Ein neues psychoakustisches Meßverfahren zur Bestimmung der Wahrnehmbarkeit von Codierfehlern

Thilo Thiede, Ernst Kabot

(Institut für Nachrichtentechnik und Theoretische Elektrotechnik, Technische Universität Berlin)

1. Zusammenfassung

Zur Einschätzung der Qualität gehöranpaßter Audiocodierer sind neue psychoakustische Meßverfahren erforderlich, die durch einen Vergleich mit dem Originalsignal wahrnehmbare Störungen erkennen und bewerten. Mit Hilfe eines Experimentiermodells, das eine gehörgerechte Filterung mit weitgehend beliebiger zeitlicher und spektraler Auflösung erlaubt, wird die Eignung verschiedener Parameter zur Abschätzung des subjektiven Höreindrucks überprüft. Weiterhin wird der Einfluß der zeitlichen Auflösung und der Anzahl der Filterkanäle auf die Leistungsfähigkeit des Modells untersucht. Zur Bewertung der Ergebnisse und zur Optimierung der Parameter des Modells werden die von der ITU und der MPEG zur Qualitätsbewertung verschiedener Audiocodierverfahren durchgeführten Hörtests herangezogen. Die hier vorgestellten Arbeiten wurden im Auftrag der Deutschen Telekom AG durchgeführt.

2. Funktionsweise von gehörrichtigen Meßverfahren

Hauptaufgabe eines gehörrichtigen Meßverfahrens ist es, wahrnehmbare Störungen eines Audiosignals von nicht wahrnehmbaren (maskierten) Störanteilen zu unterscheiden und deren Einfluß auf den subjektiv empfundenen Qualitätsverlust abzuschätzen. Dies geschieht durch einen Vergleich des zu bewertenden Signals mit dem ungestörten Referenzsignal. Hierzu werden die Eingangssignale zunächst in Kurzzeitspektren zerlegt. Anschließend wird nach einer Gewichtung der Spektrallinien mit Übertragungsfunktionen des Ohrs eine Skalentransformation von der linearen Frequenzskala auf eine gehörrichtige Tonhohenskala vorgenommen. Dies geschieht üblicherweise durch eine Gruppierung benachbarter Spektrallinien zu Frequenzbändern, deren Breite jeweils einem festgelegtem Bruchteil einer Frequenzgruppe entspricht. Durch eine Faltungsoperation mit einem idealisiertem Mithörschwellenverlauf wird eine Maskierung zwischen benachbarten Frequenzbändern berücksichtigt. Eine zeitliche Verschmierung der Signale zur Berücksichtigung der Nachverdeckung wird üblicherweise durch einen Tiefpaß erster Ordnung vorgenommen, der zu einem exponentiellen Abklingen der Maskierungswirkung nach Ende des Maskierers führt. Zur Berücksichtigung der Vorverdeckung ist meist keine weitere zeitliche Verschmierung erforderlich, da bereits durch die Fensterlänge der FFT eine ausreichende Begrenzung der zeitlichen Auflösung gegeben ist.

Je nach Meßverfahren wird als Ausgangsgröße der Pegelabstand zwischen Störung und Verdeckungsschwelle, die Häufigkeit auftretender Störungen [1], ein Schätzwert für die Lautheit bzw. Lästigkeit der Störung [2], oder die Wahrscheinlichkeit für die Hörbarkeit der Störung [3][4] angegeben.

3. Beschreibung des neuen Meßverfahrens

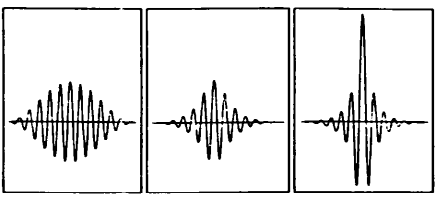
Wegen der nichtlinearen Abbildung zwischen Frequenzskala und Tonhöhenskala ist bei Verwendung einer FFT als Zeit-/Frequenztransformation die Berechnung einer vergleichsweise großen Anzahl von Spektrallinien notwendig, um eine ausreichend genaue Anpassung der Frequenzbänder an die Frequenzgruppenskala zu ermöglichen. Durch die hierdurch bedingten großen Fensterlängen wird jedoch die Genauigkeit, mit der zeitliche Verdeckungseffekte (insbesondere Vorverdeckung) nachgebildet werden können, eingeschränkt. Da durch ein langes Zeitfenster auch ein Teil der in den Hüllkurven der Ausgangssignale in den einzelnen Frequenzbändern enthaltenen Information verloren geht, können einige hiermit zusammenhängende Effekte nur schwer modelliert werden. Insbesondere bei der Modellierung von Effekten, die mit dem Richtungshören zusammenhängen, ist die Auswertung von Hüllkurveninformation erforderlich.

