

Nichtlineare Kennlinien zur Verbesserung der Sprachverständlichkeit in geräuschbehafteter Umgebung

Philipp Bulling¹, Klaus Linhard¹, Arthur Wolf¹, Gerhard Schmidt², Anne Theiß², Marco Gimm²

¹ Daimler AG, 89081 Ulm, Deutschland, Email: {philipp.bulling, klaus.linhard, arthur.wolf}@daimler.com

² Christian-Albrechts-Universität zu Kiel, 24143 Kiel, Deutschland, Email: {gus, ath, mgi}@tf.uni-kiel.de

Einleitung

Die nichtlineare Verzerrung eines Audiosignals wird in der Regel zu den unerwünschten Effekten bei der Signalverarbeitung gezählt. Allerdings gibt es Anwendungen, bei denen nichtlineare Verzerrungen gewünscht sind und zur positiven Klanggestaltung beitragen sollen. Verzerrung bedeutet hier, dass zu einem Grundton ganzzahlige Vielfache seiner Grundfrequenz hinzugefügt werden. Bei einem Grundton (= 1. Harmonische) von beispielsweise 1 kHz ist die 2. Harmonische ein Ton bei 2 kHz, die 3. Harmonische bei 3 kHz, usw. Ein klassisches Beispiel ist ein Röhrenverstärker. Dieser erzeugt bei starker Aussteuerung nichtlineare Verzerrungen mit überwiegend geradzahligem Harmonischen. Letztere werden oftmals als angenehm und warm empfunden. Darüber hinaus gibt es eine Reihe von Effektgeräten, die gezielt nichtlineare Verzerrungen erzeugen. Effektgeräte dieser Art werden oft vereinfacht Verzerrer genannt (vor allem für die elektrische Gitarre). Ein anderer Name ist *Exciter*, womit angedeutet wird, dass künstliche Obertöne (weitere Harmonische) erzeugt werden, die im Original u. U. gar nicht vorhanden waren. Der erste Exciter wurde von der Firma Apex in den 70er Jahren entwickelt. Exciter werden vor allem in der Musikproduktion für Gesang und verschiedene musikalische Instrumente eingesetzt. Die psychoakustischen Effekte des Exciters werden in [1] untersucht. Eine digitale Modellierung der analogen Hardware wird z. B. in [2] beschrieben.

In der vorliegenden Arbeit soll der Exciter zur Bearbeitung von Sprachsignalen verwendet werden. Anwendungsbeispiele sind das Freisprech-Telefonieren und Sprechanlagen. Bei der Anwendung wird berücksichtigt, dass Umgebungsgeräusch vorhanden sein kann (z. B. im Fahrzeug). Nach einer Darlegung der Algorithmen soll durch Hörtests untersucht werden, ob die Sprache nach Verarbeitung mit dem Exciter subjektiv bevorzugt wird und ob sich die Sprachverständlichkeit verbessert.

Der Exciter

Der Verzerrer zur Erzeugung der Oberwellen wird in einem Parallelpfad eingesetzt [3, 4]. In dem Hauptpfad zwischen Eingang und Ausgang sorgt eine Verzögerung für den Laufzeitgleich, der durch die Verarbeitung im Parallelpfad entstehen kann (Abb. 1). Würde die Verzögerung nicht ausgeglichen, ergäbe sich durch die Addition im Frequenzgang ein „Kammfilter-Effekt“. Die diskrete Zeit wird mit k bezeichnet. Für die Verzögerung sollen d Abtastwerte angenommen werden. Das verzerr-

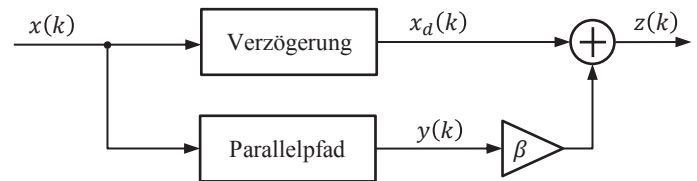


Abbildung 1: Vereinfachtes Blockschaltbild des Exciters

te Signal $y(k)$ wird mit dem Faktor β multipliziert und zu dem verzögerten Originalsignal $x_d(k) = x(k - d)$ addiert. Es gilt damit für den Ausgang $z(k)$ des Exciters

