Paper

Communication Platform for Evaluation of Transmitted Speech Quality

Andrzej Ciarkowski and Andrzej Czyżewski

Faculty of Electronics, Telecommunications and Informatics, Gdańsk University of Technology, Gdańsk, Poland

Abstract—A voice communication system designed and implemented is described. The purpose of the presented platform was to enable a series of experiments related to the quality assessment of algorithms used in the coding and transmitting of speech. The system is equipped with tools for recording signals at each stage of processing, making it possible to subject them to subjective assessments by listening tests or, objective evaluation employing PESQ or PSQM algorithms. The functionality for the simulation of distortions typical for voice communication over the Internet was implemented, making it possible to obtain reproducible, quantifiable results. An application of the presented platform for evaluation of acoustic echo canceler algorithm based on watermarking techniques, which was developed earlier is presented as an example of an effective deployment of the described technology.

Keywords—acoustic echo cancelation, doubletalk detection, echo-hiding.

1. Introduction

Development process of new algorithms for coding and improving the quality of transmitted speech entails the need to submit the results to assessments, making possible for the author of the algorithm to observe the introduced changes effect on the speech signal quality. An essential element of the assessment procedure is reproducibility of the results, which is often not feasible when working with active, "live" communication system. An obstacle in obtaining reproducible results is typically the element of randomness associated with a variable system load, choice of routes of communication and many other factors whose impact can be minimized or neglected by creating a separate, isolated platform for research and evaluation purposes only. Another need is the ability to simulate certain phenomena that typically occur randomly in communication systems, also assuming certain quantitative or qualitative parameters. This paper summarizes the design and implementation of such a system, which was conducted by the authors.

The implemented platform was practically utilized during the evaluation of the novel acoustic echo canceler (AEC) algorithm based on semi-fragile watermarking techniques, which was introduced by the authors [1], [2]. Obtained results are presented in the final part of this paper as a proof of usability of the developed system.

2. Platform Description

The developed communication platform is based on the use of typical elements, common in Internet telephony (VoIP) implementations, but extends them with some additional tools for collecting measurement data, including recording signals at the various intermediate stages of processing. An important aspect of the developed system is the automation of the results collecting, which is possible by using non-interactive execution mode. This allows to create scripts easily which enable obtaining a series of results depending on the specific parameter values. On the other hand, the interactive mode, allowing the user to manipulate UI elements facilitates the single passage, quick measurements as well as checking the behavior of algorithms while certain parameter changes over time. For this reason, the described system consists of two applications based on common software libraries, but differing with regard to the method of interaction.

Both applications are in fact the VoIP clients (terminals) communicating via standard RTP protocol and incorporating a complete communication stack thereof. The used implementation was conceived entirely at the Gdańsk University of Technology (GUT) and constitutes a fully functional realization of the RTP specification [3], together with a number of additional extensions and profiles. The entire source code associated with the implementation of the system was prepared in C++ programming language, with the support of numerous open source libraries for enhanced portability. Thus, although the primary work environment of the authors is Microsoft Windows OS, the developed libraries and non-interactive applications can be easily run on other operating systems, including Linux and MacOS X.

Figure 1 shows the architectural elements of the system. Four main groups of blocks can be identified, corresponding to consecutive stages of speech signal processing in the path from I/O interface to the transport layer. The input/output stage lists the three possible types of signal path "endpoints". In the case of the online analysis it is possible to use a real audio interface (sound card); the user of the system provides a test signal through the microphone and it gets possible to rehearse the result immediately. The signal can also be read from an audio file, what allows for a series of experiments using the same signal and differ-

