



Technische Universität Berlin
Fakultät I - Geistes- und Bildungswissenschaften
Institut für Sprache und Kommunikation
Fachgebiet Audiokommunikation
Sek. EN-8 - Einsteinufer 17c - 10587 Berlin

Exposé zur Masterarbeit
**Messung der binauralen Sprachverständlichkeit mit 3rd-order
Ambisonics-Mikrofonen**¹
Betreuer: Prof. Dr. Anselm Goertz
April 2021

Lukas Kaiser
E-Mail: Lukas.Kaiser@campus.tu-berlin.de

¹Arbeitstitel

Inhaltsverzeichnis

1 Abstract	3
2 Einleitung und Fragestellung	3
3 Stand der Forschung	4
3.1 Raumakustik	4
3.2 Psychoakustik	4
3.3 Virtuelle Akustik	5
4 Methodik	5
4.1 Arbeitspaket 1 – Planung von Simulation/Versuchen	6
4.2 Arbeitspaket 2 – Signalverarbeitung der Mikrofonarraydaten und Algorithmus zur Berechnung des binaural benefits	6
4.3 Arbeitspaket 3 – Feldversuch und ggf. Simulation akustischer Szenen . .	6
4.4 Arbeitspaket 4 – Hörversuch	7
5 Arbeits- und Zeitplan	7

1 Abstract

Das binaurale Hören verbessert die Sprachverständlichkeit, dennoch werden in der Praxis vor allem einkanalige Messungen des Speech Transmission Index (STI) durchgeführt, da die Ermittlung des binaural benefits eines räumlichen Schallfeldes bis dato mit einem hohen messtechnischen Aufwand verbunden ist. Der Einsatz moderner portabler Mikrofonarrays reduziert diesen drastisch und bietet die binaurale Erweiterung des STI außerhalb eines Forschungskontexts an. Ziel des Forschungsvorhabens ist daher die rechnerische Umsetzung eines Algorithmus zur Messung der Sprachverständlichkeit mit einem 3rd-order Ambisonics-Mikrofon. Da die Sprachverständlichkeit innerhalb eines öffentlichen Raumes auch in einer Gefahrensituation gewährleistet werden muss und indes einer strikten Normung [1] unterliegt, ist das Vorhaben für Lautsprecher herstellende Unternehmen von wirtschaftlichem und gesamtgesellschaftlich aus sicherheitstechnischem Interesse relevant. Die Validierung der Messergebnisse soll mit im Feld aufgenommen oder simulierten BRIRs in einem Hörversuch erfolgen.

2 Einleitung und Fragestellung

Der Speech Transmission Index (STI) ist ein in der Raumakustik verwendeter numerischer Deskriptor, welcher die Sprachverständlichkeit innerhalb eines Raumes beschreibt [2]. Der STI ist auch für die Planung von Beschallungsanlagen eine wichtige Kenngröße, da im Gefahrenfall Warnsignale und -hinweise zum Schutz von Personen verständlich sein müssen. [1].

Die menschliche Fähigkeit des binauralen Hörens kann die Sprachverständlichkeit verbessern [3]. Dadurch kann es zu Abweichungen zwischen perceptiver Sprachverständlichkeit und gemessenen monoauralen STI kommen. Die Verbesserung der Sprachverständlichkeit wird in der Literatur dabei durch einen binaural benefit beschrieben, welcher aus binauralen Raumimpulsantworten (BRIRs) bestimmt werden kann [4].

Mit dem Verfahren Ambisonics lassen sich räumliche Schallfelder aufnehmen und können nach einem Rendering binaural wiedergegeben werden [5]. Die messtechnische Einrichtung und die Aufstellung dafür verwendeter mehrkanaliger Mikrofonarrays ist bis dato für die Einmessung von Lautsprecheranlagen mit einem sehr hohen, meist unpraktikablen Aufwand verbunden, weswegen im Feld Messungen des STI in der Regel monoaural stattfinden. Der Einsatz von neuentwickelten 3rd-order Ambisonics-Mikrofonen, welche mit eigener Soundcard ausgeliefert und über USB direkt mit einem beliebigen PC/DAW verbunden werden können, reduzieren den messtechnischen Aufwand drastisch. Die Verfügbarkeit neuer Technologien und die aktuell fortschreitende Erforschung verschiedener Modelle zur Berechnung des binaural benefits eröffnen damit eine Erweiterung des Messverfahrens für einen Einsatz in der Praxis außerhalb eines reinen Forschungskontexts. Durch die sicherheitstechnischen Aspekte ist eine exakte Ermittlung des STIs gesamtgesellschaftlich und für Lautsprecher herstellende Unternehmen mit Empfehlung und Einsatz ihrer Produkte auch wirtschaftlich relevant.

