

Vorhersage und Messung von Detektionsschwellen für Audio-Codec-Verzerrungen

Rainer Huber

AG Medizinische Physik, Carl-von-Ossietzky-Universität Oldenburg, D-26111 Oldenburg
hub@medi.physik.uni-oldenburg.de

Einleitung

Verlustbehaftete, gehörorientierte Kodierungsverfahren zur Reduktion digitaler Audiodaten werden heutzutage vermehrt eingesetzt, wenn für die Übertragung oder Speicherung von Audiodaten nur begrenzte Kapazitäten zur Verfügung stehen. Um die Übertragungsqualität solcher Verfahren beurteilen zu können, sind konventionelle technische Maße, wie z.B. das Signal-to-Noise Ratio (SNR) ungeeignet, da sie die Wahrnehmungseigenschaften des Gehörs nicht berücksichtigen. Zur Qualitätsbeurteilung mussten daher bislang aufwendige Hörtests durchgeführt werden. Es besteht also ein Bedarf an instrumentellen, objektiven Verfahren zur Qualitätsabschätzung von Audioübertragungssystemen.

Das von der ITU (*International Telecommunication Union*) standardisierte Verfahren PEAQ (*Perceptual Evaluation of Audio Quality* [1]) sowie das von Huber und Kollmeier im letzten Jahr an dieser Stelle vorgestellte Verfahren [2] versuchen, die subjektiv empfundene Qualitätsverschlechterung eines Audiosignals gegenüber einer Referenz auf einer kontinuierlichen Skala vorherzusagen. Mit zunehmender Verbesserung der Audio-Codex und wachsenden Ansprüchen an deren Übertragungsqualität dürfte künftig jedoch weniger der *Grad* der wahrnehmbaren Qualitätsverschlechterung von Interesse sein als vielmehr nur noch die Frage, ob *überhaupt* wahrnehmbare Signaländerungen durch einen Codec bewirkt werden. D.h. die Frage nach der *Transparenz* eines Codex wird entscheidend sein. Die oben genannten instrumentellen Verfahren enthalten die Kategorie der Transparenz lediglich als Endpunkt ihrer kontinuierlichen Skalen.

Vor diesem Hintergrund wurde in der vorliegenden Arbeit untersucht, ob es mit Hilfe eines auditorischen Modells möglich ist, Detektionsschwellen von komplexen Audiosignalverzerrungen vorherzusagen. Die Detektionsschwellen könnten als „Abstand zur Transparenz“ als Maß für die Beurteilung der Übertragungsqualität von Audio-Codex herangezogen werden.

Methode

Psychoakustische Ermittlung von Detektionsschwellen

Es wurde eine psychoakustische Messung durchgeführt, an der 10 normalhörende Versuchspersonen teilgenommen haben. Die Messung bestand aus zwei Teilen. Im ersten Teil wurden die folgenden vier Audiosignale verwendet:

- „Carmen“; Dauer: 3,8 s; Pegel: 62 (76) dB SPL
- „Glockenspiel“; Dauer: 2,5 s; Pegel: 55 (66) dB SPL
- „Kastagnetten“; Dauer: 3 s; Pegel: 54 (78) dB SPL
- Sprache (männl., deutsch); Dauer: 4,1 s; Pegel: 60 (74) dB SPL

(Die in Klammern angegebenen Pegelwerte sind Spitzenpegel.)

Die Signale wurden durch drei verschiedene Signalverarbeitungen verändert:

- MPEG-1 Layer 3 („mp3“) bei 48 kbit/s/Kanal
(→ Kompression $\approx 1:14$)
- Windows Media Audio („wma“) bei 48 kbit/s/Kanal
- Rauschmodulation („mnru“): $y(t) = (1+m \cdot \text{noise}(t)) \cdot x(t)$
(„mnru“: *modulated noise reference unit*)

Die Darbietung der Signale erfolgte diotisch über Kopfhörer.

Zur Bestimmung der Detektionsschwellen wurden drei 3-AFC-Tests (AFC: *alternate forced choice*) mit adaptiver Pegelsteuerung durchgeführt. In jedem Test wurden die Audiosignale durch einen der drei Codex verändert und *interleaved* („verschachtelt“) dargeboten. Zielgröße war die Mithörschwelle der durch die Codex verursachten Verzerrungen (das erhöhte Quantisierungsrauschen). Als Maskierungsexperiment betrachtet stellt das unverzerrte Referenzsignal den Maskierer dar und die Verzerrungen das zu detektierende Signal.

