Elektrotehniški vestnik 76(5): 311-317, 2009
Electrotechnical Review: Ljubljana, Slovenija

Impact of the Echo Canceller and VAD System on Data Transmission over the GSM System Voice Channel

Zdenko Mezgec[1], Amor Chowdhury[1], Rajko Svečko[2] and Bojan Kotnik[3]

[1] *Margento R&D d.o.o., Gosposvetska cesta 84, 2000 Maribor*
[2] *Univerza v Mariboru, Fakulteta za elektrotehniko, računalništvo in informatiko, Smetanova 17, 2000 Maribor*
[3] *Ultra d.o.o., Cesta Otona Zupančiča 23A, 1410, Zagorje ob Savi*
E-pošta: zdenko.mezgec@margento.com

Abstract. The presented mobile payment system uses the GSM speech channel for data transmission. The speech channel is optimized for human speech transmission and, therefore, the transmission of modulated data is affected by various factors. The Echo Canceller and VAD systems are the factors having the greatest impact on performance of data transmission over the voice channel. For mobile phone payment, it is important for tempo-spectral characteristics of transmitted modulation signals to be similar to those of the human speech. Otherwise, these signals can be interpreted as nonspeech and blocked by the echo canceller or VAD. In this work, we investigated the impact of ETSI VAD (European Telecommunication Standards Institute - Voice Activity Detector) and Echo Canceller system on different types of test signals. By implementing the ETSI VAD system in the MATLAB workspace and various tests performed with the mobile phones and different types of test signals, determined the causes for unpredictable changes in the quality of audio data transmission. At the end of this work, some proposals for avoiding these factors and establishing a robust quality of data transmissions will be given.

Keywords: GSM system, echo cancellation, voice activity detection, data over voice, data transmission

Vpliv izničevalnika odbojev in modula za zaznavanje aktivnosti govora na prenos podatkov po govornem kanalu sistema GSM

Povzetek. Predstavljeni sistem mobilnega plačevanja temelji na prenosu podatkov prek govornega kanala sistema GSM. Govorni kanal sistema GSM je namenjen predvsem za prenos govora, zato pri prenosu zvokovno moduliranih podatkov zasledimo močan negativen vpliv izničevalnika odbojev in stopnje za zaznavanje aktivnosti govora. Za uspešno izvedbo predlaganega koncepta mobilnega plačevanja potrebujemo takšen modulacijski postopek, da bo rezultirajoči zvokovno modulirani signal čim bolj podoben časovno-spektralnim in dinamičnim karakteristikam govornega signala. V članku predstavljamo rezultate simulacije in analize odzivov standardiziranega modula za zaznavanje aktivnosti govora (ETSI VAD) na različne vhodne testne signale. Na podlagi teh analiz smo tako določili nabor priporočil za izvedbo optimalnega postopka zvokovne modulacije v smislu neobčutljivosti na neželene vplive izničevalnika odbojev in stopnje za zaznavanje aktivnosti govora,

Ključne besede: system GSM, izničevalnik odbojev, detekcija aktivnosti govora, zvokovno modulirani podatki, prenos podatkov

1 Introduction

The GSM mobile phone is an indispensable device in everyday life. Its basic function is to enable a remote voice communication. People living in the modern information technology era often come across various services such as mobile trading, mobile business, and mobile payments. All of these services usually use different interfaces, payment instruments, and/or data systems for data transmission. One of the numerous types of data transmissions is the voice-modulated data transmission over the GSM (Global System for Mobile) voice channel [11]. This principle is already incorporated in the internationally patented Margento mobile payment system, developed by Ultra, d.o.o., and Margento R&D d.o.o. [1]. The basis of the Margento system is voice-modulated data transmission between the Margento centre and the payment terminal (see Fig. 1).

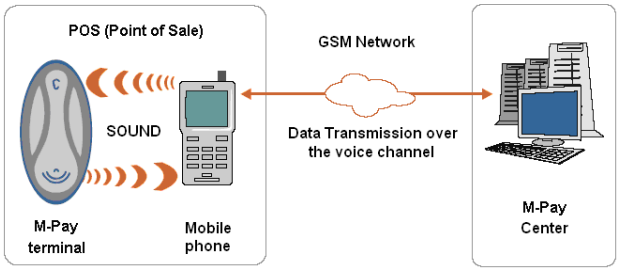

Figure 1. Concept of the Margento system.

