VILNIUS GEDIMINAS TECHNICAL UNIVERSITY

Aurimas ANSKAITIS

ANALYSIS OF QUALITY OF CODED VOICE SIGNALS

SUMMARY OF DOCTORAL DISSERTATION

TECHNOLOGICAL SCIENCES,
ELECTRICAL AND ELECTRONIC ENGINEERING (01T)

VGTU LEIDYKLA TECHNIKA

Vilnius 2009

Doctoral dissertation was prepared at Vilnius Gediminas Technical University in 2005–2009.

Scientific Supervisor

Prof Dr Habil Algimantas KAJACKAS (Vilnius Gediminas Technical University, Technological Sciences, Electrical and Electronic Engineering – 01T).

The dissertation is being defended at the Council of Scientific Field of Electrical and Electronic Engineering at Vilnius Gediminas Technical University:

Chairman

Prof Dr Habil Romanas MARTAVIČIUS (Vilnius Gediminas Technical University, Technological Sciences, Electrical and Electronic Engineering – 01T).

Members:

Prof Dr Habil Gintautas DZEMYDA (Institute of Mathematics and Informatics, Technological Sciences, Informatics Engineering – 07T),

Prof Dr Habil Romualdas NAVICKAS (Vilnius Gediminas Technical University, Technological Sciences, Electrical and Electronic Engineering – 01T),

Assoc Prof Dr Šarūnas PAULIKAS (Vilnius Gediminas Technical University, Technological Sciences, Electrical and Electronic Engineering – 01T),

Prof Dr Jonas RIMAS (Kaunas University of Technology, Physical Sciences, Informatics – 09P).

Opponents:

Prof Dr Dalius NAVAKAUSKAS (Vilnius Gediminas Technical University, Technological Sciences, Electrical and Electronic Engineering – 01T),

Dr Algimantas Aleksandras RUDŽIONIS (Kaunas University of Technology, Technological Sciences, Informatics Engineering – 07T).

The dissertation will be defended at the public meeting of the Council of Scientific Field of Electrical and Electronic Engineering in the Senate Hall of Vilnius Gediminas Technical University at 1 p. m. on 18 December 2009.

Address: Saulėtekio al. 11, LT-10223 Vilnius, Lithuania.
Tel.: +370 5 274 4952, +370 5 274 4956; fax +370 5 270 0112;
e-mail: doktor@adm.vgtu.lt

The summary of the doctoral dissertation was distributed on 17 November 2009.

A copy of the doctoral dissertation is available for review at the Library of Vilnius Gediminas Technical University (Saulėtekio al. 14, LT-10223 Vilnius, Lithuania).

VILNIAUS GEDIMINO TECHNIKOS UNIVERSITETAS

Aurimas ANSKAITIS

KODUOTO BALSO KOKYBĖS TYRIMAS

DAKTARO DISERTACIJOS SANTRAUKA

TECHNOLOGIJOS MOKSLAI,
ELEKTROS IR ELEKTRONIKOS INŽINERIJA (01T)

Vilnius LEIDYKLA TECHNIKA 2009

Disertacija rengta 2005–2009 metais Vilniaus Gedimino technikos universitete.

Mokslinis vadovas

prof. habil. dr. Algimantas KAJACKAS (Vilniaus Gedimino technikos universitetas, technologijos mokslai, elektros ir elektronikos inžinerija – 01T).

Disertacija ginama Vilniaus Gedimino technikos universiteto Elektros ir elektronikos inžinerijos mokslo krypties taryboje:

Pirmininkas

prof. habil. dr. Romanas MARTAVIČIUS (Vilniaus Gedimino technikos universitetas, technologijos mokslai, elektros ir elektronikos inžinerija – 01T).

Nariai:

prof. habil. dr. Gintautas DZEMYDA (Matematikos ir informatikos institutas, technologijos mokslai, informatikos inžinerija – 07T),

prof. habil. dr. Romualdas NAVICKAS (Vilniaus Gedimino technikos universitetas, technologijos mokslai, elektros ir elektronikos inžinerija – 01T),

doc. dr. Šarūnas PAULIKAS (Vilniaus Gedimino technikos universitetas, technologijos mokslai, elektros ir elektronikos inžinerija – 01T),

prof. dr. Jonas RIMAS (Kauno technologijos universitetas, fiziniai mokslai, informatika – 09P).

Oponentai:

prof. dr. Dalius NAVAKAUSKAS (Vilniaus Gedimino technikos universitetas, technologijos mokslai, elektros ir elektronikos inžinerija – 01T),

dr. Algimantas Aleksandras RUDŽIONIS (Kauno technologijos universitetas, technologijos mokslai, informatikos inžinerija – 07T).

Disertacija bus ginama viešame Elektros ir elektronikos inžinerijos mokslo krypties tarybos posėdyje 2009 m. gruodžio 18 d. 13 val. Vilniaus Gedimino technikos universiteto senato posėdžių salėje.

Adresas: Saulėtekio al. 11, LT-10223 Vilnius, Lietuva.
Tel.: (8 5) 274 4952, (8 5) 274 4956; faksas (8 5) 270 0112;
el. paštas doktor@adm.vgtu.lt

Disertacijos santrauka išsiuntinėta 2009 m. lapkričio 17 d.

Disertaciją galima peržiūrėti Vilniaus Gedimino technikos universiteto bibliotekoje (Saulėtekio al. 14, LT-10223 Vilnius, Lietuva).

VGTU leidyklos „Technika“ 1678-M mokslo literatūros knyga.

