ARCHITECTURE OF NON-INTRUSIVE PERCEIVED VOICE QUALITY ASSESSMENT

Elodie Bernex, Human Interface Laboratory of France Telecom R&D,

Vincent Barriac, Human Interface Laboratory of France Telecom R&D,
elodie.bernex@rd.francetelecom.com **vincent.barriac@francetelecom.com**

Keywords: perceived voice quality, non-intrusive measurement, VoIP.

Summary: This contribution deals with the measurement of vocal quality impairments perceived by users of IP telephony. The approach described in the following pages is non-intrusive measurement method (INMD like) and is based on the subjective analysis of VoIP impairments. A subjective study, employing Free Categorization [1], Magnitude Estimation [3] and statistical analyses allows the building of a library of typical impairments used as basis for detection of corruption and assessment of quality. This paper describes a possible architecture of non-intrusive voice quality assessment in the VoIP context.

NATURE OF THE PROBLEM

The types of voice quality impairments perceived using IP-telephony are different from the ones generally perceived using PSTN telephony and so the methods to assess voice quality when using PSTN are not relevant.

STATE OF THE ART

In order to assess voice transmission quality, today proposed methods, as described in [2], some are intrusive and use a comparison between a reference and a distorted signals, as it is done by a model like PESQ [6], the others compute an impairment factor from the measurements of physical parameters as for example delay, packet loss or jitter, to be used with a transmission planning model like the E model [7]. The first kind of measurement is well suited for assessment with access on both terminations of a link, but it is difficult to implement when competitors' networks are interconnected, and this is why complementary, non-intrusive measurements are needed. The second kind of measurement is a first answer to this lack, and can be very easily implemented, but it only takes into account physical parameters quite impossible to rely to voice quality perception (since the calculation of the impairment factors is not based on subjective results, as it was for the first parameters of the E model, but on a model of packet loss).

New and complementary methods for non-intrusive assessing voice quality in a more subjective way are needed.

DETAILED DESCRIPTION OF THE ASSESSMENT METHOD

The proposed approach, described in figure 1, is non-intrusive (INMD like) and is based on the subjective analysis of VoIP impairments. The goals of this non-intrusive method are: to generate alarms when the voice quality of a call decreases, and to assess the overall quality of a call with the determination of a Mean Opinion Score (M.O.S.).

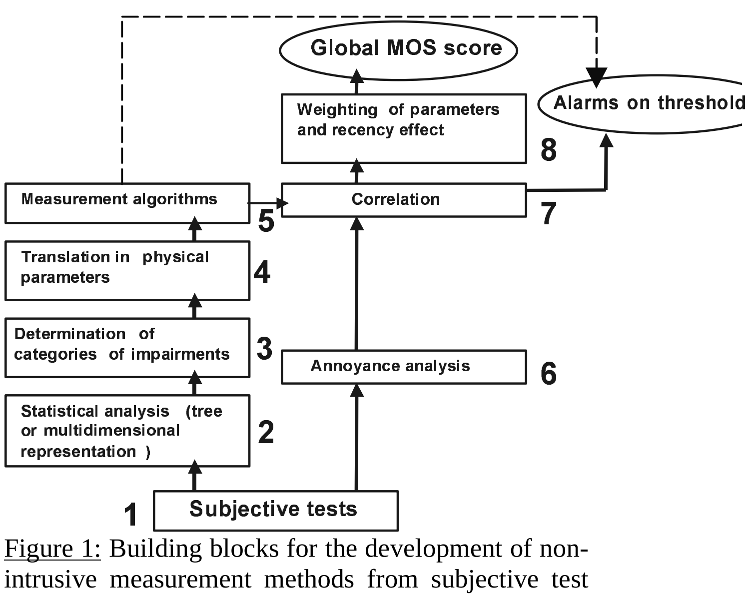

Figure 1: Building blocks for the development of non-intrusive measurement methods from subjective test results.

Library of Typical Perceived Impairments

The steps 1,2 and 3 of Figure 1 correspond to the elaboration of a library of typical impairments perceived by users of IP telephony [9].

Free categorization tests

In order to assess the speech quality perceived by the listeners, a study was designed to find all the different types of perceived voice quality impairments when calling on IP-telephony. To define and describe them subjectively, the method of "Free Categorization " was implemented on a corpus composed by speech sequences corrupted by a loss of several consecutive packets. [9]. The people taking part in this experiment were also asked to put into word each type of corruption and to quantify the annoyance they cause.

Additive tree and multidimensional scaling

The data collected during the tests were converted into dissimilarity matrices, analysed with two different methods: tree analysis (Sattah et Tversky algorithm [10]), and multidimensional scaling analysis (Guttman-Lingoes algorithm [11]). Both analysis methods give similar results (the same categories were found), but they are also complementary: additive trees show more precise groups, when multidimensional scaling representations show the dimensions of the grouping[9].

