PERCEIVED QUALITY OF AN AUDIO SIGNAL IMPAIRED BY SIGNAL LOSS: PSYCHOACOUSTIC TESTS AND PREDICTION MODEL

Ricardo R. Pastrana-Vidal, Catherine Colomes

France Telecom, Research Division
QVP Lab, Cesson Sévigné, France,

ABSTRACT

Sporadic signal loss is a temporal degradation that can be found in audio streaming over mobile internet or triple play services. Radio channel errors, packet loss or reconstruction strategy failures can lead to irregular audio discontinuities having a negative impact on the end-user quality assessment of the sequence. We therefore propose two models that predict the quality assessment from a group of listeners when an audio wideband sequence is impaired by a distribution of signal losses. The model showing the highest performance takes into account non-linear listeners' assessment characteristics based on psychoacoustic experiments.

Index Terms— Auditory quality perception, signal loss, audio streaming, modelling

1. INTRODUCTION

The advent of protocols for quasi real time communications and transmission systems with greater bandwidth has motivated an increasing deployment of real time audio services. Audio streaming is now available from PC stations and 3G mobile devices. Real time applications can only tolerate a short delay in the signal reconstruction. Unfortunately, packets of media data are transmitted over unreliable, lossy networks. Audio signal may therefore be impaired.

Packet loss or jitter could cause a sporadic or non-uniform signal loss during the decoding process because of the playout buffer time limit. Signal loss can be partially overcome by recovery techniques. However, in the presence of significant transmission errors recovery techniques may fail and audio loss is unavoidable. Furthermore, in real time audio applications a retransmission request is not always possible. For instance, retransmission is not feasible in mobile multicast streaming systems [1]. The application should therefore anticipate some data loss meaning that it is absolutely necessary to measure the quality.

When considering quality, it is essential to consider the end user quality perception because they are the final recipient. When signal loss occurs, the end-user perceives a silence followed by an abrupt "click" which has a negative impact on his quality assessment of the sequence. Nowadays, psychoacoustic experiments are the only recognized way to characterize the perceived quality.

To the best of our knowledge, the effect of signal loss on audio wide-band (like music) has still not been widely analyzed. The effect of this type of temporal discontinuity was studied using voice sequences [2]. Voran studied the impact of a reduced number of signal losses (1 to 4) each one with the same duration (either 30, 60, or 120ms). The combination of different loss duration was not included. Recently, authors in [3] quantified the effect of transmission scenarios on the quality of wide-band speech. However, both studies conducted for speech signals are not directly valid for other contents. There is evidence showing that the listener's reaction to a single audio loss is content dependent [4]. The aim of our work is twofold: to characterize the effect of signal loss on listener perception using audio wideband and to build a mathematical model able to reproduce the listener's behavior when assessing the quality of a sequence impaired by a distribution of signal losses.

2. PREVIOUS STUDY

2.1. Audibility

The first goal of our study [4] was to measure the audibility of an isolated audio loss as a function of its duration. Audibility means the strength at which a temporal discontinuity is just perceptible to listeners. Six audio excerpts were used. Those sequences are listed in Table 1.

Table 1. Audio sequences contents and short name

Content	Short name
Speech	Newsreader
Speech and Noise	Basket
Music	Guitar
Singer	Chorus
Singer, Music	Jazz
Music, Speech, atmosphere noise	Film

The impaired sequences present signal losses placed at different temporal positions presenting audio activity. Signal

losses were introduced to replace the original samples with silence. This process allows a better control over all the experiment chain.

We observed that unequivocal detection is attained at 30ms of loss duration. This result is valid for the selected temporal positions and for all contents. In addition, the detection threshold (just perceptible) varies from 1.2 to 6.1ms.

2.2. Quality impact of an isolated loss

The quality test described in [4] showed that an isolated signal loss of short duration can have a strong negative effect on quality perception and this effect is content dependent. For instance, listeners had a significant negative reaction to a discontinuity of 30ms. Quality function exhibits a fall in rating of 30%. In general, the audio loss was more annoying in music than speech contents.

3. QUALITY IMPACT OF DENSITY

Two tests were conducted to characterize the impact that audio losses at different densities have on quality. The first experiment was conceived to quantify the degradation caused by a variable number of discontinuities each one with the same duration. The second experiment studies the impact of several audio losses of different durations.