The Margento terminal is a device which enables the usage of user's GSM mobile phone as a universal payment instrument. It works similarly as the ordinary

Received 12 August 2009
Accepted 18 November 2009

Figure 2.VEND terminal (left), and MPOS terminal (right).

POS (Point of Sale) terminal, where the buyer introduces his/ her credit card to perform a payment. There are two types of the Margento terminals on the market; the MPOS and VEND terminals, which are intended for different applications (Fig. 2).

Data transmission over the GSM voice channel is a subject to various disturbances. The main reason for this is because the GSM voice channel is not intended to transmit arbitrary acoustical signals with exception of the speech signal itself [2], [3]. Little has been done to avoid these problems because of lack of competition and the patent protection afforded to the Margento [1].

The intention of this paper is to highlight some problems and solutions for improvement of the voice modulated data transmission over the GSM voice channel. In the first part of the paper, some problems which cause certain difficulties with the data transmission are introduced. Section 3 depicts the analysis and the ETSI VAD system implementation in the MALTAB programme environment. Furthermore, there are some results and system impacts shown in Section 4. In the conclusion, some solutions are suggested to increase performance of the data transmission over the GSM voice channel.

2 Description of the problem

The Margento terminal is a device which simultaneously modulates and demodulates data signals. The data communication between the Margento Centre and the Margento Terminal can be performed either in the full duplex mode, which increases the data transmission bit rate, or with the less effective and slower half duplex. The transmission quality of the audio modulated data is limited due to the following impacts and interferences:

- The impact of the Margento terminal audio coupling and surrounding acoustic environment,
- GSM system impact (speech codecs, compression, packet loss), and
- Mobile phone impact (VAD, echo canceller).

2.1 The Margento Terminal and its Surrounding Impact

Modulated data transmission runs in an unsatisfactory acoustic surrounding, because of the different shapes of the mobile phones, thus lowering the transmission quality. The consequences of the unsatisfactory acoustic connection between the Margento terminal and the mobile phone (Fig. 3) are additional signals disturbing the terminal and mobile phone communication. The terminal works in noisy surroundings, such as restaurants, shops, etc. It could happen in some areas that the surrounding signal, which is added to the modulated signal, is stronger than the modulated signal (SNR<0dB). Because of this and some surrounding noise, the VAD of the GSM phone can block the modulated signal. The terminal microphone also receives a deformed signal. In addition to the modulated signal sent by the Margento centre, the microphone also receives the surrounding noise and the noise broadcasted by the speaker inside the terminal. Likewise, this occurrence happens for the modulated signal sent to the Margento centre. The surrounding noise can thus be stronger than the modulated signal. The MPOS and VEND terminal hardware equipment are very much alike (see Fig. 2). The only difference is that the VEND terminal does not have the support for the communication between the human and the terminal (keyboard, screen, printer, etc.). Both types of terminals have the Texas Instruments 32-bit TMS320F2812 DSP processor, with the working core running on 120 MHz.

2.2 GSM System Impact

One of the most important impacts of the GSM system is the voice coder, developed and intended primary for speech transmission, and not for the transmission of arbitrary modulation signals. The GSM speech channel is adjusted to the human speech, which has very specific features. In the GSM system, three standard voice coders are applied most frequently: GSM EFR, GSM FR, and GSM FR [2], [3]. All of these voice coders affect every mobile phone equally. The impact of the GSM system and Margento terminal can be avoided by using adjusted modulated methods and by implementing the forward error correction algorithms – FEC. The major degradation factor which reduces the quality of data transmission is the impact of the mobile phone itself.

2.3 Mobile Phone Impact

The most important mobile phone impacts are:

- Echo Canceller, and
- Voice Activity Detector System – VAD.

The echo canceller and the voice activity detector are – from the audio modulated data transmission point of view – unwanted systems that disturb the quality of data transmission over the GSM voice channel. Both systems can disturb broadcasted signal up to the limit, where the signal, received by the Margento centre, is totally useless for demodulation.