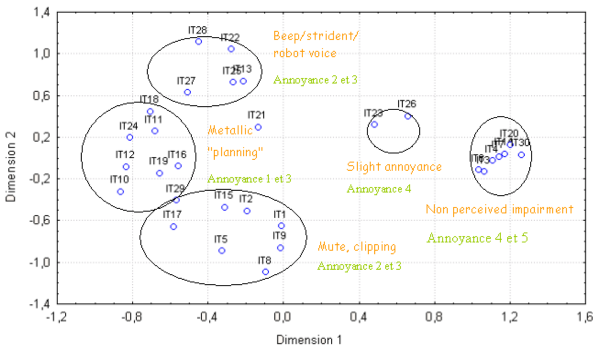

Figure 3: example of multidimensional representation for one test in dim1*dim2 plane

We find dimension 1 as being: “perceived impairment” till no “perceived impairment”(cf. figure 3), dimension 2 as being: “not metallic” (Clipping, mute) till “very metallic” (beep, high-pitching strident) (cf. figure 4), and dimension 3: “slight reducing” till “high reducing” (mute, clipping) (cf. figure 5).

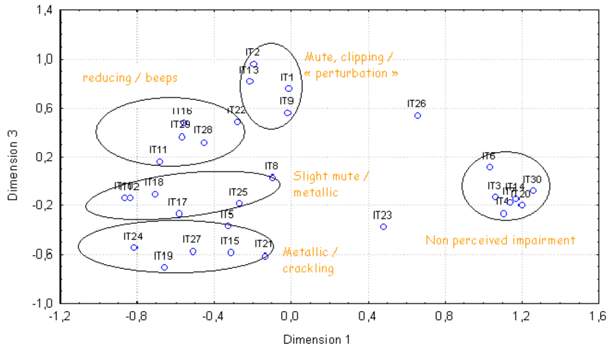

Figure 4: example of multidimensional representation for one test in dim1*dim3 plane

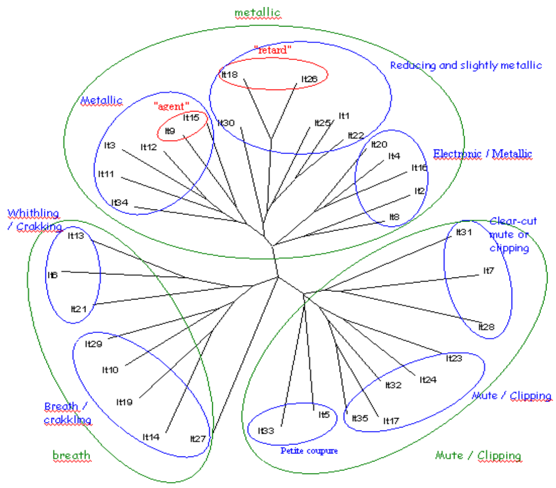

Figure 5: example of tree representation for one test

The results of this study show the existence of three main impairment categories of perceived impairments, with more or less numerous subcategories. Subjects describe them as: « Whistling, breath, crackling », « Mute » (clipping, slight mute, reducing, drop in energy…), « Metallic » (metallic, robot voice, electronic voice, beep…).

Additionally, these results show that the impairments located outside speech activity periods are not perceived when there is no background noise and when packet loss concealment mechanisms with replacement by interpolation and silence are used.

When we carry out the same study but with background noise, the results depend on the SNR: with high SNR we find approximately the same categories than without background noise; with low SNR, the results (figure 6) depend on the kind of background noise, and arranging depends on the noise and the disturbance induced: when it is not disturbing, classification depend on the noise itself (street, PC, TGV); when it is disturbing, classification depend on category of perceived impairment. We find then: “repetitive cutting”, “continuous breath”.

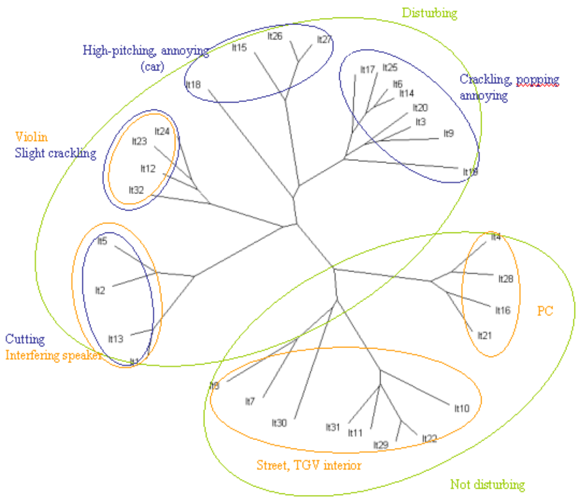

figure 6: example of tree representation for test with background noise and low SNR

Perceived impairments library

Four main categories of sporadic impairments, divided in several subcategories, and two of continuous impairments have been found. Subjects describe these categories as: « Non perceived impairment », « Clipping, attenuation », « Metallic noise, robot voice, beep », « Whistling, breath, crackling » [3] and “continuous mute”, “continuous breath”. These describe the library of impairments used for the assessment of perceived voice quality.

