Sporadic Signal Loss Impact on Auditory Quality Perception

Ricardo R. Pastrana-Vidal*°, Catherine Colomes*, Jean-Charles Gicquel*, Hocine Cherifi°

*France Telecom R&D, DIH/EQS/MAI, 4 rue de Clos Courtel, 35512 – Cesson Sévigné, France
Phone: +33 2 99 12 38 49, fax: +33 2 99 12 37 13
{ricardo.pastrana, jeancharles.gicquel, catherine.colomes}@francetelecom.fr

°LIRSIA, Université de Bourgogne, Faculté de Science Mirande, 21011 Dijon cedex, France
Phone: +33 3 80 39 68 49, fax: +33 3 80 39 58 87
cherifi@u.bourgogne.fr

ABSTRACT

Over the past few years, there has been an increasing interest in real time audio services over packet networks. For quality evaluation, it is essential to quantify user perception of the played-out audio sequence. Signal losses are one of the most common degradations in audio streaming at low bit rate. The end-user perceives a silence followed by an abrupt clipping. Cell loss in the packet networks or restitution strategy could be the origin of this perceived temporal audio discontinuity. We present a psychoacoustic experiment performed to quantify the effect of the sporadic audio loss on the overall perceived quality. First, the perceptual detection thresholds of generated temporal discontinuities were measured. Then, the quality function was estimated in relation to a single signal loss of different durations. We have found that the thresholds are content, local activity and loss duration dependent. The quality function estimated in relation to a burst of lost samples of different durations is presented. The analysis of the results is useful for audio quality metrics design and for a better understanding of time-varying quality.

Keywords: Signal loss, quality, audio streaming

1. INTRODUCTION

The advent of protocols for quasi real time communications and the increasing computer powering have motivated an increasing interest in real time audio services over packet networks. For real time applications, audio streaming is the technology solution because the data needs to be transmitted as soon as it is generated in order to deliver continuous media play out. These applications can only tolerate a short delay in the signal restitution. However, packets of media data are transmitted over unreliable, lossy networks [1]. Packet loss could produce significant temporal impairments in the received audio.

When considering quality, it is essential to quantify user perception of the played-out audio sequence. Signal losses are one of the most common degradations in audio streaming at low bit rate. The end-user perceives a silence followed by an abrupt clipping. Cell loss in the packet networks or restitution strategy could be the origin of this perceived temporal audio discontinuity. Packet loss or jitter could cause a sporadic or non-uniform signal loss at the decoding process because of the play-out buffer time limit [2]. In the following text, we will use the term temporal discontinuity as a synonym of audio loss.

Nowadays, psychoacoustic experiments are the only recognized way to characterize the perceived quality. We have therefore conducted a psychoacoustic experiment performed to quantify the effect of the sporadic audio loss on the overall perceived quality.

We present a psychoacoustic experiment performed to quantify the effect of the sporadic audio loss on the overall perceived quality. First, the perceptual detection thresholds [3] of generated temporal discontinuities were measured. Then, the quality function was estimated in relation to a single signal loss of different durations.

Six original audio contents were selected. We have inserted artificially-controlled signal losses covering a broad range of impairments. Coding degradation was not introduced in order to isolate the effects of signal losses on the quality assessment. The adopted methodology was inspired from a previous study of the effects of temporal discontinuities in video streaming [4] on quality perception.

We have therefore organized the study in two main parts: the audibility of an isolated signal loss and its effect on user quality perception.

2. AUDIBILITY EXPERIMENT

Before studying audio loss effects on quality perception, we were studying at which discontinuity strengths (duration) subjects are able to detect audio loss. Thresholds for detection were calculated in two different audio activity contexts in order to quantify the range of signal loss detection.

2.1. Apparatus

The test was carried out in a listening controlled environment. An acoustically isolated chamber was used. The environmental conditions were based on the ITU-R BS.1116-1 [5] recommendation. All audio sequences were stored and played on a PC station. The signal coming from the PC sound card undergoes a digital to analogue conversion by a 24 bits 3DLab DAC 2000 converter followed by an SRM-006t amplifier. The restitution equipment was a professional headphone STAX Signature SR-404. A software tool for audio evaluation was used to perform the audibility test. The software tool gathers participant's answers by means of an evaluation interface.

