12th IMEKO TC4 International Symposium
Electrical Measurements and Instrumentation
September 25–27, 2002, Zagreb, Croatia

AN IMPROVED PROCEDURE FOR QoS MEASUREMENT IN TELECOMMUNICATION SYSTEMS

Domenico Grimaldi

Dip. di Elettronica, Inf. e Sistemistica, Università della Calabria, 87036 Rende (CS), Italy,
Ph.: ++39 0984 494712, Fax: ++39 0984 494713, E-mail: grimaldi@deis.unical.it.

Abstract – The paper deals with (i) the characteristics of an improved signal-processing procedure for both voiced-unvoiced classification and echo parameter evaluation, and (ii) the reference signal selection to be adopted for clarity measurement. Improvements, respect to the traditional procedures, are achieved (i) in classifying the voice by means of the Learning Vector Quantization neural network, (ii) in echo parameter measurement by means of the recursive evaluation of the FIR filter coefficients estimating the impulse response of the echo path, and (iii) in the clarity measurement by using the optimized multi-sine signal.
Numerical tests by using different types of telecommunication network noise show the efficacy of the proposed approach. The test results are also compared with the results obtained by different techniques presented in literature.

Keywords: Quality of Service, Measurement in Telecommunication Systems, Signal Processing.

1. INTRODUCTION

The implementation of advanced technologies in the new generation of telecommunication systems demands new generation of measurement tools for Quality of Service (QoS) evaluation [1, 2]. In the new scenario of the telecommunication systems the QoS is strongly affected by: (i) the coexistence of wire line and wireless technologies, (ii) the connection of the traditional public telephone network to both cellular and packet networks, (iii) the conversion of the signal from analog-to-digital-to-analog, (iv) the division of the signal into multiple packets, (v) the buffering and routing to traverse multiple networks, and (vi) the use of multiple compression algorithms.

The QoS emerges as one of the most strategic area in the ultra-competitive communication market and the Voice Quality (VQ) is one of the most important components. The correct VQ measurement procedure, according to the ITU-T Recommendation P.561 [3], requires: (i) to select only the frames of voice signal into the acquired one, and (ii) to use these frames to measure all the factors affecting the VQ such as the echo, the delay, and the clarity of the voiced signal.

The echo is produced by the signal crossing from one type of network to another. At the listener the echo is the sum of the talker echo and the listener echo and it results an annoying reflection of the speaker's voice.

The delay is the amount of time a voice signal takes to travel from one caller to another in a telephone conversation. This includes the time required for signal processing. Variations in delay is referred as jitter.

The clarity is the fidelity of the signal itself. A factor that may impair is the use of multiple compression algorithms occurring when a call that origins on one type of network and terminates on another one.

In the paper, the problems to implement the measurement procedure are considered, and in particular, those concerning: (i) the selection of features which are both simple to compute and effective for the differentiation between the two classes of voiced and unvoiced signals, (ii) the accurate evaluation of the echo path parameters in-service voice connections, (iii) the evaluation of the clarity, and (iv) the reduction of the processing time for the real-time operation in the In-service Non-intrusive Measurement Devices (INMD) [4].

First of all, the classification procedure of the acquired signal as voiced and unvoiced is made both speed and accurate by using the Learning Vector Quantisation (LVQ) Artificial Neural Network (ANN) [5, 6]. This ANN is conveniently used according to the following considerations: (i) the pattern recognition problems can be adequately solved by means of the ANN, (ii) the ANN's capability of learning by means of the training set overcomes the difficult selection of both the features and the effective parameters for the classification, and (iii) the parallel architecture of the ANN is fundamental for the real-time operation of the INMD.

Successively, the accurate evaluation of the echo parameters is achieved by means of the recursive evaluation of the coefficients of the FIR filter modeling the impulse response of the echo path [7]. This recursive procedure allows the reduction of the number of voice signal samples and of the processing time without degreasing the accuracy of the echo parameter measurement.

