Vorhersage von Audioqualität mit einem psychoakustischen Modell

Rainer Huber und Birger Kollmeier

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 daher ein Bedarf an instrumentellen, objektiven Verfahren zur Qualitätsabschätzung von Audioübertragungssystemen. In dem vorliegenden Beitrag soll ein solches Verfahren vorgestellt werden. Es stellt eine Weiterentwicklung des Ansatzes von Hansen und Kollmeier dar, die mit Hilfe eines Modells der auditorischen Signalverarbeitung [1] bereits die Qualität von Sprachübertragungssystemen erfolgreich vorhersagen konnten [2]. Um auch sehr geringe Qualitätsunterschiede beliebiger HiFi-Audiosignale vorhersagen zu können, waren Erweiterungen und Modifikationen ihres Verfahrens notwendig. Diese werden, zusammen mit den Vorhersageergebnissen, im Folgenden dargestellt.

Methode

Subjektive Bewertung von Qualitätsunterschieden

Die verwendeten Audiosignale und ihre subjektiven Qualitätsbeurteilungen entstammen umfangreichen Hörtests zur Evaluation von Niedrig-Bitraten-Codecs (Kodierungs-Dekodierungsverfahren), die im Auftrag der ITU (*International Telecommunication Union*) und der MPEG (*Moving Pictures Experts Group*) durchgeführt wurden [3-7]. Die subjektive Qualitätsbewertung erfolgte gem. ITU-Empfehlung ITU-R BS.1116 [8]. Dabei handelt es sich um einen Tripel-Stimulus-Test mit verdeckter Referenz, bei dem die Versuchsperson die empfundene Verschlechterung der „globalen Audioqualität" eines Testsignals gegenüber der eines Referenzsignals auf einer fünfstufigen Skala angeben soll. Das Urteil der Versuchsperson wird letztlich in den „*Subjective Difference Grade*" (*SDG*) umgerechnet, der (praktisch) im Bereich [-4, 0] liegt.

Instrumentelle Abschätzung von Qualitätsunterschieden

Kernstück des instrumentellen Verfahrens zur Qualitätsabschätzung ist das Modell der „effektiven" auditorischen Signalverarbeitung von Dau et al. [1] mit einer Modifikation, bestehend aus dem Einfügen einer Begrenzungsstufe hinter die Adaptationsschleifen. Die Begrenzung geschieht mittels Abbildung durch die Funktion $f(x) = 500 \cdot \tanh(x)$ („*soft peak clipping*"). Diese Modifikation stellt sich als vorteilhaft für die signalunabhängige Qualitätsvorhersage heraus (s.u.). Test- und Referenzsignal werden durch dieses Gehörmodell verarbeitet. Die resultierenden Modellausgänge („interne Repräsentationen", dreidimensionale Matrizen) werden anschließend nachbearbeitet und durch Berechnung des linearen Kreuzkorrelationskoeffizienten quantitativ miteinander verglichen. Die Nachbearbeitung besteht, einem Ansatz von Beerends folgend [8], in einer asymmetrischen Angleichung der internen Repräsentation des Testsignals an die des Referenzsignals:

$$IR_{test} \to IR'_{test}: \quad t_{ijk} \to t'_{ijk} = \begin{cases} (t_{ijk} + r_{ijk})/2, & |t_{ijk}| < |r_{ijk}| \\ t_{ijk}, & |t_{ijk}| \geq |r_{ijk}| \end{cases}$$

mit: IR_{test}: interne Repräsentation des Testsignals,
t_{ijk}, r_{ijk}: Elemente der internen Repräsentationen von Test- bzw. Referenzsignal

Negative Änderungen werden dadurch weniger stark gewichtet als positive. Dem liegt die Hypothese zugrunde, dass „fehlende" Komponenten im verzerrten Signal weniger stören als „hinzugekommene" (kognitiver Effekt). Der lineare Kreuzkorrelationskoeffizient der so nachbearbeiteten internen Repräsentationen bildet das objektive Qualitätsmaß q_c. Alternativ lässt sich auch die mittlere quadratische Differenz der internen Repräsentationen berechnen; sie bildet das objektive Qualitätsmaß q_s.

Um eine signalunabhängige Qualitätsvorhersage zu ermöglichen, ist es notwendig, den Zeitverlauf der Momentanqualität zu berücksichtigen. Dies ist motiviert durch Untersuchungen anderer Empfindungsgrößen wie z.B. Lautheit oder Videoqualität, bei denen die Gesamtempfindung nichtlinear vom Zeitverlauf der Momentanempfindung abhängt [9, 10]. Deshalb wird der Parameter q_c zeitaufgelöst berechnet, indem 50ms-Frames der internen Repräsentationen sukzessiv korreliert werden. Die Folge von Kurzzeitkorrelationswerten ergibt das zeitabhängige Qualitätsmaß $q_c(t)$.

