Radosław WEYCHAN, Tomasz MARCINIAK

Poznań University of Technology, Chair of Control and System Engineering, Division of Signal Processing and Electronic Systems

Analysis of differences between MFCC after multiple GSM transcodings

Streszczenie. *Artykuł prezentuje rezultaty badań nad wpływem wielokrotnego transkodowania sygnału audio próbkowanego z szybkością 8 kSps dla standardu GSM, oraz 16 kSps. Przeanalizowane zostały uzyskane różnice między współczynnikami MFCC, otrzymane w wyniku kolejnych transkodowań. Głównym celem porównania jest sprawdzenie możliwości separacji danych oraz detekcji wykorzystywanego w transmisji kodera GSM. Do eksperymentu wykorzystana została baza nagrań sygnału mowy TIMIT, transkodowana czterokrotnie przez kodery GSM. Przeanalizowane zostały możliwości detekcji typu kodera na podstawie różnic między aproksymatami krzywoliniowymi błędów współczynników MFCC.* ***(Analiza wpływu wielokrotnego transkodowania GSM na różnice między współczynnikami MFCC).***

Abstract*: This paper presents results of studies on the effects of multiple speech transcoding operations in the case of GSM standard with 8 kSps and 16 kSps sampling rate. Differences between the MFCC coefficients obtained by successive transcoding were considered. The aim of comparisons is to check the possibility for separation and detection of the used GSM encoder. During the research we used the TIMIT database recordings, transcoded four times by GSM codecs. A possibility of encoder type detection was analyzed based on differences between the curvilinear approximations of the MFCC coefficient errors.*

Słowa kluczowe: GSM, transkodowanie, MFCC, kodowanie mowy
Keywords: GSM, transcoding, MFCC, speech encoding

Introduction

GSM speech coding is an operation that introduces distortion into a useful signal. These changes are critical in the speaker identification and speech recognition systems [1] [2]. Information about the type of speech codec can have a significant impact on the effectiveness of such systems, because we can use appropriate reference recordings during the training phase and calculate the dedicated models for detected type of GSM encoding [3]. Proposed in [4] GSM encoding detection algorithm exploits the fact that the every next encoding introduces less distortion into the speech signal [5]. Calculation of the mean square error between the coefficients of MFCC (*mel frequency cepstral* coefficients) allows to detect whether the audio signal has already undergone coding operation. The idea of the detection algorithm is shown in Fig. 1.

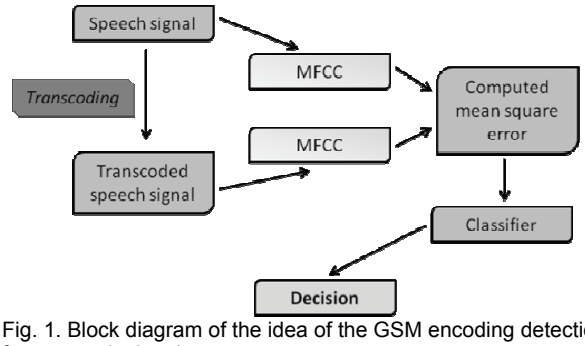

Fig. 1. Block diagram of the idea of the GSM encoding detection from speech signal

Using only one parameter, however, the correct detection of the GSM codec is not always possible due to the overlapping ranges of MSE (*mean square errors*) distributions. In this article we propose to extend the classifier to 12 separate errors between the MFCC coefficients. Analysis of the relationship between MFCC transcoded recordings allows detection of the used GSM codec. In opposite to systems described in [6] and [7], proposed solution converts the whole speech signal, including unvoiced, and especially voiced part, which are the most important in this case. In comparison to encoders used in previously mentioned papers, presented tests include also non-ACELP algorithms, like Full Rate encoder.

Determination of MSE

Software used during experiments consists of transcoding modules, and a program to determine the MSE between the MFCC coefficients.

Transcoding software consists of four independent modules, which realizes particular types of GSM standard coders/decoders: full rate (FR), enhanced full rate (EFR), adaptive multi-rate (AMR) and half rate (HR). The first one was implemented in Matlab / Simulink based on [8] and extended by the batch processing. The next three coders were written in ANSI C using example implementations [9] [10] [11], compiled using Dev C++. They have been integrated into one program, and extended by the ability to support of .wav files and also batch processing. File names for transcoding operation are downloaded from an external text document, both for implementation in Matlab language as well as in ANSI C. A general diagram of Figure 2 shows the transcoding process.