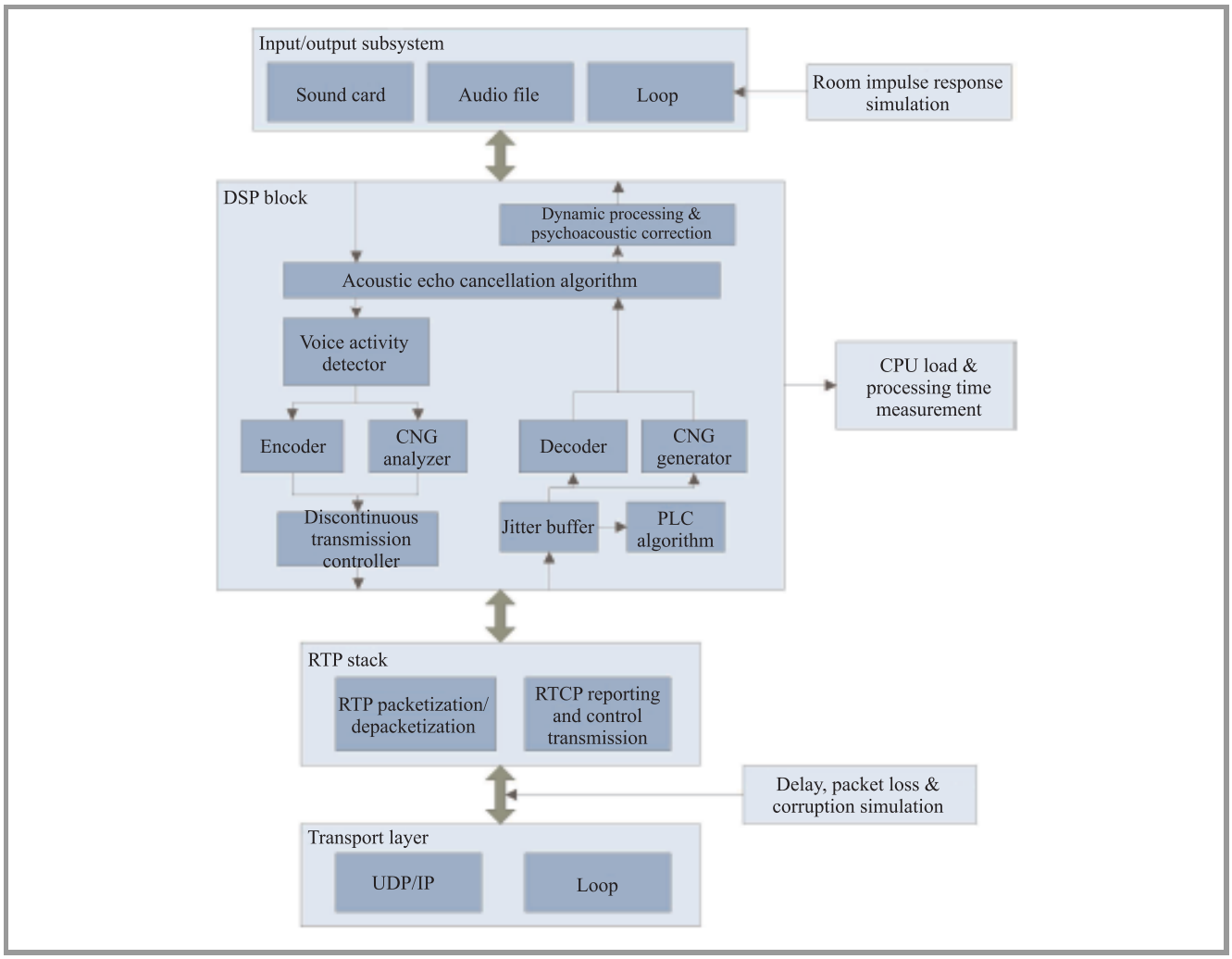

Fig. 1. Elements of communication platform for monitoring and evaluating voice transmission quality.

ent parameters of algorithms. Finally, it is also possible to use the local loop, optionally equipped with a filter that implements a specified transient response, which is particularly applicable in the case of testing algorithms for echo cancelation, because it allows simulating echoes with some specific properties using the remote terminal. It should be noted that the choice of sound file or local loop enables the analysis to be performed in the offline mode (for non-interactive applications), because it is not necessary in this situation to synchronize to periodically arriving data packets. This feature allows shortening the analysis time considerably, providing a valuable feature in case of large data series processing.

The next group of blocks makes the most important part of the system, namely the digital signal processing path. It is important to emphasize that the various elements of this group are in fact the algorithms which are subject to evaluation within the system, so that a special attention was paid to designing interfaces in such a way that blocks performing the same function using different algorithms are easily replaceable. The signal processing path is asymmetrical, with the flow oriented "towards the network" and "from the network" being different. The flow "towards the network" begins with the acoustic echo cancelation (AEC) algorithm block, which in this section shall record the signal coming from the near-end-speaker and is supposed to remove the estimated echo signal. At this stage, the system allows detailed analysis of the results of operations and performance of the following AEC algorithms:

- algorithm based on Geigel double-talk detector (DTD) and NLMS adaptive filter [4], [5];
- algorithm available in the open source speex voice codec library [6];
- algorithm based on semi-fragile watermarking DTD and NLMS adaptive filter, developed by the authors [1], [2];
- algorithm based on normalized cross-correlation DTD (NCC) and NLMS adaptive filter [7].

