

Salienz von Umweltgeräuschen: Welchen Einfluss haben Intensität, zeitlicher Verlauf und spektraler Kontrast?

Jan Selzer¹, Jörg Becker-Schweitzer¹, Michael Oehler², Sabrina Skoda¹

¹ Hochschule Düsseldorf, Institute of Sound and Vibration Engineering, 40476 Düsseldorf,

² Universität Osnabrück, Institut für Musikwissenschaft und Musikpädagogik, 49074 Osnabrück

E-Mail: jan.selzer@hs-duesseldorf.de

Einleitung

Wenn mehrere Umweltgeräusche zeitgleich das menschliche Ohr erreichen, wird meist ein Geräusch stärker wahrgenommen als die Übrigen. Das Hervortreten eines Geräusches aus einem akustischen Kontext wird Salienz genannt. Ob ein bestimmtes Geräusch in einer auditorischen Szene salient oder nicht-salient erscheint und somit vom Hörer mit höherer Aufmerksamkeit bedacht wird, ist von verschiedenen auditorischen Merkmalen abhängig. Die Merkmale *Intensität*, sowie Unterschiede im *Zeitverlauf* oder *Frequenzspektrum* fördern eine unterbewusste Aufmerksamkeitsfokussierung [1]. Die auditive Aufmerksamkeit selektiert mittels Top-Down- und Bottom-Up-Prozessen aus einer Flut von Informationen die salienten auditorischen Objekte und macht sie dem menschlichen Bewusstsein einfacher zugänglich [1][5]. Modelle zur Ermittlung salienter Stimuli basieren auf der Erstellung von Salienzkarten auf Grundlage von Merkmalsextraktion der o.g. Merkmale. Diese können mittels unterschiedlicher Filter nachgebildet werden [2][5][6]. Eine Salienzkarte zeigt die hervorstechenden Stellen innerhalb einer auditorischen Szene in Zeit-Frequenz-Darstellung. Um den Einfluss dieser Merkmale auf die menschliche Salienz bei diotisch dargebotenen „quasi-stationären“ Umweltgeräuschen zu ermitteln, wurde im Rahmen dieser Untersuchung ein Hörversuch durchgeführt. Die Versuchsergebnisse dienen als Basis zur Erstellung eines MATLAB-basierten Modells, welches auf Grundlage der untersuchten Parameter menschliche Salienzprognosen bei auditorischen Szenen bestehend aus Umweltgeräuschen erstellen soll.

Hörversuch

Durch den Hörversuch sollen obige Merkmale überprüft und ggf. weitere Merkmale für das angestrebte Salienzmodell gefunden werden. Diese sollen dazu beitragen in auditorischen Szenen, bestehend aus Umweltgeräuschen, das hervorstechende Geräusch mittels eines Salienzwerts zu ermitteln. Dazu wurde folgende Fragestellung untersucht: Welchen Einfluss haben die Merkmale *Intensität*, *zeitlicher* und *spektraler Kontrast* diotisch dargebotener Umweltgeräusche auf deren Salienz? Welche Erkenntnisse lassen sich aus diesen Merkmalen für den Aufbau eines MATLAB-gestützten Salienzmodells ziehen? Dem Hörversuch lag dabei die Hypothese zugrunde, dass eine deutlichere Ausprägung eines der Merkmale zu einer stärkeren Wahrnehmbarkeit/Salienz des entsprechenden Geräuschanteils führt.

Methodik

Der Hörversuch wurde für Experten aus dem Bereich der Akustik ausgelegt, da die komplexen Versuchsaufgaben spektrale und zeitliche Kontraste verschiedener Reize zu bewerten, eine Durchführung des Versuchs mit geschulten Ohren nahelegte. Mehrere Möglichkeiten wurden diskutiert:

- Verschicken von Audio-Files mit Kalibrierinformationen (beispielsweise im HEAD acoustics- oder PAK File-Format)
- Verschicken des gesamten Hörversuchs für die Plattform *Testing with Ears and Eyes* (iPad) [3]
- Online-Hörversuch

Trotz des Vorteils der reproduzierbaren Wiedergabewege wurde der Hörversuch nicht mit einer der ersteren Möglichkeiten, sondern zugunsten uneingeschränkter Verbreitungswege als Online-Hörversuch durchgeführt, um Erfahrung für weitere Hörversuche auf webbasierten Plattformen zu sammeln. Die in diesem Fall genutzte Plattform war SoSci Survey (<http://www.soscisurvey.de/>).