$$z(k) = x_d(k) + \beta \cdot y(k). \quad (1)$$

Der Verzerrer im Parallelpfad ist aus einer Kennlinie K_g

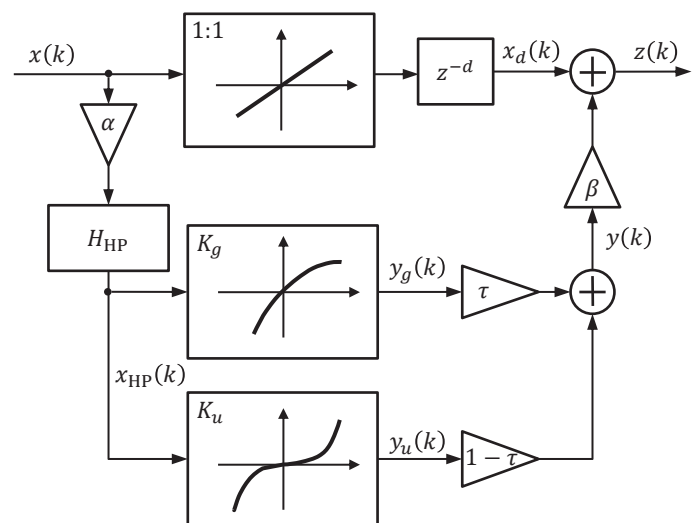


Abbildung 2: Blockschaltbild des Exciters, dargestellt als Addition von Kennlinien

für die geradzahligem Harmonischen mit dem Ausgangssignal $y_g(k)$ und einer weiteren Kennlinie K_u für die ungeradzahligem Harmonischen $y_u(k)$ aufgebaut. Der Faktor τ soll für Timbre (Klangfarbe) stehen. Er bestimmt das Mischungsverhältnis der erzeugten geradzahligem und ungeradzahligem Harmonischen (Abb. 2). Für das Ausgangssignal $z(k)$ gilt damit

$$z(k) = x_d(k) + \beta \cdot (\tau \cdot y_g(k) + (1 - \tau) \cdot y_u(k)). \quad (2)$$

Ebenfalls im Parallelpfad befindet sich ein Hochpassfilter H_{HP} mit -3 dB-Grenzfrequenz f_{HP} sowie ein Verstärkungsfaktor α . Sowohl Filterung als auch

Verstärkung werden vor der Verzerrung durchgeführt. Die Hochpassfilterung sorgt dafür, dass pegelstarke Anteile bei tieferen Frequenzen gedämpft werden und die Harmonischen hauptsächlich aus den höheren Eingangsfrequenzen gebildet werden. Mit α wird der ausgenutzte Bereich der Kennlinie bestimmt und damit die Stärke der Verzerrung. In Abb. 2 ist der direkte Pfad für $x(k)$ ebenfalls als Kennlinie gezeichnet, deren Übertragungsverhältnis zwischen Ein- und Ausgang 1 : 1 ist (also ohne Wirkung). Für die Gesamtwirkung können alle drei Kennlinien addiert werden, wobei die Gewichte β und τ zu beachten sind. Die Kennlinie für die geraden Harmonischen wird nach folgender Gleichung berechnet

$$K_g: y_g(k) = -0.5 \cdot x_{HP}(k)^2 + x_{HP}(k), \quad (3)$$

für die ungeraden Harmonischen gilt

$$K_u: y_u(k) = |x_{HP}(k)| \cdot x_{HP}(k). \quad (4)$$

Als geradzahlige Harmonische entsteht hier wegen des Parabelastes nur die erste Oberwelle, d. h. die zweite Harmonische. Für die ungeraden Harmonischen ergeben sich aufgrund der aus zwei Parabelästen zusammengesetzten Kennlinie eine Vielzahl von Oberwellen.

Überabtastung, Alias und Phasenausgleich

Bei der Erzeugung der Harmonischen wird die ursprüngliche Bandbreite überschritten, so dass vor Verwendung der Kennlinie die Bandbreite durch Überabtastung mit einem Faktor u erhöht wird. Nach der Verarbeitung wird wieder mit Faktor u reduziert. Nachfolgend soll das Über- und Unterabtasten als ideale Verarbeitung angenommen werden. Diese ideale Verarbeitung kann in der offline Simulation durch eine steilflankige, nullphasige Filterung angenähert werden. Aliaskomponenten treten somit nicht auf und es ergibt sich auch keine Laufzeit (Phase). Wird zusätzlich H_{HP} nullphasig ausgeführt, kann die Verzögerung d für den Phasenausgleich zu $d = 0$ gesetzt werden, d. h. im Folgenden gilt $x_d(k) = x(k)$.