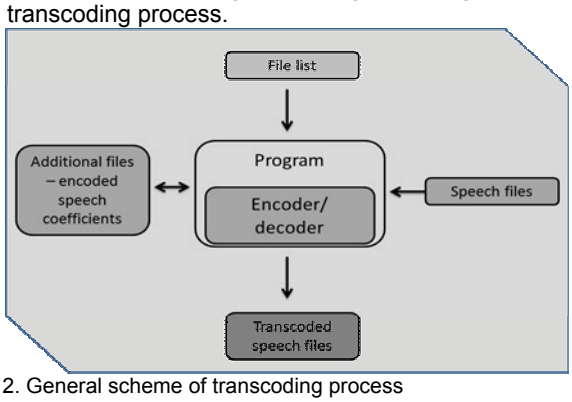

Fig 2. General scheme of transcoding process

Calculation of errors between MFCC values is realized in Matlab environment using the package Voicebox [12]. It contains a set of functions which are helpful in speech signal processing. During experiment we used 12 coefficients for each data frame of 256 samples, which is equivalent to 16 ms in the case of 16000 samples/second sampling rate. Number of filters in the filter bank was set to 29, at the lower frequency 0 Hz, while the highest 4000 Hz.

Transcoding operation was tested for speech recordings from TIMIT database [13]. It contains recordings of the speech signal of 630 speakers presenting eight main dialects of English, each of them says 10 sequences. This database is used primarily to test the efficiency of speech

recognition algorithms. The sequences of speech were recorded with a resolution of 16 bits and sampling rate 16000 samples / second. In our experiment, the sampling rate was converted to the value of 8000 samples / second.

Experimental results

Transcoding operation have been done four times on full database of 6300 files by each of transcoders, separately for both sampling frequencies. 214200 files was created, including original recordings. Between each of transcoding stages, mean MSE (*mean root square error*) values for GSM encoders have been computed. A block scheme of experiment is presented on Fig. 3. The result of each comparison is the set of 12 features, which are errors between each subsequent MFCC coefficients.

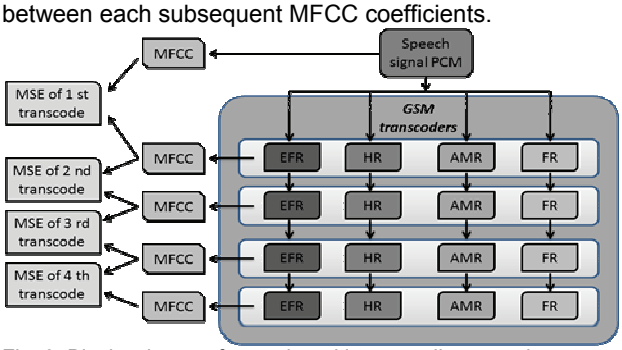

Fig. 3. Block scheme of speech multi-transcoding experiment

In order to determine curvilinear approximations from above obtained results, it has been established, that error value decreases exponentially with increasing the number of coefficient. Therefore each of results tends to hyperbolic regression function, which is described by:

$$\hat{y}_i = a + b\frac{1}{x_i} \tag{1}$$

where

$$b = \frac{\sum_{i=1}^{n}\frac{1}{x_i}\cdot y_i - \frac{1}{n}\sum_{i=1}^{n}\frac{1}{x_i}\sum_{i=1}^{n} y_i}{\sum_{i=1}^{n}\frac{1}{x_i^2} - \frac{1}{n}\left(\sum_{i=1}^{n}\frac{1}{x_i}\right)^2} \tag{2}$$

$$a = \frac{1}{n}\left(\sum_{i=1}^{n} y_i - b\sum_{i=1}^{n}\frac{1}{x_i}\right) \tag{3}$$

n - amount of MFCC coefficients, x_i - number of MFCC coefficient (i=1, 2…12), y_i - value of MSE between MFCC coefficients (*i*=1, 2…12)

Calculation of approximation using curvilinear regression allows to estimate influence of speech multi-transcoding on particular coefficients for various encoder types, and also dependency between them.

Adaptive multi-rate encoder

First analyzed GSM encoder is adaptive multi-rate (AMR). It allows to control rate of binary stream depending on occupancy of system, form 4.75 to 12.2 kbps. This encoder uses Algebraic-Code-Excited Linear Predictive (ACELP) algorithm for speech compression. All possible modes of binary stream were used to encode speech signal in experiment. The sequence of mentioned modes is presented in Table 1.

