Quadrat der Varianz einer Zufallsgröße.“ [Bit09]

- **Crest-Faktor:**

$$P^C[\ell] = \text{Crest}[\ell] = \frac{\max\{|x[k, \ell]|}\right\}}{\sqrt{\frac{1}{N} \sum |x[k, \ell]|^2}} \quad (4.10)$$

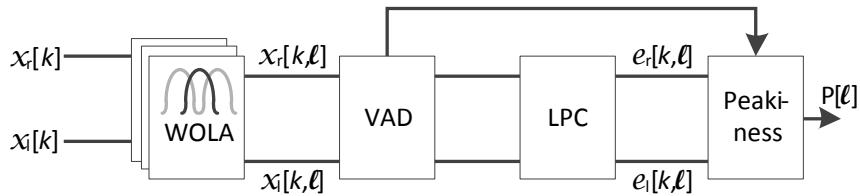
Der Crest-Faktor beschreibt das Verhältnis von Spitzenwert zu Effektivwert einer Wechselgröße. Er gibt einen Hinweis auf die Modulation der Amplitude der Wechselgröße (vgl. [Kra11]).

- **Peakigkeit:**

$$P^P[\ell] = \text{Peakigkeit}[\ell] = \frac{\sqrt{\frac{1}{N} \sum |x[k, \ell]|^2}}{\frac{1}{N} \sum |x[k, \ell]|} \quad (4.11)$$

„Die Peakigkeit ist durch das Verhältnis des Effektivwerts der Leistung zum Durchschnittswert des vollwellengleichgerichteten „ [Ben08] Signals definiert.“

Für eine stationäre Sprecherausrichtung ergibt sich eine Schätzung der ge-



**Abb. 4.10:** Blockschaltbild zur Veranschaulichung des Algorithmus zur Schätzung der Sprecherausrichtung mittels linearer Prädiktion. Pro überlappenden Zeitblock wird das LPC-Residual-Signal berechnet und ein Maß für die Peakigkeit berechnet. Die Annahme ist, dass im Residual-Signal die Peaks der frühen Reflexionen bei frontaler Sprecherausrichtung geringer ausgeprägt sind, als bei einem weg gedrehten Sprecher.

mittelten Peakigkeit bei aktiver Sprache als arithmetisches Mittel aus

$$\bar{P}_{r,l} = \mu(P_{r,l}[l_s]) \quad (4.12)$$

oder als Quantil

$$\tilde{P}_{r,l} = Q_q(P_{r,l}[l_s]) \quad (4.13)$$

mit der Peakigkeit  $P[l_s]$  der sprachenthaltenden Blöcke  $l_s$ , dem Mittelwert  $\mu(\cdot)$  und dem q-Quantil  $Q_q(\cdot)$ . Bei nicht stationären Sprecherausrichtungen bieten sich gleitende Mittelwerte an. Auch hier wurden Quantile aufgrund der Robustheit als Mittel verwendet und bei der Parameteroptimierung mittels Monte-Carlo-Simulation untersucht. Es hat sich gezeigt, dass sich das 0.7-Quantil am besten eignet. Zusätzlich wurden unterschiedliche Blöcklängen untersucht. Hier führten Blöcke in der Größenordnung von 20 ms zu den besten Ergebnissen.

Entsprechend der Position des Sprechers muss abschließend entschieden werden, von welcher Kopfseite das Maß zur Bewertung der Sprecherausrichtung herangezogen wird. Sinnvoll ist daher der Einsatz einer zusätzlichen Sprecherpositionsbestimmung wie in Abschnitt 1.5.1 vorgestellt.

## 4.2 Versuchsdurchführung: Schätzung der Sprecherausrichtung mit Signalverarbeitungsalgorithmen

Um eine erste qualitative Aussage über die Fähigkeit der vorgestellten Algorithmen zur Schätzung der Sprecherausrichtung zu erhalten, wurden die

in Abschnitt 3.2 vorgestellten Probandenhörversuche mit den Algorithmen simuliert. Die Algorithmen mussten die gleichen Aufgaben erfüllen, wie die Probanden.

Bisher existieren keine Schwell- oder Grenzwerte und Entscheidungskriterien zur Unterscheidung der Sprecherausrichtungen. Aufgrund variierender Raumeigenschaften ist davon auszugehen, dass auch diese Entscheidungskriterien raumabhängig sind. Verfahren zu Schätzung dieser Entscheidungskriterien müssen noch entwickelt werden und sind nicht Bestandteil dieser Arbeit.

### 4.2.1 Bewertung der berechneten Maße für die Ja/Nein Experimente

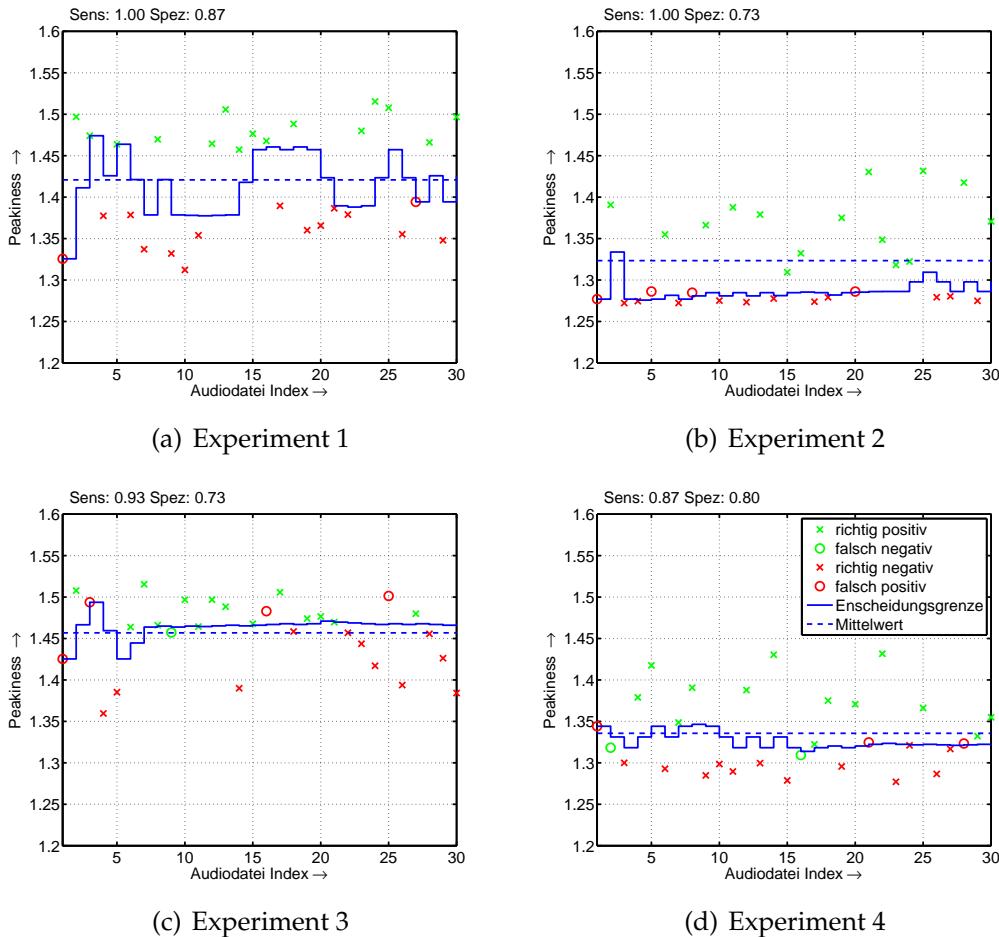
Um den Missstand eines fehlenden Entscheidungskriteriums auszugleichen, wurde für die Ja/Nein Experimente ein adaptives Verfahren zur Schätzung der benötigten Schwellwerte gewählt. Motiviert ist es aus der Annahme, dass die Probanden aufgrund der geringen Erfahrungswerte hinsichtlich der Bewertung einer Sprecherausrichtung zu Beginn der Experimente selbst keine Entscheidungskriterien besitzen. Vielmehr ist zu erwarten, dass die Probanden pro Audiosample nach verwertbaren akustischen Eigenschaften suchen. In die aktuelle Bewertung fließen dann die Erfahrungen aus den vorangegangenen Audiosamples ein, ihr Entscheidungskriterium passt sich adaptiv an.

Eine entsprechende Umsetzung dieser Annahme für die Bewertung der Algorithmen ermittelt das jeweilige Entscheidungskriterium  $E$  für die mittlere Steigung  $m$ , der Peakigkeit  $P^P$  und dem Direkt/Reflexionen Maß  $D$  aus den berechneten Maßen der vorangegangenen Audiodateien. Die Entscheidung  $E$ , ob die relative Sprecherposition  $\varphi_s$  der aktuellen Audiodatei  $N$   $0^\circ$  beträgt, wird bestimmt nach

$$E = \begin{cases} \text{wahr} & \text{falls } M(N) \geq Q_{.5}\{M[n]\} \text{ mit } n = 1 \dots N \\ \text{falsch} & \text{sonst} \end{cases} \quad (4.14)$$

mit dem Median  $Q_{.5}$ , welcher aufgrund seiner Robustheit gegenüber Ausreißern gewählt wurde. Entsprechend des verwendeten Algorithmus muss die Größe  $M$  durch die mittlere Steigung  $m$ , der Peakigkeit  $P^P$  oder dem Direkt/Reflexionen Maß  $D$  ersetzt werden.

In den Abbildungen 4.11(a) bis 4.11(d) wird das Vorgehen für die durchgeföhrten Ja/Nein Experimente am Beispiel des LPC-Ansatzes unter Verwendung der Peakigkeit  $P^P$  dargestellt. In Abhängigkeit von den voran-



**Abb. 4.11:** Ergebnisse Experiment 1: Blinde Schätzung der Kopfausrichtung im reflexionsarmen Raum ( $0^\circ$  vs.  $180^\circ$ )

gegangenen Audiodateien, wird die Entscheidungsgrenze angepasst und die daraus resultierenden richtig positiven/negativen und falsch positiven/-negativen Bewertungen dargestellt. Es zeigt sich bei allen Experimenten, dass das gewählte Verfahren für viele Signale in der Lage ist, die Sprecherausrichtung von  $\varphi = 0^\circ$  von der abweichenden Sprecherausrichtung zu unterscheiden. Allerdings führt das Verfahren zu einem Bias, da für die ersten Audiodateien keine ausreichenden Informationen für eine geeignete Schwelle vorliegen.

#### 4.2.2 Bewertung der berechneten Maße für die Zuordnungsexperimente

Die Bewertung der Audiodateien für die Zuordnungsexperimente wurde motiviert durch die Annahme, dass die Probanden in den zur Verfügung gestellten Audiodateien entsprechende akustische Merkmale suchen und dann z.B. aufsteigend den jeweiligen Sprecherausrichtungen zuweisen. Dem

entsprechend wurden bei der Bewertung durch die Audiodateien jeweils die mittlere Steigung  $m$ , die Peakigkeit  $P^P$  und das Direkt/Reflexionen Maß  $D$  für alle zu bewertenden Audiodateien in einem Testdurchlauf berechnet und absteigend sortiert. Hierbei wird angenommen, dass die entsprechenden Maße für eine Sprecherausrichtung  $\varphi_S = 0^\circ$  am höchsten ausgeprägt sind und mit steigender Sprecherausrichtung bis  $\varphi = 180^\circ$  abnehmen. Durch dieses Verfahren ist es nicht möglich, eine Sprecherausrichtung nach rechts oder links zu unterscheiden.

## **4.3 Versuchsergebnisse: Schätzung der Sprecherausrichtung mit Signalverarbeitungsalgorithmen**

Die Auswertung der Algorithmenevaluation erfolgt unter den gleichen Gesichtspunkten wie bei den Probandenhörversuchen im Abschnitt 3.3. Im Folgenden werden die Ergebnisse zusammengefasst vorgestellt und abschließend interpretiert. Bei einer durchgeföhrten Monte-Carlo-Simulation hat sich gezeigt, dass für den LPC-Ansatz das Peakigkeitsmaß nach Gleichung 4.11 die besten Ergebnisse zu erwarten ist. Aus Gründen der Übersichtlichkeit werden daher für diesen Algorithmus nur die Ergebnisse unter Verwendung dieses Maßes vorgestellt. Die Ergebnisse unter Verwendung der Kurtosis (Gleichung 4.9) und des Crest-Faktors (Gleichung 4.10) können den Anhängen A.5 und A.6 entnommen werden.

### **4.3.1 Ja/Nein Experimente**

In den folgenden Unterpunkten werden die wesentlichen Ergebnisse für die durchgeföhrten Ja/Nein Experimente mit den entwickelten Algorithmen beschrieben.

#### **4.3.1.1 Experiment 1: Unterscheidung der Kopfausrichtung im reflexionsarmen Raum ( $0^\circ$ vs. $180^\circ$ )**

Die Sensitivität, Spezifität und Korrektklassifikationsrate vom ersten Experiment der Algorithmen-Evaluation sind in Abhängigkeit von den Quellsignalen und der Sprecherposition in den Tabellen 4.1(a) bis 4.1(c) dargestellt. Eine Unterscheidung der Sprecherausrichtungen  $0^\circ$  und  $180^\circ$  ist mit den vorgestellten Algorithmen im Reflexionsarmenraum mit dem LPC-Ansatz

## Kapitel 4. Analyse der algorithmischen Fähigkeit zur Schätzung der Sprecherausrichtung

---

**Tab. 4.1:** Ergebnisse Experiment 1: Unterscheidung der Sprecherausrichtung im reflexionsarmen Raum von Signalverarbeitungsalgorithmen. Tabellen zeigen die Sensitivität, die Spezifität und die Korrektklassifikationsrate für Unterscheidung von den Sprecherausrichtungen von  $0^\circ$  und  $180^\circ$  für unterschiedliche Quellsignale.

(a) Mittlere Steigung des Frequenzgangs

Signal	SENS	SPEZ	KORR
SA000 1m männlich	0.500	0.500	0.500
SA045 1m männlich	0.400	0.486	0.475
SA090 1m männlich	0.000	0.500	0.500
SA135 1m männlich	0.000	0.487	0.475
SA180 1m männlich	1.000	0.526	0.550
SA000 1m weiblich	0.556	1.000	0.600
SA045 1m weiblich	0.500	0.000	0.500
SA090 1m weiblich	0.952	1.000	0.975
SA135 1m weiblich	0.513	1.000	0.525
SA180 1m weiblich	0.870	1.000	0.925
SA000 1m OLSA	0.800	1.000	0.875
SA045 1m OLSA	0.526	1.000	0.550
SA090 1m OLSA	0.655	0.909	0.725
SA135 1m OLSA	0.625	1.000	0.700
SA180 1m OLSA	0.792	0.938	0.850
Gesamt $\mu$	0.579	0.756	0.648
$\sigma$	0.296	0.313	0.179

(b) Modifiziertes globales Kohärenzfeld

Signal	SENS	SPEZ	KORR
SA000 1m männlich	0.826	0.941	0.875
SA045 1m männlich	0.882	0.783	0.825
SA090 1m männlich	0.952	1.000	0.975
SA135 1m männlich	0.952	1.000	0.975
SA180 1m männlich	1.000	0.952	0.975
SA000 1m weiblich	0.870	1.000	0.925
SA045 1m weiblich	1.000	0.833	0.900
SA090 1m weiblich	0.000	0.500	0.500
SA135 1m weiblich	0.864	0.944	0.900
SA180 1m weiblich	0.870	1.000	0.925
SA000 1m OLSA	0.571	1.000	0.625
SA045 1m OLSA	0.905	0.947	0.925
SA090 1m OLSA	0.000	0.474	0.450
SA135 1m OLSA	0.645	1.000	0.725
SA180 1m OLSA	0.500	0.000	0.500
Gesamt $\mu$	0.722	0.825	0.800
$\sigma$	0.329	0.287	0.190

(c) Linearen Prädiktion ( $P^P$ )

Signal	SENS	SPEZ	KORR
SA000 1m männlich	0.909	1.000	0.950
SA045 1m männlich	0.655	0.909	0.725
SA090 1m männlich	0.905	0.947	0.925
SA135 1m männlich	0.952	1.000	0.975
SA180 1m männlich	1.000	0.833	0.900
SA000 1m weiblich	0.952	1.000	0.975
SA045 1m weiblich	0.952	1.000	0.975
SA090 1m weiblich	1.000	0.870	0.925
SA135 1m weiblich	0.895	0.857	0.875
SA180 1m weiblich	0.923	0.704	0.775
SA000 1m OLSA	0.833	1.000	0.900
SA045 1m OLSA	0.741	1.000	0.825
SA090 1m OLSA	0.870	1.000	0.925
SA135 1m OLSA	0.769	1.000	0.850
SA180 1m OLSA	0.900	0.900	0.900
Gesamt $\mu$	0.884	0.935	0.893
$\sigma$	0.097	0.089	0.073

am erfolgreichsten. Hier wird eine mittlere Sensitivität von 84 % und eine Spezifität von 93 % (Tabelle 4.1(c)) erreicht. Insgesamt wurden 89 % der Audiodateien korrekt klassifiziert. Die Standardabweichungen für die Sensitivität, Spezifität und Korrektklassifikationsrate betragen weniger als 10 %.

