TRANSMULTIPLEXING SYSTEM FOR COMPRESSION OF SELECTED SIGNALS

M. ZIÓŁKO[1], A. DZIECH[2], R. BARAN[3], P. SYPKA[1], B. ZIÓŁKO[1]
[1]Department of Electronics, AGH University of Science and Technology, POLAND
e-mail: {ziolko, bziolko, sypka}@agh.edu.pl
[2]Department of Telecommunications, AGH University of Science and Technology, POLAND
e-mail: dziech@kt.agh.edu.pl
[3] Department of Electronics and Intelligent Systems, Kielce University of Technology, POLAND,
e-mail: ketrb@eden.tu.kielce.pl

Abstract: - In this paper efficient method of compression using multiplexing for two different classes of signals – contour and speech – is proposed. For a contour the contour signals in the form of two 1-D signals are combined for a transmission by a single channel and recovered by a receiver. Constraints on the synthesis and analysis FIR filters are imposed to achieve a perfect reconstruction. The optimization method of filter design for the 4-channel transmultiplexer is presented. The frequency properties of combined signal were verified to choose the compression method. For a speech signal it uses the speech recognition system to convert the voice signal into its transcripted form. Next, a speech synthesizer is used to reconstruct speech on the receiver side. Integer filters are used to realize perfect reconstruction in the transmultiplexer system. Although such system destroys the individual speech features, it provides an high compression.

Key-Words: - Transmultiplexing, Compression, Filter Banks, Contour Representation, Speech Recognition, Speech Synthesis.

1 Introduction

A transmultiplexer [1,2] is a system that combines a collection of suitably upsampled and filtered signals for the transmission by a single channel. Contours can be treated as a set of points that describe the boundary of objects (see Fig. 1). The boundary between an object and the background or boundary between overlapping objects is just an edge. Contour points are produced by edge detection and contour-tracing techniques. The edge detection is a part of a process called segmentation - the identification of regions within an image. The segmentation is an important tool for many computer vision applications e.g. robot guidance, object recognition or for instance non-contact visual inspection. Moreover, contour processing is required in topographic or weather maps preparation, character recognition, processing of medical images and image compression. The image compression based on a region identification and contour extraction, called "the second generation" image compression is nowadays a subject of interest for many scientists.

The extracted contours are generally described using Freeman's chain codes [3]. Commonly used chain codes are based on the eight-directional scheme. However, most of the methods use two vectors for contour representation. One of such methods is the generalized chain-coding scheme [4]. Some applications (e.g. Fig. 2) prefer the (x,y) Cartesian representation [5]. Others require the (α,l) polar representation [4]. These two vectors can be treated as two independent 1-D signals for the transmultiplexing and transmission. The separation of the signals should be perfect and the recovery of the contour should be performed without distortion. The main problem in transmultiplexers is the leakage of signal from one channel to another. This goal can be achieved by a choice of filters that ensures perfect reconstruction. Our paper presents the transmultiplexing scheme for four signals, which describe two contours in Cartesian representation. Each signal is treated as a 1-D signal with f_s sampling frequency. Their spectra are presented in Fig.3.