2.2. Stimuli

Six original audio contents were selected: two speech signals (news and commentaries of a basketball game), a song sequence, an instrumental music (rock), one signal containing a song and instrumental music (jazz) and the last content is a sequence mixing dialogues, background and incidental music. The six selected sequences cover a broad range of audio contents. The sequences were all of 10 sec to avoid fatiguing. The audio format was PCM, 48 000 Hz, 16 bits, stereo. We have inserted artificially-controlled signal losses covering a broad range of impairments around the audibility threshold. The impaired sequences present signal losses placed at different acoustic activity contexts (in terms of perceptual annoyance, local bandwidth and energy distribution among frequencies). The selection of activity context was done by a previous informal test. Coding degradation was not introduced in order to isolate the audibility of a signal loss alone.

The News content is a speech sequence from a male French-language newsreader. The Sport commentary contains speech from a basketball game commentator and crowd noise. For the Song content we have selected the sound of three males singing in chorus. The Instrumental content is the music of an acoustic guitar alone. The named Jazz sequence consists of a female singer and instrumental music. Dialogues, background and incidental music make up the Mixture content.

The discontinuity caused by signal loss was inserted in the reference audio sequences replacing original samples by silences (Fig. 1). Thresholds for detection were evaluated over a range of impairment durations: 0.2, 1, 5, 10, 30 and 60ms. For a given impairment duration, we have selected two different activity contexts in order to cover extreme conditions (low and high audibility).

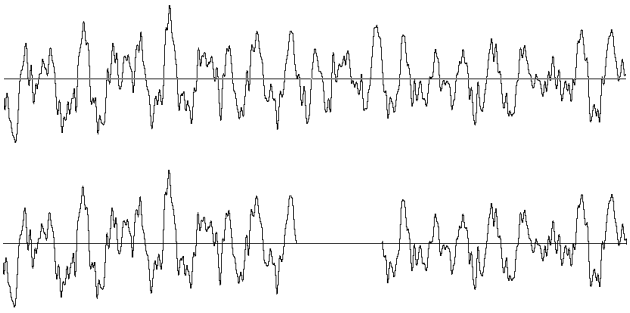

Fig. 1. Audio loss example. Upper: original instrument sequence. Bottom : sequence impaired by a signal loss

2.3. Auditors and method

Twenty six subjects, ranging from 18 to 31 years of age, participated in the experiment. All had normal hearing. Subjects were paid for their participation. The four-alternative forced choice method [3] was employed to evaluate the thresholds. The auditors were asked to identify which of 4-alternatives sequences contained the impairment (Fig. 2). Each sequence is randomly related to the selection buttons. In addition, each set of four-alternative was presented in a random order to avoid the tiredness influence caused by the presentation order. The detection test was organized in two sessions of about 24min each.

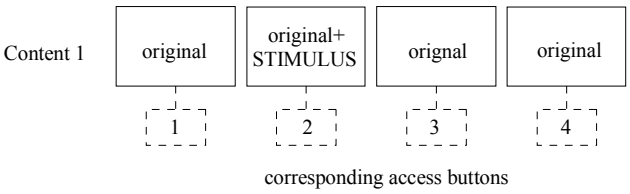

Fig. 2. Four-alternative forced choice method

2.4 Results of audibility

The first goal of our study was to measure the audibility of an isolated audio loss as a function of its duration. The probability of detection as a function of the discontinuity

strength (duration) was calculated in order to characterize this audibility. A threshold for detection was defined as the discontinuity duration such that the generated impairment was listened by 62.5% of the participants. The so defined threshold corresponds to the middle of the detection rate between 25 and 100%. It is assumed that the probability of detection tends to be 0.25 as the impairment strength tends to be zero because of the 4-alternative forced choice. When the audio loss is not audible, all of the four alternatives have the same probability of being selected as the right answer.