Das hier vorgestellte Modell wurde u.a. mit dem Ziel entwickelt, die Auswirkung verschiedener zeitlicher und spektraler Auflösungen unabhängig voneinander zu untersuchen. Weiterhin wurde ein neuer Ansatz zur Bestimmung der Drosselung der Störung durch das Nutzsignal unter Ausnutzung von Hüllkurveneigenschaften verwendet. An der Modellierung von Gehöreigenschaften, die mit dem beidohrigen Hören zusammenhängen, wird zur Zeit gearbeitet. Erste Simulationen scheinen aber darauf hinzuweisen, daß solche Effekte nur einen geringen Einfluß auf die Ergebnisse der zur Überprüfung des Modells verwendeten Hörtests hatten.

3.1. Beschreibung der gehörangepaßten Filterbank

Ebenso wie in anderen gehörrichtigen Meßverfahren, wird in dem vorliegenden Verfahren durch einen Vergleich des zu beurteilenden Testsignals mit einem fehlerfreien Referenzsignal eine Abschätzung der hörbaren Störungen im Testsignal vorgenommen. Hierzu werden sowohl Test- als auch Referenzsignal nach einer Filterung mit der Übertragungsfunktion zwischen Außen- und Innenohr durch eine gehörangepaßte Filterbank in den Frequenzbereich zerlegt. Die Dämpfungsverläufe an den Filterflanken werden dabei den aus der psychoakustischen Literatur bekannten Mithörschwellenverläufen angepaßt.

Die Filterbank besteht aus einer frei wählbaren Anzahl von Filterpaaren, die linear über einer Tonhöhenskala verteilt sind. Eine nachträgliche Transformation von einer Frequenzskala auf eine Tonhöhenskala, wie in den meisten übrigen Verfahren, ist daher nicht mehr nötig. Die Impulsantworten der Filter weisen zunächst eine cosinusquadratförmige Hüllkurve auf.



a) b) c)

Abb. 1: Impulsantworten eines Filters vor (a) und nach der komplexen Faltung für zwei Impulse verschiedener Intensität (b: geringe Intensität, c: hohe Intensität). Die Form der Hüllkurve der Impulsantwort nach der Faltung entspricht in guter Näherung dem bei Zwicker [7] dargestellten Nachverdeckungsverlauf.

Im Gegensatz zu anderen Meßverfahren wird die Faltung mit den Mithörschwellenverläufen bereits vor der Bestimmung der Hüllkurven vorgenommen. Hierzu wird die Faltungsoperation sowohl auf die Filterausgänge, die den Realteil des Signals repräsentieren, als auch auf die Filterausgänge, die den Imaginärteil des Signals repräsentieren, angewendet. Dies hat zur Folge, daß sich die Impulsantworten der Filter an die nach der Faltung vorliegenden Amplitudenfrequenzgänge anpassen. Daher kann man diese Faltung als Teil der Filterbank auffassen. Für die Steilheit der oberen Flanke der Mithörschwellen kann optional entweder ein pegelunabhängiger Verlauf oder ein pegelabhängiger Verlauf nach Terhardt [5] verwendet werden. Durch eine Wiederholung der Faltung lassen sich auch abgerundete Mithörschwellen („rounded exponentials" nach Patterson [6]) erzeugen, dies führte jedoch zu keiner nachweisbaren Verbesserung des Modells. Bei Verwendung der pegelabhängigen Modellierung ergibt sich auch eine Pegelabhängigkeit

der Impulsantworten (vgl. Abb. 1): die Impulsantwort für einen sehr lauten Impuls fällt deutlich steiler ab als die für einen vergleichsweise schwaches Signal. Dies korrespondiert mit der qualitativ ähnlich verlaufenden Pegelabhängigkeit der Nachverdeckung. Um diesen Effekt zur Modellierung der Pegelabhängigkeit der Nachverdeckung zu nutzen, müßte jedoch mit einer sehr großen Anzahl von entsprechend schmalbandigen Filtern gearbeitet werden.

