Modell zur Detektion und Quantifizierung von zeitlich variablen tonalen Komponenten

Eike Claaßen[1], Arne Oetjen, Steven van de Par,

[1] Carl-von-Ossietzky Universität Oldenburg, Acoustics Group, 26129 Oldenburg, Deutschland, Email: eike.claassen@uol.de

Einleitung

Die Tonhaltigkeit ist eine wichtige psychoakustische Kenngröße zur Beurteilung von Fahrzeuggeräuschen. Gängige Verfahren wie die DIN 45681-2005 sind jedoch durch die Länge der Analysefenster von 0,25 bis 0,53 s ungeeignet für die Analyse von nicht-stationären, tonhaltigen Komponenten [4]. Im Zuge dieser Arbeit wird die Erweiterung eines bestehenden Modells zur Detektion von Tonhaltigkeit vorgestellt und anhand von zwei Hörversuchen verifiziert. Der erste Hörversuch untersucht die wahrgenommene Tonhaltigkeit von Schmalbandrauschen in Abhängigkeit der Bandbreite, der Filtermittenfrequenz und der Filterart. Im zweiten Hörversuch wird die Tonhaltigkeit von Fahraufnahmen eines Elektrofahrzeugs untersucht.

Tonhaltigkeit von Schmalbandrauschen

In einem Hörversuch wird die wahrgenommene Tonhaltigkeit von Schmalbandrauschen(„SBR“) untersucht. Tonhaltige Signale gelten als spektral schmalbandig. Ziel dieses Hörversuchs ist es, den Einfluss verschiedener Frequenzbereiche, Bandbreiten und Filterarten auf die wahrgenommene Tonhaltigkeit zu untersuchen. Des Weiteren dienen die Erkenntnisse dieses Versuch für die spätere Parametrisierung eines Modells zur Detektion und Quantifizierung von Tonhaltigkeit.

Versuchspersonen

Es wurden acht Versuchspersonen („VP“) mit einem mittleren Alter von 26,2 Jahren in einem Altersbereich von 24 bis 30 Jahren untersucht. Die acht VPs setzen sich aus fünf weiblichen und drei männlichen zusammen. Alle VPs verfügen über ausgeprägte Erfahrung in der Durchführung von Hörversuchen und sind nach eigenen Angaben normalhörend.

Stimuli

Der Hörversuch umfasst 14 Stimuli. Die Stimuli setzen sich aus 12 SBR zusammen, die mittels Butterworth-Bandpassfilter 4. Ordnung gefiltert wurden. Die zwei verbliebenen Stimuli wurden mittels eines FFT-Bandpassfilters bzw. eines FFT-Tiefpassfilters gefiltert. Für die Butterworth-Bandpassfilter wurden Filtermittenfrequenzen f_c von 500, 1000 und 2000 Hz gewählt. Die Bandbreiten Δf für jedes f_c betrugen 10 Hz, 50 Hz, 1 bark und 2 bark. Der FFT-Tiefpassfilter wies eine f_c von 1000 Hz und eine Δf von 50 Hz auf. Für den FFT-Tiefpass wurde eine Grenzfrequenz f_g von 500 Hz gewählt. Alle Stimuli waren 2 s lang und wurden auf einen Schalldruckpegel von 65 dB SPL kalibriert und mit einer zufälligen Variation um ± 2,5 dB dargeboten.