Der mGCF-Ansatz erreicht eine mittlere Sensitivität von 72 % und eine Spezifität von 76 % (Tabelle 4.1(b)). Insgesamt wurden 80 % der Audiodateien korrekt klassifiziert. Die Standardabweichungen für die Sensitivität und Spezifität betragen  $\approx 30\%$  und für die Korrektklassifikationsrate 19 %. Viele Quellsignale können mit über 90 % korrekt klassifiziert werden. Bei einigen Signalen entsprechen die Ergebnisse der Ratewahrscheinlichkeit. Es ist dabei keine Abhängigkeit von den verwendeten Quellen erkennbar.

Das schlechteste Ergebnis erreicht die Schätzung mittels Steigung des Fre-

quenzgangs mit einer mittleren Sensitivität von 58 %, knapp oberhalb der Ratewahrscheinlichkeit, und einer Spezifität von 76 % (Tabelle 4.1(a)). Insgesamt wurden 65 % der Audiodateien korrekt klassifiziert. Zudem betragen die Standardabweichungen für die Sensitivität und Spezifität  $\approx 30\%$  und für die Korrektklassifikationsrate 18 %. Die signalabhängigen Ergebnisse variieren stark. So werden einige Signale nahezu hunderprozentig korrekt klassifiziert, bei anderen ist der Algorithmus nur in der Lage zu raten. Dabei ist keine Abhängigkeit von den verwendeten Quellen erkennbar.

### 4.3.1.2 Experiment 2: Unterscheidung der Kopfausrichtung im Büro ( $0^\circ$ vs. $180^\circ$ )

Die Sensitivität, Spezifität und Korrektklassifikationsrate vom zweiten Experiment der Algorithmen-Evaluation sind in Abhängigkeit von den Quellsignalen und der Sprecherposition in den Tabellen 4.2(a) bis 4.2(c) dargestellt.

Im halligen Büroraum ist die Unterscheidung der Sprecherausrichtungen  $0^\circ$  und  $180^\circ$  ebenfalls am erfolgreichsten mit dem LPC-Ansatz. Der Algorithmus erreicht im Mittel eine Sensitivität von 89 %, eine Spezifität von 92 % und eine Korrektklassifikationsrate von 89 % (Tabelle 4.2(c)). Die Standardabweichung liegen unterhalb von 10 %. Vergleichbare mittlere Ergebnisse erreicht der mGCF-Ansatz mit etwas größeren Standardabweichungen von  $\approx 14\%$  (Tabelle 4.2(b)).

Erneut erzielt die Unterscheidung mittels mittlerer Steigung des Frequenzgangs die schletesten Resultaten. Der Algorithmus erreicht eine mittlere Sensitivität und Spezifität von  $\approx 55\%$ , knapp oberhalb der Ratewahrscheinlichkeit (Tabelle 4.2(a)). Insgesamt wurden 61 % der Audiodateien korrekt klassifiziert. Zudem betragen die Standardabweichungen für die Sensitivität 22 %, die Spezifität 40 % und die Korrektklassifikationsrate 16 %. Die signalabhängigen Ergebnisse variieren stark. Die Signale aus der Lautsprecherbox können verhältnismäßig gut klassifiziert werden, bei natürlichen Sprechern ist der Algorithmus nur in der Lage zu raten oder schlägt komplett fehl.

### 4.3.1.3 Experiment 3: Unterscheidung der Kopfausrichtung im reflexionsarmen Raum ( $0^\circ$ vs. $90^\circ$ )

Die Sensitivität, Spezifität und Korrektklassifikationsrate vom dritten Experiment der Algorithmen-Evaluation sind in Abhängigkeit von den Quellsignalen und der Sprecherposition in den Tabellen 4.3(a) bis 4.3(c) dargestellt. Auch bei der Unterscheidung der Sprecherausrichtungen  $0^\circ$  und  $90^\circ$  im

## Kapitel 4. Analyse der algorithmischen Fähigkeit zur Schätzung der Sprecherausrichtung

---

**Tab. 4.2:** Ergebnisse Experiment 2: Unterscheidung der Sprecherausrichtung im Büro von Signalverarbeitungsalgorithmen. Tabellen zeigen die Sensitivität, die Spezifität und die Korrektklassifikationsrate für Unterscheidung von den Sprecherausrichtungen von  $0^\circ$  und  $180^\circ$  für unterschiedliche Quellsignale.

(a) Mittlere Steigung des Frequenzgangs

Signal	SENS	SPEZ	KORR
SA000 1m männlich	0.889	0.613	0.675
SA045 1m männlich	1.000	0.541	0.575
SA090 1m männlich	0.667	0.514	0.525
SA135 1m männlich	0.900	0.633	0.700
SA180 1m männlich	1.000	0.513	0.525
SA000 1m weiblich	0.500	0.000	0.500
SA045 1m weiblich	0.500	0.000	0.500
SA090 1m weiblich	0.500	0.000	0.500
SA135 1m weiblich	0.500	0.000	0.500
SA180 1m weiblich	0.500	0.000	0.500
SA000 1m OLSA	0.895	0.857	0.875
SA045 1m OLSA	0.690	1.000	0.775
SA090 1m OLSA	0.708	0.813	0.750
SA135 1m OLSA	0.870	1.000	0.925
SA180 1m OLSA	0.800	1.000	0.875
Gesamt $\mu$	0.728	0.499	0.647
$\sigma$	0.193	0.402	0.159

(b) Modifiziertes globales Kohärenzfeld

Signal	SENS	SPEZ	KORR
SA000 1m männlich	1.000	1.000	1.000
SA045 1m männlich	1.000	0.833	0.900
SA090 1m männlich	0.909	1.000	0.950
SA135 1m männlich	0.952	1.000	0.975
SA180 1m männlich	0.870	1.000	0.925
SA000 1m weiblich	0.952	1.000	0.975
SA045 1m weiblich	0.947	0.905	0.925
SA090 1m weiblich	0.875	0.594	0.650
SA135 1m weiblich	1.000	1.000	1.000
SA180 1m weiblich	1.000	1.000	1.000
SA000 1m OLSA	0.571	1.000	0.625
SA045 1m OLSA	0.952	1.000	0.975
SA090 1m OLSA	0.682	0.722	0.700
SA135 1m OLSA	0.833	1.000	0.900
SA180 1m OLSA	0.625	1.000	0.700
Gesamt $\mu$	0.878	0.937	0.880
$\sigma$	0.142	0.125	0.137

(c) Linearen Prädiktion ( $P^P$ )

Signal	SENS	SPEZ	KORR
SA000 1m männlich	0.941	0.826	0.875
SA045 1m männlich	0.952	1.000	0.975
SA090 1m männlich	0.870	1.000	0.925
SA135 1m männlich	0.905	0.947	0.925
SA180 1m männlich	0.889	0.818	0.850
SA000 1m weiblich	0.882	0.783	0.825
SA045 1m weiblich	0.952	1.000	0.975
SA090 1m weiblich	0.905	0.947	0.925
SA135 1m weiblich	0.947	0.905	0.925
SA180 1m weiblich	1.000	0.741	0.825
SA000 1m OLSA	0.952	1.000	0.975
SA045 1m OLSA	0.714	1.000	0.800
SA090 1m OLSA	0.690	1.000	0.775
SA135 1m OLSA	0.909	1.000	0.950
SA180 1m OLSA	0.818	0.889	0.850
Gesamt $\mu$	0.888	0.924	0.892
$\sigma$	0.087	0.092	0.067

reflexionsarmen Raum ist der LPC-Ansatz mit einer mittleren Sensitivität von 83 %, einer Spezifität von 75 % und einer Korrektklassifikationsrate von 80 % am erfolgreichsten (Tabelle 4.3(c)). Die Standardabweichungen sind im Vergleich zum ersten Experiment mit  $\geq 20\%$  deutlich erhöht. Einige Einzelergebnisse entsprechen der Ratewahrscheinlichkeit.

Die Ergebnisse des mGCF-Ansatzes liegen für alle drei Bewertungsmaße bei  $\approx 70\%$  mit Standardabweichungen von über 20 % (Tabelle 4.3(b)). Erneut führt der Ansatz unter Verwendung der mittleren Steigung des Frequenzganges mit Bewertungsmaßen um 55 % bis 60 % zu den schlechtesten Ergebnissen und liegt im Mittel nur knapp über der Ratewahrscheinlichkeit (Tabelle 4.3(a)).

## Kapitel 4. Analyse der algorithmischen Fähigkeit zur Schätzung der Sprecherausrichtung

---

**Tab. 4.3:** Ergebnisse Experiment 3: Unterscheidung der Sprecherausrichtung im reflexionsarmen Raum von Signalverarbeitungsalgorithmen. Tabellen zeigen die Sensitivität, die Spezifität und die Korrektklassifikationsrate für Unterscheidung von den Sprecherausrichtungen von  $0^\circ$  und  $90^\circ$  für unterschiedliche Quellsignale.

(a) Mittlere Steigung des Frequenzgangs

Signal	SENS	SPEZ	KORR
SA000 1m männlich	0.600	0.514	0.525
SA045 1m männlich	0.500	0.500	0.500
SA090 1m männlich	0.000	0.500	0.500
SA135 1m männlich	0.400	0.486	0.475
SA180 1m männlich	0.333	0.486	0.475
SA000 1m weiblich	0.500	0.000	0.500
SA045 1m weiblich	0.500	0.000	0.500
SA090 1m weiblich	0.526	1.000	0.550
SA135 1m weiblich	0.500	0.000	0.500
SA180 1m weiblich	0.500	0.000	0.500
SA000 1m OLSA	0.800	1.000	0.875
SA045 1m OLSA	0.690	1.000	0.775
SA090 1m OLSA	0.833	0.773	0.800
SA135 1m OLSA	0.741	1.000	0.825
SA180 1m OLSA	0.857	0.895	0.875
Gesamt $\mu$	0.552	0.544	0.612
$\sigma$	0.220	0.397	0.163

(b) Modifiziertes globales Kohärenzfeld

Signal	SENS	SPEZ	KORR
SA000 1m männlich	0.818	0.889	0.850
SA045 1m männlich	0.333	0.486	0.475
SA090 1m männlich	0.909	1.000	0.950
SA135 1m männlich	0.952	1.000	0.975
SA180 1m männlich	0.909	1.000	0.950
SA000 1m weiblich	0.826	0.941	0.875
SA045 1m weiblich	1.000	0.606	0.675
SA090 1m weiblich	0.600	0.514	0.525
SA135 1m weiblich	0.500	0.000	0.500
SA180 1m weiblich	0.500	0.000	0.500
SA000 1m OLSA	0.588	1.000	0.650
SA045 1m OLSA	1.000	1.000	1.000
SA090 1m OLSA	0.000	0.487	0.475
SA135 1m OLSA	0.870	1.000	0.925
SA180 1m OLSA	0.588	1.000	0.650
Gesamt $\mu$	0.693	0.728	0.732
$\sigma$	0.284	0.361	0.206

(c) Linearen Prädiktion ( $P^P$ )

Signal	SENS	SPEZ	KORR
SA000 1m männlich	0.917	0.679	0.750
SA045 1m männlich	0.500	0.000	0.500
SA090 1m männlich	0.938	0.792	0.850
SA135 1m männlich	0.909	1.000	0.950
SA180 1m männlich	0.944	0.864	0.900
SA000 1m weiblich	0.513	1.000	0.525
SA045 1m weiblich	0.500	0.000	0.500
SA090 1m weiblich	0.750	0.563	0.600
SA135 1m weiblich	0.792	0.938	0.850
SA180 1m weiblich	1.000	0.556	0.600
SA000 1m OLSA	0.952	1.000	0.975
SA045 1m OLSA	1.000	1.000	1.000
SA090 1m OLSA	0.952	1.000	0.975
SA135 1m OLSA	0.952	1.000	0.975
SA180 1m OLSA	0.950	0.950	0.950
Gesamt $\mu$	0.838	0.756	0.793
$\sigma$	0.185	0.345	0.194

### 4.3.1.4 Experiment 4: Unterscheidung der Kopfausrichtung im Büro ( $0^\circ$ vs. $90^\circ$ )

Die Sensitivität, Spezifität und Korrektklassifikationsrate vom vierten Experiment der Algorithmen-Evaluation sind in Abhängigkeit von den Quellsignalen und der Sprecherposition in den Tabellen 4.4(a) bis 4.4(c) dargestellt.

Eine Unterscheidung der Sprecherausrichtung von  $0^\circ$  und  $90^\circ$  im halligen Büroraum gelingt dem LPC-Ansatz mit einer Sensitivität, einer Spezifität und einer Korrektklassifikationsrate von jeweils 80 % oder mehr (Tabelle 4.4(c)). Die Standardabweichung aller drei Bewertungsmaße beträgt  $\approx 19\%$ . In gleicher Größenordnung liegen die mittleren Ergebnisse und die Standardabweichungen des mGCF-Ansatzes, mit Ausnahme von der Spezifität

## Kapitel 4. Analyse der algorithmischen Fähigkeit zur Schätzung der Sprecherausrichtung

---

**Tab. 4.4:** Ergebnisse Experiment 4 und 12: Unterscheidung der Sprecherausrichtung im Büro von Signalverarbeitungsalgorithmen. Tabellen zeigen die Sensitivität, die Spezifität und die Korrektklassifikationsrate für Unterscheidung von den Sprecherausrichtungen von  $0^\circ$  und  $90^\circ$  für unterschiedliche Quellsignale.

