INFORMATICA, 2010, Vol. 21, No. 4, 575–596

GSM Speech Coder Indirect Identification Algorithm

Rajko SVEČKO[1], Bojan KOTNIK[2], Amor CHOWDHURY[2], Zdenko MEZGEC[2]

[1]*University of Maribor, Faculty of Electrical Engineering and Computer Science*
Smetanova ul. 17, SI-2000 Maribor, Slovenia
[2]*Ultra d.o.o., Research Center*
Gosposvetska ul. 84, SI-2000 Maribor, Slovenia
e-mail: rajko.svecko@uni-mb.si

Received: April 2009; accepted: September 2009

Abstract. This paper presents GSM speech coder indirect identification algorithm based on sending novel identification pilot signals through the GSM speech channel. Each GSM subsystem disturbs identification pilot, while speech coder uniquely changes the tempo-spectral characteristics of the proposed pilot signal. Speech coder identification algorithm identifies speech coder with the usage of robust linear frequency cepstral coefficient (LFCC) feature extraction procedure and fast artificial neural networks. First step of speech coder identification algorithm is the exact position detection of the identification pilot signal using normalized cross correlation approach. Next stage is time-domain windowing of the input signal to convolve each frame of the input speech signal and window spectrum. Consecutive step is a short-time Fast Fourier Transformation to produce the magnitude spectrum of each windowed frame. Further, a noise reduction with spectral subtraction based on spectral smoothing is carried out. In last steps we perform the frequency filtering and Discrete Cosine Transformation to receive 24 uncorrelated cepstral coefficients per frame as a result. Speech coder identification is completed with fast artificial neural network classification using the input feature vector of 24 LFCC coefficients, giving a result of identified speech coder. For GSM speech coder indirect identification evaluation, the standardized GSM ETSI bit-exact implementations were used. Furthermore, a set of custom tools was build. These tools were used to simulate and control various conditions in the GSM system. Final results show that proposed algorithm identifies the GSM-EFR speech coder with the accuracy of 98.85%, the GSM-FR speech coder with 98.71%, and the GSM-HR coder with 98.61%. These scores were achieved at various types of surrounding noises and even at very low SNR conditions.

Keywords: GSM, speech coder, identification, pilot.

1. Introduction

GSM technology has been used worldwide for almost two decades. At the end of first quarter of 2007 there were 2,278,095,380 GSM subscribers, which represent almost 80.5% of all mobile network subscribers (Wireless Intelligence Organization information's; Scourias, 1995). With rapid growth a lot of systems were developed for GSM interaction in different approaches and for various purposes. All these systems can be

roughly divided in two parts. First part is usually just application software running on mobile phone itself, while second part presents the systems that interacts with mobile phone through various communication channels (Scourias, 1995). These systems usually don't have information about currently used mobile phone speech coder. This information is especially significant for systems that use mobile phone speech channel for data transmitting or speech encryption (Lehtonen, 2004; Ultra M-Pay Patent 1 and 2, 2002). Namely, the speech coder importantly impacts the modulation used for data transmission and consequently lack of speech coder information can significantly reduce ability to perform robust, fast and reliable data transmission (Bingham, 2000).

The accurate GSM speech coder identification procedure can be beneficially applied also in modern IVR (Interactive Voice Response) systems incorporating the automatic speech recognition (ASR) technology (Rotovnik *et al.*, 2007). Namely, the ASR system could apply specially designed and trained acoustical models in cases of lower quality GSM speech coders like GSM FR or GSM HR. Such adaptive acoustic model selection mechanism could improve the performance of overall IVR system substantially.

This paper presents GSM speech coder indirect identification based on sending novel pilot signals through GSM speech channel (Fig. 1).

Identification pilot structure (described in Section 3) is designed to achieve three main contradict objectives. First objective is robust pilot transmission through all GSM subsystems. Speech identification algorithm must reliably detect pilot signals in relatively noisy environments and all subsystems must recognize pilot signal as speech (Ibars and Bar-Ness, 2001). Second objective is to be able to identify each speech coder with checking only its unique disturbances that were made on pilot signal. Because of that, all other GSM subsystems must disturb the carefully predefined pilot signal in different way, so that identification algorithm can still be able to identify speech coder (Scourias, 1995). Third objective is that identification pilot signal should be as short as possible to leave the speech channel capacity free for modulated primary data transfer (Xiong, 2006; Mezgec *et al.*, 2009). Proposed identification pilot structure is described in Section 3.

Speech coder identification algorithm first detects presence of identification pilot signal with normalized cross correlation approach (described in Section 4). Significant part of the pilot is used for further time-domain pre-processing. First, the input signal is divided into short segments, called frames which are multiplied by a standard Hann win-

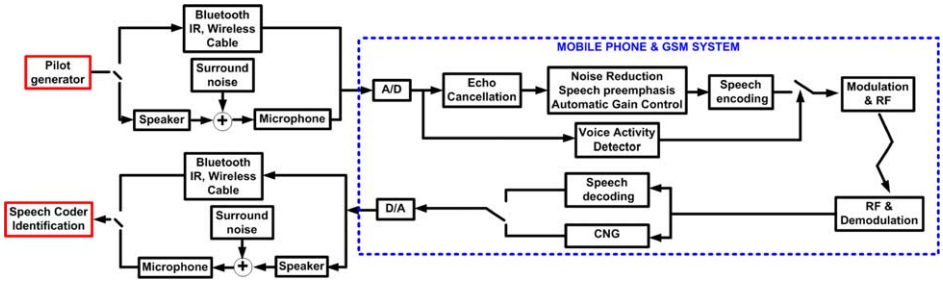

Fig. 1. Block diagram of GSM speech coder indirect identification algorithm.

dow. This operation is equivalent to convolution of the signal spectrum with the window spectrum. In consecutive step the short time FFT (Fast Fourier Transformation) is performed on each windowed frame to produce the magnitude spectrum. Further, a special attention is given to noise reduction performed by the spectral subtraction approach. Next the spectral smoothing is performed by a linear frequency filtering approach with triangular-shaped half-overlapped filters. In the last step of feature extraction algorithm the DCT (Discrete Cosine Transformation) is performed and 24 linear frequency cepstral coefficients (LFCC) are obtained as result. Feature extraction algorithm is described in Section 5. Speech coder identification is completed with the LFCC feature vector fed to input of the pre-trained ANN (Artificial Neural Network), that gives as a result identified speech coder. ANN procedure is described in Section 6.

For extensive analyses and experiments we developed different tools that simulate various controlled conditions in GSM system. Each tool operation was checked with GSM test sequences provided by ETSI (European Telecommunications Standards Institute). The results are presented and discussed in Section 7, conclusions are given in Section 8.

2. GSM Speech Coding Standards

GSM is the pan-European cellular mobile standard (Scourias, 1995). Three considered speech coding algorithms are part of this standard. The purpose of these coders is to compress the speech signal before its transmission, reducing the number of bits needed in its digital representation, while keeping an acceptable perceived quality of the decoded output speech signal. Spectrum efficiency of the GSM transmission system is increased through the use of DTX (Discontinuous Transmission), switching the transmitter on only during speech activity periods. VAD (Voice Activity Detection) is used to decide upon presence of active speech. To reduce the annoying modulation of the background noise at the receiver (noise contrast effects), CNG (Comfort Noise Generation) is used, inserting a coarse reconstruction of the background noise at the receiver (Scourias, 1995). Additionally, GSM incorporates also non-standard modules such as Echo Cancellation, Noise Reduction, Speech Preemphasis and Automatic Gain Control. As mentioned earlier, there exist three different GSM speech coders, which are referred to as the FR (full rate), HR (half rate) and EFR (enhanced full rate). Their corresponding European telecommunications standards are the GSM 06.10 (1996), GSM 06.20 (1998) and GSM 06.60 (1997). These coders work on a 13 bit uniform PCM speech input signal, sampled at 8 kHz. The input is processed on a frame-by-frame basis, with a frame size of 20 ms (160 samples). A brief description of speech coders follows.