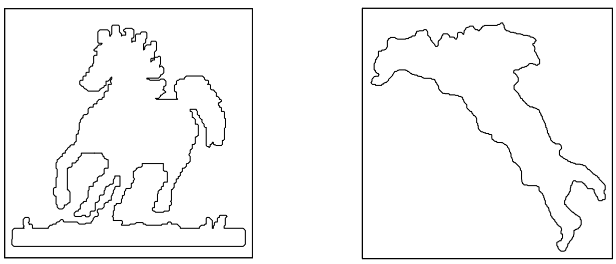

Fig. 1. Example of two contours

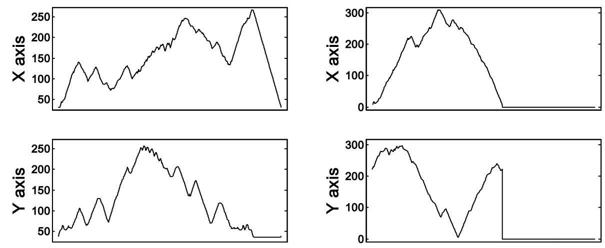

Fig. 2. 1-D representation of contours

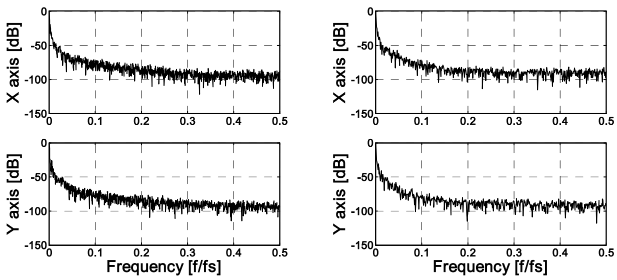

Fig. 3. Frequency spectra of 1-D contour representation

2 Transmultiplexing

A transmultiplexer combines several signals into a single signal. In contrast to the Time Division Multiplexing (TDM) and the Frequency Division Multiplexing (FDM) systems, transmultiplexers can use all samples and all frequencies to represent signals. Their main application is the simultaneous transmission of several data signals through a single channel.

Fig. 4 shows the classical structure of the 4-channel transmultiplexer. The input signals were upsampled, filtered and summed to obtain a composite signal which can be transmitted over a single transmission channel. The signal is relayed to the four channels of the detransmultiplexation part at the receiver end, where the signals must be filtered and downsampled to recover the original signals.

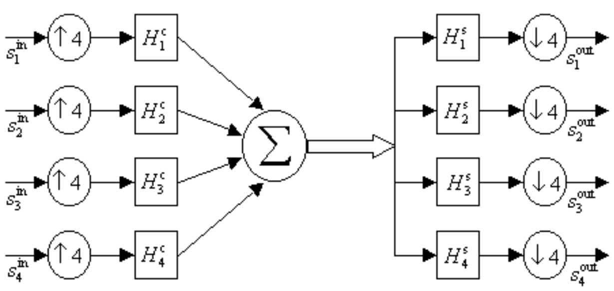

Fig. 4. A scheme of 4-channel transmultiplexer

The basic idea is the reversibility of all procedures of transmultiplexation in such a way that all output signals could be recovered as precisely as possible. For the well designed transmultiplexers, the output signal s_i^{out} approximates the input signal s_i^{in}, where i is a number of channel, $i \in \{1,2,3,4\}$. Transmultiplexer achieves the perfect reconstruction if s_i^{out} is a delayed and amplified version of s_i^{in}, namely if there exist a nonzero c and a positive integer τ such that

$$s_i^{out}(n) = c \cdot s_i^{in}(n-\tau)\,. \tag{1}$$

The dependence of output s_i^{out} from inputs s_k^{in} is described [2] in the z-transform domain by

$$\bar{s}_i^{\,out}(z) = 0.25\sum_{k=1}^{4}\bar{s}_k^{\,in}(z)\left[\sum_{m=0}^{3} H_i^{\,d}(w^m z^{0.25}) H_k^{\,t}(w^m z^{0.25})\right], \tag{2}$$

where $w = e^{-0,5\pi\cdot\sqrt{-1}}$ and $\bar{s}_i(z)$, $H^t(z)$, $H^d(z)$ stand for the z-spectrum of signal $s_i(n)$, transmultiplexer and detransmultiplexer transfer functions, respectively.

Each output signal depends on all input signals. A key point is that the constituent signals should be recoverable from the combined signal. To fulfil the perfect reconstruction condition (1), from (2) we obtain a set of equations

$$\sum_{m=0}^{3} H_i^d\left(w^m z^{0.25}\right) H_k^t\left(w^m z^{0.25}\right) = 4c\, z^{-\tau}\delta_{i,k}\,, \tag{3}$$

where

$\delta_{i,j}$ is a Kronecker function.