(a) Mittlere Steigung des Frequenzgangs

Signal	SENS	SPEZ	KORR
SA000 1m männlich	0.500	0.500	0.500
SA045 1m männlich	0.111	0.387	0.325
SA090 1m männlich	0.000	0.459	0.425
SA135 1m männlich	0.800	0.600	0.650
SA180 1m männlich	1.000	0.513	0.525
SA000 1m weiblich	0.500	0.000	0.500
SA045 1m weiblich	0.487	0.000	0.475
SA090 1m weiblich	0.500	0.000	0.500
SA135 1m weiblich	0.541	1.000	0.575
SA180 1m weiblich	0.500	0.000	0.500
SA000 1m OLSA	1.000	0.870	0.925
SA045 1m OLSA	0.625	1.000	0.700
SA090 1m OLSA	0.786	0.654	0.700
SA135 1m OLSA	0.739	0.824	0.775
SA180 1m OLSA	0.895	0.857	0.875
Gesamt $\mu$	0.599	0.511	0.597
$\sigma$	0.287	0.370	0.170

(b) Modifiziertes globales Kohärenzfeld

Signal	SENS	SPEZ	KORR
SA000 1m männlich	0.487	0.000	0.475
SA045 1m männlich	0.667	0.529	0.550
SA090 1m männlich	0.559	0.833	0.600
SA135 1m männlich	0.909	1.000	0.950
SA180 1m männlich	0.541	1.000	0.575
SA000 1m weiblich	0.800	1.000	0.875
SA045 1m weiblich	0.923	0.704	0.775
SA090 1m weiblich	0.500	0.500	0.500
SA135 1m weiblich	0.952	1.000	0.975
SA180 1m weiblich	0.741	1.000	0.825
SA000 1m OLSA	0.952	1.000	0.975
SA045 1m OLSA	0.952	1.000	0.975
SA090 1m OLSA	0.900	0.633	0.700
SA135 1m OLSA	1.000	1.000	1.000
SA180 1m OLSA	0.952	1.000	0.975
Gesamt $\mu$	0.789	0.813	0.782
$\sigma$	0.190	0.293	0.197

(c) Linearen Prädiktion ( $P^P$ )

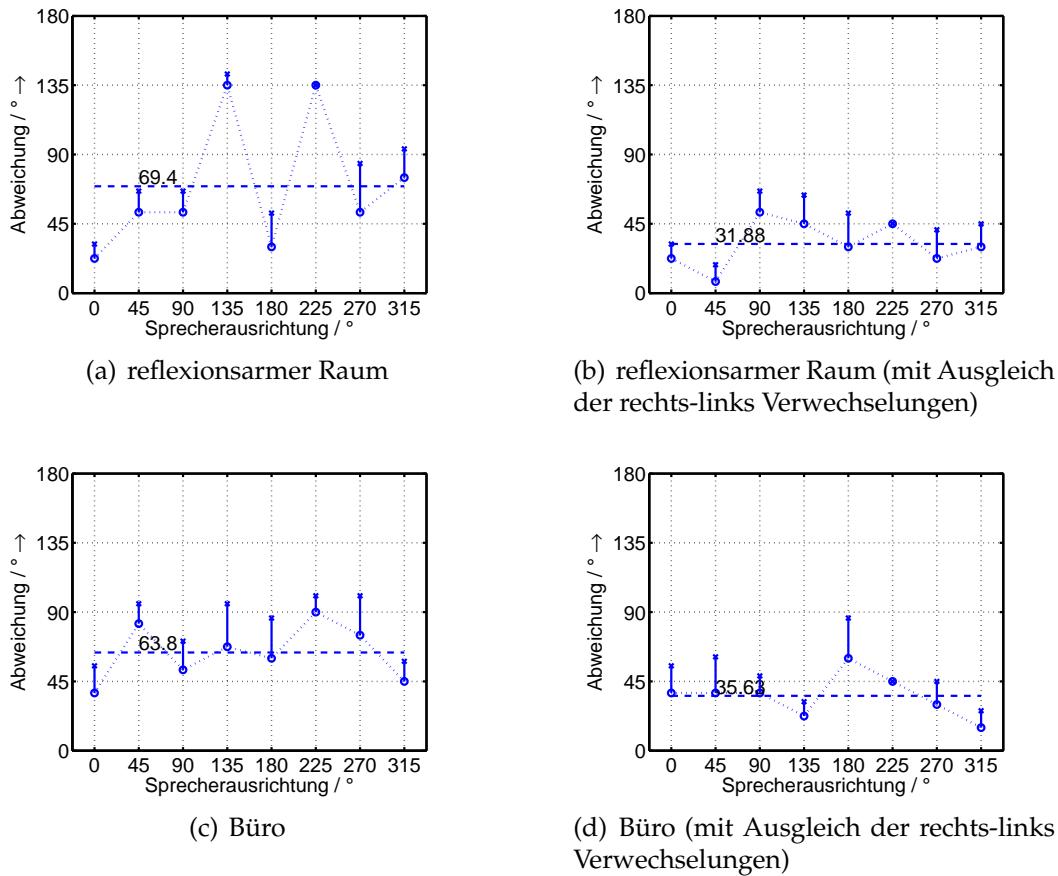
Signal	SENS	SPEZ	KORR
SA000 1m männlich	0.810	0.842	0.825
SA045 1m männlich	0.556	1.000	0.600
SA090 1m männlich	0.541	1.000	0.575
SA135 1m männlich	0.909	1.000	0.950
SA180 1m männlich	1.000	0.513	0.525
SA000 1m weiblich	0.750	0.563	0.600
SA045 1m weiblich	0.655	0.909	0.725
SA090 1m weiblich	0.833	1.000	0.900
SA135 1m weiblich	1.000	0.870	0.925
SA180 1m weiblich	0.500	0.500	0.500
SA000 1m OLSA	1.000	1.000	1.000
SA045 1m OLSA	0.909	1.000	0.950
SA090 1m OLSA	0.909	1.000	0.950
SA135 1m OLSA	1.000	1.000	1.000
SA180 1m OLSA	1.000	0.909	0.950
Gesamt $\mu$	0.825	0.874	0.798
$\sigma$	0.183	0.189	0.189

mit einer Standardabweichung von 29 % (Tabelle 4.4(b)).

Die Bewertung mittels der Steigung des Frequenzganges führt zu Ergebnissen, die mit 51 % bis 60 % im Mittel nur knapp über der Rate wahrscheinlichkeit liegen und Standardabweichungen von 17 % bis zu 37 % ausweisen (Tabelle 4.4(a)).

### 4.3.2 Zuordnungsexperimente

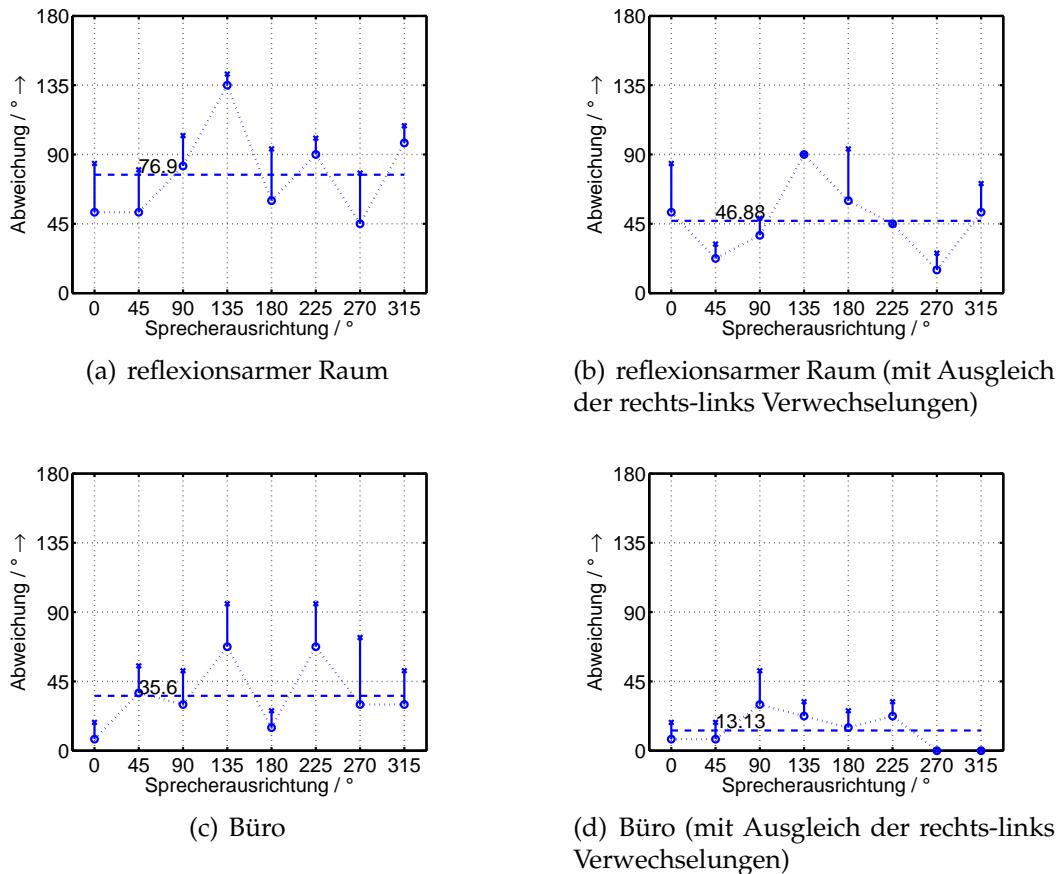
Nicht alle mit den Probanden durchgeführten Zuordnungsexperimente sind für die Evaluation der Algorithmen erforderlich. Die monotonen Experimente waren ausschließlich für eine Betrachtung des binauralen Gewinns für die Probanden entwickelt worden. Die Experimente mit den LPC-



**Abb. 4.12:** Ergebnisse Experiment 7: Dichotische Schätzung der Kopfausrichtung mit Sprachsignalen unter Verwendung des gemittelten Sprachspektrums. Grafiken zeigen die mittleren Abweichung  $a_m$  (•) und die Wurzel der mittleren quadratischen Abweichung  $a_q$  (x) und Durchschnitt der mittleren Abweichung der Probandenbewertung in Abhängigkeit von den wahren Sprecherausrichtungen.

Residual Signale sollten zeigen, ob in diesem Signal entsprechend der Annahme aus Abschnitt 1.4 nutzbare akustische Eigenschaften enthalten sind. Durchgeführt wurde daher lediglich das dichotische Experiment mit Sprachsignalen. Vorgestellt werden nur die Ergebnisse anhand der mittleren Abweichungen über den wahren Sprecherausrichtungen.

Die Abbildungen 4.12(a) und 4.12(c) zeigen die resultierenden mittleren Abweichungen zur Unterscheidung der realen Sprecherausrichtung unter Verwendung der mittleren Steigung des Frequenzganges für den reflexionsarmen Raum bzw. für den Bürraum. In beiden Räumen zeigt sich, dass der Algorithmus mit über  $63^\circ$  zu großen mittleren Abweichungen führt. Erst beim Ausgleich der auftretenden rechts-links Verwechslungen in den Abbildungen 4.12(b) und 4.12(d) sinken die mittleren Abweichungen auf  $32^\circ$  im reflexionsarmen Raum und  $35^\circ$  im Bürraum. Der Verlauf der Kurven ist nicht konstant und variiert zwischen den wahren Ausrichtungen. Der Algo-

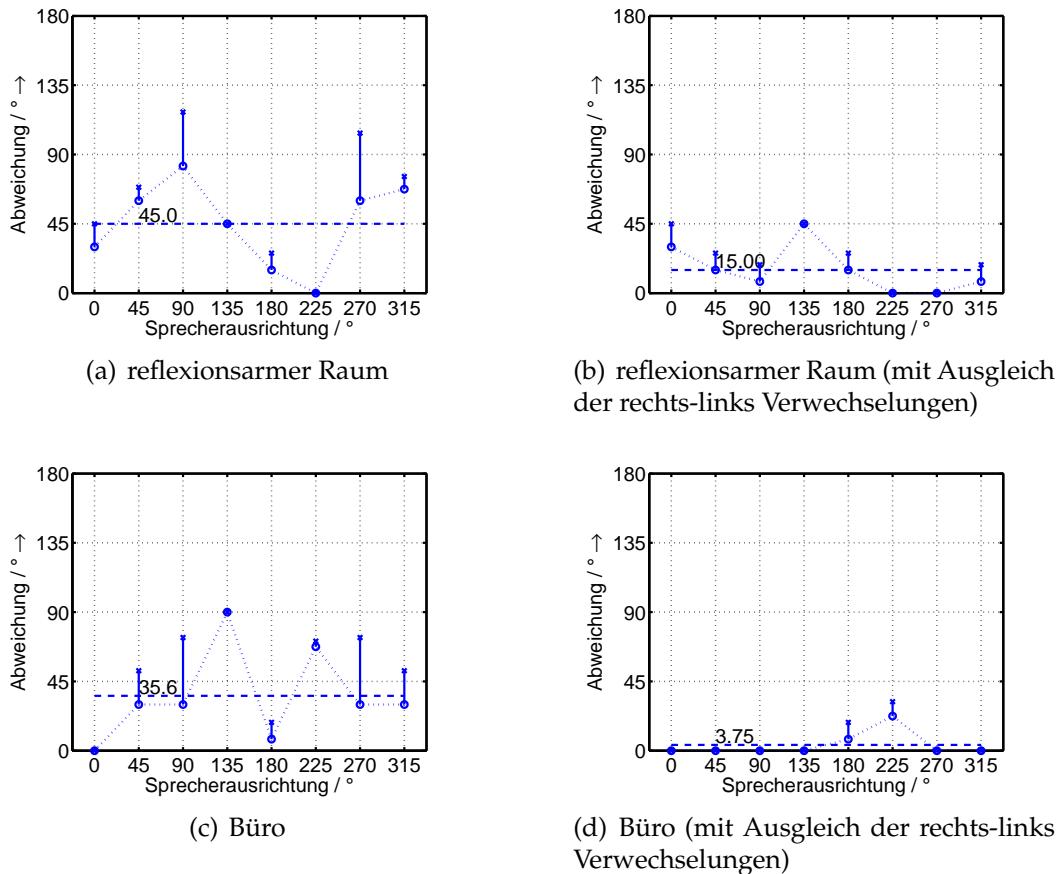


**Abb. 4.13:** Ergebnisse Experiment 7: Dichotische Schätzung der Kopfausrichtung mit Sprachsignalen unter Verwendung des modifizierten globalen Kohärenzfeldes. Grafiken zeigen die mittleren Abweichung  $a_m$  (●) und die Wurzel der mittleren quadratischen Abweichung  $a_q$  (x) und Durschschnitt der mittleren Abweichung der Probandenbewertung in Abhängigkeit von den wahren Sprecherausrichtungen.

rithmus zeigt damit deutlich schlechtere Ergebnisse als die normalhörenden Probanden, die in beiden Räumen Abweichungen von unter  $20^\circ$  erreichten.

Für den mGCF-Ansatz sind die Ergebnisse in den Abbildungen 4.13(a) und 4.13(c) entsprechend dargestellt. Diese Abbildungen zeigen ebenfalls hohe mittlere Abweichungen, sinken aber nach einem Ausgleich der rechts-links Verwechslungen. Während sich im reflexionsarmen Raum aufgrund der fehlenden Reflexionen eine Abweichung von  $47^\circ$  ergeben 4.13(b), betragen die Ergebnisse im Bürraum  $13^\circ$  (Abbildung 4.13(d)), schwanken aber in Abhängigkeit von der wahren Sprecherausrichtung. Der Algorithmus kann, wie erwartet, mit dem globalen Kohärenzfeld im halligen Bürraum ein Verhältnis aus Direktschall und Reflexionen bilden und für die Unterscheidung der Sprecherausrichtung nutzen.

Die Abbildungen 4.14(a) und 4.14(c) enthalten die mittleren Abweichungen für den LPC-Ansatz. Auch hier zeigen sich starke Schwankungen der Ergeb-



**Abb. 4.14:** Ergebnisse Experiment 7: Dichotische Schätzung der Kopfausrichtung mit Sprachsignalen unter Verwendung der linearen Prädiktion (Peakigkeit). Grafiken zeigen die mittleren Abweichung  $a_m$  (●) und die Wurzel der mittleren quadratischen Abweichung  $a_q$  (x) und Durchschnitt der mittleren Abweichung der Probandenbewertung in Abhängigkeit von den wahren Sprecherausrichtungen.

nisse in Abhängigkeit von der wahren Sprecherausrichtung. Erst durch den Ausgleich der rechts-links Verwechslungen sinken die Abweichungen auf  $15^\circ$  im reflexionsarmen Raum (Abbildung 4.14(b)) bzw. auf  $4^\circ$  im Bürroraum (Abbildung 4.14(d)). Der Algorithmus erreicht damit in beiden Räumen bessere Ergebnisse als die normalhörenden Probanden.

### 4.3.3 Erkenntnisse aus den Messergebnissen der durchgeführten Experimente zur algorithmischen Bewertung der Sprecherausrichtung

Die bei allen Algorithmen auftretenden rechts-links Verwechslungen wurden erwartet, da die entwickelten Algorithmen im derzeitigen Entwicklungsstadium keine Auswertung möglicher Informationen enthalten. Die folgenden Beschreibungen beschränken sich daher auf die Ergebnisse mit

Ausgleich der rechts-links Verwechslungen.

In Abschnitt 2.4 wurde gezeigt, dass die Klangverfärbung der Sprachsignale eine auswertbare akustische Eigenschaft für die Schätzung der relativen Sprecherausrichtung darstellt. Der darauf basierende Algorithmus ist im derzeitigen Entwicklungsstadium nicht in der Lage, diese Eigenschaft zufriedenstellend auszunutzen. Es wurde davon ausgegangen, dass die Abstrahlcharakteristiken im Mittel zwischen männlichen und weiblichen Sprechern keine signifikanten Unterschiede ergeben (vgl. 2.3). Einzig aufgrund unterschiedlicher Kopfgrößen der verwendeten Sprecher sind dennoch individuelle Unterschiede zu erwarten, was zu unterschiedlicher richtungsabhängiger Tiefpassfilterung der Sprachsignale führt. Weiterhin ist davon auszugehen, dass die Übertragungsfunktion der Mundabstrahlung, z. B. durch unterschiedliche Betonung und Artikulation, ebenfalls individuell ist und somit zwischen den unterschiedlichen Sprechern variiert.

