ELECTRONICS AND ELECTRICAL ENGINEERING
ISSN 1392 – 1215 2011. No. 3(109)
ELEKTRONIKA IR ELEKTROTECHNIKA

TELECOMMUNICATIONS ENGINEERING
T 180
TELEKOMUNIKACIJŲ INŽINERIJA

Method for Conversational Voice Quality Evaluation in Cellular Networks

A. Kajackas, A. Anskaitis, D. Gursnys
Telecommunications Engineering Department, Vilnius Gediminas Technical University,
Naugarduko str. 41, LT-03227 Vilnius, Lithuania, e-mails: algimantas.kajackas@el.vgtu.lt, aurimas.anskaitis@el.vgtu.lt, darius.gursnys@el.vgtu.lt

Introduction

Wireless and cellular communication is highly unique in a sense that signal strength constantly varies when the user changes his location. There exist critical localities in the network where QoS degradation is unavoidable. In cellular networks these localities are usually located near the edges of network cells where signal level is naturally lower.

Scientific literature pays adequate attention to QoS evaluation methods [1] and means in cellular communications conditions. Typical QoS evaluation tasks are well established in ITU-T G.1000:

- QoS surveys with the purpose to determine customer's requirements;
- Measurement of achieved QoS indicators;
- Investigation of QoS indicators and their dependencies on different quality impairment factors [2–5].

However the list presented does not include solutions how to measure and account user's perceived quality in real time. This work is concentrated on voice quality evaluation in mobile networks. An attempt is made to fill an existing scientific uncertainty – the method for evaluation and accounting of varying speech quality is proposed.

The paper shows that real time evaluation of speech quality may be performed using indicators based on count of lost voice data frames and voice activity indication.

Metrics for speech quality evaluation

Throughout the history of voice telephony subjective and objective voice quality assessment techniques have been used extensively in order to evaluate the quality of voice communication systems and/or its components. Subjective criteria were applied to determine the required voice spectrum bandwidth and allowed distortion levels. Expert quality assessments were also employed when modern codecs emerged, despite the fact they seriously distort the waveform of a signal.

An experience gathered during MOS testing later has been extended for creating objective quality evaluation algorithms which try to imitate human assessments of perceived speech quality. Well known quality evaluation algorithms are presented in ITU-T P.862 (PESQ) and P.564 (3SQM) recommendations. PESQ algorithm is invasive one – it uses original and degraded signals while calculating quality score for the degraded signal. PESQ is considered the most precise objective quality evaluation method currently in existence [3].

It is important to note that speech quality measurements using MOS and PESQ require special conditions. It is clear that these methods are inapplicable for evaluation of speech quality during real telephone conversation.

Speech quality impairing factors

QoS provided by network operators is limited by underlying technology. Let us call this limited quality level QoS_0. When communication conditions meet certain criteria then experienced quality is not significantly different than QoS_0.

Under real circumstances the quality experienced by particular user is a stochastic process characterized by time varying function $Q(t)$. Taking into account that communication conditions can only degrade resulting quality, the function can be expressed as difference

$$Q(t) = Q_0 - \Delta Q(t), \quad (1)$$

here Q_0 - maximum voice quality offered by the service provider and $\Delta Q(t)$ describes the quality degradation due to changing communication conditions.

In cellular networks voice is transmitted using circuit switching technologies. It means that the main quality impairment factor is packet loss. The delay variations are insignificant. For such conditions the function $\Delta Q(t)$ can be expressed using the following functional

$$\Delta Q(t) = \Phi(v(t), \rho, \pi)\,, \tag{2}$$

here $v(t)$ is voice signal, ρ is parameter, describing chosen codec, π – packet loss trace. Using the analogy with well-known E model [11], $\Delta Q(t)$ can be expressed as a sum of quality impairing components

$$\Delta Q(t) = \Phi_1(v(t), \rho) + \Phi_2(v(t), \pi)\,. \tag{3}$$

Functionals $\Phi_1(.)$, $\Phi_2(.)$ describe the relation between quality and separate quality impairing factors ρ and π.

Codec mode ρ affects voice quality because when communication conditions change, also changes and AMR codec type. AMR is the most widely used voice codec in current mobile communications [9]. The impact of codec mode may be expressed using data from 3GPP document [9].

In cases when the same codec is used during the whole conversation formula (4) can be further simplified by leaving only the second term. Now quality degradation depends only on packet loss pattern.