Introduction

Topicality of the problem. Voice transmission systems are constantly developed, implemented, and widely used. Current telecommunication systems divide voice stream into equal length segments and those are coded by a codec. The result of coding is data packets – so called frames. The frames are transmitted over the network. Disturbances in communication channel causes bit errors some of which can not be corrected. The frame with such errors is marked as bad frame and is said to be lost. Delays in IP networks are translated into frame losses too because severely delayed frame is worthless for the far end user. Frame losses reduces overall voice quality.

Current telecommunication systems are designed according to ITU E-800 recommendation and underlying QoS conception. QoS is characterized as end user perception of service quality. Practical construction of communication systems tries to maintain acceptable average quality. This approach works well when communication conditions are fixed. In mobile communication scenarios there are groups of users who constantly get poor quality. So monitoring and measurement of de facto perceived voice quality is very important task.

Analysis of voice signals was developed for a long time now. Very popular task is speech recognition. But in current systems we have one more dimension of complexity – voice is coded and sometimes decoded with frame losses. So it is important to describe the features of voice signals in conjunction with voice coding.

The fact is that voice quality evaluation is important not only for academics. Big players in telecommunications also solve tasks related to quality of service. Examples are „Nokia", „Nortel Networks" and others.

Taking into account presented data, it is obvious that voice quality research is of big importance for academical and practical purposes.

The object of research. The object of research is a quality of coded voice.

Aim and tasks of the work. The aim of the work is to improve voice quality evaluation algorithms.

The tasks of the work are:

1. Construction of the means for measurement of voice quality of short voice signals.
2. To define the concept of value of coded voice segment and to choose corresponding value metrics.
3. To measure distributions of frame values in standard voice.
4. To establish limits of distortions created by different codecs.

5. To investigate inertia of wide spread codecs. To establish the length of impact of one lost frame.
6. To investigate possibilities of frame value prognosis in real time.

Methodology of research

Dissertation relies on the methods of statistics, linear algebra and signal analysis. Hypotheses are verified using models of phenomenons under investigation and underlying simulations.

Scientific novelty

1. The method of voice signal synthesis was created which allows to apply PESQ algorithm for short signals.
2. Conception of informational voice frame value is created. The methodology for evaluation of informational frame value was proposed.
3. Statistical characteristics of the impact of lost frame on voice signal over time were estimated.
4. The algorithms for frame value prognosis in real time were constructed.

Practical value

The results of research can be applied when creating and optimizing next generation of voice telecommunication systems. Results also can be applied as algorithms in end user equipment for quality accounting.

Defended propositions

1. Algorithm described in ITU P.862 recommendation (PESQ) is not suited for quality evaluation of short voice signals.
2. Proposed signal extension method allows to use PESQ for short signals.
3. Informational value of a frame can be calculated as a difference between qualities when frame is received successfully and when the frame is lost.
4. Distribution of frame value has a shape of asymmetrical bell.
5. The biggest distortions in the coded signal are observed exactly one frame apart from the lost frame.
6. It is possible to predict voice frame value in real time. Correlation coefficient obtained with exact frame value is 0.6.

The scope of the scientific work. The scientific work consists of the introduction, 4 chapters, conclusions, list of literature, list of publications and addendum. The total scope of the dissertation – 105 pages, 45 pictures, 13 tables.

1. Voice quality evaluation tasks and methods

In this chapter voice evaluation tasks in telecommunications are analysed. Factors affecting voice quality are presented.

In all times voice transfer was one of the most important parts of telecommunications. Voice quality evaluation is of first importance too. In analogue telecommunication systems the main criteria for quality evaluation was signal to noise ratio. With the development of modern voice systems, criteria for quality evaluation have changed too. Currently the main factor which decreases voice quality is packet loss and coding. Because of this many methods for quality evaluation are available currently.

Classification of voice quality assessment methods is shown in Fig. 1.

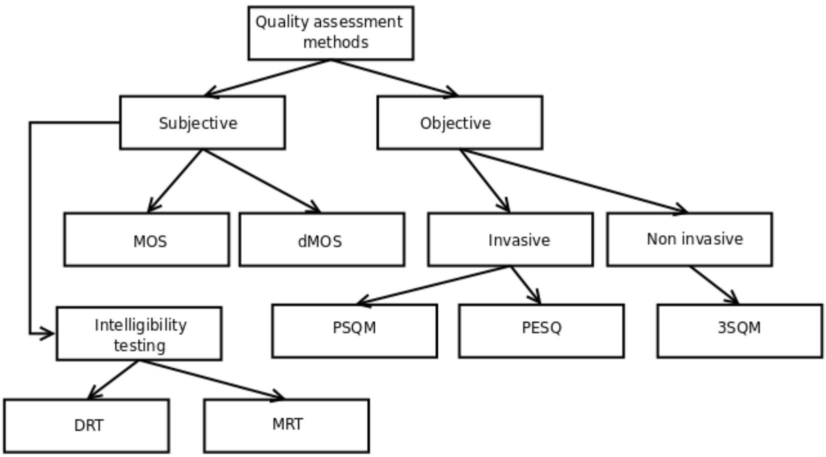

Fig. 1. Classification of voice evaluation methods

There are two main groups of voice quality evaluation methods – objective and subjective ones.

All subjective methods rely on the opinion of a group of listeners. The main method in this group is called Mean Opinion Score (MOS). The tests here are performed as follows. The selected group of listeners are asked to evaluate given record on a scale of one to five. Final score is formed as an average of scores given by individual listeners. It is obvious that this method is not reliable in a sense that different tests yield different results. Nonetheless the scores of MOS evaluations are used for calibration of known objective quality evaluation methods.