Neben dem Vorteil der leichten Verbreitungsmöglichkeit führt ein Online-Hörversuch jedoch auch zu Nachteilen, denn eine Kalibrierung der Audioinhalte ist nicht gegeben. Außerdem gibt es Abweichungen in der Art der Wiedergabe: Verschiedene Kopfhörer mit unterschiedlichen Wiedergabesystemen führen zu unterschiedlich physikalisch evozierten Reizen. Auch die räumlichen Gegebenheiten bei den Versuchspersonen vor Ort divergieren. Um dennoch eine Vergleichbarkeit herzustellen, wurde bei dieser Wahrnehmungsaufgabe die Hörschwelle der jeweiligen Testperson als Bezugspunkt gewählt. So sollte jede Versuchsperson vor Beginn des Hörversuchs ihre Wiedergabeeinheit mithilfe eines dafür entwickelten Klangs einstellen. Für diese Aufgabe wurde ein Glockenklingeln ausgewählt, welches in zwei Pegelstufen mit einem Unterschied von 3 dB hintereinander abgespielt wurde (Abbildung 1). So sollte der lautere der beiden Klänge gerade noch wahrgenommen werden – unter besonderer Berücksichtigung der Transienten – der leisere Klang jedoch nicht mehr gehört werden. Zusätzlich wurden, in vertretbarem Rahmen, Beschränkungen für die Hörversuchsdurchführung auferlegt. So wurde von den Versuchspersonen erwartet, den Test in einem ruhigen Raum durchzuführen. Weiterhin wurde ein geschlossener Kopfhörer und eine stufenlos regelbare Wiedergabeeinheit (z.B. externe Soundkarte, USB-Interface o.ä.) benötigt, um den Pegel für die Hörschwelle genau einstellen zu können.

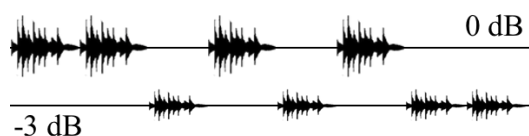


Abbildung 1: Schematische Darstellung des Glockengeräuschs in zwei Pegelstufen. Das lautere der beiden Geräusche sollte gerade so wahrgenommen werden, das leisere nicht mehr.

Der Versuchsablauf war wie folgt: Nach Einstellung des Wiedergabepiegels erhielt der Teilnehmer eine kurze Übersicht über die Fragestellung im Versuch mit Erläuterung dieser. Anschließend begann der Versuch. Alle 21 Szenen (à 12 Sekunden Länge) wurden randomisiert und einmalig dargeboten, um Sequenzeffekte zu vermeiden und eine spontane Bewertung zu erzwingen. Die Versuchsteilnehmer sollten sich die jeweiligen Reize vollständig anhören, das hervorstechende Geräusch mittels eines Dropdown-Menüs wählen und nachfolgende Bewertungen mittels Schieberegler abgeben:

1. Empfundene Lautheit des hervorstechenden Geräuschs (Extrema: laut/leise)
2. Spektrale Unterschiede des hervorstechenden Geräuschs zu den übrigen Geräuschen der Szene mit der Fragestellung: *Wie stark ist der spektrale Unterschied zwischen dem ausschlaggebenden Geräusch und den übrigen Geräuschen?* (Extrema: schwach/stark)
3. Zeitliche Unterschiede des hervorstechenden Geräuschs zu den übrigen Geräuschen der Szene (*Wie stark hat sich das ausschlaggebende Geräusch zeitlich von den übrigen Geräuschen abgehoben?* Extrema: schwach/stark)

Zudem sollte stichpunktartig die Wahl des salienten Geräuschs kurz begründet werden. Nach dem Versuchteil folgten Abschlussfragen zum Alter, Geschlecht, eine Selbstauskunft über Gehörbeeinträchtigungen und zu Kopfhörer/ Wiedergabesystem.