The voice clarity basically is a subjective measurement because it depends on individual perceptions. The ITU-T Recommendation P.861 standardizes the Perceptual Speech Quality Measurement (PSQM) algorithm for objective testing within the voice bandwidth of 300-3400 Hz [8]. The PSQM provides a relative score corresponding to how a statistically large number of human listeners would react. In order to overcome the difficulties arising from the need to select numerous and statistical correct speeches and, moreover, to permit to test with a reference signals with assigned characteristics, the optimize multi-sine signal is proposed. This signal is easily to generate and to reproduce because it is the sum

of equal amplitude sinusoidal waves at equal spaced frequencies and different phases. Moreover, it is easily to adapted in the PSQM algorithm to substitute the voice signal. In this manner the inconvenient of several approaches presented in the literature are overcome. In particular, (i) the problems of selecting the features both simple to compute and effective for the differentiation between the two classes voiced and unvoiced [9]-[12] are solved, (ii) the difficulties of the pattern recognition in presence of either a high-level noise or an impulsive noise [13] are avoided, (iii) the requests of high accuracy and rapidity of both classification of the signal [14] and echo parameter measurement [7] are addressed, and (iv) both the experimental complexity of relatively large number of human speeches and the referable voiced signal for clarity quantify are overcome.

Tests according to the international recommendations, and using experimental voice signals were performed. Some results showing (i) the improvements achieved respect to the traditional ones, and (ii) the performance of the proposed measurement procedure versus the S/N ratio in the case of different noise types are presented.

2. VOICE QUALITY MEASUREMENT PROCEDURE

The proposed procedure for VQ parameter measurement operates in successive steps according to the general scheme of Fig. 1. The voice signal is acquired and classified as belonging to voiced and unvoiced frames. Successively, the voiced frames are processed in order to evaluate both the echo and the delay parameters. The clarity measurement is executed in independent way by using the multi sine as the reference input signal.

In particular, the multi sine signal parameters as (i) the common amplitude, (ii) the equal spaced frequencies and (iii) the corresponding phases are set on the base of the suitable equivalence criteria. This equivalence must ensure that the PSQM algorithm furnishes the same relative score in the case the input is either the voice or the multi sine signal.

According to the operating mode of the measurement procedure, the following considerations hold:

1 the voiced-unvoiced classification block influences the successive ones and, consequently it requires a very robust and speed algorithm;
2 the number of samples of the voiced signal must be set according to the accuracy and speed desired from the user blocks;
3 the echo and delay procedures operate in parallel way and, consequently, in the same time interval;
4 the multi sine signal is sent in the telecommunication network separately and successively to the speech signal.

In the following the procedures for voiced unvoiced classification and echo parameter measurement are separately shown and tested. Moreover, the numerical results validating the use of the multi sine for clarity measurement and the corresponding values of the parameters of the multi sine signal are also given.

3. VOICED-UNVOICED CLASSIFICATION

The block scheme of the procedure for voiced-unvoiced is shown in Fig. 2. The voice signal, sampled at a frequency rate of 8 kHz, is sent to the pre-processing block where it is subdivided

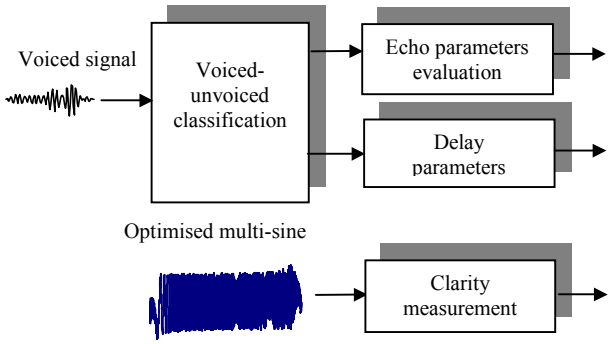

Fig. 1. Block scheme of the proposed procedure for VQ measurement.

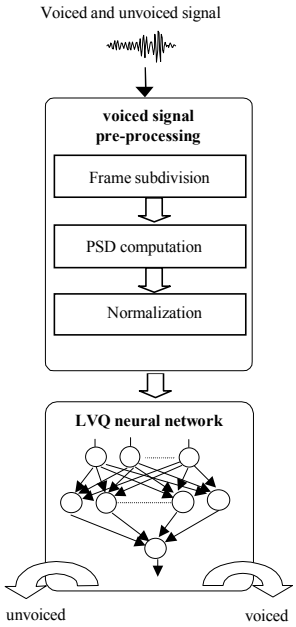

Fig. 2. Block scheme for voiced-unvoiced classification.

in successive and separate frames. Each one is constituted by 3072 samples. In this manner the length of each frame is greater than the length of the inter-syllabus pause. This last length, according to [3], is approximately up to 350 ms, corresponding to 2800 samples.

