Voice Quality Measurement in Telecommunication Networks by Optimized Multisine Signals

Domenico Luca Carnì, Domenico Grimaldi

Dept. of Electronics, Computer and System Sciences, University of Calabria, 87036 Rende – CS, Italy
Ph.:+39 0984 494712, Fax: +39 0984 494713, {dlcarni, grimaldi}@deis.unical.it

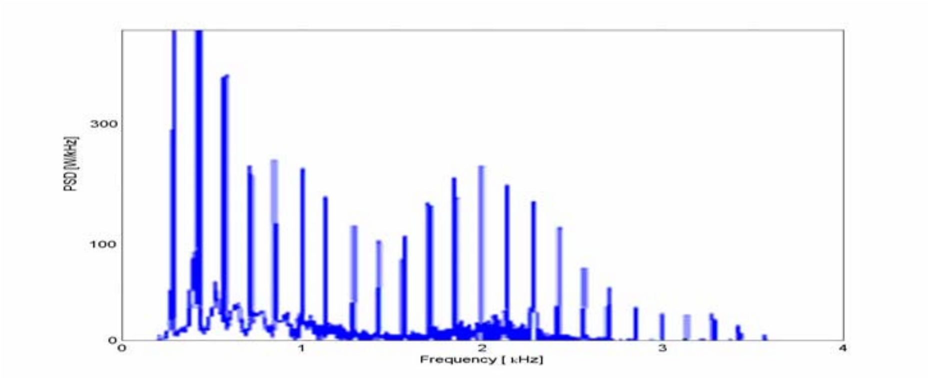
Abstract- The paper deals with the test signal selection to be adopted for the Voice Quality Measurement (VQM) in telecommunication networks. It is proposed to use the optimized multisine as test signal in place of both the natural and the artificial voices, suggested by the ITU-T recommendations. The effectiveness of the optimized multisine to the VQM is investigated by using four algorithms: (i) Perceptual Speech Quality Measurement, (ii) Perceptual Speech Quality Measurement plus, (iii) Perceptual Evaluation of Speech Quality, and (iv) Perceptual Analysis Measurement System.
Numerical tests are performed in order to compare the output index of the four algorithms feed by the optimized multisine, the natural and the artificial voice signals. In order to take into account the effects of the actual telecommunication networks, the Modulated Noise Reference Unit is used to corrupt both the multisine and the voices. The tests show the effectiveness of the proposed approach, based on the use of the optimized multisine, respect to the traditional procedures.

I. Introduction

The Voice Quality Measurement (VQM) is one important components in order to evaluate the Quality of Service [1] in VoIP systems, Public Switched Telephone Network (PSTN), Asynchronous Transfer Mode (ATM) networks, Frame Relay, and Wireless Networks [2], [3]. The voice quality is a subjective measure because it depends on individual perceptions [4]. In order to overcome the problems arising from the subjective measurement, the ITU-T recommendations (i) standardize different algorithms for the objective VQM within the bandwidth of 300-3400 Hz [5]-[8], and (ii) define the listening quality scale. The standardized algorithms are based on the comparison of the samples of original unprocessed signal with the samples of the degraded version. The results of these algorithms are indexes that can be mapped into the listening quality scale. In particular, the indexes can be mapped into the Mean Opinion Score (MOS) [4]. Four different algorithms are proposed from the ITU-T on the basis of the operating conditions of the telecommunication network. In particular:

1. the Perceptual Speech Quality Measurement (PSQM) predicts subjective quality of speech codecs without requiring subjective testing. It provides a relative score corresponding to how a statistically large number of human listeners would react [6];
2. the PSQM plus [7] is a modified version of the PSQM and it is able to be used in the case of remarkable distortion of the signal as (i) packet loss, and (ii) time clipping;
3. the Perceptual Evaluation of Speech Quality (PESQ) [8] provides an objective measurement of subjective listening tests on telephony systems in the case of (i) codec distortion, (ii) transmission error, (iii) packet loss, (iv) multiple transcoding, (v) environmental noise, and (vi) variable delay;
4. the Perceptual Analysis Measurement System (PAMS) [9] predicts the overall subjective listening quality without requiring actual subjective testing, avoiding a very expensive and time consuming process. It is able to operate, also, in the case of (i) packet loss, and (ii) variable delay.