Table. 1. Periodic sequence of modes for particular speech frames encodings

Frame number	1	2	3	4	5	6	7
Mode [kbps]	12.2	10.2	7.95	7.4	6.7	5.9	5.15
Frame number	8	9	10	11	12	13	14
Mode [kbps]	4.75	5.15	5.9	6.7	7.4	7.95	10.2

Figure 4 and 5 presents mean square errors between particular MFCC coefficients for subsequent transcoding. Error values decrease with the increase in the number of coefficient, regardless of transcoding numbers. Derived curves show, that for subsequent encodings and decodings of speech signal possibility to extract and parameterize useful signal decreases, therefore the greatest difference between lines takes place in case of original speech signal processing. In Fig. 4 and 5, axis of ordinates has been scaled logarithmically.

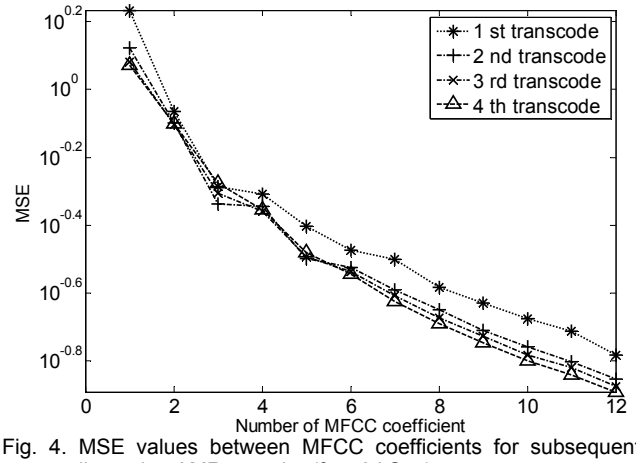

Fig. 4. MSE values between MFCC coefficients for subsequent transcoding using AMR encoder (fs = 8 kSps)

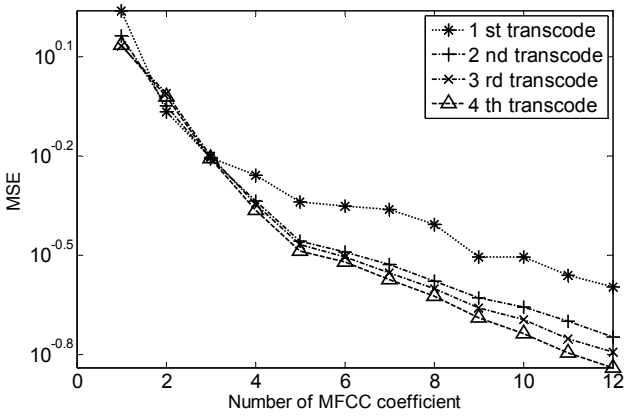

Fig. 5. MSE values between MFCC coefficients for subsequent transcoding using AMR encoder (fs = 16 kSps)

Approximations of computed curves with the use of curvilinear regression are presented respectively in Figures 6 and 7 and represented by equations given in Tables 2 and 3.

Table 2. Approximation equations of MFCC errors for fs = 8 kSps

Number of transcoding	Approximation equation
1	$\hat{y}_1 = 0.0479 + 1.6483\frac{1}{x}$ (4)
2	$\hat{y}_2 = 0.0623 + 1.3055\frac{1}{x}$ (5)
3	$\hat{y}_3 = 0.0761 + 1.1988\frac{1}{x}$ (6)
4	$\hat{y}_4 = 0.0780 + 1.1843\frac{1}{x}$ (7)

Table 3. Approximation equations of MFCC errors for fs = 16 kSps

Number of transcoding	Approximation equation
1	$\hat{y}_1 = 0.1546 + 1.5456\frac{1}{x}$ (8)
2	$\hat{y}_2 = 0.0910 + 1.4209\frac{1}{x}$ (9)
3	$\hat{y}_3 = 0.0920 + 1.3707\frac{1}{x}$ (10)
4	$\hat{y}_4 = 0.0722 + 1.3965\frac{1}{x}$ (11)

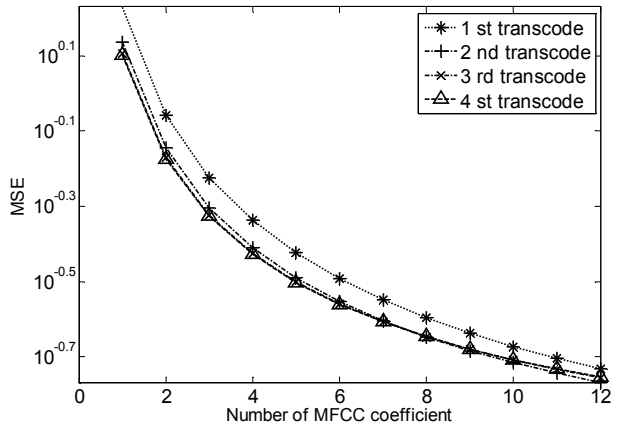

Fig. 6. Curvilinear approximations of computed curves for AMR encoder (fs= 8 kSps)