Stimuli

Die im Versuch verwendeten Geräusche wurden teils an der HSD monophon aufgezeichnet und teils aus einer Geräuschdatenbank gewählt und sollten verschiedene Klangeigenschaften abdecken. Es sollten tonale, rauschhafte, impulshafte, aber auch zeitvariante oder Geräusche mit hoher Lautheit für den Versuch genutzt werden. Eine Übersicht der ausgewählten monophonen Umweltgeräusche liefert die folgende Tabelle:

Tabelle 1: Auswahl der Umweltgeräusche

Ozean	Glocke	Möwen	Singvogel	Grill
Motorrad	Verkehr	Lüfter	Uhr	Telefon
Tastatur	Drucker	Computermaus		

Eine auditorische Szene, bestehend aus der Addition dreier Umweltgeräusche, wurde in je drei Varianten realisiert. Dazu wurde eines der Geräusche in jeder Szene in einem Parameter variiert. Einerseits wurde der Pegel in 6 dB-Stufen, beispielsweise beim Geräusch des Singvogels, verändert. Der

zeitliche Verlauf wurde verändert. Zum Beispiel wurde das Geräusch der Uhr in normaler und doppelter Geschwindigkeit, sowie arrhythmisch realisiert. Spektrale Strukturen, beispielsweise im Glockenklingeln, wurden verändert, sodass dieses in einer der Stufen dumpf und perkussiv erscheint und in der letzten Variation tonal ist und ein Nachklingen wahrnehmbar wird. Hinzu kamen drei weitere Szenen, die Geräusche mit nur einer Klangeigenschaft (rauschhaft, impulshaft, tonal, im zweiten Abschnitt der Tabelle 2) enthielten und in keinem der genannten Merkmale variiert wurden. Die auditorischen Szenen sind in Tabelle 2 zusammengefasst. In der ersten Spalte der Tabelle sind die Parameter bezeichnet, in dem das Geräusch aus der zweiten Spalte in drei Stufen variiert wurde.

Tabelle 2: Gestaltete auditorische Szenen

Pegel	Singvogel	Grill	Verkehr
	Lüfter	Uhr	Computermaus
Zeitlich	Motorrad	Möwen	Verkehr
	Uhr	Tastatur	Drucker
Spektral	Glocke	Möwen	Ozean
	Telefon	Tastatur	Verkehr
Rauschhaft	Lüfter	Drucker	Verkehr
Impulshaft	Tastatur	Uhr	Computermaus
Tonal	Glocke	Möwen	Vogel

Ergebnisse und Beobachtungen

Der Versuch war für den Zeitraum von Anfang Oktober bis Ende November 2016 freigeschaltet. Für den Versuch meldeten sich 12 Experten zurück, von denen sechs den Versuch vollständig durchführten (m: 5, w: 1, $\bar{\sigma}$ Alter: 40,5, SD: 10,7). Eine statistisch aussagekräftige Auswertung ist aufgrund dieser kleinen Stichprobe nicht sinnvoll. Jedoch kann auf Basis der vorhandenen Ergebnisse bereits eine Tendenz bereitgestellt werden. So wurden mehrere Geräusche teilweise einstimmig, teilweise mehrheitlich als hervorstechend gekennzeichnet, die sich über die psychoakustischen Parameter Lautheit, Schärfe und Tonhaltigkeit, sowie über die textliche Begründung der Versuchspersonen erklären lassen. Die Geräusche „Singvogel“, „Lüfter“ und „Telefon“ (vgl. Tabelle 3) besitzen einen hohen Lautheitswert gegenüber den übrigen Geräuschen der jeweiligen Szene, was sich auch in der Bewertung der Lautheit durch die Versuchsteilnehmer widerspiegelt. Das Geräusch „Telefon“ weist gleichzeitig den höchsten Schärfewert des Gesamtreizes auf. In der Tonhaltigkeit (ΔL nach DIN 45681 [5]) stechen die Geräusche „Lüfter“, „Glocke“ und wiederum „Telefon“ in ihrer jeweiligen Szene hervor. Das Geräusch „Motorrad“ sticht in keinem der ausgewerteten psychoakustischen Parameter hervor, wurde dennoch als salient gekennzeichnet. Dies wurde von den Versuchsteilnehmern in der Bewertung des Parameters „Zeit“ deutlich gemacht und in der textlichen Begründung durch die „Periodizität und Technizität“ des Signals begründet. Allerdings kann davon ausgegangen werden, dass bei diesem Geräusch Top-down-Prozesse Einfluss auf die Auswahl hatten. Der „Glocken“-Gesamtstimulus ist ebenfalls interessant. Der Glockenklang ist besonders tonhaltig, wurde dennoch durch die Versuchsteilnehmer aufgrund der „Periodizität“ und des „An

und Aus“-Charakters des Geräuschs als salient begründet. Diese Tendenzen können bereits in dem auf diesem Versuch basierenden Saliensmodell verwendet werden. Dabei werden die Merkmale *Intensität*, *spektrale* und *zeitliche Kontraste* wie in [2] und [5] beschrieben, beibehalten. Zusätzliche Merkmale zur Berechnung der Salienswerte werden auf Basis der Tendenzen hinzugezogen. So kann u.a. die Periodizität und die Tonhaltigkeit ebenfalls als mitverantwortliches Merkmal in das Saliensmodell überführt werden.