Another element in the processing path is the voice activity detector (VAD), which is used in case the employed speech codec lacks this feature. Its task is to determine whether the currently processed block of data has the characteristics

of voice activity, therefore, whether it is desirable to switch the system to the comfort noise encoding mode (CN). This technique is commonly employed in VoIP systems for bandwidth savings, especially in connection with so-called discontinuous transmission (DTX) mode, involving suppression of the transmission of packets representing the noise with characteristics similar to the memorized state.

The packet classified as active voice is passed to the speech encoder. The choice of audio coding algorithm is determined by the parameters of the application; in the case of interactive applications the codec can be changed during the session. At the present the system supports the following voice coding algorithms:

- PCM with a resolution of 8 and 16 bits/sample,
- ITU-T G.711 A-law and μ-Law [8],
- ITU-T G.722 [9],
- ITU-T G.723.1 [10],
- ITU-T G.726 (in versions 16, 32, 40 and 24 kbit/s) [11],
- ITU-T G.729 (with annexes B, D and E) [12];
- IMA ADPCM (DVI4) [13],
- ETSI GSM 06.10 [14],
- Speex [6],
- Internet low bitrate codec (iLBC, RFC3951) [15].

Figure 2 shows an example screen from the interactive application containing a list of codecs available in the current audio path configuration.

The last block of "towards the network" flow is aforementioned discontinuous transmission controller; whose function is to suppress transmission of the encoded packet in response to a signal from the VAD algorithm, or the sole codec, provided it supports that feature (e.g., G.729B, Speex).

The signal processing path in the direction "from the network" begins with the anti-jitter buffer algorithm. At this stage adaptive-length jitter-buffer implemented in the SpeexDsp library may be used interchangeably with the generic algorithm developed at the GUT. It cooperates with the packet loss concealment (PLC) algorithm, whose purpose is to recover (or to interpolate) packets which were lost during the transmission. The system provides a functionality for simulation of packet loss at a preconfigured ratio, which allows for the evaluation of the quality of speech transmission in a lossy environment. The simulator accepts the probability of packet loss as an input parameter, and the reproducibility of the loss pattern is achieved through the use of a dedicated MLS pseudo-random number generator with user-supplied seed. At this stage, the PLC algorithm described by the ITU-T G.711 Appendix I recommendation has been implemented, as well as several algorithms built-into some speech codecs [16].

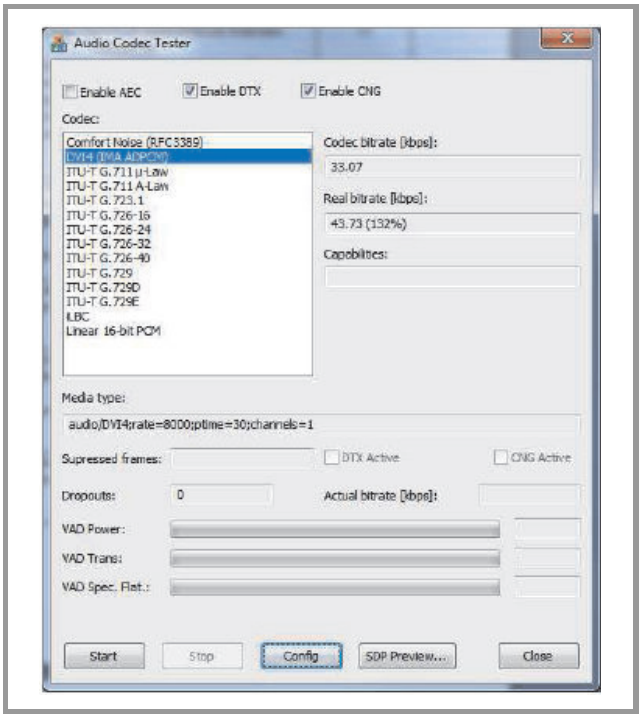

Fig. 2. Application interface allowing interactive selection of audio coding algorithm during the session.

Packets leaving the jitter-buffer are subjected to decoding and the decoded speech signal is delivered to the AEC algorithm, which interprets it as a model of the signal coming from the far-end speaker. Before returning to the output device, the postprocessing is performed, which includes the correction of the dynamics provided by a programmable noise gate, expander, compressor, and limiter blocks.