Das schlechtere Abschneiden des mGCF im reflexionsarmen Raum war zu erwarten, da hier keine sinnvolle Aussage über das Verhältnis des Direktschalls zu den (fehlenden) Reflexionen getroffen werden kann. Im halligen Bürorraum führt der Algorithmus hingegen zu besseren Ergebnissen. Die Sprecherausrichtung wird mit einer mittleren Abweichung von  $13^\circ$  erfolgreicher unterschieden, als bei den normalhörenden Probanden mit  $20^\circ$  Abweichung.

Vielversprechend ist das Ergebnis des LPC-Ansatzes, da er in allen getesteten Situationen zu guten bis sehr guten Ergebnissen geführt hat. Interessant ist das Abschneiden vor allem im reflexionsarmen Raum, da bei der Entwicklung des Verfahrens vor allem auf den enthaltenen Direktschallanteil und die frühen Reflexionen eingegangen wurde. Ein Erklärungsansatz für die guten Ergebnisse ohne vorhandene Reflexionen ergibt sich aus der richtungsabhängigen Tiefpassfilterung der Quelle. Je stärker die Tiefpassfilterung bei steigender Sprecherausrichtungen ausfällt, desto weniger spitz ausgeprägt sind die im Zeitsignal enthaltenen Glottisimpulse, da hierfür hohe Frequenzen erforderlich sind. Das verwendete Peakigkeitsmaß  $P^P$  ist dadurch sowohl im reflexionsarmen Raum als auch im Büro in der Lage, die Sprecherausrichtung zu unterscheiden. Die mittlere Abweichung ist mit  $15^\circ$  im reflexionsarmen Raum vergleichbar mit der Abweichung der Probandentests, im halligen Bürorraum mit  $4^\circ$  deutlich besser.

---

# Kapitel 5

## Zusammenfassung, Erkenntnisse und Ausblick

In diesem Kapitel werden die Inhalte der Arbeit zusammengefasst und die gewonnenen Erkenntnisse beschrieben. Zum Abschluss wird ein Ausblick auf mögliche weitere Schritte gegeben, in dem Verbesserungsvorschläge der Algorithmen und Probandenhörversuche vorgestellt werden.

### 5.1 Zusammenfassung

In dieser Arbeit wurden die menschlichen und algorithmischen Fähigkeiten zur Schätzung der relativen Sprecherausrichtung analysiert.

Einleitend wurden die raumakustischen Grundlagen und die Effekte interauraler Laufzeitdifferenz und Pegeldifferenz beschrieben. Beide Themen sind für das Verständnis der akustischen Einflüsse auf die Mikrofonsignale in Abhängigkeit von der Sprecherausrichtung äußerst relevant. In Hinblick auf die entwickelten Hörgerätealgorithmen wurden das Signalmodell für mehrkanalige Richtmikrofone sowie die zugrundeliegenden Verfahren und Modelle beschrieben.

Die Fähigkeiten zur Schätzung der relativen Sprecherausrichtung wurden bei normalhörenden Menschen bisher in nur wenigen wissenschaftlichen Untersuchungen behandelt und für Schwerhörige sind bisher keine Ergebnisse bekannt. Drei der veröffentlichten Untersuchungen wurden ausführlich beschrieben und die Ergebnisse vorgestellt und kritisch betrachtet. Dabei zeigte sich bereits, dass die Wahrnehmung der relativen Sprecherausrichtung von einigen akustischen Eigenschaften abhängen kann. Darauf wurde die Abstrahlcharakteristik des menschlichen Kopfes mithilfe des *Transmission-Line-Matrix-Methode* in einem reflexionsarmen Raum und ei-

nem halligen Raum simuliert und die resultierende Schallausbreitung auf mögliche akustische Effekte untersucht. Hieraus ergaben sich einige auswertbare akustische Eigenschaften, wie z. B. die richtungsabhängigen Pegeldifferenzen, interaurale Laufzeitdifferenzen und Pegeldifferenzen und Einflüsse auf Direktschall und frühe Reflexionen in Abhängigkeit von der Sprecherausrichtung.

Einige dieser akustischen Einflüsse wurden mit Hilfe von Hörversuchen bei normalhörenden Probanden getestet, um ein besseres Verständnis für die Wahrnehmung dieser Effekte zu erlangen. Gleichzeitig wurden die Ergebnisse als Referenz für die Erkennungsgleistung der entwickelten Algorithmen verwendet. Bei den Hörversuchen wurden Ja/Nein Experimente und Zuordnungsexperimente mit unterschiedlichen Quellsignalen in unterschiedlichen Räumen durchgeführt. Anders als bei den zuvor vorgestellten, veröffentlichten Untersuchungen konnten durch das verwendete Testsetup mit Kunstkopfaufnahmen einige negative Effekte, welche die Probanden in ihrer Bewertungen beeinflussen können, minimiert werden. Hierzu gehören beispielsweise auftretende Geräusche während der Drehung der Quelle oder unterschiedliche Pausendauern zwischen den Ausrichtungsänderungen. Zusätzlich konnten die Experimente auch mit monotoner Darbietung erfolgen. Dies ist in einer Insitu-Situation nicht möglich ist.

Im Rahmen der vorangegangenen Untersuchungen hat sich gezeigt, dass die richtungsabhängige Klangverfärbung und das Verhältnis zwischen dem Direktschall und den frühen Reflexionen wichtige und verwertbare akustische Unterschiede in Abhängigkeit von der Sprecherausrichtung hervorrufen. Mit diesen Erkenntnissen konnten drei unterschiedliche Algorithmen für binaural gekoppelte Hörgeräte entwickelt werden. Der erste Algorithmus versucht die mittlere Steigung der Sprachanteile zu schätzen, um so die richtungsabhängige Klangverfärbung auszunutzen. Der zweite Algorithmus basiert auf dem globalen Kohärenzfeld. Hier konnte festgestellt werden, dass der Direktschall und die frühen Reflexionen erkennbar und auswertbar sind. Der dritte Algorithmus bestimmt das LPC-Residualsignal und bestimmt das Verhältnis zwischen Direktschall und die frühen Reflexionen anhand der Peakigkeit des Signals. Dies ist möglich, da mit Hilfe des LPC-Verfahrens näherungsweise nur der Einfluss des Raumes und das Glottissignal im Residual enthalten sind. Durch den verminderten Direktschallanteil bei einem abgewandten Sprecher sinkt das Verhältnis zwischen Direktschall und Reflexionen. Die Glottisimpulse werden als Folge bei steigender Sprecherausrichtung im Residual-Signal geringer. Der verwendete

LPC-Ansatz nutzt diese Signaländerungen

### 5.2 Erkenntnisse

In den durchgeföhrten Simulationen zeigte sich, dass nur geringe binaurale Effekte durch unterschiedliche Sprecherausrichtungen auftreten. Größere und verwertbare Unterschiede in den richtungsabhängigen Signalen haben sich in der spektralen Verfärbung und dem Verhältnis zwischen dem Direktschall zu den frühen Reflexionen gezeigt. Unterstützt werden diese Erkenntnisse durch die Hörversuche mit normalhörenden Probanden. Im reflektionsarmen Raum wird ausschließlich die Klangverfärbung der Signale als akustische Eigenschaft genutzt und in halliger Umgebung auch der Einfluss der Reflexionen. Mittels monotischen und dichotischen Test konnte gezeigt werden, dass die Probanden keinen signifikanten Gewinn durch die dichotische Darbietung erreichen. Einen Einfluss unterschiedlicher Sprecher auf die Erkennungsleistung konnte nicht erkannt werden.

Für die algorithmische Umsetzung wurden binaural gekoppelte, mehrkanalige Hörgeräte vorausgesetzt. Bisherige Verfahren setzten große Mikrofonarrays oder räumlich verteilte Mikrofone voraus. Diese Algorithmen können aufgrund des maximal möglichen Mikrofonabstands, der durch den Kopfdurchmesser begrenzt wird, nicht oder nur eingeschränkt eingesetzt werden. Somit sind neue Ansätze erforderlich. Ziel dieser Arbeit war es daher, Verfahren zu suchen und zu entwickeln und für die gegebene Aufgabe zu evaluieren.

Die bei allen Algorithmen auftretenden rechts-links Verwechslungen wurden erwartet, da die entwickelten Algorithmen im derzeitigen Entwicklungsstadium keine Auswertung möglicher Informationen enthalten.

Die Evaluation der vorgestellten Algorithmen hat gezeigt, dass der LPC-Ansatz gut geeignet ist, um die unterschiedlichen Sprecherausrichtungen zu unterscheiden, sogar besser als die normalhörenden Probanden.

### 5.3 Ausblick

Die Evaluation des ersten Algorithmus unter Verwendung der mittleren Steigung des Frequenzganges fällt unerwartet schlecht aus, obwohl die Probandenversuche gezeigt haben, dass die Klangveränderung eine entscheidende Eigenschaft für die Unterscheidung der Sprecherausrichtung darstellt. Eine Überprüfung und Optimierung des Verfahrens sollte daher noch erfolgen.

Sehr gut hat der LPC-Ansatz die unterschiedlichen Sprecherausrichtungen unterscheiden können. Daher sollte dieser Ansatz verfolgt und optimiert werden. Eine mögliche Optimierung stellt die Untersuchung weiterer Peakigkeitsmaße dar. Weiterhin wurde bei dem Algorithmus nur ein VAD eingesetzt. Die Einflüsse auf das Residual wurden nur bei Impulsanregung vorgestellt, so dass es sinnvoll erscheint, einen Detektor einzusetzen, der die impulsartige Anregung detektiert und nur in diesen Abschnitten die Peakigkeit des Signals zu verwerten.

Im aktuellen Entwicklungsstadium sind die Algorithmen nicht in der Lage, die relative Sprecherposition absolut zu bestimmen. Im Rahmen der Evaluation wurde bisher die Fähigkeit untersucht, anhand der verwendeten Maße die vorgegebenen Sprecherausrichtungen zu unterscheiden. Bisher wurden die Algorithmen nur in zwei Räumen verwendet. Es ist davon auszugehen, dass unterschiedliche Raumeigenschaften auch die Fähigkeiten der Algorithmen beeinflussen. Daher ist es erforderlich entsprechende, raumabhängige Schwellwerte zur Bewertung zu entwickeln und zu verwenden, z. B. unter Verwendung des Hallradius.

Bisher wurden die Algorithmen nur einzeln betrachtet. Es wird erwartet, dass die Kombination der Algorithmen zu besseren Erkennungsraten führt. Weiterhin wurden die Algorithmen bisher nur dahingehend überprüft, ob die verwendeten Maße geeignet sind, unterschiedliche Sprecherausrichtungen sicher zu unterscheiden. Es wurden im Rahmen dieser Arbeit keine Untersuchen durchgeführt, um eine Sprecherausrichtung direkt zu schätzen. Die Schätzung der Ausrichtung ist jedoch entscheidend für den Einsatz in Hörgeräten und muss daher weiter untersucht werden. Ungelöst ist bisher auch die Unterscheidung der Sprecherausrichtung nach links oder rechts. Allerdings erscheint es für eine Hauptsprecherdetektion im ersten Schritt ausreichend zu sein, zu detektieren, ob der Sprecher in Richtung des Empfängers blickt oder nicht.

Zudem wurden bisher nur einzelne Sprecher in ruhigen Umgebungen betrachtet. Der Einsatz möglicher Algorithmen in Multisprecher-Situationen mit und ohne Einfluss von zusätzlichen Störgeräuschen sind das Ziel, derzeit jedoch noch nicht erreicht. Somit steckt noch viel Forschungspotenzial im Bereich der algorithmischen Schätzung von der relativen Sprecherausrichtung.

---

# **Anhang A**

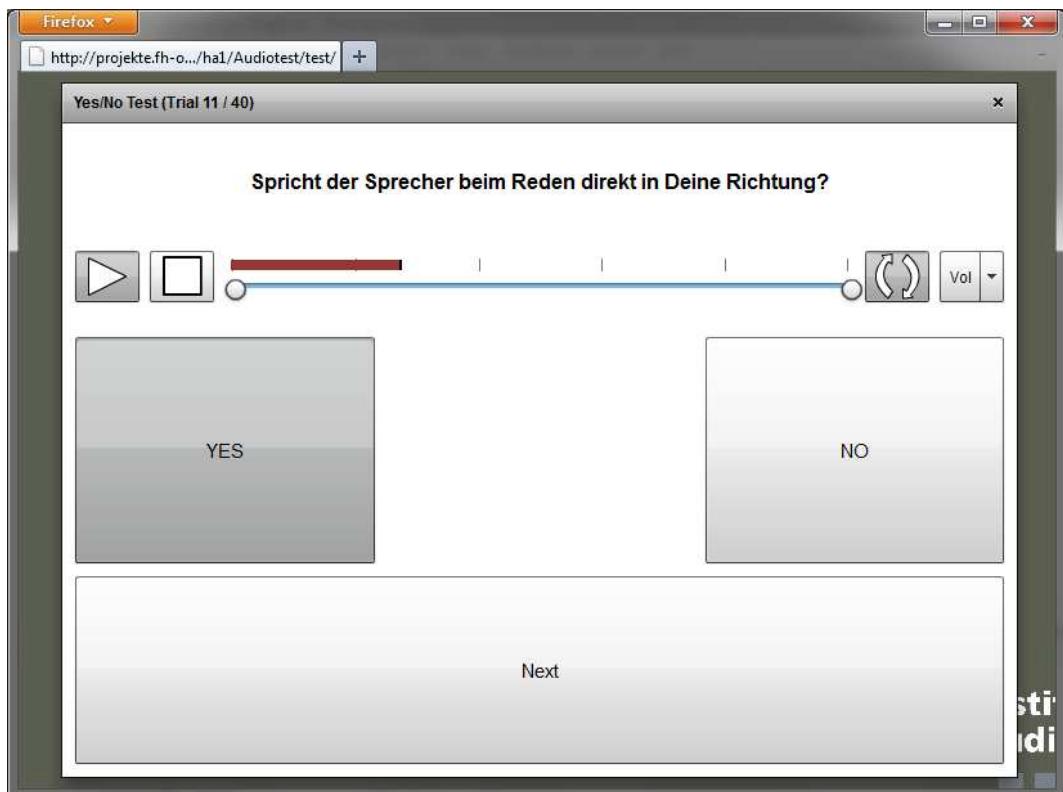
## **Anhang**

### **A.1 Grafische Benutzer Oberflächen der durchgeführten psychoakustische Experimente zur Schätzung der Sprecherausrichtung mit normalhörenden Probanden**

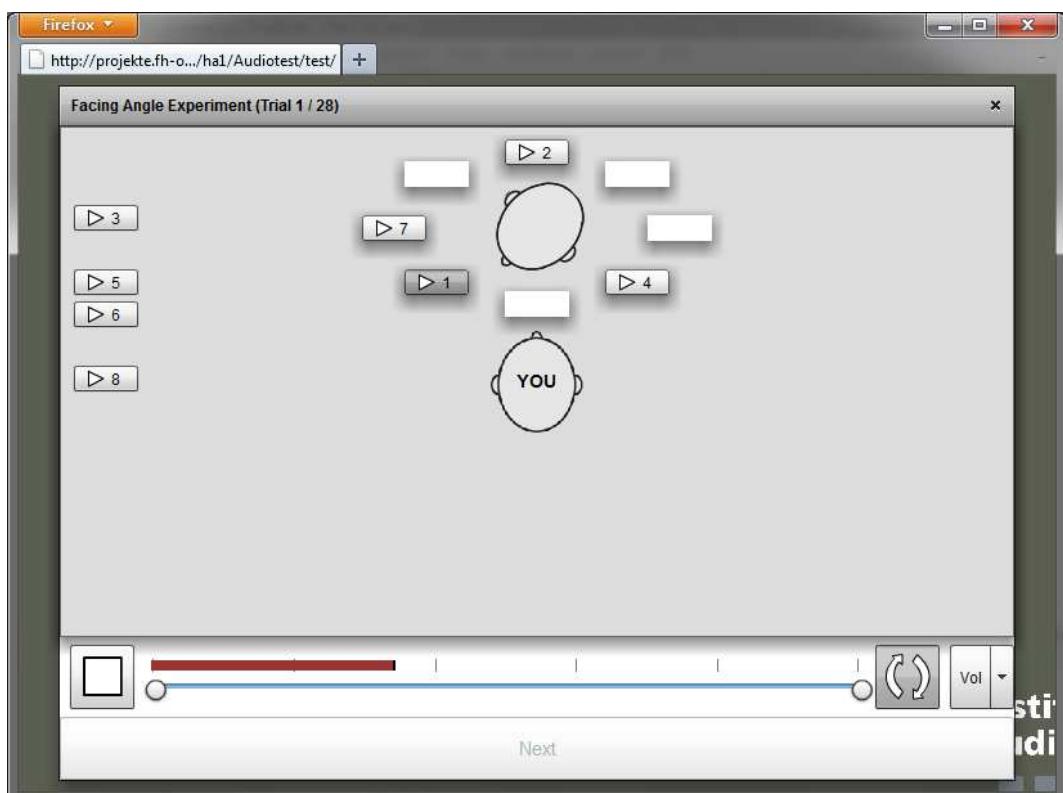
Die vorgestellten Experimente erfolgten über eine Online-Hörversuchs Applikation, welche in Adobe Flex implementiert und am Institut für Hörtechnik und Audiologie (Oldenburg) entwickelt wurde. Diese Applikation stellt eine Auswahl von Hörversuchssetups bereit, z. B. Ja/Nein-Experimente (siehe Abbildung A.1), MUSHRA, Paarvergleiche und Zuordnungsexperimente (siehe Abbildung A.2). Als Audiomaterial können sowohl MP3- als auch WAV-Dateien verwendet werden. Bei den durchgeführten Experimenten wurden ausschließlich WAV-Dateien genutzt. Der Vorteil der verwendeten Online-Applikation besteht darin, dass sie von jedem internetfähigem Rechner mit aktuellem Adobe Flash Plugin ausgeführt werden kann.