2.1. *Full Rate Speech Coder*

The FR coder was standardized in 1987. This coder belongs to the class of RPE-LTP (Regular Pulse Excitation – Long Term Prediction) linear predictive coders (Hanzo *et al.*, 1999). At the encoder part, a frame of 160 speech samples is encoded as a block of 260 bits, leading to a bit rate of 13 kbps (GSM 06.10 Full Rate (FR) Vocoder, 1996).

The decoder maps the encoded blocks of 260 bits to output blocks of 160 reconstructed speech samples. The GSM FR channel supports 22.8 kbps. Thus, the remaining 9.8 kbps are used for error protection. The FR coder is described in GSM 06.10 down to the bit level and a set of digital test sequence for verification is also given.

2.2. *Half Rate Speech Coder*

The HR coder standard was established to cope with the increasing number of subscribers. This coder is a 5.6 kbps VSELP (Vector Sum Excited Linear Prediction) coder from Motorola (GSM 06.20 Half Rate (HR) Vocoder, 1998; Hanzo *et al.*, 1999). In order to double the capacity of the GSM cellular system, the HR channel supports 11.4 kbps. Therefore, 5.8 kbps are used for error protection. The measured output speech quality for the HR coder is slightly degraded to the quality of the FR coder. The normative GSM 06.06 gives the bit-exact ANSI-C code for this algorithm, while GSM 06.07 gives a set of digital test sequences for compliance verification.

2.3. *Enhanced Full Rate Speech Coder*

The EFR coder was the latest on field implemented. This coder is intended for utilization in the EFR channel, and it provides a substantial improvement in quality compared to the FR or HR coder. The EFR coder uses 12.2 kbps for speech coding and 10.6 kbps for error protection. The speech coding scheme is based on ACELP (Algebraic Code Excited Linear Prediction; Hanzo *et al.*, 1999). The bit exact ANSI-C code for the EFR coder is given in GSM 06.53 and the verification test sequences are given in GSM 06.54 (1997).

3. Identification Pilot Structure

3.1. *Objectives*

Identification pilot signal structure is designed to achieve three main contradict objectives.

First objective is robust pilot transmission through the speech transmission channel of all GSM subsystems. As mentioned before, the GSM system is primarily optimized to transmit the speech signal from one point to the other. As GSM transcoding (the process of coding and decoding) modifies the speech signal in a pitch-asynchronous manner, it is likely to have a strong influence on phase, amplitude, and spectral characteristics of the transmitted signal together with other perturbations introduced by the mobile cellular network (channel errors, background noise). Moreover, the speech channel of the GSM is primarily intended for transmission of speech and DTMF signals only. If pilot signal is to be transmitted over the GSM, it must be in principal similar to the tempo-spectral characteristics of the speech signal. Moreover, due to the characteristics of the RPE-LTP, VSELP and ACELP, the signal should mimic the properties of the sustained

voiced speech (pitch frequency and higher harmonics, formants, etc.) for best transmission properties. If this is achieved, then VAD will also detect pilot signal as speech frame and will not activate the DTX mode. For robust speech coder identification we must also think about scenario where surrounding noise is present, which might occur when system transmits/receives voice to/from mobile phone through speaker and microphone (Fig. 1).

Second objective is to be able to identify each speech coder with checking only unique disturbances that were made on pilot signal. Due to different limited order of the autoregressive speech production modeling and characteristics of RPE-LTP, VSELP and ACELP, number of pilot frequencies and their positions are combined so that each speech coder uniquely disturbs same input signal (Chow *et al.*, 1991; Hanzo *et al.*, 1999).

Third objective is that identification pilot signal should be as short as possible so that also identification process uses speech channel as least as possible.

3.2. *Identification Pilot Structure*

The proposed identification pilot signal g[n] consists of two parts.

First part represents the signal $cc[n]$ that is actually a sum of two sinusoidal waveforms with linear frequency increase (so called chirp signals):

$$\begin{aligned} cc[n] = & \sin\left(2\pi\left(f_{1S} + \frac{(f_{1E} - f_{1S})(n-1)}{(S-1)}\right)\frac{n}{f_S}\right) \\ & + \sin\left(2\pi\left(f_{2S} + \frac{(f_{2E} - f_{2S})(n-1)}{(S-1)}\right)\frac{n}{f_S}\right), \end{aligned} \tag{1}$$

where f_S is a sampling frequency of 8 kHz. $f_{1S} = 400$ Hz and $f_{2S} = 600$ Hz are the starting frequencies of the chirp signals. The frequencies $f_{1E} = 600$ Hz and $f_{2E} = 800$ Hz are the ending frequencies of the mentioned chirp signals. $S = 640$ represents the chirp signal length in samples ($n = 1, 2, \ldots, 640$).

Second part represents the signal $ttt[n]$ that consists of three sinusoidal waveforms with different constant frequencies and amplitudes:

$$\begin{aligned} ttt[n] = & \frac{1}{2}\sin\left(2\pi f_3 \frac{n}{f_S}\right) + \frac{1}{1.5}\sin\left(2\pi f_4 \frac{n}{f_S}\right) \\ & + \frac{1}{2}\sin\left(2\pi f_5 \frac{n}{f_S}\right), \end{aligned} \tag{2}$$

where f_S is a sampling frequency of 8 kHz. $f_3 = 2000$ Hz, $f_4 = 2550$ Hz and $f_5 = 3200$ Hz are frequencies of sinusoidal waveforms. The length of $ttt[n]$ signal is 640 samples ($n = 1, 2, \ldots, 640$).

Both parts $cc[n]$ and $ttt[n]$ are multiplied with modified root-raised cosine window w_m[n] and afterwards added together:

$$g[n] = ttt[n]w_m[n] + cc[n]w_m[n] = tttw[n] + ccw[n]. \tag{3}$$

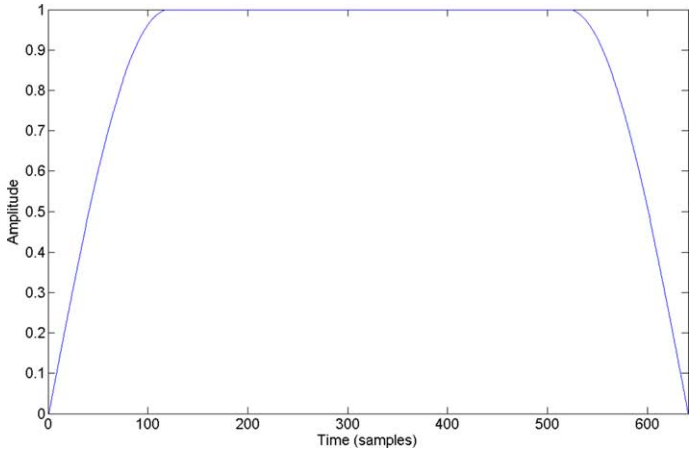

Fig. 2. Modified root raised cosine window wRRC[n].

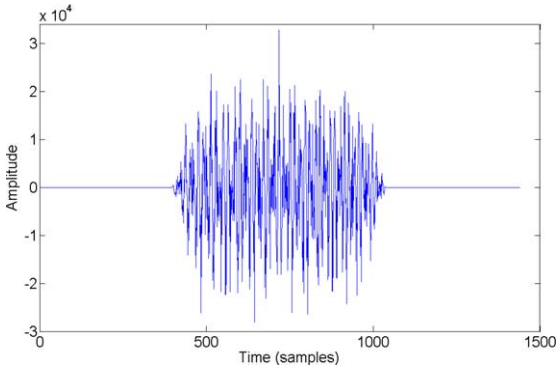

Fig. 3. Waveform of the proposed pilot signal $g[n]$.