The Power Spectral Density (PSD) of the 3072-sample frame is estimated as the mean value of the PSD of 12 sub-frames, each one of 256 samples, corresponding to the interval time of the voice signal of length equal to 32ms. This interval ensures a quasi-stationary condition for the voice signal [10].

According to the Discrete Fourier Transform (DFT) properties, the resulting PSD is univocally identified by 128 samples. Before sending to the ANN, the 128 samples constituting the PSD are normalized. Fig. 3 shows the voice signal and the corresponding normalized PSD. The PSD shape associated to voiced is different from that associated to unvoiced and can be classified by means of the ANN. The ANN classifier is based on the LVQ neural network.

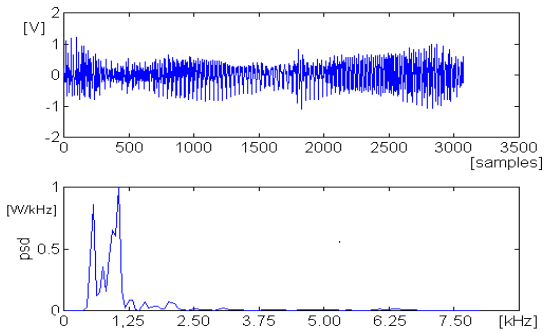

Fig. 3. Voiced signal (upper) and the corresponding PSD (lower).

3.1 LVQ neural network for voiced-unvoiced classification

Given N classes of d-dimensional vectors and a vector $x \in \Re^d$, the LVQ neural network classifies x by individualizing the class to which it belongs. For this purpose, the ANN needs a training phase, in which a reference vector y, called the codebook, is determined for each class. Next, the LVQ network classifies a vector individualizing the codebook, which best matches it.

The used matching function is the Euclidean distance:

$$d_e(x,y) = \sum_{j=1}^{n} (x_j - y_j)^2 . \quad (1)$$

This solution shows the better performance if compared with the normalized dot product [15]. In the hidden layer the codebooks are compared with the input vector and the matching function is computed according to (1). In the output layer, the winning codebook is computed as the codebook with the minimum d_e and the voiced-unvoiced classification is determined. The ANN is trained by using the LVQ2 algorithm [5]. This works as follows: let x be a vector of the training set and m_w the winning codebook, this is updated according to:

$$m_w(t+1) = m_w(t) \pm \alpha(t)[x - m_w(t)]. \quad (2)$$

The plus sign is used if x and m_w belong to the same class, the minus sign otherwise. This process is repeated for several iterations for each vector of the training set. The $\alpha(t)$ training coefficient (ranging from 0 to 1) is empirically determined. A larger value is usually adopted during the first iterations, while a smaller one is preferred for the last ones.

3.2 Training phase

The voice signals constituting the training set were organized according to the indications of the Annex A to the P.561 Recommendation for testing the INMD [3].

In particular, 6 different conversations were recorded in a controlled environment without impairments. The talkers were 5 males and 5 females, whose 6 English and 4 Italian native speakers. The conversations varies from about 150 s to 240 s.

From the acquired voice signal, 1170 patterns were obtained: 787 patterns were used in the training phase and 383 patterns in the test one.

The architecture of the LVQ neural network was empirically determined. Tab. 1 shows, for the different configurations of the LVQ neural network's hidden layer, the number of the epochs, the value of the training coefficient $\alpha(t)$ and the minimum error in the testing phase. The configuration characterized by 4 neurons in the hidden layer offers correct classification of the 383 patterns after 10^5 epochs performed in the training phase using α=0.05.

Tab. 1. Epochs, α and classification error values versus the number of neurons in the hidden layer of the LVQ neural network with 128 neurons in the input layer.