## Anhang A. Anhang

---



**Abb. A.1:** Screenshot der in Adobe Flex entwickelten Online-Hörversuchs Applikation des Instituts für Hörtechnik und Audiologie. Dargestellt ist die Oberfläche eines Ja/Nein Experimentes.



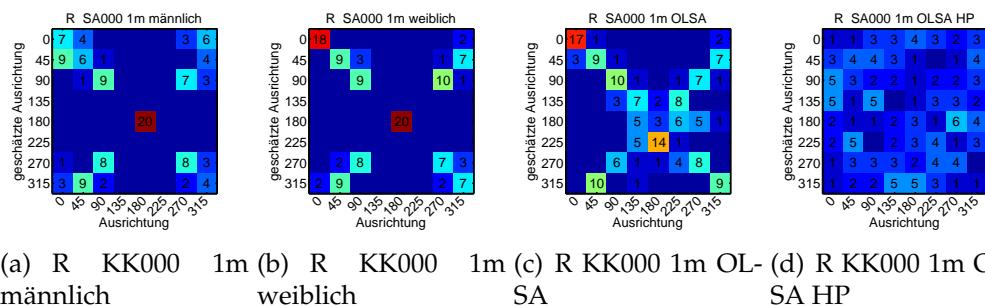
**Abb. A.2:** Screenshot der in Adobe Flex entwickelten Online-Hörversuchs Applikation des Instituts für Hörtechnik und Audiologie. Dargestellt ist die Oberfläche eines Zuordnungsexperimentes.

## A.2 Ergebnisse der Hörtests mit normalhörenden Probanden

Die folgenden Abbildungen enthalten die Verwechslungsmatrizen der Experimente zur Schätzung der Sprecherausrichtung mit normalhörenden Probanden. Die Bezeichnungen der verwendeten Audiofiles in den Abbildungen und Tabellen der Ergebnisse sind so gewählt, dass daraus die jeweilige Aufnahmesituation ersichtlich ist. Jede Bezeichnung beginnt mit der Sprecherausrichtung in Grad ( $SA000, \dots, SA360$ ), gefolgt vom Abstand zwischen Quelle und Empfänger (1m oder 2m). Anschließend folgt das Quellsignal (*männlich, weiblich, OLSA, etc.*). Für den Fall, dass es sich um ein Signal mit einem Ausgleich der quellenabhängigen Tiefpasscharakteristik handelt, wird zusätzlich ein *HP* an die Bezeichnung angehängt. So ergibt sich z. B. für den hochpassgefilterten männlichen Sprecher im Abstand von 1 m und einer Sprecherausrichtung von  $135^\circ$  die Bezeichnung  $SA135\ 1m\ \text{männlich}\ HP$ .

Auf der x-Achse der Verwechslungsmatrizen befinden sich die wahren Sprecherausrichtungen, auf der y-Achse die von den Probanden zugeordneten Ausrichtungen.

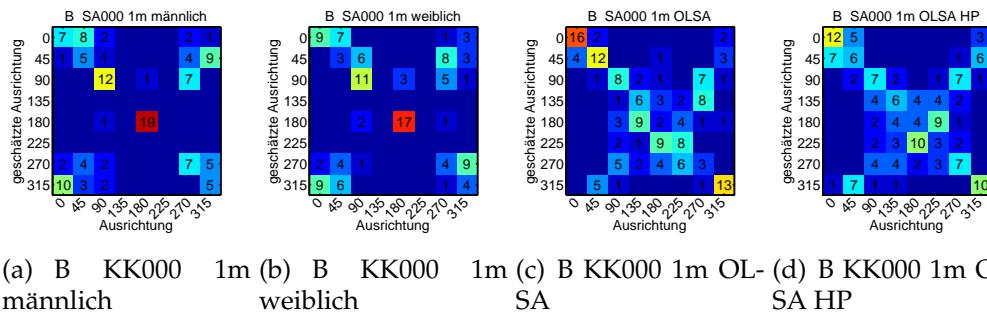
### A.2.1 Experiment 5: Monotische Schätzung der Kopfausrichtung mit Sprachsignalen



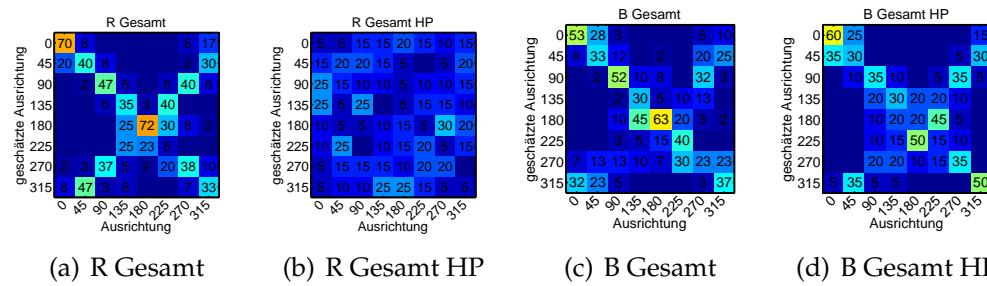
**Abb. A.3:** Verwechslungsmatrizen Experimente 5: Monotische Schätzung der Sprecherausrichtung mit Sprachsignalen von normalhörenden Probanden im reflexionsarmen Raum in Abhängigkeit des Quellsignals

## Anhang A. Anhang

---

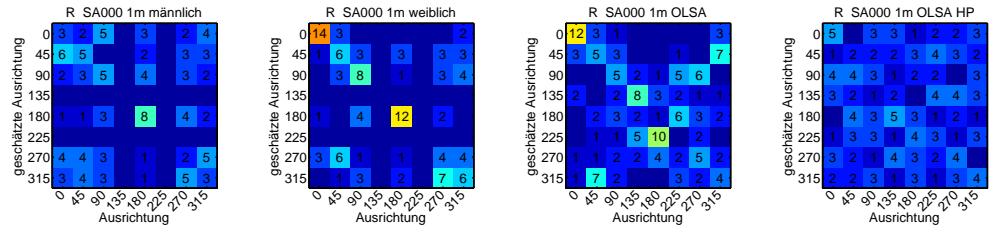


**Abb. A.4:** Verwechslungsmatrizen Experimente 5: Monotische Schätzung der Sprecherausrichtung mit Sprachsignalen von normalhörenden Probanden im halligen Büroraum

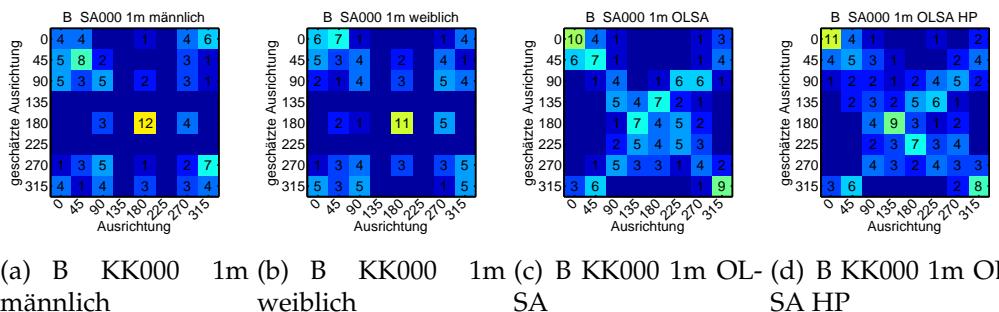


**Abb. A.5:** Verwechslungsmatrizen Experimente 5: Monotische Schätzung der Sprecherausrichtung mit Sprachsignalen von normalhörenden Probanden in Abhängigkeit von den Räumen und einer möglichen Hochpassfilterung

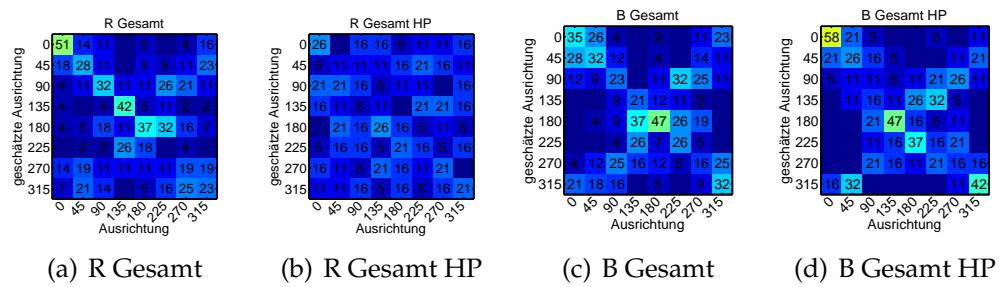
## A.2.2 Experiment 6: Monotische Schätzung der Kopfausrichtung mit LPC-Residual-Signalen



**Abb. A.6:** Verwechslungsmatrizen Experimente 6: Monotische Schätzung der Sprecherausrichtung mit LPC-Residual-Signalen von normalhörenden Probanden im reflexionsarmen Raum in Abhängigkeit des Quellsignals

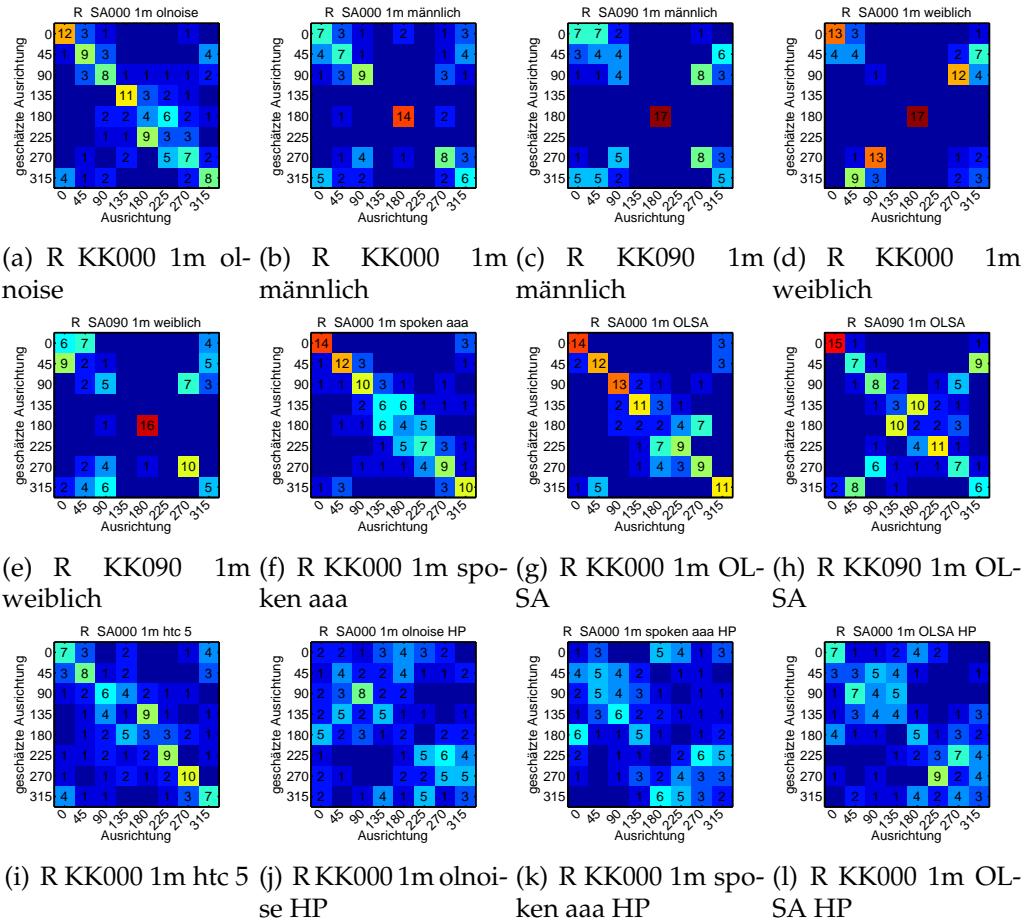


**Abb. A.7:** Verwechslungsmatrizen Experimente 6: Monotische Schätzung der Sprecherausrichtung mit LPC-Residual-Signalen von normalhörenden Probanden im halligen Büroraum in Abhängigkeit des Quellsignals



**Abb. A.8:** Verwechslungsmatrizen Experimente 6: Monotische Schätzung der Sprecherausrichtung mit LPC-Residual-Signalen von normalhörenden Probanden in Abhängigkeit von den Räumen und einer möglichen Hochpassfilterung

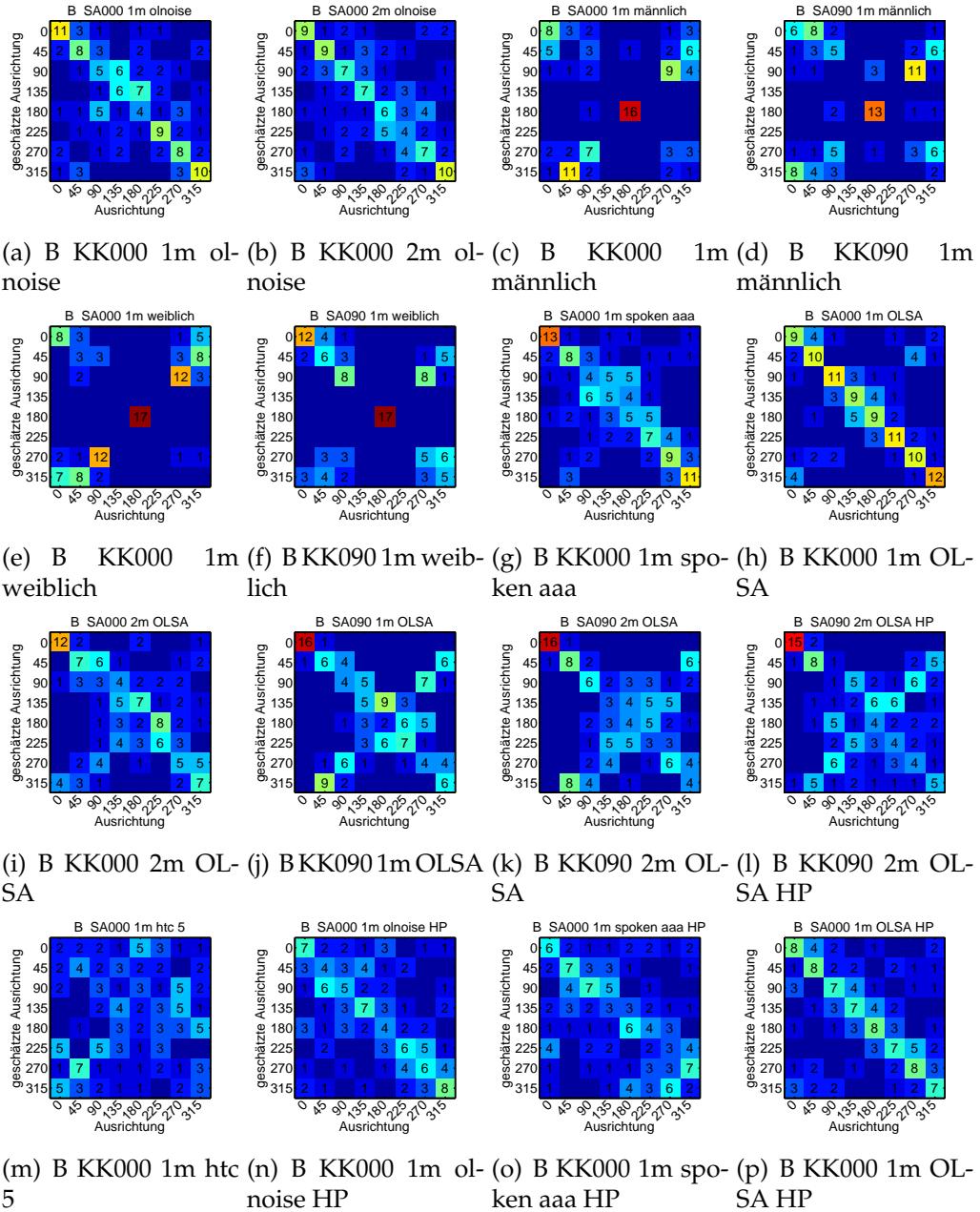
### A.2.3 Experiment 7: Dichotische Schätzung der Kopfausrichtung mit Sprachsignalen



