Sprachmaskierung im Fahrzeuginnenraum

Jan Rennies[1], Lena Schell-Majoor[1], Andreas Volgenandt[1],
Christian Volkmar[2], Niklas Schmincke[3], Stefan Behr[4]

[1]*Fraunhofer IDMT, Hör-, Sprach- und Audiotechnologie, Oldenburg, E-Mail:jan.rennies@idmt.fraunhofer.de*
[2]*IAV GmbH, München*
[3]*Volke Consulting Engineers GmbH & Co. Planungs KG, München*
[4]*BMW Group, München*

Einleitung

In manchen Anwendungen ist es wünschenswert eine möglichst unverständliche Übertragung von Sprache zwischen Fahrzeuginsassen innerhalb eines Kfz sicher zu stellen. Ein wirksames Mittel hierfür kann das Abspielen von Maskiersignalen sein. Dabei ist jedoch zu beachten, dass neben der gewünschten Maskierwirkung auch Akzeptanz-probleme auftreten können, beispielsweise wenn das Maskiersignal zu laut ist, unangenehm klingt und über längere Zeiträume aktiviert ist. In dieser Studie wurden daher 16 verschiedene Signale mit unterschiedlicher zeitlicher und spektraler Struktur hinsichtlich ihrer Maskierwirkung und ihrer Lästigkeit bzw. ihres Störcharakters in realen Hörsituationen aus dem Fahrzeug-innenraum untersucht. Zur Quantifizierung der Sprachverständlichkeit wurden psychometrische Funktionen des Oldenburger Satztests für alle Störgeräusche ermittelt. Anschließend wurde die Lästigkeit von gleichmäßig maskierenden Signalen mittels eines Skalierungsverfahrens bewertet.

Akustisches Szenario

In allen Messungen dieser Studie wurde das folgende Szenario betrachtet: Der Zielsprecher befand sich beifahrerseitig im Fond des Fahrzeugs, d.h. schräg rechts hinter dem Fahrer, welcher den Zuhörer repräsentierte. Die Maskiersignale wurden über kopfnahe Lautsprecher beim Fahrer abgespielt mit dem Ziel, die Sprache des Zielsprechers zu verdecken. Für alle Übertragungswege wurden binaurale kopfbezogene Übertragungsfunktionen in einem SUV gemessen. Die Klangdarbietung in dieser Studie erfolgte anschließend über Sennheiser HD 650 Kopfhörer in einer schallisolierten Hörkabine.

Experiment 1: Messung der Maskierwirkung

Im ersten Experiment wurde die Maskierwirkung verschiedener Signale bei 15 normalhörenden Probanden gemessen.

Stimuli

Als Sprache diente der Oldenburger Satztest [1]. Dieser besteht aus Fünfwortsätzen mit der festen syntaktischen Struktur *Name Verb Zahlwort Adjektiv Objekt* (z.B. Peter hat fünf nasse Sessel), wobei für jedes Wort zehn Alternativen zur Verfügung stehen. Die Alternativen können beliebig kombiniert werden, so dass die Sätze quasi kontextlos sind, d.h., es ist den Probanden nicht möglich von einem verstandenen Wort auf das nächste zu schließen. Ebenso wenig können Sätze eingeprägt werden, so dass der Test beliebig oft wiederholt werden kann. Für diese Art Satztest ist eine Gewöhnung der Probanden an das Sprachmaterial notwendig. Für jede Messkondition wurde eine zufällig ausgewählte Liste von 20 Sätzen verwendet. Für die Gewöhnung wurden drei Listen verwendet, die nicht in die Auswertung einfließen. Sprache wurde bei einem festen Pegel von 59,8 dB SPL (wie alle hier berichteten Pegel und SNR bezogen auf das rechte Ohr des Zuhörers) dargeboten. Dies entspricht einem für diesen Übertragungsweg zu erwartenden Sprachpegel unter Berücksichtigung der Tatsache, dass der Sprecher in realen Fahrsituationen aufgrund des Lombard-Effekts mit erhöhtem Stimmaufwand sprechen würde [2].