Neuron number	epochs	α	error
2	10000	0.05	14
2	50000	0.05	6
2	100000	0.05	4
2	200000	0.05	4
2	100000	0.01	4
4	100000	0.05	0

3.3 Experimental results

The efficacy of the proposed algorithm has been verified by means of many experimental tests.

Three different types of noise were used to corrupt the voice signal: stationary noise, impulsive noise and Modulated Noise Reference Unit (MNRU).

The voice signal of different conversations in a controlled environment without impairments of male and female talkers is recorded. Only the meaningful results for each type of noise will be reported in the following. Fig. 4 shows the block scheme used in the tests.

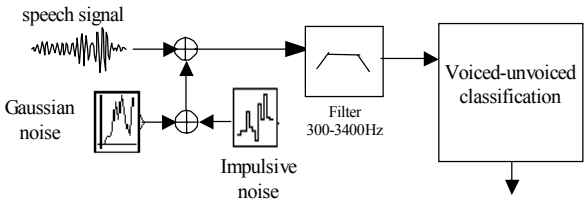

Fig. 4. Block scheme for experimental tests in the case of stationary and impulsive noise.

3.4 Stationary noise

The stationary noise is suggested by ITU-T Recommendation P.561 in order to simulate the actual telephone channel.

The gaussian noise has been varied so as to obtain different values of the Signal-to-Stationary Noise Ratio (SSNR). In particular, the SSNR has been evaluated as the ratio between the power of the voice signal evaluated by considering only the time portion containing useful voice, and the stationary wideband noise variance [13]. In Tab.2 the percentage error values versus different values of SSNR are compared. The SSNR is evaluated on 358 frames of voice signal for the proposed algorithm (column 2) and the technique proposed in [13] (column 3), respectively. In the case of the proposed algorithm, the stationary noise affects in a negligible way the voiced-unvoiced classification. In particular, for SSNR=1 dB the classification error is equivalent to the classification error of the technique [13] for SSNR=40 dB.

3.5 *Impulsive noise*

The impulsive noise is characterized by a burst with a high amplitude and a short length.

In this case the impulsive noise is added to the gaussian noise (Fig. 4). This last has been varied so as to obtain different values of the SSNR.

The impulsive noise has been varied so as to obtain different values of the Arrival Rate of Impulse (ARI) [13]. In all the tests the impulse amplitude is set equal to 2V. Fig. 5 shows the percentage error value versus different SSNR values, evaluated on 358 frames of voice signal, for different ARI values. In Fig. 5 the error values corresponding to the presence of the stationary noise only (ARI=0) is also reported.

The impulse noise affects in a not negligible way the voiced-unvoiced classification. In particular, at low values of SSNR the classification error increases if compared with the values corresponding to ARI=0. However, tests show that the proposed algorithm is always more accurate than the technique proposed in [13].

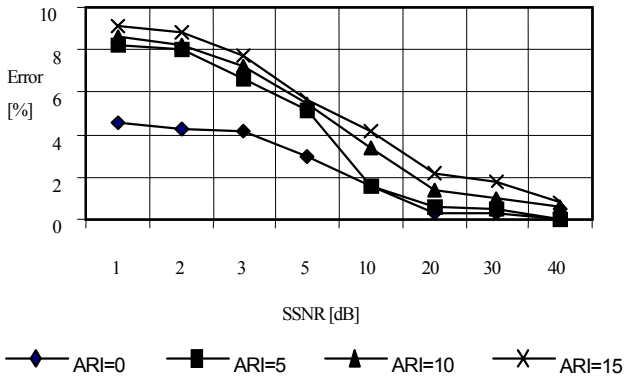

Fig. 5. Percentage error for different values of ARI versus SSNR signal, and the stationary wideband noise variance.

3.6 *Modulated Noise Reference Unit*

The (MNRU) is used in order to model the distortion and noise that characterize the digital process in the telecommunication networks [8], [17].

The block scheme is similar to that of Fig. 4 in which the noise generator is the MNRU generator, only. The signal y(i), obtained by adding the MNRU to the voice signal, is:

$$y(i) = x(i)[1+10^{-Q/20} N(i)], \qquad (3)$$

where x(i) is the voice, N(i) random noise, Q the ratio between the power of the voiced signal and the power of the modulated noise. Fig. 6 shows the percentage error versus the Q parameter. The MNRU affects in a not negligible way the voiced-unvoiced classification. In particular, for negative values of Q the classification error increases if compared with the values corresponding to SSNR and ARI=0.