Ziel der Masterarbeit ist die Entwicklung und Umsetzung eines Algorithmus zur Messung des binauralen STIs mit 3rd-order Ambisonics-Mikrofonen. Dies umfasst zum einen die rechnerische Verarbeitung der räumlichen Audiodaten, Rendern von BRIRs sowie Berechnung des binaural benefits für eine jede Messung. Die Aussagekraft des berechneten binauralen STI muss mit einem Hörversuch validiert werden. Für diesen sollen gezielt Messungen und/oder Simulationen betrachtet werden, bei welchen nicht zu vernachlässigende Unterschiede zwischen STI und binauralem STI bestehen.

3 Stand der Forschung

Im folgenden Kapitel wird unterteilt in die Bereiche Raumakustik (Kapitel 3.1), Psychoakustik (Kapitel 3.2) und virtuelle Akustik (Kapitel 3.3) auf den für das Vorhaben relevanten und aktuellen Stand der Forschung eingegangen.

3.1 Raumakustik

Die Raumakustik als Teil der technischen Akustik befasst sich mit Schallfeldern innerhalb zumindest teilweise geschlossener Räume. Durch Reflexionen des Raumes wird das Signal durch die Übertragungsfunktion von Quelle zu Empfänger verzerrt [2]. Der Speech Transmission Index (STI) ist dabei ein raumakustischer Deskriptor der die Sprachverständlichkeit an einer bestimmten Stelle von Quelle und Empfänger des Raumes beschreibt. Der STI kann nach [6] aus der Raumimpulsantwort (RIR) bestimmt werden und bewertet dabei die Veränderung der Modulationstiefe des Signals durch Reflexionen des Raumes. Bei einer geringen Modulationstiefe werden vor allem energieschwache Silben mit stärker werdenden Nachhall zunehmend verdeckt [7]. Der STI unterliegt bei der Planung von Beschallungsanlagen in öffentlichen Räumen strengen sicherheitstechnischen Aspekten. Nach der DIN VDE 0833-4 (Gefahrenmeldeanlagen für Brand, Einbruch und Überfall) [1] muss dieser größer als 0.5 sein, um eine zeitnahe und effiziente Evakuierung innerhalb von Gebäuden sicherzustellen. Untersuchungen zeigen, dass die Sprachverständlichkeit durch das binaurale Hören (siehe Kapitel 3.2) verbessert wird [3]. Auch wenn es entwickelte und applizierbare Modelle für eine solche binaurale Erweiterung (auch binaural benefit genannt) gibt, so fehlt es an der Übertragung in die Praxis an praktikablen Lösungen mit geringem Messaufwand. In [8] konnten signifikante Zusammenhänge zwischen Nachhallzeit im Oktavband 250 Hz und verschiedenen binauralen Sprachverständlichkeitsmaßen festgestellt werden. Die Nachhallzeiten wurden durch Simulation verschiedener akustischer Umgebungen ermittelt.

3.2 Psychoakustik

Die Psychoakustik als Teil der Psychophysik umfasst die Wahrnehmung von Schalleignissen durch den Menschen. Das der Psychoakustik zuzuordnende binaurale Hö-