According to the ITU-T recommendations, the procedure for the VQM uses as test signal either the natural voices selected from the technicians, or the artificial voices coded in the recommendation [10]. Both these two test signals involve some inconvenient. Indeed, the use of the natural voices requires to select numerous and statistical correct mother speeches for each test, both men and women, in order to make the VQM independent from the selected test signals. The use of the artificial voices overcome the previous difficulties and the VQM can be performed by using the same signals, always. Nevertheless, others problems are introduced because the voice signal is characterized by the trend of the Power Spectral Density (PSD) not uniformly distributed in the telephone band, as shown in Fig.1 (upper). In this manner the telecommunication system is tested in different conditions at every frequency. Moreover, the test signals constituted by natural and artificial voices can not be adapted and modified within the bandwidth of 300-3400 Hz as regarding (i) the spectral content in the frequency domain, and (ii) the frequency resolution of the spectral content. In order to overcome these inconvenient and to permit to test with reference signals with assigned characteristics, in the paper the optimized multisine signal is proposed to be used in the place of both the natural and artificial voices. This



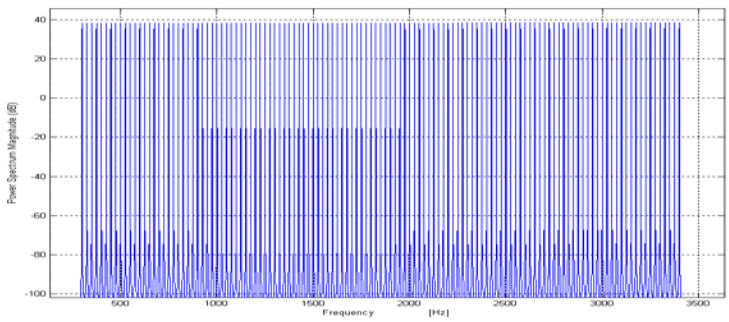


Fig.1 PSD of the artificial voice (upper) and the optimized multisine (lower) in the telephone band [300, 3400] Hz.

signal is easily to generate. As matter of fact, the multisine is the sum of sine waves characterised by (i) assigned amplitude, (ii) assigned spaced frequencies, and (iii) different phases to avoiding the peaks in the time domain causing the saturation of the electronic devices. Fig.1 shows the PSD of the optimized multisine signal in the telephone band. It can be noted that the PSD is constant at all the frequencies included in the signal.
On the basis of these considerations, the effectiveness of the optimized multisine signal to VQM is investigated. In particular, the output index of the four algorithms PSQM, PSQM plus, PESQ and PAMS feed by (i) the optimized multisine, (ii) natural voice signal, and the (iii) artificial voice signal are examined and compared.
The paper is organised as it follows. The characteristic aspects and the features of the multisine signal are discussed in order to justify the use for the VQM. Successively, numerical comparing tests are performed in MATLAB environmental by using as test signals (i) the natural voice signals, (ii) the artificial voice signals, and (iii) the optimized multisine. The test signals are degraded by means of the Modulated Noise Reference Unit (MNRU), according to the ITU-T recommendations [11].

II. Multisine signals

The multisine signal consists of series of sine waves, typically with a constant spacing frequency. One property of the optimized multisine signal is similar to the that of the artificial voice signal in the time domain: there is the presence of discrete amplitude variation. The advantages of the multisine signal respect to the artificial voice consist of: (i) the mathematic function can be used for the generation of the series of sine waves at different frequencies and imposed phase, (ii) the VQM is independent from the utilised speech samples, and (iii) the frequency spectrum has constant amplitude. The mathematic function of the multisine waveform is:

$$x(t) = \sum_{i=1}^{N} A_i \cos\left(2\pi f_i t + \phi_i\right), \tag{1}$$

where $f_i = l_i * f_0$, l_i is positive integer, and $f_{min} \le f_i \le f_{max}$. In order to reduce the crest factor, the phase can be determined according to the Schroeder expression:

$$\phi_i = -\frac{i(i-1)}{N}\pi \,. \tag{2}$$

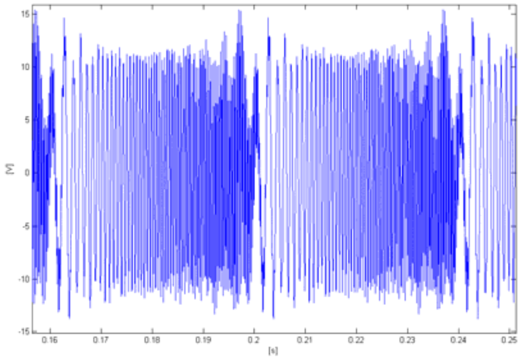


Fig.2 Trend of the optimised multisine in the time domain.