The modified root-raised cosine window $w_m[n]$ (Fig. 2) is defined with (4), where $N = 120$.

$$w_m[n] = \begin{cases} \left(\cos\left(2\pi N \frac{N+n-1}{2N-1}\right)\right)^2, & n \leqslant 120, \\ 1, & 120 < n \leqslant 520, \\ \left(\cos\left(2\pi N \frac{N+(640-n)}{2N-1}\right)\right)^2, & n > 520. \end{cases} \tag{4}$$

The waveform and spectrogram of the proposed identification pilot signal are presented in Figs. 3 and 4.

As described above, the identification pilot signal consists of five different sinusoidal waveforms. The whole signal is used for feature extraction procedure and GSM speech coder identification. However, only the two chirp sinusoidal waveforms are used for pilot search based on cross correlation principle.

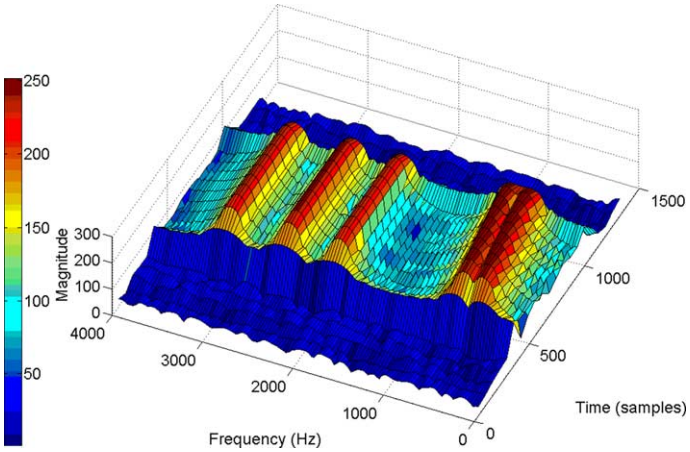

Fig. 4. Spectrogram of the proposed pilot signal $g[n]$.

4. Identification Pilot Search

In the first part of speech coder identification procedure the accurate positions of pilot sequence in the input signal is detected. Signal $ccw[n]$ (the chirp signals multiplied with $w_m[n]$) is used for this purpose. Namely, the chirp signal $ccw[n]$ is well correlated only with itself. It is well-known that the cross-correlation of $ccw[n]$ with any other arbitrary signal produces very low scores. Moreover, the autocorrelation produces high values only at full alignment of the two chirp signals (Fig. 5).

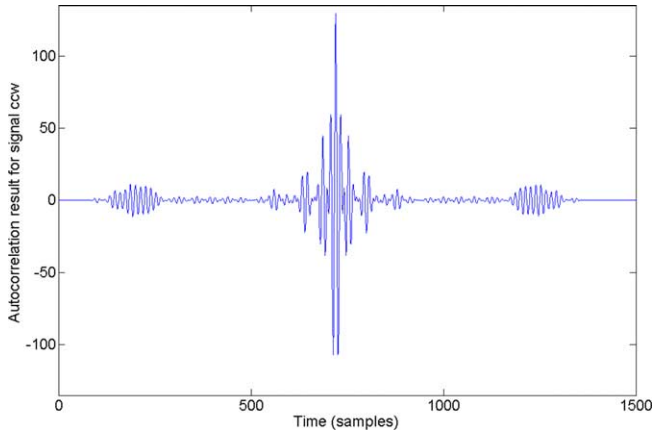

Fig. 5. Autocorrelation result for signal $ccw[n]$.

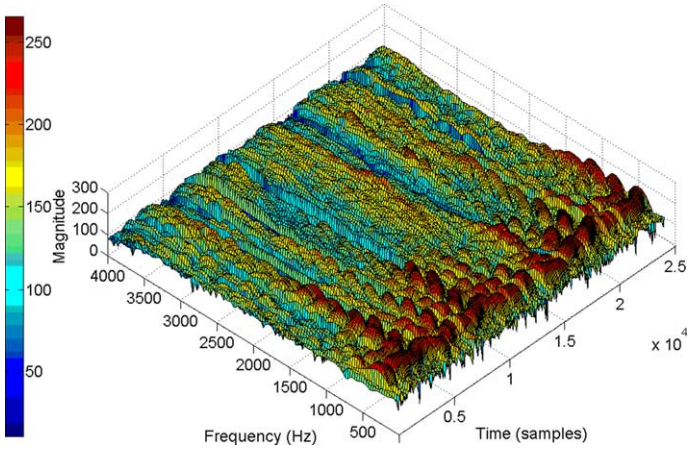

Fig. 6. Spectrogram segment of female speech signal.

The procedure is performed as follows; first the cross-correlation $x_1[n]$ between the prototype signal $ccw[n]$ and the input signal $s[n]$ is performed:

$$x_1[n] = \sum_{i=1}^{640} s[n+i] \cdot ccw[i]. \tag{5}$$

In order to produce the energy-normalized cross-correlation estimates, we need to determine the auto-correlation peak value for signal $ccw[n]$. Figure 5 presents autocorrelation result with peak value $PV = 129.9843$. Next, the energy-normalized cross correlation estimate $X_1[n]$ is determined:

$$X_1[n] = \frac{x_1[n]}{PV}. \tag{6}$$

In the last step, the candidates of the center positions are searched:

$$\textit{AllPeaks}[n] = \begin{cases} 1, & \text{if } X_1[n] \geqslant G, \\ 0, & \text{if } X_1[n] < G, \end{cases} \tag{7}$$

where G represents the a priori determined decision threshold. Finally, the most probable center position peak is searched using simple local-maximum peak search procedure (the local maximums are searched inside the 320 samples long time intervals). Unfortunately, the threshold G has to be determined empirically. The empirical procedure starts with set of 7 arbitrary signals. For that we choose *white noise*, *blue noise*, *brown noise*, *restaurant noise*, *road noise*, and *male* and *female* speech signals. Figure 6 shows example of chosen female noise.

In order to empirically determine the most appropriate value of threshold G, a special amplitude-decaying signal $cs[m]$ is constructed. The signal $cs[m]$ with the size of

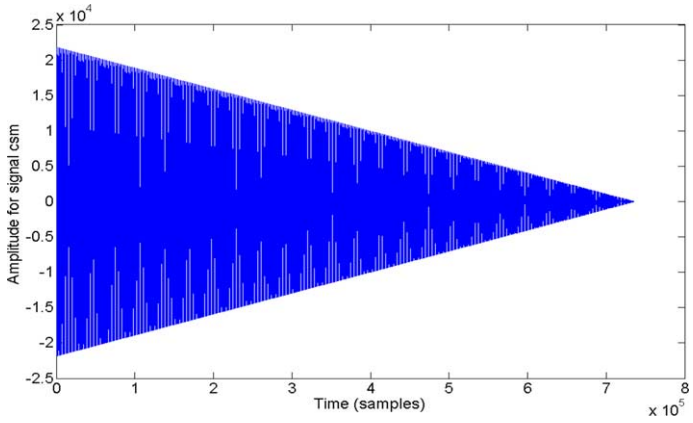

Fig. 7. $csm[n]$ signal for empirical search of threshold G.