Als Maskiersignale dienten elf Maskierer, die sich in ihren zeitlichen und spektralen Eigenschaften unterschieden und in Tabelle 1 zusammen gefasst sind. Diese umfassten stationäre Maskierer mit drei unterschiedlichen sprachgefärbten (speech-shaped noise, *ssn*) Langzeitspektren, eine unverständliche Überlagerung von zehn simultanen Sprechern (MultiTalkerBabble, *mtb*) mit denselben drei Langzeitspektren, drei sehr stark dynamische Maskierer, die adaptiv aus der gesprochenen Sprache generiert wurden und verschiedenen Formen von unverständlichen Echos bzw. Überlagerungen der Sprache mit sich selbst entsprechen (*babble_einfach, vocal_delay, babble_mehrfach*). Zusätzlich wurden zwei erkennbare Naturgeräusche verwendet (Wasserfall und Flussrauschen).

Neben diesen isolierten Maskierern wurden fünf Maskiererkombinationen verwendet, wobei die einzelnen Maskiersignale so zusammen gemischt wurden, dass jeder einzelne Maskierer klar erkennbar blieb, so dass eine verändert klangliche Assoziation erfolgte als bei den isolierten Maskierern.

Methode

Nach jedem dargebotenen Satz wiederholte der Proband die verstandenen Wörter. Mit dem adaptiven Verfahren des Oldenburger Satztests [1] wurden durch systematisches Verändern des Maskierpegels die beiden Punkte der psychometrischen Funktion bestimmt, die einer Wortverständlichkeit von 20% bzw. 80% entsprechen. Dadurch konnte für jeden Maskierer und Probanden die psychometrische Funktion durch Fitten einer sigmoiden Funktion ermittelt werden, so dass die zu erwartende Sprachverständlichkeit als Funktion des Maskierpegels berechnet werden konnte. Die Reihenfolge der Maskierer und Messpunkte wurde für jeden Probanden randomisiert.

Tabelle 1: Übersicht über die verwendeten Maskiersignale

Maskierer	Beschreibung
ssnts	Stationäres sprachgefärbtes Rauschen, gleiches Langzeitspektrum wie Zielsprecher
ssnltas	Stationäres sprachgefärbtes Rauschen, Long-term average speech spectrum (LTASS) wie in [3]
ssnsii	Stationäres sprachgefärbtes Rauschen, Spektrum gewichtet gemäß SII [4], d.h. verstärkte Maskierung wichtiger Sprachfrequenzkomponenten
mtbts	Multitalker babble (N=10), gleiches Langzeitspektrum wie Zielsprecher
mtbltas	Multitalker babble (N=10), LTASS wie in [3]
mtbsii	Multitalker babble (N=10), Spektrum gewichtet gemäß SII [4]
babble_ einfach	Stark zeitlich adaptiv, Babble aus Zielsprache
vocalDelay	Stark zeitlich adaptiv, verzögerte Zielsprache
babble_ mehrfach	Stark zeitlich adaptiv, Babble aus Zielsprache, mehrfach überlagert
waterfall	Wasserfallgeräusch von freesounds.org
river	Flussgeräusch von freesounds.org

Ergebnisse

Abbildung 1 veranschaulicht die für den Maskierer *ssnsii* ermittelten individuellen (grau) und parametrisch gemittelten (schwarz) psychometrischen Funktionen, d.h. den Zusammenhang zwischen Maskierpegel und Sprachverständlichkeit. Wie erwartet fällt die Verständlichkeit deutlich mit steigendem Maskierpegel. Bei ca. 68 dB SPL (entspricht einem SNR von ca. -8 dB) wird bei diesem Maskierer im Mittel noch jedes zweite Wort verstanden. Ab ca. 72 dB SPL kann von einer hinreichend schlechten Verständlichkeit ausgegangen werden.

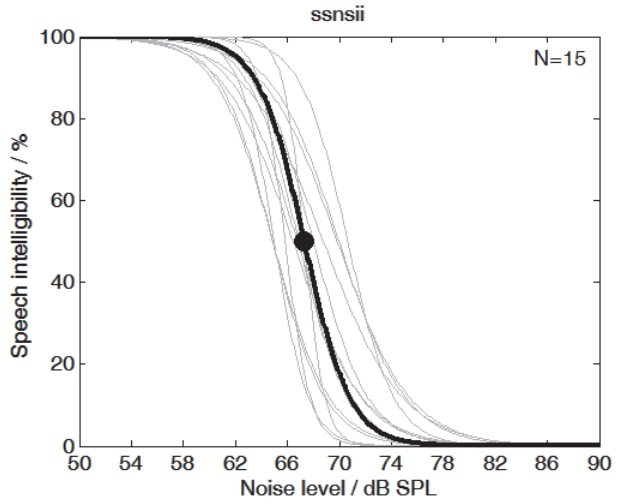

Abbildung 1: Individuelle (grau) und parametrisch gemittelte psychometrische Funktion (schwarz) für den Maskierer *ssnsii*.