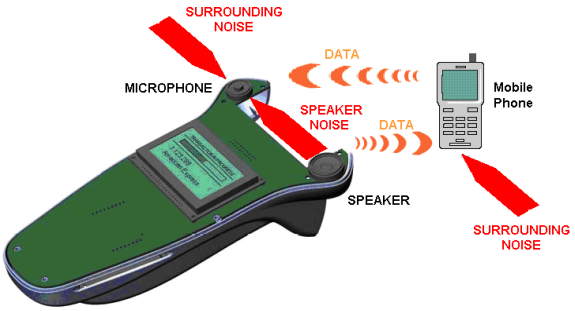

Figure 3.Margento terminal and corresponding acoustic impacts and interferences.

2.3.1 Echo Canceller

The basic function of the echo canceller is to eliminate all the signals, which are not similar to speech, and the sub signals, which appear along the speech transmission. One of the unwanted speech signal characteristics is echo [10]. The echo occurrence and development of the echo reduction had appeared long before the mobile phoning. The first demands to eliminate the echo started in the 1960s, more precisely, in satellite telecommunication. These echoes can be divided into the hybrid and the acoustic echoes.
The hybrid echo elimination is carried out well with the FIR (Fine Impulse Response) in mobile phone technology [14]. The acoustic echo is generated in the analogue and digital device. The echo cancellers are realized with complex algorithms, which have to predict the echo path or the delay and echo strength. The mathematical acoustic echo model is needed for the qualitative canceller. Model approximation is adaptive, which means that the canceller parameters adjust to the certain mobile phone environment system by themselves. The canceller adapts or adjusts the parameters continuously. The final parameter's adaptation is called the convergent time [20]. The canceller's efficiency depends on the strength and maximal echo signal delay, received by the microphone. The echo signal estimation is subtracted from the original signal and the person, who is the source of the speech, cannot hear his/ her own speech echo.

2.3.2 Voice Activity Detector

Generally speaking, the human speech consists of sequences of silences and sounds. The silence parts are between the words and sentences, and/or when we do not talk and just listen to the speaking person [4], [5], [6], [18]. Most of the time, the mobile phone user is in the areas where there are various acoustic background noises. The signal, detected by the mobile microphone, is the sum of the surrounding noise, or the surrounding noise and speech. In most speech applications there are utilized algorithms to automatically distinguish between useful speech and background noise (or silence). These are the so called the Voice Activity Detectors, VAD. The VAD systems are used in different communication systems. Quite often, VAD is one of the most important components in the mobile phone and in similar systems. The communication quality increases with the quality of speech detection; however, it can disable normal communication in extremely noisy surroundings. Without the VAD system, there is no qualitative communication in modern mobile networks. The VAD system in connection with the echo canceller has become the integral component of every mobile phone. The primary intention is to estimate the tempo-spectral characteristics of the input speech signal, captured by the microphone [7]. The noise spectrum is not stationary and can be changed quite quickly; therefore, the changes have to be followed properly with the noise spectrum estimations. The noise estimation can be carried out any time when the speech is not active and the microphone detects only the noise signal. At that time, the system estimates the spectrum and cancels the signal as long as only the noise is in the input [8]. Furthermore, there are also some algorithms which continuously estimate the noise spectrum and do not need any speech pauses for noise estimation [9]. The VAD system is efficiently used for the discontinuous speech transmission. The problem of the VAD systems is that they can decrease the speech quality. It can happen that VAD can misinterpret the speech for the noise and in these circumstances the mobile phone will not transmit the speech. Moreover, the high quality VAD can increase the perceptive quality of the transmitted speech and simultaneously reduce the power consumption of the mobile device [16]. Namely, there is no need to transmit the nonspeech frames through the GSM network. Where the surrounding noise strength is low in comparison with the speech strength (big SNR-Signal to Noise Ratio), VAD can quite quickly and easily detect segments in the signal where there is no speech. And there is quite the opposite situation in the case of low SNR. In such cases quite complex and adaptive VAD algorithms are needed to deal with these very dynamic situations [15].