Parametrierung des Exciters

Es werden zwei verschiedene Exciter-Varianten getestet, wobei im Folgenden als Bezeichnung für die Varianten A und B verwendet wird (vgl. Tabelle 1). Die beiden Varianten repräsentieren zwei unterschiedlich parametrierte Versionen des vorgestellten Algorithmus. Bei Variante B wird durch $\alpha = 4$ die Verzerrung deutlich gesteigert.

Tabelle 1: Getestete Varianten

Var	Parameter					
	α	β	f_{HP}	τ	u	d
A	1.8	1.0	2 kHz	0.8	4	0
B	4.0	1.0	2 kHz	0.8	4	0

Abb. 3 zeigt den gemessenen Frequenzgang (Kurve für Signal $z(k)$) der Variante A. Um diesen messen zu können,

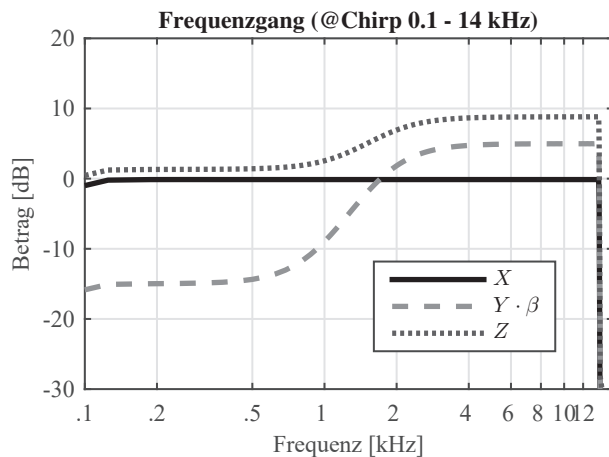


Abbildung 3: Frequenzgang der Variante A gemessen mit einem Chirp-Signal von 100 Hz bis 14 kHz

wird an Stelle der oben gezeigten Kennlinien jeweils die 1 : 1 Kennlinie eingesetzt. Als Hochpass wurde ein Base-Shelf Filter mit $f_{HP} = 2$ kHz und -20 dB Bass-Dämpfung verwendet. Die Verstärkung in den hohen Frequenzen (ca. +9 dB) ergibt sich durch die Multiplikation mit $\alpha=1.8$ und die Addition von $x(k)$ und $y(k)$.

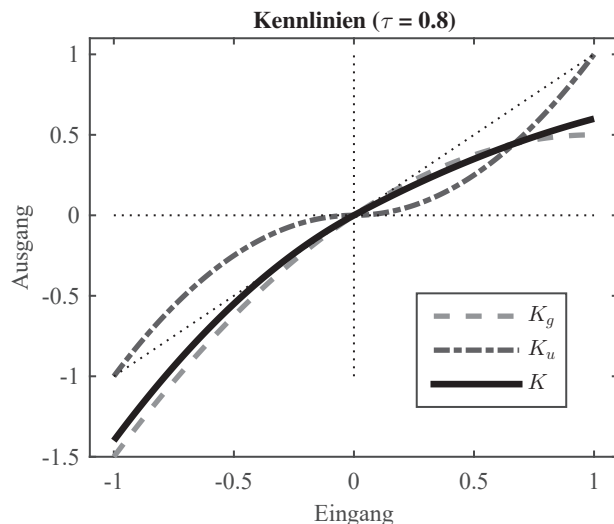


Abbildung 4: Kennlinie K als Addition der geraden und ungeraden Anteile, K_g und K_u , mit $\tau = 0.8$

Abb. 4 zeigt die Kennlinie $K = K_g + K_u$ für das Signal $y(k)$. Dabei wurde $H_{HP} = 1$ gesetzt und ebenso $\alpha = 1$ gewählt. Für das Signal $z(k)$ ergibt sich eine Gesamtkennlinie die infolge der Addition von $y(k) \cdot \beta$ zu $x(k)$ noch stärker an die 1 : 1 Diagonale angenähert ist.