Um zu verhindern, daß Pegelunterschiede und Unterschiede im Frequenzgang zwischen Test- und Referenzsignal als additive Störungen behandelt und dadurch überbewertet werden, wird nach der Betragsbildung ein Angleich des Pegels und des Frequenzganges von Test- und Referenzsignal vorgenommen. Anschließend wird zur Berücksichtigung der an den Hörzellen auch ohne Schalleinwirkung vorhandenen Grunderregung ein frequenzabhängiger Offset zu den Ausgangswerten in den einzelnen Filterkanälen addiert.

Die zeitliche Verschmierung der Signale wird in zwei Schritten ausgeführt. Zunächst werden die Signale mit einem cosinusquadratförmigen Zeitfenster gefaltet, das in erster Linie zur Berücksichtigung der Vorverdeckung dient. Für die Länge dieses Zeitfensters hat sich ein Wert von ca. 8 ms als günstig erwiesen. Nachverdeckung wird durch einen Tiefpaß erster Ordnung modelliert, wobei die Zeitkonstanten je nach Frequenzband zwischen 8 ms und ca. 100 ms liegen.

3.2. Berechnete Ausgangsparameter

Aus den gehörangepaßten Darstellungen von Test- und Referenzsignal wird eine Reihe von objektiven Parametern zur Bewertung der auftretenden Störungen berechnet. Die wichtigsten Ausgangsparameter sind

- die Kreuzkorrelation zwischen den spezifischen Lautheitsmustern nach Zwicker [7] von Test- und Referenzsignal
- die mittlere Kreuzkorrelation zwischen den zeitlichen Hüllkurven von Test- und Referenzsignal in den einzelnen Filterkanälen
- ein Maß für die durch das Referenzsignal gedrosselte Lautheit der Störung (siehe unten)

sowie Meßgrößen für die durch den fortlaufenden Pegel- und Frequenzgangangleich kompensierten langsamen Verstärkungsschwankungen und linearen Verzerrungen.

Zur Bestimmung der gedrosselten Störlautheit wurden einige Abwandungen der von Zwicker [7] angegebenen Lautheitsgleichung getestet. Diese hatten alle die Eigenschaft , daß sich der Ausgangswert bei sehr schwachem Maskierer ($E_{Ref} = 0$) der spezifischen Lautheit der Störung nähert, während bei schwachen Störungen ($E_{Test} \approx E_{Ref}$) eine Drosselung proportional zur Stärke des Maskierers erfolgt. Die besten Ergebnisse wurden dabei mit einem Ausdruck der Form

$$NL(f_i,t) = \left(\frac{1}{s_{test}} \cdot \frac{E_{HS}}{E_0}\right)^{0.23} \cdot \left[\left(1 + \frac{\max\left(s_{test} \cdot E_{test} - s_{ref} \cdot E_{ref}, 0\right)}{E_{HS} + \beta \cdot s_{ref} \cdot E_{ref}}\right)^{0.23} - 1\right]$$

erzielt. Dabei steht E_{HS} für die Grunderregung und β bestimmt die Drosselung durch das Referenzsignal. Die Größen s_{test} und s_{ref} ersetzen den bei Zwicker [7] auftretenden Schwellenfaktor und werden aus der Hüllkurvenmodulation in dem jeweiligen Filterkanal bestimmt.

$$\beta = \exp\left(-\frac{E_{test} - E_{ref}}{E_{ref}}\right)$$

4. Überprüfung des Modells

Das beschriebene Meßverfahren wurde anhand einer Reihe von Hörtests überprüft, die in den vergangenen Jahren von der MPEG und der ITU zum Test gehörangepaßter Audiocodierverfahren durchgeführt worden sind. Die von dem Modell gelieferten Ausgangsparameter wurden mit den Ergebnissen der Hörtests verglichen. Hierzu wurden Abbildungsfunktionen bestimmt, die den Wertebereich der Modellparameter auf die bei diesen Hörtests verwendete „5 Grade Impairment Scale" umrechnen. Anschließend wurden der Standardfehler und der Kreuzkorrelationskoeffizient zwischen den Modellparametern und den Ergebnissen der Hörtests ermittelt. Die höchsten Korrelationen mit den Hörtestergebnissen ergaben sich dabei für die mittlere gedrosselte Störlautheit.

