

## Subjektive Präferenzen eines Stereo-Vollband-Freisprechsystems

Jens Friedrich<sup>1,2</sup>, Arthur Wolf<sup>2</sup>, Klaus Linhard<sup>2</sup>, Stephan Senkbeil<sup>1</sup>, Gerhard Schmidt<sup>1</sup>, Harald Schnepf<sup>2</sup>

<sup>1</sup> Christian Albrechts Universität zu Kiel, 24143 Kiel, jefr/stse/gus@tf.uni-kiel.de

<sup>2</sup> Daimler AG, 89081 Ulm, E-Mail: arthur.wolf/klaus.linhard/harald.schnepf@daimler.com

### Einleitung

Die Mehrzahl der heutigen Freisprechsysteme überträgt ein einkanaliges Monosignal mit einer beschränkten akustischen Bandbreite. Das flächendeckend ausgebaute und breitbandige mobile Datennetzwerk würde jedoch eine neuartige Voice-over-IP (VoIP) Technik mit leistungsfähigeren Codecs unterstützen. Das Ziel dieser Arbeit ist die Untersuchung subjektiver Präferenzen eines prototypischen Stereo-Vollband-Freisprechsystems in einem Fahrzeug. Die Untersuchung zeigt verschiedene Nutzerpräferenzen im Hinblick auf die räumliche Audiowiedergabe in einem Fahrzeuginnenraum und die Audiobandbreite eines Telefonsignals.

Zuerst werden die theoretischen Grundlagen erläutert. Dies beinhaltet die Übersicht über die zwei untersuchten Audiobandbreiten für die mobile Telefonie. Weiterhin werden zwei Stereophoniemethoden und ein Wiener-Filter-Ansatz für das Modul der Geräuschreduktion erläutert. Ebenfalls wird auf die Auswertungsmethode des vergleichenden Hörtests eingegangen. Im Hauptteil wird die Evaluierung drei verschiedener Untersuchungen detailliert dargestellt. Diese drei Untersuchungen zeigen die Einflüsse der Geräuschreduktion auf das von Nutzern bevorzugte Wiedergabeformat und auf die bevorzugte Audiobandbreite sowie die von Nutzern präferierte Stereophoniemethode auf. Abschließend werden die Ergebnisse zusammengefasst und daraus ein Fazit abgeleitet.

### Grundlagen

Die Audiobandbreite ist ein wichtiges Kriterium, um die empfundene Qualität eines Telefonates zu beschreiben. Daher versuchen die Telekommunikationsunternehmen diese zu erweitern. Der heutige Standard ist ein bandbegrenzt Signal zwischen 300 Hz und 3,4 kHz, das sogenannte Schmalbandsignal (engl. *narrowband*, kurz: NB). Im Vergleich zum menschlichen Hörspektrum von 30 Hz bis 20 kHz stellt das Schmalband nur einen kleinen Ausschnitt dar [4]. Um diese Bandbegrenzung zu erweitern, wurde z.B. ein neuer VoIP-basierter Codec, der sog. Opus-Codec entwickelt. Dieser unterstützt neben der zweikanaligen Stereoübertragung auch höhere Audiobandbreiten bis zu 24 kHz, dem sog. Vollbandsignal (engl. *fullband*, kurz: FB) [2].

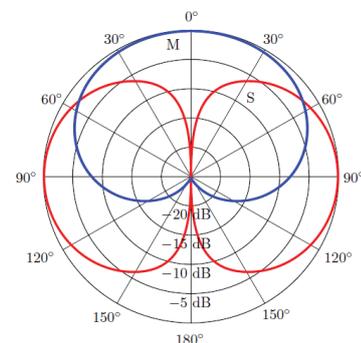
In einem Fahrzeug liegt der Fokus der Telefonie auf dem Freisprechsystem. In Fahrzeugen der Mittel- und Oberklasse stehen heute für die Freisprechtelefonie meist zwei Mikrophone und für die Wiedergabe die Lautsprecher des Audiosystems zur Verfügung. Dieses Setup ist die Basis für eine Untersuchung eines prototypischen Stereo-Vollband-Freisprechsystems.