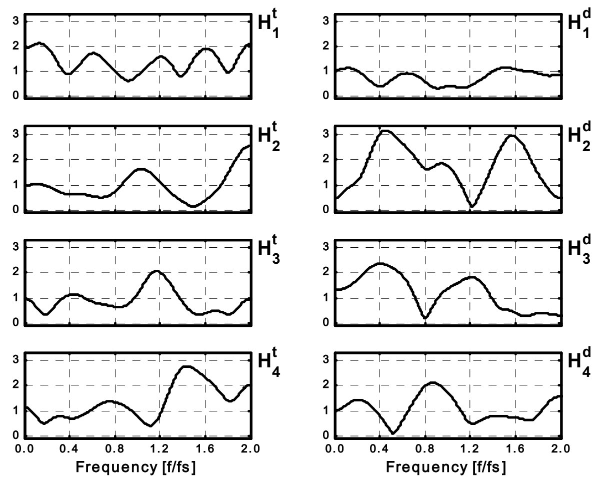

Fig. 5. Amplitude characteristics of H_k^t and H_i^d filters

Under assumption that all filters are FIR type of order I

$$H(z)=\sum_{j=0}^{I} h(j)\, z^{-j}, \tag{4}$$

conditions (1) and (3) are equivalent to

$$\sum_{j=\max\{0,2n-I\}}^{\min\{2n,I\}} h_i^d(j)\, h_k^t(4n-j) = c\delta_{i,k}\delta_{n,\tau} \tag{5}$$

where $h_i^d(j)$ means the j-th coefficient of the filter $H_i^d(z)$.

Similarly $h_k^t(4n-j)$ stands for the coefficients of the $H_k^t(z)$ filter. The assumption that all filters are of order I gives us restriction for sum (5), i.e. it gives an additional condition $0 \le 2n-j \le I$, which is equivalent to $\max\{0,2n-I\} \le j \le \min\{2n,I\}$.

3 Filter design

For each pair $i,k \in \{1,2,3,4\}$ of filter numbers, condition (5) gives $1+I/2$ equations, under assumption that I is an even number. For the considered case it results in the system of $16(1+I/2)$ equations.

Conditions (5) must be followed while filters designing to accomplish the perfect reconstruction (1) with the delay τ. A computer program, which minimizes the quantity criterion

$$Q=\sum_{i=1}^{4}\sum_{k=1}^{4}\sum_{n=0}^{I}\left(\sum_{j=\max\{0,2n-I\}}^{\min\{2n,I\}} h_i^d(j)\, h_k^t(4n-j) - c\delta_{i,k}\delta_{n,\tau}\right)^2 \tag{6}$$

was used to find the FIR filters which fulfil (5).

Each solution obtained in this way depends on the starting point of minimization procedure and usually reaches only a local minimum of (6). To obtain the solution, which fulfils some additional assumptions, important from the technical point of view, it is necessary to modify criterion (6). There should be added an additional term which should depend on filters parameters $h^t(j)$ and $h^d(j)$ in such a way that its minimization should lead to the realization of the technical requirements.

Some examples were computed and analysed to verify the presented above method of transmultiplexer filters design. One of them is presented below. The following properties were assumed:

1. transmultiplexer consists of 4-channels,
2. all filters are of $I=20$ order,
3. shifting $\tau=2$,
4. amplification $c=1$.