Tabelle 3: Psychoakustische Parameter der Umweltgeräusche, Auswahl der salient bestimmten Geräusche, je an erster Stelle in jeweiliger Szene.

Szene	Geräusch	Lautheit [sone]	Schärfe [acum]	Tonhaltigkeit [dB]
1	Singvogel	6,1	2,7	0,0
	Grill	2,2	4,4	0,0
	Verkehr	3,6	1,5	0,0
2	Lüfter	7,2	1,2	19,0
	Uhr	1,7	0,4	2,4
	Computer-maus	2,1	2,1	0,0
3	Motorrad	4,7	2,0	0,0
	Möwen	6,4	2,9	0,0
	Verkehr	3,6	1,5	0,0
5	Glocke	7,0	1,7	27,5
	Möwen	6,4	2,9	0,0
	Ozean	9,5	2,6	0,0
6	Telefon	5,5	3,5	25,3
	Tastatur	2,6	2,9	0,9
	Verkehr	3,6	1,5	0,0

Saliensmodell

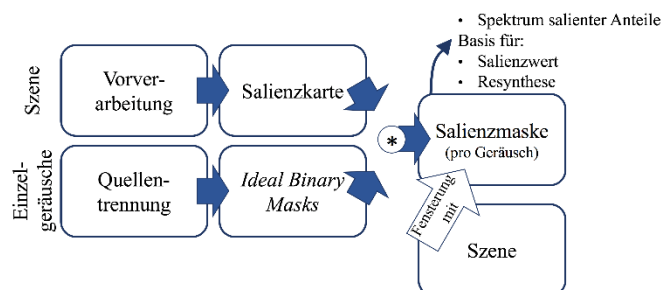


Abbildung 2: Schematische Darstellung des Saliensmodells

Das Saliensmodell lässt sich in zwei parallele Zweige unterteilen. Zu Beginn steht die gehörgerechte Vorverarbeitung des vorliegenden Gesamtgeräusches und Überführung dessen in den Zeit-Frequenz-Bereich. Dazu wird eine *Gammatone*-Filterbank mit 128 Filtern verwendet, die gleichmäßig auf der ERB-Skala im hörbaren Bereich von 20 Hz bis 20 kHz verteilt sind [5]. Anschließend wird für diese Darstellung des Signals eine Salienskarte erstellt, die sich grundlegend nach den Erkenntnissen von Kayser et al., Kalinli et al. und De Coensel et al. richtet [2][5][6].

Parallel dazu wird eine Quellentrennung mittels der Spektrogramme der Einzelsignale durchgeführt, sodass pro Geräusch eine Maske (*Ideal Binary Mask*) erstellt wird.

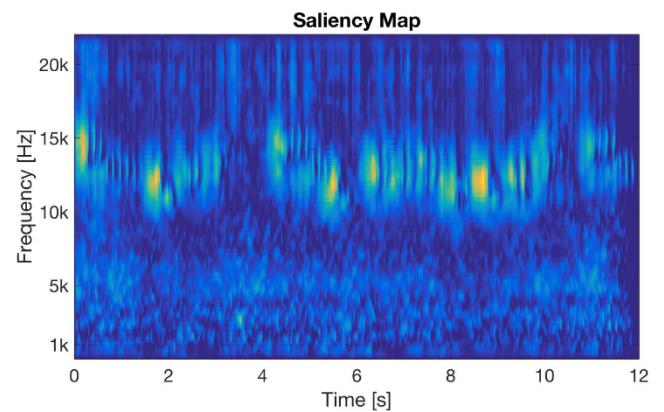


Abbildung 3: Salienskarte der Geräuschumgebung „Singvogel, Grill, Verkehr“. In der dritten Dimension werden die salienten Anteile im Gesamtgeräusch angezeigt.

Durch die Multiplikation der Masken mit der Salienskarte entstehen Saliensmasken für jedes auditorische Objekt. Diese beinhalten die salienten Anteile des jeweiligen Geräusches aus der Szene. Werden diese Saliensmasken auf das Gesamtgeräuschszenario angewandt, so können anschließend die Spektrogramme der salienten Anteile der einzelnen Geräusche betrachtet werden. Auf Basis dieser Spektrogramme kann zusätzlich eine Resynthese der einzelnen Geräusche durchgeführt werden.