Die Abbildung der Zeitreihe $q_c(t)$ auf die Gesamtqualität geschieht durch Berechnung einiger deskriptiver statistischer Parameter, Hauptkomponentenanalyse dieser Größen und schließlich Interpretation der resultierenden Hauptkomponenten durch ein künstliches neuronales Netz.

Die verwendeten statistischen Parameter sind die ersten vier statistischen Momente bzw. daraus abgeleitete Größen (Mittelwert, Standardabweichung, Schiefe, Wölbung), der Modalwert, mehrere Quantile (0.5-, 0.05-, 0.01-Quantil) sowie das untere Quartil der Menge der lokalen Minima (Minima von Intervallen der Länge 1s). Das letztgenannte Maß weist die größte (signalunabhängige) Korrelation mit den subjektiven Qualitätsurteilen auf.

Die Menge der so extrahierten Einzelparameter wird durch Hinzunahme der Gesamtqualitätsmaße q_c und q_s auf 11 erweitert. Eine Hauptkomponentenanalyse (durch *Singular Value Decomposition*) dieser Größen zum Zwecke der Dimensionsreduzierung ergibt fünf transformierte Parameter, die zusammen 97,6% der Gesamtvarianz erklären.

Die letztendliche Abbildung auf die vorherzusagende Gesamtqualität erfolgt durch ein dreischichtiges *Feed-Forward*-Netz mit fünf Neuronen in der verdeckten Schicht. Die subjektiven Qualitätsurteile stellten im Training des Netzes die Zielausgaben dar.

Ergebnisse

Signalabhängige Qualitätsvorhersage

Abb. 1 zeigt die Qualitätsvorhersagen durch den Parameter q_c für die Signale *Stimmpfeife* und *Glockenspiel*. Die Qualitätsurteile clustern entlang signalspezifischer Geraden, d.h. die Qualität wird für jedes Signal getrennt gut vorhergesagt, nicht jedoch für verschiedene Signale gemeinsam.

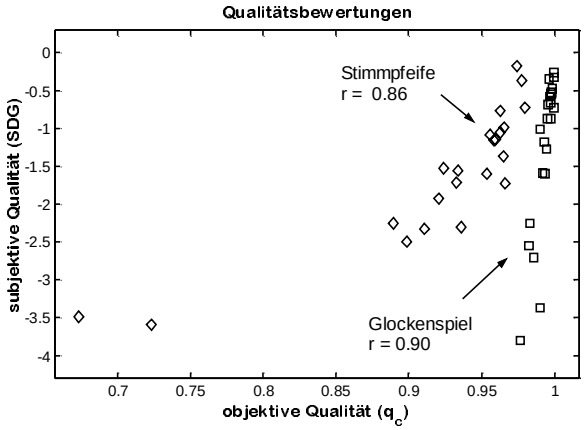

Abb. 1: *Qualitätsvorhersagen durch das objektive Maß q_c (r: linearer Korrelationskoeffizient)*

Dank der guten signalabhängigen Qualitätsvorhersage lässt sich die Qualität von Übertragungssystemen durch Mittelung über mehrere signalweise Einzelbewertungen recht sicher vorhersagen. Eine solche Vorhersage zeigt Abb. 2. Hier wurde über jeweils fünf signalspezifische Bewertungen gemittelt.

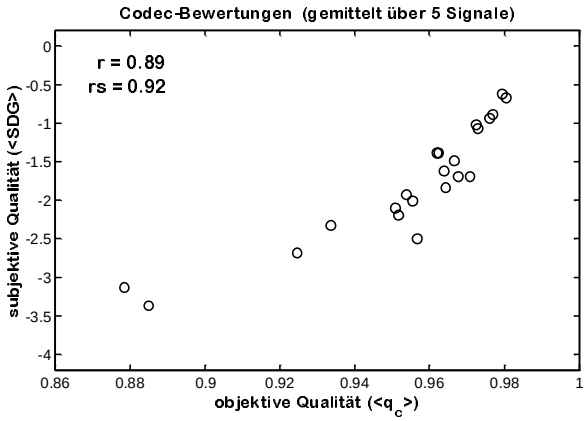

Abb. 2: *Vorhersage von Codec-Qualitätsbewertungen durch Mittelung über fünf signalweisen Einzelbewertungen (rs: Rangkorrelationskoeffizient nach Spearman)*

Signalunabhängige Qualitätsvorhersage

Die in Abb. 1 erkennbare Diskrepanz zwischen den objektiven Qualitätsbewertungen verschiedenartiger Signale konnte durch die oben beschriebene Hinzufügung einer Begrenzungsstufe in das Perzeptionsmodell bereits deutlich verringert werden. Diese Diskrepanz bestand vor allem zwischen wenig und stark fluktuierenden Signalen. Einhüllendenfluktuationen werden durch die Adaptationsstufe des Perzeptionsmodells (zu?) stark kontrastiert. Die Begrenzungsstufe mildert diesen Effekt etwas ab. Ein weiterer Teil der Diskrepanz verschwand durch die Berücksichtigung des Verlaufs der Momentanqualität $q_c(t)$. Stärker fluktuierende Signale zeigen zumeist auch hier größere Variationen, denen durch z.B. bloße Mittelwertbildung nicht genügend Rechnung getragen wird.

