Lautheits- und Entfernungswahrnehmung von Sprache in virtuellen Räumen

Stephanie Blank[1], Stephan Töpken[2], Steven van de Par[3]

[1] CvO Universität Oldenburg, Abteilung Akustik, Cluster of Excellence Hearing4all, 26111 Oldenburg, Email: stephanie.blank@uol.de

[2] CvO Universität Oldenburg, Abteilung Akustik, Cluster of Excellence Hearing4all, 26111 Oldenburg, Email: stephan.toepken@uol.de

[3] CvO Universität Oldenburg, Abteilung Akustik, Cluster of Excellence Hearing4all, 26111 Oldenburg, Email: steven.van.de.par@uol.de

Einleitung

Für die wahrgenommene Lautheit eines Sprechers scheint die Abnahme des Schalldruckpegels am Ohr mit der Entfernung zwischen Sprecher und Hörer als ausschlaggebend [1], jedoch können auch Lautheitskonstanzeffekte sowohl für Sprache als auch für Rauschen auftreten [1, 2]. Möglicherweise könnten darüber hinaus kognitive Faktoren wie die angenommene Quellleistung und die Sprechanstrengung die Lautheitswahrnehmung beeinflussen.

In dieser Studie wird untersucht, ob innerhalb einer virtuellen akustischen Umgebung die wahrgenommene Lautheit vom Pegel nahe der Quelle (Quellleistung) oder dem Pegel am Ohr abhängt und welchen Einfluss die Sprecherentfernung und Sprechanstrengung auf die Lautheits- und Entfernungswahrnehmung haben. Im Gegensatz zu [1] und [2] werden keine real aufgenommenen BRIRs verwendet, sondern die Stimuli mit virtuellen BRIRs erzeugt [3, 4, 5]. Daher war ein weiteres Ziel zu prüfen, ob mit diesen Voraussetzungen und dieser Art von psychoakustischen Hörexperimenten ebenfalls perzeptiv plausible Ergebnisse resultieren.

Methode

Innerhalb von zwei Hörexperimenten hatten sieben Probandinnen die Aufgabe die Lautheit in zwei virtuellen Umgebungen (Raum L und FH) von zuvor generierten Sprachstimuli im Verhältnis zu einem Referenzsignal nach der Methode der Magnitude Estimation [6] zu bewerten und die Entfernung zu einer Sprecherin nach der Absolute Magnitude Estimation-Methode [7] einzuschätzen. Der Hörversuch zur Lautheitswahrnehmung bestand aus zwei Messterminen mit mindestens 24 Stunden Abstand zwischen den Durchführungen. Die Entfernungsbeurteilung erfolgte am Ende des zweiten Termins. Es nahmen sieben weibliche Studierende der Universität Oldenburg an den Versuchen teil. Allesamt hatten bereits Erfahrung mit der Methode der Magnitude Estimation unter Verwendung eines Referenzsignals innerhalb eines anderen Experiments [8]. Das Alter der Probandinnen ohne bekannte Hörbeeinträchtigungen betrug 21 bis 29 Jahre.

Stimuli

Zur Erzeugung der Sprachstimuli wurden zuvor von einer Sprecherin der Satz „Hallo, jetzt bin ich hier.“ in drei verschiedenen Sprechanstrengungen (▼ leise, ♦ mittel, ▲ laut) mit einem Pegel von etwa 48, 54 und 60 dB SPL (RMS) (gemessen in 1 m Abstand durch ein Referenzmikrofon) in einer Hörkabine (Abb. 1) aufgenommen. Diese Anregungssignale mit einer Länge von 2 Sekunden wurden im Pegel angehoben und entsprechen den ’natürlichen’ Konditionen in Verbindung von aufgenommener Sprechanstrengung und Pegelwert (• auf der Diagonalen in Abb. 2). Zur Untersuchung

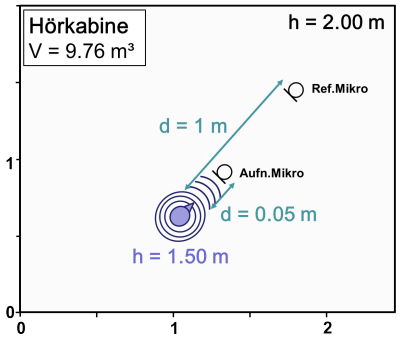

Abbildung 1: Grundriss der Hörkabine aus der realen Aufnahmesituation mit zusätzlichem Referenzmikrofon.