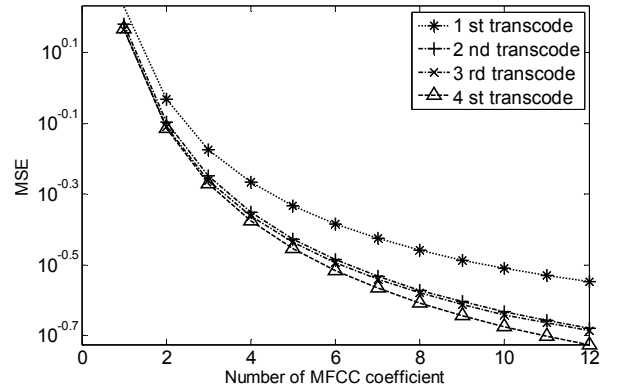

Fig. 7. Curvilinear approximations of computed curves for AMR encoder (fs = 16 kSps)

Transcodings number 2, 3 and 4 gives very similar results, as evidenced by equations (5) - (7) and (9) – (11). An interesting result is growing saturate value, while slope decreases for 8000 samples / second sampling rate in opposite to 16000 samples / second . It results from high influence of 3rd , 4th and 5th MFCC values, which increase with number of transcoding (cf., Fig. 4). Error intervals and slopes in computed equations are similar. A greater influence on higher MFCC can be observed for the first transcoding in case of 16000 samples / second sampling frequency.

Enhanced full rate encoder

Enhanced full rate (EFR) encoder uses the same algorithm as AMR – ACELP, however the rate of binary stream is constant and amounts to 12,2 kbps. EFR encoder have higher computational complexity than FR, which in a mobile device can potentially result in the increase of energy consumption.

Figure 8 and 9 presents computed errors between particular MFCC coefficients for subsequent transcodings, respectively for GSM standard 8 kSps and 16 kSps sampling frequency. Generally, errors decrease with the number of coefficient increases, as in case of AMR encoder. It can be noticed that error values saturate with the fourth transcoding in both cases. Error intervals are very similar as well. It can be observed, that EFR introduces more distortion to MFCC number 3-12, in the case of 16 kSps sampling rate.

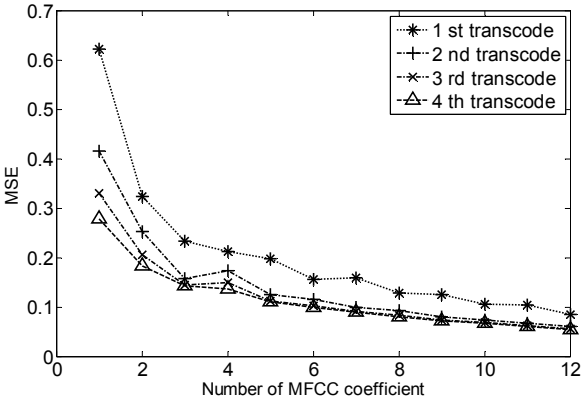

Fig. 8. MSE values between MFCC coefficients for subsequent transcodings using EFR encoder (fs = 8 kSps)

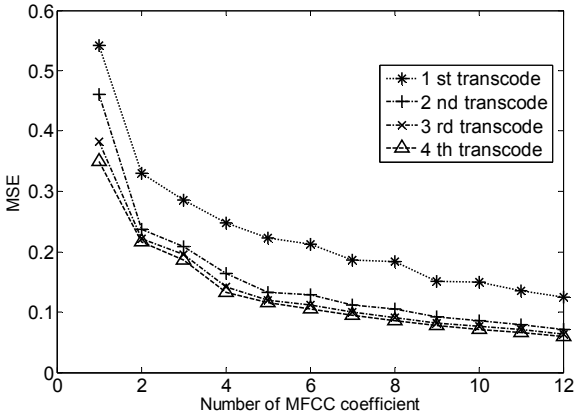

Fig. 9. MSE values between MFCC coefficients for subsequent transcodings using EFR encoder (fs = 16 kSps)