There exist many objective voice quality evaluation methods. The most popular and widely used is PESQ algorithm standardized by ITU P.862 recommendation. There are also less popular algorithms – PSQM, 3SQM. The algorithm compares two signals – original and degraded – and analysing differences between these signals calculates score which is highly correlated (correlation up to 0.93) with MOS scores.

Intelligibility testing works by asking listeners what word was played from a group of words. Depending on the groups of words, few methods exist – DRT (Diagnostic Rhyme Test) and MRT (Modified Rhyme Test) are the most popular ones. Lithuanian word lists for intelligibility testing were created in the dissertation.

All intrusive objective quality evaluation algorithms work using the same basic scheme. This scheme is depicted in Fig. 2. The signals (original and degraded) are preprocessed in the first step. After this windowing and feature extraction follows. The last step calculates integral difference between corresponding feature vectors. This calculation is optimized to yield results similar to MOS ones.

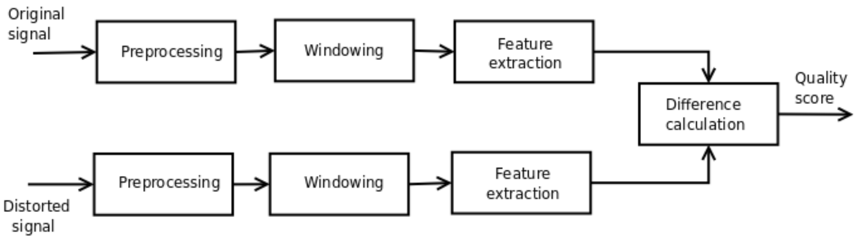

Fig. 2. Working scheme of objective voice quality evaluation algorithms

2. Value of voice segment

Current telecommunication systems divide voice signal into equal length intervals called frames. Let us call these frames after coding C_i. It is obvious that different frames have not equal value for speech perception. Some frames may be lost without noticeable impact on speech quality (silence frames) while others are very important and their loss may impact word intelligibility.

Frame loss is a complex process at a signal level. The fact of frame loss is signalled to decoder by lower telecommunication layers (channel layer in mobile communications). Then decoder employs the following strategy to construct the signal for a missing frame. Parameters of the last successfully received frame are used for current frame (which is lost). For subsequent lost frames the energy of reconstructed signal is reduced also. This algorithms reconstructs the signal of the lost frame well if signal is relatively stationary.

The main concepts in this chapter is informational frame value. We hypothesize that every frame could be assigned a value which shows the importance of the frame for overall voice quality. By definition frame value is the difference between voice quality when frame is received succesfully and voice quality when frame is lost. Frame value shows us the amount of voice quality degradation due to loss of this frame.

According to above given definition, frame value is calculated as:

$$V_i = \Delta Q = Q_0 - Q_i \,. \tag{1}$$

Here Q_0 is a voice quality after coding-decoding of a signal. The quality may be calculated using any objective algorithm. Our choice is PESQ algorithm. Q_i is a voice quality after coding-frame error simulation-decoding. From this definition it can be seen that value of a frame depends on the codec in use. In our experiment AMR (Adaptive Multi Rate) family of codecs was used. Results will be presented for AMR 12.2 kb/s codec.

While performing experiments it was noted that PESQ algorithm is not completely precise when used for very short signals (0.5 s). So special signal formation technique was developed. Imagine that we have one lost frame in the middle of a degraded signal and original signal is obtained by simply coding decoding initial signal. In this case the original and degraded signals at the beginning and at the end will be equal. Now, what happens if window of quality measurement is displaced a little, say, by 1 ms? From human evaluation point of view (MOS) the result should be the same in both measurements. But PESQ algorithm gives a little different result in both cases. This is because of windowing nature of PESQ algorithm. We solved this problem by proposing signal extension method before PESQ measurement. It means that we still use PESQ, but the signal given to algorithm is constructed as shown in Fig. 3. This construction of composite signal $x_{\Sigma}(t)$ may be described using formula:

$$x_{\Sigma}(t) = \sum_{i=1}^{N-1} x_{\text{init}}(t - iT_{\text{init}}), \qquad (2)$$

where x_{init} is initial signal with 16/N ms silence added to the end. N is number of signal repetitions used and in our case it equals 8.

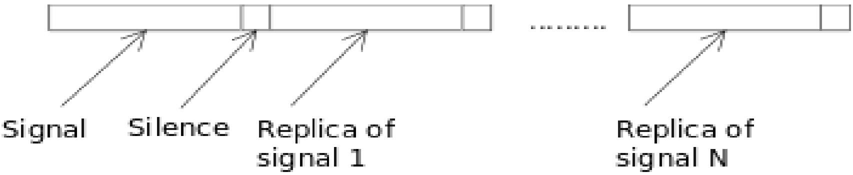

Fig. 3. Construction of extended signal

Precision of PESQ using such extended signal is much improved. The improvement may be seen in Fig. 4. In the picture is shown two curves – one corresponds to original PESQ measurements when measurement window is displaced and the other represent performance of PESQ with modified signal given for evaluation. It is easy to see that algorithm works much more consistently with modified signals.