3.1. Listeners and method

The quality ratings from twenty five normal hearing participants were gathered using the MUSHRA (Multi Stimulus test with Hidden Reference and Anchors) method [5]. A numerical scale (0-100) for rating overall quality is used. This scale is related to five quality categories (bad, poor, fair, good and excellent) that are uniformly distributed.

The signal presentation time is limited to 10s in order to avoid the forgiveness effect [6].

3.2. Stimuli

The six original excerpts used here were the same as the contents described in the previous experiments ($\oint$ 2.1) in order to maintain the content's consistency. Coding degradation was not introduced in order to isolate the effect of audio loss density on the quality assessment and to facilitate modeling.

Duration choice: We restricted loss durations to 30 and 150ms in order to reduce the test duration and to guarantee loss detection. In this way the study was exclusively concentrated on the listener's assessment mechanisms.

Density choice: We pointed out that a single audio loss can cause a dramatic fall in quality of 30 percent. Furthermore, a quality test evaluating audio transmitted over a simulated IP network [7] showed that some temporal discontinuities, from 3 to 5, which have a duration ranging from 200 to 500ms, lead to a quality considered as "Bad". At the sight of these results, the defined numbers of discontinuities are: 1, 3, 5 and 8.

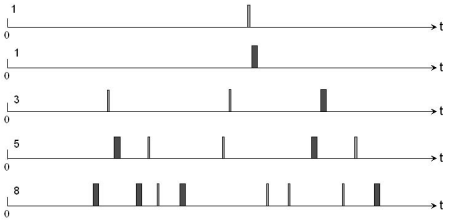

Fig. 1. Impairment profiles used in "density test-2". A bar represents a signal loss. Fine bars: loss duration of 30ms. Large bars: loss duration of 150ms.

Density test-1, losses of the same duration: the degradation profiles exclusively contain losses of 30ms or 150ms.

Density test-2, losses of different duration: the profiles are composed by discontinuities of 30 and 150ms (Figure 1).

Low quality anchor: for the "density test-1" we used a profile of discontinuities (audio losses of $\{360, 430, 469, 615, 630\}$ ms) taken from a streaming audio simulation test previously conducted. For the "density test-2" we used a profile of 8 discontinuities of 150ms each one as a low quality anchor. Complementary anchor: a bandwidth limitation of 3.5 kHz in the original sequences as recommended by the MUSHRA method. This is a common condition for all the tests.

4. QUALITY RESULTS

The data of the experiment consists of a mean opinion score (MOS) and the 95% confidence intervals for each stimulus calculated after a post-screening of the participants. The MOS over all contents $\overline{c} = \{c_1, c_2, \cdots, c_6\}$ for the d_j degradation is calculated as follows:

$$MOS_{\overline{c},d_j} = \frac{1}{6}\sum_{i=1}^{6}\left[\frac{1}{A}\sum_{a=1}^{A} OS_{c_i,d_j,a}\right] \quad (1)$$

with $OS_{c_i,d_j,a}$ = score given by the listener a for the degradation d_j applied to the content c_i and A = total number of listeners.

The MOS are plotted as a function of the number of audio losses or as a function of the cumulated duration of all losses:

$$T_c = \sum_{p=1}^{L} t_p \quad (2)$$

where L is the total number of signal losses in the sequence and t_p is the duration of the p discontinuity.

Comparing MOS from the three tests as a function of the cumulated losses duration (Fig. 2), we can see that the perceived quality exhibits a dramatic fall given a reduced number of discontinuities. For instance, three losses of 30ms are

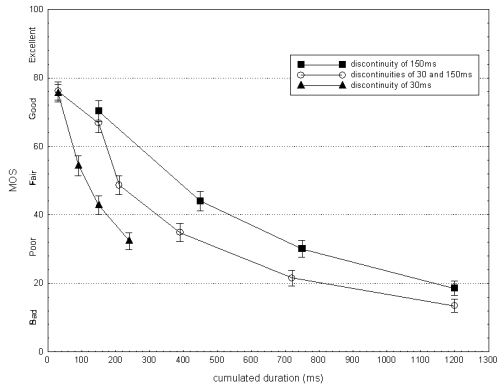

Fig. 2. Quality vs. cumulated duration: three tests comparison.

enough to lose 46 points (over 0 to 100 scale) attaining the "Fair" category. It is interesting to note that it is not possible to establish a univocal relationship between cumulated impairment duration and mean opinion scores. For instance, a cumulated duration of 150ms obtained by the combination of 5 losses of 30ms or by a single loss of 150ms can respectively give a quality of 42 "Fair" or 66 "Good". This is a surprising result because participants were more annoyed by several short discontinuities than a single one with the same cumulated duration.