ren beruht auf der Auswertung von interauralen Zeit- und Pegeldifferenzen (ITDs und ILDs). [9] Bei der binauralen Demaskierung können räumlichgetrennte Schallquellen auch in der Wahrnehmung separiert und so etwaige Störsignale unterdrückt werden [10]. Die in Kapitel 3.1 genannte binaurale Erweiterung der Sprachverständlichkeit wird dabei in der Literatur durch einen binauralen benefit beschrieben. Die Berechnung kann mittels binauralen Raumimpulsantworten (BRIRs) durchgeführt werden [4]. Positiv wirken dabei vor allem die binauralen Effekte better-ear-listening und binaural unmasking [3]. Einfluss auf die binaurale Sprachverständlichkeit haben dabei auch die Orientierung des Kopfes mit einer Verbesserung des Speech Transmission Thresholds (SRT) von bis zu 8 dB [11]. In einer anderen Studie wurde der Einfluss von Head-Related Transfer Function (HRTF) (Verzerrung und Filterung eines Signals durch Kopf, Ohr und Torso des Menschen) mit einer Verbesserung des SRTs von bis zu 5.5 dB bewertet [12]. Eine Segmentierung von Nutz- und Störsignal, welche eine zeitliche Trennung von Direktschall und frühen eintreffenden Reflexionen von diffusen Nachhall und späten Reflexionen vornimmt, zeigte eine Verbesserung des SRTs um 1 dB [13]. Eine vollständige Beschreibung des binaural benefits durch binaurale Demaskierung steht dennoch aus. Gründe dafür könnten nach [13] eine wahrnehmungsbezogene Raumanpassung und binaurale Nachhallunterdrückung sein.

3.3 Virtuelle Akustik

Ein Teil der virtuellen Akustik ist die Aufnahme und Wiedergabe räumlicher Schallfelder, wofür das Ambisonics-Verfahren mit einem mehrkanaligen Mikrofonarray genutzt werden kann. Aus aufgenommenen RIRs können mit Hilfe verschiedener Verfahren BRIRs gerendert werden [14]. Dazu gehören die Spatial Decomposition Method (SDM) [15] und das Higher-order Spatial Impulse Response Rendering (HO-SIRR) [16]. Bei der SDM werden räumliche RIRs in verschiedene Bildquellen encodiert und daraus eine BRIR generiert. Bei HO-SIRR werden Laufzeitunterschiede und die Diffusität des Schallfeldes für die Berechnung von BRIRs herangezogen. Für beide Methoden gibt es frei verfügbare Implementierungen in der Programmiersprache Matlab (SDM [17]; HO-SIRR [14]). Für die Simulation eines räumlichen Eindrucks eines Schallfeldes sind auch die HRTFs von entscheidender Bedeutung, um ein für die Person natürliches und räumliches Klangfeld zu erzeugen. Diese können von der offenen Datenbank FABIAN [18] abgerufen werden. Teil der virtuellen Akustik ist auch die Auralisation virtueller Räume. Dabei werden mit verschiedenen numerischen Simulationsansätzen akustische Szenen virtuell auralisiert und der Einfluss der Szene hörbar gemacht [19]. Vorteil des Verfahrens ist, dass verschiedene akustische Umgebungen kosteneffizient simuliert und schon in der Planungsphase Konfigurationen oder Änderungen evaluiert und überprüft werden können.

4 Methodik

Das Forschungsvorhaben wird folgend in verschiedene Arbeitspakete unterteilt.

4.1 Arbeitspaket 1 – Planung von Simulation/Versuchen

Für die Durchführung von Feldversuchen und ggf. Simulationen akustischer Szenen (siehe Kapitel 4.3) sollen im Voraus bereits Räume und Anordnungen von Beschallungsanlagen identifiziert werden, bei welchen es wahrscheinlich zu einer starken Abweichung zwischen STI und binauralen STI kommen kann. Ergebnis des Arbeitspakets 1 ist ein Katalog von Räumen, Lautsprecheranordnungen und Raumpositionen bei denen ein hoher binaural benefit wahrscheinlich ist.