The resulting signal is also denoted as optimised multisine.
To utilize the multisine waveform as test signal for the telephone system, the band must to be limited into the telephone band. Therefore it is: f_{min}=300Hz and f_{max}=3400 Hz. Fig.2a shows the trend in the time domain of the optimised multisine generated according to (1) and (2). The corresponding PSD is shown in Fig.1.

III. Numerical tests

The effectiveness of the multisine signal to the VQM was examined by comparing the output index of the PSQM, PSQM plus, PESQ and PAMS algorithms feed by the optimized multisine and the voice signals, respectively. The voice signals were the natural and artificial voice signals suggested by the ITU-T P.861.
In order to model the distortion and the noise characterizing the digital process in the telecommunication networks [4], [7] the MNRU is used. MNRU performs a pre-established degradation of the signal quality. As shown in Fig.3, each of the algorithms PSQM, PSQM plus, PESQ and PAMS (i) operates by comparing the reference signal with the degraded one, and (ii) evaluates the index corresponding to the imposed value of the MNRU. The degraded signal furnished to the algorithms is:

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

where x(i) is the test signal (voice or multisine), N(i) is the random noise, Q is the ratio between the power of the test signal and the power of the modulated noise.

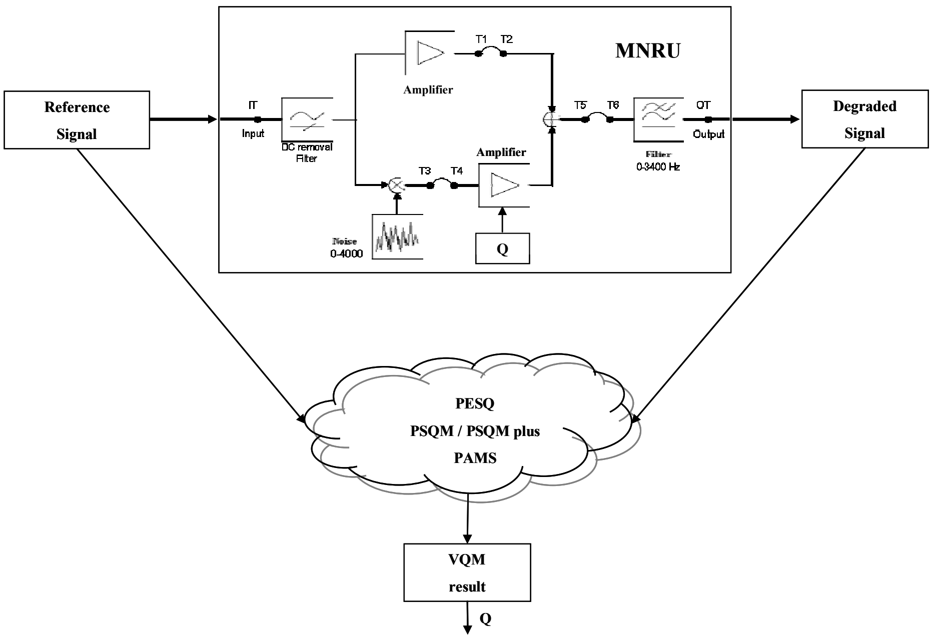


Fig.3 Block scheme to evaluate the multisine signal effectiveness for the VQM in comparison with other test signals.