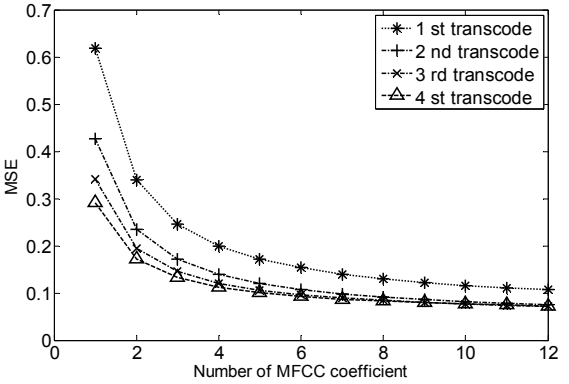

Fig. 10. Curvilinear approximates of curves computed for EFR encoder (fs = 8 kSps)

Figure 10 and 11 shows approximations of plotted lines for both sampling frequencies, represented respectively by equations (12)-(15) and (16)-(19) in Table 4 and 5. Similar to AMR encoder (GSM standard 8 kSps), the saturation increases in the case of second and next transcodings because of using the same algorithm. Slope, which is more than twice less than in case of AMR encoder in both cases, is the main coefficient to distinguish calculated curves,. Presented error intervals are very similar. Also greater influence on MFCC coefficients number 3-12 can be observed during the first transcoding with 16 kSps.

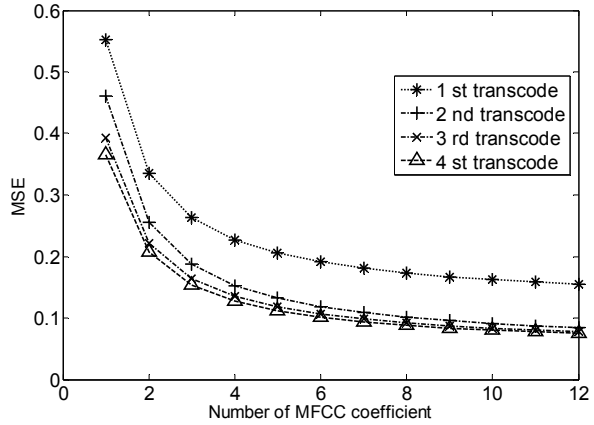

Fig. 11. Curvilinear approximates of curves computed for EFR encoder (fs = 16 kSps)

Table 4. Approximation equations of MFCC errors for fs = 8 kSps

Number of transcoding	Approximation equation
1	$\hat{y}_1 = 0.0603 + 0.5577\frac{1}{x}$ (12)
2	$\hat{y}_2 = 0.0435 + 0.3836\frac{1}{x}$ (13)
3	$\hat{y}_3 = 0.0471 + 0.2941\frac{1}{x}$ (14)
4	$\hat{y}_4 = 0.0526 + 0.2390\frac{1}{x}$ (15)

Table 5. Approximation equations of MFCC errors for fs = 16 kSps

Number of transcoding	Approximation equation
1	$\hat{y}_1 = 0.1187 + 0.4341\frac{1}{x}$ (16)
2	$\hat{y}_2 = 0.0501 + 0.4108\frac{1}{x}$ (17)
3	$\hat{y}_3 = 0.0491 + 0.3437\frac{1}{x}$ (18)
4	$\hat{y}_4 = 0.0484 + 0.3164\frac{1}{x}$ (19)

Full rate encoder

Full rate (FR) encoder is the first speech compression algorithm used in GSM mobile system. It uses Regular Pulse Excitation - Long Term Prediction (RPE-LTP) algorithm, producing binary stream of 13 kbps.

Determined MSE between particular MFCC coefficients for subsequent transcoding operations are presented in Fig. 12 and 13. Logarithmic scale of axes enables a better result presentation. Similarly to previous results, the error values decrease with following encodings and decodings. During the 1st transcoding a significant influence on 6th MFCC can be observed when speech is sampled at 8 kSps, and less significant influence on the 9th coefficient at 16 kSps. It can be observed, that saturation ratio of the second transcoding is much lower than the first one in both cases, being the next parameter to distinguish this encoder from others. However, general differences between subsequent MFCC errors for every transcoding are much smaller.

Curvilinear approximations are presented in Fig. 14. And 15., marked respectively with equations (20)-(23) and (24)-(27) in Tables 6 and 7. Differences against previous encoders are being seen in error values as well as in the dependence between them. It is confirmed by slopes and expected saturation of error values. As in previous encoders, the first transcoding influences the signal more than in the case of 16 kSps sampling frequency.