In Abb. 3 ist die Qualitätsvorhersage für verschiedene Signale durch das beschriebene neuronale Netz dargestellt. (Die Vorhersageleistung bezieht sich ausschließlich auf solche Signalarten, die nicht im Trainingsdatensatz des Netzes enthalten waren.) In Anlehnung an das subjektive Qualitätsmaß wird die vorhergesagte Qualität mit „*Objective Difference Grade – ODG*“ bezeichnet.

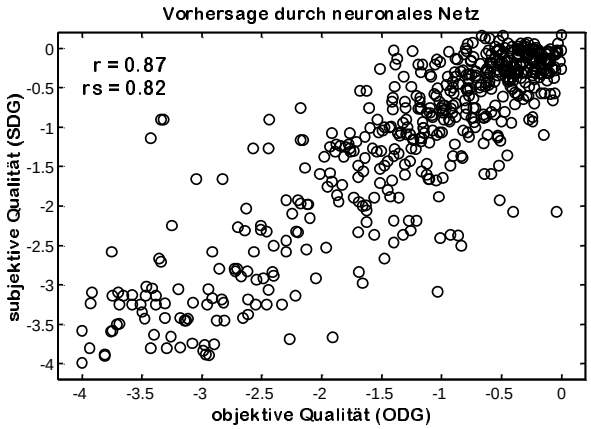

Abb. 3: *Qualitätsvorhersagen für verschiedenartige Signale durch ein neuronales Netz*

Die Vorhersageleistung des Verfahrens für die hier verwendeten Testdaten ist damit von vergleichbarer Qualität wie die des ITU-Standards PEAQ (*Perceptual Evaluation of Audio Quality* [12]) für die in [13] verwendete Test-Datenbasis.

Literatur

[1] Dau, T., Kollmeier, B. und Kohlrausch, A. "Modeling auditory processing of amplitude modulation", J. Acoustical Soc. Am., vol. 102, no. 5, pp. 2892 - 2905, 1997.

[2] Hansen, M. und Kollmeier, B. "Using a quantitative psychoacoustical signal representation for objective speech quality measurement" Proc. ICASSP '97, p.1387-1390

[3] ISO/IEC/JTC 1/SC 2/WG 11 MPEG/Audio test report, Document MPEG90/N0030, October 1990.

[4] ISO/IEC/JTC 1/SC 2/WG 11 MPEG/Audio test report, Document MPEG91/N0010, June 1991

[5] ITU-R, CCIR Listening Tests - Basic Audio Quality of Distribution and Contribution Codecs, Sweden, CCIR-Doc. 10-2/24 (1992).

[6] ITU-R, CCIR *Listening Test - Network Verification Tests without Commentary Codecs*, Canada and Italy, Doc. 10-2/43 (1993).

[7] Meares, D.J., Kim, S-W, "NBC time/frequency module subjective tests: overall results", ISO/IEC JTC 1/SC 29/WG 11 N0973 MPEG95/208, July 1995.

[8] ITU-R Rec. BS.1116 „Methods for the subjective assessment of small impairments in audio systems including multichannel sound systems“

[9] Beerends, J. G. „Modelling cognitiv effects that play a role in the perception of speech quality. Workshop ‚Speech quality assessment‘“, Ruhr-Universität Bochum, Nov. 1994

[10] Fastl, H. „Evaluation and measurement of the perceived average loudness“, in: Contributions to Psychological Acoustics, Vol V, 205-216, Oldenburg, BIS (1991)

[11] Hamberg, R. und de Ridder, H. „Time-varying image quality: Modelling the relation between instantaneous an overall quality“, IEEE Transactions on Systems, Man, and Cybernetics, part A. (eingereicht). Gegenwärtige Version: IPO-Manuskript Nr. 1234 (1997)

[12] ITU-R Rec. BS-1387, „Method for Objective Measurement of Perceived Audio Quality“, ITU, Genf, Schweiz (1999)

[13] Treurniert, W. C. und Soulodre, G.A. „Evaluation of the ITU-R Objective Audio Quality Measurement Method“, J. Audio Eng. Soc., Vol. 48, No. 3, March 2000