Durchführung

Der Hörversuch findet in einer akustisch abgeschirmten Hörkabine statt. Die Darbietung der Stimuli erfolgt mittels eines Sennheiser HD 650 Kopfhörers. Auf Basis der Point-of-Subjective-Equality („PSE“) werden der VP in Form eines Paarvergleichs jeweils einer der zuvor beschriebenen Stimuli sowie ein Referenzgeräusch dargeboten. Das Referenzgeräusch besteht aus einem Sinuston mit einer festen Frequenz von 1 kHz und einem Breitbandrauschen. Der Schalldruckpegel betrug 65 dB SPL und wurde synchron zum Stimulus mit einer zufälligen Variation um ± 2,5 dB dargeboten. Das Anpassungskriterium des Hörversuchs ist der Signalrauschabstand („SNR“). Die Aufgabe der VP ist die Beurteilung, ob das dargebotene Stimulusgeräusch oder das Referenzgeräusch stärker tonhaltig ist. Innerhalb des Hörexperiments fanden fünf Wiederholungen für die Schwellenbestimmung durchgeführt. Der Hörversuch wurde von jeder VP zwei mal, an unterschiedlichen Tagen durchgeführt. Die Instruktion der VP erfolgt anhand eines festgelegten Instruktionstextes, im Anschluss wurden mittels einer grafischen Nutzeroberfläche Einführungsbeispiele bereitgestellt.

Auswertung

Um VPs mit inkonsistentem Antwortverhalten von der Auswertung auszuschließen wurde ein Ausschlusskriterium basierend auf der mittleren absoluten Abweichung („MAD “) gebildet. VPs mit einer MAD $\geq$ 15 dB wurden von der weiteren Auswertung ausgeschlossen, dies betraf zwei der acht VPs. Die Ergebnisse des Hörversuchs sind in Abbildung 1 dargestellt. Es ist zu erkennen, dass die wahrgenommene Tonhaltigkeit einer Sättigung zu hohen Bandbreiten unterliegt. Dies zeigt sich bei den Bandpassrauschen mit Bandbreiten von 1 und 2 bark, sowie dem Tiefpassrauschen. Hinsichtlich der maximalen wahrgenommenen Tonhaltigkeit zeigen sich keine Unterschiede bzgl. der f_c zwischen den Bandpassrauschen mit einer Δf von 10 Hz. Für die Bandpassrauschen mit einer Δf von 50 Hz zeigt sich eine Zunahme der wahrgenommen Tonhaltigkeit hin zu höheren f_c. Ein Vergleich der Bandpassrauschen mit einer f_c von 1 kHz und einer Δf von 50 Hz zeigt, dass der mittels FFT-Bandpass gefilterte Stimulus als stärker tonhaltig wahrgenommen wird, als der mittels Butterworth-Bandpass gefilterte Stimulus.

Tonhaltigkeit von Fahrzeuggeräuschen

In einem zweiten Hörversuch wird die Tonhaltigkeit von Fahrzeuggeräuschen untersucht. Wie bereits beschrieben stellt die Tonhaltigkeit eine wichtige Beurteilungsgröße bei der Auswertung akustischer Emissionen von Fahrzeugen dar. Die verwendeten Fahraufnahmen stammen aus der Aufnahme eines Beschleunigungsvorganges eines Elektrofahrzeugs, simultan gemessen an fünf verschiedeneren Positionen.

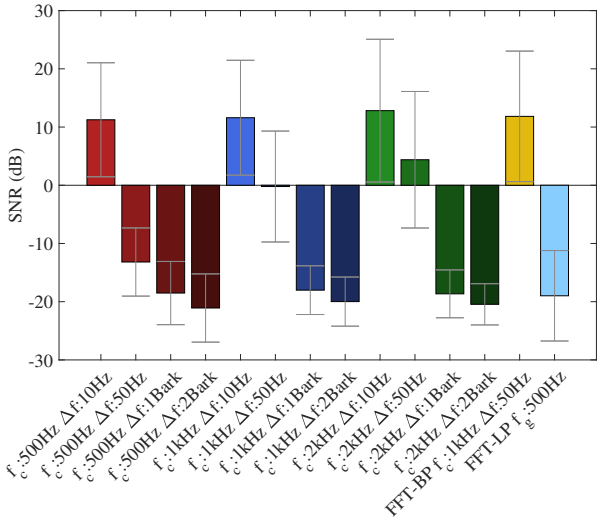

Abbildung 1: Mittelwert für den SNR über beide Messdurchläufe mit Standardfehler über die sechs ausgewerteten VPs für die Untersuchung zur Tonhaltigkeit von SBR.