Physical Characterization

Step 4 corresponds to the physical characterisation of subjective impairments categories determined in steps 1, 2 and 3.

We use for the physical characterization, very simple describing voice parameters like energy, pitch, MFCC, spectrum. We correlate them with the subjective categories of impairments with the help of several statistic methods to find which physical parameters are able to explain the different impairment categories of the library. Each impairment category is characterised by its own physical indicators to ensure a reliable detection in the speech signal.

Detection of the Library’s Impairments

Step 5 corresponds to the detection of the library’s impairments in the speech signal by the means of different signal processing algorithms.

The signal is extracted through a physical interface (E1/T1, Ethernet, etc.). It is then analysed by a means of different signal processing algorithms. These algorithms measure the magnitude of signal physical parameters (levels, delay, bandwidth, pitch, etc.) and the variation of these amplitudes, and, using decision criteria, detect the degradations of the library elaborated in step 3. For example, for the detection of "mute", the parameters associated to this impairment category are: the duration of null signal and the variation of the energy of the speech signal at both ends of the lost packet sequence. We use those very simple parameters and several thresholds on them to ensure the detection of only perceived mutes.

The parameters recommended in ITU-T P.561 Recommendation [5] (i.e. active speech level, background noise level, echo path loss and delay) can also be evaluated at this step of the process.

Impairment Rating

The rating of the voice quality impairment degree on a subjective scale is now necessary (step 6).

We evaluate the degree of each kind of impairments by the mean of Magnitude Estimation test. We used as corpus of these test the whole corpus of the Free Categorisation test, which have already given us all the preceding results.

Correlation Step

All the impairments of the library are now characterised by physical and subjective indicators. A correlation step (block 7 on Figure 1) is then necessary to use objective measurement results to model the users’ perception of voice quality.

When the measurement algorithms detect a corruption, the corresponding level of impairment impact is estimated thanks to the correlation function.

Alarms

Then it is possible to trigger off alarms in relation with the exceeding of a threshold (or a combination of several thresholds). Alarms can also be trigger off after the step 5, but without information on the impairment degree in this case. It is also useful to know which type of perceived degradation has been detected on the studied call, and must be taken into account in the overall quality rating.

Mean Opinion Score

The last step determines a Mean Opinion Score for each transmission direction by processing the impairment degrees, the frequencies and time locations of all the impairments encountered in a call, by level functions [4].

CONCLUSION

It is very useful to obtain as output of a quality estimation process a combination of both overall subjective measurement and sporadic objective indicators: the M.O.S. score (calculated by using a psycho-acoustical model like the P.862 [6] standard, but with no reference signal) is directly linked to the user's perception of the quality, and so gives a direct overall indication of the quality of delivered service. If computational models like the E-model (ITU-T G.107 [7]) or the Call Clarity Index (ITU-T P.562 [8]) are used rather than psycho-acoustical ones, parameters like the P.561 ones will be accurately taken into account, but not necessarily the others like the ones taken in the library elaborated in steps 1 to 3; the alarms on thresholds are needed to undertake actions improving the quality, based on a detailed knowledge of the cause of the impairments. Having only a single quality score, without this information, has no interest for practical network supervision purposes.

The possible architecture proposed in this paper allows the both measurement ways and moreover it should provides more accurate subjective results in a non-intrusive way than methods taking into account only network parameters to compute impairment factor.

REFERENCES

[1] Maffiolo, V. "De la caractérisation sémantique et acoustique de la qualité sonore de l'environnement urbain". *Thèse de l'Université du Mans soutenue le 6 octobre 1999.*

[2] "Quality of Service (QoS) measurement methodologies", ETSI Telecommunications and Internet Protocol Harmonization Over Networks (TIPHON) Technical Standard 101-329 Part 5 (2000)

[4] L. Gros (may, 2000). "Study of the relationship between instantaneous and overall subjective speech quality for time varying quality speech sequences: influence of recency effect". UIT-T Com12-D139 (fev2001).

[5]"Recommendation In-service Non-intrusive Measurement Device - Voice Service Measurements" UIT-T P561

[6]" Recommendation. Perceptual evaluation of speech quality (PESQ), an objective method for end-to-end speech quality assessment of narrowband telephone networks and speech codecs", UIT-T P862

[7]" Recommendation.The E-Model, a computational model for use in transmission planning ", UIT-T G.107

[8]"Recommendation Analysis and interpretation of INMD voice-service measurements" UIT-T P562

[9] Bernex, E and Barriac, V (fev 2001). UIT-T COM12 023_ww9 Non-intrusive assessment of perceived voice quality impairements

[10] Sattah, S. and Tversky, A. (1977). "Additive similarity tree". Psychometrika – vol.42, n°3, 319-345.

[11] Lingoes, J. -C. (1973). The Guttman-Lingoes nonmetric program series", Ann Arbor. Mathesis.