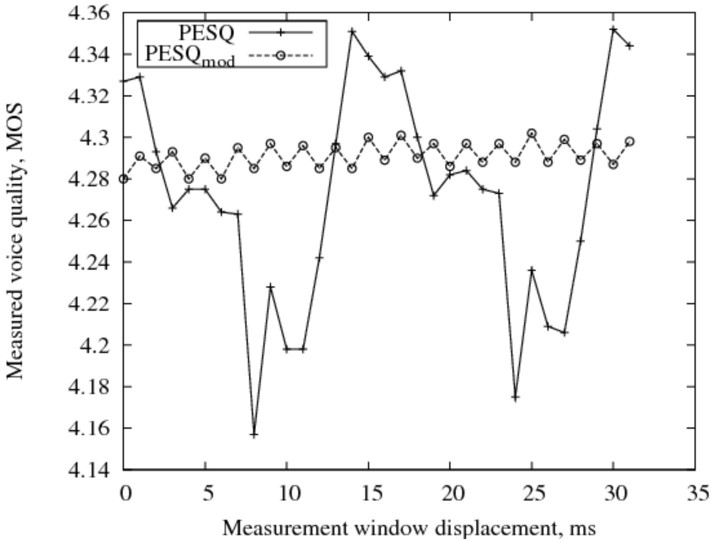

Fig. 4. Performance of original PESQ and PESQ with modified signal as an input

Statistical frame values for two speech types were determined. 2000 frames were used from records of fast speech (CNN news) and slow speech (interview about sports). Frame values were calculated for all these frames and histograms were constructed. Resulting histograms are shown in Fig. 5.

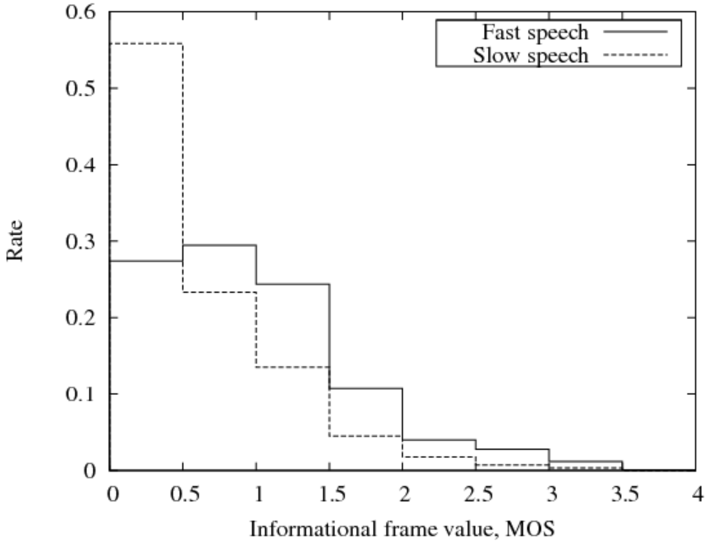

Fig. 5. Frame value distributions for fast and slow speech records

It can be seen that there are more low valued frames in slow speech and vice versa. Explanation for this lies in frame substitution algorithm which is used when frame is lost. When speech is slow, corresponding signal is more stationary in a sense and more neighbouring frames have similar frequency content. This results in better masking of frame error.

3. Analysis of the impact of voice microdistortions

In this chapter two phenomenons will be analysed. The first one is the impact of coding on quality for different AMR codecs (12.2 and 4.75 will be used). Second question is about the impact of lost frame on the signal following this frame.

It is frequently stated something like “AMR 12.2 codec gives average quality of 4.3“. Of course this statement is correct. But what does it hide under the word “average“? There is statistical distribution. Our purpose is to find this distribution for particular codec. Distribution will be calculated using very short sentences (0.5 s). Distortion for every record was calculated using two metrics – differential PESQ score and LPC disturbances. LPC disturbances are calculated for every 20 ms frame using formula:

$$D_i = \frac{1}{N}\sum_{n=1}^{N}\sqrt{\sum_{i=1}^{10}(A_i(n) - B_i(n))^2}\ , \tag{3}$$

where N is the number of frames in signal, $A_i(n)$ and $B_i(n)$ are LPC coefficients from n-th frame in original and distorted signals.

Full experiment for the evaluation of coding distortions may be described as follows:

- For every n take different signal of length 0.5 s.
- Code and decode this signal using AMR 12.2 and AMR 4.75 codecs.
- Calculate LPC distortion of coding for every signal using formula 3. Draw histogram from resulting data.
- Calculate PESQ distortion of coding for every signal. Draw histogram from resulting data.

Histograms showing LPC distortions are depicted in Fig. 6. Similar histograms for PESQ are shown in Fig. 7.

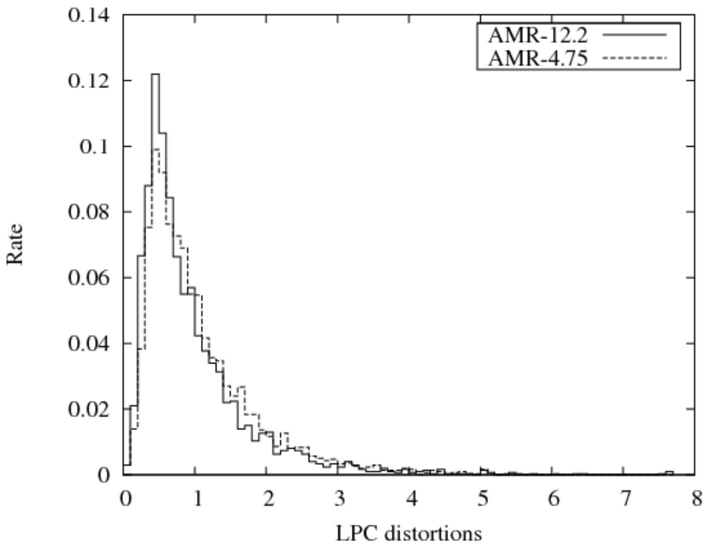

Fig. 6. Coding distortions of LPC coefficients

From figures it can be seen that LPC distortions are very small yet PESQ scores differs greatly between different codecs. It is easy to explain this phenomenon – PESQ algorithm uses psychoacoustic processing which raises separate distortions to a sixth degree. This discards all smaller distortions over the signal and amplifies extreme values. One other thing to note is quite a big spread of separate histograms. For example even AMR 12.2 codec may give quality of 1.5 MOS (4.5 minus dMOS value of 3) scores on some records. Such quality level is qualified as unacceptable usually. Though on average this codec is of excellent quality.