The VAD systems and echo cancellers render the audio modulated data transmission over the voice channel because the voice modulated data signals, which are FSK (Frequency-Shift-Keying), ASK (Amplitude-Shift Keying), QAM (Quadrature Amplitude Shift-Keying), OFDM (Orthogonal Frequency-Division Multiplexing), etc., have tempo-spectral characteristics usually much different than the speech signal. Therefore, it is necessary to analyse the VAD's and echo canceller's impact on the voice modulated data signals. The documentation and specifications about the GSM ETSI VAD system are publicly available and thus the ETSI VAD simulator can be implemented, i.e. in the MATLAB environment. However, it should be noted that the exact simulation of the echo canceller cannot be realized because –in

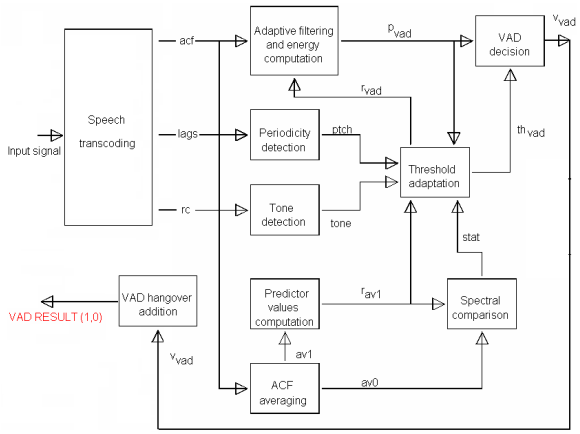

Figure 4.VAD system diagram according to the ETSI standard.

opposite to the VAD– there is no official standard for the echo canceller. Namely, each mobile-phone manufacturer has its own secret algorithm for the echo cancellation. Only the testing on the real objects can be carried out in order to analyse the VAD together with the echo canceller.

3 Implementation of the GSM ETSI-VAD system within MATLAB

3.1 Description of the GSM ETSI VAD system

The GSM VAD system proposed by the ETSI institution is basically a signal-energy detector based on the adaptation adjustment to the determined speech-nonspeech threshold (Fig. 4). The VAD system uses the following characterizations and facts about the speech and surrounding noise:

- Speech is a discontinuous signal,
- The speech signal spectrum changes in short periods of time from 20 to 30 ms [19],
- The surrounding noise is usually more stationary than speech,
- The surrounding noise spectrum is usually changed with longer average sequences of time, depending on the speech,
- The amplitude dynamics of the speech is more expressive than the dynamics of the surrounding noise,
- The surrounding noise usually has a spectrum similar to the white or coloured noise which differs from the spectrum of the speech. The tempo-spectral characteristics of the speech are usually more complex.

The VAD system is roughly divided into nine components, where the logical decision about the speech presence or absence is based on various parameters (see Fig. 4), calculated in 160 samples (20 ms) long intervals (GSM speech coders operate at the sampling frequency of 8000 Hz) [13]. The input signal is filtered through the adaptive noise reduction filters, which reduce the level of the surrounding noise in the captured waveform. Filter coefficients are calculated every four frames (80 ms) through average coefficient autocorrelation. This method contributes to better surrounding noise cancellation [12]. The energy threshold and adaptive filter coefficients are adapted only when there is no speech in the input signal. Namely, in this case the input signal energy is usually low, or the input signal spectrum is stationary and it does not include the periodical component, which could be the result of the special network information tone. The signal stationarity is verified by the LHR (Likelihood Ratio) measurement between the average linear-predictive coefficients of the last four frames and the temporary LPC (Linear prediction Coding) filters. The LPC filter coefficients represent the formants of the human vocal tract. In the linear prediction, it is presumed that the signal sample value at a certain moment of time is the linear combination of the determined number of past signal samples. When these coefficients are calculated, human speech can be produced with the reverse filtering of the received signal. In case the result of LHR method is smaller than the fixed threshold, the signal is stationary. The presences of periodical components are calculated every 5 ms. In order to prevent the pauses between syllables (smaller speech quality) to be misclassified as noise or silence, the VAD system is making decisions every five frames (100 ms). Moreover, hangover is applied thus marking few noise-only or silence frames after every, sufficiently long speech interval, as speech. The GSM VAD system is adjusted to the principle of making decisions such as »Wrong-Safety«, which means that whenever the system is in doubt about the input signal classification, VAD makes a decision that the input signal is the human voice (preference of false acceptance rather than true rejection) [17].