Verzerrungen werden üblicherweise durch einen Klirrfaktor erfasst, der sich bei einem sinusförmigen Eingangssignal ergibt. Hier wurde nachfolgende Definition verwendet (Harmonic Distortion, HD_n)

$$HD_n = \frac{\sqrt{P_n}}{\sqrt{P_0}}. \quad (5)$$

P_n ist die Leistung der n -ten Harmonischen, P_0 die Leistung des Grundtons. Abb. 5 zeigt das Spektrum und die

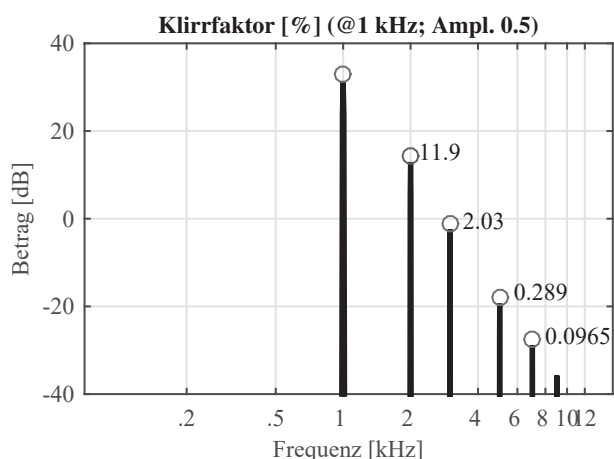


Abbildung 5: Klirrfaktor, gemessen mit Sinus 1 kHz und Amplitude 0.5

Klirrfaktoren für das Signal $z(k)$ wobei $H_{HP} = 1$ verwendet wird ($\alpha = 1.8$; $\beta = 1$; $\tau = 0.8$). Als Eingangssignal dient ein Sinus mit Frequenz 1 kHz und Amplitude 0.5. Der Klirrfaktor beträgt für die 2. Harmonische beispielsweise 11.9%. Halbiert man die Amplitude des Eingangssignals $x(k)$, werden sich auch die Klirrfaktoren bei sonst gleichen Parametern etwa halbieren.

Evaluation

Zur Evaluation der Exciter-Varianten werden zwei Hörtests durchgeführt. Bei beiden Tests kommt ein CMOS Test (*Comparison Mean Opinion Score*) zum Einsatz [5]. Bei diesem wird jeweils das unverarbeitete Signal dem mit dem Exciter verarbeiteten Signal gegenübergestellt. Den Probanden ist nicht bekannt, welches das verarbeitete Signal ist. Die Reihenfolge, in der die Signale präsentiert werden, ist zufällig. Im ersten Test wird die Sprachqualität der verarbeiteten Signale im Vergleich zu den unverarbeiteten Signalen beurteilt. Hierfür wird der Parametersatz A verwendet. Die Fragestellung lautet „Bewerten Sie, welches der beiden Sprachsignale besser klingt“. Im zweiten Test wird mit dem Parametersatz B überprüft, ob sich die Sprachverständlichkeit nach Verarbeitung mit dem Exciter verbessert. Die Fragestellung ist dementsprechend „Bewerten Sie, bei welchem der beiden Sprachsignale die Sprachverständlichkeit besser ist“.

Die beiden CMOS Tests werden dreistufig durchgeführt, die entsprechenden Antwortmöglichkeiten lauten:

- Signal A wird bevorzugt
- Kein Unterschied hörbar
- Signal B wird bevorzugt

An dem Test nahmen 18 normalhörende Probanden (4 weibliche, 14 männliche) in einem Alter von 22 bis 44 Jahren teil. Die Darbietung der einzelnen Hörproben wurde mit Hilfe von geschlossenen Kopfhörern (AKG K271) durchgeführt.

Szenarien

Die Sprachqualität wird in drei Szenarien evaluiert. In allen drei Szenarien werden die Signale mit dem Parametersatz A verarbeitet:

- S1: Klare Sprache (kein Hintergrundgeräusch)
- S2: Sprache mit Geräusch nach einer Geräuschreduktion
- S3: Klare Sprache, Wiedergabe in Geräuschumgebung

Die Sprachverständlichkeit wird in einem vierten Szenario S4 evaluiert. Hier werden die Signale mit dem Parametersatz B (gesteigerte Verzerrung) verarbeitet:

- S4: Klare Sprache, Wiedergabe in sehr lauter Geräuschumgebung

Alle Szenarien verwenden die gleichen Sprachdaten, welche in ca. 50 cm Abstand aufgezeichnet wurden (Mikrofon mit linearem Frequenzgang). Die Sprecherdauer je Sprecher beträgt zw. 5 bis 10 s. Es gibt fünf Sprecher (zwei weibliche und drei männliche).