**Abb. A.9:** Verwechslungsmatrizen Experimente 7: Dichotische Schätzung der Sprecherausrichtung von normalhörenden Probanden im reflexionsarmen Raum in Abhängigkeit des Quellsignals

## Anhang A. Anhang

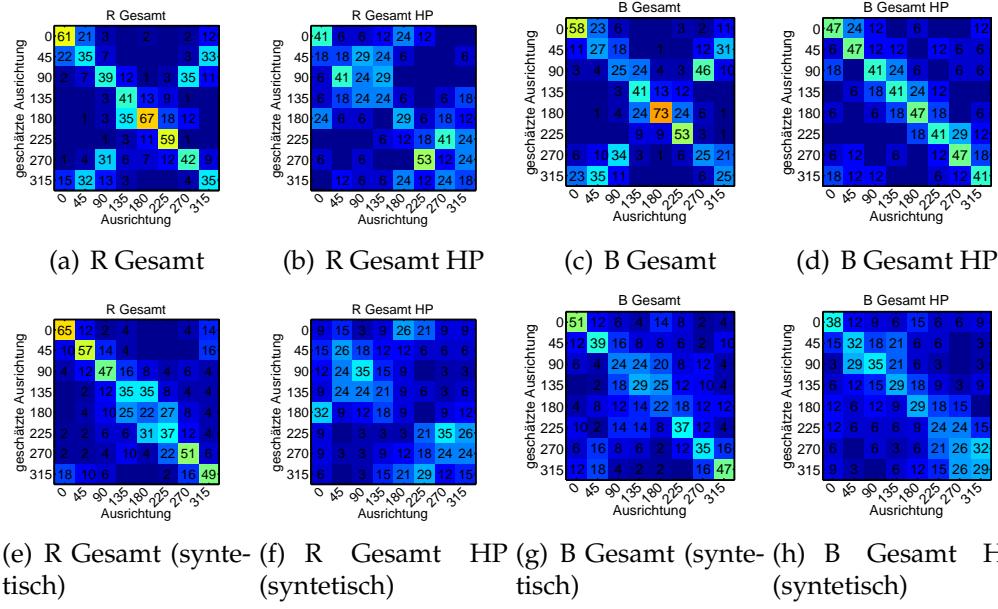
---



**Abb. A.10:** Verwechslungsmatrizen Experimente 7: Dichotische Schätzung der Sprecherausrichtung von normalhörenden Probanden im halligen Bürroraum in Abhängigkeit des Quellsignals

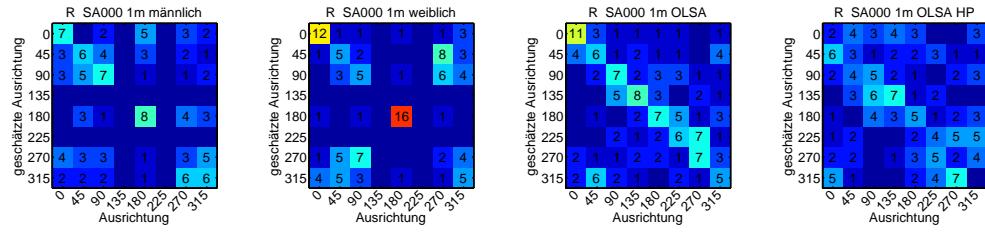
## Anhang A. Anhang

---

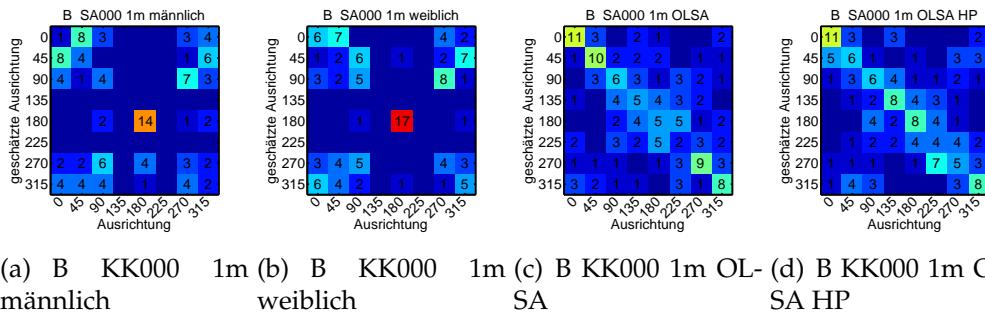


**Abb. A.11:** Verwechslungsmatrizen Experimente 7: Dichotische Schätzung der Sprecherausrichtung von normalhörenden Probanden von normalhörenden Probanden in Abhängigkeit von den Räumen und einer möglichen Hochpassfilterung

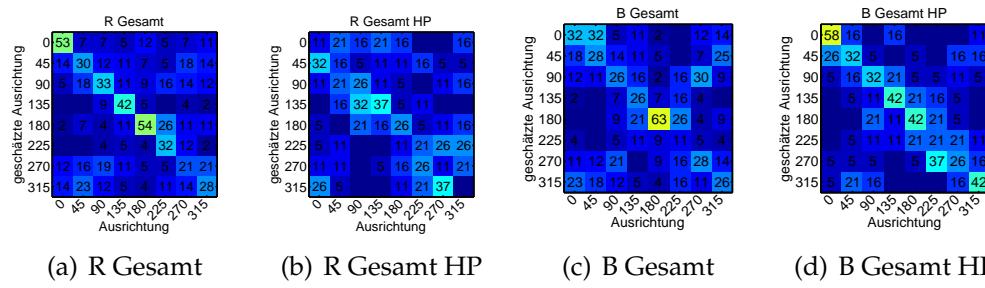
### A.2.4 Experiment 8: Dichotische Schätzung der Kopfausrichtung mit LPC-Residual-Signalen



**Abb. A.12:** Ergebnisse Experimente 8: Dichotische Schätzung der Sprecherausrichtung mit LPC-Residual-Signalen von normalhörenden Probanden im reflexionsarmen Raum in Abhängigkeit des Quellsignals



**Abb. A.13:** Ergebnisse Experimente 8: Dichotische Schätzung der Sprecherausrichtung mit LPC-Residual-Signalen von normalhörenden Probanden im halligen Büroraum in Abhängigkeit des Quellsignals

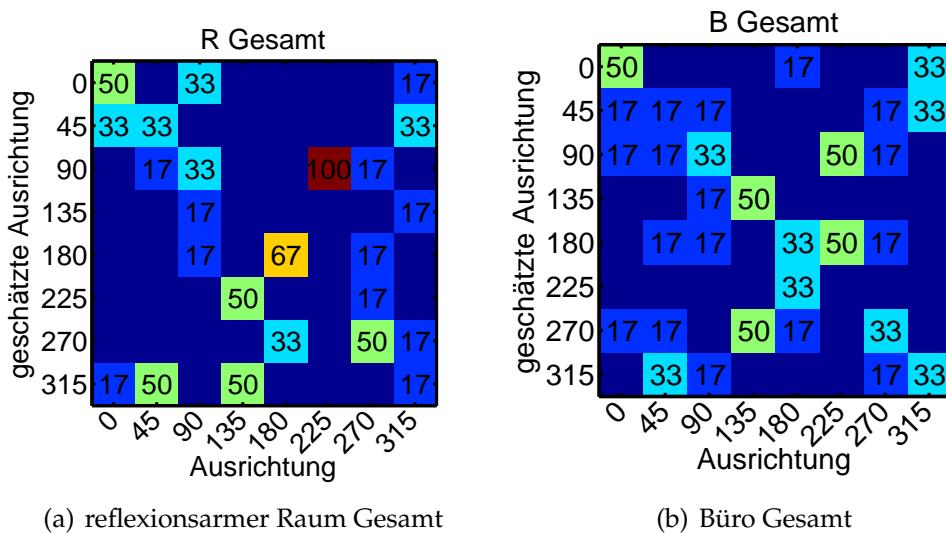


**Abb. A.14:** Ergebnisse Experimente 8: Dichotische Schätzung der Sprecherausrichtung mit LPC-Residual-Signalen von normalhörenden Probanden von normalhörenden Probanden in Abhängigkeit von den Räumen und einer möglichen Hochpassfilterung

### A.3 Ergebnisse der Algorithmen-Evaluation zur Schätzung der Sprecherausrichtung unter Verwendung der mittlere Steigung des Frequenzgangs

Die folgenden Abbildungen enthalten die Verwechslungsmatrizen der Experimente zur algorithmischen Schätzung der Sprecherausrichtung mittels Steigung des Leistungsdichtespektrums. Auf der x-Achse der Verwechslungsmatrizen befinden sich die wahren Sprecherausrichtungen, auf der y-Achse die von den Algorithmen zugeordneten Ausrichtungen.

**Experiment 7: Dichotische Schätzung der Kopfausrichtung mit Sprachsignalen**

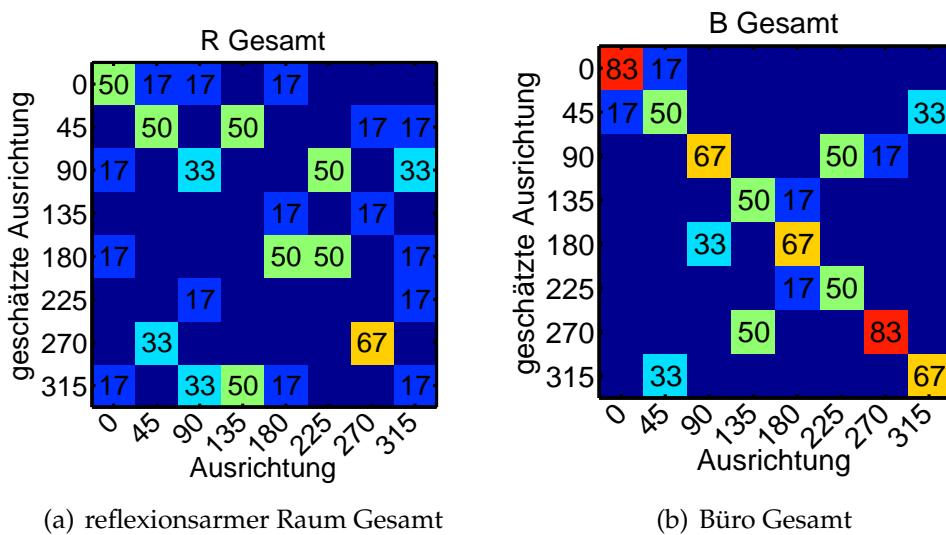


**Abb. A.15:** Verwechslungsmatrizen Experimente 7: Dichotische Schätzung der Sprecherausrichtung unter Verwendung der mittlere Steigung des Frequenzgangs in Abhängigkeit von den Räumen.

## A.4 Ergebnisse der Algorithmen-Evaluation zur Schätzung der Sprecherausrichtung unter Verwendung des globalen Kohärenzfeldes

Die folgenden Abbildungen enthalten die Verwechslungsmatrizen der Experimente zur algorithmischen Schätzung der Sprecherausrichtung unter Verwendung des globalen Kohärenzfeldes. Auf der x-Achse der Verwechslungsmatrizen befinden sich die wahren Sprecherausrichtungen, auf der y-Achse die von den Algorithmen zugeordneten Ausrichtungen.

**Experiment 7: Dichotische Schätzung der Kopfausrichtung mit Sprachsignalen**



**Abb. A.16:** Verwechslungsmatrizen Experimente 7: Dichotische Schätzung der Sprecherausrichtung unter Verwendung des globalen Kohärenzfeldes in Abhängigkeit von den Räumen.

## A.5 Ergebnisse der Algorithmen-Evaluation zur Schätzung der Sprecherausrichtung unter Verwendung der linearen Prädiktion (Kurtosis $P^K$ )

Die folgenden Abbildungen und Tabellen enthalten die Ergebnisse der Experimente zur algorithmischen Schätzung der Sprecherausrichtung unter Verwendung der linearen Prädiktion (Kurtosis  $P^K$ ). Die Bezeichnungen der verwendeten Audiofiles in den Abbildungen und Tabellen der Ergebnisse sind so gewählt, dass daraus die jeweilige Aufnahmesituation ersichtlich ist. Jede Bezeichnung beginnt mit der Sprecherausrichtung in Grad ( $SA000, \dots, SA360$ ), gefolgt vom Abstand zwischen Quelle und Empfänger ( $1m$  oder  $2m$ ). Anschließend folgt das Quellsignal (*männlich, weiblich, OLSA, etc*). Für den Fall, dass es sich um ein Signal mit einem Ausgleich der quellenabhängigen Tiefpasscharakteristik handelt, wird zusätzlich ein *HP* an die Bezeichnung angehängt. So ergibt sich z. B. für den hochpassgefilterten männlichen Sprecher im Abstand von  $1m$  und einer Sprecherausrichtung von  $135^\circ$  die Bezeichnung *SA135 1m männlich HP*.

Auf der x-Achse der Verwechslungsmatrizen befinden sich die wahren Sprecherausrichtungen, auf der y-Achse die von den Probanden oder Algorithmen zugeordneten Ausrichtungen.

### A.5.1 Experiment 1: Unterscheidung der Kopfausrichtung im reflexionsarmen Raum ( $0^\circ$ vs. $180^\circ$ )

**Tab. A.1:** Ergebnisse Experiment 1: Unterscheidung der Sprecherausrichtung im reflexionsarmen Raum unter Verwendung der linearen Prädiktion (Kurtosis  $P^K$ ). Tabellen zeigen die Sensitivität, die Spezifität und die Korrektklassifikationsrate für Unterscheidung von den Sprecherausrichtungen von  $0^\circ$  und  $180^\circ$  für unterschiedliche Quellsignale.

Signal	SENS	SPEZ	KORR
SA000 1m männlich	0.909	1.000	0.950
SA045 1m männlich	0.800	1.000	0.875
SA090 1m männlich	0.690	1.000	0.775
SA135 1m männlich	0.952	1.000	0.975
SA180 1m männlich	0.950	0.950	0.950
SA000 1m weiblich	0.923	0.704	0.775
SA045 1m weiblich	1.000	1.000	1.000
SA090 1m weiblich	1.000	0.909	0.950
SA135 1m weiblich	0.895	0.857	0.875
SA180 1m weiblich	0.944	0.864	0.900
SA000 1m OLSA	0.870	1.000	0.925
SA045 1m OLSA	0.800	1.000	0.875
SA090 1m OLSA	0.909	1.000	0.950
SA135 1m OLSA	0.769	1.000	0.850
SA180 1m OLSA	0.867	0.720	0.775
Gesamt $\mu$	0.885	0.934	0.893
$\sigma$	0.088	0.103	0.074

### A.5.2 Experiment 2: Unterscheidung der Kopfausrichtung im Büro ( $0^\circ$ vs. $180^\circ$ )

**Tab. A.2:** Ergebnisse Experiment 2: Unterscheidung der Sprecherausrichtung im Büro unter Verwendung der linearen Prädiktion (Kurtosis  $P^K$ ). Tabellen zeigen die Sensitivität, die Spezifität und die Korrektklassifikationsrate für Unterscheidung von den Sprecherausrichtungen von  $0^\circ$  und  $180^\circ$  für unterschiedliche Quellsignale.