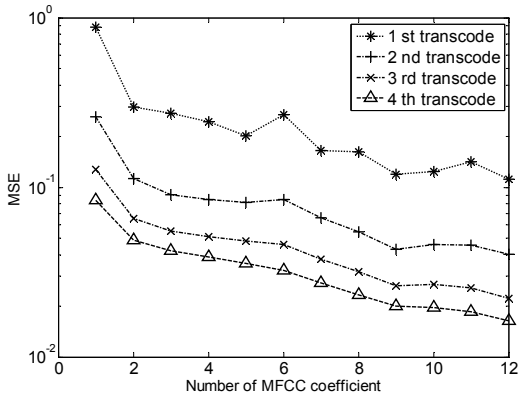

Fig. 12. MSE values between particular MFCC coefficients for subsequent transcodings using FR encoder (fs= 8 kSps)

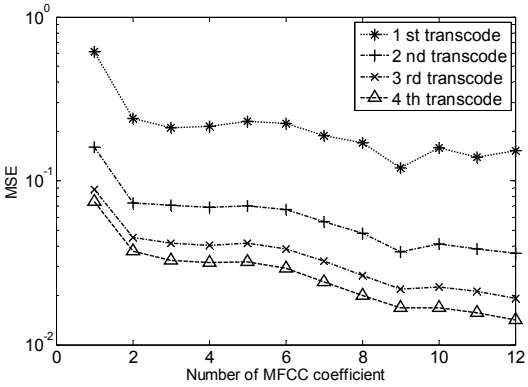

Fig. 13. MSE values between particular MFCC coefficients for subsequent transcodings using FR encoder (fs= 16 kSps)

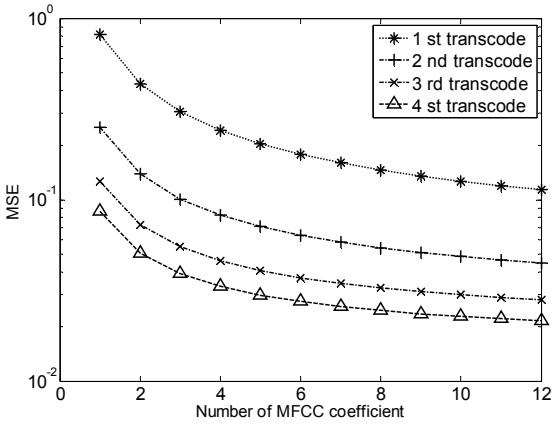

Fig. 14. Curvilinear approximates of computed curves for FR encoder (fs= 8 kSps)

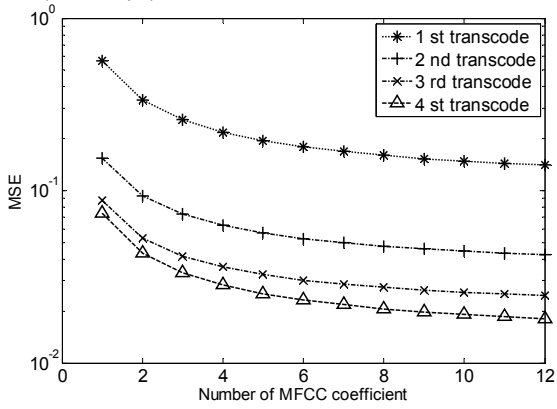

Fig. 15. Curvilinear approximates of computed curves for FR encoder (fs= 16 kSps)

Table 6. Approximation equations of MFCC errors for fs = 8 kSps

Number of transcoding	Approximation equation
1	$\hat{y}_2 = 0.0498 + 0.7685\frac{1}{x}$ (20)
2	$\hat{y}_2 = 0.0263 + 0.2238\frac{1}{x}$ (21)
3	$\hat{y}_3 = 0.0193 + 0.1073\frac{1}{x}$ (22)
4	$\hat{y}_4 = 0.0157 + 0.0702\frac{1}{x}$ (23)

Table 7. Approximation equations of MFCC errors for fs = 16 kSps

Number of transcoding	Approximation equation
1	$\hat{y}_2 = 0.1017 + 0.4648\frac{1}{x}$ (24)
2	$\hat{y}_2 = 0.0323 + 0.1221\frac{1}{x}$ (25)
3	$\hat{y}_3 = 0.0188 + 0.0687\frac{1}{x}$ (26)
4	$\hat{y}_4 = 0.0130 + 0.0609\frac{1}{x}$ (27)

Half rate encoder

Fourth algorithm of GSM speech encoding used in our researches is half rate (HR). It uses VSELP (Vector-Sum Excited Linear Prediction) algorithm, generating binary stream of 5,6 kbps.