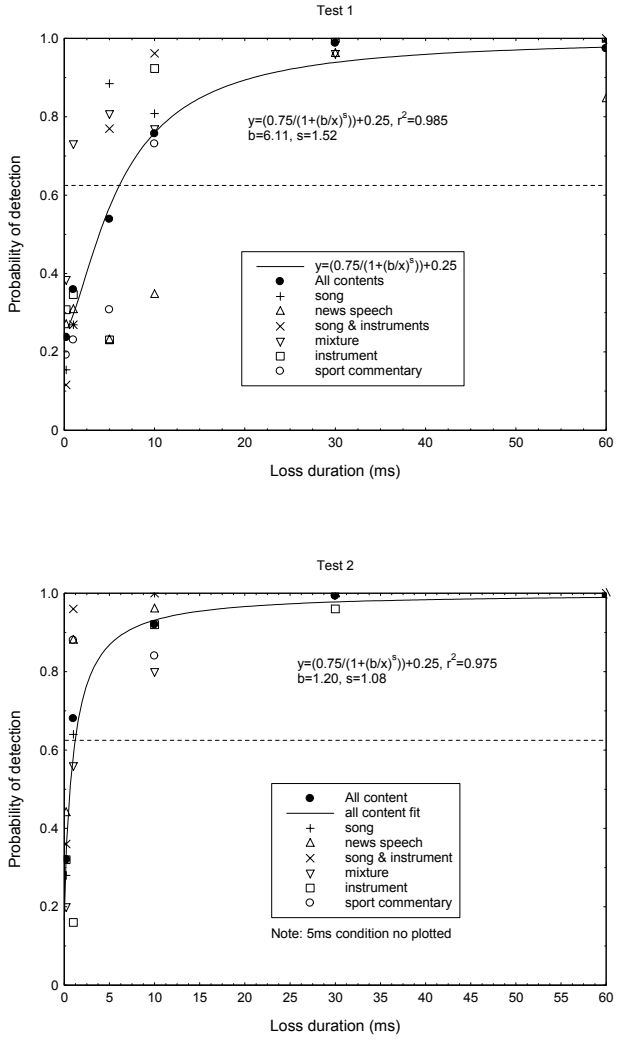

Fig. 3. Probability of detection as a function of audio loss duration: results by content and the fitting by logistic function. Upper graph for low audibility context and bottom graph for high audibility context.

The probability of detection as a function of discontinuity duration by contents is plotted in Fig. 3. The top graph corresponds to the test of low audibility context and the bottom graph shows the results of the high audibility context. The results of the overall detection task were fitted by a logistic function:

$$y = ((p_{max} - p_{min}) / (1 + (b/x)^s)) + p_{min} .$$

In the case of low audibility context (Fig. 3) the Newsreader content presents the lowest probabilities of detection for most of the loss durations (Table 1). The sequence composed of dialogues, background and incidental music shows the highest audibility of audio losses.

Table 1. Mean probability of all loss durations for each content. Low audibility context.

Content	**Mean Probability**
Newsreader	0.49
Sport commentary	0.57
Instrument	0.63
Song & Instruments	0.68
Song	0.68
Mixture	0.77

For most of the contents, a signal loss of 10ms is detectable (with the exception of News content). The unequivocal detection, close to the probability of 1, is attained for a discontinuity of 30ms. This result is valid for all contents. We have found that the detection threshold corresponds to loss strength of 6ms.

At the bottom of Fig. 3, we have the probabilities of detection in high audibility context. For clarity proposes the 5ms condition has not been plotted because two sequences containing speech exhibited quite a low probability (0.24 and 0.28). A discontinuity of 10ms has a high probability of being detected. This is valid for most of the contents. For all selected contents, the unequivocal audibility is attained for a loss duration of 30ms. The data fitting line shows that the audibility of an isolated signal loss is more significant in this context. The threshold of audibility corresponds to a discontinuity of 1.2 ms.

Table 2. Mean probability of all loss durations for each content. High audibility context.

Content	**Mean Probability**
Instrument	0.65
Mixture	0.71
Sport commentary	0.72
Newsreader	0.75
Song	0.80
Song & Instruments	0.89

In Table 2, the contents are arranged by the overall probability of detection computed over the six loss durations. We can see that the Instrument (guitar) content presents the lowest audibility and Song & Instruments (Jazz) has the highest. The content ranking is quite different to the first experiment (low audibility context). Only the Song (chorus) sequence has kept the same ranking.

