Perzeptive Evaluation eines räumlichkeitsbewahrenden Beamformers

Mareike Buhl[1,3,*], Steven van de Par[2,3], Stephan M. A. Ernst[1,3]
[1] CvO Universität Oldenburg, Medizinische Physik, 26111 Oldenburg
[2] CvO Universität Oldenburg, AG Akustik, 26111 Oldenburg
[3] Cluster of Excellence 'Hearing4all'
∗ Email: mareike.buhl@uni-oldenburg.de

Einleitung

In Situationen mit mehreren Sprechern sowie weiteren Hintergrundgeräuschen, sogenannten Cocktail-Party-Situationen, soll zu jedem bestimmten Zeitpunkt meist nur einem der vielen Sprecher gefolgt werden. Die Extraktion dieses Zielsprechers sowie die Unterdrückung der Störquellen ist für normalhörende Personen in der Regel gut möglich [1], für Schwerhörende ist diese Aufgabe jedoch schwieriger. Eine Klasse der Hörgerätealgorithmen versucht daher das Signal-Rausch-Verhältnis (SNR) von zu verstehendem Sprecher zum Rest der Situation zu verbessern. Zur Umsetzung dieser Störgeräuschreduktion werden beispielsweise Beamformer verwendet [2][3]. Nach [3] ist für gute Sprachverständlichkeit auch der Erhalt der räumlichen Informationen der gesamten akustischen Szene wichtig. Der Binaural Linearly Constrained Minimum Variance (BLCMV) Beamformer [3][4] verspricht sowohl SNR-Verbesserung als auch Erhalt binauraler Cues. Dieser Algorithmus wird in dieser Studie umfassend evaluiert. Dabei werden der Einfluss verschiedener Einstellungen des Beamformers auf das binaurale Signal, die resultierende Sprachverständlichkeit und die empfundene Änderung des Quellenortes untersucht.

Methoden

Binaural Linearly Constrained Minimum Variance (BLCMV) Beamformer

Der BLCMV-Algorithmus [3][4] nutzt zwei Beamformer, um binaurale Eigenschaften zu repräsentieren. Ziel der Signalverarbeitung ist es, die am linken und rechten Referenzmikrofon ankommenden Signalanteile des Zielsprechers unbeeinflusst zu erhalten, während gleichzeitig Störquellen unterdrückt werden und so die Rauschleistung minimiert wird. Für mathematische Details wird auf [3] verwiesen. Der Erhalt von binauralen Informationen, d.h. interauraler Pegeldifferenz (ILD) und interauraler Phasendifferenz (IPD), wird durch eine gezielte Beibehaltung eines Anteils der Störquellen realisiert. Dadurch wird allerdings die vom Beamformer durchgeführte SNR-Verbesserung reduziert, es wird folglich ein Kompromiss zwischen SNR-Verbesserung und Erhalt binauraler Cues erreicht. Die Wichtung des Kompromisses zu mehr SNR-Verbesserung oder mehr binauraler Korrektheit ist parametrisch einstellbar über den Faktor $0 < \eta \leq 1$, der den Anteil der erhaltenen Störquelle angibt. Je kleiner η, desto weniger werden binaurale Cues erhalten und desto mehr SNR-Verbesserung wird erreicht.

Stimuli-Generation

Um den Einfluss des Beamformers auf die binaurale Wahrnehmung ohne den Einfluss der SNR-Änderung zu untersuchen, wurde zunächst für mit verschiedenen η-Einstellungen des Algorithmus verarbeitete Signale die ILD sowie die IPD nach [9] berechnet. Die verwendeten Testsignale waren hierbei eine Mischung aus einem Zielsprecher aus 0° (Satzmaterial aus dem Oldenburger Satztest [5]) und einem Störsprecher aus 15° (International Speech Test Signal [8]). Tabelle 1 und 2 zeigen die hieraus bestimmten Faktoren Q_{ILD} und Q_{IPD} (Gleichung 1, analog für IPD), welche die Veränderung der ILD und IPD im Vergleich zur unverarbeiteten Situation repräsentieren.

$$Q_{ILD}(k,\eta) = \left\langle \frac{ILD_{k,\eta}(t)}{ILD_{k,ref}(t)} \right\rangle_t \quad (1)$$

Die beobachteten Änderungen wurden anschließend in

Tabelle 1: Faktoren $Q_{IPD}(k,\eta)$ für einige IPD-relevante Frequenzkanäle mit der jeweiligen Zentralfrequenz f_c.

