XVII IMEKO World Congress
Metrology in the 3rd Millennium
June 22–27, 2003, Dubrovnik, Croatia

VOICE TRANSMISSION QUALITY MEASUREMENT BASED ON WAVELET TRANSFORMATION

Tomáš Dresler, Jan Holub, Radislav Šmíd

Department of Measurement K338, Faculty of Electrical Engineeing,
Czech Technical University, Prague, Czech Republic

Abstract – A new method for objective voice quality measurements is described in the article. It is based on wavelet transformation that enables intrinsically also time localisation of eventual impairments. In comparison with other methods (based on P.861 and P.862 algorithms), the described method saves about one half of operations needed for achievement of comparable results, thus saving computation power and time.

Keywords: Voice transmission quality measurements, wavelet transform.

1. INTRODUCTION

Measurement of voice transmission quality in the telecommunication industry is an important field from of all common technical, economic and law points of view. Perceptual audio coding and perceptual quality assessment techniques have generally the need to account for the properties of the human auditory system. Drawing on research on hearing, mathematical models may be constructed that reproduce key properties such as loudness perception and masking. This measurement are important for companies maintaining the transmission lines of any kind (metallic, optical, wireless etc.)

a) Human-like perception response, so the operator/technology vendor can estimate negative user reactions and prevent customer migration in technical and logistic way, in advance
b) Information and statistics of network operation, transmission line load and voice transmission errors for particular place, season, daytime, weather and other influences,
c) Possibility to comparison of various technologies, devices and solutions, that are offered by other companies, in order to consider their quality and choose the best trade-off,
d) Vindication of choice of particular technology.

Methods for objective measurement of voice transmission quality are normalised (for example ITU-T P.861, ITU-T P.862). Unfortunately, they are always aligned to currently widely used transmission technologies, thus they need not affect all transmission impairments or their combinations and sometimes differ from expected results (based on listening tests). Although these methods are based on human-like error perception evaluation and they are being developed and optimised for many years, there are still possibilities to enhance them from both accuracy and necessary computation power point of view.

The reason listed above is challenge for us to develop new method of measurement, based on comparison between source and sound signal distorted during transmission. The overlapped frequency spectrum analysis (based on FFT) is used for comparison of voice samples in the ITU-T P.862. This is our point of application of wavelet transformation, that can locate required information from differences between voice samples (source and distorted one).

2. WAVELET TRANSFORM

Continuous wavelet transform (CWT) [6], [7], provides an alternative to the classical Short-Time Fourier Transform (STFT) for the analysis of non-stationary signals. In the contrast to the STFT, which uses a constant analysis window, the CWT uses short windows at high frequencies and long windows at long frequencies (constant relative bandwidth). In a CWT, the notion of scale is introduced as an alternative to frequency (the scale can be understood as a reciprocal value of frequency in the FFT for specific analysing function - wavelet).

Proper choice of scales can set sensitivity in specific frequency domains (e.g. in Bark scale [2]). The last but not least benefit is that any of the source signals can be analyzed at once, not part by part with overlapping. This significantly contributes to computation power savings.

3. COMPUTATION STEPS

The algorithm consists of the following steps:

1. Raw time alignment
2. Amplitude alignment
3. DWT calculation
4. Variable delay compensation
5. Psychoacoustics model application

3.1. Raw Time and Amplitude Alignment

The input and output speech samples are aligned based on cross-correlation of absolute values of the samples. An eventual portion at the beginning and end of each sample that does not match any part of the second one is cancelled. Both samples are of the same length at the end of this step.

Both input and output speech samples are divided point by point by their overall mean value to equalize them to the same level.

3.2. Variable Delay Compensation

The compensation of variable delay is performed using DWT coefficients of the samples, see Tab.1 and also an example at Fig. 1. In all of our experiments, we have used "dmey" wavelet that is a FIR based Approximation of the Meyer Wavelet. Meyer wavelet ensures orthogonal analysis.