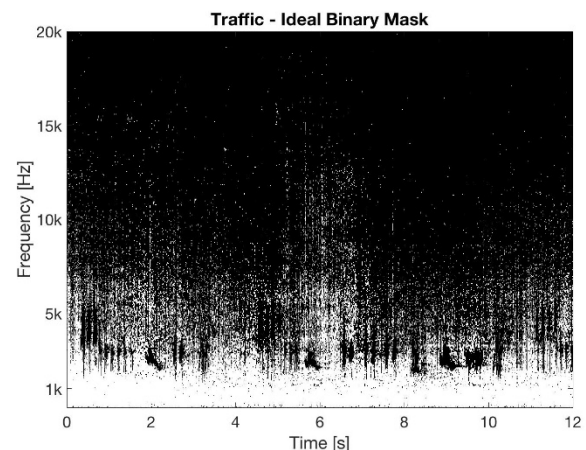


Abbildung 4: *Ideal Binary Mask* des Verkehrsgeräusches aus der Geräuschumgebung „Singvogel, Grill, Verkehr“. Weiß zeigt die „durchlässigen“ Stellen mit Wert 1, schwarz den Wert 0 an.

Die gesamte Quellentrennung und Resynthese wird mittels der *IoSR*-Toolbox [7] durchgeführt. Bei Untersuchung von monauralen Stimuli kommt es dazu, dass die Quellentrennung nicht einwandfrei funktioniert, wodurch Signalanteile anderer Signale in der Resynthese vorkommen. Nach der Resynthese der einzelnen Geräusche der Szene „Singvogel, Grill, Verkehr“ wird beispielsweise deutlich, dass im breitbandigen Verkehrsgeräusch die salienten Anteile dieses Geräusches durch das Signal des Vogels hervorgerufen werden (Abbildung 4). Die gleiche Beobachtung kann in der Resynthese und im Spektrum des Grillgeräusches gemacht werden. Der Grund für diese Abweichungen in der Maskenerstellung ist die partielle Verdeckung einzelner Zeit-Frequenz-Blöcke durch die breitbandigen Geräusche. Weisen diese in dem jeweiligen Block einen höheren Amplitudenwert

auf, so wird der Block dem Objekt mit dem höheren Wert zugeordnet. Dabei beinhaltet dieser Block zusätzliche Informationen des eigentlich salienten Signals. Auf Basis der Saliensmasken werden die Salienswerte für jedes Objekt gebildet (Einzahlwert, einheitenlos, in den Grenzen zwischen 0 und 1). Ein Objekt, welches innerhalb einer auditorischen Szene als hervorstechend wahrgenommen wird, soll den höchsten Salienswert dieses Szenarios erhalten. Aufgrund der Maskenerstellung entsteht hier eine Diskrepanz zwischen dem eigentlichen salienten Geräusch und der Bewertung dessen durch den Computer. Um das saliente Geräusch auch mit dem höchsten Wert zu beziffern, sollen zusätzliche akustische Merkmale zur Ermittlung des Salienswerts beitragen. Durch die häufige Nennung der zeitlichen Komponente, insbesondere als „*Periodizität*“, wurde diese als weiteres Merkmal hinzugefügt. Wird eine Periodizität (ermittelt über Autokorrelation des Signals) detektiert, so erfährt der Salienswert eine Aufwertung von bis zu 0,2 Punkten. Die *Fluktuation des Schalldruckpegels über der Zeit* wurde ebenfalls im Versuch für die Detektion eines salienten Schalls benannt. Daher führt dieses zusätzliche Merkmal ebenfalls zu einer Aufwertung des Salienswertes. Keines der Geräusche im Versuch wurde aufgrund seiner *Breitbandigkeit* als salient bezeichnet. Der Salienswert liefert nach Berechnung mit der Saliensmaske jedoch häufig hohe Werte für ebendiese Geräusche, daher wird bei der Ermittlung von Breitbandigkeit der Salienswert des Streams um maximal 0,2 Punkte abgewertet. Ein weiteres Merkmal ist die Tonhaltigkeit eines Geräusches. So wurde diese nicht explizit durch die Versuchsteilnehmer als Grund für die Aufmerksamkeitsfokussierung genannt, jedoch sind häufig Geräusche mit hoher Tonhaltigkeit auch als salient gekennzeichnet worden. Die Berechnung der Tonhaltigkeit wird mit dem *PATool* der Universität Erlangen-Nürnberg direkt in MATLAB durchgeführt [8]. So können bereits Saliensprognosen in auditorischen Szenen, bestehend aus Umweltgeräuschen erstellt werden.