η / f_c [Hz]	0.1	0.3	0.6	1.0
569	0.53	0.82	0.93	0.98
660	0.64	0.90	0.98	1.02
761	0.75	0.98	1.04	1.07
874	0.55	0.89	0.97	1.00
1000	0.51	0.84	0.95	0.99
1140	0.70	0.96	1.05	1.09
1296	0.69	0.87	0.93	0.95
1470	0.71	0.89	0.93	0.94

Tabelle 2: Faktoren $Q_{ILD}(k,\eta)$ für einige ILD-relevante Frequenzkanäle mit der jeweiligen Zentralfrequenz f_c.

η / f_c [Hz]	0.1	0.3	0.6	1.0
1879	1.42	1.12	0.98	0.92
2119	1.17	1.14	1.04	0.99
2387	0.88	1.05	1.05	1.04
2685	0.78	1.06	1.12	1.14
3017	0.55	0.99	1.11	1.16
3387	0.84	0.96	0.97	0.96
3799	1.15	1.15	1.02	0.95
4259	0.43	0.95	1.06	1.10

neu erzeugte Stimuli so eingebaut, dass die Änderung binauraler Parameter isoliert von der SNR-Veränderung betrachtet werden konnte. Die Manipulation erfolgte in Gammatonfiltern [10] und war angelehnt an den Algorithmus in [11]. Die so gewonnenen Stimuli für $\eta_{eq} = (0.1, 0.3, 0.6, 1.0)$ wurden mit vollständig HRTF-verarbeiteten Signalen [12] (drei Konditionen für drei verschiedene Raumrichtungen des Störsprechers) sowie den vom original BLCMV-Beamformer verarbeiteten Signalen (bei $\eta = (0.1, 0.3)$) verglichen. Tabelle 3 gibt einen Überblick über alle Signalkonditionen.

Tabelle 3: Überblick über verwendete Signalkonditionen. Neben den im folgenden verwendeten Bezeichnungen ist angegeben, welche Parametereinstellungen verwendet und ob binaurale Cues oder SNR im Vergleich zur Ausgangskondition $H_{15°}$ manipuliert wurden.

Kondition	Bereich	Manipulation	
		bin	SNR
H_{ϕ}	$\phi = (0°, 15°, 30°)$		
$Q_{BIN,\eta_{eq}}$	$\eta_{eq} = (0.1, 0.3, 0.6, 1.0)$	x	
B_{η}	$\eta = (0.1, 0.3)$	x	x

Experiment I: Sprachverständlichkeit

Zur Messung der Sprachverständlichkeit in den verschiedenen Signalkonditionen wurde die geschlossene Version des Oldenburger Satztests [5] genutzt, d.h. alle möglichen Wortalternativen wurden angegeben. Die Sprachverständlichkeitsschwellen SRT50 und SRT80 wurden von 10 normalhörenden Versuchspersonen (∅ 24.6 Jahre) gemessen. Für jeden dargebotenen Satz musste der Proband in einer GUI auswählen, welche Wörter gehört wurden. Basierend auf dem jeweiligen Prozentsatz der korrekt verstandenen Wörter wurde der Pegel des Zielsprechers verändert, sodass die Sprachverständlichkeit adaptiv auf 50 % bzw. 80 % angepasst wurde. Der Pegel des Störsprechers betrug konstant 65 dB SPL. Für jede zu bestimmende Schwelle wurde eine 20 Sätze lange Testliste gemessen.

Experiment II: Lokalisierung

In einer zweiten Messung wurde die Veränderung des wahrgenommenen Azimuthwinkels des Störsprechers im Vergleich zur Kondition $H_{15°}$ erfasst. 10 normalhörende Versuchspersonen (∅ 25.9 Jahre) bewerteten die Signale mithilfe der subjektiven Bewertungsmethode CoDiCl (Combined Discrimination and Classification) [13]. Dabei lautete die Aufgabe, den Konditionen zugeordnete Buttons anhand der Frage "Bitte bewerten Sie die Winkeländerung der Sprecherin im Vergleich zur Referenz $H_{15°}$" zwischen den Referenzbuttons für $H_{0°}$, $H_{15°}$ und $H_{30°}$ anzuordnen. Die Bewertung erfolgte in kategorialen Einheiten (CU) im Bereich von -50 CU (starke Linksverschiebung) bis 50 CU (starke Rechtsverschiebung).