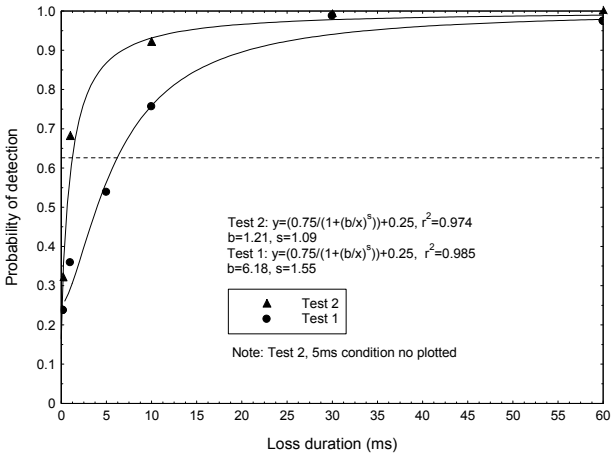

Fig. 4. Probability of detection as a function of audio loss duration. Results of two activity contexts: top graph corresponds to high activity context, bottom graph to low activity.

Fig. 4 let us compare the two audibility contexts. Unequivocal detection is attained at 30ms of loss duration. This result is valid for both audibility context and for all selected contents. In addition, the threshold for detection varies from 1.2 to 6.11ms.

After the audibility results, the question now is how signal loss durations close to and greater than the audibility threshold affect auditory quality perception.

3. QUALITY EXPERIMENT

The main goal of this experiment was to characterize the effect of an audio loss of different duration on perceived quality. The experiment was organized in two test sessions: low and high annoyance contexts. The selected temporal positions (context) were taken from the previous audibility experiment.

3.1. Apparatus

The test apparatus for the quality assessment experiment was the same as the material described in the audibility experiment. A software tool was used to carry out the test. The listener assesses explicit reference and N test conditions (including hidden reference and anchors) using a software slider placed over the selection button.

3.2. Stimuli

The six original sequences used for the quality experiment were the same as the contents described in the audibility experiment. We have inserted an artificially-controlled signal loss covering a broad range of impairments. The impaired sequences present signal losses placed at different acoustic activity contexts. The duration of the impairments was greater than or equal to the detection thresholds and they were selected with enough perceptual distance. Coding degradation was not introduced in order to isolate the effects of signal losses on the quality assessment.

The quality experiment was organized in two tests following the same approach as the audibility experiment, i.e., one test for impairments placed in low annoyance activity context and the second test for higher annoyance. The selection of these temporal contexts was done a priori by an informal inspection conducted by one of the authors.

The impact of a single audio loss was evaluated as a function of impairment duration. In order to cover a broad range of possible situations, we have selected values from durations close to the detection thresholds of up to 20% information loss: 5, 10, 30, 150, 550, 1000, 2000ms. An intermediate anchor was added, a signal loss of 30ms. The greatest duration was used as a low quality anchor.

3.3. Auditors and method for quality test

The quality ratings from 20 subjects were gathered using the MUSHRA (Multi Stimulus test with Hidden Reference and Anchors) method [6]. A numerical scale (0-100) for rating overall quality is used. This scale is related to five quality categories (bad, poor, fair, good, excellent) that are uniformly distributed. In this method the assessor is able to switch between the N signals of the trial (explicit reference, hidden reference, hidden anchor and conditions under test). He is also able to change the current scores as many times as he wants. This flexibility of access leads to an implicit comparison process increasing stability ratings.

For each trial the listener is asked to assess signals using a software slider placed over the selection button. These selection buttons are randomly related to test sequences in order to avoid ordering effect. Furthermore, trials are randomly ordered for each subject. Individual adjustment of listening level was allowed to each participant at the beginning of the test. Further level adjustment was not allowed.

Subjects assessed two tests (one session per test). Each test consists of six audio contents, ten signals by content (1 known reference, 1 intermediate anchor, 1 hidden reference and 7 condition test). The signal presentation time is 10s. The total duration of each test is approximately 20min. Assessors were invited to take a 10 min break.

3.4. Results of quality

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 participants. Mean opinion scores for the low (test 1) and high (test 2) activity context are plotted in Fig. 5.