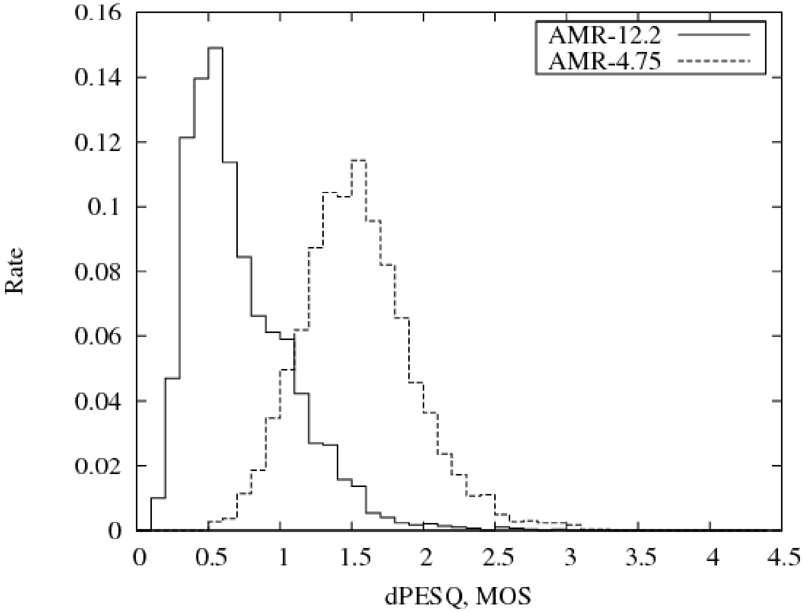

Fig. 7. Coding distortions as PESQ degradation

Frame loss distorts resulting signal over time. Our task here is determination of the character of this distortion. For this purpose experiment calculated LPC distortions in current (lost one) and subsequent frames. LPC distortion was calculated as Euclidean distance between frames. Average results of these distortions over time are presented in Fig. 8.

The largest distortion is observed in the frame directly after the lost frame. It is fundamental result of this research and it reflects statistical properties of the speech. This phenomenon may be explained by the fact that one frame after the loss and one frame before the loss (last good frame used for substitution) are statistically dissimilar. Also 40 ms is not long enough time for synchronization of decoders. Subsequent frames are less distorted because this synchronization takes place. The lost frame is distorted less (than the frame after loss) because it is statistically similar to the last good frame and substitution works statistically well.

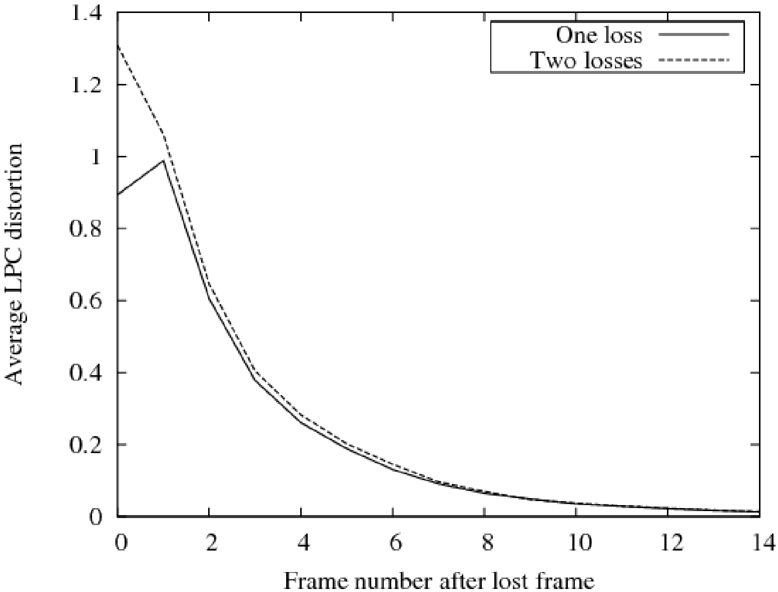

Fig. 8. Distortions over time after lost frame

Two lost frames in a row cause only a little bigger signal distortions over time. This shows that one frame loss in essence desynchronizes decoders almost to the highest degree possible.

4. Search of other means for frame value estimation

The method for frame value determination presented in chapter 3 requires a lot of computer time. This method has one more drawback – it can not be used in real time, because measurement should contain frame under investigation with its surroundings. In this chapter analysis of alternative less computationally complex methods will be presented. All presented methods will try to predict frame value using spectral feature vectors from current and previous frames. Prediction accuracy will be assessed using correlation coefficients between true and predicted frame values.

Linear prediction using spectral coefficients was the most successful from linear predictions. The highest correlation coefficient obtained using linear prediction was 0.31 with training data and 0.24 with testing data.

The most successful of tried prediction methods was the one using k-nearest neighbours' algorithm. This algorithm works by selecting *k* nearest neighbours from analysed point in independent variables. Algorithm output is an average from corresponding *k* dependent variables. In our experiments *k=10* was used. In our case independent variables are LPC coefficients calculated from training data. Dependent variables are corresponding frame values calculated from the same data. Algorithm output gives approximation of frame value given LPC coefficients of current and previous frames.

Scatter plot of the best prediction obtained during experimentation is shown in Fig. 9. This prediction was obtained using testing data. Correlation coefficient is 0.6. This is not particularly good result but is sufficient for many practical applications. One of such applications could be frame priority marking before sending the frame.