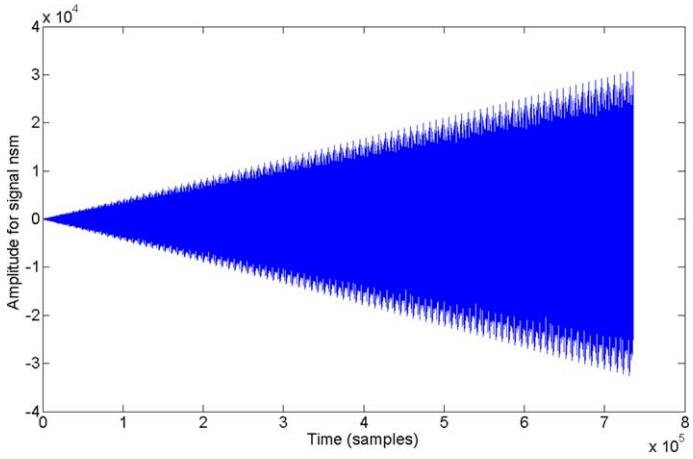

Fig. 8. $nsm_1[n]$ signal (white noise).

$S = 736,000$ samples consists of 512 consecutive repetitions of identification pilot signals. Between them there are inserted silences with average length of 700 samples. Finally, the signal $cs[m]$ is multiplied with linear amplitude decreasing factor (Fig. 7):

$$csm[m] = cs[m] \cdot \left(1 - \frac{(m-1)}{S}\right), \tag{8}$$

where $m = 1, 2, \ldots, S$.

Next we prepared seven signals $ns_1[m],\ ns_2[m], \ldots, ns_7[m]$ with size $S = 736,000$ samples, each of them consists one of arbitrary noisy signals, that we already mentioned. Each signal $ns_X[m]$ is then multiplied with linear amplitude increasing factor. This operation is represented in Fig. 8 and with (9).

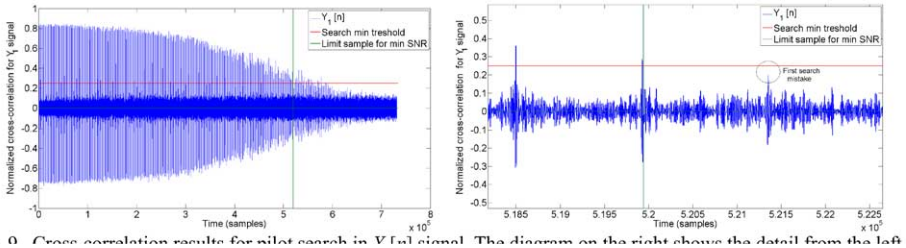

Fig. 9. Cross-correlation results for pilot search in $Y_1[n]$ signal. The diagram on the right shows the detail from the left diagram.

$$nsm_X[m] = \frac{m}{S} \cdot ns_X[m], \tag{9}$$

where $x = 1, 2, \ldots, 7$ and $m = 1, 2, \ldots, S$.

Further we make 7 pairs of signals:

$$Y_X[n] = csm[n] + nsm_X[n]. \tag{10}$$

These 7 signals $Y_X[n]$ are ready for final step where each of them is sent through 3 different speech coders. All together there are 21 resulting signals and 7 original signals that are still left for reference. With this procedure we have managed to simulate controlled gradual decreasing of SNR to check the peak values of cross-correlation result.

Figure 9 represents cross-correlation results for pilot search in $Y_1[n]$ signal. There we decrease min searching threshold G to the point where cross-correlation result with arbitrary signal doesn't exceed this limit. This guaranty, that when threshold G is exceeded, only pilot signals are found. The cross-correlation result decreases with decreasing SNR. The first point where cross-correlation result doesn't exceed threshold G at pilot signal position is denoted as limit sample for minimal SNR. This limit is used to estimate the minimal SNR. Figure 10 shows the computed SNR for the signal $Y_1[n]$. SNR is calculated for all 28 signals with the following procedure. Signal $Y_X[n]$ is divided into overlapping frames of the length 10 ms (80 samples). The frame shift interval for SNR calculation is 5 ms (40 samples) long. SNR for each frame is calculated with the following equation:

$$\mathrm{SNR}_X[s](\mathrm{dB}) = 10 \log \frac{\sum_{n=1+((s-1)\cdot 40)}^{n=(s+1)\cdot 40} csm[n]^2}{\sum_{n=1+((s-1)\cdot 40)}^{n=(s+1)\cdot 40} nsm_X[n]^2}, \tag{11}$$

where $x = 1, 2, \ldots, 7$ and $s = 1, 2, \ldots, \frac{S}{40}$.

With the procedure described above we empirically determined the threshold G and the lowest limit of SNR for different arbitrary noisy signals, to successfully detect identification pilot signal.

Table 1 shows Min SNR and thresholds G for all 28 signals. The black-marked threshold G is applied as an overall and universal threshold value that is used in final cross-correlation search in different environments. The results are also showing that individual

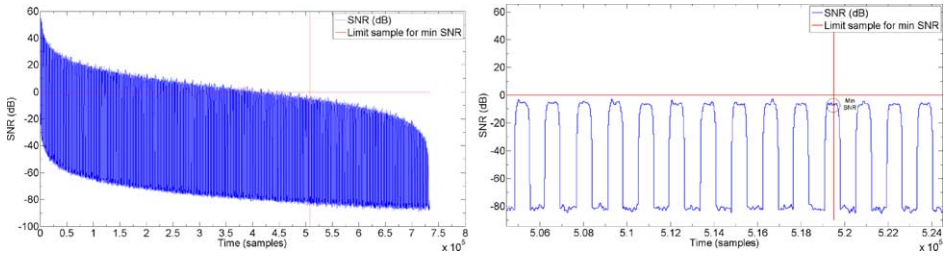

Fig. 10. SNR results for $Y_1[n]$ signal. The diagram on the right shows the detail from the left diagram.

Table 1

Minimal snr values and treshold g for succesfull cross-correlation search of identification pilot signal in different environments

Arbitrary test signals	Y_X signals		Y_X signals with GSM-FR		Y_X signals with GSM-EFR		Y_X signals with GSM-HR	
	G	SNR(dB)	G	SNR(dB)	G	SNR(dB)	G	SNR(dB)
Blue noise	0.06	−19.28	0.09	−16.06	0.07	−11.34	0.19	−1.60
Brown noise	0.05	−16.02	0.07	−13.89	0.06	−15.60	0.23	−5.59
Female speech	0.32	1.99	0.33	2.40	0.34	−2.34	0.35	2.81
Male speech	0.32	1.05	0.35	2.25	0.28	2.67	0.33	10.31
Restaurant noise	0.23	−1.28	0.22	−3.37	0.34	−3.27	**0.41**	5.97
Street noise	0.20	−7.56	0.20	−7.16	0.26	−4.03	0.27	−1.10
White noise	0.19	−8.13	0.20	−7.92	0.25	−5.47	0.25	−1.03

speech coders disturb each arbitrary signal uniquely. Pilot signals can be detected at SNR -16.06 dB, with added blue noise and through GSM-FR speech coder. The worst detection is with male speech test signal through speech coder GSM-HR and there pilot signals can be detected only at SNR -10.31 dB.

Speech coder identification algorithm first detects presence of identification pilot signal with cross correlation approach as described above. Empirically determined threshold G was found to be 0.41 to ensure that only pilot signals will be successfully found and none arbitrary signals would be miss-detected as a pilot signal. The center position of identification pilot signal is searched using simple local-maximum peak search procedure that was already described.