Proposed algorithm for speech quality evaluation

In mobile communication systems essential voice quality deterioration is created by loss of transmitted voice frames. In GSM and UMTS systems lost frames are marked as BFI (Bad Frame Indicator) at the receiver. The frame is marked as lost when OSI channel layer finds CRC (Cyclic Redundancy Check) failure. When such frame arrives to the receiver, special decoding strategy – so called erased frame substitution – is performed. This somewhat decreases speech quality degradation. Unfortunately some quite noticeable quality degradation remains. Currently, the end user equipment can't assess voice quality. There is no way to apply PESQ when real conversation is going on, because of PESQ requires reference signal which is available only at transmitting side. Our proposed algorithm tries to fill this gap in quality evaluation. During long time investigations we have noted [8] that voice quality impairment in every 320 ms duration syllable can be approximated using only two parameters: number of lost frames in this syllable (N_F) and number of silent frames (N_S) in the same syllable. In this paper presented metric for speech quality evaluation is non-intrusive and parameter based. This metric uses only to indicators: number of lost frames N_F and number of silent frames N_S.

The proposed algorithm for voice quality evaluation based on two stages: training stage and decision making stage. The similar concept is also used in speech and speaker recognition [12, 13].

Training stage. The structure of tool for training stage is shown in Fig. 1. For every signal in training signal database (every signal is of 320 ms duration) three parameters were calculated. The number of lost frames N_F, number of silent frames N_S and quality degradation value Δq. The first two parameters are already discussed; the third is calculated using modified PESQ algorithm [8] shown in Fig. 1. We should note that it is not possible to calculate standard PESQ for 320 ms signal when there are many lost frames in this interval. This is because PESQ (and modified PESQ) requires certain amount of active voice frames and many lost frames in such interval make this interval almost completely silent. Because of this peculiarity of PESQ the method shown in picture was chosen. In particular, we added 1 s of active voice before and after the investigated signal. Quality score was measured for this composite signal.

For experimentation voice signals from real conversations were used. They contained variable percentage of silence – from 30% to 60%. Percentage of silence was measured using voice activity detection algorithm (VAD) [10].

Training data consisted of 80000 speech samples. Frame losses were generated independently from 1 frame loss to 16 frame losses (one voice frame contains 20 ms of voice signal). 16 frame losses means that all frames in the investigated interval are lost. AMR 12.2 codec was used for signal coding-decoding and frame loss simulation.

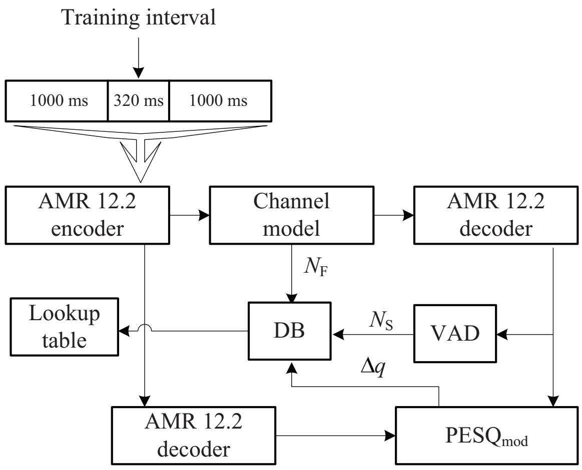

Fig. 1. The structure of training tool

Full collection of training data is 80000 triplets $\Delta q_i(N_F, N_S)$. Using this data 17x17 matrix – lookup table – was created. To get particular degradation value we average triplets $\Delta q_i(N_{Fb}, N_S)$ as follows

$$\Delta\overline{q}(N_F, N_S) = \frac{1}{M}\sum_{i=1}^{M} \Delta q_i(N_F, N_S)\,, \tag{4}$$

here M is the number of measurements for particular element from lookup table. Some points of empirical function $\Delta\overline{q}(N_F, N_S)$ are listed in Table 1.

Table 1. Quality degradation lookup table

N_F	N_S							
	0	1	2	3	…	14	15	16
0	0	0	0	0		0	0	0
1	0.34	0.44	0.35	0.32		0.11	0.10	0.04
2	0.53	0.65	0.59	0.55		0.21	0.16	0.09
4	0.81	0.98	0.93	0.95		0.42	0.42	0.14
…								

It can be seen that the more silent intervals are in the syllable, the less quality is degraded. Also quality degradation increases with the number of lost frames.

It should be noted that absolute quality degradation values are relatively small. This is because measurement signal length is 2.32 s and signal distortions occur only in short 320 ms duration segment.

The empirical function lookup table in decision stage is used as a short-time speech quality measure when N_F and N_S are stated.

Decision stage. Decision making is the key of the algorithm – it predicts quality degradation of a received signal. The structure of algorithm for calculation of real time voice quality impairments is shown in Fig. 2.

The calculation of voice quality impairments are performed in 4 steps:

1. Segmentation of voice signal using windows of 320 ms duration and 50% overlap. Such segmentation is similar to PESQ algorithm used [3].