4. ECHO PARAMETER MEASUREMENT

Several aspects exist in measuring echo in telecommunication systems. Relevant interest in characterizing echo involves measuring both the loss of the echo attenuation α, and the echo delay τ in the voiced classified signal without influence of different signals. A suitable procedure able to estimate these echo parameters is proposed in [7] and its flow diagram is included in

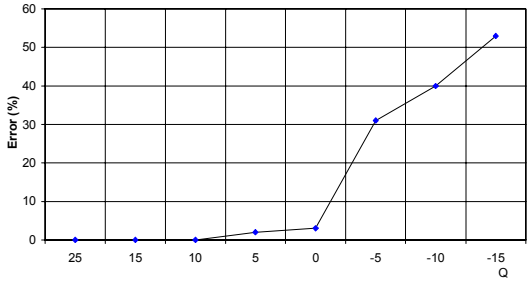

Fig. 6. Percentage error versus the parameter Q.

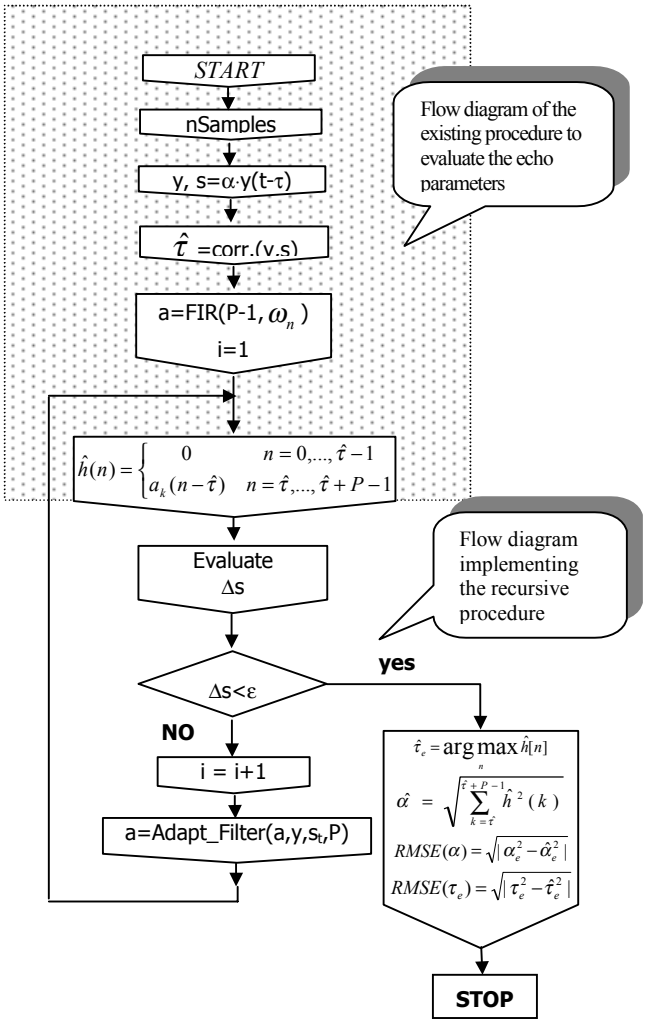

Fig.7 Flow diagram for echo attenuation and delay evaluation.

the dashed area of Fig. 7. The procedure is based on the esteem of the impulse response associated to the echo path. The esteem is conveniently made by using the FIR filter.

Therefore, denoting by h(n) the impulse response of FIR filter and by P the length of the FIR filter, the echo parameters are given by:

$$\tau = \arg\max[h(n)], \qquad \alpha = \sqrt{\sum_{k}^{P-1} h^2(n)} \qquad (4)$$

In order (i) to increase the esteem accuracy of the echo parameters, (ii) to reduce the sample number of the signal, and (iii) to minimize the impact of ADC characteristics [18] of the acquisition board, the recursive procedure is proposed and implemented. The fundamental steps are:

step 1. acquisition of the input signal to telephone-type network y(k), k=1,…, N and the echo signal s(k)=y(k)*h(k);

step 2. evaluation of the coefficients of FIR filter with impulse response h(n) according to [7];

step 3. reconstruction of the echo signal $s_r(k)$ and comparison with that previously acquired s(k);

step 4. end of the iterations if $\Delta s = \sum_{i=1}^{N} |s(i) - s_r(i)| \le \varepsilon$.