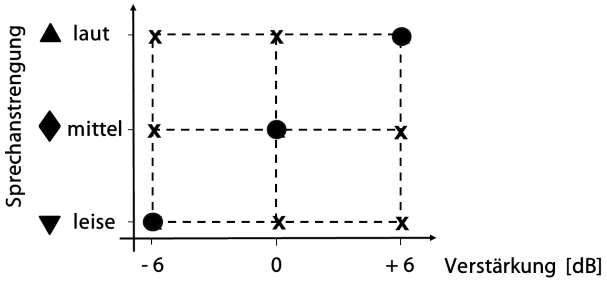

Abbildung 2: Pegelvariationen der aufgenommenen Sprachsignale. Die Konditionen auf der Diagonale (•) entsprechen den originalen Pegeln bei den jeweiligen Sprechanstrengungen (’natürliche’ Konditionen).

des Einflusses der Sprechanstrengung im Zusammenhang mit dem Pegel auf die Lautheits- und Entfernungswahrnehmung [1] erfolgte zusätzlich, ausgehend von der mittleren Sprechanstrengung (0 dB Verstärkung), eine Modifizierung der Pegel um + oder -6 dB über die Sprechanstrengungen hinweg. Insgesamt entstanden so neun verschiedenen Anregungssignale (Abb. 2), welche innerhalb der raumakustischen Simulation mit RAZR als Eingangssignale des Senders mit Sprecher-Richtcharakteristik dienen. Nach der Auralisierung der generierten Anregungssignale bei verschiedenen Sender-Hörer-Positionen in den zwei virtuellen Umgebungen durch RAZR ergaben sich 63 Stimuli für beide Hörexperimente. 27 Stimuli für Raum L (Abb. 3) und 36 für Raum FH (Sprechanstrengung x Verstärkung x Entfernung). Diese wurden im Vorfeld erzeugt.

Das Referenzsignal zur Bestimmung der Lautheitsurteile wurde als ein Rauschen mit Langzeitsprachspektrum (SSN (Speech Shaped Noise)) generiert [9]. Hierzu wurde die trockene Sprachaufnahme der ’natürlichen’ mittleren Sprechanstrengung verwendet. Nach der Auralisierung ergab sich für

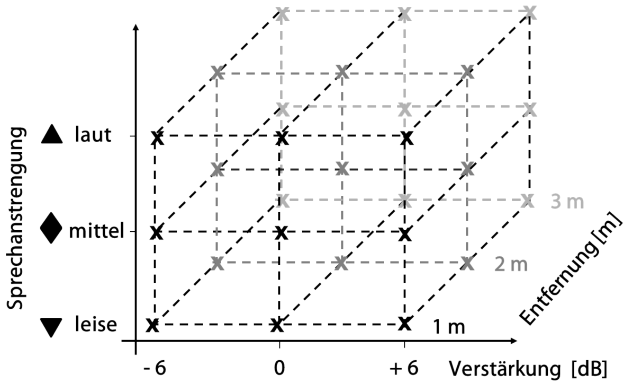

Abbildung 3: Messkonditionen beispielhaft für Raum L mit den Faktoren: Sprechanstrengung (▼ leise, ♦ mittel, ▲ laut), Pegelstufen (-6, 0, +6 dB) und Sprecher-Hörer-Entfernungen (1, 2, 3 m in Grauton-Abstufungen dargestellt).

das SSN für beide Räume ein identischer Pegel im Direktschall von 62 dB SPL (RMS) und im Gesamtschall (in 1 m virt. Abstand) wie der mittleren Sprechanstrengung (Tab. 1). Die Referenzsignale stimmen sowohl im Langzeitsprachspektrum, in der Richtcharakteristik, in den binauralen und räumlichen Einflüssen als auch im Pegel mit den Gegebenheiten der auralisierten Stimuli mittlerer Sprechanstrengung in 1 m Abstand überein.