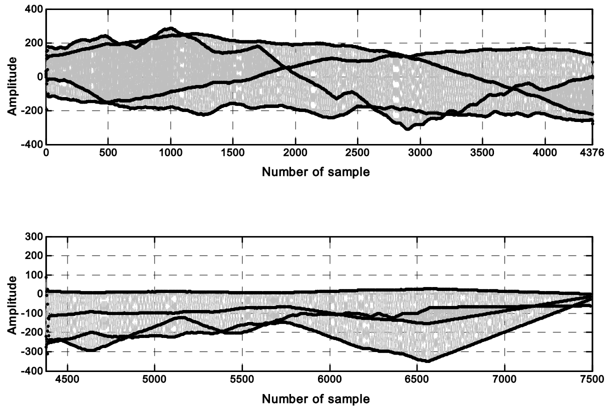

Fig. 6. Signal in transmission line

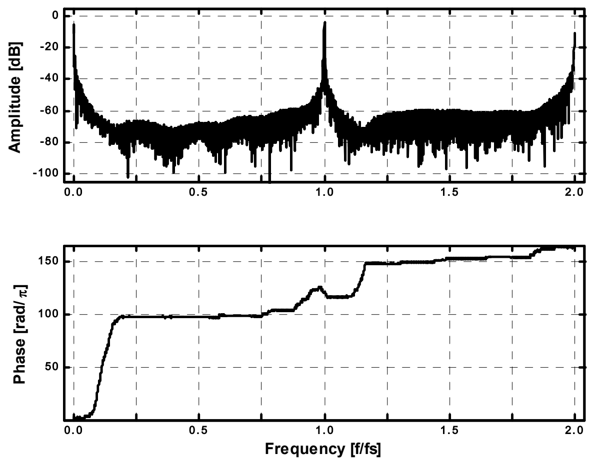

Fig. 7. Frequency spectra of combined signal

All values of filter parameters were equal to 1 at the beginning of the minimization procedure. A minimum value $Q_{\min}=2.5855\cdot 10^{-5}$ of the quantity criterion (6) was obtained. The set of conditions (5) consists of 176 equations for the considered case. The amplitude characteristics of all filters are presented in Fig. 5. The frequency band has width equal to 2 because the upsampling procedure increases the sampling frequency four times.

4 Compression

The spectra (see Fig. 3) of 1-D contour representations have special properties. The lack of high frequencies in the contour signals leads to small values of the transmitted signal spectrum for some frequencies (see Fig. 7). The location of these frequencies depends on the number of contours and can be easily determined. This allows applying some effective compression methods to reduce the number of bits of the transmitted signal.

Transmultiplexed signal is presented in Fig. 6. It was split into two waveforms to show some details. The lengths of vectors, which represent contours (see Fig. 2), are different. The upper plot of Fig. 6 presents the combined signal, which carries information connected with two contours presented in Fig. 1. The lower plot represents the remained part of the first contour and the zero signals instead of information connected with the second contour. Plots presented in Fig. 6 looks like four different waveforms but in practice it is only one waveform and value of signal change rapidly between two neighbouring samples.

Our fundamental concept of compression is to split up the frequency band of a signal spectrum and then use fewer bits to represent the less important frequencies, and more bits for the more important ones. If the values of spectra have large differences between frequency bands some compression gain must be obtained. The spectrum of combined signals, presented in Fig. 7, was analyzed to verify the frequency properties of transmultiplexed signal. Some frequencies appeared frequently in the spectrum while the other frequencies are almost not detectable. This property allows applying effectively both the lossless and the lossy compression methods to reduce the number of bits of the transmultiplexed signal. The wavelet packet algorithm (see Fig. 8) generates a set of subband spectra that are derived from a composed signal. By using a filter bank the subband spectra are produced by cascading the filtering and downsampling operations. Subsequent levels in the tree are constructed by applying recursively the wavelet transform, step A-approximation to the low and D-detail to the high pass filter. Three-level packets were generated by using the discrete Meyer wavelet. Each successive stage decomposes its input vectors twice. Each output vector has half of the number of samples of the input vector.

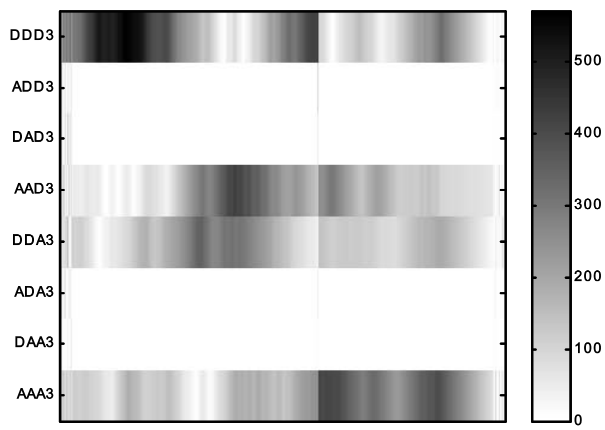

Fig. 9. Frequency decomposition of transmultiplexed signal

Thus, the frequency domain of the signal spectrum is partitioned twice and the transform with three stages has 2^3 spectra. To recover the signal from the wavelet spectrum, the inverse discrete wavelet filter bank (the right part of Fig. 8) is used. The absolute values of the wavelet packet decomposition are presented in Fig. 9. The frequency is plotted on the vertical axis. 99,87 % of the spectrum energy is transmitted in subbands which constitute a half of the whole band. It is easy to distinguish important and unimportant bands. The lossy compression consists in removing spectrum parts: ADD3, DAD3, ADA3 and DAA3 (see Fig. 8 and Fig. 9). For the case of lossless compression these spectra bands are represented by lower number of bits than other parts of spectrum.

