

Sprachverständlichkeit in räumlich symmetrischen binauralen Störgeräuschsituationen

Wiebke Schubotz¹, Thomas Brand¹, Stephan D. Ewert¹

¹Medizinische Physik and Cluster of Excellence Hearing4all, Carl von Ossietzky-Universität Oldenburg, D-26111 Oldenburg, E-Mail: wiebke.schubotz@uni-oldenburg.de

Einleitung

Beim binauralen Hören treten zwischen den Ohren Pegel- und Phasenunterschiede auf (interaural level/phase differences; ILD/IPD), die sich je nach Position der Schallquelle relativ zum Kopf des Hörenden verändern. Diese interauralen Merkmale („cues“) sind besonders hilfreich, wenn zwei Schallquellen getrennt werden sollen, z.B. ein Sprachsignal (Target) und ein Störgeräusch (Maskierer). Für frontale Schallquellen treten keine ILDs und IPDs auf, jedoch für seitlich vom Kopf positionierte Schallquellen. Hier treten bei hohen Frequenzen (> 1.5 kHz) große ILDs durch Kopfabschattung auf, während bei tiefen Frequenzen (< 1.5 kHz) IPDs genutzt werden können. Das führt dazu, dass die Separation von zwei Schallquellen durch interaurale Merkmale einfacher wird, je weiter beide räumlich voneinander getrennt sind. Für Sprachverständlichkeit zeigt sich, dass gemessene Sprachverständlichkeitsschwellen (Speech Reception Thresholds, SRTs) sinken, je weiter Target und Maskierer räumlich auseinanderliegen. Dieser Effekt wird spatial release from masking (SRM) genannt.

SRTs in asymmetrischen Hörsituationen (Maskierer nur auf einer Seite des Kopfes präsentiert) können häufig durch ein „better ear“ erklärt werden, da es in solch einer Situation ein Ohr gibt, das ständig einen besseren Signal-Rausch-Abstand (SNR) aufweist. In symmetrischen Maskiersituationen (Maskierer wird auf beiden Seiten des Kopfes präsentiert) tritt so ein better ear Effekt nicht auf. Es gibt hingegen nur kurze spektro-temporale Bereiche („glimpses“; [1]), die einen hohen SNR aufweisen und diese können für beide Ohren gleichermaßen auftreten. Studien wie z.B. [2] schlagen vor, dass glimpses ausgenutzt werden und es ein schnelles Umschalten zwischen beiden Ohren gibt („better ear glimpsing“).

In dieser Studie wurde Sprachverständlichkeit in Situationen mit räumlich symmetrischen Maskierern für vier unterschiedliche Maskierer gemessen, sowie untersucht, inwiefern bestimmte interaurale cues (ILD, glimpses) die Sprachverständlichkeit in diesen Störgeräuschsituationen beeinflussen. Die unterschiedlichen Maskierer wurden angewendet, um zu untersuchen, ob der beobachtete SRM in allen Maskiersituationen in gleicher Größe auftritt. Die gewählten unterschiedlichen Maskierer bieten verschieden große glimpses und somit einen guten Einblick in die Nutzung dieser binauralen cues.

Methoden

In einem adaptiven Verfahren wurden die Schwellen bestimmt, bei denen die Probanden die Hälfte der im Maskierer präsentierten Wörter verstehen konnten. Der

Maskierpegel lag bei 65 dB SPL, der Pegel des Sprachsignals wurde während der Messung variiert. Insgesamt nahmen vier junge, normalhörende Probanden an den Versuchen teil. Zu Beginn der Messung wurden zwei Satzlisten mit jeweils 20 Targetsätzen in einem Cafeteria-Störgeräusch dargeboten, um die Probanden an das Sprachmaterial und die Messung zu gewöhnen.

Sprachmaterial

Als Target wurden Sätze aus dem Oldenburger Satztest (OLSA, [3]) verwendet, einem Matrixtest mit männlichem Sprecher, in dem die Targetsätze zufällig generiert werden können. Alle Sätze hatten dieselbe Satzstruktur (Name-Verb-Zahlwort-Adjektiv-Objekt), waren aber nicht vorher-sagbar. Für jeden Maskierer wurden unterschiedliche Testlisten mit jeweils 20 Sätzen benutzt. Das Sprachsignal wurde während der gesamten Messreihe frontal präsentiert (0°), wobei die Sprachsignale mit head related transfer functions (HRTFs) aus der Datenbank von [4] gefaltet wurden.