Virtuelle Akustik

Für die Durchführung der Hörexperimente I und II wurde die RAZR Version 0.93 nach Wendt [3] mit der Erweiterung einer Quellcharakteristik durch Steffens [4] verwendet. RAZR ist eine umfassende, perzeptiv plausible raumakustische Simulation zur effizienten Berechnung und Auralisierung von (B)RIRs. Der Direktschall als auch die frühen Reflexionen von Räumen werden mit Hilfe der ISM Methode modelliert [10], dabei gilt zur Vereinfachung die standardmäßige Annahme einer Schuhkarton-Form. Zur Generierung des Nachhalls gehen die Reflexionen Nter-Ordnung in eine (binaurale) Erweiterung eines FDN-Modells ein [11]. Neben der Wahl einer sprecherähnlichen Abstrahlcharakteristik auf der Senderseite [4] wurde eine HRTF-Datenbank zur Erzeugung binauraler RIRs auf der Empfängerseite in RAZR genutzt. Die Messung der HRTF-Datenbank fand mit einem Cortex MK2 Kunstkopf nach dem Verfahren aus [12] statt. Die Festsetzung des ISM erfolgte auf N=3 und des FDN auf N=12. Die Synthetisierung der BRIRs erfolgte für zwei unterschiedliche raumakustische Umgebungen (Abb. 4): einem Labor (’Raum L’) [3] sowie einem Seminarraum der Jade Hochschule Oldenburg (’Raum FH’) [5]. Es wurden drei oder vier Sender-Empfänger-Positionen mit einer Orientierung zueinander diagonal in den zwei virtuellen Umgebungen (Raum L: 1 m, 2 m und 3 m; Raum FH: 1 m, 2 m, 4 m und 8 m) gewählt. Die Schallquelle hat eine Höhe von 1.50 Metern entsprechend der Mundposition der stehenden Sprecherin aus der Aufnahmesituation, ebenso wie die Empfänger. Zusätzlich wurde eine Entzerrung auf den Frequenzgang des HD 650 Kopfhörers vorgenommen.

Für Raum L und FH sind die Schallpegel der auralisierten Signale in 1 m Abstand für Direkt- und Gesamtschall unter Anregung der ’natürlichen’ Sprachsignale leiser, mittlerer und lauter Sprechanstrengung (-6, 0, +6 dB Verstärkung) in der Tabelle 1 aufgetragen. In beiden virtuellen Räumen ist der gleiche Direktschallpegel (in 1 m virt. Abstand) vorhanden, d.h. der gleiche Schallpegel wie von einem Sprecher in einem quasi-anechoischen Raum in 1 Meter Abstand am Hörer erzeugt. Die Abweichungen im Gesamtschallpegel erfolgen durch den individuellen Raumeinfluss je nach virtueller Umgebung.