5. PREDICTION MODELS

The results from the impact of audio losses on quality perception led us to propose two prediction models to reproduce the listener assessment of an audio sequence impaired by signal losses.

5.1. PeMoAL Model

We propose a prediction model combining the quality function for an isolated audio loss, the discontinuities density as a function of their duration (histogram) and a weighting function depending on loss density:

$$\widehat{Q} = 100 - d_{total} \quad (3)$$

$$d_{total} = \min\left\{d_{cumulated}, d_{max}\right\} \quad (4)$$

$$d_{cumulated} = \left[\sum_{t_p=T_{min}}^{T_{max}} d_{t_p}\right]^{1/2} \quad (5)$$

$$d_{t_p} = n(t_p) \times [\widehat{e}(t_p)]^{p(n(t_p))} \quad (6)$$

$$\widehat{e}(t_p) = 100 - \widehat{q}(t_p) \quad (7)$$

$$\widehat{q}(t_p) = m_{max} - \frac{(m_{max} - m_{min})}{(1 + (b/t_p)^s)} \quad (8)$$

$$p(n(t_p)) = -a \times \ln(n(t_p)) + c, \quad (9)$$

d_{total} is the global degradation in the 10s sequence. This term corresponds to the cumulated impact of all audio losses $d_{cumulated}$ limited to the maximal degradation d_{max} accounting for boundary scale effect. In practice, the MOS will be greater than 0 even for extreme degradations. On the right side of (5) t_g is the loss duration, T_{min} is the minimal duration of a perceptible discontinuity ($T_{min} > threshold$) and T_{max} corresponds to the analysis time window of 10sec.

The d_{t_p} term is the calculated contribution from all discontinuities having a duration of t_p; the term $n(t_p)$ corresponds to the loss duration distribution, i.e., the number of discontinuities as a function of their duration (histogram).

The expression $\widehat{q}(t_p)$ is the quality function for an isolated signal loss with a duration of t_p. This function was obtained by fitting experimental assessment data for a single loss. In this range, participants showed an unequivocal probability of detection. The constants m_{max} and m_{min} are the quality limits found in the psychoacoustic experiment.

The impact of each discontinuity is weighted by the power function $p(n(t_p))$ that depends on the distribution of loss durations. By means of this variable exponent, the impact of each discontinuity is indeed less significant when the number of audio losses of the same duration increase. This function was obtained by fitting the data of optimized exponents for several loss densities of a constant duration.

5.2. Iso-Densi Model

A second model is proposed: $\widehat{Q}_{iso} = \alpha(l) \times \ln(T_c) + d$. The variable T_c corresponds to the losses cumulated duration (2), $\alpha(l) = -0.7971 \times l - 3.8689$. $\alpha(l)$ is the modulation function width regards to the number of signal losses l. This second model is interesting because of its simplicity.

5.3. Models' Evaluation

Model performance was evaluated by comparing model predictions against quality assessment scores from a group of participants. In order to evaluate the accuracy and monotonic property of model predictions, Pearson's correlation coefficient (r), determination coefficient (r^2), Spearman's rank order correlation (r_s) and standard prediction error (e) were used. Performance criteria were calculated over the test conditions cited in previous sections. Five degradation profiles, from an independent test, were added to improve the evaluation data base. This independent test evaluates the quality of different coders under a simulated audio streaming system.

5.3.1. Global evaluation

Firstly, we will analyse the correlation over all contents, i.e. the correlation between the pairs $(MOS_{\overline{c},dj}, \widehat{Q}_{c,dj})$, where $MOS_{\overline{c},dj}$ is the assessment score for content set (1) and $\widehat{Q}_{c,dj}$ is the model quality estimation. The evaluation data base consists of 33 impairment conditions from 6 experiments.