In the subsequent step the frames of speech signal are fed into the encoder, which produces payload according to the profile corresponding to the applied RTP audio codec. The encoded payload is passed to the RTP stack in order to append the RTP header. This process is called packetization, and the analogous operation "isolating" the payload of the RTP packet received from the transport layer is called depacketization. During depacketization the data obtained from the transport layer is reviewed for accuracy and continuity of the timestamp and sequence number, which allows the detection of loss of the packet, its repetition or change the order. The additional role of RTP stack package is to identify the sender, discarding packets received outside the current session, whose presence may indicate an attack attempt. Also certain statistical measurements are carried out, such as estimation of round-trip delays, delay fluctuation (jitter), packet loss ratio. These data are collected for the control of communication within the RTP session, which is carried out using the RTCP protocol.

RTP transport layer is typically based on sending UDP datagrams over an IP network. In many cases, however, the desired behavior is to work in a "loop", then sent datagrams are transmitted immediately back to the RTP stack. The developed system supports both of these modes. In

UDP/IP mode the system terminal may communicate not only with identical terminal, but also with any RTP client equipped with compatible codecs. This allows the system to use the endpoint for the analysis of data obtained from external applications, such as the popular streaming server VLC. The use of the loop mode allows for a convenient evaluation of algorithms, whose behavior is not dependent on using a distributed configuration. An important complement to the system are the aforementioned packet loss simulator and a "delay line" generating a programmed delay of the packet arrival, useful in a research of acoustic echo cancelation and buffering. The purpose of this delay block is the simulation of round-trip delays introduced by the network, which do not occur in the "loopback". These delays, which can range from single milliseconds up to seconds, are typically responsible for the perceived annoyance of the acoustic echo affecting quality of VoIP calls.

3. Application to Acoustic Echo Cancellation Evaluation

3.1. Robustness Analysis of AEC Algorithms under Time-Variable Echo Conditions

A standard, widely-accepted in the literature method of objective testing, based on the detection theory, was used during the study. This method is based on plotting the receiver operating characteristic (ROC) curves representing the probability of false alarm and miss as a function of relative signal levels of the near- and far-end speakers (NFR, near-to-far ratio). Its detailed description can be found in the literature [17]. During the research a modification of the method was proposed, whose purpose was the simulation of time-varying echo conditions (time-variable echo delay, changing room impulse response).

The evaluation was based on 5 speech excerpts in Polish. 4 recordings of length 1s (2 women, 2 men) represented the near-end speakers, and the recording of the length of 5 seconds (male voice) served as a far-end speaker signal. Fragments were sampled at the frequency of 8 kHz, consistent with common telephony applications. During the evaluation a constant echo delay of 40 ms was applied, with variable component added according to the characteristic plotted in Fig. 3.

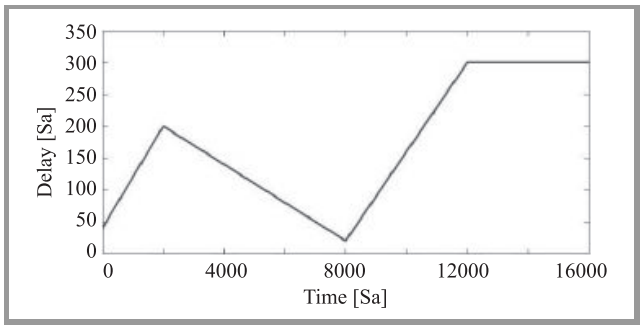

Fig. 3. Characteristics of variation of echo delay time.

The simulation of variable room impulse response involved a weighted sum of 2 impulse responses which were acquired in different locations at the same room. The variation was a simple linear transition from $h_1(n)$ to $h_2(n)$ over the time span of 4 s. The impulse responses are presented in Fig. 4.

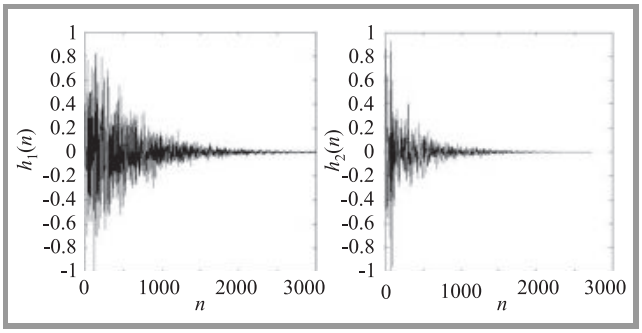

Fig. 4. Acoustic path impulse responses $h_1(n)$ and $h_2(n)$ for the research.