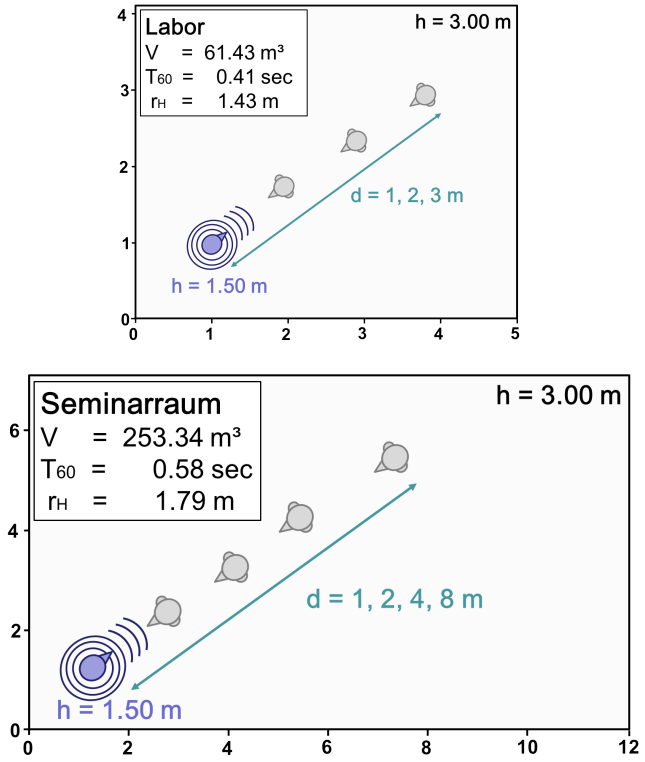

Abbildung 4: Draufsicht der innerhalb der zwei Hörexperimente und der Simulation mit RAZR verwendeten virtuellen Umgebungen. Oben: ein Labor der Universität Oldenburg (Raum L) [3] mit drei Sender-Empfänger-Positionen; Unten: ein Seminarraum der Jade Hochschule Oldenburg (Raum FH) [5] mit vier Sender-Empfänger-Positionen. Die Höhe der Schallquelle (lila) und der Empfänger (grau) beträgt h = 1.50 Meter.

Pegel @1 m [dB (SPL)]	Sprechanstrengungen		
	▼ leise	♦ mittel	▲ laut
Direkt	56	62	68
Raum L	63	69	75
Raum FH	60	66	72
Verstärkung	-6	0	+6

Tabelle 1: Direktschall- und Gesamtschallpegel in 1 Meter virtuellem Abstand frontal zum Sprecher in Raum L und FH unter Anregung der ’natürlichen’ Sprachsignale leiser, mittlerer und lauter Sprechanstrengung (-6, 0, +6 dB Verstärkung) aus Abb. 2. Die 0 dB Verstärkung entspricht einem Direktschallpegel von 62 dB SPL (RMS) an der Hörerposition in 1 m virtuellem Abstand unter Anregung des ’natürlichen’ mittleren Sprachsignals.

Versuchsaufbau und -durchführung

In zwei Hörexperimenten wurden die generierten Stimuli den sieben Probandinnen binaural über Kopfhörer (HD 650) in einer schallisolierten, einwandigen Hörkabine dargeboten. Der Darbietungspegel am Ohr der Probandinnen wurde auf 60 dB SPL kalibriert. Als Kalibrierungssignal (für den

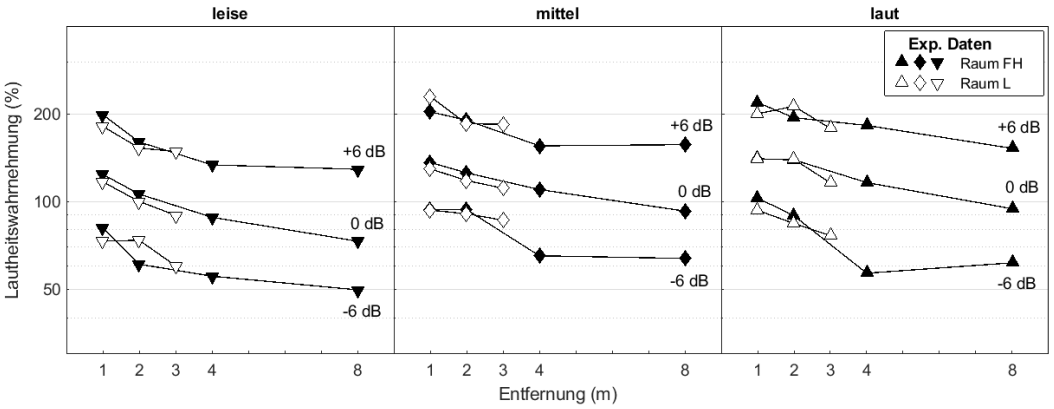

Abbildung 5: Ergebnisse der Lautheitswahrnehmung des virtuellen Raums L (*weiße Symbole*) und Raum FH (*schwarze Symbole*). Darstellung der geometrischen Mittelwerte und geometrischen Standardfehler über alle Probandinnen und beide Testdurchführungen intraindividuell gemittelt. Die geometrischen Standardfehler sind so gering, dass sie unterhalb der Markergröße liegen.