Die Untersuchung dieser Arbeit zeigt die präferierte Audiobandbreite und Wiedergabeformate für Freisprechanwendungen in einem Fahrzeug. Hierzu wurden Szenarien definiert, die typische Telefonumgebungen darstellen. Die Empfangsseite ist immer ein Fahrzeuginnenraum. Die Sendeseite unterscheidet sich in fünf verschiedene Umgebungen:

1. Fahrzeug ↔ Fahrzeug
2. Fahrzeug ↔ Büro
3. Fahrzeug ↔ Konferenzraum
4. Fahrzeug ↔ Straße
5. Fahrzeug ↔ Kantine.

In jedem Verbindungsszenario sprechen ein oder mehrere Personen auf der Sendeseite und die Wiedergabe erfolgt im Fahrzeug auf der Empfangsseite. Die Probanden dieser Untersuchung bewerten die Audiowiedergabe im Fahrzeug. Aus praktischen Erwägungen wurde die Audiowiedergabe durch die entsprechende binaurale Kopfhörer-Wiedergabe ersetzt.

Das erste Szenario stellt ein Telefonat aus einem Fahrzeug mit Freisprechsystem zu einem anderen Fahrzeug mit Freisprechsystem dar. Im Fahrzeug auf der Sendeseite befinden sich zwei Personen, Fahrer und Beifahrer. Im zweiten Szenario spricht nur eine Person in einem ruhigen Büroraum. Das dritte Szenario simuliert ein Gespräch dreier Personen in einem Konferenzraum, die aus unterschiedlichen Richtungen zum mittig platzierten Stereo-Mikrophon sprechen. Im vierten Szenario steht die sprechende Person an einer befahrenen Straße. Im fünften Szenario spricht eine Person in einer vollbesetzten Kantine.

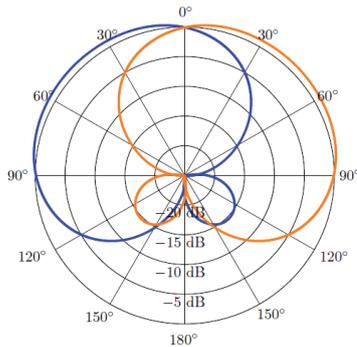


**Abbildung 1:** Richtdiagramm zweier Mikrophone (M: Nierencharakteristik (blau), S: Acht-Charakteristik (rot)) in MS-Anordnung.

Die Stereoaufnahmen für diese Untersuchungen wurden im Szenario Fahrzeug mit zwei Mikrophenen im Dachhimmel

eines Fahrzeuges im Abstand von jeweils 20 cm über dem Fahrer und dem Beifahrer platziert. Für die vier anderen Szenarien wurden die Aufnahmen simultan mit einem MS-Mikrophon (MS) und zwei omnidirektionalen Mikrophonen (Abstand der Kapseln 6 cm) durchgeführt.

Das MS-Mikrophon besteht aus einer Kapsel mit Nierencharakteristik (Mittenkanal M) und einer Kapsel mit Acht-Charakteristik (Seitenkanal S), wie in Abbildung 1 dargestellt. Durch gewichtete Addition der beiden Kanäle lässt sich eine effektive Charakteristik der Stereoaufnahme erzeugen (siehe Abbildung 2) die für die zweikanalige Stereowiedergabe geeignet ist [3].



**Abbildung 2:** Effektive Richtcharakteristik der MS-Stereophonie durch Summen- und Differenzbildung der beiden Kanäle.

In der Signalverarbeitungskette wurde zum Zwecke dieser Untersuchung das Modul der Geräuschreduktion näher betrachtet. Da das Umgebungsgeräusch am Mikrophon nicht isoliert betrachtet werden kann muss aus dem Gesamtsignal das Geräusch geschätzt und anschließend reduziert werden. Da durch eine ungenaue Schätzung des Geräusches tonale Reststörungen auftreten können, beinhaltet die Kennlinie der Geräuschreduktion aus dem Wiener-Filteransatz die Stellgröße Maximaldämpfung  $H_{min}$ . Die Filterkennlinie der Geräuschreduktion berechnet sich daraus zu:

$$\hat{H}(e^{j\Omega_\mu}, k) = \max \left\{ H_{min}, 1 - \frac{K_{over} \cdot \hat{S}_{bb}(e^{j\Omega_\mu}, k)}{\hat{S}_{xx}(e^{j\Omega_\mu}, k)} \right\} \quad (2)$$