4.2 Arbeitspaket 2 – Signalverarbeitung der Mikrofonarraydaten und Algorithmus zur Berechnung des binaural benefits

Zur Bestimmung des STI und binaural benefits sollen mit einem Zylia ZM-1 (3rd-order Ambisonics-Mikrofon) [20] sweepbasierte Raumimpulsantworten gemessen werden. Das Mikrofonarray besitzt 19 omnidirektionale Kapseln und kann über USB direkt mit einer beliebigen DAW verbunden werden. Durch den Einfluss der Kopforientierung auf den binaural benefit nach [12] sollen auch in [18] verfügbare HRTFs berücksichtigt werden. Anschließend werden auf Grundlage der verfügbaren Implementierungen für SDM [17] und HO-SIRR [21] BRIRs gerendert. Mit diesen kann für verschiedene Kopforientierung der STI und jeweilige binaural benefit mit dem Modell von [4] bestimmt werden. Auch der Frequenzgang des Mikrofons soll für die Berechnungen entzerrt werden, wodurch eine Ermittlung des Frequenzgangs aller Mikrofone und verschiedener Einfallrichtungen notwendig ist. Ergebnis des Arbeitspaketes ist ein eigenständiges Modul in der Programmiersprache Python, welches anhand von 20 Eingangssignalen (1 Anregungssignal, 19 Mikrofone) STI und binaural benefit im Feld für beliebige Kopforientierungen bestimmen kann. Die Berechnung der binauralen Sprachverständlichkeit soll auch unter der Berücksichtigung der Modellerweiterung von [12] stattfinden, wobei die BRIR zeitlich in Nutz- und Störsignal zerlegt wird.

4.3 Arbeitspaket 3 – Feldversuch und ggf. Simulation akustischer Szenen

Das Messsystem inklusive Software soll im Feld in verschiedenen Räumen mit unterschiedlichen Beschallungsanlagen erprobt werden. Dafür sollen vor allem akustische Umgebungen betrachtet werden, bei welchen nach Analyse in Kapitel 4.1 ein hoher binaural benefit zu erwarten ist und der Einfluss verschiedener Beschallungsanlagen auf die binaurale Sprachverständlichkeit überprüft werden. Mit dem Feldversuch soll ebenfalls sichergestellt werden, ob das entwickelte Modul und Messsystem für den Einsatz im Feld geeignet ist sowie Messungen mit geringem Aufwand durchgeführt werden können. Im Feldversuch sollen sowohl RIRs und BRIRs ermittelt werden, welche für die Validierung der Messergebnisse durch einen Hörversuch, wie in Kapitel 4.4 beschrieben, verwendet werden sollen. Sollte eine Vergrößerung des Datensatzes notwendig sein, können zusätz-

lich zu den tatsächlichen Messergebnissen räumliche RIRs aus simulierten vergleichbaren akustischen Szenen und Gesamtanordnungen hinzugezogen werden.

4.4 Arbeitspaket 4 – Hörversuch

Mit den im Feldversuch ermittelten und/oder simulierten RIRs und BRIRs soll vergleichend der STI und binaural benefit mittels Hörversuch empirisch validiert werden und die Eignung der beiden verwendeten Renderings überprüft werden. So kann sichergestellt werden, ob Messsystem und -software zuverlässige Ergebnisse generieren. Ergebnis des Hörversuchs sind mit dem Oldenburger Satztest ermittelte Speech Reception Thresholds, welche zu Korrelation mit den Messwerten genutzt werden können.

5 Arbeits- und Zeitplan

In Tabelle 1 sind, unterteilt in einzelne Meilensteine, die in Kapitel 4 beschriebenen Arbeitspakete und deren Abschlussdaten festgehalten.

Tabelle 1: Arbeits- und Zeitplan des Forschungsvorhabens.

Fällig bis	Aufgabe	Arbeitspaket
19. April 2021	Erstellung des Exposés	-
3. Mai 2021	Katalog zu akustischen Szenen mit hohem zu erwartenden binaural benefit	1
17. Mai 2021	Implementierung einer Schnittstelle zur Eingabe der Messdaten	2
8. Juni 2021	Implementierung einer Methode zur Berücksichtigung der HRTF	2
15. Juni 2021	Implementierung der Renderverfahren SDM und HO-SIRR	2
25. Juni 2021	Implementierung der STI und binaural benefit Berechnung nach [4]	2
12. Juli 2021	Implementierung des U/D-Ansatzes nach [13]	2
1. August 2021	Code Beta-Version	2
16. August 2021	Planung und Vorbereitung von Feldversuchen und Simulationen	3
20. September 2021	Durchführung der Feldversuche und Simulationen	3
1. Oktober 2021	Auswertung der Feldversuche und Simulationen	3
18. Oktober 2021	Planung und Vorbereitung des Hörversuches	4
1. November 2021	Durchführung des Hörversuchs	4
7. November 2021	Auswertung des Hörversuchs	4
14. November 2021	Finalisierung der schriftlichen Ausfertigung	-
19. November 2021	Abgabe der Thesis	-