Messverstärker) diente das SSN bei omnidirektionaler Abstrahlung in einer virtuellen Hörkabine in RAZR (ISM: N=3 und FDN: N=12) auralisiert. Es ergab sich ein Schallpegel von 60 dB SPL (RMS), in einem virtuellen Abstand von 1 m gemessen.

In separaten Versuchen hatten die Probandinnen die Aufgabe, die empfundene Lautheit der Sprachsignale gegenüber der SSN-Referenz zu bewerten sowie die Entfernung zur Schallquelle absolut einzuschätzen, ohne dass sie mit der Instruktion gezielt auf die Bewertung der Lautheit der Quelle oder des Schalls am Ohr hingewiesen wurden [13]. Am ersten Tag der Hörexperimente fanden die Untersuchungen zur Lautheitswahrnehmung statt. Drei Probandinnen beurteilten Raum L zuerst, woraufhin Raum FH folgte. Für die anderen vier Probandinnen war die Reihenfolge der Beurteilung umgekehrt. Die Darbietung der Konditionen fand jeweils individuell randomisiert statt. Vor jeder Präsentation eines virtuellen Raums erfolgte ein Training mit ausgewählten, randomisierten Stimuli. Dieses Vorgehen wurde am zweiten Testtag wiederholt. Erst danach erfolgte das zweite Hörexperiment zur Entfernungswahrnehmung mit der gleichen zuvor festgelegten Reihenfolge der virtuellen Räume. Für die Lautheitsbewertung wurde stets das SSN-Referenzsignal vor dem Sprachstimulus präsentiert. Die Probandinnen hatten daraufhin die Aufgabe, die empfundene Lautheit des Stimulus im Verhältnis zum Signal zu bewerten und in Prozent anzugeben, wobei das SSN als 100%-Referenz anzusehen war. Beim Entfernungsexperiment erfolgte nur eine Präsentation des Stimulus und die Aufgabe war eine absolute Entfernungsschätzung in Metern. Eine Wiederholung der Signale war in beiden Fällen lediglich einmal möglich.

Ergebnisse

Lautheitswahrnehmung

Vor der Mittelung der Lautheitsurteile wurde pro virtuellem Raum eine 3-faktorielle ANOVA mit Messwiederholung der logarithmierten Messdaten aller Probandinnen und beider Testtage durchgeführt. Dabei erfolgte eine Benferroni-Korrektur des Signifikanzniveaus auf $\alpha = 0.025$. Die Messreliabilität (Korrelation zwischen Test und Testwiederholung) ist für beide virtuelle Umgebungen sehr hoch (Raum L: $r = 0.97$; Raum FH: $r = 0.98$) und die Messwiederholung ist statistisch nicht signifikant ($F(1, 6) = 0.512$. $p = 0.501$), weshalb die Messdaten beider Testtage intraindividuell für alle Probandinnen geometrisch gemittelt wurden. In Abbildung 5 sind die Lautheitsverhältnisse auf einer logarithmischen Skala über die Entfernung (linear) pro Sprechanstrengung aufgetragen. Die geometrischen Standardfehler sind so gering, dass sie unterhalb der Markergröße liegen. Die 100%-Linie entspricht dabei der SSN-Referenz, gegenüber welcher die Lautheitsurteile abzugeben waren.