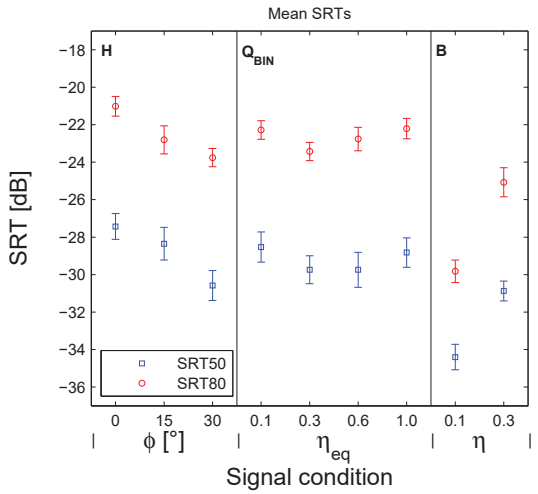

Abbildung 1: Mittelwerte und Standardfehler für SRT50 und SRT80. Dargestellt sind von links nach rechts die Signalkonditionen H_{ϕ}, $Q_{BIN,\eta_{eq}}$ und B_{η}.

Ergebnisse

Experiment I: Sprachverständlichkeit

Abbildung 1 zeigt Mittelwerte und Standardfehler für die Sprachverständlichkeitsschwellen SRT50 und SRT80. Grundsätzlich liegen die SRT80-Schwellen etwa 6 dB höher als die SRT50-Schwellen, die relative Beziehung der unterschiedlichen Konditionen ist jedoch vergleichbar für 50 % und 80 % Sprachverständlichkeit.

Innerhalb der H-Konditionen zeigt $H_{0°}$ mit ca. -27.5 dB die höchste SRT50-Schwelle. Für $H_{15°}$ liegt die Schwelle etwa 1 dB niedriger, für $H_{30°}$ weitere 2 dB niedriger. Die Sprachverständlichkeit ist wie zu erwarten höher, je größer die räumliche Trennung in Bezug auf den Azimuthwinkel zwischen Ziel- und Störsprecher ist.

In der Kondition $Q_{BIN,0.1}$ wird eine ähnliche Sprachverständlichkeit wie in $H_{15°}$ erreicht, während sich für $\eta_{eq} = 0.3$ und 0.6 eine Verbesserung der Schwelle zeigt. In $Q_{BIN,1.0}$ ist ebenfalls ein ähnlicher Wert wie für $H_{15°}$ zu beobachten.

In den BLCMV-verarbeiteten Konditionen zeigt sich für $\eta = 0.3$ gegenüber $H_{15°}$ eine leicht verbesserte Sprachverständlichkeit vergleichbar mit der in $H_{30°}$, während für $\eta = 0.1$ noch einmal 4 dB niedrigere Schwellen erreicht werden. Die maximale SNR-Verbesserung durch den Beamformer führt also zu der besten Sprachverständlichkeit aller acht betrachteten Konditionen.

Experiment II: Lokalisierung

Abbildung 2 zeigt Medianwerte der Bewertung der wahrgenommenen Winkeländerung im Vergleich zur Azimuthposition des Störsprechers in Kondition $H_{15°}$.

Die als Referenz verwendeten H-Konditionen konnten von neun Probanden erfolgreich zugeordnet werden, eine Versuchsperson verortete die $H_{30°}$-Kondition nur bei etwa 40 CU, was aber noch einer starken Rechtsverschiebung der Störsprecherposition entspricht. Somit waren

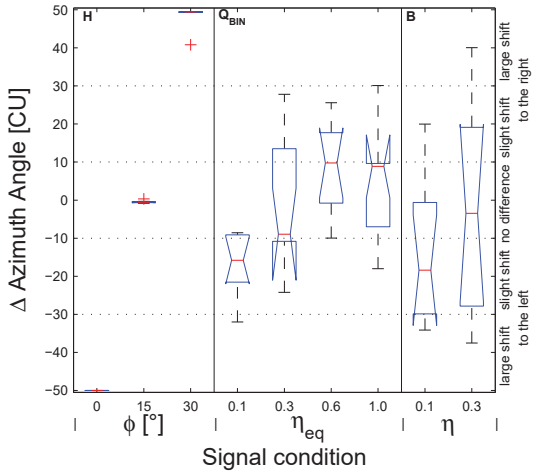

Abbildung 2: Boxplotdarstellung der Lokalisationsergebnisse. Dargestellt sind von links nach rechts die Signalkonditionen H_{ϕ}, $Q_{BIN,\eta_{eq}}$ und B_{η}.

die Probanden in der Lage, die veränderten räumlichen Eigenschaften im Signal wahrzunehmen.