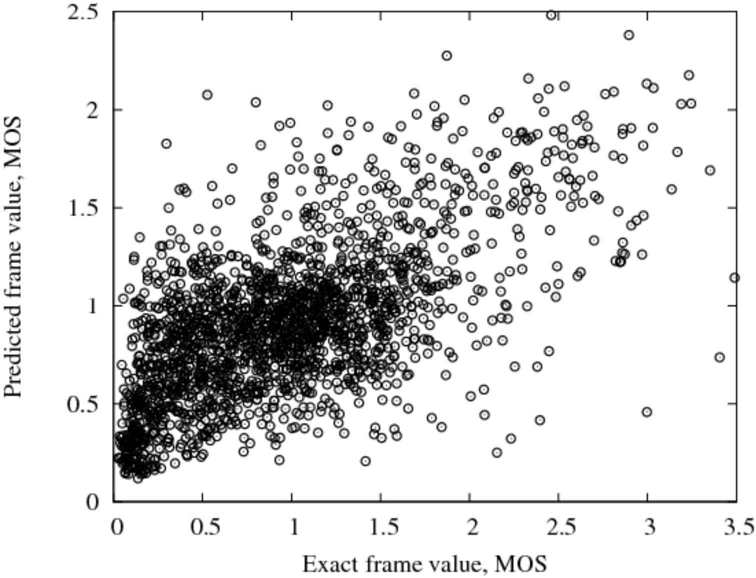

Fig. 9. Scatter plot of k-nearest neighbours' frame value prediction

General conclusions

1. It is proved in dissertation that PESQ algorithm described in ITU P.862 rec. is not well suited for evaluation of voice quality of short voice signals. It was proposed to solve this problem by composing signal from fragments of the original signal repeated *N* (usually 8 or 16) times. In this way the uncertainty of measurement results can be reduced on average 8 times.
2. The concept of frame value was proposed. Also the algorithms for determination of frame value are presented. The measurement means for frame value determination were created.
3. Distributions of frame values were calculated. These show the rate of important and not so important frames in typical speech.
4. Quality distortions resulting from voice coding were investigated. It was noted that distortion highly depends on the signal in use. AMR-12.2 codec creates quality degradation between 0.5 and 3.5 MOS points. It means that quality of codec should not be characterised using one number.
5. Voice quality variances for two popular codecs were calculated. It is proved that psychoacoustic distortions are much bigger than spectrum ones. It is shown that the biggest distortions fall to the frame immediately after the lost frame.
6. The linear and non linear algorithms for frame value prediction were created. They can work in real time. The best correlation coefficient achieved with exact frame values was 0.6. This precision is sufficient if the purpose is to choose better coding scheme for more important frames.

List of published works on the topic of the dissertation
In the reviewed scientific periodical publications

1. Kajackas, A.; Anskaitis, A. 2005. Investigation the ability of objective measures of the perceptual speech quality in mobile networks, *Electronics and Electrical Engineering* 7(63): 10–15.
2. Kajackas, A.; Anskaitis, A.; Guršnys, D. 2005. Individual Quality of Service Concept in Next Generations Telecommunications networks, *Electronics and Electrical Engineering* 4(60): 11–16.
3. Kajackas, A.; Anskaitis, A.; Guršnys, D. 2008. Peculiarities of Testing the Impact of Packet Loss on Voice Quality, *Electronics and Electrical Engineering* 2(82): 35–40.
4. Kajackas,A.; Anskaitis, A. 2009. An Investigation of the Perceptual Value of Voice Frames, *Informatica* (in print).

In the other editions

5. Kajackas, A.; Anskaitis, A.; Guršnys, D.; Pavilanskas, L. 2005. Estimation of QoS Dynamics in the Wireless Networks, in *The 4th COST 290 meeting. October 13–14*. Prieiga per internetą: <http://www.cost290.org/td2005/tds/td05050.pdf>.
6. Kajackas, A.; Anskaitis, A.; Guršnys, D. 2009. User Perceived Quality of Service, in *Traffic and QoS Management in Wireless Multimedia Networks: Cost 290 Final Report: Lecture Notes in Electrical Engineering*, 31: 125–128. Ed. by E. Koucheryavy *et al*.
7. Kajackas, A.; Anskaitis, A.; Guršnys, D. 2009. Estimating Individual QoS, in *Traffic and QoS Management in Wireless Multimedia Networks: Cost 290 Final Report: Lecture Notes in Electrical Engineering* 31: 180–183. Ed. by E. Koucheryavy *et al*.

About the author

Aurimas Anskaitis was born in Vilnius, on 24 September 1980.

Bachelor degree in Electrical Engineering and Electronics from Faculty of Electronics, Vilnius Gediminas Technical University, 2003. Master of Science in Electrical and Electronics Engineering from Faculty of Electronics, Vilnius Gediminas Technical University, 2005. In 2005–2009 – PhD student of Vilnius Gediminas Technical University. Assistant in Department of Telecommunication Engineering of Vilnius Gediminas Technical University at present.

KODUOTO BALSO KOKYBĖS TYRIMAS

Mokslo problemos aktualumas

Balso perdavimo sistemos nuolat aktyviai tobulinamos, diegiamos, ir plačiai naudojamos. Visose šiuolaikinėse sistemose balso signalas skaidomas į vienodos trukmės atkarpas, tos atkarpos koduojamos, suformuojami duomenų paketai ir perduodami. Ryšių kanaluose pasitaikantys trukdžiai sukuria klaidas, dėl kurių kai kurie paketai negali būti ištaisyti ir ištrinami. Tokie iškraipymai, kai prarandami paketai ypatingai būdingi radijo mobiliojo ryšio sistemoms. IP telefonijos sistemose pasitaikantys dideli paketų vėlinimai transformuojasi taip pat į paketų praradimus.

