

Erhaltung der räumlichen Wahrnehmung bei Störgeräuschreduktion in Hörgeräten

Menno Müller¹, Joachim Thiemann¹, Daniel Marquardt², Simon Doclo², Steven van de Par¹

¹ Acoustics Group, ² Signal Processing Group, Universität Oldenburg, 26129 Oldenburg, Deutschland,

Email: menno.mueller@uni-oldenburg.de

Einleitung

In Situationen mit mehreren anwesenden Schallquellen sind Normalhörende in der Lage eine bestimmte Schallquelle zu fokussieren und alle anderen zu unterdrücken (Cocktailparty Effekt [1]). Die räumliche Trennung der Schallquellen ist dabei hilfreich für das auditorische System. Es nutzt die binauralen cues, d.h. die interauralen Zeit- (ITD) und Pegeldifferenzen (ILD) und es kommt zu räumlicher Demaskierung. Bei Hörgeschädigten ist dieses oftmals reduziert und das nötige Signal-to-Noise-Ratio (SNR) für Sprachverstehen in individueller Weise erhöht [1, 2], was ihnen erschwert solche Situationen zu meistern. Abhilfe schaffen Störgeräuschalgorithmen welche die Hintergrundgeräusche (HG) reduzieren. Dabei kann die räumliche Wahrnehmung der Hintergrundgeräusche verloren gehen, was die Lokalisation von Schallquellen erschwert. In dieser Studie wurden drei Algorithmen entwickelt und evaluiert, die gleichermaßen eine Störgeräuschreduktion und Erhaltung der binauralen cues der HG ermöglichen.

Algorithmen

Abb. 1 zeigt das Blockdiagramm des Algorithmus', welcher hier als "selective binaural beamformer (SBB)" bezeichnet wird.

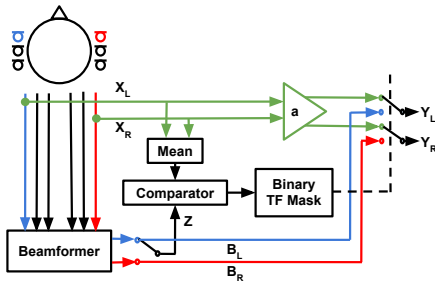


Abbildung 1: Blockdiagramm des "Selective Binaural Beamformer"

Kern des Algorithmus' ist ein binauraler Beamformer [3]. Dieser lässt ein Signal aus einer bestimmten wählbaren Zielrichtung praktisch ungefiltert passieren und reduziert Signale aus anderen Richtungen. Im zweikanaligen Ausgangssignal bleiben die binauralen cues des Zielsignals (ZS) erhalten, während die der HG verloren gehen. Um dies zu verhindern, wird das Beamformerausgangssignal mit dem Eingangssignal aus den vorderen beiden Mikrofonen (Referenzmikrofone) auf Basis einer binären Maske im Zeitfrequenzbereich gemischt, welche sich folgender-

maßen berechnet:

$$BM(f, n) = \begin{cases} ZS & |\bar{X}(f, n)| \leq |Z(f, n)| \\ HG & |\bar{X}(f, n)| > |Z(f, n)| \end{cases} \quad (1)$$

Die Grundidee der binären Maske basiert auf einem Vergleich mit dem Kanal des Beamformer-Ausgangs Z dessen Referenzmikrofon näher an der Zielquelle ist und dem Mittelwert \bar{X} aus den Eingangssignalen X_L und X_R der beiden Referenzmikrofone. Da der Beamformer das ZS praktisch ungefiltert passieren lässt und die HG reduziert, können mithilfe von Formel 1 Annahmen getroffen werden welche Zeitfrequenzanteile zum ZS und welche zum HG gehören. Es ergeben sich verschiedene Anwendungsmöglichkeiten für die binäre Maske:

1. Beim SBB werden die Zeitfrequenzanteile des ZS durch den Beamformer-Ausgang ersetzt und die des HG durch das um den konstanten Faktor α gedämpfte Eingangssignal. Durch die gleiche Dämpfung beider Kanäle um α bleiben die binauralen cues des HG erhalten. Für α wurde der Wert 0,3 verwendet.
2. Eine andere Möglichkeit ist die binäre Maske als Filterfunktion direkt auf die Eingangssignale der Referenzmikrofone anzuwenden. Zeitfrequenzanteile des ZS passieren ungefiltert und Anteile des HG werden um den Faktor α gedämpft. Der Beamformer dient dabei nur zur Berechnung der binären Maske. Die gleiche Filterung beider Referenzkanäle garantiert die Erhaltung der binauralen cues. Diese Anwendung ist ein eigenständiger Algorithmus und wird als "selective attenuator" (SA) bezeichnet.
3. Ein dritter Algorithmus ist der "Time Difference of Arrival Simulator" (TS) welcher wie der SBB das Beamformerausgangssignal mit dem Eingangssignal mischt. Hierbei wird allerdings nur der zur Berechnung der binären Maske verwendete Ausgangskanal Z verwendet und diesem eine Richtung aufgeprägt. Dies geschieht über eine Verdopplung des Signals zu 2 Kanälen und einer Aufprägung eines Phasenunterschieds zwischen den Kanälen basierend auf der "Time Difference of Arrival" zwischen 2 Mikrofonen im Freifeld. Dieser Algorithmus erlaubt die Kontrolle über die binauralen cues des ZS.

Evaluation

In der Evaluation wurden die Algorithmen mit einem Algorithmus von Lotter et al. (Lot06) verglichen [4]. Dieser ist sehr ähnlich zum SA aber unterscheidet sich in der Filterfunktion. Als Testsignale wurde eine Cafeteria-Szene bestehend aus zwei Sprechern und Hintergrundgeräuschen simuliert. Beide Sprecher sprachen gleichzeitig mit gleichem Pegel und befanden sich in 0° und -30° Richtung gegenüber des Hörgeräteträgers. Das Hinter-

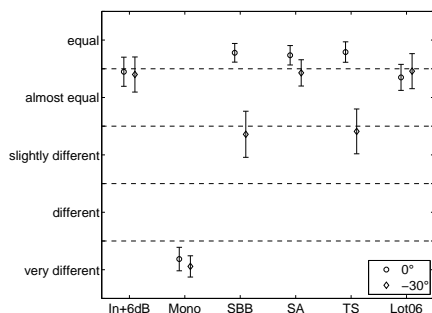


Abbildung 2: Ergebnisse des subjektiven Vergleichstests zur Erhaltung der binauralen cues.

grundgeräusch war eine Audioaufnahme aus einer Cafeteria und im Pegel 6 dB höher als die Sprecher. Ziel war es, jeweils einen der beiden Sprecher zu verbessern.

Für die objektive Evaluation wurde das Signal-to-Interfer+Noise-Ratio Enhancement (SINRE) berechnet. Es handelt sich dabei um die Differenz des SINR zwischen Ein- und Ausgangssignal in dB. Das SINR ist das Verhältnis der Leistung des Zielsprechers geteilt durch die Leistung des Störsprechers und der Hintergrundgeräusche. In Tabelle 1 sind die SINRE Werte für die verschiedenen Algorithmen und Zielsprecher dargestellt. Es ist erkennbar, dass SBB und TS zu höheren Werten führen als der SA. Das liegt daran, dass beide Algorithmen den Beamformerausgang in das Gesamtausgangssignal integrieren. In Zeitfrequenzanteilen die dem ZS zugeordnet werden ist somit das Rauschen reduziert. Beim SA bleiben diese Anteile ungefiltert was das Rauschen nicht reduziert. Alle drei Algorithmen zeigen höhere SINRE-Werte mit dem Zielsprecher auf 30° während Lot06 sich umgekehrt verhält, was an der unterschiedlichen Filterfunktion liegt.