With regards to MOS over all contents (Fig. 5 and Fig. 6), we can see that a discontinuity of 5ms has a negligible effect on quality. The perceived quality is still considered as excellent. This result is consistent with the probality of detection of such discontinuity. Duration of 5ms is in the range of the detection thresholds (1.2 to 6.11 ms).When the discontinuity strength is 10ms, greater than the detection thresholds, quality degradation becomes significant. Quality exhibits a reduction of 8 and 14 points (over the 0-100 scale) for test 1 and test 2 respectively.

Subjects had a significant negative reaction to a discontinuity of 30ms. Quality function exhibits a fall in rating of 1 ½ MOS category (Fig. 6). This result is related to the probability of detection found in the audibility experiment. Most of the subjects were able to detect the audio loss of this strength for all contents. After 150ms (Fig. 5), quality function shows a slower decreasing rate. Ratings attain the bad category at 2000ms of information loss.

Comparing both tests, we can see that the results of the high annoyance test show a slightly lower quality than test-2. The differences are greater for loss durations of 5 and 10ms. It is interesting to note that these differences are supported by the respecting probabilities of detection (Fig. 4).

Table 3. Mean Opinion Scores over all conditions. Low and high annoyance context (test1 et 2). Contents are listed in increasing order.

	Test1		Test2
Content	**MOS**	**Content**	**MOS**
Instrument	53.86	Instrument	53.28
Song & Intrum	55.24	Song	53.73
Song	55.42	Song & Intrum	54.48
Mixture	59.75	Mixture	57.21
Sport comment	60.49	Sport comment	58.32
Newsreader	61.98	Newsreader	60.73

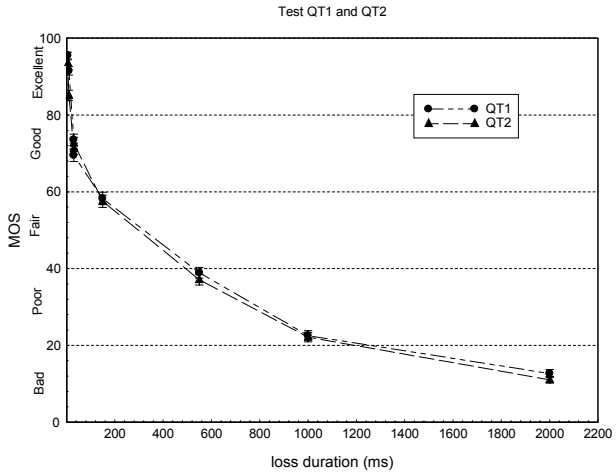

Fig. 5. Mean opinions scores as a function of audio loss duration. Results from low and high activity context (tests 1 and 2) calculated over all contents.

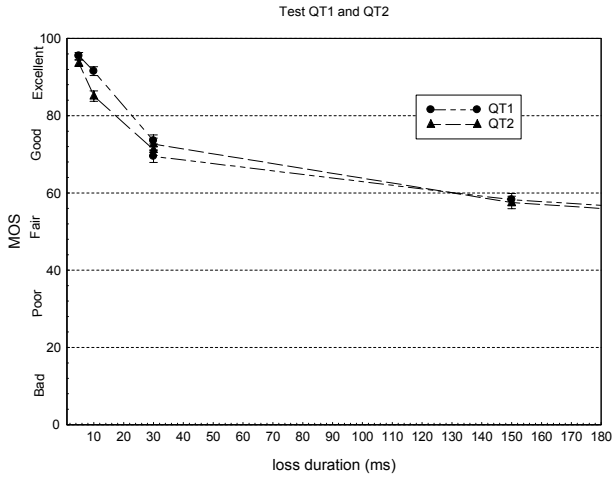

Fig. 6. Detailed results of test-1 and test-2. Range duration up to 150ms.

Now we will look at the assessment results by content from test 1 and 2 (Fig. 7). A solid line joins the means ratings over all contents. For both tests, the content less affected by audio loss was the Newsreader sequence (speech alone) (Table 3). The inserted temporal discontinuities were more annoying for the Instrument content (guitar).

Looking at Table 3 it seems that audio loss has a greater impact on music (guitar, jazz, chorus) than on speech signals (commentary either alone or mixed with

crowd noise). This finding is valid for both tests even if the semantic contexts were significantly different. However, this result has to be confirmed by other future experiments.

When a loss duration is greater than or equal to 550ms, the inter content variability is reduced for both tests. It seems therefore that annoyance is content independent for such discontinuities.

