Study of the relationship between subjective conversational quality, and talking, listening and interaction qualities: towards an objective model of the conversational quality

Marie Guéguin[1,2], Valérie Gautier-Turbin[1], Laëtitia Gros[1], Vincent Barriac[1], Régine Le Bouquin-Jeannès[2], and Gérard Faucon[2]

[1]France Télécom R&D, TECH/QVP, Lannion, France
[2]Laboratoire Traitement du Signal et de l'Image, INSERM U642, Université de Rennes 1, France

{marie.gueguin, valerie.gautierturbin, laetitia.gros, vincent.barriac}@francetelecom.com,
{regine.le-bouquin-jeannes, gerard.faucon}@univ-rennes1.fr

Abstract

This paper relates a subjective test designed to study the relationship between conversational quality and talking, listening and interaction qualities when facing delay and echo. The results show that subjects were not as much disturbed by echo-free delay, and that the relationship between listening, talking and conversational qualities depends on several factors, like the type of the considered impairments, for instance. A relationship based on linear combination has been determined. A comparison has been made between the estimated conversational quality scores and the subjective conversational quality scores, which showed these are very close.

Keywords

Conversational quality, objective model, subjective assessment

1 Introduction

Subjective tests are the only way to assess perceived speech quality in telecommunications, they are although complex, cost- and time-consuming. Objective measures have been introduced to model the results of these tests and to predict users' perception of speech quality in a given context. Among them, the ITU-T PESQ [1] models speech quality in listening context (mainly impacted by speech distortion, background noise and packet loss) and PESQM, proposed by Appel and Beerends in [2], models speech quality in talking context (mainly impacted by echo and sidetone distortion). However there is, to our knowledge, no objective model of the conversation context, in which two persons, whether it is face to face or on telephone, alternate talking and listening roles [3]. The ease of this alternation depends on the ease of interaction between the two persons. Considering this relationship, one wonders whether, on a quality assessment point of view, this relationship also exists. The question is then to know whether and how the conversational quality can be decomposed in different components corresponding to these different roles, as it is suggested by several sources like [4]:

- listening quality,
- talking quality,
- interaction quality (mainly impacted by end-to-end delay and distortion due to double-talk situation).

Our objective here is then to study the relationship between conversational quality and talking, listening and interaction qualities on a subjective point of view by using the results of a new subjective test specially designed for this issue, and to propose a model of the conversational quality. This is a preliminary study concerning two impairments.

2 Method

Our approach, given in Figure 1, consists in combining two scores: the talking quality score and the listening quality score given by subjects during a subjective test. It also takes into account the main impairment impacting the interaction quality, *i.e.* the delay, by using the knowledge on the impact of the delay on users' judgment, assessed during subjective tests. That combination of three components (talking quality, listening quality and delay) is not an obvious and simple juxtaposition. Indeed, the conversational quality is more or less influenced by one of the three components,

depending on the impairments affecting the communication. When only a listening impairment is present, for example packet losses, the conversational quality score will be essentially correlated with the listening quality score.

By introducing a decision system, our approach takes into account the influence of the type of impairment on that combination. The decision system weights the influence of the three components on the conversational quality score, depending on subjective tests results.

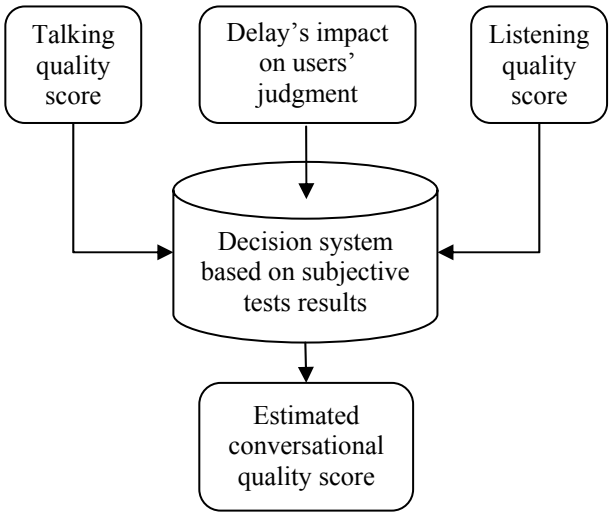

Figure 1. Estimation of the conversational quality score

Consequently, subjective tests are necessary to determine, depending on the impairments, the relationship that links conversational quality score to talking quality score, listening quality score and delay value.

3 Subjective test

A subjective test, described in detail in [5], has been conducted to study the relationship between the conversational quality score and the three components (talking quality score, listening quality score and delay value). It is a conversation-opinion test using a new test methodology, since no methodology already exists to assess this relationship.