Versuchspersonen

Für den Hörversuch wurden 15 VPs mit einem mittleren Alter von 26,8 Jahren in einem Altersbereich von 21 bis 30 Jahren untersucht. Die 15 VPs setzen sich aus acht weiblichen und sieben männlichen zusammen. 10 VPs verfügen über ausgeprägte Erfahrung in der Durchführung von Hörversuchen. Alle VPs sind nach eigener Angabe normalhörend.

Stimuli

Der Hörversuch umfasst 14 Stimuli. 10 Stimuli stammen aus Fahraufnahmen eine Elektrofahrzeugs, die während eines Beschleunigungsvorgangs an fünf verschiedenen Messpositionen aufgenommen wurden. Die Messpositionen umfassen das Chassis, das Getriebe, den Inverter, das Nahfeld des Inverters sowie die Kabine. Die Stimuli wurden nach DIN 45631 in ihrer Lautheit normalisiert. Die vier weiteren Stimuli entstammen dem Hörversuch zur Ermittlung der Tonhaltigkeit von SBR. Ausgewählt wurden hier die mittels Butterworth-Bandpass gefilterten SBR bei einer Δf von 50 Hz und einer f_c von 500 Hz, 1000 Hz und 2000 Hz, sowie das mittels FFT-Tiefpass gefilterte Rauschen mit einer f_g von 500 Hz. Alle Stimuli waren 2 s lang und wurden auf einen Schalldruckpegel von 65 dB SPL kalibriert und mit einer zufälligen Variation um ± 2,5 dB dargeboten.

Die Spektrogramme der Fahraufnahmen des Elektrofahrzeugs sind in Abbildung 2 dargestellt. Insbesondere die Messpositionen am Chassis und dem Inverter Nahfeld zeigen hier ausgeprägte hochfrequente Anteile.

Durchführung

Die Durchführung des Hörversuchs erfolgt analog zur der des Hörversuchs zur Tonhaltigkeit von SBR.

Auswertung

Die Auswertung der Rohdaten zeigte, dass sich die VPs hinsichtlich ihrer Bewertung in zwei Gruppen unterscheiden ließen. Eine Gruppe bewertete die Testkonditionen vorrangig im positiven SNR-Bereich, die andere Gruppe bewertete ausschließlich im negativen SNR-Bereich. Ein weiterer Unterschied zwischen den beiden Gruppen lag im Dynamikbereich der Bewertungen. Als Dynamikbereich wird der dB-Wert zwischen der maximalen und der minimalen Bewertung des SNR-Pegels über alle Testkonditionen bezeichnet. Die Teilnehmergruppe welche die Stimuli ausschließlich im negativen Bereich bewertete wies einen sehr niedrigen Dynamikbereich über alle Testkonditionen aus. Um die Aussagekraft der Auswertung zu erhöhen wurde ein neues Ausschlusskriterium bzgl. des Dynamikbereichs festgelegt. VPs, deren Dynamikbereich unterhalb des halben mittleren Dynamikbereichs über alle VPs (12,9 dB) liegen, werden von der weiteren Auswertung ausgeschlossen. Durch Anwendung dieses Kriteriums wurden fünf VPs ausgeschlossen. Das aus dem Hörversuch zur Bestimmung der Tonhaltigkeit von SBR entnommene Ausschlusskriterium hinsichtlich der mittleren absoluten Abweichung von maximal 15 dB wurde von allen VPs unterschritten. Die Auswertung des Hörversuchs ist in Abbildung 3 dargestellt. Die Auswertungen der Fahrzeugaufnahmen zeigt, dass die Messpositionen am Chassis und am Inverter Nahfeld als am stärksten Tonhaltig wahrgenommen wurden. Es handelt sich hierbei um die Geräusche, die, wie in Abbildung 2 zu erkennen, den stärksten hochfrequenten Anteil aufweisen. Als schwächer tonhaltig wahrgenommen werden die Messpositionen Getriebe, Kabine und Inverter.
Unter den synthetischen Geräuschen, die aus dem Hörversuch zur Tonhaltigkeit von SBR entnommen wur-