Insgesamt zeigt sich für die Lautheitswahrnehmung für beide Räume ein signifikanter Einfluss von Pegel, Sprechanstrengung und Entfernung ($p < 0.025$). Die Lautheit nimmt über die Entfernung ab, was auf eine Abhängigkeit der Lautheitswahrnehmung vom Pegel am Ohr hindeutet. Ein höherer Pegel sowie ebenso eine höhere Sprechanstrengung gehen einher mit einer höheren Lautheitsbewertung, wobei der Effekt mit Anstieg der Sprechanstrengung abnimmt. Zwischen den drei Pegelkonditionen ist innerhalb einer Sprechanstrengung ein Offset mit teilweise parallelem Verlauf zu erkennen. D.h. eine Erhöhung des Pegels um 6 dB wurde über die Entfernungskonditionen hinweg als annähernd gleich stark jedoch geringer als eine Verdopplung der Lautheit wahrgenommen. Eine Halbierung der Lautheit über die Entfernung ist nicht zu beobachten.

Entfernungswahrnehmung

Für die logarithmierten Messdaten der Entfernungswahrnehmung wurde ebenfalls pro virtuellen Raum eine 3-faktorielle ANOVA mit Messwiederholung aller Probandinnen durchgeführt, wobei wieder eine Benferroni-Korrektur des Signifikanzniveaus ($\alpha = 0.025$) erfolgte. In Abbildung 6 sind die geometrischen Mittelwerte (pro virtuellen Raum über alle Probandinnen (N=7)) und geometrischen Standardfehler auf einer linearen Skala über die tatsächliche Entfernung pro Sprechanstrengung abgebildet.

Für die Entfernungswahrnehmung zeigte sich ebenfalls ein signifikanter Effekt von Pegel, Sprechanstrengung und Entfernung ($p < 0.025$). Es erfolgt ein Abfall der Entfernungsschätzung mit höherem Pegel, jedoch in Teilen ein Anstieg mit höherer Sprechanstrengung. D.h., bei einem höheren Pegel wird die Quelle von den Probandinnen als Näher empfunden, jedoch mit höherer Sprechanstrengung als weiter entfernt bewertet. Diese Beobachtung ist zwischen leiser und mittlerer Sprechanstrengung stärker ausgeprägt und sinkt zur Lauten. Es zeigt sich insgesamt eine erhöhte Entfernungsschätzung mit Anstieg in der Distanz bis zu 4 m. Bei geringeren Entfernungen sind tendenziell korrekte Schätzungen (*gepunktete Diagonale*) zu erkennen, wobei für höhere Entfernungen (> 4 m) eine starke Unterschätzung vorhanden ist.

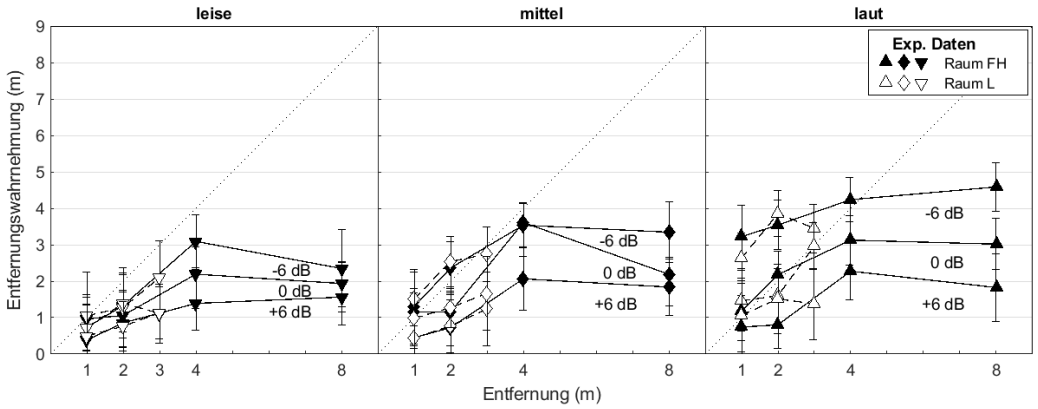

Abbildung 6: Ergebnisse der Entfernungswahrnehmung des virtuellen Raums L (*weiße Symbole*) und Raum FH (*schwarze Symbole*). Darstellung der geometrischen Mittelwerte und geometrischen Standardfehler über alle Probandinnen intraindividuell gemittelt. Die *gepunktete Diagonale* stellt dabei die theoretisch korrekte Entfernungsschätzung dar.