5. Feature Extraction Procedure

Once we know the central position of disturbed identification pilot signal in an input signal $y[l]$, feature extraction procedure starts. The length of the whole identification pilot signal is exactly 640 samples. For further time-domain preprocessing stage (windowing) we use only the middle, most significant part of this sequence with the length of 512 samples:

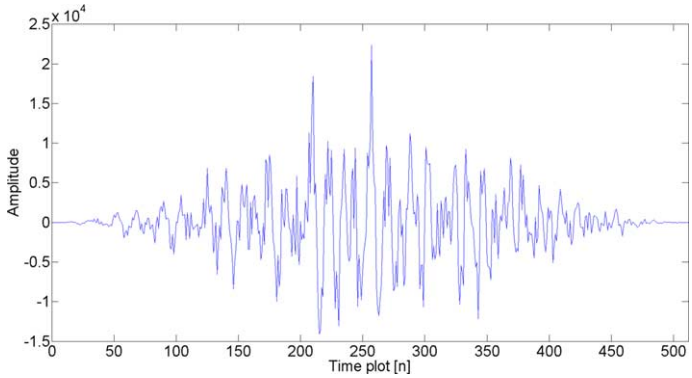

Fig. 11. Hann-windowed significant part of the identification pilot signal.

$$ip[l] = y[l + X], \tag{12}$$

where $X = \frac{640}{2} - \frac{512}{2}$ and $l = 1, 2, \ldots, 512$.

Significant part of the pilot signal is used for further time-domain preprocessing stage. The windowing is implemented by multiplying a signal with the standard Hann window $w_{\mathrm{Hann}}[l]$ which acts as a convolution of the signal spectrum with the window spectrum (see Fig. 11):

$$w_{\mathrm{Hann}}[l] = 0.5 - (1 - 0.5)\cos\left(\frac{2\pi l}{512}\right), \quad \text{where } l = 1, 2, \ldots, 512, \tag{13}$$

$$ipw[l] = w_{\mathrm{Hann}}[l] \cdot ip[l], \quad l = 1, 2, \ldots, 512. \tag{14}$$

In a consecutive step a Fast Fourier Transformation with the order of $N = 512$ is performed, producing the magnitude spectrum |*IPW*[k]| where k represents the frequency bin index running from 0 to 255 (Fig. 12):

$$f = k\frac{f_S}{\mathrm{FFT}_{\mathrm{ORD}}}, \qquad f_S = 8000 \text{ Hz},$$
$$\mathrm{FFT}_{\mathrm{ORD}} = N = 512, \quad k = 0, 1, \ldots, 255. \tag{15}$$

The next operation step is intended for simple noise reduction and spectral smoothing before the final step of discrete cosine transformation. A well-known 32 channel ($NumChan$) linear filter bank analysis with triangular shaped and half overlapped filters (Fig. 13) is performed on the signal |*IPW*[k]| (Rotovnik *et al.*, 2007). The central frequencies of the triangular filters are calculated with the following equation:

$$f_C[i] = \frac{i}{(NumChan + 1)} \cdot \frac{f_S}{2}, \tag{16}$$

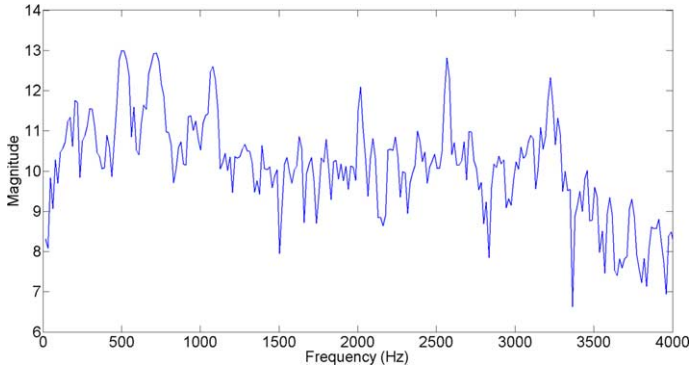

Fig. 12. The short-time spectrum instance of disturbed identification pilot signal.

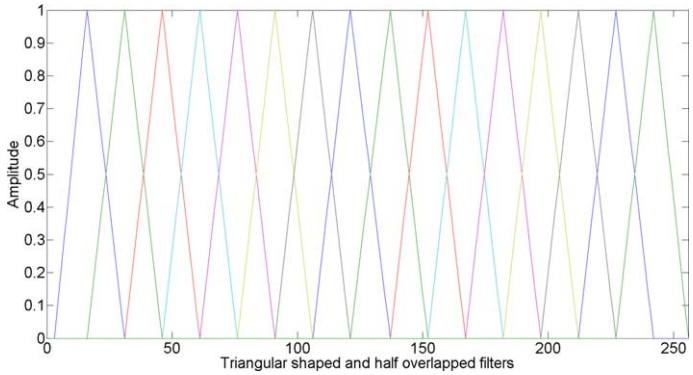

Fig. 13. Triangular filters for filter bank analysis.

where $i = 1, 2, \ldots, NumChan$ denotes a filter-bank channel index. Further we calculate central frequencies bin indexes:

$$k_C[i] = 0.5 + \left(f_C[i] \cdot \frac{\mathrm{FFT}_{\mathrm{ORD}}}{f_S} \right), \quad i = 1, 2, \ldots, 32. \tag{17}$$

Next we calculate half overlapped triangular shapes that represent filter banks and apply them on the magnitude spectrum |*IPW*[k]|. Afterwards, the filter-bank outputs $f_{\mathrm{bank}}[i]$ are subject to a natural logarithmic function, producing log filter-bank magnitudes $f_{ln}[i]$:

$$f_{\mathrm{ln}}[i] = \ln(1 + f_{\mathrm{bank}}[i]). \tag{18}$$

Figure 14 shows result of linear filter bank analysis and nonlinear transformation on magnitude spectrum |*IPW*[k]|.

Figure 15 presents results comparison for a clear identification pilot signal disturbed only by three GSM speech coders. The results present main differences in reaction of

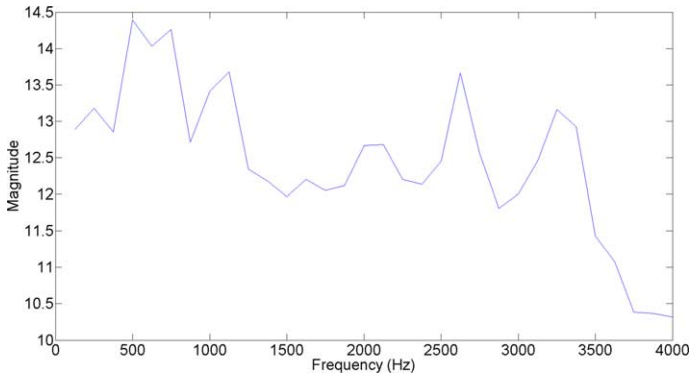

Fig. 14. Linear-frequency filter bank analysis and nonlinear transformation on disturbed identification pilot signal.

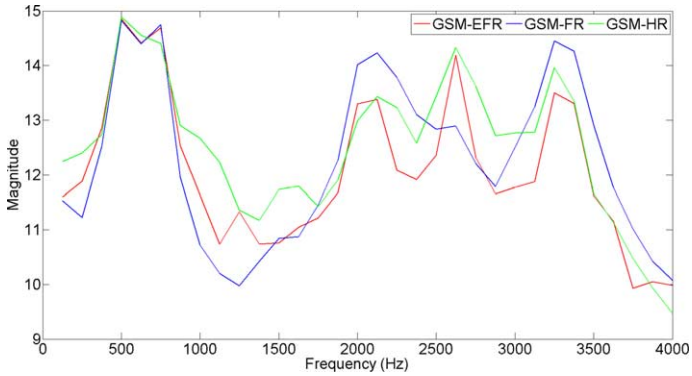

Fig. 15. Unique disturbances of identification pilot signal components, made by three GSM speech coders.

particular speech coder's autoregressive speech production model to identification pilot signal. Similar differences are still found when speech coder frame synchronization changes and also when arbitrary signal (surrounding noise) is added.