In Abb. 2 werden Ergebnisse für den MPEG-Hörtest von 1990 sowie für den ITU-Hörtest von 1993 gezeigt. Die zur Abbildung der gedrosselten Störlautheit auf die Hörtestskala verwendete Sigmoid-Funktion ist im Diagramm eingezeichnet. Es werden für fast alle zur Verfügung stehenden Hörtestdaten Korrelationen erreicht, die mit den besten Ergebnissen anderer objektiver Meßverfahren vergleichbar sind.

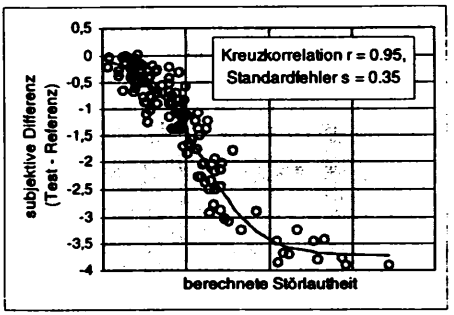


Abb. 2: .Ergebnisse für drei Hörtests
(MPEG 1990: Musicam + SB-ADPCM, Kopfhörerwiedergabe, ITU 1993: kaskadierte Codecs, ITU 1993: „commentary stereo")

Der Einfluß der spektralen und zeitlichen Auflösung auf die Vorhersagegenauigkeit der Hörtestergebnisse war für die meisten Teststücke unerwartet gering. Bei einer Erhöhung der Filteranzahl von 80 auf über 120 Teilbänder ergab sich für keinen der zur Überprüfung zur Verfügung stehenden Hörtestdatensätze eine merkliche Verbesserung. Eine Verringerung der Filteranzahl auf weniger als 50 Teilbänder hatte nur bei wenigen Datensätzen eine merkliche Verschlechterung der Korrelationen zur Folge. Eine Verkürzung des zur Modellierung der Vorverdeckung verwendeten Zeitfensters auf unter 8 ms verringerte die Korrelationen mit den Hörtestergebnissen deutlich, während die Verwendung eines längeren Zeitfensters nur eine geringfügige Verschlechterung zur Folge hatte.

5. Ausblick

Eine Erweiterung des Modells um Effekte des beidohrigen Hörens (BMLDs. „stereo unmasking") ist vorgesehen. Bei der Berechnung der gedrosselten Störlautheit und der hier auftretenden Schwellenfaktoren sind noch Verbesserungen möglich. Auch die bei der Bestimmung des Pegel- und Frequenzgangangleichs vorgenommen Unterscheidung zwischen Frequenzgangänderungen und additiven Störungen ist noch nicht optimal.

6. Literatur

[1] Brandenburg, K. H.; Sporer, Th.: NMR and Masking Flag: Evaluation of Quality Using Perceptual Criteria. Proceedings of the AES 11th International Conference, Portland, Oregon, USA, 1992, S. 169-179.

[2] Beerends, J. G.; Stemerdink, J. A.: A Perceptual Audio Quality Measure Based on a Psychoacoustic Sound Representation. J. AES, Vol. 40(1992), No. 12, December, S. 963-978.

[3] Colomes, C.; Lever, M.; Dehery, Y. F.: A Perceptual Objective Measurement System (POM) for the Quality Assessment of Perceptual Codecs. Beitrag zur 96th AES Convention, Amsterdam, February 1994, Preprint 3801.

[4] Paillard, B.; Mabilleau, P.; Morissette, S.; Soumagne, J.: PERCEVAL: Perceptual Evaluation of the Quality of Audio Signals. J. AES., Vol. 40(1992), No. 1/2, January/February , S. 21-31.

[5] Terhardt, E.: Calculating Virtual Pitch. Hearing Research, Vol. 1(1979), S. 155-182.

[6] Patterson, R. D.: Auditory Filter Shapes Derived with Noise Stimuli. J. Acoust. Soc. Am., Vol. 59, No. 3, March 1976, pp. 640-654.

[7] Zwicker, E.; Feldtkeller, R.: Das Ohr als Nachrichtenempfänger. Stuttgart: Hirzel Verlag, 1967.