Literatur

- [1] “Gefahrenmeldeanlagen für brand, einbruch und überfall - teil 4: Festlegungen für anlagen zur sprachalarmierung im brandfall,” Oct. 2014.
- [2] H. Kuttruff, *Room acoustics*. Boca Raton: CRC Press/Taylor & Francis Group, sixth edition ed., 2017.
- [3] J. C. Middlebrooks, J. Z. Simon, A. N. Popper, and R. R. Fay, eds., *The Auditory System at the Cocktail Party*. Springer International Publishing, 2017.
- [4] S. Jelfs, J. F. Culling, and M. Lavandier, “Revision and validation of a binaural model for speech intelligibility in noise,” *Hearing research*, vol. 275, no. 1-2, pp. 96–104, 2011.
- [5] J. Ahrens and S. Spors, “Applying the ambisonics approach to planar and linear distributions of secondary sources and combinations thereof,” *Acta Acustica united*

- with Acustica*, vol. 98, no. 1, pp. 28–36, 2012.
- [6] Schroeder, Manfred Robert, “Modulation transfer functions: Definition and measurement,” *Acta Acustica united with Acustica*, no. 49, pp. 179–182, 1981.
 - [7] W. Ahnert and H.-P. Tennhardt, “Raumakustik,” in *Handbuch der Audiotechnik* (S. Weinzierl, ed.), pp. 181–266, Berlin, Heidelberg: Springer Berlin Heidelberg, 2008.
 - [8] Amirmohammad Ala, “Speech intelligibility and reverberation at low frequencies: Masterarbeit,” 2020.
 - [9] T. S. Huang, T. Kohonen, M. R. Schroeder, H. K. V. Lotsch, E. Zwicker, and H. Fastl, *Psychoacoustics*, vol. 22. Berlin, Heidelberg: Springer Berlin Heidelberg, 1999.
 - [10] W. E. Kock, “Binaural localization and masking,” *The Journal of the Acoustical Society of America*, vol. 22, pp. 801–804, Nov. 1950.
 - [11] J. A. Grange and J. F. Culling, “The benefit of head orientation to speech intelligibility in noise,” *The Journal of the Acoustical Society of America*, vol. 139, no. 2, pp. 703–712, 2016.
 - [12] O. Kokabi, F. Brinkmann, and S. Weinzierl, “Prediction of speech intelligibility using pseudo-binaural room impulse responses,” *The Journal of the Acoustical Society of America*, vol. 145, no. 4, p. EL329, 2019.
 - [13] O. Kokabi, F. Brinkmann, and S. Weinzierl, “Segmentation of binaural room impulse responses for speech intelligibility prediction,” *The Journal of the Acoustical Society of America*, vol. 144, no. 5, p. 2793, 2018.
 - [14] L. McCormack, V. Pulkki, A. Politis, O. Scheuregger, and M. Marschall, “Higher-order spatial impulse response rendering: Investigating the perceived effects of spherical order, dedicated diffuse rendering, and frequency resolution,” *Journal of the Audio Engineering Society*, vol. 68, no. 5, pp. 338–354, 2020.
 - [15] Sakari Tervo, Jukka Pätynen, Antti Kuusinen, and Tapio Lokki, “Spatial decomposition method for room impulse responses,” *Journal of the Audio Engineering Society*, no. 61, pp. 16–27, 2013.
 - [16] J. Merimaa and V. Pulkki, “Spatial impulse response rendering i: Analysis and synthesis,” *Journal of the Audio Engineering Society*, vol. 53, pp. 1115–1127, december 2005.
 - [17] S. Tervo, “Sdm toolbox.” <https://www.mathworks.com/matlabcentral/fileexchange/56663-sdm-toolbox>, 2018. Abgerufen am 16.04.2021.
 - [18] F. Brinkmann, “The fabian head-related transfer function data base,”
 - [19] M. Vorländer, *Auralization*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008.
 - [20] “Zylia zm-1.” <https://www.zylia.co/white-paper.html>, 2018. Abgerufen am 17.04.2021.

- [21] Leo McCormack, Archontis Politis, and Ville Pulkki, “Ho-sirr.” <https://github.com/leomccormack/HO-SIRR>, 2020. Abgerufen am 16.04.2021.