Abbildung 2 zeigt die mittleren psychometrischen Funktionen für die elf isolierten Maskiersignale. Es zeigen sich deutliche Unterschiede hinsichtlich der Maskierwirkung der unterschiedlichen Signale. So ist bspw. bei einem Maskierpegel von 80 dB SPL für die *ssn*- und *mtb*-Maskierer kein Sprachverstehen mehr möglich, während bei *river* noch ca. jedes zweite Wort verstanden wird.

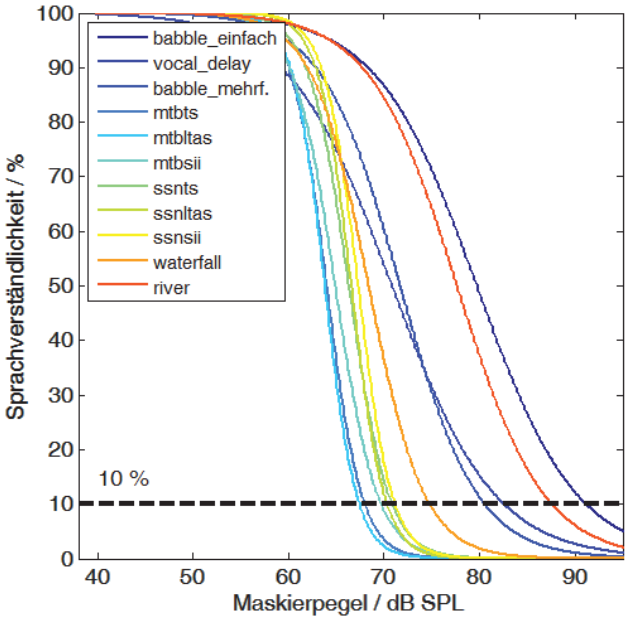

Abbildung 2: Sprachverständlichkeit als Funktion des Maskierpegels für die elf isolierten Maskierer.

Dies verdeutlicht auch die Betrachtung der Maskierpegel, die notwendig sind, um eine bestimmte Maskierwirkung zu erzielen. Die gestrichelte Linie in Abbildung 2 markiert eine feste Wortverständlichkeit von 10%. Abbildung 3 zeigt die Maskierpegel, die notwendig sind, um diese Verständlichkeit zu erzielen. Fehlerbalken veranschaulichen interindividuelle Standardabweichungen um den Mittelwert. Es zeigt sich, dass die *mtb*-Maskierer die beste Maskierwirkung haben, gefolgt von den *ssn*-Maskierern, die einen ca. 3 dB höheren Pegel bei gleicher Verständlichkeit benötigen. Die beiden Naturgeräusche sowie die stark dynamischen Maskierer fallen in der Maskierwirkung deutlich ab und benötigen deutlich höhere Pegel. Beispielhafte Ergebnisse für die Maskiererkombinationen sind in Abbildung 4 dargestellt. Hier zeigt sich, dass bei den gewählten Kombinationen und Mischverhältnissen die Maskierwirkung immer durch den besten Maskierer dominiert wird, d.h. zum einen, dass die Maskierwirkung von *mtbts* durch die Mischung mit anderen Signalen nicht verbessert wird. Zum anderen ist die Mischung von *river* mit *mtb* ein deutlich besserer Maskierer als *river* alleine. Gleiches gilt für *babble_mehrfach.*

Experiment 2: Messung von Lästigkeit und Lautheit

Im zweiten Experiment wurden für ein Subset der in Experiment 1 verwendeten Signale die Lästigkeit und die Lautheit durch 20 normalhörende Probanden bewertet.