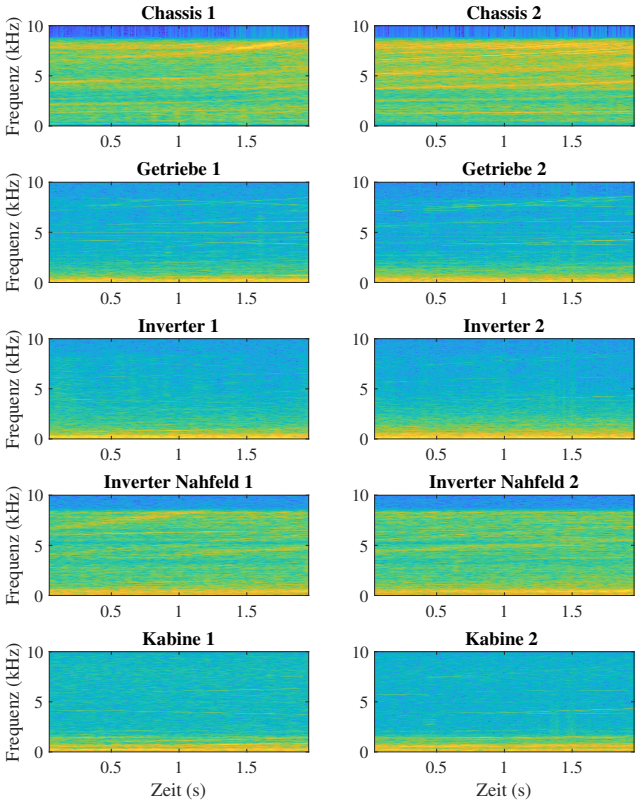

Abbildung 2: Spektrogramme der Fahraufnahmen, aufgenommen beim Beschleunigungsvorgang eines Elektrofahrzeuges an fünf verschiedenen Messpositionen.

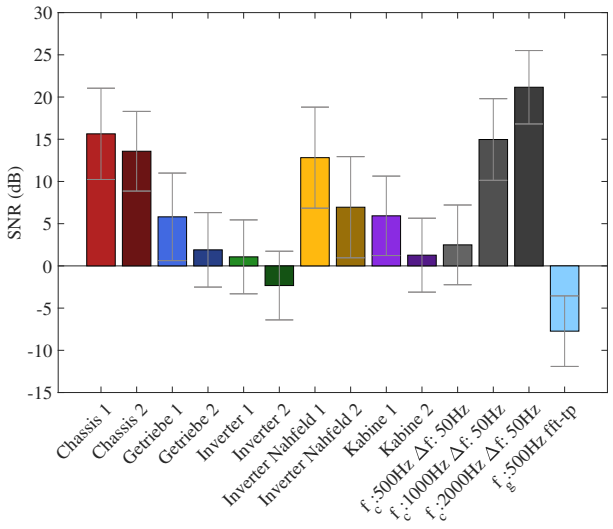

Abbildung 3: Mittelwert für den SNR über beide Messdurchläufe mit Standardfehler über die 10 ausgewerteten VPs. Die x-Achse zeigt die 14 verschiedenen Messkonditionen.

den, zeigte sich, dass die Ergebnisse des Hörversuchs tendenziell verifiziert wurden. Unter den mittels Butterworth-Bandpass gefilterten SBR mit einer Δf von 50 Hz bestätigt sich die bereits zuvor festgestellte Zunahme der wahrgenommenen Tonhaltigkeit hin zu höheren f_c. Das mittels FFT-Bandpass gefilterte Tiefpassrauschen wurde als am wenigsten tonhaltig wahrgenommen. Auffällig ist, dass der Wertebereich der SNR-Beurteilungen um ca. 15,2 - 16,8 dB vom ersten Hörversuch abweicht. Ursächlich hierfür kann zum einen die Auswahl der VPs sein, die nicht wie beim ersten Hörversuch ausschließlich aus erfahrenen VPs bestehen. Zum anderen betraf das Ausschlusskriterium hinsichtlich des Dynamikbereichs ausschließlich VPs, die alle Stimuli deutlich im negativen SNR-Bereich bewertet haben.