3.2 Implementation in the MATLAB Programme environment

The GSM VAD system, which was standardised by ETSI, has been implemented in the MATLAB programme environment. The program code is written in the M-file. ETSI has prescribed 16 test vectors (the input and output signals) used to check VAD algorithm implementation suitability and compliance with the standard. The algorithm suitability of the MATLAB implementation has been checked with test vectors so that the ETSI VAD system impact analysis on voice modulated data signal can be performed. The most important VAD algorithm's intermediate signal results are shown in the diagrams in Fig. 5 in order to understand the operation of the ETSI VAD system. The selected input test signal is the standardized ETSI test vector number 5 (female speech). Fig. 5 shows the following signals (from top to bottom):

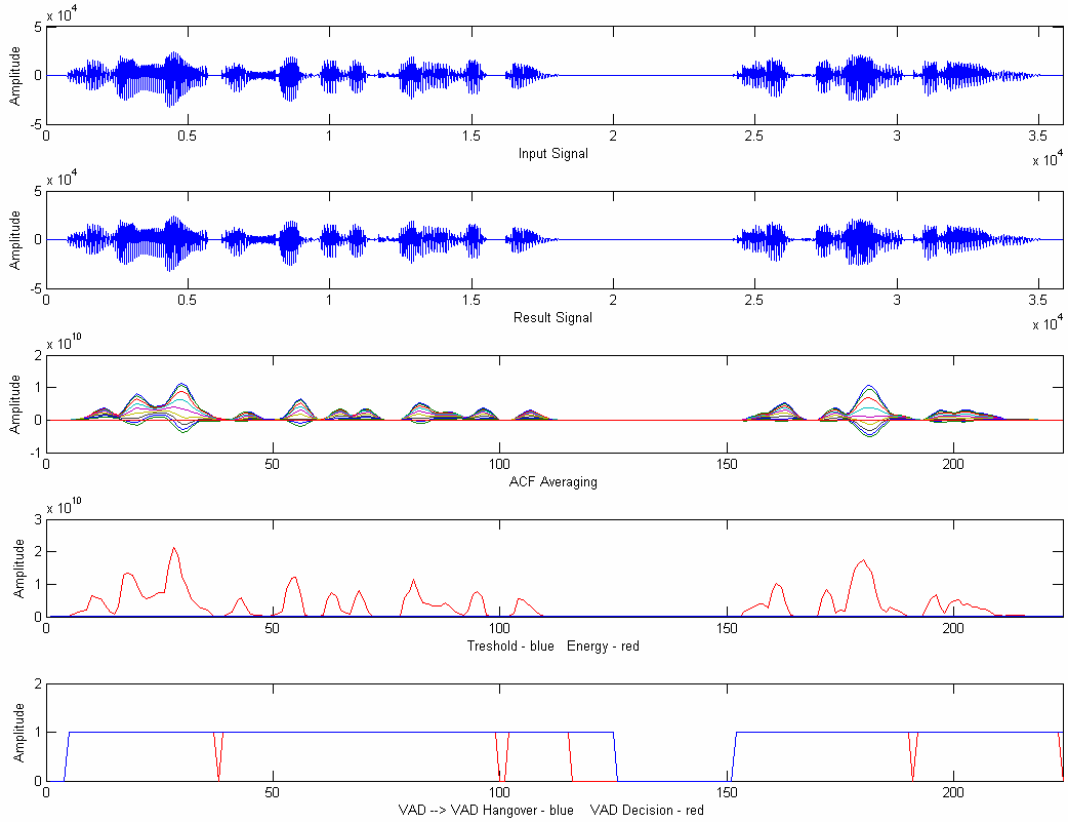

Figure 5.VAD system intermediate operation signals (see description in Section 3.2)

1. The input test signal,
2. The output signal- The ETSI VAD system result,
3. The computed result of an average autocorrelation or input signal energy,
4. The output of the adaptive algorithm for the adjustment of the threshold decision, and
5. The VAD system result and the result of the postprocessed VAD decision algorithm (hangover).