Tabelle 1: SINRE Werte der einzelnen Algorithmen für beide getesteten Zielrichtungen (ZR)

ZR	SBB	SA	TS	Lot06
0°	2.47 dB	2.22 dB	2.38 dB	2.66 dB
-30°	2.97 dB	2.47 dB	3.86 dB	2.24 dB

Zusätzlich wurde zur subjektiven Bewertung ein Vergleichstest zwischen dem Eingangssignal und den Ausgangssignalen der Algorithmen mit acht normalhörenden Probanden durchgeführt. In einem Durchlauf des Experimentes wurde gefragt, ob die Position der Schallquellen erhalten bleibt. Abb. 2 zeigt die Ergebnisse. Eines der Testsignale ist das Eingangssignal mit um 6 dB im Pegel erhöhten Zielsprecher. Obwohl die räumliche Wiedergabe gleich ist zeigt sich, dass die Probanden die Verbesserung des Zielsprechers schon als veränderte räumliche Darbietung empfunden haben. Zum Vergleich wurde ein Mono-Signal, welches keine räumliche Information enthält, als sehr unterschiedlich wahrgenommen. Alle Algorithmen führen zu einer guten Erhaltung der binauralen cues mit dem Zielsprecher auf 0° . Bei -30° zeigt sich jedoch eine veränderte Darbietung beim SBB und TS. Dies ist auf die Integration des Beamformerausgangssignals in das Gesamtausgangssignal zurückzuführen. Da im Beamfor-

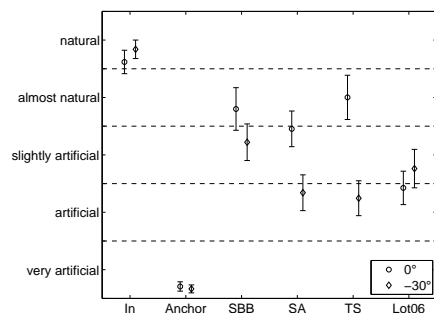


Abbildung 3: Ergebnisse des Vergleichstests zur Natürlichkeit der Ausgangssignale.

merausgang noch Anteile vom Hintergrundgeräusch und Störsprecher enthalten sind, kommt es insgesamt zu einer räumlichen Verschiebung in Richtung des Zielsprechers.

In einem zweiten Durchlauf des Experimentes bewerteten die Probanden die Natürlichkeit der Signale. Abb. 3 zeigt die Ergebnisse. Ein Testsignal ist das Eingangssignal selbst, was als natürlich empfunden wurde. Zum Vergleich wurde ein artefaktreiches Signal mithilfe einer sehr stark dämpfenden binären Maske erzeugt, welches als sehr künstlich empfunden wurde. Die Ausgangssignale der Algorithmen erreichen nicht die Natürlichkeit des Eingangssignals. Der Grund dafür liegt wahrscheinlich darin, dass weder für die binäre Maske noch für die Filterfunktion von Lot06 eine Glättung verwendet wurde. Dies kann zu starken Unterschieden in der Filterung zwischen benachbarten Zeitfrequenzanteilen führen.

Zusammenfassung und Ausblick

Es wurden drei beamformer-basierte Algorithmen mit binärer Maske zur Störgeräuschreduktion und Erhaltung der räumlichen Wahrnehmung entwickelt. Die binäre Maske dient dabei zur Unterscheidung zwischen Ziel- und Störsignal. Insgesamt sind Beamformer und binäre Maske die entscheidenden Einflussfaktoren auf die Performance der Algorithmen. Die zukünftige Forschung richtet sich daher auf die Optimierung der binären Maske und des Beamformers.

Literatur

- [1] Bronkhorst, A. W.: The cocktail party phenomenon: A review of research on speech intelligibility in multiple-talker conditions. *Acta Acustica united with Acustica* 86(1) (1990), 117-128
- [2] Peissig, J. et al.: Directivity of binaural noise reduction in spatial multiple noise-source arrangements for normal and impaired listeners. *The Journal of Acoustical Society of America* 101(3) (1997), 1660-1670
- [3] Doclo, S. et al.: Acoustic beamforming for hearing aid applications. *Handbook on Array Processing and Sensor Networks* (2008), 269-302
- [4] Lotter, T. et al.: Dual-channel speech enhancement by superdirective beamforming. *EURASIP Journal on Applied Signal Processing*, (2006), 1-14