Maskierer

Spektro-temporale Manipulation

Die Sprachverständlichkeitsmessungen wurden mit vier unterschiedlichen Maskierern durchgeführt, einem stationären, zwei modulierten und einem sprachähnlichen Maskierer. Als sprachähnlicher Maskierer wurde das International Speech Test Signal (ISTS, [5]) verwendet, welches mit Hilfe des STRAIGHT Algorithmus [6] spektral so bearbeitet wurde, dass die Grundfrequenz (F_0) mit der des OLSA Sprechers übereinstimmte ($F_0 = 110$ Hz). Dadurch wurde eine männliche Version des ISTS erzeugt (ISTSmale). Durch eine FFT-Analyse und anschließende Randomisierung der Phasen, wurde ein stationäres Rauschen generiert, welches dasselbe Langzeitspektrum wie das ISTSmale aufweist (STAT). Auf dieses Rauschen wurde dann entweder eine 8-Hz Sinusmodulation (SAM-STAT) oder eine breitbandige Spracheinhüllende (BB-STAT) aufgeprägt um reguläre (SAM-STAT) und irreguläre, sprachähnliche Modulationen (BB-STAT) zu erzeugen. In beiden Fällen ist die resultierende Modulation kohärent über alle Frequenzgruppen. Die breitbandige Spracheinhüllende war die Hilbertinhüllende eines breitbandigen Sprachsignals, welches sich aus zehn OLSA Sätzen zusammensetzte. Die Spektrogramme der einzelnen Maskierer sind in Abbildung 1 gezeigt.

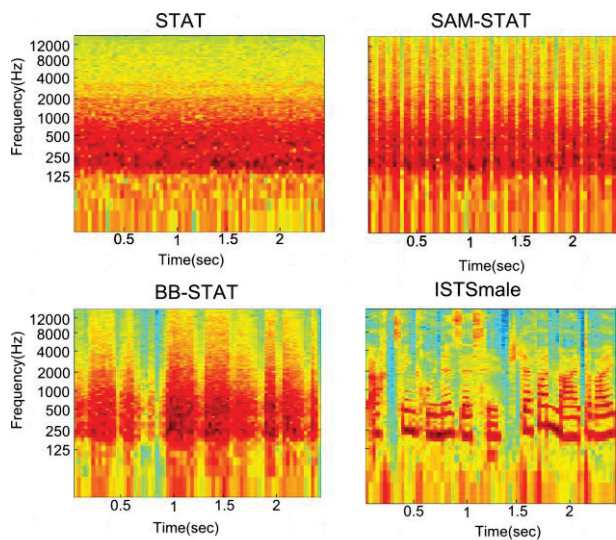


Abbildung 1: Spektrogramme der vier Maskierer, für die Sprachverständlichkeitsschwellen bestimmt wurden. Von links oben nach rechts unten sind dies ein stationärer Maskierer (STAT), ein Maskierer mit kohärenten, regulären 8-Hz Sinusmodulationen (SAM-STAT), ein Maskierer mit kohärenten, irregulären Modulationen (BB-STAT), sowie ein sprachähnlicher Maskierer (ISTSmale), der eine Version des ISTS ([5]) mit männlichen Sprechern darstellt. Alle vier Maskierer haben das gleiche Langzeitspektrum.

HRTF Manipulation

Alle Maskierer wurden mit HRTFs aus [4] gefaltet, um verschiedene Maskierpositionen zu realisieren. Zum einen wurde der Maskierer aus einer Position von 0° präsentiert (frontal), zum anderen aus -60° und $+60^\circ$ (räumlich symmetrisch), wie in Abbildung 2 dargestellt ist.