The ETSI VAD system implementation in the MATLAB environment enables the analysis of disturbances over the other test signals, which replicate the real-world scenarios in the operating Margento (Margento) system. Six new test signals were used (each of them of length of 30 seconds):

- White noise,
- QAM modulated signal at 400bps (carrier at 800Hz),
- FSK 200bps (carriers at 800 Hz and 1100Hz),
- ASK 200bps (carrier at 800Hz),
- LPC Modulation based on the LCP speech synthesis, and
- Speech signal (male, female voice).

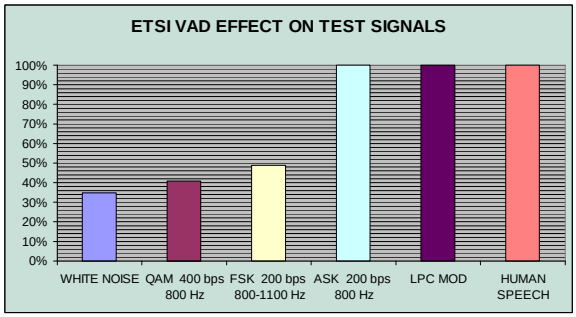

Figure 6. Simulated ETSI VAD performance using different test signals.

The analysis results of the ETSI VAD system show the percentage of the total input signal length which was marked as “speech” and, therefore, wasn’t cut-off by the VAD system. The results presented in Fig. 6 show that the white noise, QAM, and FSK test signals are more often classified as “non-speech”. It can also be seen that the human speech and the LPC modulated signal are - as expected - entirely marked as “speech” by ETSI VAD. Therefore, the proposed digital modulation scheme must have similar tempo-spectral characteristics as human speech in order to be properly transmitted through the GSM speech channel.

4 Real GSM system experimental results and discussion

The ETSI VAD system performance has been checked also in the real system with six test signals which were transmitted over the voice channel of various GSM mobile phones. The experiment was carried out in the supervised environment with the high SNR signal quality. Each mobile phone has been tested multiple

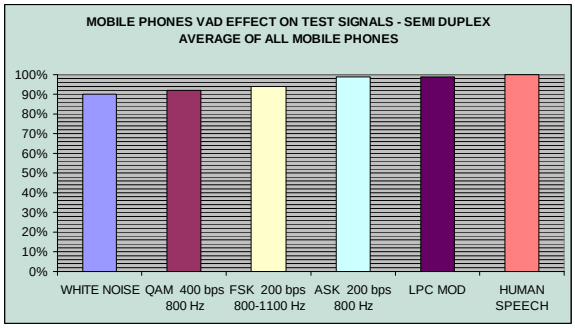

Figure 7. Average GSM phone VAD+echo canceller impact at half-duplex communication over the GSM speech channel.

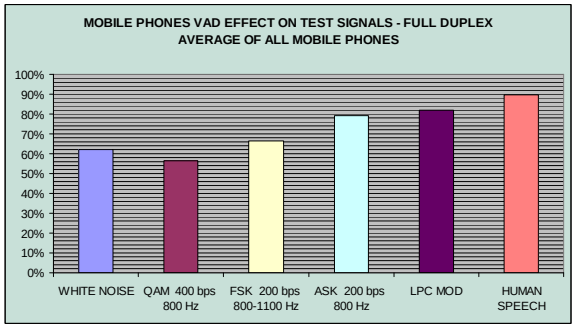

Figure 8. Average GSM phone VAD+echo canceller impact at full-duplex communication over the GSM speech channel.

times in the two phase test and the average of the results was taken into consideration in each phase. The whole testing was divided into the following two phases:

- one-way communication (half-duplex)
- simultaneous both way communication (full duplex).