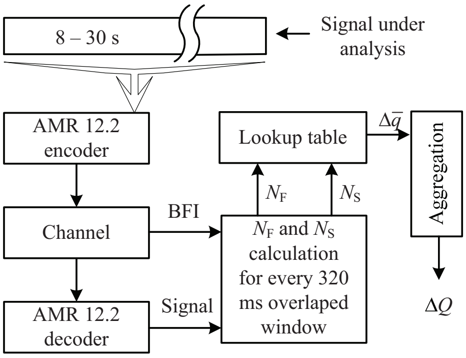

Fig. 2. Algorithm for calculation of voice quality impairments

2. For every *j*-th voice segment the number of lost frames N_F shall be obtained from lower communication layer and number of silent frames N_S shall be obtained using VAD.

3. The value of voice quality impairment in a given *j*-th segment $\Delta\bar{q}_j$ are obtained from lookup table as element in N_F-th row and N_S-th column.

4. The final decision stage – calculation of quality impairment ΔQ for full investigated signal using formula

$$\Delta Q = c \cdot \sqrt{\frac{1}{N}\sum_{j=1}^{N} \Delta\bar{q}_j^2}\ , \tag{5}$$

here N is a number of 320 ms syllables and overlapped by 50 % syllables in the signal. The length of investigated signal may be arbitrary.

The multiplier c in (6) is constant which was calculated minimizing RMSE between predicted scores and PESQ scores for the set of long sets of training signals. In this particular case c is 3.4.

The proposed method for speech quality prediction can be used to test of voice quality impairments when the main source of distortions is frame losses. It is possible to use the algorithm in IP networks but delays should be taken into account separately.

Algorithm verification

The verification of proposed algorithm was performed using records with different silence percentage. There were 3 such records groups, each of 250 records. Duration of every record was 32 s. The records consist of 50 % male and 50 % female voices.

For every record in every group the two frame loss models were used: independent Bernoulli errors and grouped Gilbert errors.

Voice records for verification and training steps were recorded by different speakers.

Voice quality degradation ΔQ value was calculated using proposed algorithm and formula (6). The second quality degradation ΔQ_{PESQ} value was measured using standard PESQ and the same voice records with the same frame losses.

The result of such measurements is data pairs (ΔQ_{PESQ}, ΔQ). The data for the first experiment is shown in Fig. 3. It can be seen that quality degradation prediction is quite precise.

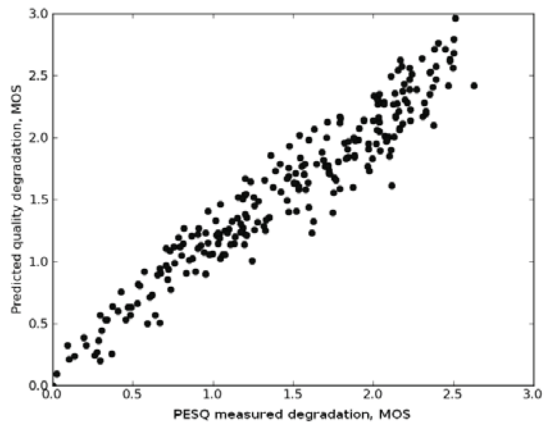

Fig 3. Prognosis of voice quality degradation

In Table 2 listed verification data when frame losses were simulated using Bernoulli model (frame loss rate between 0 and 30%).

Table 2. Algorithm verification data for Bernoulli losses

Silence %	30	50
Correlation	0.985	0.977
RMSE	0.136	0.208

In Table 3 listed verification data when frame losses were simulated using Gilbert model. In particular, grouping parameter (average number of consecutive lost frames) was 3.

Table 3. Algorithm verification data for Gilbert losses

Silence %	30	50
Correlation	0.980	0.974
RMSE	0.215	0.177

Obtained results show that proposed algorithm is sufficiently precise and robust. Similar performance metrics are obtained using different frame loss models and different active voice percentage in test records. In 3GPP document [9] it is said that quality variations of up to 0.2 MOS scores are statistically insignificant. Because of this we think that proposed algorithm is ready for real quality measurements.

Conclusions

In this paper the new non-intrusive parameter based speech quality evaluation method and algorithm for use in wireless voice communication systems was created. It is based on count of lost voice data frames and voice activity indication.

Speech quality measurement results of proposed algorithm show high correlation and low RMSE for the test cases with results obtained under same conditions using PESQ algorithm.

Proposed algorithm can be implemented in mobile stations and used for speech quality evaluation by real conversation. This way using real time voice quality evaluation measures network operators may offer services of different quality. The payment for telecommunication services may become dependent on the measured service quality.