Im zweiten Teil des Experiments wurden zur Erweiterung der Messdatenbasis zwei weitere Audiosignale (eine elektrische Bass-Gitarre und eine weibliche Sprecherin) verwendet. Daneben wurden zwei der Signale aus dem ersten Hörtest („Carmen“ und „Kastagnetten“) zu Vergleichszwecken erneut gemessen. Die Codex blieben gleich. Aufgrund der Erfahrungen aus dem ersten Teil des Experiments wurde die adaptive Messprozedur zudem um zwei zusätzliche Umkehrpunkte verlängert.

Generierung der Testsignale

Die Größe, die im Experiment eingeregelt wurde, war der relative Pegel des Quantisierungsrauschens bzw. das Signal-zu-Rausch-Verhältnis. Um dies skalieren zu können, wurde das Differenzsignal aus Referenzsignal und verarbeitetem Signal gebildet, skaliert und wieder zum Referenzsignal hinzu addiert:

$$\text{Testsignal} = \text{Referenzsignal} + X \cdot \text{Differenzsignal}$$

$$\text{mit: Differenzsignal} = \text{Codecsignal} - \text{Referenzsignal}$$

Bei einem Skalierungsfaktor von $X = 1$ ergibt sich als Testsignal somit gerade wieder das Codec-Ausgangssignal, für andere Werte ändert sich das Signal-Rausch-Verhältnis entsprechend: $\Delta \text{SNR} = -20 \cdot \log X$ dB

Simulation

Das Experiment wurde unter Verwendung des Perzeptionsmodells von Dau et al. [3] simuliert. Das Modell ist in Abb. 1 dargestellt. Die Vorverarbeitungseinheit stellt ein quantitatives Modell der „effektiven“ Signalverarbeitung im auditorischen System dar. Die anschließende Entscheidungseinheit des Modells simuliert einen menschlichen Beobachter in einem AFC-Experiment. (Für eine ausführliche Beschreibung des Modells siehe [3].)

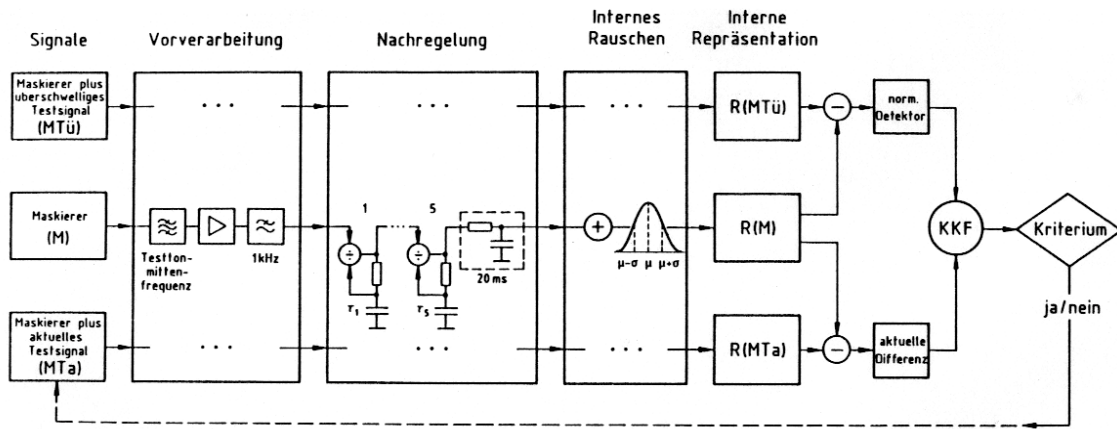


Abb. 1: Schematische Darstellung des psychoakustischen Modells mit optimalem Detektor als Entscheidungsstufe. Die Signale werden vorverarbeitet (Basilmembranfilterung, Halbwellengleichrichtung), nachgeregelt (Pegeladaptation) und liegen schließlich als interne Repräsentationen vor. Die Differenz der Repräsentationen von Maskierer plus aktuellem Signal ($R(MTa)$) und Maskierer allein ($R(M)$) wird mit der normierten überschwelligen Detektorrepräsentation korreliert und der berechnete Korrelationskoeffizient mit einem Schwellenkriterium verglichen. (Aus [4])