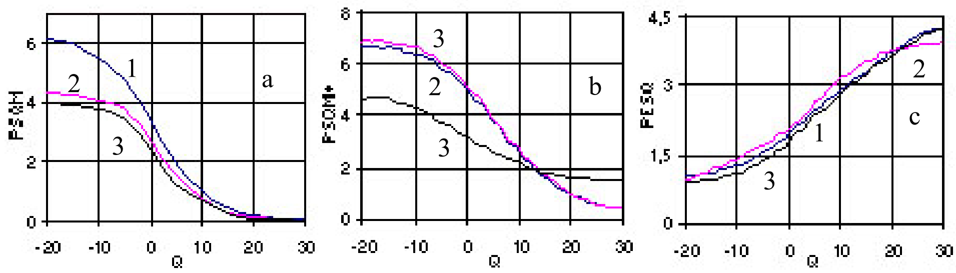


Fig.4 Trend of the mean index evaluated by a) PSQM, b) PSQM plus and c) PESQ versus Q for artificial voice signal (1), optimized multisine signal (2) and natural voice signal (3).

The characteristics of the natural and the artificial voice signals are defined into the standard, the characteristic parameters of the multisine are:

- ✓ amplitude A_m=0.7 V,
- ✓ frequency uniformly distributed in the range [300, 3400] Hz with step equal to 5Hz. The result is a periodic waveform with a period equal to the lowest common multiple of its frequency components.
- ✓ phase uniformly distributed according to (2).

The comparison is made in the following conditions:

- ➢ the multisine and voice signals are sampled at the 8kHz and 12 effective number of bits,
- ➢ the signals are corrupted by MNRU,
- ➢ the noise amplitude added to multisine and to voice signals was different in order to realize the common value of Q,
- ➢ the multisine and voice signals are conditioned to realize spoken level equal to –26.15 dBov,
- ➢ the multisine and voice signals are conditioned to realize the hearing level equal to 78 dB Sound Pressure Level (SPL).

As shown in Fig.4, the algorithms furnish practically the same output values at different values of the superimposed MNRU when the input is the voice or the multisine signal, respectively. Similar results are obtained from the PAMS algorithm.

Other tests are performed by superimposing impulsive and Gaussian noise to both the multisine and the artificial voice signal. In these tests the output values are different according to the different test signals. Indeed, the multisine signal having harmonic components in the high frequency of the telephone pass band, is more sensitive than the voice signal to the high frequency disturbances.

These tests show the efficacy of the proposed approach based on the use of the optimized multisine.

Conclusions

The use of the optimized multisine signal as test signal to the Voice Quality Measurement (VQM) is proposed. The basic differences among the natural and artificial voice signals used in the tests according to the ITU-T recommendations and the multisine signal are introduced and analysed.

The effectiveness of the optimized multisine to the VQM is investigated by using the four algorithms (i) Perceptual Speech Quality Measurement, (ii) Perceptual Speech Quality Measurement plus, (iii) Perceptual Evaluation of Speech Quality, and (iv) Perceptual Analysis Measurement System.

The effectiveness of the multisine signal was examined by comparing the output index of these algorithms feed by the optimized multisine signal, the natural voice signal and the artificial voice signal. The MNRU is used in to model the distortion and the noise characterizing the digital process in the telecommunication networks.

The tests highlight that the proposed approach can be conveniently used for the VQM.

References

[1] D. Grimaldi, "An improved procedure for QoS measurement in telecommunication systems", Proc. of 12th IMEKO TC-4 International Symposium, Zagreb, Croatia, September 25-27, 2002, pp.362-367.

[2] http://www.gl.com

[3] http://www.ascomm.com

[4] ITU-T Rec. P.800, "Method for subjective determination of transmission quality", 1996.

[5] ITU-T Rec. P.561 "In-service non-intrusive measurement device – Voice service measurement", 1996.

[6] ITU-T Rec. P. 861 "Objective quality measurement of telephone band (300-3400Hz) speech codes", 1998.

[7] ITU Study Group 12 – Contribution 20, "Improvement of the P.861 Perceptual Speech Quality Measure", 1999.

[8] ITU-T Rec. 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 codes", 2001.

[9] A.W. Rix, R. Reynolds, M.P. Hollier, "The perceptual analysis measurement system for robust end-to-end speech quality assessment", Proc. of IEEE Conf. On Acoustics, Speech and Signal Processing (ICASSP), Vol. 3 , 5-9 June 2000, pp.1515–1518.

[10] ITU-T Rec.P.50 "Artificial voices", 1999.

[11] ITU-T Rec. P.810 "Telephone transmission quality Modulated Noise Reference Unit (MNRU)", 1996.