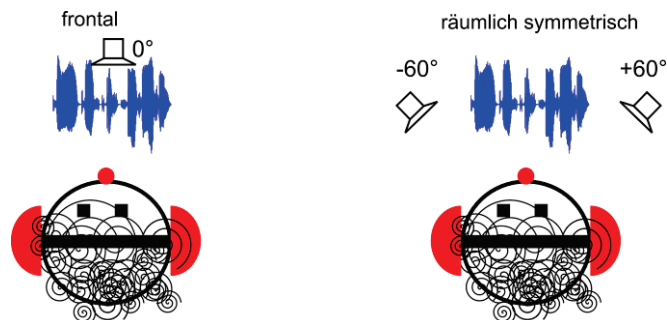


Abbildung 2: Räumliche Konfigurationen der Sprachverständlichkeitsmessungen. Es wurden SRTs für frontales Target und Maskierer (frontal) gemessen und für den Fall, in dem das Target frontal und der Maskierer bei -60° und $+60^\circ$ positioniert war (räumlich symmetrisch). Dafür wurden die Signale mit head related transfer functions (HRTFs) aus [4] gefaltet.

Dabei wurden je zwei statistisch unabhängige Maskierer präsentiert. Insgesamt wurden drei verschiedene HRTF Manipulationen durchgeführt, um die verschiedenen interauralen cues darzubieten. Zunächst wurden die originalen HRTFs aus [4] genutzt, um die frontale und symmetrische Maskierpositionen zu realisieren und alle interauralen Merkmale der originalen HRTFs darzubieten (full HRTF).

Für die Darbietung von Pegelunterschieden (ILD only) wurden die HRTFs so bearbeitet, dass die IPD eliminiert und nur die originalen ILDs dargeboten wurden. Für die dritte Manipulation wurden aus den Signalen beider Ohren die glimpses bestimmt, die den jeweils besseren SNR aufwiesen. Dabei wurde sich weitestgehend an der Verarbeitung aus [2] orientiert: Beide Maskiersignale wurden mit einer Gammafilterbank vierter Ordnung analysiert (50 Hz-16 kHz), die Ausgänge der einzelnen Filter dabei in 20-ms Fenster unterteilt und dann derjenige glimpse (linkes oder rechtes Ohr) bestimmt, der weniger Maskierenergie (also einen besseren SNR) aufwies. Diese „better ear glimpses“ des rechten und linken Ohres wurden dann zu einem monauralen Signal zusammengefasst (ideal monaural better ear mask, IMBM). Die IMBM-manipulierten Signale wurden während der Messung diotisch dargeboten (vgl. [2]). Da bei der Generierung der IMBM nur der Pegel des Maskierers an beiden Ohren eine Rolle spielt, wird in dieser Konfiguration auch keinerlei IPD Information dargeboten.

Ergebnisse

Die Abbildungen 3-4 zeigen die Ergebnisse der Sprachverständlichkeitsmessungen mit den vier verschiedenen Maskierern. Im oberen Teil von Abbildung 3 sind die gemessenen SRTs der frontalen (gefüllte Symbole) und symmetrischen (offene Symbole) Maskierpositionen für die Manipulationen full HRTF (ILD und IPD Information) und ILD only (nur ILD Information) zu sehen. Im unteren Teil sind die daraus resultierenden Werte für SRM gezeigt. Der Verlauf der SRTs ist für beide Maskierpositionen (frontal, symmetrisch) ähnlich, die größten SRTs werden für den stationären Maskierer gemessen, niedrigere Schwellen hingegen für die modulierten Maskierer. Dieser Verlauf tritt auch für die anderen Manipulationen (ILD only; IMBM, siehe Abbildung 4) auf. Für den sprachähnlichen Maskierer liegen die SRTs bei der ILD only Kondition etwas über denen der full HRTF Kondition, bei der IMBM jedoch leicht darunter. Bei Betrachtung der SRTs der symmetrischen Maskierposition (offene Symbole) in Abbildung 3, ist erkennbar, dass diese den gleichen Verlauf wie die SRTs der co-located Position aufweisen, jedoch generell um einige dB nach unten verschoben sind. Dieser Unterschied (SRM) beträgt für die full HRTF Kondition für den stationären Maskierer etwa 4 dB, 6 dB für beide modulierte Maskierer und etwa 10 dB für den sprachähnlichen Maskierer. Für die ILD only Kondition sind die SRMs um 1-2 dB kleiner. Generell zeigt sich, dass die Streuung der Messwerte für die ILD only Kondition abnimmt.

