Testing of Objective Audio Quality Assessment Models on Archive Recordings Artifacts

Bc. Martin Zalabáč

1Department of Radioelectronics, Czech Technical University in Prague, Technická 2, 166 27 Praha, Czech Republic
zalabmar@fel.cvut.cz

Abstract. This paper presents results of testing of objective audio quality assessment model PEMO-Q in its original form and with use of CASP auditory model as a substitute of model included in the assessment model. Both versions are tested on artifacts modelled after those present in archive recordings, such as noise and distortion by rectifying.

Keywords
Poster2015, audio, restoration, PEMO-Q, CASP

1. Introduction

Objective audio quality assessment models offer an interesting alternative to subjective quality testing because of time and resource expensiveness of such subjective tests compared to a signal analysis algorithm and because of invariable outputs of objective assessment models. However, because of subjectivity of human hearing, it is very difficult to specify accuracy of such models.

The subjects of testing are artifacts present in digitized archive recordings including both artifacts present in the original analog media and artifacts which emerged in the process of digitizing. The goal is to analyze possibilities of use of assessment models in detection of audible quality decrease caused by such artifacts mainly to identify the best method for digitization of certain signals.

It should be also noted that models presented in this paper are based on comparison of reference and testing signals. Unfortunately many of archive recordings have their original sound media no longer available. Because of this, subsequent tests are done not just by the recommended approach, but also using test signal as a reference and the reference as a test signal to examine the behavior of the models for use of two signals of unknown properties.

2. Artifacts

There are five types of artifacts included in presented test. All of those artifacts are more thoroughly described in [1]. The first tested artifact is noise, or intrusive sound present in the signal along with relevant sounds. Such noise can be present in the original recording due to imperfection of recording equipment or it can appear in the process of digitizing.

The second artifact is change of spectral characteristics. Such changes can appear in any point of processing due to frequency dependence of real analog equipment. Another possible artifact is non-linear distortion, such as compression, rectifying or hard-clipping.

The two remaining artifacts are most commonly related to physical damage of the original media and have an impulse nature. The first one is an impulse error, which appears as a relatively high level signal impulse or noise signal added to or substituting the original signal. The last tested artifact is an impulse loss, which can appear as an empty space in signal or a point with an audibly missing fragment of the original sound.

3. Tested Models

Both tested models are based on objective audio quality assessment model PEMO-Q defined in [3] with block diagram shown in figure 1. The core block of this model is an auditory model which obtains internal representation values of reference and test signals (both signals are synchronized in time, amplified on the same level and processed independently by the model). These internal representation are then evaluated by the model backend with output values $PSM$ (or Perceptual Similarity Measure) and time-dependent $PSM_t$. The latter is then mapped to ODG, or Objective Difference Grade, for the possibility to predict the subjective difference grade.
3.1. Auditory Model PEMO-Q

The first auditory model follows the definition in [3] and its block diagram is shown in figure 2. In the beginning, the input signal is passed through basilar membrane filterbank composed of 35 4th order gammatone filters. Center frequencies are equally spaced on an Equally Rectangular Bandwidth scale (or ERB scale) with one filter per ERB with range from 235 to 14500 Hz. The bandwidth of each filter is also 1 ERB. Outputs of each filter are in subsequent steps processed individually.

The next step is simulated transformation of mechanical oscillations to neural firing rates of the inner haircells done by halfwave rectification and a low-pass filter set on 1 kHz. Such transformation preserves the envelope of frequencies above this threshold, and amplitude and phase below this frequency.

The output of the haircell block is limited to a minimal threshold set to the value of $10^{-9}$ and processed by a cascade of five nonlinear feedback loops with a dividing element and a lowpass filter. This should model a temporal masking and an adaptation. The filters are defined by time constants ranging from 5 to 500 ms. This approximates logarithmic compression, with stationary inputs transformed to its 32th root and rapid changes processed almost linearly.

Finally, the signals are processed by a modulation filterbank. It is a set of eight filters with fixed bandwidth of 5 Hz and center frequencies of 5 and 10 Hz for first two filters. The remaining bandpass filters are logarithmically scaled, with a constant Q-value of 2 and overlapping at -3 dB. For easier processing, the Hilbert envelopes are calculated (with only the real part of result being used) and the signal is downsampled to a sampling frequency at least six times of the center frequency of each modulation channel. Output of the whole model is an internal representation consisting of 35x8 subsignals.

3.2. Auditory model CASP

Model CASP, specified in [4], offers an alternative auditory model to one included in PEMO-Q and its base structure is identical. Haircell transformation, the adaptation cascade and the modulation filterbank are also unchanged. However, the basilar membrane block is different and several new elements are included.

The gammatone filterbank modelling basilar membrane was substituted by a DRNL filterbank described in [5] with modifications from [4]. Single DRNL filter is divided into linear and nonlinear parts. The linear part consists of a linear gain element, a cascade of first order gammatone filters, and a cascade of lowpass filters tuned to the same frequency. In the nonlinear part, both cascades are also present, but tuned to a different frequency. A linear gain element is not present, and a broken stick compression element is put between the two sets of gammatone filters.

A completely new element in CASP compared to PEMO-Q is a filter modelling outer and middle ear frequency characteristics before the basilar membrane filterbank. Output of this filter is a stapes velocity representation and the filter itself is a 512-point finite impulse response filter specified in [4]. Other new elements are, as described in [2] and [4], a gain element increasing the signal level by 50 dB between the haircell and the adaptive cascade blocks, a 150 Hz lowpass filter before the modulation filterbank and also a much lower threshold before the adaptive phase, corresponding with a lower output of the DRNL filters.