Die Maximaldämpfung  $H_{min}$  setzt der Dämpfung des Geräusches eine untere Grenze. Dies verhindert eine komplette Unterdrückung des Hintergrundgeräusches und belässt ein Restgeräusch im Signal. Dies hat den Effekt, dass das Sprachsignal natürlicher erhalten bleibt und dass Sprachartefakte durch das Restgeräusch maskiert werden.

Auf weitere Bezeichnungen der Zwischengrößen, sowie der Herleitung der Filterkennlinie aus Gleichung (2) wird hier verzichtet und auf [1] verwiesen. Es wird für die Untersuchung dieser Arbeit o.g. Standardverfahren der Geräuschreduktion verwendet, wobei im weiteren Verlauf der Untersuchung nur die Maximaldämpfung  $H_{min}$  eine Rolle spielt.

Den insgesamt 15 Probanden wurden in dem CMOS-Test jeweils zwei Audiosignale A und B vorgespielt, die

vergleichend bewertet werden sollten. Um neben der absoluten Präferenz für eines der beiden Signale auch eine Gewichtung des als besser bewerteten Signals zu bekommen wurde die Bewertungsskala abgestuft. Die Bewertung erfolgt anhand einer siebenstufigen Skala mit Werten zwischen -3 und +3. Die einzelnen Stufen sind in Tabelle 1 dargestellt.

**Tabelle 1:** Siebenstufige Bewertungsskala samt deren Gewichtungswerten des durchgeführten CMOS-Tests

Bewertung	Gewichtung
A ist viel besser als B	-3
A ist besser als B	-2
A ist etwas besser als B	-1
A und B sind nahezu gleich	0
A ist etwas schlechter als B	+1
A ist schlechter als B	+2
A ist viel schlechter als B	+3

Aus diesen Gewichtungswerten wurden drei verschiedene Mittelwerte errechnet. Der Gesamtmittelwert einer Untersuchung  $V_{Avg}$  (siehe Abbildungen 3-7: rote Linie) errechnet sich aus der Summe über alle Abstimmungswerte aller Probanden  $N_{Prob}$ :

$$V_{Avg} = \frac{1}{N_{Prob} \cdot N_t} \sum_{m=0}^{N_{Prob}-1} \sum_{k=0}^{N_t-1} V_m(k) \quad (1)$$

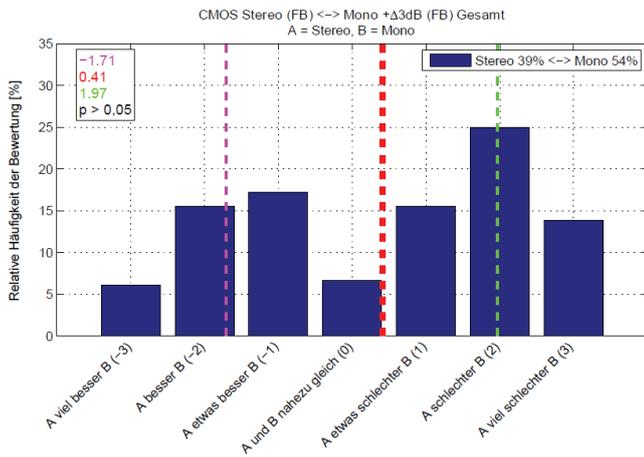
Jeder Proband hat  $N_t$  verschiedene Vergleiche pro Untersuchungsgegenstand bewertet.  $V_m(k)$  repräsentiert den Abstimmungswert des Probanden  $m$  für den Vergleich  $k$ .

Ergänzend wurden in den Abbildungen der Ergebnisse noch die beiden Merkmalsmittelwerte  $M_A$  und  $M_B$  dargestellt. Diese stellen die gemittelten Abstimmungsergebnisse für die einzelnen Merkmale A und B getrennt dar. Sie errechnen sich aus der Anzahl und der Gewichtung der für jeweils ein Merkmal positiven Abstimmungen (für  $M_A$  Gewichtungen -3, -2, -1; für  $M_B$  Gewichtungen +3, +2, +1) und sind in den Abbildungen 3-7 in Magenta ( $M_A$ ) und Grün ( $M_B$ ) dargestellt.