TABLE 1: Scales of DWT and corresponding number of samples (Y is number of samples of the speech sample) for 8 kSa/s sampling frequency

Scale	Frequency range [Hz]	Number of Samples
B1	0...125	Y/32
B2	125...250	Y/32
B3	250...500	Y/16
B4	500...1000	Y/8
B5	1000…2000	Y/4
B6	2000…4000	Y/2

Time alignment procedure has to be performed at the beginning of all voice quality measurements based on two sound files comparison like P.86x or PAMS. It basically means that the optimal shift of received file is found by means of correlation between input (transmitted) and output (received) files. Since contemporary codecs used especially in mobile networks does not necessarily keep the waveform but only spectrum amplitude information (phase information is lost), the correlation is usually performed not directly on waveforms but on their envelopes (calculated either by means of amplitude demodulation or by means of Hilbert transform). Such a correlation process is (in case of standard methods) performed either on the whole record (P.861) or recursively on its portions to find changes in transmission delay (called delay jitter) that occur in packet transmission technologies (voice over Internet protocol etc.).

Due to the average speech frequency occupation, the best scales for the delay examination are B3, B4, B5 that means the frequency range 250…2000 Hz. The example is given in Figure 2.

Delay is estimated on blocks at each scale B3.. B5 separately by means of segmented cross-correlation and combined using median function applied on the corresponding time points of all the relevant scales (see Figure 3). The fatal problem is low time resolution at lower scales caused by decimation (see Tab.1). Using Continuous Wavelet Transform (CWT) instead of DWT could solve this trouble but the computation would take more time in that case.

3.3. Samples DWT Comparison and Psychoacoustics Model Application

Frame powers on corresponding scales and time shifts are compared and positive and negative differences are collected separately. Also "speech" and "silence" time periods comparisons are stored separately (that gives 2x2 sums and 2x2 counters). As voice activity detector (VAD), a simple energy threshold approach is used. Differences at different scales are weighted according to the simplified human ear sensitivity (see Table 2). The masking effect can be eventually considered at this step by highlighting the most powerful scale at each time position (we did not test it since neither PAMS nor PESQ is considering masking as well).

The 4 calculated differences (speech-positive, speech-negative, silence-positive and silence-negative), normalized by relevant number of frames, serves as an input to psychoacoustics model. It contains proper weighting of differences identified in silence periods against speech periods and also positive versus negative differences. The final result is recalculated to MOS-like scale covering range 1…5. The recalculation formula has been found by least square fit to listening test results that were available for the speech samples used.

TABLE 2: Scales of DWT and corresponding Bark scales and averaged gains

DWT Scale	Bark Scale	Gain
B1	0-4	1E-5
B2	5-7	1E-3
B3	8-15	0.3
B4	16-27	0.9
B5	28-41	1
B6	42-55	0.8

4. COMPUTATION POWER SAVINGS

There are two independent principles contributing to computation savings: CWT implementation of time alignment procedure and avoiding time overlaps during FFT calculation.

The computation power is saved in our approach due to the fact that the same wavelet outputs that are further used for quality estimations are correlated without any mediation operations (like enveloping).

The second principle of saving (avoiding overlapping that is necessary for FFT procedures) is obvious. In standard methods, both original and received files are segmented into (usually) 16ms long packets with 50% overlap that are then processed by FFT. The overlapping is necessary not to miss any short time effect that can potentially occur just on the border between two neighboring packets (in case of non-overlapped packetisation). This means that each time-domain sample is processed twice by FFT. In our wavelet-based approach no such overlapping is necessary.