3.3. Evaluation

When the values of the internal representation has been calculated, an assimilation step is done through all three domains (time, frequency and modulation band) with following equation,

$$\hat{y}_{tfm} = \begin{cases} \frac{y_{tfm} + x_{tfm}}{2}, & |y_{tfm}| < |x_{tfm}| \\ y_{tfm}, & |y_{tfm}| \geq |x_{tfm}| \end{cases},$$

(1)
where $x$ and $y$ are the elements of internal representations of the reference signal and the tested signal respectively. This step follows an assertion that the subjectively ,,missing" components are less disturbing than the ,,additive" ones by suppressing of the elements with a lower value compared to the reference.

The $PSM$ output value is, as specified in [3], calculated by the process of cross-corellation though time and frequency domains with weighted sum through modulation band domain done by subsequent equations,

$$r = \frac{\sum_{t,f}(x_{tf} - \bar{x})(y_{tf} - \bar{y})}{\sqrt{\sum_{t,f}(x_{tf} - \bar{x})^2 \sum_{t,f}(y_{tf} - \bar{y})^2}}.$$

$$PSM = \sum_{m} w_m r_m,$$  \hspace{1cm} (3)

with the weight specified by

$$w_m = \frac{\sum_{t,f} y_{tf}^2}{\sum_{t,f} y_{tf}^2}.$$  \hspace{1cm} (4)

The values $\bar{x}$ and $\bar{y}$ are mean values of internal representations through time and frequency.

The time dependent value $PSM_t$ is obtained by short time (10 ms) cross-correlations, weighting by an ,,instant audio activity" - moving averages of the same short-time frames, and obtaining the 5% quantile of those weighted values. Both $PSM$ and $PSM_t$ measures attain values from 0 to 1, with 1 indicating identity. $PSM_T$ is mapped to $ODG$ by this function

$$ODG(x) = \begin{cases} \max\{-4, \frac{a}{x - b} + c\}, & x < x_0 \\ d \cdot x - d, & x \geq x_0 \end{cases}$$  \hspace{1cm} (5)

with values $a = -0.22$, $b = 0.98$, $c = -4.13$, $d = 16.4$ and $x_0 = 0.864$.

4. Test Description

The test itself is realised in MATLAB by modelling of the artifact on the reference signal and obtaining the output values of both versions of model as a dependence of ,,intensity" of the chosen artifact’s influence.

The effect of noise is modelled by an additive pink noise with the range of levels from -100 to 20 dB with step of 2 dB. Impulse errors are mimiced by 1 ms long noise substituting the signal. These errors are uniformly distributed and their ,,intensity" is specified by a density compared to the total length of the signal with range set from no errors to 10% with step of 0.2%. A fixed length, an uniform distribution and a variable density are also the properties of a signal loss modelling, simulated by removal of 250ms long fragments of signal with a range from 0 to 90% with the step of 2%.

For simplicity and for possibility to use a single variable, the only examined change of spectral characteristics is bandwidth limiting modelled by a first order Butterworth low-pass filter with cutoff frequency in logarithmic range from 20 to 0.1 kHz. For similar reasons the distortion artifact is mimiced by a halfwave rectification with a range from minimal value of -1 (or no distortion) to 0 (or full halfwave rectification) with the step of 0.02.

Described models of artifacts are applied on three signals. The first one (Sig1) is five seconds long log-sweep sine with range from 20 to 20000 Hz and level of 0.8. The remaining two (Sig2 and Sig3) are sound samples from the SQAM collection [6]: a male speech in English (sample no. 50) and a sample of orchestra (first 15 seconds of sample 68).

5. Results

The results of the described test for signal loss is shown in figure 3, for rectification in figure 4 and for rest of artifacts in the appendix. $PSM_t$ values are ommitted, as $ODG$ values carry the same information.

For signal loss, the models fail to detect the degrading quality of signals, as with high density of missing fragments.
the output levels begin to rise. This issue is much less significant in an inverted testing (using testing signal as a reference). In case of lowpass filtering, the assessed quality fall with frequency, although with local extremes and slightly higher sensitivity in inverted testing.

Impulse errors are detected with relation similar to exponential. In case of noise levels, unmodified PEMO-Q values start to fall above -50 dB, in case of orchestra above -10 dB. The time-dependent values appear to have local extremes above this level. Values of model using CASP start to fall at very high levels of noise, above 0 dB.

Distortion by rectifying appears on the results of log-sweep sine at the amplitude of the signal with very sudden fall. In case of „real” signals, values start to fall at almost full halfwave rectification. The difference is especially small with pure PEMO-Q and the speech signal.

6. Summary and Conclusion

In most cases, objective audio quality assessment model PEMO-Q with or without auditory model CASP detected increasing influence of the artifacts as a loss of quality. The only exception is signal loss, although rectification of speech signal with use of unmodified PEMO-Q shown highly non-monotonous behavior. This can be the result of modelled masking.

Inverted evaluation, or using testing signal as a reference, shown minor differences except for signal loss, where the results were monotonous. Possible explanation is the influence of assimilation, where the signal loss was evaluated as nonsignificant compared to a signal addition.

The actual accuracy of output relations still remains unknown, as it can not be determined without comparison with subjective tests. The correctness of implemented models was also not verified, which would be appropriate for further research.

Acknowledgements

This work was supported by the Grant Agency of the Czech Technical University in Prague, grant No. SGS14/204/OHK3/3T/13.

References


324–336.


Appendix - Output relations of remaining artifacts

About Authors...

Bc. Martin Zalabák was born in Prague and is studying Multimedia Technology on the Czech Technical University. He obtained his bachelor degree in 2013. The subject of his bachelor thesis was “Sound Effects Real-Time Implementation”.

CASP - Error Density

PEMO-Q - Error Density

PEMO-Q - Error Density (inv.)

CASP - Error Density (inv.)