This new methodology determined the listening, talking and conversational qualities on both sides of a vocal link. According to ITU-T Recommendation P.800 [6], the conversation-opinion test involved couples of non-expert subjects (A and B). For each tested condition, the test was split in three phases. During the first phase, subject A reads a text and subject B listens, to assess talking quality on side A and listening quality on side B. During the second phase, roles are inverted. During the third phase, subjects have a free conversation (using the conversation scenarios developed in [7]), to assess conversational quality on both sides. At the end of each phase, both subjects were asked to assess the overall quality on the recommended five-point MOS scale [6] (5 = Excellent, 4 = Good, 3 = Fair, 2 = Poor, 1 = Bad).

The test conducted here with that new methodology examined the quality in presence of delay and electric echo, using eight test conditions (given in Table 1) and fifteen couples of non-expert subjects. These two impairments were chosen because they have, if considered separately and / or combined, an impact on each of the three qualities: listening (echo during double talk), talking (echo + delay) and interaction (delay).

Condition	One-way delay (ms)	Echo level attenuation (dB)
1	0	No echo
2	0	25
3	200	No echo
4	200	25
5	400	No echo
6	400	25
7	600	No echo
8	600	25

Table 1. Test conditions

4 Results

Figure 2 shows mean opinion scores obtained, according to the situation (Talking, Listening and Conversation), the presence of echo and the one-way delay. The curves have been offset horizontally for clarity.

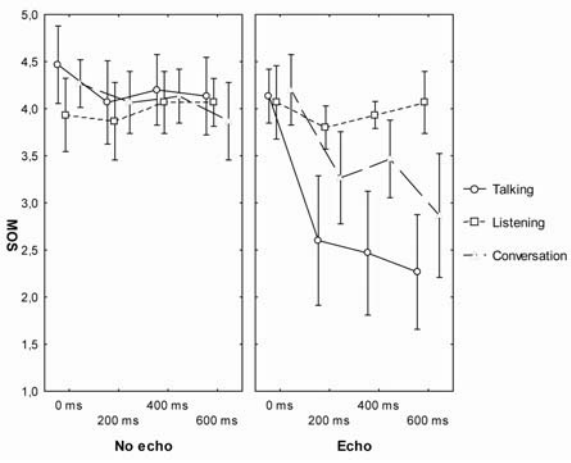

Figure 2. Mean opinion scores of the subjects

In the case with echo-free delay (Figure 2, left), subjects' judgment is almost constant, whatever the delay or the situation, which is confirmed by a HSD Tukey test. These results show that, for values between 0 and 600 ms, one-way echo-free delay has little impact on subjects' judgment, despite the interactivity of the conversation situation. So, echo-free delay with values between 0 and 600 ms does not need to be considered in our model.

In the case with echo and delay (Figure 2, right), subjects' judgment depends on the situation. This indicates that there is a difference between the talking situation and the conversation situation, subjects being more disturbed by echo in the talking situation than in an interactive situation. It can be explained by the fact that in talking situation subjects are more attentive to the quality and to its judgment, whereas in the situation of conversation (interactive) the attention of the subjects is shared between the task of conversation and the task of judgment of the quality.

As a first approach, we choose to estimate the conversational quality score with a linear combination of the talking quality score and the listening quality score:

$$\widehat{\mathrm{MOS}}_{\mathrm{conversation}} = \alpha \times \mathrm{MOS}_{\mathrm{talking}} + \beta \times \mathrm{MOS}_{\mathrm{listening}} . \quad (1)$$

The two parameters α and β have been calculated to maximize the correlation between the subjective conversational scores and the estimated conversational scores. The correlation is calculated by:

$$r = \frac{\sum\left[\left(x_i - \bar{x}\right)\cdot\left(y_i - \bar{y}\right)\right]}{\sqrt{\sum\left(x_i - \bar{x}\right)^2 \cdot \sum\left(y_i - \bar{y}\right)^2}} \quad (2)$$

where x_i is the subjective conversational score for condition i, and $\bar{x}$ is the mean of the x_i. y_i is the estimated conversational score for condition i, and $\bar{y}$ is the mean of the y_i.

The optimal values of α and β, and the corresponding optimal correlation r are given in Table 2 for the two cases studied in this test. It is to note, however, that here the training and validation databases are the same, because these are the only data we have.

The correlation value indicates that, for the impairments studied in this test, the conversational subjective score is well estimated with a linear combination of talking quality scores and listening quality scores.