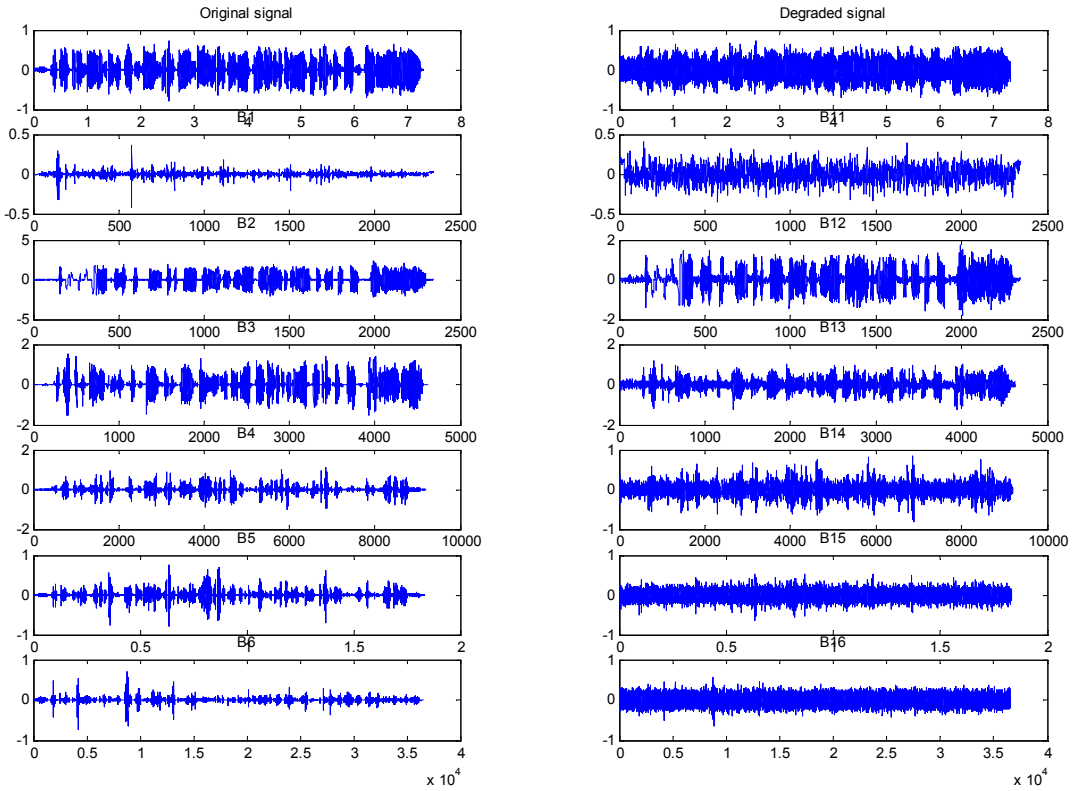

Fig. 1. DWT of the original (left) and transmitted (right) speech samples

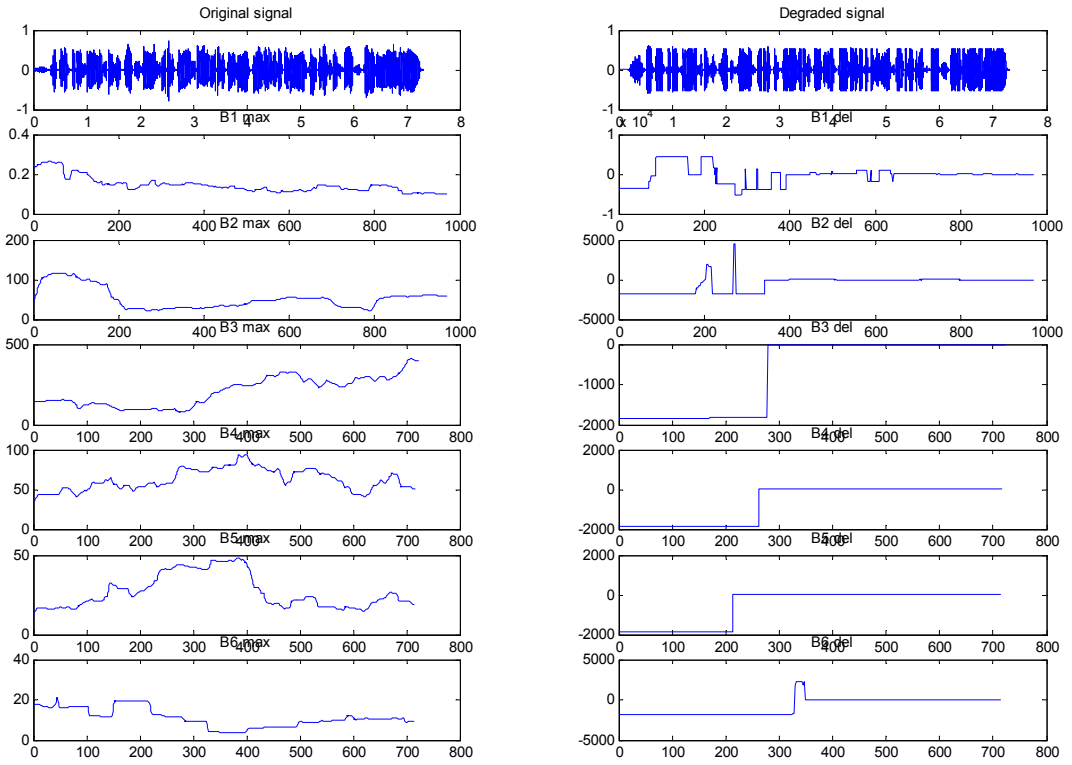

Fig. 2. Maxima of segmented crosscorrelations (left) and resulting delays (right) at various DWT scales (single simulated delay change at position 280)

5. MESUREMENT RESULTS

The final version of the algorithm has been tested on 130 speech samples fulfilling the P.80 requirements. Those samples were obtained partly on real transmissions in GSM networks, partly by artificial distortion (noise, amplitude and temporal clipping, echo, harmonic and non-harmonic distortion using Matlab Toolbox described in [4], avoiding changes in transmission delay. The correlation between results of our calculation and listening tests were for all sample subsets (noisy samples, clipped samples etc.) higher than 0.85. The maximum absolute difference of M.O.S. (Mean Opinion Score) results was 0.3 (M.O.S. ranges from 1=worst quality to 5=best quality, see [1] or P.80). This is fully comparable to accuracy of the relevant ITU standard [2].But the calculation according our approach took about 40% time on the same HW platform (common PC, PIII, 1GHz, 256 MB RAM).

6. CONCLUSIONS

A wavelet transformation based method for objective voice quality measurements is presented. As shown on preliminary results, it reduces the necessary time for calculation by more than one half while keeping comparable accuracy. Additionally, the information about time location of eventual impairment can be obtained form the result easier than in case of standard methods. However, to correctly suppress variable signal delay, it is highly recommendable to use CWT instead of DWT or to apply conventional approach (as given in PESQ).