Betrachtet man die SRTs und SRM für die Darbietung der binauralen glimpses in Abbildung 4, so zeigt sich, dass die SRTs des frontalen Maskierers einen ähnlichen Verlauf wie bei full HRTF und ILD only zeigen, jedoch einen anderen Verlauf bei der symmetrischen Maskierposition. Dadurch ergeben sich deutliche Unterschiede in den resultierenden SRM, die nun 2 dB für den stationären, 3 dB und 6 dB für die modulierten und nur noch 6 dB für den sprachähnlichen Maskierer betragen.

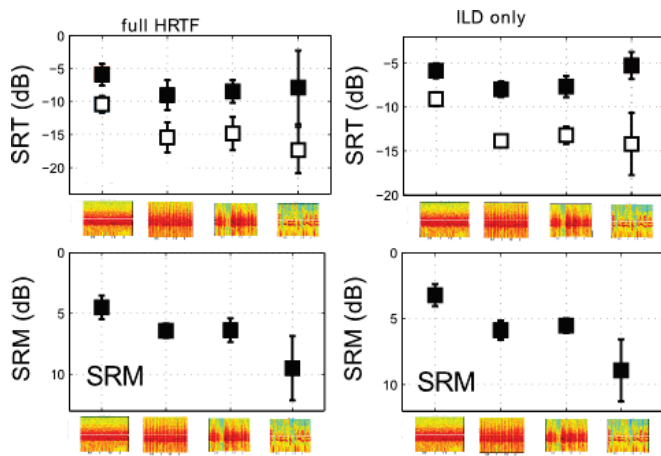


Abbildung 3: Gemessene Sprachverständlichkeitsschwellen für die HRTF Manipulationen full HRTF und ILD only mit den dazugehörigen Standardabweichungen, sowie die gemessenen Werte des SRM (untere plots). Die SRTs der frontalen Maskierposition sind mit gefüllten Symbolen, die der symmetrischen mit offenen Symbolen dargestellt. Der SRM ergibt sich als Differenz beider SRTs. Positive SRMs (nach unten aufgetragen) ergeben sich für niedrigere SRTs im symmetrischen Maskierer.

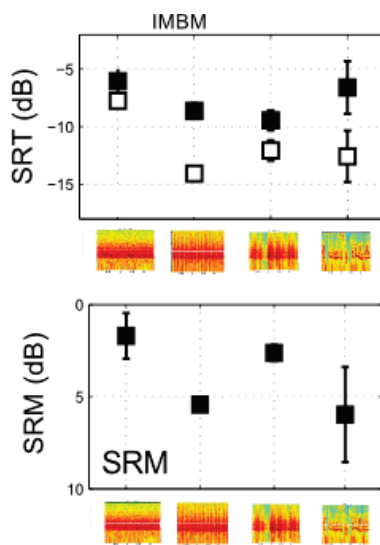


Abbildung 4: Gemessene Sprachverständlichkeitsschwellen mit Standardabweichungen, sowie die dazugehörigen Werte für SRM (unterer plot) für die Kondition IMBM. Die SRTs der frontalen Maskierposition sind vergleichbar zu denen der anderen HRTF Manipulationen, die SRTs der symmetrischen Maskierposition sind jedoch größer. Daher sind die resultierenden Werte für SRM kleiner als die der anderen HRTF Manipulationen.