Kai perduodant duomenis paketai prarandami, tai perduodamo balso kokybė pablogėja bet kokioje telekomunikacinėje sistemoje.

Kuriant šiuolaikines telekomunikacijų sistemas prisilaikoma 1994 m. Tarptautinės telekomunikacijų sąjungos (ITU) E.800 rekomendacijoje suformuluotos paslaugų kokybės (QoS) sampratos, apibūdinamos, kaip vartotojų pasitenkinimą atspindintis paslaugos savybių bendras efektas. Prisilaikant šios sampratos, kuriamos telekomunikacijų sistemos, metodikos ir priemonės, kurios palaiko nustatytą paslaugų kokybės lygį, vertinant visų vartotojų paslaugų kokybės vidurkį. Reikia atkreipti dėmesį, kad paplitusios paslaugų kokybės vidutinių rodiklių vertinimo metodikos sukurtos fiksuotojo telefoninio ryšio sistemoms, kuriose visų vartotojų sąlygos yra panašios, visiems vartotojams teikiamų paslaugų kokybės rodikliai laikytini vienodais.

Mobiliojo ryšio tinkluose realiai vartotojui suteikiamų paslaugų kokybė nėra vienoda. Dėl vartotojo judrumo ir kintančių tinklo apkrovų ji nuolatos kinta ir ne visada, ir ne visur būna pakankamai aukšta.

Balso signalų savybių tyrimas seniai yra plėtojamas. Labai dažnai balso signalų savybės analizuojamos siejant jas su kalbos atpažinimo uždaviniais. Tačiau šiuolaikinėse sistemose, kai balso kodavimas taip paplitęs, iškyla keletas specifinių uždavinių. Kalbos savybių aprašymas tandeme su koderiu yra vienas iš uždavinių. Dar tiksliau – koderio įnešamų efektų balso signalams analizė. Siedami šias tyrimo kryptis su moderniomis telekomunikacijomis, gauname svarbų vartotojams ir operatoriams uždavinį – koduoto balso kokybės nustatymą.

Apie tai, kad vartotojų patiriama balso kokybė yra svarbus rodiklis kalba ne tik akademiniai sluoksniai. Didieji telekomunikacinių sistemų kūrėjai taip pat sprendžia uždavinius, susijusius su paslaugų kokybe: „Nokia“, „Nortel Networks“ ir kiti.

Iš pateiktų faktų matome, kad balso kokybės tyrimai yra plačiai vykdomi, jais vadovaujamasi kuriant telekomunikacijų sistemas, jas prižiūrint. Įvertinus koduoto balso perdavimų telekomunikacijose paplitimą, manome, kad darbas yra aktualus ir atliekamas laiku.

Tyrimų objektas

Darbo tyrimų objektas – koduoto balso kokybė.

Darbo tikslas ir uždaviniai

Darbo tikslas – ištirti galimybes ir sukurti priemones signalo mikrostruktūros pokyčių įtakai balso kokybei vertinti bei sukaupti informaciją apie koduotos kalbos segmentų informacinės vertės pasiskirstymus.

Suformuluotiems tikslams pasiekti darbe sprendžiami uždaviniai:

1. Sukurti matavimo priemonę trumpų balso signalo atkarpų kokybei vertinti.
2. Apibrėžti koduoto balso segmentų vertės sampratą ir parinkti vertės metrikas.
3. Išmatuoti bendrinės kalbos balso segmentų verčių skirstinius.
4. Nustatyti skirtingų koderių sukuriamų iškraipymų ribas.
5. Ištirti paplitusių koderių inertiškumą, nustatyti kiek laiko pastebima prarastų paketų įtaka sekantiems segmentams.
6. Ištirti galimybes balso paketo informacinę vertę prognozuoti realiu laiku.

Tyrimų metodika

Darbe taikomi statistikos, tiesinės algebros, signalų analizės metodai. Hipotezės tikrinamos sudarant reiškinių modelius ir atliekant simuliacijas.

Mokslinis naujumas

Rengiant disertaciją buvo gauti šie elektros ir elektronikos inžinerijos mokslui nauji rezultatai:

1. Sukurtas balso testams skirtas signalų sintezės būdas, išplečiantis galimybes taikyti PESQ algoritmą trumpų signalo atkarpų kokybei vertinti.
2. Suformuluota balso paketo informacinės vertės samprata. Pasiūlyta metodika ir sudaryti algoritmai balso paketo informacinei vertei skaičiuoti.
3. Nustatytos statistinės charakteristikos, parodančios, kaip balso paketo praradimas veikia signalą laike.
4. Sudaryti algoritmai perduodamo paketo vertei prognozuoti realiu laiku.

Praktinė vertė

Tyrimų rezultatai gali būti panaudojami kuriant ir optimizuojant naujų kartų balso perdavimo telekomunikacines sistemas. Rezultatai taip pat gali būti panaudojami galinio vartotojo įrangoje kuriant kokybės apskaitos modulius.

Ginamieji teiginiai

1. ITU P. 862 rekomendacija aprašytas PESQ algoritmas nėra tinkamas trumpų signalų kokybei vertinti.
2. Sukonstruotas signalų išplėtimo metodas leidžia taikyti PESQ algoritmą trumpiems signalams.