Modellierung von Tonhaltigkeit

Der in dieser Arbeit gewählte Ansatz zur Modellierung von Tonhaltigkeit basiert auf einer früheren Arbeit von Peter Volk [1]. Das Modell basiert auf dem Perception Model (PEMO) von Dau et al. [2]. Aus dem PEMO entnommen werden die Gammatone-Filterbank und die Halbwellengleichrichtung. Ein Überblick über das gesamte Modell ist in Abbildung 4 dargestellt.

Im ersten Schritt erfolgt die Filterung mittels einer Gammatone-Filterbank in die einzelnen Frequenzgruppen. Die Bandbreite der einzelnen Filter beträgt 0,75 ERB, mit jeweils 5 Filtern pro 0,75 ERB. Hieraus resultieren 191 Filter. Im zweiten Schritt wird die Einhüllende x_i' als Betrag des analytischen Signals des Ausgangssignals x_i, des jeweiligen auditorischen Filters berechnet:

$$x_i' = |x_i + \mathcal{H}(x_i), | \tag{1}$$

wobei $\mathcal{H}(x_i)$ die Hilberttransformierte des Signals x_i bezeichnet. Die Einhüllende wird mit einem Butterworth Tiefpassfilter 2. Ordnung bei einer Grenzfrequenz von 0,2 bark relativ zur Mittenfrequenz des auditorischen

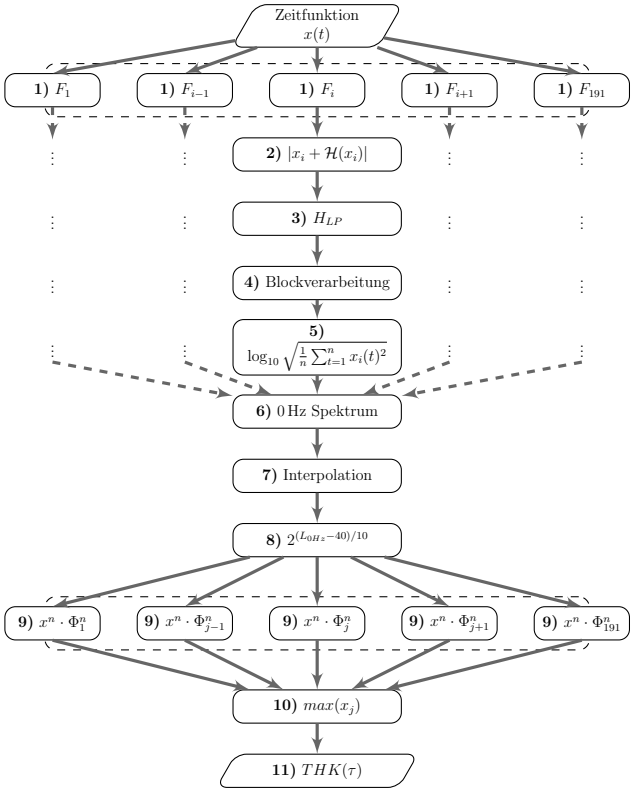

Abbildung 4: Flussdiagramm des Modells zur Detektion und Quantifizierung von Tonhaltigkeit (THK).