For maintaining consistency with the results presented in the literature the evaluation was conducted for the probability of false alarm $P_f \in \{0.1, 0.3\}$. Robustness of AEC methods based on different DTD algorithms was evaluated through the execution of the same test, first with the constant echo time and the impulse response, and then, with variable ones. Increased probability of DTD algorithm miss (i.e., not detecting the doubletalk when it is present) in these conditions determines the susceptibility of the DTD algorithm to the variability the echo path characteristics. The results obtained are presented in Fig. 5 and in Fig. 6.

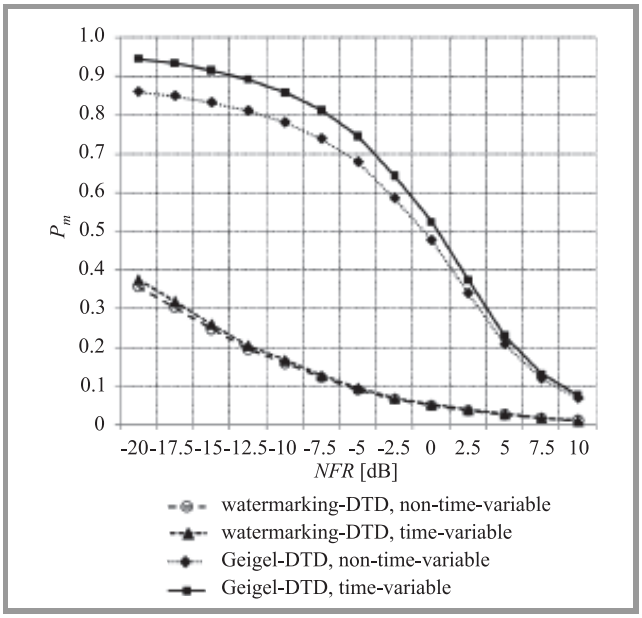

Fig. 5. Probability of DTD algorithm miss for the probability of false alarm $P_f = 0.1$ while running in the variable and "stable"conditions.

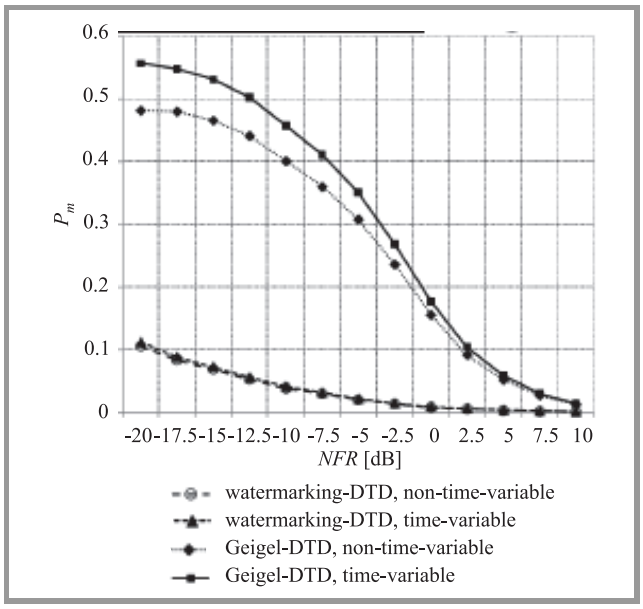

Fig. 6. Probability of DTD algorithm miss for the probability of false alarm $P_f = 0.3$ while running in the variable and "stable"conditions.

Both evaluated algorithms demonstrated a performance deterioration in the "variable" conditions, however, the scale of this phenomenon is different. For comparison, a relative deterioration measure (RDM) was derived which determines how the algorithm performs in "variable"conditions relating to the "stable" ones.

$$\mathrm{RDM} = \frac{P_{m,variable}}{P_{m,stable}} \tag{1}$$

This coefficient values were plotted in Fig. 7.

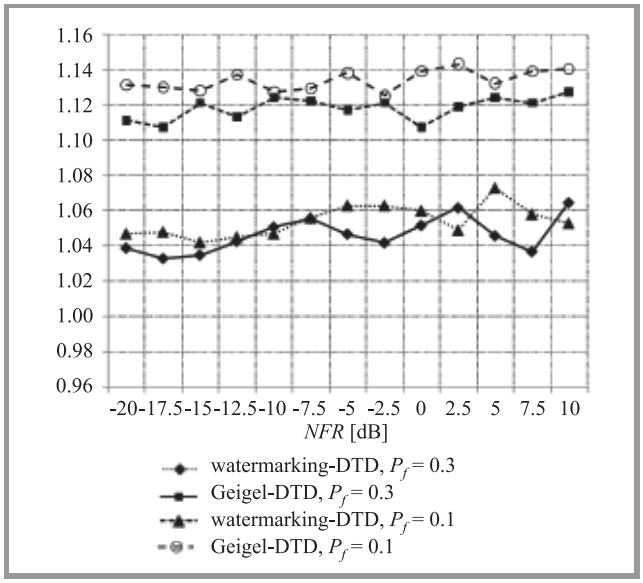

Fig. 7. The relative increase in the DTD algorithm probability of miss while working under time-variable echo delay and changing room impulse response.