3. Paketo informacinę vertę galima nustatyti kaip skirtumą, tarp balso kokybių, kai paketas priimtas sėkmingai ir kai priimtas su klaida.
4. Paketo kokybės skirstinys yra nesimetriško varpo formos.
5. Praradus paketą, didžiausi iškraipymai stebimi pakete, nutolusiame nuo prarastojo per 20 ms.
6. Balso paketo vertę galima prognozuoti realiame laike. Gaunamas koreliacijos koeficientas su tikrąja paketo verte yra 0,6.

Darbo apimtis

Disertaciją sudaro įvadas, keturi skyriai ir rezultatų apibendrinimas.

Darbo apimtis yra 105 puslapių, neskaitant priedų, tekste panaudotos 31 numeruota formulė, 45 paveikslai ir 13 lentelių. Rašant disertaciją buvo panaudota 76 literatūros šaltiniai. Pirmajame skyriuje nagrinėjamas problemos aktualumas, formuluojamas darbo tikslas ir uždaviniai. Aptariami tyrimų metodai ir mokslinis darbo naujumas.

Įvade pristatomas darbo naujumas, aktualumas, aptariamas autoriaus indėlis, formuluojami darbo tikslai.

Pirmas skyrius yra bendro apžvalginio pobūdžio – apžvelgiami balso kokybės vertinimo metodai, jų privalumai ir trūkumai, aptariamas ITU rekomendacija P.862 įteisintas kokybės vertinimo PESQ algoritmas ir parodomos jo taikymų ribos. Kaip savarankiška dalis čia pristatyti autoriaus sudaryti sąrašai lietuviškų žodžių, skirtų kalbos suprantamumo tyrimams.

Antrame skyriuje parodoma, kaip galima išplėsti kokybės vertinimo PESQ algoritmo taikymo ribas. Čia įvedama koduoto balso paketo vertės sąvoka, nustatomi statistiniai paketų vertės skirstiniai.

Trečiame skyriuje nagrinėjami specifiniai koduotos kalbos iškraipymai ir kodavimo parametrų įtaka balso kokybei. Parodoma, kad kodavimo iškraipymų dydis priklauso nuo kalbos signalo ir kinta plačiose ribose. Skyriuje visapusiškai ištirta prarastų paketų įtaka.

Ketvirtame skyriuje nagrinėjamas paketo vertės nustatymo realiu laiku uždavinys. Pasiūlyti metodai viršutiniams paketo vertės nustatymo metodų tikslumo rėžiams nustatyti, sudaryti skaičiavimų atžvilgiu efektyvūs algoritmai, paketo vertei prognozuoti.

Bendrosios išvados

1. Pasiūlytas specialus tiriamųjų signalų sudarymo būdas, kuris leidžia trumpų balso atkarpų kokybės PESQ būdu vertinimo rezultatų sklaidą sumažinti vidutiniškai 8 kartus.

2. Sukurta balso segmentų vertės matavimo metodika ir matavimo priemonė, kurie sudaro galimybes įvertinti kiekvieno atskiro paketo potencialų indėlį į šnekos kokybę.
3. Sudaryti pavyzdiniai balso paketų vertės skirstiniai, parodantys kiek šnekoje yra vertingų ir kiek mažai svarbių paketų. Nustatyta, kad paketo vertės skirstinys yra asimetriškos formą.
4. Ištirti įvairiais koderiais koduotos šnekos iškraipymų dydžiai. Pastebėta, kad koderio perteikiamos šnekos kokybė labai priklauso nuo paties koduojamo signalo. Kokybės kitimo ribos, kai šneka koduojama AMR-12.2 koderiu, yra nuo 0,5 iki 3 MOS vienetų. Tai reiškia, kad bazinio kodavimo sukuriamo iškraipymo vienu skaičiumi apibūdinti negalima.
5. Nustatyta, kad girdimieji iškraipymai turi daug didesnes standartines deviacijas, nei spektriniai. Įrodyta, kad AMR koderiai paketo praradimo atveju labiausiai iškraipo po prarandamojo einantį balso segmentą.
6. Sukurti algoritmai paketo vertei prognozuoti pritaikant spektrinius požymius. Pasiektas paketo vertės prognozės koreliacijos koeficientas su tikrąja paketo verte yra 0,6. Toks tikslumas leidžia parinkti skirtingą kodavimo būdą skirtingos informacinės vertės balso paketams.

Trumpos žinios apie autorių

Aurimas Anskaitis gimė 1980 m. rugsėjo 24 d. Vilniuje.

2003 m. įgijo elektronikos inžinerijos bakalauro laipsnį Vilniaus Gedimino technikos universiteto Elektronikos fakultete. 2005 m. įgijo elektros ir elektronikos inžinerijos mokslo magistro laipsnį Vilniaus Gedimino technikos universiteto Elektronikos fakultete. 2005–2009 m. – Vilniaus Gedimino technikos universiteto doktorantas. Šiuo metu dirba asistentu Vilniaus Gedimino technikos universiteto Telekomunikacijų inžinerijos katedroje.

Aurimas ANSKAITIS

ANALYSIS OF QUALITY
OF CODED VOICE SIGNALS

Summary of Doctoral Dissertation
Technological Sciences, Electrical and Electronic Engineering (01T)

Aurimas ANSKAITIS

KODUOTO BALSO KOKYBĖS TYRIMAS

Daktaro disertacijos santrauka
Technologijos mokslai, elektros ir elektronikos inžinerija (01T)

2009 11 09. 1,5 sp. l. Tiražas 70 egz.
Vilniaus Gedimino technikos universiteto
leidykla „Technika“, Saulėtekio al. 11, LT-10223 Vilnius
http://leidykla.vgtu.lt
Spausdino UAB „Baltijos kopija“,
Kareivių g. 13B, 09109 Vilnius
http://www.kopija.lt