Zusammenfassung und Ausblick

Ein Hörversuch zur Bestimmung des Einflusses der akustischen Merkmale *Intensität*, *zeitlicher* und *spektraler Unterschiede* wurde online durchgeführt. Die Erkenntnisse aus dem Versuch zeigen, dass stets mehr als ein Merkmal für das Hervorrufen von Aufmerksamkeit verantwortlich ist. Um diese Aussage statistisch überprüfbar zu machen, muss eine größere Stichprobe untersucht werden. Anhand der Versuchsergebnisse konnten weiterhin Tendenzen bzgl. der Charakterisierung eines salienten Geräusches bestimmt werden. Diese Tendenzen wurden mit den psychoakustischen Parametern Lautheit, Schärfe und Tonhaltigkeit verknüpft, sodass anhand dieser Prädiktoren die Aufmerksamkeit prognostiziert werden kann. Ab- und Zuschläge auf die Salienswerte werden auf Basis der zusätzlichen Merkmale *Periodizität*, *Tonhaltigkeit*, *Fluktuation des Schalldruckpegels über der Zeit* und *Breitbandigkeit* berechnet. Berechnungsgrundlage für die Salienswerte sind die Saliensmasken. Diese werden durch Multiplikation der *Ideal Binary Masks* jedes einzelnen Geräusches aus der auditorischen Szene mit der Salienskarte der Szene berechnet.

Die neu hinzugefügten Merkmale sind auf rein experimenteller Basis, aufgrund vermehrter Nennungen im Hörversuch, erstellt worden. Inwiefern die Ab- und Zuschläge auf den Salienswert der realen Aufmerksamkeit entsprechen, muss in zukünftigen Studien überprüft werden. So sollen einzelne Versuche gestaltet werden, die den Einfluss auf die Saliens ebendieser Merkmale validieren und überprüfen. Zusätzlich muss das Modell anhand von weiteren auditorischen Szenen mit anderen Arten von Umweltgeräuschen auf seine Gültigkeit überprüft werden. Zukünftige Untersuchungen können zeigen, inwiefern die Lästigkeit oder Angenehmheit eines Geräusches dazu beiträgt, ob dieses als salient eingestuft wird. Auch den Einfluss von Persönlichkeitseigenschaften und Lärmempfindlichkeit auf die Charakterisierung spezifischer Geräusche als salient gilt es weiter zu untersuchen. Der Ausbau dieses Modells auf eine binaurale Saliensdetektion würde die Quellentrennung verbessern und dazu beitragen, genauere Salienswerte zu bilden.

Literatur

- [1] Steffens, J.: Wie viel Realität braucht der Mensch? Dissertation TU Berlin, 2013.
- [2] Kayser, C., Petkov, C. I., Lippert, M., Logothetis, N. K.: Mechanisms for Allocating Auditory Attention: An Auditory Saliency Map. *Current Biology* 15, 2005, 1943–1947.
- [3] Nievelstein, F., Leckschat, D., Epe, C., Skoda, S.: Erstellung einer intuitiven, modularen Versuchsumgebung für wissenschaftliche Hörversuche in iOS. *DAGA 2015*, 1300–1301.
- [4] DIN 45681: Akustik – Bestimmung der Tonhaltigkeit von Geräuschen und Ermittlung eines Tonzuschlages für die Beurteilung von Geräuschmissionen, Beuth Verlag, 2005.
- [5] De Coensel, B. und Botteldooren, D.: A model of saliency-based auditory attention to environmental sound. *ICA 2010*, 1–8.
- [6] Kalinli, O. und Narayanan, S.: A Saliency-Based Auditory Attention Model with Applications to Unsupervised Prominent Syllable Detection in Speech. *Interspeech 2007*, 1941–1944.
- [7] Institute of Sound and Recording, University of Surrey: Matlab Toolbox (Version 2.4.1) for auditory modelling, signal processing, sound source separation, statistics, plotting, etc.
URL: <http://www.tonmeister.co.uk/software/index.php/>
- [8] Münsterjohann, S., Becker, S., Grohmann, T., Dittmar, R.: Contribution of Acoustic and Structural Sources to the Tonal Sound Radiation of Side Channel Blowers. *Proceedings of ASME Turbo Expo*, 2014.