3.2. Objective and Subjective Evaluation of Watermarking-Based DTD Algorithm in Relation to NCC Algorithm

The implementation of the normalized cross-correlation DTD algorithm made it possible to conduct a comprehensive evaluation of DTD algorithm developed by the authors against the algorithm representing current state of the art. In the first phase of evaluation, the objective tests were carried out in accordance to the methodology proposed in the literature [17]. The test set used was the same as for the previously presented robustness analysis. Consequently, the listening tests were conducted to investigate as to how DTD misses made by the various algorithms affect the subjective opinion on speech quality.

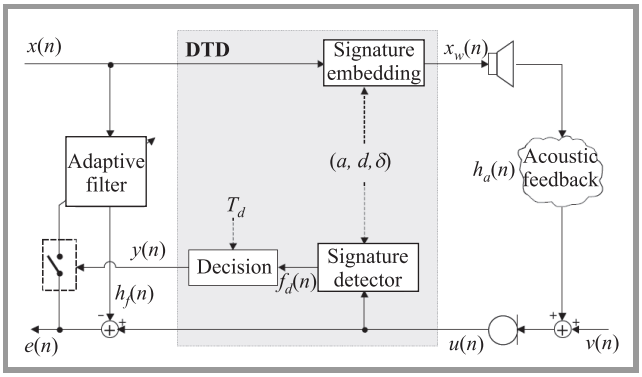

Fig. 8. Acoustic echo cancelation system employing watermarking-based DTD algorithm and adaptive filter.

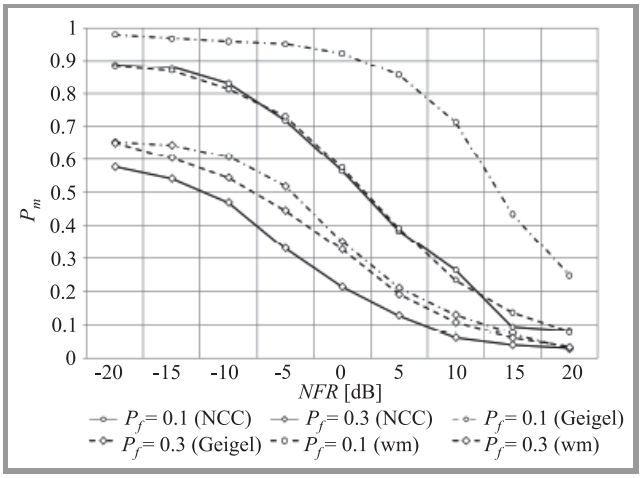

Fig. 9. Probability of DTD algorithm miss in the presence of background noise of –30 dB.

Both DTD algorithms were combined with the NLMS adaptive filter of length $L = 512$ to create a working AEC system during the tests. The example setup of such system with watermarking-based DTD algorithm is depicted in Fig. 8 and for the NCC algorithm only the grayed box labeled DTD is different. The length of the window used by the NCC algorithm to estimate the correlation coefficients between $x(n)$ and $u(n)$ was $W = 500$.

Both of these values were chosen in consistency with the research published in the literature by the authors of the NCC algorithm [7], [18], [19]. The results of objective tests carried out in the first phase are presented in Fig. 9 and Fig. 10.

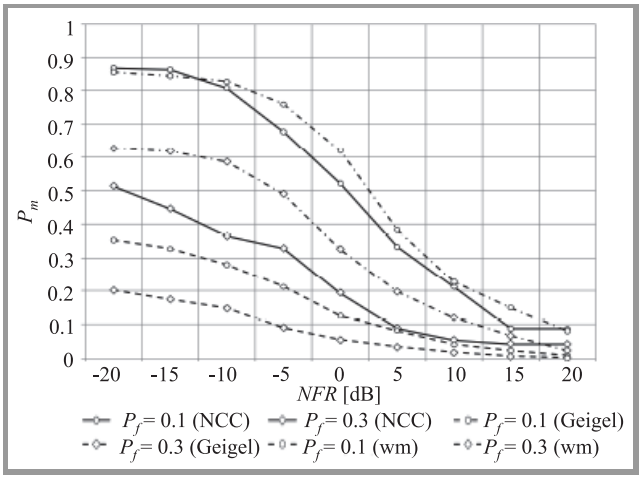

Fig. 10. Probability of DTD algorithm miss in the presence of background noise of –60 dB.