Signal	SENS	SPEZ	KORR
SA000 1m männlich	0.941	0.826	0.875
SA045 1m männlich	0.952	1.000	0.975
SA090 1m männlich	0.870	1.000	0.925
SA135 1m männlich	0.826	0.941	0.875
SA180 1m männlich	0.900	0.900	0.900
SA000 1m weiblich	0.882	0.783	0.825
SA045 1m weiblich	0.870	1.000	0.925
SA090 1m weiblich	0.905	0.947	0.925
SA135 1m weiblich	0.947	0.905	0.925
SA180 1m weiblich	1.000	0.741	0.825
SA000 1m OLSA	0.909	1.000	0.950
SA045 1m OLSA	0.769	1.000	0.850
SA090 1m OLSA	0.645	1.000	0.725
SA135 1m OLSA	0.909	1.000	0.950
SA180 1m OLSA	0.818	0.889	0.850
Gesamt $\mu$	0.876	0.929	0.887
$\sigma$	0.087	0.087	0.065

### A.5.3 Experiment 3: Unterscheidung der Kopfausrichtung im reflexionsarmen Raum ( $0^\circ$ vs. $90^\circ$ )

**Tab. A.3:** Ergebnisse Experiment 3: Unterscheidung der Sprecherausrichtung im reflexionsarmen Raum unter Verwendung der linearen Prädiktion (Kurtosis  $P^K$ ). Tabellen zeigen die Sensitivität, die Spezifität und die Korrektklassifikationsrate für Unterscheidung von den Sprecherausrichtungen von  $0^\circ$  und  $90^\circ$  für unterschiedliche Quellsignale.

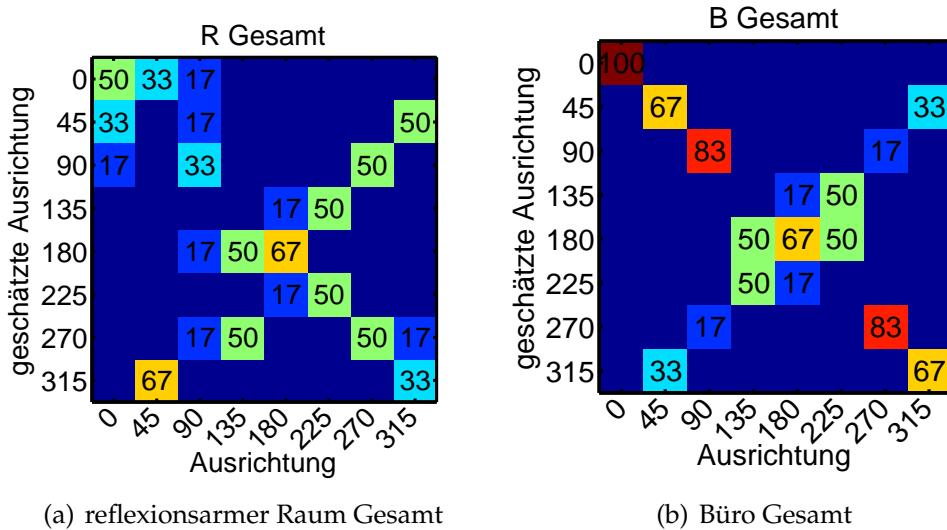
Signal	SENS	SPEZ	KORR
SA000 1m männlich	0.952	1.000	0.975
SA045 1m männlich	0.474	0.000	0.450
SA090 1m männlich	0.514	0.600	0.525
SA135 1m männlich	0.500	0.000	0.500
SA180 1m männlich	0.720	0.867	0.775
SA000 1m weiblich	0.667	0.514	0.525
SA045 1m weiblich	0.500	0.000	0.500
SA090 1m weiblich	0.846	0.667	0.725
SA135 1m weiblich	0.900	0.900	0.900
SA180 1m weiblich	1.000	0.556	0.600
SA000 1m OLSA	0.947	0.905	0.925
SA045 1m OLSA	1.000	1.000	1.000
SA090 1m OLSA	0.950	0.950	0.950
SA135 1m OLSA	0.950	0.950	0.950
SA180 1m OLSA	0.500	0.500	0.500
Gesamt $\mu$	0.761	0.627	0.720
$\sigma$	0.214	0.370	0.213

### A.5.4 Experiment 4: Unterscheidung der Kopfausrichtung im Büro ( $0^\circ$ vs. $90^\circ$ )

**Tab. A.4:** Ergebnisse Experiment 4 und 12: Unterscheidung der Sprecherausrichtung im Büro unter Verwendung der linearen Prädiktion (Kurtosis  $P^K$ ). Tabellen zeigen die Sensitivität, die Spezifität und die Korrektklassifikationsrate für Unterscheidung von den Sprecherausrichtungen von  $0^\circ$  und  $90^\circ$  für unterschiedliche Quellsignale.

Signal	SENS	SPEZ	KORR
SA000 1m männlich	0.789	0.762	0.775
SA045 1m männlich	0.556	1.000	0.600
SA090 1m männlich	0.500	0.000	0.500
SA135 1m männlich	0.909	1.000	0.950
SA180 1m männlich	0.923	0.704	0.775
SA000 1m weiblich	0.714	0.545	0.575
SA045 1m weiblich	0.679	0.917	0.750
SA090 1m weiblich	0.870	1.000	0.925
SA135 1m weiblich	1.000	0.833	0.900
SA180 1m weiblich	0.500	0.500	0.500
SA000 1m OLSA	1.000	1.000	1.000
SA045 1m OLSA	0.909	1.000	0.950
SA090 1m OLSA	0.909	1.000	0.950
SA135 1m OLSA	1.000	1.000	1.000
SA180 1m OLSA	1.000	0.909	0.950
Gesamt $\mu$	0.817	0.811	0.807
$\sigma$	0.184	0.282	0.183

### A.5.5 Experiment 7: Dichotische Schätzung der Kopfausrichtung mit Sprachsignalen



**Abb. A.17:** Verwechslungsmatrizen Experimente 7: Dichotische Schätzung der Sprecherausrichtung unter der linearen Prädiktion (Kurtosis  $P^K$ ) in Abhängigkeit von den Räumen.

## A.6 Ergebnisse der Algorithmen-Evaluation zur Schätzung der Sprecherausrichtung unter Verwendung der linearen Prädiktion (Crest Faktor $P^C$ )

Die folgenden Abbildungen und Tabellen enthalten die Ergebnisse der Experimente zur algorithmischen Schätzung der Sprecherausrichtung unter Verwendung der linearen Prädiktion (Crest Faktor  $P^C$ ). Die Bezeichnungen der verwendeten Audiofiles in den Abbildungen und Tabellen der Ergebnisse sind so gewählt, dass daraus die jeweilige Aufnahmesituation ersichtlich ist. Jede Bezeichnung beginnt mit der Sprecherausrichtung in Grad ( $SA000, \dots, SA360$ ), gefolgt vom Abstand zwischen Quelle und Empfänger ( $1m$  oder  $2m$ ). Anschließend folgt das Quellsignal (*männlich, weiblich, OLSA, etc*). Für den Fall, dass es sich um ein Signal mit einem Ausgleich der quellenabhängigen Tiefpasscharakteristik handelt, wird zusätzlich ein *HP* an die Bezeichnung angehängt. So ergibt sich z. B. für den hochpassgefilterten männlichen Sprecher im Abstand von  $1m$  und einer Sprecherausrichtung von  $135^\circ$  die Bezeichnung *SA135 1m männlich HP*.

Auf der x-Achse der Verwechslungsmatrizen befinden sich die wahren Sprecherausrichtungen, auf der y-Achse die von den Algorithmen zugeordneten Ausrichtungen.

### A.6.1 Experiment 1: Unterscheidung der Kopfausrichtung im reflexionsarmen Raum ( $0^\circ$ vs. $180^\circ$ )

**Tab. A.5:** Ergebnisse Experiment 1: Unterscheidung der Sprecherausrichtung im reflexionsarmen Raum unter Verwendung der linearen Prädiktion (Crest Faktor  $P^C$ ). Tabellen zeigen die Sensitivität, die Spezifität und die Korrektklassifikationsrate für Unterscheidung von den Sprecherausrichtungen von  $0^\circ$  und  $180^\circ$  für unterschiedliche Quellsignale.

Signal	SENS	SPEZ	KORR
SA000 1m männlich	0.909	1.000	0.950
SA045 1m männlich	0.909	1.000	0.950
SA090 1m männlich	0.690	1.000	0.775
SA135 1m männlich	0.741	1.000	0.825
SA180 1m männlich	0.952	1.000	0.975
SA000 1m weiblich	0.923	0.704	0.775
SA045 1m weiblich	1.000	0.909	0.950
SA090 1m weiblich	1.000	0.870	0.925
SA135 1m weiblich	0.905	0.947	0.925
SA180 1m weiblich	0.947	0.905	0.925
SA000 1m OLSA	0.760	0.933	0.825
SA045 1m OLSA	0.769	1.000	0.850
SA090 1m OLSA	0.909	1.000	0.950
SA135 1m OLSA	0.800	1.000	0.875
SA180 1m OLSA	0.895	0.857	0.875
Gesamt $\mu$	0.874	0.942	0.890
$\sigma$	0.097	0.084	0.067

### A.6.2 Experiment 2: Unterscheidung der Kopfausrichtung im Büro ( $0^\circ$ vs. $180^\circ$ )

**Tab. A.6:** Ergebnisse Experiment 2: Unterscheidung der Sprecherausrichtung im Büro unter Verwendung der linearen Prädiktion (Crest Faktor  $P^C$ ). Tabellen zeigen die Sensitivität, die Spezifität und die Korrektklassifikationsrate für Unterscheidung von den Sprecherausrichtungen von  $0^\circ$  und  $180^\circ$  für unterschiedliche Quellsignale.

Signal	SENS	SPEZ	KORR
SA000 1m männlich	0.938	0.792	0.850
SA045 1m männlich	0.952	1.000	0.975
SA090 1m männlich	0.870	1.000	0.925
SA135 1m männlich	0.833	1.000	0.900
SA180 1m männlich	0.952	1.000	0.975
SA000 1m weiblich	0.938	0.792	0.850
SA045 1m weiblich	0.870	1.000	0.925
SA090 1m weiblich	0.905	0.947	0.925
SA135 1m weiblich	0.952	1.000	0.975
SA180 1m weiblich	1.000	0.741	0.825
SA000 1m OLSA	0.909	1.000	0.950
SA045 1m OLSA	0.714	1.000	0.800
SA090 1m OLSA	0.690	1.000	0.775
SA135 1m OLSA	0.909	1.000	0.950
SA180 1m OLSA	0.810	0.842	0.825
Gesamt $\mu$	0.883	0.941	0.895
$\sigma$	0.089	0.096	0.068

### A.6.3 Experiment 3: Unterscheidung der Kopfausrichtung im reflexionsarmen Raum ( $0^\circ$ vs. $90^\circ$ )

**Tab. A.7:** Ergebnisse Experiment 3: Unterscheidung der Sprecherausrichtung im reflexionsarmen Raum unter Verwendung der linearen Prädiktion (Crest Faktor  $P^C$ ). Tabellen zeigen die Sensitivität, die Spezifität und die Korrektklassifikationsrate für Unterscheidung von den Sprecherausrichtungen von  $0^\circ$  und  $90^\circ$  für unterschiedliche Quellsignale.

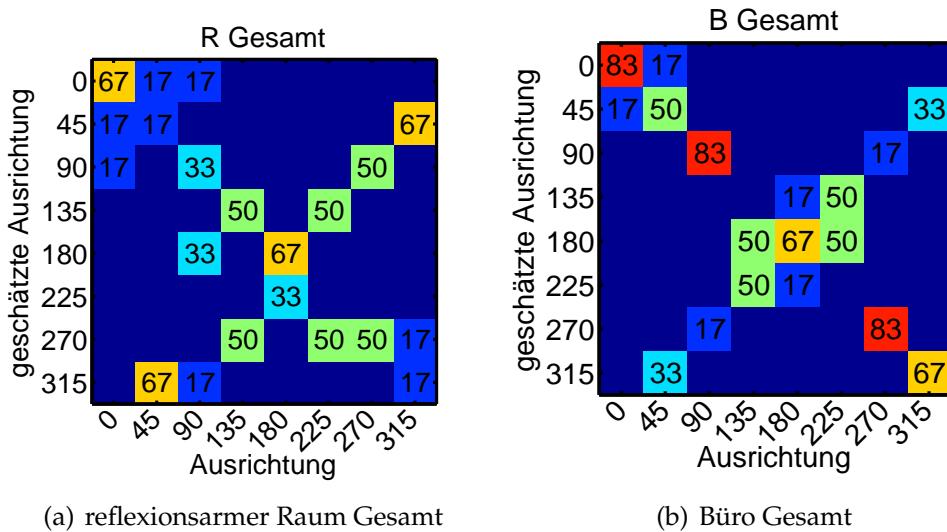
Signal	SENS	SPEZ	KORR
SA000 1m männlich	0.500	0.000	0.500
SA045 1m männlich	0.500	0.000	0.500
SA090 1m männlich	0.513	1.000	0.525
SA135 1m männlich	0.500	0.000	0.500
SA180 1m männlich	0.500	0.000	0.500
SA000 1m weiblich	1.000	0.526	0.550
SA045 1m weiblich	0.818	0.621	0.675
SA090 1m weiblich	0.667	0.529	0.550
SA135 1m weiblich	0.952	1.000	0.975
SA180 1m weiblich	0.800	0.543	0.575
SA000 1m OLSA	0.923	0.704	0.775
SA045 1m OLSA	1.000	0.952	0.975
SA090 1m OLSA	0.950	0.950	0.950
SA135 1m OLSA	0.944	0.864	0.900
SA180 1m OLSA	0.500	0.500	0.500
Gesamt $\mu$	0.738	0.546	0.663
$\sigma$	0.216	0.384	0.195

### A.6.4 Experiment 4: Unterscheidung der Kopfausrichtung im Büro ( $0^\circ$ vs. $90^\circ$ )

**Tab. A.8:** Ergebnisse Experiment 4 und 12: Unterscheidung der Sprecherausrichtung im Büro unter Verwendung der linearen Prädiktion (Crest Faktor  $P^C$ ). Tabellen zeigen die Sensitivität, die Spezifität und die Korrektklassifikationsrate für Unterscheidung von den Sprecherausrichtungen von  $0^\circ$  und  $90^\circ$  für unterschiedliche Quellsignale.