ACKNOWLEDGEMENTS

This project is supported by Grant Agency of the Czech Republic, Advanced Measurements in Mobile Network under reg. no. 102/01/1355.

REFERENCES

[1] ETR 250 – Transmission and Multiplexing (TM), Speech Communication Quality from Mouth to Ear for 3,1 kHz Handset Telephony Across Networks, technical report, ETSI, July 1996

[2] ITU-T P.861, Objective Quality Measurement of Telephone-band (300-3400 Hz) speech codecs, ITU-T, February 1998

[3] ITU-T P.862, Perceptual Evaluation of Speech Quality, ITU-T, February 2001

[4] Holub J.: Advanced Measurement in Mobile Networks, 4th Concertation Meeting of Mobile, Wireless and Satellite IST Projects, Brussel, March 2001

[5] Rix, A., Beerends, J.G., Hollier, M.P., Hekstra, A. P.: Perceptual Evaluation of Speech Quality (PESQ) - a new method for speech quality assessment of telephone networks and codecs. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Salt Lake City, May 2001

[6] Teolis, A. *Computational Signal Processing with Wavelets*. Birkhäuser, Boston, 1998.

[7] J. Lewalle. Tutorial on continuous wavelet analysis of experimental data. Technical report, Syracuse University, April 1995.

AUTHOR(S): Tomáš Dresler, Jan Holub, Radislav Šmíd, Department of Measurement K338, Faculty of Electrical Engineeing, Czech Technical University, , Technicka 2, CZ-16627 Prague, Czech Republic, phone: +420 2 2435 2131, fax: +420 2 3333 9929, e-mail: holubjan@feld.cvut.cz