Das Szenario S2 verwendet Fahrgeräusch bei 140 km/h, welches nach Gewichtung mit dem Faktor 0.2 (-14 dB) zur Sprache addiert wird. Damit wird eine ideale Geräuschreduktion simuliert. Die Summe aus Sprache und Geräusch wird dann mit dem Exciter verarbeitet. Bei den Szenarien S3 und S4 wird das Geräusch erst nach der Verarbeitung mit dem Exciter addiert. Szenario S3 verwendet ebenfalls Fahrgeräusch bei 140 km/h, jedoch ohne Gewichtung. Für Szenario S4 wird Fahrgeräusch bei 190 km/h verwendet und die Sprachdaten vor der Geräusch-Addition im Pegel soweit abgesenkt, dass die Sprachverständlichkeit etwa im Bereich 70 % bis 90 % liegt. Bei den Szenarien S1 - S3 werden Sprache und Geräusch mit einem Hochpass mit 100 Hz Grenzfrequenz gefiltert. Für das Szenario S4 wird für das Geräusch eine Grenzfrequenz von 50 Hz verwendet um ein realistisches Umgebungsgeräusch zu erhalten.

Es wurde auf eine gleichmäßige Aussteuerung der Signale geachtet. Ein evtl. höherer dB(A)-Pegel des Ausgangssignals, der sich durch die Verarbeitung ergibt, wird vor dem Abhören wieder ausgeregelt.

Auswertung

Die Auswertung des oben beschriebenen Hörtests ist in den Abbildungen 6 bis 9 gezeigt. Die vier Szenarien wurden einzeln ausgewertet. Da der Exciter, wie beschrieben, vor allem Auswirkung auf die hohen Frequenzen hat, wurde zudem zwischen männlichen und weiblichen Sprechern unterschieden.

Bei den Qualitätstests (S1 - S3) zeichnet sich ab, dass ohne bzw. mit wenig Hintergrundgeräusch (S1 und S2) das unverarbeitete Signal bevorzugt wird. Vor allem bei den weiblichen Sprechern schneidet der Exciter im Vergleich zum Originalsignal in den Szenarien S1 und S2 schlecht ab (ca. 36 % Differenz bei S1 und ca. 17 % Differenz bei S2, vgl. Abb. 6 und 7). In Szenario S3 (verarbeitetes Signal wird in Geräuschumgebung wiedergegeben) ist das Verhältnis ausgewogener. Bei den weib-

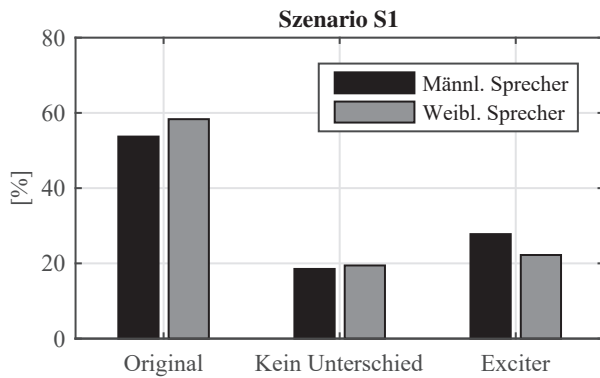


Abbildung 6: Auswertung Szenario S1 (Qualität), Parametersatz A

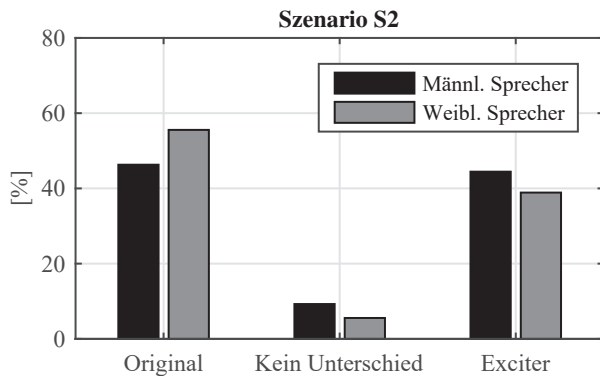


Abbildung 7: Auswertung Szenario S2 (Qualität), Parametersatz A

lichen Sprechern sind die Prozentzahlen identisch, bei den männlichen Sprechern beträgt die Differenz zugunsten des unverarbeiteten Signals ca. 6% (Abb. 8). Auf die Sprachqualität hat der vorgestellte Algorithmus also gemäß des durchgeführten Tests keine positive Auswirkung.