Stimuli

Als Sprache dienten erneut Sätze des Oldenburger Satztests. Diese wurden jedoch so geschnitten und zusammen gefügt, dass Sprachfragemente entstanden, die eher einer realen Konversation entsprechen als die feste Satzstruktur des ursprünglichen Tests. Insbesondere wurden Pausen von mehreren Sekunden eingefügt (entspricht längerem Zuhören) sowie einzelne Silben / Wörter (entspricht kurzen Antworten). Insgesamt wurden auf diese Weise Sprachstimuli von 20 s Dauer erzeugt. Der Sprachpegel lag erneut fest bei 59,8 dB SPL.

Als Maskiersignale diente ein Subset der Maskiersignale aus Experiment 1. Als Auswahlkriterium diente zum einen die ermittelte Maskierwirkung. Zum anderen wurden solche Signale ausgewählt, die deutlich unterschiedliche Klangassoziationen wecken. Entsprechend wurde neben den beiden Naturgeräuschen, den stark dynamischen Maskierern und unterschiedlichen Maskiererkombinationen jeweils nur eine Variante von *mtb*- und *ssn*-Maskierer ausgewählt. Alle Maskiersignale wurden bei dem Pegel dargeboten, bei dem gemäß Experiment 1 eine ausreichend schlechte Sprachverständlichkeit von 10% zu erwarten war. Entsprechend der unterschiedlichen Maskierwirkung bedingte dies deutliche Unterschiede im Maskierpegel zwischen 68,1 dB SPL (*mtbts*) und 92,0 dB SPL (*babble_einfach*). Alle Stimuli wurden den Probanden mit einer Dauer von >20 s dargeboten, um näher an realen Gesprächsdauern zu sein als bei typischen Darbietungsdauern von Sprachtests von wenigen Sekunden.

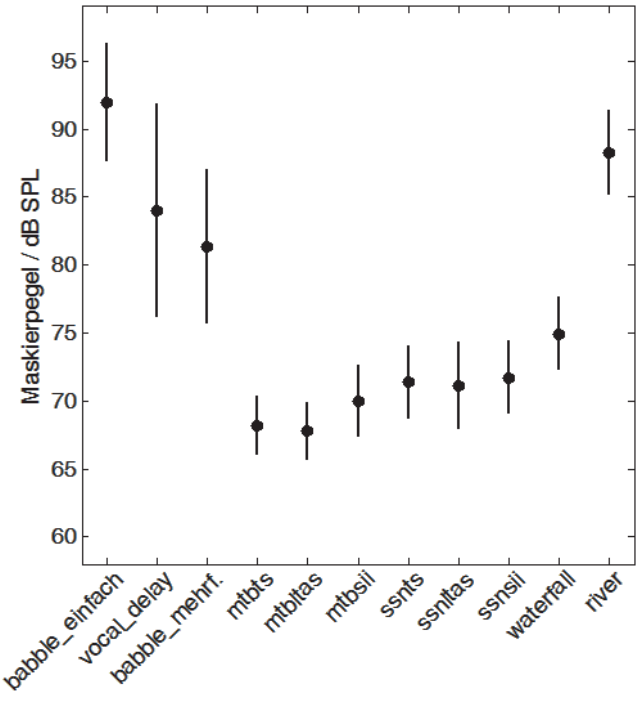

Abbildung 3: Maskierpegel bei einer festen Sprachverständlichkeit von 10% für die isolierten Maskierer. Fehlerbalken repräsentieren Standardabweichungen.