The purpose of this testing was the analysis of VAD and echo canceller impact on the signal reception. Fig. 7 shows the average VAD system impact with various mobile phones on the above mentioned test signals with half-duplex communication. The results show, that the mobile phones pass over the test signals over the GSM's voice channel almost entirely. There were actually some difficulties with the Siemens SL55 and BENQ Athena mobile phones. Fig. 8 show the average VAD system impact of various mobile phones at full-duplex communication. It has to be mentioned that all mobile phones were set to their maximum volume performance and the Margento centre was sending the test signal FSK 300 bps (carriers at 1650Hz, 1850Hz) with the relative amplitude of 30.000 at 16 bit separation. Consequently, the echo canceller disturbed the input signal up to the limit where the VAD system algorithm quickly interpreted the test signal as the surrounding noise. Even some unexpected speech cuttings were noticed with some mobile phones and consequently the average speech pass was decreased/ lowered from over 90% to 70%. The results show that the echo canceller impact is very important for further analysis. Therefore, the mobile phone Siemens SL45 has been chosen because of the highest fluctuations of the result. Fig. 9 shows the combined echo canceller and VAD system impact at the volume changing with the data transmission towards from the Margento centre to the mobile phone (consequently the Margento terminal). The volume of the full-duplex transmission is divided into three levels with the half duplex result as a reference. The results show a strong signal volume dependence of the VAD and echo canceller system performance. Thus the operation of the echo canceller depends on the data transmission volume. It can also be seen that the difference between the signal pass during one way communication and simultaneous both way communication is minimal, up to 56% at the FSK test signal and 64% at the ASK test signal. In order to improve the data transmission quality over the voice channel of the real GSM system, the new adaptive algorithm is proposed (see Fig. 10).

At the beginning of the data transmission, there is always simultaneous both-way communication (full duplex). In the situation where the bit-error rate (BER) is increased, the data transmission volume is decreased. Consequently, the negative echo canceller impact is decreased (see Fig. 9). In the situation where the signal volume decrease has not the expected impact, the transmission mode is switched to half duplex. This will eliminate the echo canceller impact. However, the overall data transmission time will be increased.

5 Conclusion

In this paper we studied the effects of VAD and echo canceller modules on audio modulated digital data transmission of the GSM speech channel. First, we analyzed and implemented the ETSI VAD procedure in the MATLAB environment. Next, the simulated VAD was evaluated using different modulation signals in order to check the VAD's output decision. It was found that the best performance is achieved when the modulation signals has speech-alike tempo-spectral characteristics. Such signals will usually not be cut-off by the VAD. Next, we performed real GSM system tests (VAD+echo canceller) using multiple mobile phones at

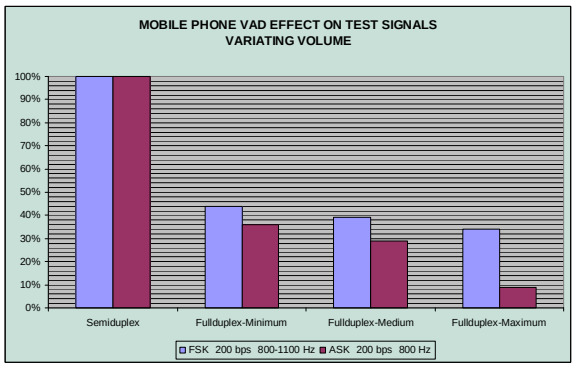

Figure 9. Signal volume dependence of average GSM phone VAD+echo canceller impact at full-duplex communication over the GSM speech channel.

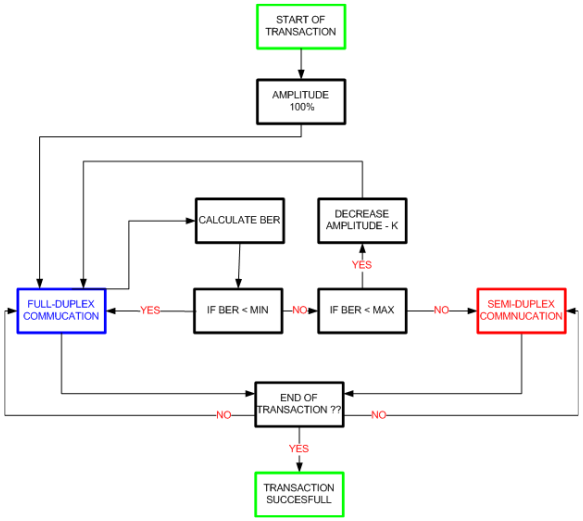

Figure 10. Proposed adaptive data transmission scheme to deal with the negative effects of VAD and echo canceller.

two different transmission mode: half duplex and full duplex. It was observed that the half duplex mode performs more or less flawlessly, while the half duplex mode in some cases degrades the transmission of the modulated signals due to the combined impact of the VAD and echo canceller. We observed that the VAD algorithm is standardized by ETSI and as such implemented by all manufacturers. However, the echo canceller is not standardized. Therefore, the performance of different mobile phones is varying at full duplex communication. Furthermore, the higher volume of the modulation signal transmitted in one direction degrades the reception of the signal in the opposite direction more strongly than the lower volume. Finally, a new adaptive data transmission scheme is proposed to optimise the data transmission performance over the GSM speech channel and to compensate the negative impacts of mobile phone's VAD and echo canceller.