The presented plots were obtained at different levels of background noise in the microphone signal. This allowed assessing the vulnerability of specific algorithms for this type of disturbance. The resulting graphs for the NCC algorithm significantly differ from those published by its authors. This discrepancy may stem from the fact that in the literature description [17] the algorithm has been combined with a real adaptive filter, but the studies were based on the use of the actual impulse response, which was previously used to simulate the echo signal. Therefore, experiments carried out using the developed system are able to model the actual phone call conditions in a more realistic way.

Table 1
MOS values for DTD algorithms; background noise level -30 dB, NFR $= 0$, $P_f = 0.1$

Test signal	MOS
Reference signal (near speaker)	4.83
Reference signal (microphone signal)	1.25
AEC algorithm w/ DTD NCC	3.92
AEC algorithm w/ Geigel DTD	2.58
AEC algorithm w/ watermarking DTD	3.75

Table 2
MOS values for DTD algorithms; background noise level -60 dB, NFR $= 0$, $P_f = 0.1$

Test signal	MOS
Reference signal (near speaker)	4.92
Reference signal (microphone signal)	1.25
AEC algorithm w/ DTD NCC	3.75
AEC algorithm w/ Geigel DTD	3.08
AEC algorithm w/ watermarking DTD	4.33

Table 3
MOS values for DTD algorithms; background noise level -30 dB, NFR $= -15dB$, $P_f = 0.1$

Test signal	MOS
Reference signal (near speaker)	4.75
Reference signal (microphone signal)	1.16
AEC algorithm w/ DTD NCC	2.0
AEC algorithm w/ Geigel DTD	1.25
AEC algorithm w/ watermarking DTD	1.84

Table 4
MOS values for DTD algorithms; background noise level -60 dB, NFR $= -15dB$, $P_f = 0.1$

Test signal	MOS
Reference signal (near speaker)	4.83
Reference signal (microphone signal)	1.16
AEC algorithm w/ DTD NCC	1.84
AEC algorithm w/ Geigel DTD	1.84
AEC algorithm w/ watermarking DTD	3.75

Mean opinion score (MOS) values were obtained in effect of listening tests based on the sound files stored during the objective tests phase, therefore both tests were performed using identical test signals. MOS values presented in Tables 1–4 are the mean of the ratings issued by 12 experts (Ph.D. students and employees of the GUT, Multimedia Systems Department).

4. Summary

The developed system provides a useful tool for comprehensive analysis of various aspects of the voice coding, transmission and quality enhancement algorithms. It has been designed and implemented during the research work, which sought to develop new algorithms for coding and improving the quality of speech transmitted over the Internet. Currently available functionality of the system provides a significant facilitation of the research process, what was practically demonstrated by the results of the evaluation of watermarking-based DTD algorithm proposed and developed by the authors.

Acknowledgement

The research was funded by the Polish Ministry of Science and Higher Education within the grant no. PBZ-MNiSW-02/II/2007.

References

[1] G. Szwoch, A. Czyżewski and A. Ciarkowski, "A double-talk detector using watermarking", *J. Audio Eng. Soc.*, vol. 57, pp. 916–926, 2009.

[2] A. Czyżewski and G. Szwoch, "Method and Apparatus for Acoustic Echo Cancellation in VoIP Terminal". International Patent Application No. PCT/PL2008/000048, 2008.

[3] H. Schulzrinne, et al, "RTP: A Transport Protocol for Real-Time Applications", RFC 3550, IETF, 2003.

[4] D. L. Duttweiler, "A twelve-channel digital echo canceler", *IEEE Trans. Commun.*, vol. 26, pp. 647–653, 1978.

[5] S. M. Kuo, B. H. Lee and W. Tian, "Adaptive echo cancelation", in *Real-Time Digital Signal Processing: Implementations and Applications* (2nd ed), S. M. Kuo *et al.*, Eds. Chichester: Wiley, 2006, pp. 443–473.

[6] J.-M. Valin, "The Speex Codec Manual", May 09, 2011, http://speex.org/docs/manual/speex-manual/

[7] J. Benesty, D. R. Morgan, J. H. Cho, "A new class of doubletalk detectors based on cross-correlation", *IEEE Trans. Speech Audio Process.*, vol. 8, pp. 168–172, 2000.