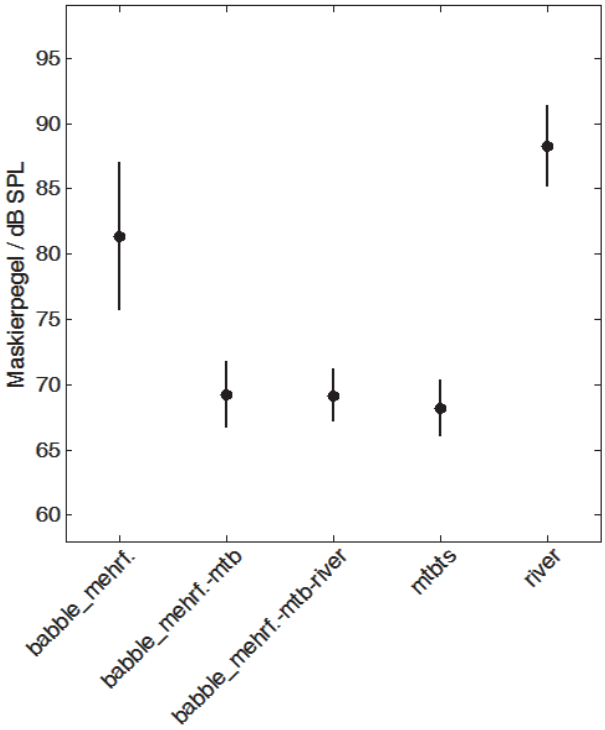

Abbildung 4: Maskierpegel bei einer festen Sprachverständlichkeit von 10% für kombinierte Maskierer. Fehlerbalken repräsentieren Standardabweichungen.

Methode

Zu Beginn wurden die Probanden durch schriftliche Instruktionen mit dem Anwendungsszenario vertraut gemacht und bekamen anschließend 5-sekündige Ausschnitte der später enthaltenen Signale zur Referenzbildung dargeboten. Im eigentlichen Experiment musste jedes Geräusch mindestens einmal in voller Länge gehört werden, bevor eine Bewertung erfolgen konnte. Die Reihenfolge der Maskierer war zufällig. Die Bewertung von Lautheit und Lästigkeit erfolgte anhand einer Skala von 0 („nicht laut / lästig“) bis 50 („sehr laut / lästig“). Die Probanden gaben ihre Bewertungen für Lautheit und Lästigkeit mithilfe zweier Slider entlang der Skalen ein, bevor sie zum nächsten Stimulus fortfahren konnten.

Ergebnisse

Abbildung 5 zeigt die mittleren Bewertungen von Lautheit (blau) und Lästigkeit (orange). Bei Betrachtung der Lautheit fällt auf, dass diese stark zwischen den einzelnen Maskierern variiert, was angesichts der großen Pegelunterschiede nicht verwundert. Die isoliert dargebotenen stark dynamischen Maskierer werden als vergleichsweise laut wie lästig bewertet, während sprachgefärbtes Rauschen und Multitalker-Babble geringe Bewertungen erhalten. Mit zwei Ausnahmen gibt es eine große Übereinstimmung zwischen der mittleren Lautheit und Lästigkeit. Einzig die Bewertungen des Maskieres *river* sowie der Kombination von *river* und ssn weisen relativ eine geringere Lästigkeit als Lautheit auf. Ebenso fällt auf, dass Bewertungen der kombinierten Maskierer, die *ssnts* oder *mtbts* enthalten, sehr ähnlich zu den Bewertungen der isoliert dargebotenen *ssnts*- und *mtbts*-Maskierer sind.

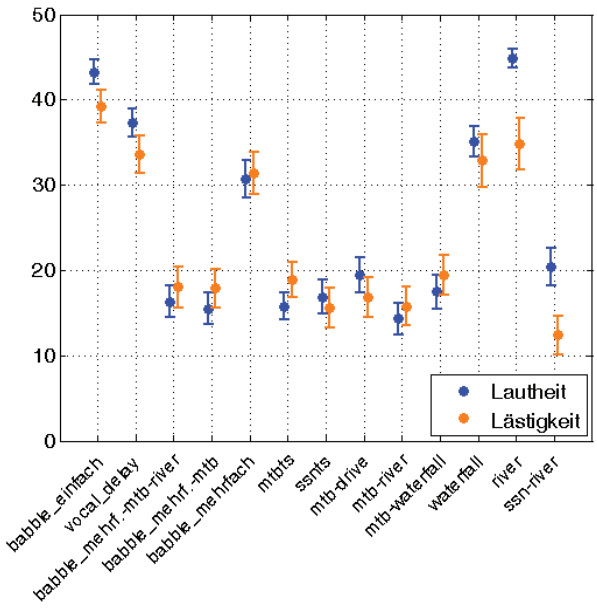

Abbildung 5: Bewertungen der Lautheit (blau) und Lästigkeit (orange) in Experiment 2. Fehlerbalken repräsentieren interindividuelle Standardfehler um den Mittelwert.