In the last step of proposed feature generation algorithm the discrete cosine transformation (DCT) is performed on 32 linear filter bank magnitudes $f_{ln}[i]$ to produce the resulting 24 linear frequency cepstral coefficients $LFCC[j]$, where $j = 1, 2, \ldots, 24$. With the usage of DCT a number of coefficients in the final output feature vector is effectively decreased. Furthermore, the output coefficients are simultaneously de-correlated. Figure 16 presents the 24 linear-frequency cepstral coefficients for a clear identification pilot signal disturbed by the three individual GSM speech coders.

6. ANN-Based Classifier

For GSM speech coder identification based on extracted feature vector of cepstral coefficients, an artificial neural network is applied (Fast Artificial Neural Network Library).

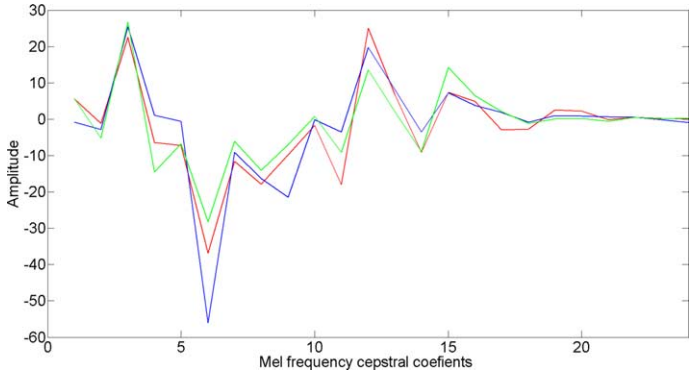

Fig. 16. Linear-frequency cepstral coefficients for ident.pilot signal disturbed by three GSM speech coders.

The artificial neural network (ANN) is a mathematical or computational model based on the principles of biological neural networks. It consists of an interconnected group of artificial neurons. In most cases an ANN is an adaptive system that changes its structure based on external or internal information that flows through the network during the learning phase (Fast Artificial Neural Network Library). The topology of the ANN-based GSM speech coder classifier and the corresponding training procedure are described in the following section.

7. Experiments and Results

In previous sections we described all steps needed to produce the feature vectors consisting of 24 linear frequency cepstral coefficients (LFCC). These feature vectors are finally used as input for the ANN-based classifier. For ANN training and evaluating of the speech coder identification results a set of 3584 identification pilot signals was constructed. These pilot signals were disturbed with different arbitrary signals in with various strengths, similar to the test signals that were used for empirical search of threshold G(controlled SNR decreasing). These signals were sent through all three GSM speech coders resulting in totally 10,752 disturbed pilot signals. Furthermore, we made 10 different time-offsets (silence insertion at the beginnings) of these signals to achieve different GSM frame synchronizations. With these changes we produced 107520 pilot signals, which were made in controlled environment. The last parts of signals were recorded in controlled real environment with Rohde&Schwartz equipment CMU200.

CMU200 is a special instrument used for universal testing of modern communication technologies, among them also GSM technologies (GSM400, 850, 900, 1800 and 1900) on various layers (RF, modulations, power and spectral measurements; Rohde/Schwarz CMU200 Universal Radio Communication Tester Datasheet, 2008). There is also ability to switch on/off different modules, such as VAD, DTX, individual speech coders, etc. This device has also its own radio frequency transmitter/receiver and for extra control there is a RF shielded box that was used for recording of identification pilot signals on different GSM terminals (mobile phones).

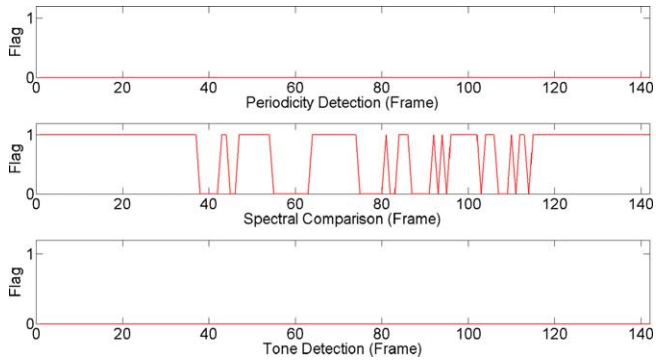

Fig. 17. Main VAD variables for 5 test pilot signals Spectral comparison, periodicity and tone detection.

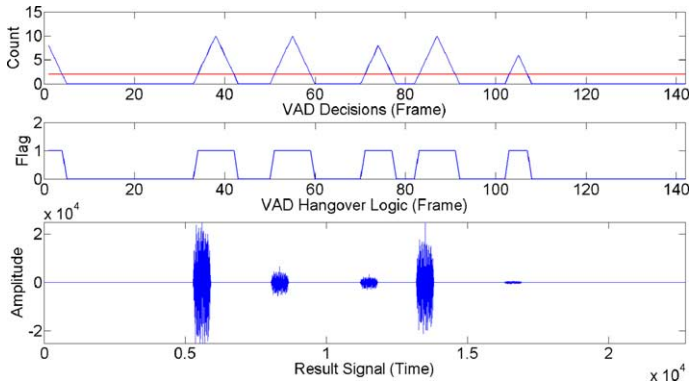

Fig. 18. Main VAD variables for 5 test pilot signals VAD decisions, hangover logic and result signal.

The structure of identification pilot signals is specially designed so that VAD doesn't recognize it as non-speech signal. For controlled testing and developing of appropriate identification pilot structure, special VAD testing tools were developed in MATLAB environment. Code implementation was checked with ETSI provided test vectors (ETSI EN 300 730 v7.0.1, 2000).

The results show that VAD always detects proposed pilot signals as speech signal and doesn't activate DTX. Figures 17 and 18 show the main VAD variables behaviour for 5 identification pilot signals with different amplitudes (see Fig. 19).

Figure 17 shows that VAD algorithm didn't find any tones or periodicities in the tested signal. Figure 18 presents VAD decisions and finally VAD hangover logic that is sent toward DTX. It is clear that VAD in the presence of all 5 versions of proposed pilot signals immediately recognizes signal as speech frame. These results prove also that the definition of proposed pilot signals assures a reliable transmission of the pilot signals through the speech channel of the GSM system.

With described cross-correlation algorithm, 72,945 identification pilot signals were found at threshold $G = 0.41$, as described in previous sections. The right diagram in Fig. 20 shows minimal SNR values for successful cross-correlation search of identifica-

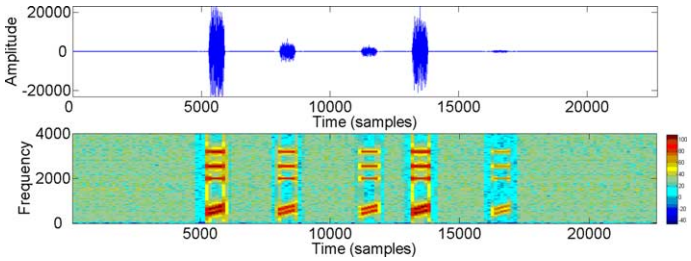

Fig. 19. Five different pilot signals for VAD test: waveform (above) and corresponding spectrogram (below).