Fig. 7 shows the complete flow diagram of the modified procedure. Before the iterative procedure, both **τ** and **α** were estimated. The first esteem is given by means of correlation procedure for the delay **τ** and of FIR coefficient computation h(n), once the cut-off frequency ω_n is set, for the attenuation **α**. Successively, the iterative procedure evaluates the new coefficient h(n) of the FIR filter in the adap_Filter block. The FIR filter coefficients were modified by means of the descendent gradient method. The iterative procedure ends if either Δs<ε, or the variation of Δs in successive iteration is unappreciable. The final evaluation of τ and α is computed by the final values of the FIR filter, according to (4). The procedure gives also the Root Mean Square Error of α, *RMSE(α)*, and τ, *RMSE(τ)*, referred to the first esteem, respectively.

4.1 *Experimental results*

The procedure for echo parameter evaluation was experimental tested by using sampled voice signal at the frequency rate of 8 kHz. Fig.8 shows the window of 12.5 s width. The resulting signals processed by the voice-unvoiced block, Fig. 1, were constituted by the previous one added to its self once delayed and attenuated. In all the tests ε was equal to 10^{-3}.

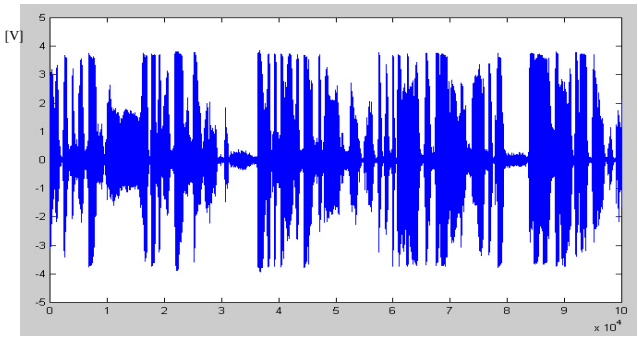

Fig. 8. Window of 100 ksample of the voiced signal used for experimental tests.

The uncertainty of time delay τ evaluation was always included in the sampling time interval.

The uncertainty of evaluation of the attenuation α was dependent on the number N of the frame's samples. Numerous tests were performed in order to investigate about this dependence. In particular, it was examined the trend versus different values of both the attenuation and the delay of the parameter RMSE(α) defined as:

$$\mathrm{RMSE}(\alpha) = \sqrt{\alpha^2 - \alpha_0^2} \qquad (5)$$

where α_0 is the fixed a priori attenuation. In Tab. 3 is shown RMSE(α) for two different values of N.

Tab. 4 shows the trend of the RMSE values once the number of frame's samples was established equal to 3072.

Tab. 3. Comparison of RMSE(α) versus different values of α and τ in the case N=1k and N=5k.

α [dB]	τ [ms]									
	10		30		50		70		90	
	N		N		N		N		N	
	1k	5k	1k	5k	1k	5k	1k	5k	1k	5k
6.000	0.141	0.115	0.132	0.108	0.124	0.106	0.119	0.104	0.112	0.101
16.000	0.189	0.151	0.171	0.147	0.159	0.141	0.134	0.121	0.126	0.116
26.000	0.191	0.153	0.173	0.149	0.160	0.143	0.135	0.122	0.127	0.118
36.000	0.193	0.153	0.174	0.149	0.161	0.144	0.136	0.122	0.128	0.118
46.000	0.193	0.153	0.175	0.149	0.160	0.144	0.136	0.122	0.128	0.118

Tab. 4. RMSE(α) for different values of delay τ and attenuation α in the case of N=3072 samples.