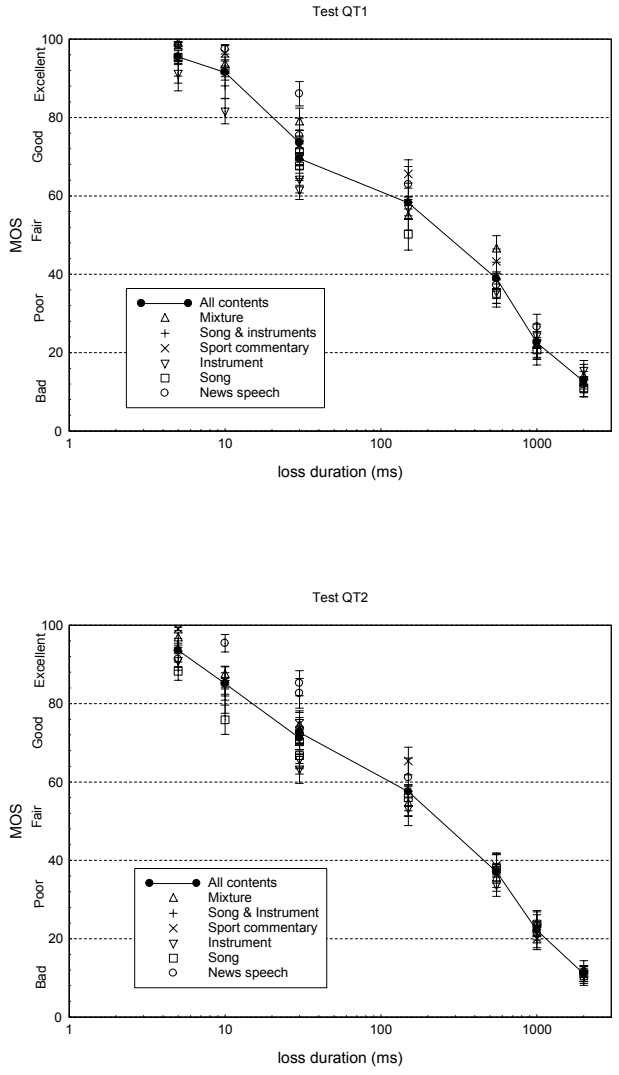

Fig. 7. Quality as a function of loss duration for each content. From top to bottom: low annoyance (Test 1) and high annoyance (Test 2) context. Solid line joins the overall quality results.

4. CONCLUSION

Audibility: thresholds for detection were measured over a range of discontinuity durations and temporal positions. We have found that the audibility of a signal loss is duration, activity context and content dependent. Detection thresholds vary from 1 to 6 ms. A discontinuity duration of 30ms is completely audible for all tested contents.

Quality impact: the assessment results showed how perceived quality could be impaired by an isolated signal loss in a 10s audio sequence. Subjects show a strong negative reaction to an isolated audio loss. The impact of such temporal discontinuities is duration, content and activity context dependent

The quality function estimated in relation to a burst of lost samples of different durations is presented. The analysis of the results is useful for audio quality metrics design and for a better understanding of time-varying quality.

REFERENCES

[1] C. Perkins, *RTP: Audio and Video for the Internet*, Addison Wesley, USA, 2003, chap. 2, pp. 15-48.

[2] M. Kalman, E. Steinbach, B. Girod, "R-D optimized media streaming enhanced with adaptive media playout," *ICME*, Vol. 1, IEEE, Lausanne, pp. 869 - 872, August 2002.

[3] B. Farell, and D. Pelli, *Psychophysical methods, or how to measure a threshold and why*. In J. G. Robson and R. H. S. Carpenter (Eds.), *A Practical Guide to Vision Research*, Oxford University Press, New York, 1998.

[4] R.R. Pastrana-Vidal, J.C. Gicquel, C. Colomes, and H. Cherifi, "Sporadic Frame Dropping Impact on Quality Perception," *Human Vision and Electronic Imaging IX*, SPIE, vol. 5292, San José CA, January 2004.

[5] ITU-R Recommandation BS.1116-1, *Methods for the subjective assessment of small impairments in audio systems including multichannel sound systems*, 1997.

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