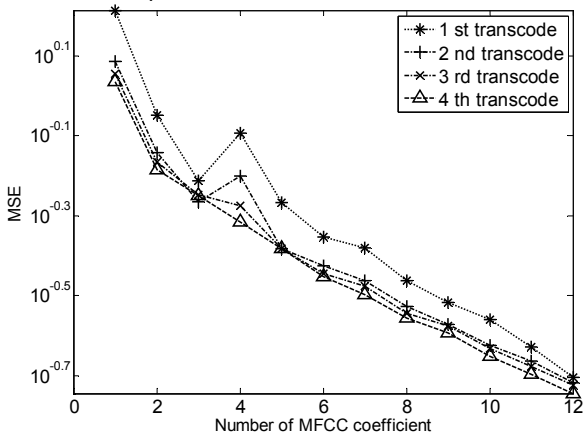

Fig. 16. MSE values between MFCC coefficients for subsequent transcoding using HR encoder (fs = 8 kSps)

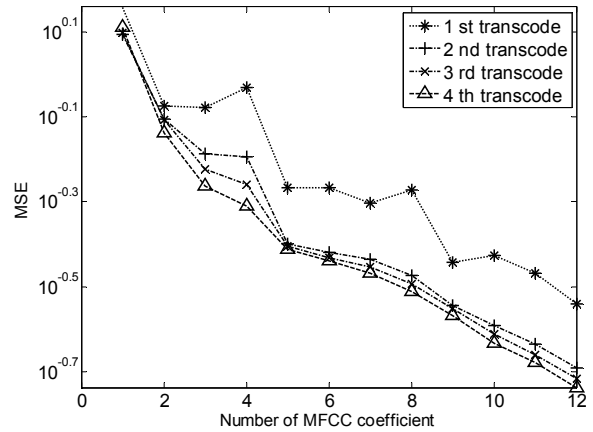

Fig. 17. MSE values between particular MFCC coefficients for subsequent transcoding using HR encoder (fs = 16 kSps)

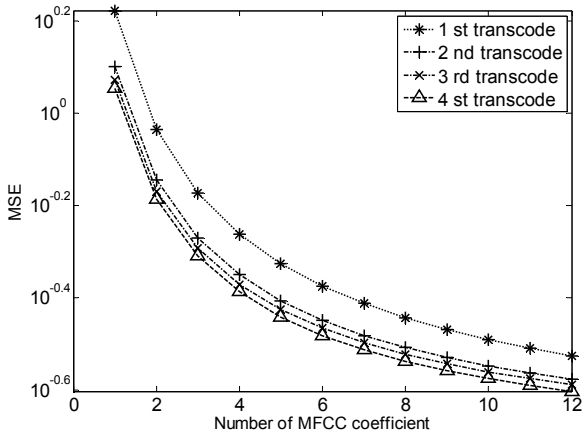

Fig. 18. Curvilinear approximates of computed curves for HR encoder (fs = 8 kSps)

Differences between computed MFCC coefficients are presented in Fig. 16 and 17. This algorithm generates MSE errors similar to AMR encoder; major influence is observed for the fourth MFCC coefficient in the case of 8 kSps, and for the fourth and eighth in the case of 16 kSps. Error values decreases with number on transcoding trending to saturate in the fourth curve neighborhood. The first transcoding introduces more distortion into the speech signal (16 kSps) than in previous cases.

Approximation of given results are presented on Fig. 18 and 19 with particular curve equations (28)–(31) and (32)–(35) in tables 8 and 9. Expected saturate values are similar to AMR encoder, but saturation is the major coefficient that allows to distinguish these approximations from AMR encoder. Similarly it refers to both sampling cases besides the first one.

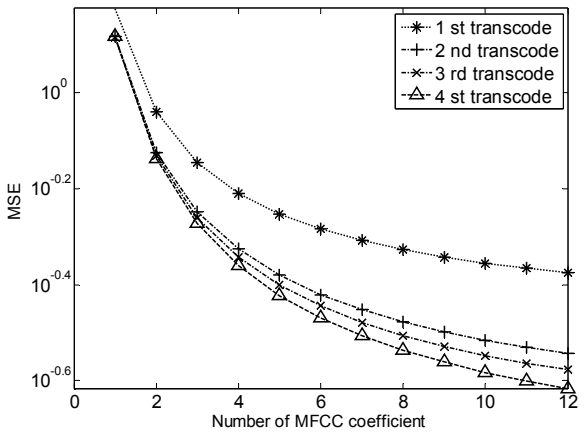

Fig. 19. Curvilinear approximates of computed curves for HR encoder (fs = 16 kSps)

