CFA/DAGA'04, Strasbourg, 22-25/03/2004

Temporal aspects in the prediction of perceived audio quality differences

Rainer Huber, Birger Kollmeier
Medical Physics Section, Carl-von-Ossietzky-University of Oldenburg, D-26111 Oldenburg, Germany
Email: rainer.huber@uni-oldenburg.de

Introduction

Computational methods for the perceptual evaluation of lossy audio processing systems are required, if subjective listening tests are either too expensive or not applicable (e.g. for real-time monitoring). One approach to assess the transmission quality of such systems adequately to human perception is to estimate the perceived audio quality of the systems output signal relative to the input signal. According to [1], the *basic audio quality* of an audio object is a *single, global attribute to judge any and all differences between the reference and the object.* In this sense, the more similar the input (= reference) and output signal are perceived, the better the audio quality of the output signal and therefore the transmission quality of the system is. According to this understanding of audio quality, it is not an *absolute* quality measure, but a quality *difference* measure. A common concept to predict the perceived audio quality (difference) by computational means is to compare simulated internal representations of reference and test signals obtained with an auditory model. This concept was also applied in the present work, employing the model of the „effective“ signal processing in the auditory system by Dau et al. [2]. As will be shown in the following, the overall correlation coefficient of the internal representations of test and reference signals serves quite well to predict subjective ratings of audio quality, as long as different kinds of audio signals are considered separately. However, if different kinds of audio signals are mixed, the temporal course of the instantaneous audio quality has to be taken into account as well.

Method

A block diagram of the signal processing is shown in Figure 1. The method represents an expansion of the speech quality measure q_C, introduced by Hansen and Kollmeier [3]. It is based on a psychoacoustically validated, quantitative model of the "effective" peripheral auditory processing by Dau et al. [2]. In order to assess the audio quality of a given test signal relative to a reference signal, both test and reference signals are processed by the auditory model and thereby transformed into corresponding internal representations. To model cognitive aspects of auditory perception, the internal representation of the test signal is further processed: Based on an approach by Beerends [4], the internal representation is partly assimilated to the corresponding reference by halving negative deviations from the reference, assuming that „missing“ components in the distorted signal are perceived less disturbing than „additional“ ones. The linear cross correlation coefficient of the assimilated internal representations estimates the perceived overall similarity between reference and test signals and is denoted as *Perceptual Similarity Measure* (PSM), which is quite similar to the speech quality measure q_C of Hansen and Kollmeier. Another more refined measure is derived by computing a sequence of short-time cross correlation values of the two internal representations, yielding PSM(t) (with $t = n \cdot 10$ ms). It serves to predict the perceived *instantaneous* audio quality of the test signal. PSM(t) is weighted by the moving average of the internal representation of the test signal („loudness“ weighting) and finally mapped onto the overall quality measure PSM_t by calculating the (lower) 5% quantile. This step models the relation between the perceived instantaneous and overall audio quality. It is known from other fields of psychophysics that this relation is highly complex. Human observers rather tend to focus on extreme occurrences in the temporal course of a considered psychophysical quantity than integrating over time linearly [5].

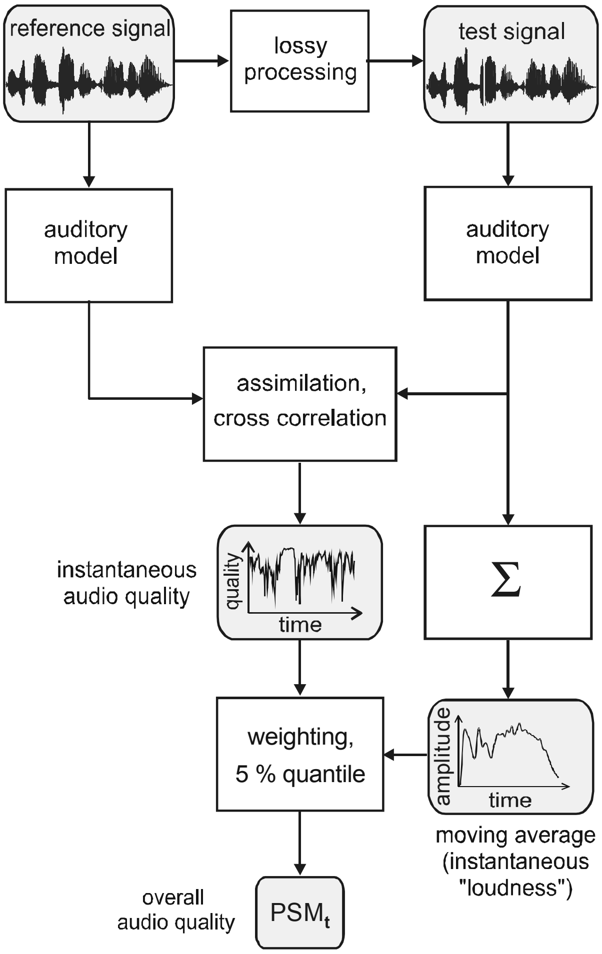

Figure 1: Block diagram of the method for audio quality estimation (see text for description).

Results

The audio quality measures PSM and PSM$_t$ were applied to 439 subjectively rated audio signals. This database emerged from listening tests that were carried out on behalf of the International Telecommunication Union (ITU) and the Moving Pictures Experts Group (MPEG) to evaluate various low-bit rate audio codecs. The subjective assessment was performed according to the ITU-R recommendation BS.1116 [1], yielding ratings of the degradation of the *basic audio quality* on a continuous impairment scale. This scale ranges from 0 = *imperceptible* to –4 = *very annoying*. Intermediate anchor points are: -1 = *perceptible but not annoying*, -2 = *slightly annoying* and –3 = *annoying*. The mean over all subjects is denoted as S*ubjective Difference Grade* (SDG)

Results obtained with PSM

Exemplary results obtained with the measure PSM are shown in Figure 2. While correlations between measured and predicted audio qualities are good if different kinds of signals (e.g. castanets, bag pipe, speech etc.) are considered separately, an apparent signal dependency of the SDG-PSM relation leads to a rather poor correlation if different signals are mixed (cf. lower right panel of Figure 2). Rapidly fluctuating signals such as castanets generally show steeper slopes in the SDG-PSM plane than rather stationary signals, i.e. the audio qualities of these signals are overestimated.