Ergebnisse

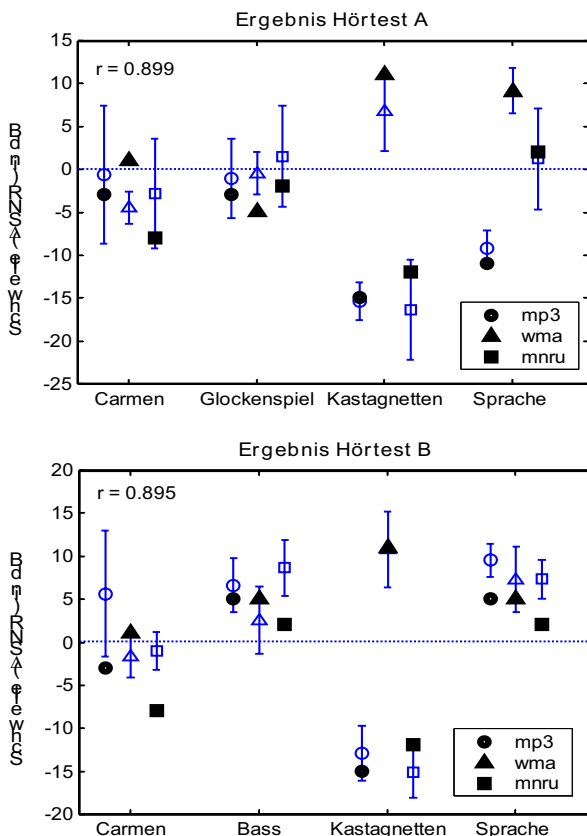


Abb. 2: Gemessene (offene Symbole) und simulierte (gefüllte Symbole) Mithörschwellen von Codec-bedingten Audiosignalverzerrungen

Die Ergebnisse der Messungen sowie der Simulationen sind in Abb. 2 dargestellt. Ein Δ SNR-Wert von 0 dB bedeutet, dass das Testsignal gerade dem Ausgangssignal des Codecs entspricht. (Im Fall der Rauschmodulation bedeutet es ein SNR von 35 dB.) Bei negativen Schwellenwerten mussten die

Verzerrungen verstärkt werden, um wahrnehmbar zu werden; der entsprechende Codec war demnach transparent. Allgemein gilt: Je höher die Werte, desto empfindlicher das Gehör.

Wie zu erkennen ist, variieren die Schwellen mit der Art des Signals und der Verarbeitung z.T. erheblich (v.a. beim Kastagnettensignal). Zudem lässt sich bei den Signalen, die in beiden Hörtesten gemessen wurden, eine Erhöhung der Schwellen im Hörtest B feststellen, was auf einen Trainingseffekt hindeutet, evtl. auch auf die erhöhte Anzahl von Umkehrpunkten in der adaptiven Messung im Hörtest B zurückzuführen ist.

Die Übereinstimmung zwischen gemessenen und simulierten Werten ist insgesamt gut: Die simulierten Werte liegen zumeist innerhalb der interindividuellen Standardabweichungen. Auch in Relation zum gesamten Wertebereich der Schwellen erscheint die Übereinstimmung gut, was sich durch einen hohen linearen Korrelationskoeffizienten ("r" in Abb. 2) von 0,9 zeigt.

Schlussfolgerung

Das Perzeptionsmodell erscheint geeignet, Detektionsschwellen von Codec-bedingten Audiosignalverzerrungen vorherzusagen und damit Aussagen über die zu erwartende Transparenz/Nicht-Transparenz eines Codecs sowie darüber hinaus über den "Abstand zur Transparenzschwelle" machen zu können.

Der Autor dankt Torsten Dau für seine Beratung.

Literatur

- [1] ITU-R Rec. BS-1387, „Method for Objective Measurement of Perceived Audio Quality“, ITU, Genf, Schweiz (1999)
- [2] R. Huber, B. Kollmeier: „Vorhersage von Audioqualität mit einem psychoakustischen Modell“, Fortschritte der Akustik DAGA 2001
- [3] T. Dau, D. Püschel, A. Kohlrausch: „A quantitative model of the 'effective' signal processing in the auditory system. I. Model structure“, JASA 99(6), 3615-3622 (1996)
- [4] T. Dau: „Der Optimale Detektor in einem Computermodell zur Simulation von psychoakustischen Experimenten“, Diplomarbeit, Universität Göttingen (1992)