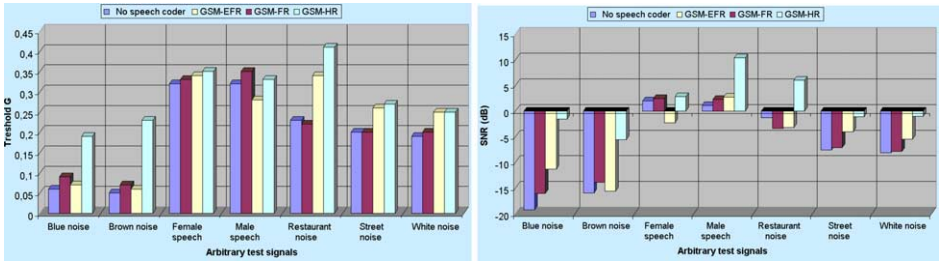

Fig. 20. Minimal SNR and treshold G for successful identification pilot search in different environments and speech coders.

tion pilot signal disordered with seven different environmental noises. It is evident from the diagram, that blue and brown noises are least critical in determination of exact position of pilot signals. This is mainly due to the spectral characteristics of these two noises which have a minor overlap with the spectrum of identification pilot signals. In opposition, the noises denoted as male, female, and restaurant have more important impact on identification pilot search procedure. However, even with these noises, the correct determination of pilot signals can be performed with SNR well below of 5 dB.

For indirect GSM speech coder identification an ANN with 4 layers has been constructed and trained. The training procedure is completed when the desired error falls below 0.0005 during the training epochs. The trained ANN has 24 neurons in input layer, 120 neurons in the first hidden layer, 60 neurons in the second hidden layer, and 4 neurons in the output layer. All together there are 9988 connections between 208 neurons (Fast Artificial Neural Network Library).

The results for indirect GSM speech coder identification show that the average GSM speech coder identification score is 98.72 %.

The accuracy of GSM-EFR coder identification is 98.85%, and the performances of GSM-FR, and GSM-HR coders are 98.71% and 98.61% respectively. Figure 21 presents the diagram of GSM speech coder identification performance comparison.

There exist several feature extraction and data classification algorithms in literature. However, little work has been done in explicit GSM speech coder identification. In Scholz *et al.* (2004) speech codec detection by spectral harmonic-plus-noise decomposition has been utilised. In this approach a short-time speech spectrum is decomposed into a har-

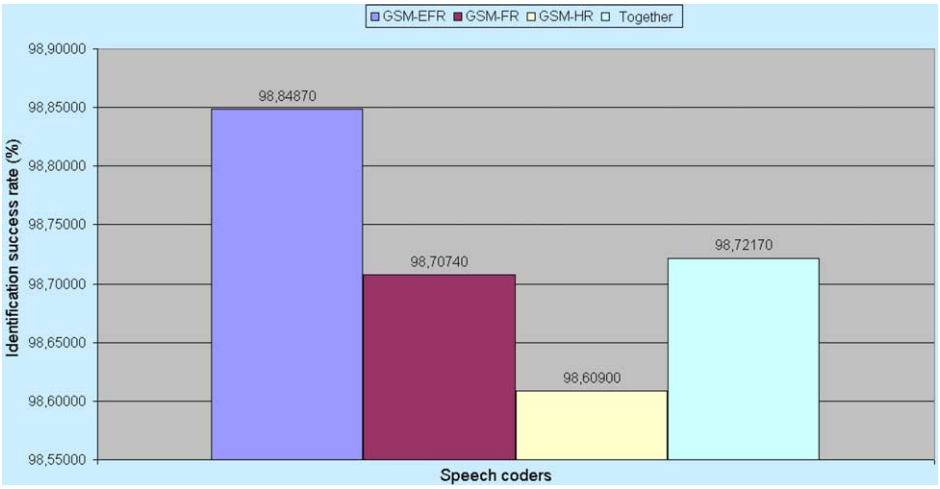

Fig. 21. Indirect GSM speech coder identification success.

monic and noise component. Then, the codec-specific distortions are revealed and thus the speech codec is identified. The authors reported an error rate of less than 8% when identifying five different codecs (not only GSM codecs). The main disadvantage of this algorithm is the fact, that the speech signal itself can be contaminated by the environmental additive noise prior to speech coding thus reducing the codec classification score. Another approach is presented in (Thorsten, 2004). Here only the GSM-FR is detected by analyzing the codec-specific spectral attenuation at 2700 Hz. The GSM-FR codec classification score of less than 5% is reported in preliminary results. The drawback of this method is that it is intended only for GSM-FR codec identification. In this paper presented GSM speech coder indirect identification algorithm can identify all three GSM speech codecs (GSM-EFR, GSM-FR, and GSM-HR) with a high score (accuracy above 98%) and is highly immune to environmental noise.

Several data classification algorithms and approaches can be utilised to perform various classification tasks. These approaches are hidden Markov models, Gaussian mixture models (GMM), Artificial neural networks (ANN) and others. Hidden Markov models are usually applied when the time component must be present in classification. This is usually the case with automatic speech recognition. However, when the data classification does not depend on the time component, then the GMM or ANN approaches are mostly used. In our case we decided to apply ANN for data classification. The main reason to use ANN is its strong ability to adapt to the nonlinear properties of GSM coder classification tasks. Namely, the ANN can more accurately model the specific spectral characteristics of different GSM codecs and perform precise separation of feature vectors in the multidimensional space. The advantage of ANN against GMM is also the ability to perform the training faster and more accurate using less training material. Moreover, the ANN training procedure is less affected by the errors in the training set. The choice of most appropriate feature vectors has also very strong impact on the classification performance. The cepstral coefficients are well proven in automatic speech recognition tasks. Since the

mel-warping of the frequency spectrum is only useful for speech analysis, we applied linear filterbank separation. The main advantage of the cepstral representation is its energy independence and decorrelation therefore it was adopted from the MFCC feature extraction. All these modification led to linear-frequency cepstral coefficients (LFCC) which are utilised in the proposed GSM speech coder identification task.

8. Conclusion

This paper presented a novel GSM speech coder indirect identification algorithm. The correct identification of GSM speech coder can be beneficially applied in several speech and signal processing applications. For example, the speech coder identification can be used in IVR (interactive voice response) systems incorporating automatic speech recognition to select different pre-trained acoustical models, adapted to the specific tempo-spectral characteristics of particular GSM speech coders (Rotovnik *et al.*, 2007). Furthermore, the GSM speech coder identification can be used for speech quality assessment purposes to evaluate the quality of service of particular segment of considered GSM network. Nevertheless, the GSM speech coder identification based on sending novel pilot signals through the speech channel of the GSM system can be applied to adapt particular components (modulation/demodulation method, bandwidth etc. (Sun, 2001; Rohde/Schwarz CMU200 Universal Radio Communication Tester Datasheet, 2008) of a GSM speech channel data transmission systems (like mobile transaction system M-Pay (Ultra M-Pay Patent 1 and 2, 2002)) thus making these systems more robust and reliable. Furthermore, the proposed energy-independent procedure for accurate determination of the position of the pilot signal can be used for data synchronisation purposes thus making the coherent OFDM (Orthogonal Frequency Division Multiplexing) data transmission method a possible choice in above mentioned data transmission methods (Batra *et al.*, 2004; Gurprakash and Arokiaswami, 2003; Schmidl *et al.*, 1997; Zhengdao and Giannakis, 2000; Rohde/Schwarz CMU200 Universal Radio Communication Tester Datasheet, 2008).

The proposed GSM speech coder indirect identification algorithm consists of several parts. First, very specific tempo-spectral characteristics of the pilot signal have been defined. The constructed pilot signal is then sent through the unknown GSM speech coder, where it becomes uniquely modified. At the receiver side a special cross-correlation based method is applied to find the precise position of the pilot signal in the received signal waveform. Next, the LFCC (linear-frequency cepstral coefficients) are extracted from the received pilot signal. Finally, the pre-trained ANN classifier is applied to determine the most probable GSM speech coder based on the previously computed LFCC vector.

The experimental results show that the overall GSM speech coder identification accuracy is 98.72% thus making the described procedure an ideal choice for its applicative usage in several applications.

References

Batra, A. *et al.* (2004). Multi-band OFDM physical layer proposal for IEEE 802.15 Task Group 3a. In: *IEEE Document P802.15-04/0493r1*, Texas Instruments, Vol. 9.