Innerhalb der Q_{BIN}-Konditionen ist der Trend zu beobachten, dass für kleineres η_{eq}, d.h. für stärkere Verfälschung der binauralen Informationen, der Störsprecher mehr nach links verschoben wahrgenommen wird. Dabei wurde die Kondition mit $\eta_{eq} = 0.1$ mit -16 CU im Median bewertet, dies entspricht der Kategorie "leichte Linksverschiebung" und somit einer Verkleinerung des wahrgenommenen Winkels zwischen den beiden Sprechern um einige Grad. Für $\eta_{eq} = 0.3$ liegt die Bewertung mit -9 CU am Übergang von "kein Unterschied" zu "leichte Linksverschiebung", während $\eta_{eq} = 0.6$ (10 CU) und 1.0 (9 CU) am Übergang von "kein Unterschied" zu "leichte Rechtsverschiebung" eingeordnet wurden. Der Verlauf der Konditionen in Abhängigkeit von η_{eq} entspricht den vor der Stimuli-Generation beobachteten Änderungen der ILD und IPD.

Die Beamformer-verarbeiteten Konditionen wurden im Median mit -19 CU ("leichte Linksverschiebung") und -3 CU ("kein Unterschied") bewertet. Für alle Werte von $\eta < 0.3$ wird also der wahrgenommene räumliche Eindruck des Störsprechers verfälscht. Es ist bemerkenswert, dass in diesen beiden B-Konditionen eine besonders große Streuung zwischen den Versuchspersonen zu beobachten ist.

Diskussion

Sprachverständlichkeit

Verschiedene Faktoren wie räumliche Trennung von Ziel- und Störsprecher und durch den Beamformer angewendete SNR-Verbesserung beeinflussen die Sprachverständlichkeit. Die höchste Sprachverständlichkeit wird erreicht, wenn der BLCMV-Algorithmus mit $\eta = 0.1$ arbeitet. In dieser Kondition beträgt die SNR-Verbesserung 17 dB im Vergleich zur unverarbeiteten Situation, allerdings rücken die beiden Sprecher bezüglich des Azimuthwinkels zusammen, der ursprüngliche räumliche Eindruck kann nicht bewahrt werden. In der $B_{0.3}$-Situation, in der noch 10 dB SNR-Verbesserung erreicht werden, kann der räumliche Eindruck dagegen gut erhalten werden. Die SRT50 in der Kondition $B_{0.1}$ ist etwa 4 dB besser als in $B_{0.3}$, d.h. der Gewinn an SNR-Verbesserung drückt sich nicht vollständig in SRT-Verbesserung aus.

Die vergleichbare Sprachverständlichkeit in $B_{0.3}$ und $H_{30°}$ signalisiert, dass sowohl durch Vergrößerung der Winkeldifferenz (binauraler Gewinn) als auch durch die 10 dB SNR-Verbesserung eine Sprachverständlichkeitsverbesserung im Vergleich zu der Kondition $H_{15°}$ erreicht werden kann. Im Vergleich der Q_{BIN}- zu den B-Konditionen können Einflüsse des Erhalts binauraler Cues und der SNR-Verbesserung isoliert diskutiert werden. Von $Q_{BIN,0.1}$ zu $Q_{BIN,0.3}$ tritt eine Verbesserung der SRT um etwa 1.5 dB auf, während beim Übergang von $B_{0.1}$ auf $B_{0.3}$ eine Verschlechterung um 4 dB zu verzeichnen ist. Der Effekt der Änderung der SNR-Verbesserung beim Wechsel von $\eta = 0.3$ auf 0.1 ist demnach größer, d.h. es kann mehr Sprachverständlichkeit gewonnen werden, als durch die gleichzeitige Verfälschung binauraler Informationen verloren geht.

Lokalisierung

In den H-Konditionen konnte der Störsprecher jeweils sicher lokalisiert werden. In diesen Situationen stehen den Probanden die kompletten in der HRTF enthaltenen Informationen zur Verfügung, d.h. sowohl binaurale Cues als auch monaurale spektrale Informationen.

Durch die frequenzspezifische Manipulation von ILD und IPD sowohl in den Q_{BIN}- als auch in den B-Konditionen wird die genaue Lokalisation erschwert bzw. es tritt eine größere Streuung über die Versuchspersonen auf. Diese ist am größten in den Beamformer-verarbeiteten Situationen. Eine Deutung hierfür könnte sein, dass durch die frequenzabhängige Veränderung der binauralen Cues eine Vergrößerung der individuell wahrgenommenen Quellenbreite auftritt.

Sowohl für $Q_{BIN,0.1}$ (-16 CU) als auch für $B_{0.1}$ (-19 CU) wird in etwa dieselbe Winkelveränderung des Störsprechers angegeben. Durch die künstliche Manipulation der binauralen Cues konnte also die durch den Beamformer erzeugte Veränderung der binauralen Cues geeignet imitiert werden. Dasselbe gilt auch für die beiden Konditionen mit $\eta = 0.3$, hier ist jeweils der originale Winkeleindruck des Störsprechers erhalten.