α [dB]	τ [ms]				
	10.2	30.8	50.2	70.7	90.9
6.000	0.127	0.119	0.112	0.110	0.107
16.000	0.174	0.159	0.157	0.140	0.135
26.000	0.175	0.161	0.159	0.141	0.137
36.000	0.175	0.161	0.159	0.141	0.137
46.000	0.175	0.161	0.159	0.141	0.137

5. CLARITY MEASUREMENT

The signal used for clarity measurement was the optimized multi-sine shown in Fig. 9.

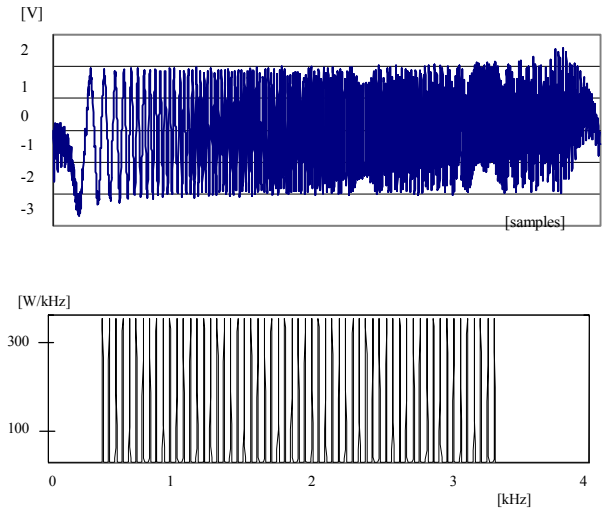

Fig. 9. Optimised multi-sine signal (upper) and the corresponding PSD (lower) in the band [300, 3400] Hz.

The effectiveness of this signal to clarity measurement can be examined by comparing the output values of the Perceptual Speech Quality Measurement (PSQM) algorithm feed by the optimized multi sine and the voiced signal, respectively. The voiced signal was the artificial voiced signal suggested by the ITU-T P.861 to test the PSQM algorithm. In order to obtain equivalence of the response in the cases that the input of the PSQM algorithm is the multi sine signal and the artificial voiced signal, respectively, the values of the multi sine parameters are: amplitude A_m=0.7 V, frequency uniformly distributed in the

range [300, 3400] Hz with step equal to 5 Hz, phase uniformly distributed in the range [14600, 13700] rad with step equal to 1 rad.

The comparison was made in the simulation environment in the following conditions:

1. both the multi sine and voiced signals were sampled at the 8kHz and 12 effective bit,
2. both the signals were corrupted by gaussian noise,
3. the noise amplitude added to multi sine and voiced signal, respectively, were different in order to realize the common value of the SNR,
4. both the multi sine and voiced signals were conditioned to realize spoken level equal to –26.15 dBov,
5. both the multi sine and voiced signals were conditioned to realize the hearing level equal to 78 dB SPL.

As shown in Fig. 10, the PSQM algorithm furnishes practically the same output values at different values of the superimposed gaussian noise when the input is the voiced or the multi-sine signal, respectively.

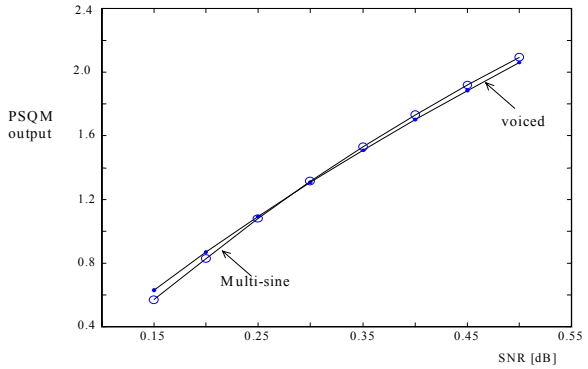

Fig. 10. A dimensional output values of the PSQM algorithm versus the SNR in the case the input is the multi sine and the voiced signal, respectively.

Tab. 5. PSQM and MOS values for different values of Q parameter of MNRU noise in the case of multi-sine and voiced input signal.