Bingham, J.A.C. (2000). *ADSL, VDSL, and Multicarrier Modulation*. Wiley, New York, pp. 160–183.

Chow, J.S., Tu, J.C., Cioffi, J.M. (1991). A discrete multitone transceiver system for HDSL applications. *IEEE J. Select. Areas Commun.*, 8, 895–908.

ETSI EN 300 730 v7.0.1 (2000). Digital cellular telecommunications system-voice activity detector for enhanced full rate speech traffic channels (GSM 06.82). *ETSI Standard Documentation*.

Fast Artificial Neural Network Library. `http://leenissen.dk/fann/`, accessible August 2008.

GSM 06.10 Full Rate (FR) Vocoder (1996). Regular Pulse Excitation – Long Term Prediction Linear Predictive Coder (RPE-LTP). `http://www.etsi.com`.

GSM 06.60 Enhanced Full Rate (EFR) Vocoder (1997). Algebraic-Code-Excited Linear Predictive (ACELP). `http://www.etsi.com`.

GSM 06.20 Half Rate (HR) Vocoder (1998). Vector-Sum Excited Linear Prediction (VSELP). `http://www.etsi.com`.

Gurprakash, S., Arokiaswami, A. (2003). OFDM modulation study for a radio-over-fiber system for wireless LAN (IEEE 802.11a). In: *Proceedings ICICS-PCM*, Singapore.

Hanzo, L., Somerville, F.C.A, Woodard, J.P. (1999). *Voice Compression and Communications*. IEEE Press/Wiley. pp. 54–72.

Ibars, C., Bar-Ness, Y. (2001). Comparing the performance of coded multiuser OFDM and coded MC-CDMA over fading channels. In: *Proceedings IEEE GLOBECOM Conference*, Vol. 2, pp. 881–885.

Lehtonen, K. (2004). *Digital Signal Processing and Filtering – GSM Codec*. Helsinki University of Technology.

Mezgec, Z., Chowdhury, A., Kotnik, B. (2009). Implementation of PCCD-OFDM-ASK robust data transmission over GSM speech channel. *Informatica*, 20(1), 51–78.

Rohde/Schwarz CMU200 Universal Radio Communication Tester Datasheet. `http://www2.rohde-schwarz.com/product/CMU200.html`, accessible August 2008.

Rotovnik, T., Sepesy Maucec, M., Kacic, Z. (2007). Large vocabulary continuous speech recognition of an inflected language using stems and endings. *Speech Commun.*, 49(6), 437–452.

Schmidl, T.M. *et al.* (1997). Robust frequency and timing synchronization for OFDM. *IEEE Trans. Commun.*, 45, 1613–1621.

Scholz, K., Leutelt, L., Heute, U. (2004). Speech-codec detection by spectral harmonic-plus-noise decomposition. *Proc. Sig., Syst. Comput.*, 2, 2295–2299.

Scourias, J. (1995). *Overview of the Global System for Mobile Communications*. `http://www.shoshin.uwaterloo.ca/~jscouria/GSM/gsmreport.html`, accessed 4 August 2008.

Sun, Y. (2001). Bandwidth-efficient wireless OFDM. *IEEE J. Selec. Areas Commun.*, 19(11), 124.

Thorsten, L. (2002). Comfort noise detection and GSM-FR codec detection for speech-quality evaluations in telephone networks. In: *Proceedings ICSLP*, pp. 309–312.

Ultra M-Pay Patent 1 and 2 (2002). WO 02/33669, WO 03/088165.

Xiong, F. (2006). *Digital Modulation Techniques*. Artech House Publishers, pp. 20–142.

Zhengdao, W., Giannakis, G.B. (2000). Wireless multicarrier communications. *IEEE Sig. Proc. Mag.*, 5, 1220.

R. Svečko received his MSc degree in electrical engineering in 1984 and PhD in control in 1989 both at University of Maribor, Slovenia. He works as an associates professor and researcher at University of Maribor, Faculty of Electrical Engineering and Computer Science, Slovenia.

A. Chowdhury received his MSc degree in electrical engineering in 1997 and PhD in robust control in 2001 both at University of Maribor, Slovenia. Since 2005 he is a head of Research Center at Ultra and beside this he is still working at University of Maribor, Faculty of Electrical Engineering and Computer Science, Slovenia.

B. Kotnik obtained his BSc degree in electrical engineering in 2000 and PhD in automatic speech recognition in 2004 both at University of Maribor, Slovenia. He is a developer and researcher in the fields of digital signal processing, digital modulation and demodulation algorithms, and statistical methods for data classification at Ultra Research Center, Maribor, Slovenia.

Z. Mezgec received his BSc degree in electrical engineering in 2004 and PhD in adaptive communications in 2008 both at University of Maribor, Slovenia. He started his research work in 2004 and since 2008 he has been working as chief of embedded system development at Ultra d.o.o. Research Center, Maribor, Slovenia.

GSM kalbos koderio netiesioginio identifikavimo algoritmas

Rajko SVEČKO, Bojan KOTNIK, Amor CHOWDHURY, Zdenko MEZGEC

Straipsnyje pateikiamas GSM kalbos koderio netiesioginio identifikavimo algoritmas, grįstas originalaus identifikavimo kontrolinio signalo pasiuntimu per GSM kalbos kanalą. Kiekviena GSM posistemė iškraipo identifikavimo kontrolinį signalą, tuo tarpu kalbos koderis vieninteliu būdu pakeičia pasiūlyto kontrolinio signalo laikines-spektrines charakteristikas. Kalbos koderio identifikavimo algoritmas identifikuoja kalbos koderį naudodamas robastinę tiesinės dažnių skalės kepstro koeficientų požymių išskyrimo procedūrą ir greitus dirbtinius neuroninius tinklus. Pirmas kalbos koderio identifikavimo algoritmo žingsnis yra identifikavimo kontrolinio signalo tikslios padėties nustatymas naudojant normalizuotos tarpusavio koreliacijos metodą. Kitas etapas yra įėjimo signalo dauginimas iš lango funkcijos laiko srityje, kad galima būtų atlikti kiekvieno įėjimo signalo kadro ir lango funkcijos spektrų sąsūką. Tolimesniame žingsnyje yra naudojama trumpalaikė Greitoji Furjė transformacija, kuri sukuria kiekvieno padauginto iš lango funkcijos kadro amplitudinį spektrą. Toliau yra atliekamas triukšmo pašalinimas, naudojant spektro išlyginimu grįstą spektro atėmimą. Paskutiniuose etapuose atliekama filtracija dažnių srityje ir naudojama diskretinė kosinuso transformacija, kad kadrams kaip rezultatas būtų gauti 24 nekoreliuoti kepstro koeficientai. Kalbos koderio identifikavimas yra baigiamas naudojant greitą dirbtinių neuronų tinklo klasifikatorių, kurio įėjimo požymių vektorius yra 24 tiesinės dažnių skalės kepstro koeficientai. Tinklo išėjimo rezultatas yra identifikuotas kalbos koderis. GSM kalbos koderio netiesioginio identifikavimo įvertinimui buvo naudojamos GSM ETSI tikslaus bitų skaičiaus realizacijos. Be to, buvo sukurta keletas papildomų įrankių. Šie įrankiai buvo naudojami modeliuoti ir valdyti įvairias GSM veikimo sąlygas. Galutiniai rezultatai rodo, kad pasiūlytas algoritmas identifikuoja GSM-EFR kalbos koderį 98,85GSM-FR kalbos koderį 98,71pasiekti esant įvairių tipų aplinkos triukšmams ir netgi esant labai mažam signalo/triukšmo santykiui.