Bands gefiltert. Für die zeitliche Verarbeitung des Signals wird eine Blocklänge von 50 ms mit 25 % Überlapp verwendet, hieraus ergeben sich Zeitschritte von τ=12,5 ms. Für alle 191 Frequenzbänder und jeden Zeitschritt wird der Pegel des Effektivwerts L_i über die n Abtastwerte des Zeitausschnitts gemäß

$$L_i = \log_{10}\sqrt{\frac{1}{n}\sum_{t=1}^{n} x_i(t)^2} \tag{2}$$

berechnet. Die Einzahlwerte aus jedem Frequenzband werden anschließend zu einem Vektor L_{0Hz} zusammengesetzt, der die Stärke des tiefpassgefilterten Erregungsmusters über die Frequenz, das sogenannte „0 Hz Spektrum“ zeigt, es gilt

$$L_{0Hz}(i) = L_i. \tag{3}$$

Die Auflösung der 0 Hz Spektren wird durch kubische Interpolation um den Faktor 10 erhöht. Im Anschluss werden die Pegelwerte L_{0Hz} in ein Maß N_{0Hz} umgerechnet, das sich an der Lautheit orientiert. Hierbei wurde nach Fastl und Zwicker [3] die vereinfachte Beziehung

$$N_{0Hz} \approx 2^{(L_{0Hz}-40)/10} \tag{4}$$

verwendet. Mithilfe des Skalarproduktes wird anhand des aktuellen Erregungsspektrums x und einer Template-Datanbank tonhaltige Komponenten detektiert und quantifiziert. Die verwendete Template-Datenbank besteht aus den 0 Hz Erregungsmustern von 191 Sinustönen mit den Mittenfrequenzen der Frequenzgruppen. Vorher

wird die fünfte Potenz der Spektren berechnet (n=5), es gilt

$$x'(j) = x^n \cdot \Phi_j^n. \quad (5)$$

Aus der resultierenden Ähnlichkeitsmatrix x' wird anschließend der zeitliche Verlauf der detektierten Tonhaltigkeit durch Maximalwertbildung über die Frequenzgruppen gebildet. Ein geeigneter Einzahlwert für die Tonhaltigkeit eines Gesamtsignals ergibt sich durch die Mittelung des zeitlichen Verlaufs der Tonhaltigkeit.

Verifizierung des Modells

Zur Verifizierung des Modells werden die Stimuli und Referenzgeräusche der durchgeführten Hörversuche durch das Modell analysiert und mit den Ergebnissen der VPs verglichen. Die Stimuli umfassen die den VPs dargebotenen Signale. Die Referenz-Geräusche entsprechen einem 1 kHz Sinuston in Breitbandrauschen, dessen SNR entsprechend der Ergebnisse des jeweiligen Hörversuchs angepasst wurde.

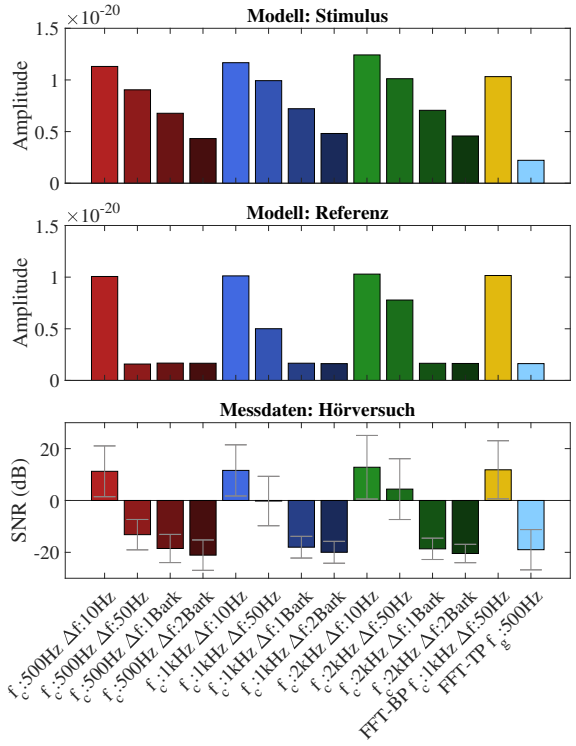

Abbildung 5: Vergleich der Modell-Ergebnisse mit den Ergebnissen des Hörversuchs zur Tonhaltigkeit von SBR.