Weiterhin wird in der Legende der Abbildungen 3-7 ein Wahrscheinlichkeitswert  $p$  angegeben der das Signifikanzniveau der Verteilung beschreibt. Hierzu wurde ein Einstichproben-t-Test durchgeführt und die t-Werte mit der Student'schen t-Verteilung verglichen.

### Ergebnisse

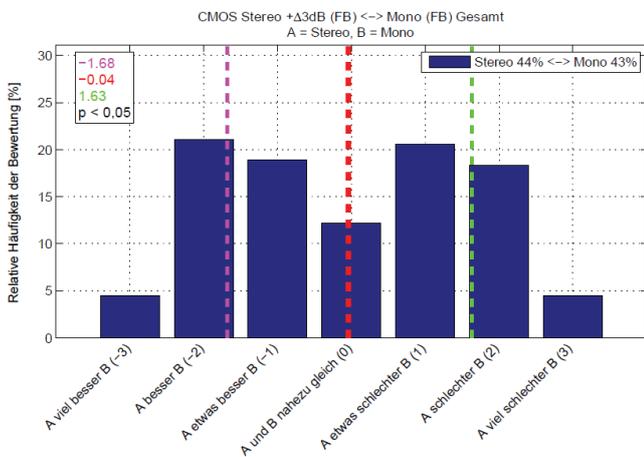
Für den Einfluss der Geräuschreduktion auf das bevorzugte Wiedergabeformat soll beurteilt werden, ob die Stärke der Maximaldämpfung einen Präferenzunterschied in der räumlichen Wiedergabe der verschiedenen Szenarien verursacht.



**Abbildung 3:** CMOS Gesamtergebnis für das präferierte Wiedergabeformat bei +Δ3 dB stärkerer Maximaldämpfung der Geräuschreduktion für das Monosignal gegenüber dem Stereosignal. Der Wert und die Linie in Rot repräsentieren den Gesamtmittelwert  $V_{Avg}$ . Der Wert und die Linie in Magenta zeigen den Merkmalsmittelwert  $M_A$  und der Wert und die Linie in Grün den Merkmalsmittelwert  $M_B$ .

Verglichen wurden jeweils ein Monosignal und ein Stereosignal. Die Geräuschreduktionsstufe der Maximaldämpfung für die beiden Signale variierten in dieser Untersuchung in einem relativen Unterschied von  $H_{min} = +\Delta 3$  dB. Dies bedeutet, dass entweder das Geräusch des Monosignals relativ um +Δ3 dB (siehe Abbildung 3) oder das des Stereosignals um +Δ3 dB (siehe Abbildung 4) stärker gedämpft wurde. So standen beispielsweise in Abbildung 3 Signalpaare mit einer absoluten Maximaldämpfung von -6 dB (Stereo) und -9 dB (Mono) sowie -3 dB (Stereo) und -6 dB (Mono) gegenüber. In Abbildung 4 gilt dies entsprechend umgekehrt.

Die Gesamtergebnisse über alle Szenarien gemittelt zeigen nur einen geringen Einfluss der Maximaldämpfung auf die Präferenz des Wiedergabeformates.

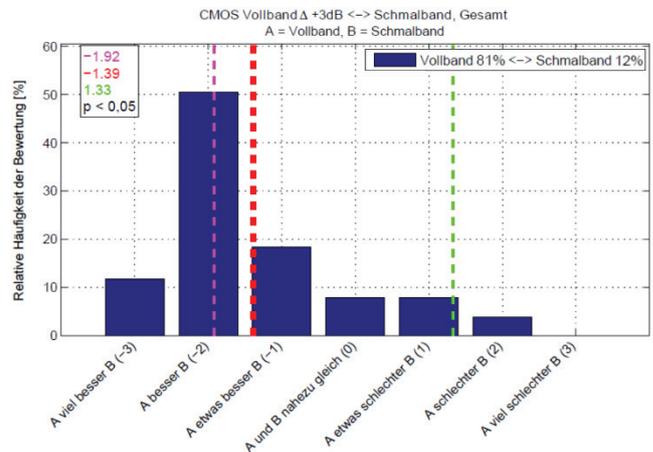


**Abbildung 4:** CMOS Gesamtergebnis für das präferierte Wiedergabeformat bei +Δ3 dB stärkerer Maximaldämpfung der Geräuschreduktion für das Stereosignal gegenüber dem Monosignal.