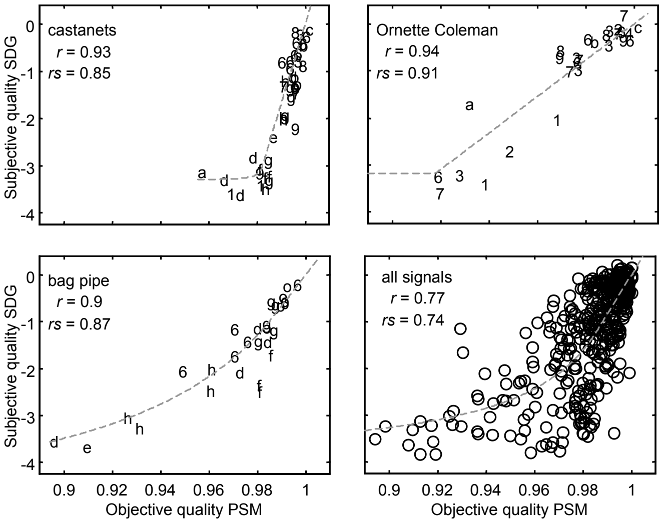

Figure 2: Audio quality predictions obtained with the quality measure PSM, for three different kinds of audio signals separately and 28 kinds of audio signals mixed (439 signals altogether). Mean subjective ratings are given as *Subjective Difference Grades,* SDG. Letters and digits represent different audio codecs. Fit functions of three degrees of freedom are plotted as dashed lines. *r* and *rs* denote the linear and the rank correlation coefficient, respectively.

Results obtained with PSM$_t$

The signal dependency of the quality prediction is strongly reduced if the relation between the perceived instantaneous and overall audio quality is modeled as well. This is realized by the measure PSM$_t$. The results obtained with this measure are shown in Figure 3 for all signals of the database.

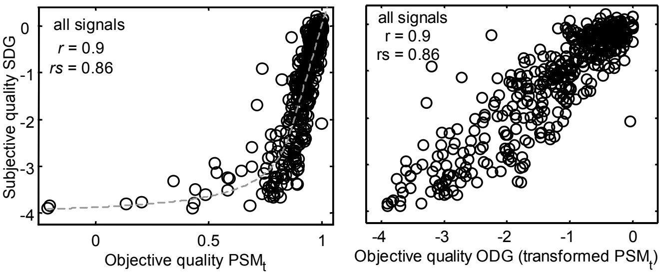

Figure 3: Audio quality predictions obtained with the quality measure PSM$_t$ (left panel), for all audio signals of the database (439 items). Right panel: same data, but transformed abscissa, using the fit function indicated by the dashed line in the left panel. The transformed quality measure is called *Objective Difference Grade* (ODG).

Influence of correlation interval and quantiles

The influence of the length of the correlation interval Δt that is used to compute the instantaneous audio quality PSM(t) = PSM($n \cdot \Delta t$) on the prediction performance was investigated in more detail. Moreover, different quantile measures that are used to map the sequence of instantaneous quality PSM(t) onto the overall audio quality PSM$_t$ were tested as well. Figure 4 shows the prediction performance of PSM$_t$, quantified by the linear correlation between measured and predicted data, as a function of the correlation interval Δt using the 3%, 5% and 10% quantile.

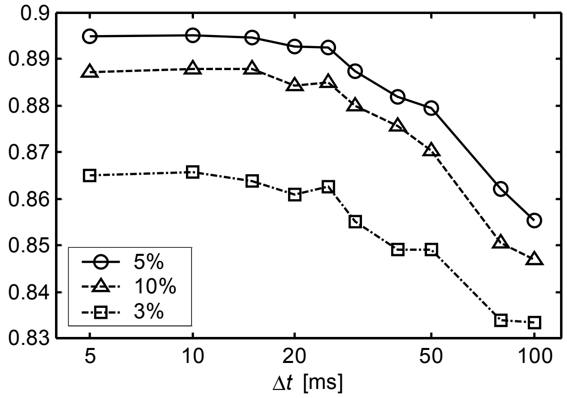

Figure 4: Correlation between subjective quality ratings and PSM$_t$ for different correlation intervall lengths Δt and quantiles.

References

[1] ITU-R Rec. BS.1116: Methods for the subjective assessment of small impairments in audio systems including multichannel sound systems. ITU, Geneva, Switzerland, 1997

[2] Dau, T., Kollmeier, B. und Kohlrausch, A.: Modeling auditory processing of amplitude modulation. J. Acoust. Soc. Am. 1997, vol. 102, no. 5, p. 2892 – 2905

[3] Hansen, M. und Kollmeier, B.: Objective modelling of speech quality with a psychacoustically validated auditory model. J. Audio Eng. Soc., vol. 48, May 2000, p. 395-409

[4] Beerends, J. G.: Modelling cognitive effects that play a role in the perception of speech quality. Workshop ”Speech quality assessment”, Ruhr-Universität Bochum, Nov. 1994

[5] Fastl, H.: Evaluation and measurement of the perceived average loudness. Contributions to Psychological Acoustics, Vol V, p. 205-216, BIS, Oldenburg, 1991