Q	Voiced signal		Multi-sine signal	
	PSQM	MOS	PSQM	MOS
-5	4.47	-5	3.78	-6
0	3.26	0	2.77	0
5	2.03	5	1.67	5
10	1,09	10	0.88	10
15	0.54	15	0.43	15
20	0.27	20	0.20	20

Moreover, tests were performed by superimposing impulsive and gaussian noise to both multi-sine and artificial voiced signal. In these tests, differently form the previous ones, the PSQM output values were different according to the different input signals. Indeed, the multi sine signal, having harmonic components in the high frequency of the telephone pass band, is more sensitive than the voiced signal to the high frequency disturbances. Other tests were performed by superimposing MNRU noise. Also in these tests the a-dimensional output values of PSQM were different, but, as shown in Tab. 5, the difference is reduced once computed the Mean Opinion Score (MOS).

6. CONCLUSIONS

In the paper the improved signal-processing algorithms for Voice Quality measurement is presented. According to the International Recommendations, this algorithm is able: (i) to classify the voice signal into voiced-unvoiced, (ii) to evaluate the echo parameters, and (iii) to measure the voice clarity.

Improvements are achieved owing (i) the use of the Learning Vector Quantization neural network, (ii) the adaptive evaluation of the coefficients of the FIR filter estimating the impulse response of the echo path, and (iii) the use of multi sine signal in place of the voiced signal.

ACKNOWLEDGMENTS

The author wish to thank to Prof. Pasquale Daponte for his helpful suggestions and active interest during all the phases of the research.

REFERENCES

[1] http://www.gl.com
[2] http://www.ascomm.com
[3] ITU-T Rec. P.561 "In-service Non-intrusive measurement device – Voice service measurement", 1996.
[4] D.B. Ramsden, "In-service non-intrusive measurement on speech signals", Proc. of GLOBECOM'91, pp.1761-1764.
[5] T. Kohonen, *Self-organizing maps*, Springer, 1997.
[6] A. Aiello, P. Daponte, D. Grimaldi, "Neural approach to voiced-unvoiced-silence analysis for quality measurements in telecommunication systems, Proc. of NAISO Con. on Inf. Sc. Innov. ISI'2001, March 17-21, 2001 Dubai (UAE), pp.892-896.
[7] M. Bertocco, P. Paglierani, "In-service non intrusive measurement of echo parameter in telephone-type network", IEEE Instr. and Measur. Tech. Conf. IMTC'98, St. Paul, Minnesota (USA), May 18-21, 1998, pp.614-617.
[8] ITU-T Rec. P. 861 "Objective quality measurement of telephone band (300-3400Hz) speech codes", 1998.
[9] D.A. Krubsack, R.J. Niederjohn, "An autocorrelation pitch detector and voicing decision with confidence measures developed for noise corrupted speech", IEEE Trans. on Signal Processing, vol.39, 1991, pp.319-329.
[10] L.R. Rabiner, "Application of voice processing to telecommunications", Proc. IEEE, vol.82, No.4, 1994, pp.199-228.
[11] B.S. Atal, L.R. Rabiner, "A pattern recognition approach to voiced-unvoiced-silence classification with application to speech recognition", IEEE Trans. on Acoust., Speech Signal Processing, vol. ASSP-24, 1976, pp.201-212.
[12] B.S. Atal, L.R. Rabiner, C.E. Schmidt, "Evaluation of a statistical approach to voiced-unvoiced-silence analysis for telephone quality speech", Bell Syst. Tech. Journal, vol.56, No.3, 1977, pp.455-482.
[13] M. Bertocco, P. Paglierani, "Non-intrusive measurement of impulsive noise in telephone-type networks", IEEE Transaction on Instrumentation and Measurement, vol.47, No.4, 1998, pp.864-868.
[14] L. Liao, M.A. Gregory, "Algorithms for speech classification", Fifth International Symposium on Signal Processing and its Applications, ISSPA'99, Brisbane, Australia, 22-25 August, 1999, pp.623-627.
[15] B. Yuhas, N. Ansari, *Neural networks in telecommunications*, Kluwer Academic Publishers, 1997.
[16] http://www.ni.com/catalog/pdf/1mhw286a.pdf
[17] ITU-T Rec. P.810 "Telephone transmission quality Modulated Noise Reference Unit (MNRU)", 1996.
[18] P. Arpaia, P. Daponte, L. Michaeli, "The influence of architecture on ADC modelling", IEEE Trans. on Instr. and Meas., vol.48, N.50, Oct. 1999, pp. 956-967.