Zusammenfassung und Ausblick

Die Ergebnisse zeigen für beide Experimente einen signifikanten Einfluss von Pegel, Entfernung und Sprechanstrengung. Insgesamt stimmen die Ergebnisse der Hörversuche bis zu 4 m mit den Ergebnissen aus der Literatur mit real gemessenen BRIRs überein [1, 2]. Dies weist auf eine Bewertung des Pegels am Ohr für die Lautheitswahrnehmung hin. Es konnte gezeigt werden, dass die virtuell erzeugten BRIRs und somit RAZR sich gut für Experimente zur Lautheits- und Entfernungswahrnehmung eignen. Die Untersuchungen zum Einfluss der Pegel nahe der Quelle bzw. der in Teilen konstanten Bewertungen über 4 Meter hinaus könnten in einer Weiterführung der Experimente mit der neusten Version von RAZR und größeren raumakustischen Umgebungen, Variationen im DRR und Sprachaufnahmen mit 'natürlicher' Sprechanstrengung [14] weiter geführt werden.

Literatur

[1] Berthomieu, G. (2019). Influence de la position d'une source sur le niveau sonore perçu. PhD-Thesis. Acoustique [physics.class-ph]. Université de Bretagne Occidentale - Brest. Français.

[2] Zahorik, P., Wightman, F. L. (2001). Loudness constancy with varying sound source distance. Nat. Neurosci., 4, 78-83.

[3] Wendt, T., van de Par, S., Ewert, S. D. (2014). A computationally-efficient and perceptually-plausible algorithm for binaural room impulse response simulation. J. Audio Eng. Soc., 62 (11), 748-766.

[4] Steffens, H., van de Par, S., Ewert, S. D. (2019). Perceptual relevance of speaker directivity modelling in virtual rooms. Proc. 23rd International Congress on Acoustics ICA 2019, 2651-2658.

[5] Budnik, A., Steffens, H., Ewert, S. D., van de Par, S., Blau, M. (2019). Dynamische binaurale Auralisierung von Klassenraumakustik - Einfluss der Quellenrichtcharakteristik. Fortschritte der Akustik - DAGA 2019, Rostock, 1450-1453. Dega e.V., Berlin.

[6] Stevens, S. S. (1956). The Direct Estimation of Sensory Magnitudes: Loudness. Am. J. Psychol., 69(1), 1-25.

[7] Stevens, S. S.(1975) Psychophysics: Introduction to its Perceptual, Neural and Social Prospects. Wiley, New York.

[8] Schaffert, M., Oetjen, A., van de Par, S. (2020). Rauigkeitswahrnehmung wird durch spektrale Prominenz beeinflusst. Fortschritte der Akustik - DAGA 2020, Hannover, 800-803, Dega e.V., Berlin.

[9] `https://www.mathworks.com/matlabcentral/fileexchange/55701-speech-spectrum-shaped-noise`, Stand: 10.3.2020.

[10] Allen, J. B., Berkley, D. A. (1979). Image method for efficiently simulating small room acoustics. J. Acoust. Soc. Am., 65(4), 943-950.

[11] Jot, J.-M., Chaigne, A. (1991). Digital delay networks for designing artificial reverberators. In 90th AES Conv., Audio Eng. Soc., preprint 3030.

[12] Brinkmann, F., Lindau, A., Weinzierl, S., van de Par, S., Müller-Trapet, M., Opdam, R., Vorländer, M. (2017). A High Resolution and Full-Spherical Head-Related Transfer Function Database for Different Head-Above-Torso Orientations. J. Audio Eng. Soc., 65, 841-848.

[13] Berthomieu, G., Koehl V., Paquier M. (2021). Does Loudness Relate to the Strength of the Sound Produced by the Source or Received by the Ears? A Review of How Focus Affects Loudness. Front. Psychol., 12, 1-13.

[14] Eriksson, A., Traunmüller, H. (2002). Perception of vocal effort and distance from the speaker on the basis of vowel utterances. Percept. & Psychophys., 64, 131-9.