Table 8. Approximation equations of MFCC errors for fs = 8 kSps

Number of transcoding	Approximation equation
1	$\hat{y}_1 = 0.1728 + 1.4946\frac{1}{x}$ (28)
2	$\hat{y}_2 = 0.1743 + 1.0857\frac{1}{x}$ (29)
3	$\hat{y}_3 = 0.1739 + 1.0047\frac{1}{x}$ (30)
4	$\hat{y}_4 = 0.1692 + 0.9604\frac{1}{x}$ (31)

Table 9. Approximation equations of MFCC errors for fs = 16 kSps

Number of transcoding	Approximation equation
1	$\hat{y}_2 = 0.3243 + 1.1715\frac{1}{x}$ (32)
2	$\hat{y}_2 = 0.1939 + 1.1118\frac{1}{x}$ (33)
3	$\hat{y}_3 = 0.1701 + 1.1345\frac{1}{x}$ (34)
4	$\hat{y}_4 = 0.1451 + 1.1640\frac{1}{x}$ (35)

Conclusions and further work

In this paper the influence of speech multi-transcoding on the MFCC coefficients has been presented. The relationship between them and the expected saturate values have also been computed. Curves calculated with the use of the curvilinear regression show, that extension of classifier to 12 coefficients gives potential possibilities to determine a type of used encoder type in case of large variation between them. It has also been confirmed, that MSE values between MFCC coefficients decrease with number of transcoding, irrespective of encoder type. This trend is independent from sampling frequency, however computed errors are grater in case of the first transcoding when input signal is sampled with 16 kSps.

We plan to use the calculated curves approximation and the classifier enhanced to 12 coefficients to the GSM encoder type detection. Comparing the obtained curves and the distributions of the mean square errors between the MFCC coefficients with the models of the GSM coder seem to be the proper way to detect the type GSM encoder used in the voice transmission. We also plan to improve effectiveness of encoder detection by silence removal from the speech and maximizing the content of the useful signal in the tested samples.

This work was supported by INDECT and partly by DS 2012 project.

REFERENCES

[1] Grassi S., Besacier L., Dufaux A., Ansorge M., Pellandini F., Influence of gsm speech coding on the performance of text-independent speaker recognition, *Proc. of EUSIPCO 2000*, (2000), 437-440

[2] Lilly B.T., Paliwal K.K., Effect of speech coders on speech recognition performance. Proc. of ICSLP, (1996), 2344-2347

[3] Nemat. S. Abdel Kader, Effect of GSM system on text independent speaker recognition, *Journal of Theoretical and Applied Information Technology*, June 2008, 442-449

[4] Dąbrowski A., Drgas S., Marciniak T., Detection of GSM speech coding for telephone call classification and automatic speaker recognition, *Proc. of ICSES 2008 International Conf. on Signals and Electronic Systems*, (2008), 415-418

[5] Radosław Weychan, Tomasz Marciniak, Adam Dąbrowski, Influence of signal segmentation in GSM coding detection, *Elektronika*, (2010), nr 3, 94-97

[6] Zhou J., Garcia-Romero D., Espy-Wilson C., Automatic Speech Codec Identification with Applications to Tampering Detection of Speech Recordings, *INTERSPEECH 2011*,(2011), 2533-2536

[7] Scholz K., Leutelt L., Heute U., Speech-Codec Detection by Spectral Harmonic-Plus-Noise Decomposition, Proc of Systems and Computers Conf., Vol.2 (2004), 2295 - 2299

[8] Full rate encoder documentation http://www.3gpp.org/ftp/Specs/archive/06_series/06.60/0660-801.zip

[9] Enhanced full rate ANSI C source code http://www.3gpp.org/ftp/Specs/archive/26_series/26.073/26073-900.zip

[10] Adaptive multi-rate ANSI C source code http://www.3gpp.org/ftp/Specs/archive/06_series/06.53/0653-801.zip

[11] Half rate ANSI C source code http://www.3gpp.org/ftp/Specs/archive/06_series/06.06/0606-801.zip

[12] Speech processing toolbox for Matlab http://www.ee.ic.ac.uk/hp/staff/dmb/voicebox/voicebox.html

[13] TIMIT database website http://www.ldc.upenn.edu/Catalog/CatalogEntry.jsp?catalogId=LDC93S1

***Authors**: mgr inż. Radosław Weychan, dr inż. Tomasz Marciniak, Poznań University of Technology, Chair of Control and Systems Engineering, Division of Signal Processing and Electronic Systems, ul. Piotrowo 3a, 60-965 Poznań, Poland, e-mail: tomasz.marciniak@put.poznan.pl*