Signal	SENS	SPEZ	KORR
SA000 1m männlich	0.750	0.667	0.700
SA045 1m männlich	0.526	1.000	0.550
SA090 1m männlich	0.500	0.000	0.500
SA135 1m männlich	0.714	1.000	0.800
SA180 1m männlich	0.941	0.826	0.875
SA000 1m weiblich	0.667	0.529	0.550
SA045 1m weiblich	0.792	0.938	0.850
SA090 1m weiblich	0.864	0.944	0.900
SA135 1m weiblich	1.000	0.870	0.925
SA180 1m weiblich	0.500	0.500	0.500
SA000 1m OLSA	1.000	1.000	1.000
SA045 1m OLSA	0.909	1.000	0.950
SA090 1m OLSA	0.909	1.000	0.950
SA135 1m OLSA	1.000	1.000	1.000
SA180 1m OLSA	1.000	0.833	0.900
Gesamt $\mu$	0.805	0.807	0.797
$\sigma$	0.187	0.281	0.186

### A.6.5 Experiment 7: Dichotische Schätzung der Kopfausrichtung mit Sprachsignalen

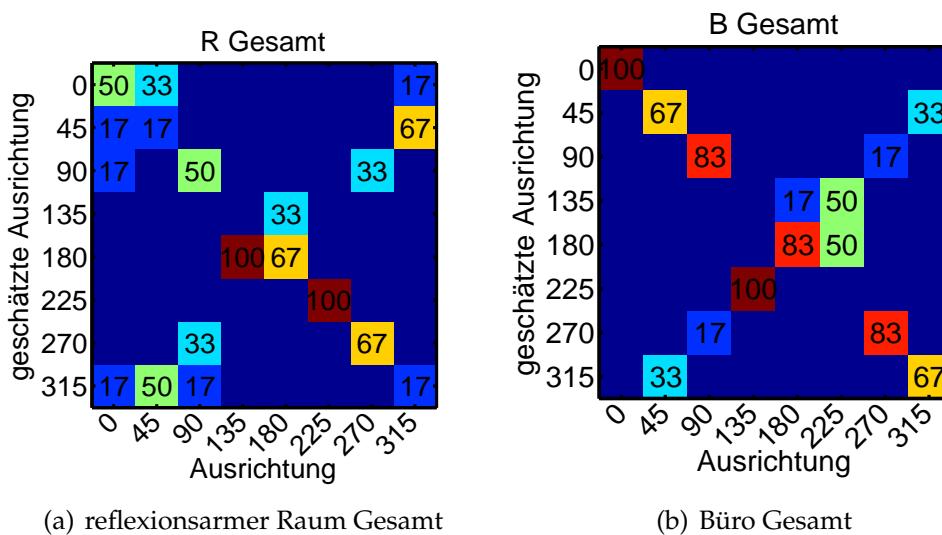


**Abb. A.18:** Verwechslungsmatrizen Experimente 7: Dichotische Schätzung der Sprecherausrichtung unter der linearen Prädiktion (Crest Faktor  $P^C$ ) in Abhängigkeit von den Räumen.

## A.7 Ergebnisse der Algorithmen-Evaluation zur Schätzung der Sprecherausrichtung unter Verwendung der linearen Prädiktion (Peakigkeit $P^P$ )

Die folgenden Abbildungen und Tabellen enthalten die Ergebnisse der Experimente zur algorithmischen Schätzung der Sprecherausrichtung unter Verwendung der linearen Prädiktion (Peakigkeit  $P^P$ ). Auf der x-Achse der Verwechslungsmatrizen befinden sich die wahren Sprecherausrichtungen, auf der y-Achse die von den Algorithmen zugeordneten Ausrichtungen.

**Experiment 7: Dichotische Schätzung der Kopfausrichtung mit Sprachsignalen**



**Abb. A.19:** Verwechslungsmatrizen Experimente 7: Dichotische Schätzung der Sprecherausrichtung unter der linearen Prädiktion (Peakigkeit  $P^P$ ) in Abhängigkeit von den Räumen.

---

# Anhang B

## Abkürzungen

### Abkürzungen:

dB	dezi Bell (logarithmische Maßeinheit)
dB-A	A-gewichteter Pegel in dezi Bell (logarithmische Maßeinheit)
DFT	Diskrete Fourier Transformation
DTFT	Zeitdiskrete Fourier Transformation engl. <i>Time-Discrete Fourier-Transform</i>
AKF	Autokorrelationsfunktion
CSP	cross-power spectrum phase
et al.	et alii
FFT	schnelle Fourier-Transformation engl. <i>Fast Fourier-Transform</i>
$f_s$	Abtastrate
GCC-PHAT	allgemeine Kreuzkorrelation mit Phasentransformation
GCF	Global Coherence Field
HRTF	head related transfer function, Kopfübertragungsfunktion
HP	hochpassgefiltert
HTC	Harmonischer Tonkomplex
IDFT	Inverse Diskrete Fourier Transformation
IDTFT	Inverse Zeitdiskrete Fourier Transformation
ILD	interaural level difference, interaurale Pegeldifferenz
ITD	interaural time difference, interaurale Laufzeitdifferenz
IPD	interaurale Phasendifferenz
KKF	Kreuzkorrelationsfunktion
KLDS	Kreuzleistungsdichtefunktion
KORR	Korrektklassifikationsrate
LP	lineare Prädiktion
LPC	lineare Prädiktionskoeffizienten
MAFA	minimum audible facing angle
mGCF	modifiziertes Global Coherence Field

## Anhang B. Abkürzungen

---

MMA	minimum audible angle
oGCF	oriented Global Coherence Field
RIA	Raumimpulsantwort
RMS	Effektivwert
SENS	Sensitivität
SPEZ	Spezifität
OLSA	Oldenburger Satztest
TDE	time delay estimation
VAD	Sprachpausendetektion engl. <i>Voice Activity Detection</i>
vgl.	vergleiche
z.B.	zum Beispiel
$\pi$	3.14159265359...

---

# Literaturverzeichnis

- [AB76] Allen, J. B. und Berkley, D. A.: *Image method for efficiently simulating small-room acoustics*. The Journal of the Acoustical Society of America, 60(S1):S9–S9, 1976.
- [Bar08] Bartsch, C.: *Schätzung der Kopfausrichtung durch mehrkanalige digitale Signalverarbeitungsalgorithmen*. Diplom-Arbeit, Fachhochschule Oldenburg / Ostfriesland / Wilhelmshaven, 2008.
- [BC06] Bentler, R. und Chiou, L.: *Digital Noise Reduction: An Overview*. Trends in Amplification, 10(2):67–82, Januar 2006.
- [Ben08] Benesty, J.: *Springer handbook of speech processing*. Springer, 2008.
- [Bit02] Bitzer, J.: *Mehrkanalige Geräuschunterdrückungssysteme - eine vergleichende Analyse*. Doktorarbeit, Dept. of Telecommunications, Universitaet Bremen, Aachen, September 2002.
- [Bit09] Bittner, S.: *Iterative Korrektur von nichtlinearen Verzerrungen und Phasenstörungen in Mehrträgersystemen*. Jörg Vogt Verlag, 2009.
- [Bla97] Blauert, J.: *Spatial hearing: the psychophysics of human sound localization*. The MIT Press, Cambridge, Massachusetts, Revised Edition Auflage, 1997.
- [BMS01] Bitzer (Meyer), J. und Simmer, K.: *Superdirective Microphone Arrays*. In: Brandstein, M. und Ward, D. (Herausgeber): *Microphone Arrays*, Kapitel 2, Seiten 19–37. Springer, Berlin, Heidelberg, New York, May 2001.
- [BOS05] Brutti, A., Omologo, M. und Svaizer, P.: *Oriented global coherence field for the estimation of the head orientation in smart rooms equipped with distributed microphone arrays*. Interspeech, Seiten 2337–2340, 2005.

- [CFSC<sup>+</sup>08] Canton-Ferrer, C., Segura, C., Casas, J. R., Pardàs, M. und Hernando, J.: *Audiovisual head orientation estimation with particle filtering in multisensor scenarios*. EURASIP J. Adv. Signal Process, 2008, January 2008.
- [CG91] Chandler, D. W. und Grantham, D. W.: *Minimum audible movement angle in the horizontal plane as a function of stimulus frequency and bandwidth, source azimuth, and velocity*. Division of Hearing and Speech Sciences, Vanderbilt University School of Medicine, Nashville, 1991.
- [CW02] Chu, W. und Warnock, A.: *Detailed Directivity of Sound Fields Around Human Talkers*. Technischer Bericht, Institute of Research in Construction, 2002.
- [DF39] Dunn, H. und Farnsworth, D.: *Exploration of Pressure Field Around the Human Head During Speech*. JASA, 10:184–199, 1939.
- [DHB01] DiBiase, J., Harvey, S. und Brandstein, M.: *Robust Localization in Reverberant Rooms*. In: *Microphone Arrays*. Springer, 2001.
- [DM98] Duda, R. O. und Martens, W. L.: *Range dependence of the response of a spherical head model*. J. Acoust. Soc. Am., 104, 1998.
- [EL11] Lombard, Étienne: *Le signe de l'élévation de la voix*. Ann. Mal. Oreil. Larynx, 37:101–119, 1911.
- [Fra70] Frant, G.: *Acoustic Therory of Speec Production*. Mouton & Co., The Hague, Niederlande, 2. Auflage, 1970.
- [Fra08] Franz, S.: *Multimikrofon-Algorithmen zur Geräuschreduktion für die Anwendung in Hörgeräten: Implementierung und Evaluation auf der Echtzeitplattform Master Hearing Aid*. Bachelor Thesis, Fachhochschule Oldenburg / Ostfriesland /Wilhelmshaven, 2008.
- [GMvdP<sup>+</sup>09] Georganti, E., May, T., Par, S.van de , Härmä, A. und Mourjopoulos, J.: *Single channel sound source distance estimation based on statistical and source specific features*. Audio Engineering Society, 2009.
- [Gör08] Görne, T.: *Tontechnik*. Hanser Verlag, Januar 2008.
- [IEC05] IEC TC29/WG13: *DIN EN 60 118-7:2005*, 2005.

- [JB71] Johns, P. B. und Beurle, R. L.: *Numerical solutions of 2-dimensional scattering problems using a transmission-line matrix.* Proceeding of the IEE, 118:1203–1208, 1971.
- [KC76] Knapp, C. und Carter, G.: *The generalized correlation method for estimation of time delay.* IEEE Transactions on Acoustics, Speech and Signal Processing, 24(4):320– 327, August 1976.
- [Kra11] Kramme, R.: *Medizintechnik: Verfahren - Systeme - Informationsverarbeitung.* Springer, Mai 2011.
- [KTNM10] Kato, H., Takemoto, H., Nishimura, R. und Mokhtari, P.: *On the Human Ability to Auditorily Perceive Human Speaker's Facing Angle.* Universal Communication Symposium, 2010.
- [Mak75] Makhoul, J.: *Linear Prediction: A Tutoeial Review.* Proceedings of the IEEE, Band 63, 1975.
- [Mar95] Martin, R.: *Freisprecheinrichtungen mit mehrkanaliger Echokompensation und Störgeräuschunterdrückung.* Doktorarbeit, Institut für Nachrichtengeräte und Datenverarbeitung, TH Achen, Verlag der Augustinus Buchhandlung, Achen Deutschland, 1995.
- [McK86] McKendree, F.: *Directivity Indices of Human Talkers in English Speech.* Proceedings Inter-Noise 86, Seiten 911–916, 1986.
- [Mil58] Mills, A. W.: *On the Minimum Audible Angle.* J. Acoust. Soc. Am., 30:237–246, 1958.
- [Mil60] Mills, A. W.: *Lateralization of high frequency tones.* J. Acoust. Soc. Am, 32:132–134, 1960.
- [MK02] Marzinzik, M. und Kollmeier, B.: *Speech Pause Detection for Noise Spectrum Estimation by Tracking Power Envelope Dynamics.* IEEE TRANSACTIONS ON SPEECH AND AUDIO PROCESSING, 10:109–118, 2002.
- [Moo82] Moore, B. C. J.: *An Introduction to the Psychology of Hearing.* Academic Press, 1982.
- [Moo03] Moore, B. C. J.: *An Introduction to the Psychology of Hearing.* Academic Press, 2003.

- [MP78] Moreno, A. und Pfretzschnner, J.: *Human Head Directivity in Speech Emission: A New Approach.* Acoustics Letters, 1:78–84, 1978.
- [Neu03] Neuhoff, J. G.: *Twist and Shout: Audible Facing Angles and Dynamic Rotation.* ECOLOGICAL PSYCHOLOGY, 15:335–351, 2003.
- [NRV01] Neuhoff, J. G., Rodstrom, M.-A. und Vaidya, T.: *The audible facing angle.* Acoustical Society of America, Seiten 109–114, 2001.
- [NYN09] Nakano, A. Y., Yamamoto, K. und Nakagawa, S.: *DIRECTIONAL ACOUSTIC SOURCE'S POSITION AND ORIENTATION ESTIMATION APPROACH BY A MICROPHONE ARRAY NETWORK.* IEEE, Seiten 606–611, 2009.
- [Old06] Oldenburg, H.: *OLSA - Oldenburger Satztest - Handbuch und Hintergrundwissen.* HörTech gGmbH, Oldenburg, 2006. [http://www.hoertech.de/web/dateien/HT.IE.007-Handbuch\\_und\\_Hintergrundwissen.OLSA.00.1.pdf](http://www.hoertech.de/web/dateien/HT.IE.007-Handbuch_und_Hintergrundwissen.OLSA.00.1.pdf).
- [RB09] Ruhland, M. und Blau, M.: *Lambda 2.0.* <http://www.hoertechnik-audiologie.de/web/file/Forschung/Software.php#lambda>, 2009. Visited: 14.02.2012.
- [Roe06] Roeske, P.: *Sprecherlokalisierung in gestörter Umgebung.* Diplom-Arbeit, Fachhochschule Oldenburg / Ostfriesland / Wilhelmshaven, 2006.
- [Roh08] Rohdenburg, T.: *Development and objective perceptual quality assessment of monaural and binaural noise reduction schemes for hearing aids.* Doktorarbeit, Universität Oldenburg, Uhlhornsweg 49-55, 26129 Oldenburg, 2008.
- [Sla92] Slatky, H.: *Algorithmen zur richtungsselektiven Verarbeitung von Schallsignalen - die Realisierung eines binauralen Cocktail-Party-Prozessor-Systems.* Doktorarbeit, Fakultät Elektrotechnik an der Ruhr-Universität Bochum, 1992.
- [SRHE09] Schroeder, J., Rohdenburg, T., Hohmann, V. und Ewert, S. D.: *Classification of Reverberant Acoustic Situations.* DAGA, 2009.

## Literaturverzeichnis

---

- [SS04] Sachar, J. M. und Silverman, H. F.: *A baseline algorithm for estimating talker orientation using acoustical data from a large-aperture microphone array*. In: *IEEE International Conference on Acoustics, Speech, and Signal Processing, 2004. Proceedings. (ICASSP '04)*, Band 4, Seiten iv–65– iv–68 vol.4. IEEE, Mai 2004.
- [VHH98] Vary, P., Heute, U. und Hess, W.: *Digitale Sprachsignalverarbeitung*. B. G. Teubner, Stuttgart, 1998.
- [WS62] Woodworth, R. S. und Schlosberg, H.: *Experimental psychology*. Holt, Rinehard and Winston, New York, 1962.

---

# Danksagung

An dieser Stelle möchte ich den folgenden Personen danken:

- Prof. Dr. Ing. Jörg Bitzer für die sehr gute Betreuung und Unterstützung während der gesamten Masterarbeit, ebenso wie für die vielen Gespräche und Anregungen die mich immer ein Stück näher ans Ziel gebracht haben,
- Prof. Dr. ir. Simon Doclo für die ebenfalls sehr gute Betreuung und den konstruktiven Gesprächen und Anregungen während den Arbeitsgruppentreffen,
- meiner kleinen Familie Nicole Lüdke und Arwen Fenna Franz für die Kraft und Unterstützung während der gesamten Dauer des Studiums und während der Masterarbeit,
- Prof. Dr.-Ing. Frank Wallhoff für seine Geduld und Unterstützung,
- Nicole Lüdke, Carolin Duda und Johanna Stever für das Lesen und Korrigieren dieser Arbeit,
- den Studenten der Studiengänge *Hörtechnik und Audiologie* und *Assitive Technologien* und den Mitarbeitern am *Institut für Hörtechnik und Audiologie* für die Unterstützung bei den durchgeführten Aufnahmen, dem Bewerten der umfangreichen Hörversuche und für die anregenden und weiterführenden Gesprächen während der gesamten Masterarbeits-Phase,
- meinen Eltern und Brüdern für die mehr als hilfsbereite Unterstützung während des gesamten Studiums.

# Eigenständigkeitserklärung

Titel der Arbeit:

**Analyse der menschlichen und  
algorithmischen Fähigkeit zur  
Schätzung der relativen Sprecherausrichtung**

Hiermit versichere ich, dass ich diese Arbeit selbstständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe. Außerdem versichere ich, dass ich die allgemeinen Prinzipien wissenschaftlicher Arbeit und Veröffentlichung, wie sie in den Leitlinien guter wissenschaftlicher Praxis der Carl von Ossietzky Universität Oldenburg festgelegt sind, befolgt habe.

Oldenburg, der 30. März 2012

---

Ort, Datum

Sven Franz