6 Literatura

[1] Ultra Margento patent 1 and 2, WO 02/33669, WO 03/088165, 2002
[2] ETSI EN 301 245 v4.1.1, "Digital cellular telecommunications system (Phase 2)-enhanced full rate speech transcoding (GSM 06.60)", 2000
[3] ETSI EN 300 730 v7.0.1, "Digital cellular telecommunications system-voice activity detector for enhanced full rate speech traffic channels (GSM 06.82)", 2000
[4] L. Hanzo, F.C.A Somerville, J. P. Woodard, "Voice compression and communications", 1999
[5] Huan M. Huerta, "Speech recognition in mobile environments", 2000
[6] Mark Marzinzik, Birger Kollmeier, "Speech pause detection for noise spectrum estimation by tracking power envelope dynamics", 2002
[7] Wang Fan, Zheng Fang, Wu Wenhu, "Speech detection in non-stationary noise based on 1-f process", 2002
[8] J. Rosca, R. Balan, N. P. Fan, "Multichannel voice detection in adverse environments", 2002
[9] J. Ramirez, J. C. Segura, C. Benitez, A. Rubio, "Efficient voice activity detection algorithms using long-term speech information", 2003
[10] Peter Eneroth, "Stereophonic acoustic echo cancellation", 2001
[11] John Scourias, "Overview of the global system for mobile communications", 1995
[12] Richard V. C., Peter Kroon, "Low bit-rate speech coders for multimedia communication", 1996
[13] L. Besacier, S. Grassi, A. Dufaux, M. Ansorge, F. Pellandini, "GSM speech coding and speaker recognition", 2000
[14] Jan Mark De Han, "Filter bank design for digital speech signal processing", 2004
[15] Kristo Lehtonen, "Digital signal processing and filtering – GSM Codec", 2004
[16] Arvind Raman Kizhanatham, "Detection of cochannel speech and usable speech – GSM Codec", 2002
[17] Khaled El-Maleh, Peter Kabal, "Comparison of voice activity detection algorithms for wireless personal communications systems", 1997
[18] Mark D. Skowronski, "Biologically inspired noise-robust speech recognition for both man and machine", 2004
[19] F. Beritelli, S. Casale, A. Cavallaro, "A robust voice activity detector for wireless communications using soft computing", 1998
[20] Peter Eneroth, Tomas Gansler, "A frequency domain adaptive echo canceller with post-processing residual echo suppression by decorrelation", 1997

Zdenko Mezgec received his BSc degree in Electrical Engineering in 2004 and Ph.D. in 2009 both at University of Maribor at University of Maribor, Slovenia. He has been working at Margento R&D as Chief of Embedded systems development.

Amor Chowdhury received his MSc degree in Electrical Engineering in 1997 and PhD in Robust Control in 2001 both at University of Maribor, Slovenia. Since 2008 he is CEO of Margento R&D d.o.o., and beside this he is still working at University of Maribor, Faculty of Electrical Engineering and Computer Science, Slovenia.

Rajko Svečko received his MSc degree in Electrical Engineering in 1984 and PhD in Robust Control in 1998 both at University of Maribor, Slovenia. He works as a associates professor and researcher at University of Maribor, Faculty of Electrical Engineering and Computer Science, Slovenia.

Bojan Kotnik obtained his B.Sc. degree in Electrical Engineering in 2000 and Ph.D. in Automatic Speech Recognition in 2004 both at University of Maribor, Slovenia. His research domains are in the fields of digital signal processing, digital modulation and demodulation algorithms, and statistical methods for data classification. Since 2008 he has been working as Chief Scientific Officer at Margento R&D.