In Szenario S4 zeigt sich jedoch, dass die Sprachverständlichkeit in lauter Geräuschumgebung mit dem Exciter deutlich verbessert werden kann. 88,9% der Probanden gaben an, die weiblichen Sprecher mit Exciter besser zu verstehen, bei den männlichen Sprechern waren es ca. 57% (Abb. 9).

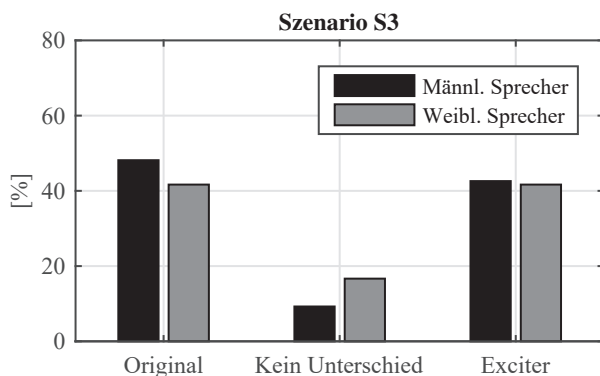


Abbildung 8: Auswertung Szenario S3 (Qualität), Parametersatz A

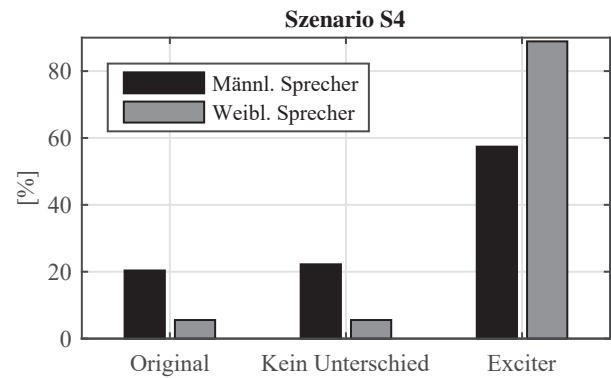


Abbildung 9: Auswertung Szenario S4 (Verständlichkeit), Parametersatz B

Zusammenfassung

Es wurde ein Algorithmus zur Nachbildung eines Exciters vorgestellt. Basierend auf zwei unterschiedlichen Kennlinien können mithilfe des Algorithmus geradzahlige und ungeradzahlige Harmonische zu einem Signal hinzugefügt werden. Ziel war es, sowohl Sprachqualität als auch Sprachverständlichkeit bei Anwendungen in Geräuschumgebung zu verbessern.

Insgesamt kann festgestellt werden, dass mit dem Exciter für Sprache interessante Effekte darstellbar sind. Im vorliegenden CMOS Test für die Sprachqualität wurde im direkten Vergleich nach einem pauschalen Urteil („*Welches Signal klingt besser?*“) gefragt. Damit sind einzelne Attribute wie z. B. „*heller*“ oder „*frischer*“ nur indirekt erfassbar, bzw. wurden nicht abgefragt. Hinzu kommt die bereits sehr gute Qualität der unverarbeiteten Sprachsignale und dass drei der fünf Sprecher eine geschulte Stimme haben. Unter diesen Voraussetzungen ist es schwierig eine subjektive Verbesserung der Qualität zu erreichen. Sehr deutlich und zugunsten des Exciters fiel hingegen die Befragung nach der Verbesserung der Sprachverständlichkeit bei sehr lautem Umgebungsgeräusch aus. Neben dem nichtlinearen Anteil bei höheren Frequenzen hat hier auch die lineare Anhebung der Höhen einen positiven Beitrag geleistet.

Literatur

- [1] J. Chalupper: Aural Exciter and Loudness Maximizer: What's Psychoacoustic About "Psychoacoustic Processors"? 109th AES Convention, Los Angeles, 2000.
- [2] P. Shekar, J. Smith: Modeling the Harmonic Exciter. 135th International AES Convention, New York, 2013.
- [3] Apex Systems Ltd.: Aural Exciter and Optical Big Bottom - Instruction Manual. Sun Valley, 2001.
- [4] C. Knoppel: Signal Distortion Circuit and Method of Use. Patent No. US4150253 A, 1979.
- [5] ITU-T-Empfehlung P.800 (08/96), URL: <http://www.itu.int/rec/T-REC-P.800-199608-I/en>