It is then necessary to apply a mapping function to minimize the mean squared error between the estimated conversational scores and the subjective conversational scores. This mapping function has to be monotonic to preserve the ordering. We decide to use a linear mapping function:

$$y = a \cdot x + b . \quad (3)$$

	Echo-free delay	**Echo + delay**
α	9.6	8.8
β	-8.75	-9.5
r	0.837	0.948

Table 2. Optimal parameters α, β and correlation *r* for the two cases

Indeed we have few data, so a classic 3rd-order polynomial mapping function would not be appropriate (4 points for 4 freedom degrees).

The optimal parameters a and b, and the corresponding mapping coefficient R^2 are given in Table 3 for the two cases studied in this test.

Table 3 shows that the mapping almost corresponds to an offset of about 4.

	Echo-free delay	**Echo + delay**
a	0.072	0.074
b	3.68	4.37
R²	0.7	0.9

Table 3. Optimal values of the parameters *a*, *b* and *R²* of the linear mapping function for the two cases

Table 4 shows the correlation r (*cf.* equation (2)), the mean squared error (*MSE*) and the mean absolute residual error (*MARE*) between the estimated conversational score and the subjective conversational score, after application of the linear mapping function.

	Echo-free delay	**Echo + delay**
r	0.837	0.948
MSE	0.006	0.024
MARE (in MOS)	0.064	0.127

Table 4. Final performances after application of the linear mapping function

The results presented in Table 4 show that, after applying the linear mapping function, the conversational quality score is very well estimated from talking and listening quality scores. The final mean absolute error (*MARE*) between estimated and subjective scores is very low (<0.15 MOS) in both cases.

Figure 3 (up) shows the final mapping between estimated and subjective conversational scores, in each case. Figure 3 (down) shows the estimated and subjective conversational scores, in each case.

5 Conclusion

In this study we show that the conversational quality score could be estimated with a linear combination of talking quality score and listening quality score, for two specific impairments, and that the echo-free delay, for one-way values from 0 to 600 ms, has only little impact on subjects' judgment. One can imagine using the parameters determined in this study to estimate the conversational quality of a voice link presenting the same impairments (type and level) than those studied here. One can also imagine using the same approach (test methodology and model) to extend these results to other impairments.

Given the lack of data, these conclusions need, however, to be confirmed by further studies on the same impairments first and on other ones then.

Finally, one can envisage to use the model presented here, estimating the conversational quality score from the talking and listening quality scores on a subjective level, to estimate the conversational quality on an objective level from objective models of the talking and listening qualities.

6 References

[1] ITU-T Recommendation P.862, "Perceptual evaluation of speech quality (PESQ), an objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs", February 2001.

[2] R. Appel and J.G. Beerends, "On the quality of hearing one's own voice", *J Audio Eng Soc*, 50, 237-248, April 2002.

[3] D.L. Richards, Telecommunication by Speech: The Transmission Performance of Telephone Networks, Butterworths, London, 1973.

[4] ITU-T COM 12-55, "A subjective/objective test protocol for determining the conversational quality of a voice link", July 2003.

[5] ITU-T COM 12-D.45, "Report on a new subjective test on the relationships between listening, talking and conversational qualities when facing delay and echo", January 2005.

[6] ITU-T Recommendation P.800, "Methods for subjective determination of transmission quality", August 1996.

[7] ITU-T COM 12-35, "Development of scenarios for a short conversation test", December 1997.

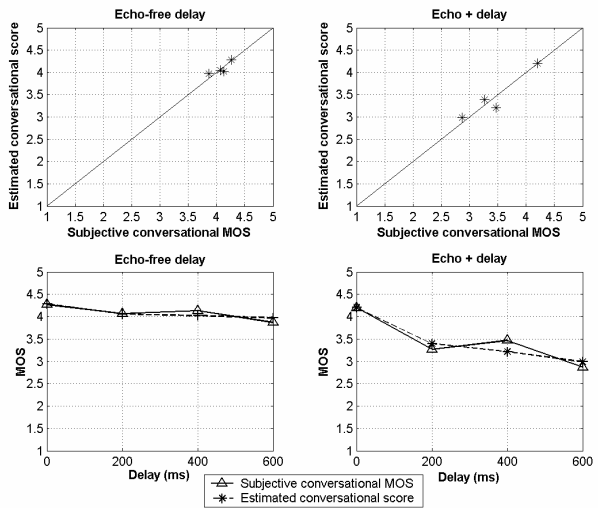

Figure 3. Up: final mapping between estimated and subjective conversational scores. Down: estimated and subjective conversational scores