Die Ergebnisse in Abbildung 5 und 6 zeigen, dass insbesondere die Beurteilung der Stimuli Übereinstimmungen mit den Ergebnissen des Hörversuchs aufweisen. So kann die Abhängigkeit der Tonhaltigkeit von der Δf gut für Stimuli mit SBR nachgebildet werden, die Abhängigkeit von der f_c ist jedoch sehr schwach ausgeprägt. Die Referenzgeräusche zeigen ein starkes Sättigungsverhalten in Abhängigkeit der Δf. Die Modelldaten der Stimuli der Fahrzeuggeräusche zeigen ebenfalls Übereinstimmungen mit dem Hörversuch. Die relativen Unterschiede der Tonhaltigkeit werden hierbei außer für das Inverter Nahfeld

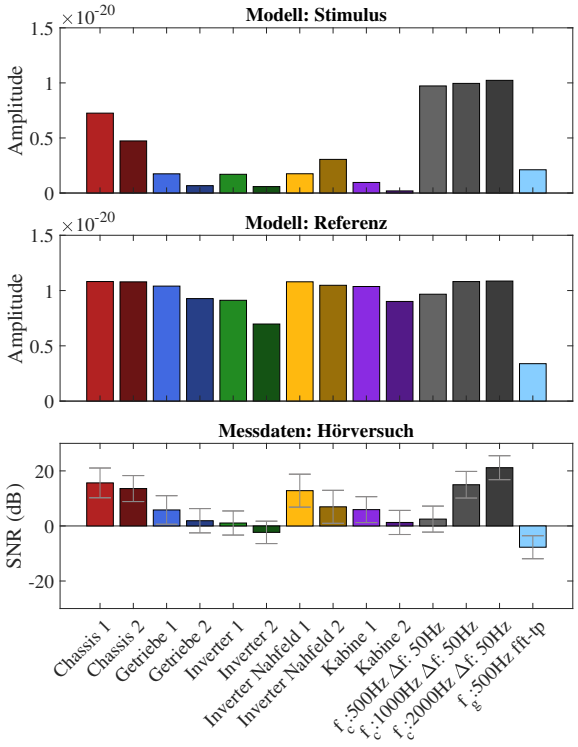

Abbildung 6: Vergleich der Modell-Ergebnisse mit den Ergebnissen des Hörversuchs zur Tonhaltigkeit von Fahrzeuggeräuschen.

korrekt nachgebildet. Die Referenzkonditionen zeigen jedoch abermals größere Abweichungen.

Fazit

Die Hörversuche zeigen, dass die wahrgenommene Tonhaltigkeit von SBR primär von der Δf abhängig ist. Weitere Einflussfaktoren sind die f_c, sowie die Flankensteilheit des zugrundeliegenden Filters. Bei der Beurteilung von Fahrzeuggeräuschen zeigt sich ein deutlicher Unterschied in der wahrgenommenen Tonhaltigkeit in Abhängigkeit der Messposition. Besonders Aufnahmen am Chassis und dem Inverter Nahfelt wurden als stark tonhaltig wahrgenommen. Das optimierte Modell zur Beurteilung von Tonhaltigkeit zeigt gute Übereinstimmung insbesondere mit den Stimuli, während die Beurteilung der Referenzgeräusche größere Abweichungen aufweisen.

Literatur

[1] Volk, P.: Wahrnehmung & Ausgeprägtheit zeitlich variabler, tonaler Signale: Hörversuche und Modellierung, (2004)

[2] Dau, T.; Püschel, D.;Kohlrausch, A. .: A quantitative model of the effective signal processing in the auditory system. i. model structure. The Journal ofthe Acoustical Society of America, (1996), S. 3615—3622

[3] Fastl, H.; Zwicker E., Psychoacoustics - Facts and Models (Springer-Verlag Berlin), 3 ed.

[4] Oetjen, A. ; Par, S. van d. ; Weber, R.; Letens, U.: Verbesserte Tonhaltigkeitsberechnung für instationäre Geräusche.; Proc DAGA, 2014