Das Signal mit relativ stärkerer Maximaldämpfung wird grundsätzlich etwas besser bewertet als das Signal mit relativ höherem Restgeräuschpegel. Die Präferenz für ein Wiedergabeformat ist indifferent, was in beiden Abbildungen 3 und 4 an den relativ geringen Abstimmungen für die neutrale Wertung „A und B sind nahezu gleich“ zu erkennen ist.

Für den Einfluss der Geräuschreduktion auf die bevorzugte Audiobandbreite wurde untersucht, wie sich die Präferenz für eine der beiden dargestellten Audiobandbreiten in Abhängigkeit der Maximaldämpfung ändert. Es wurde jeweils ein Schmalbandsignal und ein Vollbandsignal verglichen.

Die Geräuschreduktionsstufe der Maximaldämpfung betrug auch hier relativ zueinander +Δ3 dB. Entweder wurde das Geräusch des Schmalbandsignals relativ um +Δ3 dB (siehe Abbildung 5) oder das des Vollbandsignals um +Δ3 dB (siehe Abbildung 6) stärker gedämpft.

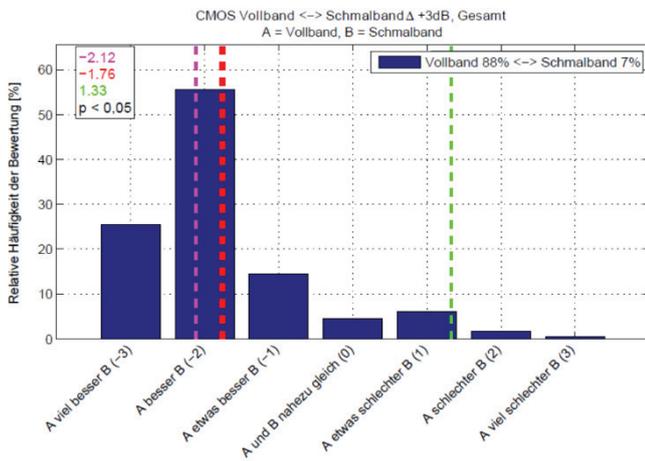


**Abbildung 5:** CMOS Gesamtergebnis für die präferierte Bandbreite bei +Δ3 dB stärkerer Maximaldämpfung der Geräuschreduktion für das Vollbandsignal gegenüber dem Schmalbandsignal.

Die Ergebnisse dieser Untersuchung zeigen auch hier nur einen geringen Einfluss der Geräuschreduktion auf die bevorzugte Audiobandbreite.

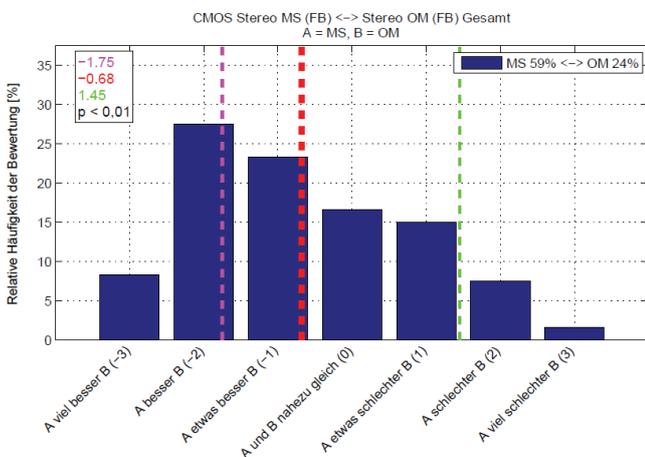
Das Vollbandsignal wird unabhängig von der Maximaldämpfung der Geräuschreduktion in beiden Bewertungen stärker präferiert. Die Geräuschreduktion hat auch hier nur geringen Einfluss auf die Präferenz.