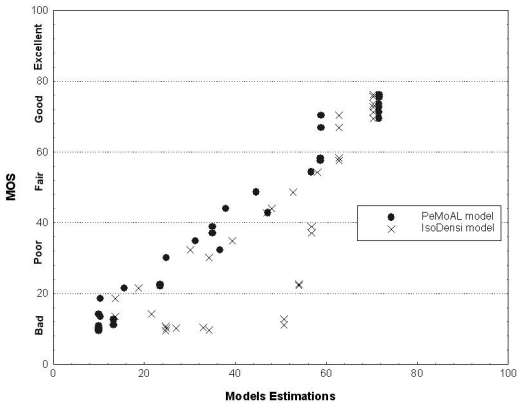

Fig. 3. Scatter plots of mean content MOS and models predictions

By comparing the correlations we can see that the PeMoAL model presents a higher performance than the IsoDensi model. PeMoAL: $r = 0.98$ and $r_s = 0.97$. IsoDensi: $r = 0.84$ and $r_s = 0.84$. We have observed (outliers in Fig. 3) that the IsoDensi model is not well adapted for predicting the impact of an isolated audio loss of long duration (greater or equal to 550ms) nor to estimate the impact of a series of severe discontinuities.

5.3.2. Evaluation by content

The model performance is then evaluated with regards to each content, i.e. the correlation between the pairs $(MOS_{c_i,dj}, \widehat{Q}_{ci,dj})$ where $MOS_{c_i,dj}$ is the mean opinion score for the content c_i having the degradation d_j:

$$MOS_{c_i,d_j} = \frac{1}{A}\sum_{a=1}^{A} OS_{c_i,d_j,a} \qquad (10)$$

o $OS_{c_i,d_j,a}$ is the score from participant a corresponding to the content c_i having the impairment d_j.

The correlation for each content (Table 2) was calculated excluding modelling conditions. In order to improve the analysis, the estimation standard error (e) was also computed.

Table 2. PeMoAL correlation by content type excluding conditions used for modelling.

Content	Determination	Spearman	stdr error
Singer,Music	0,9720	0,9643	4,5051
Speech,Noise	0,9852	1	3,2773
Speech	0,9885	1	2,9024
Mixture	0,9917	0,9832	2,3262
Singer	0,9923	1	2,3692
Music	0,9967	1	1,4816

In general, the PeMoAL model shows a significant performance for every content. PeMoAL is better adapted for music and song sequences. The estimation error increases when sequences contain speech or mixtures.

6. CONCLUSION

Listeners exhibit a strong negative reaction to a single audio loss of short duration. This reaction increases when the audio signal is impaired by a series of losses. The quality function varies in a non-linear manner as a function of number and duration of every signal loss present in the sequence.

An assessment quality prediction model for audio losses was proposed. The model predictions show a high correlation with listeners' ratings. The model can cope with different temporal impairment distributions: isolated and variable density losses.

The prediction model was conceived to estimate the mean effect of signal losses for a set of different audio contents. An extension of this model, adapted to the type of contents and the semantic context, shall include audio activity around the signal loss.

The psychoacoustical experiments and the introduced prediction model are useful for audio quality metric design, for monitoring audio steaming and for a better understanding of time-varying quality.

7. REFERENCES

[1] S. Jumisko-Pyykkö, V. Kumar, and J. Korhonen, "Unacceptability of instantaneous errors in mobile television: from annoying audio to video," in *MobileHCI*, NY, USA, 2006, pp. 1–8, ACM Press.

[2] S. Voran, "Perception of temporal discontinuity impairments in coded speech - a proposal for objective estimators and some subjective test results," in *MESAQIN*, Prague, 2003.

[3] S. Moller, A. Raake, N. Kitawaki, A. Takahashi, and M. Waltermann, "Impairment factor framework for wideband speech codecs," *ASL Processing, IEEE Transactions on*, vol. 14, no. 6, pp. 1969–1976, 2006.

[4] R.R. Pastrana-Vidal, C. Colomes, JC. Gicquel, and H. Cherifi, "Sporadic signal loss impact on auditory quality perception," in *MESAQIN*, Prague, 2004, pp. 37–42.

[5] ITU-R BS.1534-1, "Method for the subjective assessment of intermediate audio quality level of coding systems," Recommendation, ITU, June 2001.

[6] L. Gros, *Evaluation subjective de la qualité vocale fluctuante*, Ph.D. thesis, Université de la Méditerrane Aix-Marseille II, 2001.

[7] C. Colomes, M. Varela, and JC. Gicquel, "Subjective audio tests: Quality of some codecs when used in IP network," in *MESAQIN*, Prague, 2004.