Zusammenfassung

Die vorgestellten Experimente zur perzeptiven Evaluation des BLCMV-Beamformers zeigen, dass für maximale Sprachverständlichkeit ein möglichst hohes Signal-Rausch-Verhältnis am wichtigsten ist. Folglich sollte für den Beamformer die Einstellung $\eta = 0.1$ gewählt werden. Allerdings ist die Bedeutung von korrekter Lokalisation im Alltag nicht zu unterschätzen: in Gefahrensituationen oder allgemein zur Orientierung ist

die ungestörte räumliche Wahrnehmung sehr wichtig bzw. nützlich. Auch in Bezug auf die Qualität eines Hörgerätealgorithmus wird der Erhalt der gesamten räumlichen akustischen Szene vom Hörer wertgeschätzt. Nach [3] trägt der Erhalt der binauralen Informationen auch zu guter Sprachverständlichkeit bei.

Möglichst hohe Sprachverständlichkeit und korrekte binaurale Cues sind nicht gleichzeitig, d.h. mit einer einzigen Algorithmuseinstellung, zu erreichen. So hängt die beste Einstellung des Algorithmus stark von der jeweiligen Situation ab. Bei einem darauf abgestimmten Einsatz des BLCMV-Beamformers erlaubt dieser die einfache parametrische Anpassung zur Verwirklichung des jeweils optimalen Kompromisses.

Literatur

[1] Peissig, J. und Kollmeier, B. (1997): Directivity of binaural noise reduction in spatial multiple noise-source arrangements for normal and impaired listeners. Journal of the Acoustical Society of America 101, Nr. 3, S. 1660-1670

[2] Bronkhorst, A. (2000): The Cocktail Party Phenomenon: A Review of Research on Speech Intelligibility in Multiple-Talker Conditions. Acta Acustica united with Acustica 86, Nr. 1, S. 117-128

[3] Hadad, E., Gannot, S., Doclo, S. (2012): Binaural Linearly Constrained Minimum Variance Beamformer for Hearing Aid Applications. International Workshop on Acoustic Signal Enhancement, Proceedings of IWAENC 2012, S. 1-4

[4] Marquardt, D., Hadad, E., Gannot, S., Doclo, S. (2014): Optimal binaural LCMV beamformers for combined noise reduction and binaural cue preservation. IEEE, S. 288-292

[5] Wagener, K., Kühnel, V., Kollmeier, B. (1999): Entwicklung und Evaluation eines Satztests für die deutsche Sprache I: Design des Oldenburger Satztests. Zeitschrift für Audiologie/Audiological Acoustics 38, S. 4-15

[6] Wagener, K., Brand, T., Kollmeier, B. (1999): Entwicklung und Evaluation eines Satztests für die deutsche Sprache II: Optimierung des Oldenburger Satztests. Zeitschrift für Audiologie/Audiological Acoustics 38, S. 44-56

[7] Wagener, K., Brand, T., Kollmeier, B. (1999): Entwicklung und Evaluation eines Satztests für die deutsche Sprache III: Evaluation des Oldenburger Satztests. Zeitschrift für Audiologie/Audiological Acoustics 38, S. 86-95

[8] Holube, I., Fredelake, S., Vlaming, M., Kollmeier, B. (2010): Development and analysis of an International Speech Test Signal (ISTS). International Journal of Audiology 49, S. 891-903

[9] Dietz, M., Ewert, S., Hohmann, V. (2011): Auditory model based direction estimation of concurrent speakers from binaural signals. Speech Communication, Nr. 53, S. 592-605

[10] Hohmann, V. (2002): Frequency analysis and synthesis using a Gammatone filterbank. Acta Acustica united with Acustica 88, Nr. 3, S. 433-442

[11] Kollmeier, B. und Peissig, J. (1990): Speech intelligibility enhancement by interaural magnification. Acta Otolaryngol. Suppl. 469, S.215-223

[12] Kayser, H., Ewert, S., Anemüller, J., Rohdenburg, T., Hohmann, V., Kollmeier, B. (2009): Database of Multichannel In-Ear and Behind-the-Ear Head-Related and Binaural Room Impulse Responses. Eurasip Journal on Advances in Signal Processing

[13] Völker, C., Bisitz, T, Huber, R., Vormann, M., Ernst, S.M.A. (2015): Perceptual Evaluation Methods - applicable for elder and technical inexperienced participants? Journal of International Advanced Otology. 12th European Federation of Audiology (EFAS) Congress, Istanbul, Turkey, Abstract Book p. 54