Für den Vergleich der Bandbreite wird das Vollbandsignal mit einer relativ geringeren Geräuschreduktionsstufe mit  $V_{Avg} = -1.76$  gegenüber dem Schmalbandsignal bewertet. Das Vollbandsignal mit relativ höherer Stufe wird mit  $V_{Avg} = -1.39$  gegenüber dem Schmalbandsignal bewertet. Hier wird die relativ geringere Geräuschreduktionsstufe insgesamt besser bewertet und somit wird ein erhöhter Restgeräuschpegel in der Wiedergabe des Vollbandsignals von den Probanden positiv gewertet.



**Abbildung 6:** CMOS Gesamtergebnis für die präferierte Bandbreite bei  $+3\text{dB}$  stärkerer Maximaldämpfung der Geräuschreduktion für das Schmalbandsignal gegenüber dem Vollbandsignal.

Die Untersuchung nach der bevorzugten Stereophoniemethode soll zeigen, welche Stereo-Aufnahmetechnik bei der räumlichen Wiedergabe im Fahrzeug bevorzugt wird. Von den Probanden wurden jeweils eine Aufnahme mit einem Mitte-Seite-Mikrofon (MS) und eine der omnidirektionalen Mikrophone (OM) verglichen. Beide Aufnahmen waren Vollbandaufnahmen (FB). Die Ergebnisse sind in Abbildung 7 dargestellt.



**Abbildung 7:** CMOS Gesamtergebnis für die präferierte Stereophoniemethode MS gegenüber OM.

Die Ergebnisse zeigen ein Übergewicht ( $V_{Avg} = -0.68$ ) für die MS-Methode. Diese Aufnahmetechnik zeichnet sich gegenüber der OM-Methode durch eine bessere Auflösung der Stereobreite aus. Dies ist bedingt durch die Richtwirkung des Mittenkanals (M) zum Sprecher hin und der räumlichen Aufnahme der Umgebungsgereusche des orthogonal positionierten Acht-Mikrophons des Seitenkanals (S).

## Zusammenfassung

In der vorliegenden Studie wurden die Einflüsse der Maximaldämpfung der Geräuschreduktion auf das subjektive Nutzerempfinden im Hinblick auf

Wiedergabeformat und Audiobandbreite sowie von Probanden bevorzugte Stereophonietechniken für Freisprechanwendungen untersucht. Hierbei zeigt sich, dass die Stärke der Maximaldämpfung auf das bevorzugte Wiedergabeformat nur einen geringen Einfluss hat. Eine stärkere Reduktion des Hintergrundgeräusches wird hier sowohl für die Stereo- als auch Monowiedergabe positiv bewertet. Für die bevorzugte Audiobandbreite tritt ein gegenteiliger Effekt ein. Neben der grundsätzlichen Präferenz für ein Vollbandsignal gegenüber dem Schmalbandsignal wird jeweils das Signal mit einer schwächeren Stufe der Maximaldämpfung und somit einem höheren Anteil an Restgeräusch leicht besser bewertet. Für die bevorzugte Stereophoniemethode konnte gezeigt werden, dass die MS-Stereophonietechnik für die räumliche Wiedergabe eines Stereo-Telefonsignals am fernen Ende des Telefonates gegenüber der Aufnahme mit zwei omnidirektionalen Mikrofonen besser bewertet wird.

Wir danken der Firma QNX für die freundliche Unterstützung und Diskussion.

## Literatur

- [1] Hänslers, E. und Schmidt, G.: Acoustic Echo and Noise Control: A Practical Approach, Wiley-IEEE Press, 2004
- [2] Valin, J.-M., Maxwell, G., Terriberry, T. B. und Vos, K.: High-Quality, Low-Delay Music Coding in the Opus Codec. Audio Engineering Society Convention 135 (2013)
- [3] Weinzierl, S.: Handbuch der Audiotechnik, Springer, 2008
- [4] Blauert, J. und Braasch, J.: Räumliches Hören. Handbuch der Audiotechnik, Springer, 2008