Zusammenfassung und Diskussion

In allen Sprachverständlichkeitsmessungen zeigt sich, dass Sprachverständlichkeit größer ist, wenn es Lücken im Maskierer gibt, in denen das Target unmaskiert

wahrgenommen werden kann. Dies ist aus monauralen Studien (bspw. [7]) bekannt und kann hier für binaurale Maskierer mit symmetrischer Anordnung (frontal und symmetrisch) verifiziert werden. Desweiteren zeigt sich, dass, besonders für die full HRTF Kondition, die gemessenen SRM den Literaturwerten aus [8] und [9] entsprechen. Für stationäre Maskierer werden dort 3-4 dB, für sprachähnliche Maskierer bis zu 12 dB gemessen. Die ermittelten SRM fallen etwas geringer aus (vorallem für den stationären Maskierer), wenn nur Information über die ILDs dargeboten wird. Die Werte werden teilweise deutlich kleiner für den Fall der Darbietung binauraler glimpses (IMBM). Hier sinkt der SRM auf etwa 2 dB für den stationären Maskierer und auf 6 dB für den sprachähnlichen. Für die modulierten Maskierer fallen die gemessenen SRM sehr unterschiedlich aus: regelmäßige glimpses wie im Fall des SAM-STAT Maskierers führen zu einem vergleichbaren SRM von ca. 6 dB wie bei full HRTF und ILD only, während unregelmäßige glimpses (BB-STAT) zu einer deutlichen Verringerung des SRM führen.

Generell, lässt sich ein großer Teil der gemessenen SRTs und SRM-Werte durch die Darbietung von ILDs zwischen beiden Ohren erklären, da die Abweichungen zwischen der full HRTFs und ILD only Manipulation gering sind. Wenn binaurale glimpses dargeboten werden, gibt es jedoch größere Abweichungen. Diese Ergebnisse stehen im Gegensatz zu den Ergebnissen von [2] und weisen darauf hin, dass ein schnelles Umschalten zwischen beiden Ohren („better ear glimpsing“) nicht ausreicht, um ein ähnliche Ergebnisse wie mit der full HRTF zu erreichen. Die fehlende IPD Information ist als Grund für die schlechteren SRMs bei IMBM unwahrscheinlich, da sie auch in der ILD only Manipulation nicht vorhanden war. Ein wesentlicher Unterschied zwischen IMBM und ILD only bleibt die seitliche Lokalisation der Maskierer im Falle von ILD only, während die diotische Darbietung bei IMBM zu einer rein frontalen Wahrnehmung führt. Dieses Ergebnis muss weiter untersucht und weitere Probandenmessungen durchgeführt werden, um die bisherigen Ergebnisse zu untermauern.

Literatur

- [1] Cooke, M. (2006). A glimpsing model of speech perception in noise. *The Journal of the Acoustical Society of America*, 119(3), 1562-1573.
- [2] Brungart, D. S., & Iyer, N. (2012). Better-ear glimpsing efficiency with symmetrically-placed interfering talkers. *The Journal of the Acoustical Society of America*, 132(4), 2545-2556.
- [3] Wagener, K., Brand, T., and Kollmeier B. (1999). „Entwicklung und Evaluation eines Satztests für die deutsche Sprache III: Design, Optimierung und Evaluation des Oldenburger Satztests,“ *Z. Audiol.* 38 (3), 86-95.
- [4] Kayser, H., 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*, 2009.

[5] Holube, I., Fredelake, S., Vlaming, M., & Kollmeier, B. (2010). Development and analysis of an international speech test signal (ISTS). *International Journal of Audiology*, 49(12), 891-903.

[6] Kawahara, H., Morise, M., Takahashi, T., Nisimura, R., Irino, T., & Banno, H. (2008, March). TANDEM STRAIGHT: A temporally stable powerspectral representation for periodic signals and applications to interference-free spectrum, F0, and aperiodicity estimation. In *Acoustics, Speech and Signal Processing, 2008. ICASSP 2008. IEEE International Conference on* (pp. 3933-3936). IEEE.

[7] Bronkhorst, A. W. (2000). The cocktail party phenomenon: A review of research on speech intelligibility in multiple-talker conditions. *Acta Acustica united with Acustica*, 86(1), 117-128.

[8] Jones, G. L., & Litovsky, R. Y. (2011). A cocktail party model of spatial release from masking by both noise and speech interferers. *The Journal of the Acoustical Society of America*, 130(3), 1463-1474.

[9] Marrone, N., Mason, C. R., & Kidd Jr, G. (2008). Tuning in the spatial dimension: Evidence from a masked speech identification task. *The Journal of the Acoustical Society of America*, 124(2), 1146-1158.