Zusammenfassung und Fazit

Die systematische Messung der Sprachverständlichkeit ergab, dass besonders durch Multitalker-Babble und stationäres sprachgefärbtes Rauschen eine gute Maskierwirkung erreicht werden konnte. Hierbei spielte die genaue Färbung des Langzeitspektrums (angepasst an Zielsprecher, LTASS oder SII-gewichtet) eine untergeordnete Rolle. Das Multitalker-Babble hatte im Vergleich zum stationären Sprachrauschen dabei eine noch höhere Maskierwirkung, was mit Ergebnissen aus früheren Studien übereinstimmt [5, 6] und auf die erhöhte Sprachähnlichkeit aufgrund der auftretenden Sprach-modulationen zurückzuführen ist. Die für eine ausreichende Maskierwirkung notwendigen Maskierpegel lagen bei ca. 68 dB SPL, so dass deutlich negative SNR notwendig sind (< -8 dB). Nichtsdestotrotz können diese Pegel auch bei längerer Expositionsdauer als unbedenklich angesehen werden.

Wurden diese Maskierer mit anderen Geräuschen kombiniert (z.B. Wasserfall oder Flussrauschen), so brachte dies keinen Gewinn bzgl. der Maskierwirkung, weckte jedoch andere Assoziationen des Höreindrucks. Diese drückten sich nicht unmittelbar in einer geringeren Lautheit oder Lästigkeit aus, könnten aber in der Praxis verwendet werden, um den Geräuschcharakter über die Zeit zu ändern, um damit bei längerem Hören durch selbstgewählte Signalwechsel ggf. eine erhöhte Akzeptanz beim Hören zu bewirken.

Die schlechte Maskierwirkung sowie hohe empfundene Lästigkeit der stark dynamischen Maskierer deuten darauf hin, dass ein zu schnelles adaptives Anpassen an den Zielsprecher weder im Sinne der Verständlichkeitsreduktion noch der Nutzerakzeptanz vorteilhaft ist.

Für eine praxisnahe Validierung der Ergebnisse dieser Studie ist eine Implementation der Maskiererwiedergabe im realen Fahrzeug für eine Feldstudie notwendig. Diese wird weitere Aufschlüsse darüber geben, ob sich bspw. probandenabhängige Variationen zwischen Ohr- und Lautsprecherpositionen oder leichte Kopfbewegungen auf die Maskierwirkung auswirken. Diese Effekte sind durch die hier verwendete Kopfhörerwiedergabe ausgeschlossen. Ebenso sollte untersucht werden, inwieweit sich fahrsituationsbedingte Innenraumgeräusche (z.B. Fahrgeräusche oder Lüftung) positiv auf die Maskierwirkung auswirken können, um damit ggf. die notwendigen Maskierpegel weiter zu verringern.

Literatur

[1] Wagener, K., Brand, T. und Kollmeier, B. (1999). „Entwicklung und Evaluation eines Satztests für die deutsche Sprache III: Evaluation des Oldenburger Satztests," Z. Audiol. 38, 86-95.

[2] Jung, O. (2000). "On the Lombard Effect induced by Vehicle Interior Driving Noises, Regarding SPL and Long-term Average Speech Spectrum," Acta Acust United Acust 98, 334-341.

[3] Byrne, D., Dillon, H., Arlinger, S., Wibraham, K., Cox, R., Hagerman, B., Hetu, R., Kei, J., Lui, C., Kiessling, J., Kotby, M., Nasser, N., El Kholy, Y. W., Nakanishi, O. H., Powell, R., Stephens, D., Meredith, R., Sirimanna, T., Tavatkiladze, G., Frolenkov, G., Westerman, S. und Ludvigsen, C. (1994). "An international comparison of long-term average speech spectra," J. Acoust. Soc. Am. 96, 2108-2120.

[4] ANSI (1997). "Methods for the Calculation of the Speech Intelligibility Index," Standards Secretariat, Acoustical Society of America, American National Standard, S3.5-1997.

[5] Simpson, S.A. und Cooke, M. (2005). "Consonant identification in N-talker babble is a nonmonotonic function of N (L), " J. Acoust. Soc. Am. 118, 2775-2778.

[6] Hochmuth, S., Brand, T., Jürgens, T. und Kollmeier, B. (2015). "Influence of noise type on speech reception thresholds across four languages measured with matrix sentence tests," International Journal of Audiology, Early Online 1-9. DOI: DOI: 10.3109/ 14992027. 2015.1046502.