[8] "Pulse code modulation (PCM) of voice frequencies". ITU-T Recommendation G.711 (11/88), Int. Telecomm. Union, Geneva, Switzerland, 1988.

[9] "7 kHz audio-coding within 64 kbit/s". ITU-T Recommendation G.722 (11/88), Int. Telecomm. Union, Geneva, Switzerland, 1988.

[10] "Dual rate speech coder for multimedia communications transmitting at 5.3 and 6.3 kbit/s". ITU-T Recommendation G.723.1 (05/06), Int. Telecc. Union, Geneva, Switzerland, 2006.

[11] "40, 32, 24, 16 kbit/s Adaptive Differential Pulse Code Modulation (ADPCM)". ITU-T Recommendation G.726 (12/90), Int. Telecomm. Union, Geneva, Switzerland, 1990.

[12] "Coding of speech at 8 kbit/s using conjugate-structure algebraic-code-excited linear prediction (CS-ACELP)". ITU-T Recommendation G.729 (01/07), Int. Telecomm. Union, Geneva, Switzerland, 2007.

[13] H. Schulzrinne and S. Casner, "RTP Profile for Audio and Video Conferences with Minimal Control", RFC 3551, IETF, 2003.

[14] "GSM Full Rate Speech Transcoding". Europ. Telecomm. Standards Inst. (ETSI), Sophia Antipolis, France, Recommendation GSM 06.10, 1992.

[15] S. Andersen *et al.*, "Internet Low Bit Rate Codec (iLBC)", RFC 3951, IETF, 2004.

[16] "A High Quality Low Complexity Algorithm for Packet Loss Concealment with G.711". ITU-T Recommendation G.711 Appendix I (09/99), Int. Telecomm. Union, Geneva, Switzerland, 1999.

[17] J. H. Cho, D. R. Morgan and J. Benesty, "An objective technique for evaluating doubletalk detectors in acoustic echo cancelers", *IEEE Trans. Speech Audio Process.*, vol. 7, pp. 718–724, 1999.

[18] S. L. Gay and J. Benesty, "An introduction to acoustic echo and noise control" in *Acoustic Signal Processing for Telecommunication*, S. L. Gay and J. Benesty, Eds. Norwell: Kluwer, 2000, pp. 1–18.

[19] J. Benesty *et al.*, *Advances in Network and Acoustic Echo Cancellatation*. Berlin: Springer, 2001.

[20] M. Baughner *et al.*, "The Secure Real-time Transport Protocol (SRTP)". RFC 3711, IETF, 2004.

Andrzej Ciarkowski was born in 1979 in Gdańsk. In 1998–2003 he studied at the Gdańsk University of Technology, where in 2003 he graduated at the Sound Engineering Department. His thesis was related to design of custom, high quality USB audio interface. Since that time he has been a member of the research staff at the Multimedia Systems Department as a Ph.D. student (2003–2008) and is currently working on Ph.D. thesis. Current subjects of his research includes multimedia communications over Internet Protocol, what is reflected by the subject of his Ph.D. thesis, relating to acoustic echo cancelation in VoIP systems.
E-mail: rabban@sound.eti.pg.gda.pl
Multimedia Systems Department
Gdańsk University of Technology
Narutowicza st 11/12
80-233 Gdańsk, Poland

Andrzej Czyżewski is the Head of the Multimedia Systems Department of the Gdańsk University of Technology at the Faculty of Electronics, Telecommunications and Informatics. He received his M.Sc. degree in Sound Engineering from the Gdańsk University of Technology in 1982, his Ph.D. degree in this domain in 1987 and his D.Sc. degree in 1992 from the Cracov Academy of Mining and Metallurgy. In December 1999 Mr. President of Poland granted him the title of Professor. In 2002 the Senate of the Gdańsk University of Technology approved him to the position of Full Professor. He is a member of the Acoustic Committee of the Polish Academy of Sciences, IEEE, International Rough Set Society and Fellow of the Audio Engineering Society. As a researcher, together with his team designed a number of software applications and digital devices; several of which were produced commercially in Poland. The subjects of the mentioned projects concern new methods of speech recognition, audio restoration, beamformers, anti-noise filters, speech communication system for military aircraft pilots, environmental noise monitoring system, advanced surveillance monitoring systems and others.
E-mail: andcz@sound.eti.pg.gda.pl
Multimedia Systems Department
Gdańsk University of Technology
Narutowicza st 11/12
80-233 Gdańsk, Poland