Acknowledgement

Research was supported by the Research Council of Lithuania. Project Contract No. AUT-01/2010.

References

1. **Adomkus T., Kalvaitis E.** Investigation of VoIP Quality of Service using SRTP Protocol // Electronics and Electrical Engineering. – Kaunas: Technologija, 2008. – No. 4(84). – P. 85-88.
2. **Kajackas A., Anskaitis A.** Investigation the ability of objective measures of the perceptual speech quality in mobile networks // Electronics and Electrical Engineering. – Kaunas: Technologija, 2005. – No. 7(63). – P. 10–15.
3. **Kajackas A., Anskaitis A., Guršnys D.** Peculiarities of Testing the Impact of Packet Loss on Voice Quality. // Electronics and Electrical Engineering. – Kaunas: Technologija, 2008. – No. 2(82). – P. 35–40.
4. **Pastrana R., et al.** Sporadic Signal Loss Impact on Auditory Quality Perception. Online: http://wireless.feld.cvut.cz/mesaqin2004/contributions.html.
5. France Telecom Study of the relationship between instantaneous and overall subjective speech quality for time-varying quality speech sequences: influence of a recency effect. ITU Study Group 12 Contribution D.139, 2000.
6. ANSI/ASA3.2-2009. Method for Measuring the Intelligibility of Speech over Communications System.
7. **Karlsson A., et al** Radio link parameter based speech quality index-SQI // Proc. IEEE Workshop on Speech Coding, 1999. – P. 147-149.
8. **Anskaitis A., Kajackas A.** The Tool for Quality Estimation of Short Voice Segments // Electronics and Electrical Engineering. – Kaunas: Technologija, 2010. – No. 8(104). –P. 97–102.
9. 3GPP TR26.975 V5.0.0 (2002). Performance characterization of the Adaptive Multi-Rate (AMR) speech codec.0
10. **Tüske Z., Mihajlik P., Tobler Z., Fegyó T.** Robust Voice Activity Detection Based on the Entropy of Noise-Suppressed Spectrum // Interspeech, 2005.
11. **ITU-T rec. G.107.** The E-model, a computational model for use in transmission planning.
12. **Šalna B., Kamarauskas J.** Evaluation of Effectiveness of Different Methods in Speaker Recognition // Electronics and Electrical Engineering. – Kaunas: Technologija, 2010. – No. 2(98). – P. 67–70.
13. **Tamulevičius G., Arminas V., Ivanovas E. and Navakauskas D.** Hardware Accelerated FPGA Implementation of Lithuanian Isolated Word Recognition System // Electronics and Electrical Engineering. – Kaunas: Technologija, 2010. – No. 3(99). – P. 67–70.

Received 2011 01 05

A. Kajackas, A. Anskaitis, D. Gursnys. Method for Conversational Voice Quality Evaluation in Cellular Networks // Electronics and Electrical Engineering. – Kaunas: Technologija, 2011. – No. 3(109). – P. 105–108.

In this paper, method for evaluation of varying conversational speech quality in wireless communications is proposed. The proposed algorithm evaluates quality degradations using indicators based on count of lost frames and voice activity indications. The correctness of proposed algorithm is investigated by comparison of test results with results obtained using PESQ algorithm under same conditions. The achieved average correlation coefficient is 0.975. This result is independent of frame loss model and percentage of silence in test sentences. Proposed algorithm can be implemented in mobile stations and used for speech quality evaluation by real conversation. Ill. 3, bibl. 13, tabl. 3 (in English; abstracts in English and Lithuanian).

A. Kajackas, A. Anskaitis, D. Guršnys. Mobiliesiems tinklams skirta realaus pokalbio balso signalo kokybės vertinimo priemonė // Elektronika ir elektrotechnika. – Kaunas: Technologija, 2011. – Nr. 3(109). – P. 105–108.

Realiose telekomunikacijų sistemose nėra galimybės naudoti standartizuoto PESQ algoritmo. Šiame straipsnyje pasiūlytas realaus laiko balso kokybės vertinimo metodas gali būti įdiegtas į judriojo ryšio mobiliąsias stotis ir naudojamas perduoto balso kokybei vertinti. Kokybė nustatoma pagal ištrintus paketus ir signalo nebuvimo intervalus. Pasiūlytu metodu gautų balso kokybės pablogėjimo įverčių tarpusavio koreliacija su PESQ matavimų rezultatais siekia 0,975. Šis rezultatas gautas nepriklausomai nuo prarandamų paketų modelio ar signalo nebuvimo intervalų skaičiaus. Il. 3, bibl. 13, lent. 3 (anglų kalba; santraukos anglų ir lietuvių k.).