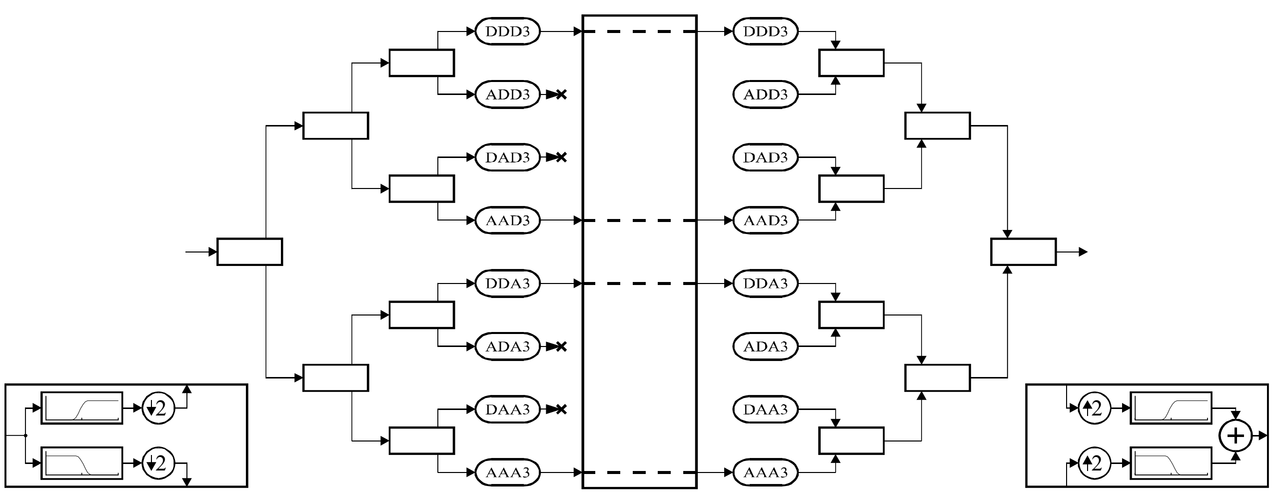

Fig. 8. Scheme of decomposition, compression and restoration of transmultiplexed signal

5 Speech transmultiplexing

In the transmitter, the M input signals were upsampled, filtered and summed to obtain a composite signal. At the receiver end, the composite signal is relayed to four channels of the separation part, where the signal is filtered and downsampled to recover the original input signals. The basic idea [6,7] is the reversibility of all procedures.

There is a growing tendency to equip transmultiplexers in reversible integer-to-integer filter banks [8]. Signals are then processed in finite-precision arithmetic. Due to this property, transmultiplexers of this type can be applied to transmit lossless compressed signals, low memory is needed and complexity of computations is slight. Such filters are needed in the case when written text is transmitted.

6 Speech recognition system

Voiced sounds are produced by modulation of the air flow from the lungs by vibration of vocal cords. Speech is an input in speech recognition system while the sequence of written words is an output. Speech recognition is a problem investigated by many scientists [9, 10, 11].

The time dependent amplitude and the frequency characteristics of a speech signal change in the time domain by continuous reconfiguration of human's voice-tract resonant chambers. There is a number of dynamically altered parameters in the speech signal, which make the analyses and modeling difficult.

Speech signal should be segmented in a certain manner, before analyzing. The signal contained in the obtained segments can be more precisely processed because their characteristics are more constant. The most effective method seems to be constant segmentation for 20 [ms] length blocks. However, good segmentation can be based on energy fluctuations: rises and falls.

The appropriate representation of speech segments is the next important problem. The original time-varying signal representation is not useful in speech recognition. The signal transformation is necessary for an efficient system. Finding an accurate transformation is a fundamental problem. The choice of transform depends on the purpose of analysis.

Usually, methods, which are based on the cosine (DCT) or the wavelet (DWT) transform, are used. This way the frequency properties of speech can be analyzed. The analysis of energy distribution for the different frequencies seems to be the best solution to distinguish the phonemes.

Fig.10 presents the waveform for a single word spoken in the Polish language, its time-frequency analysis which is based on DWT and the energy fluctuations.

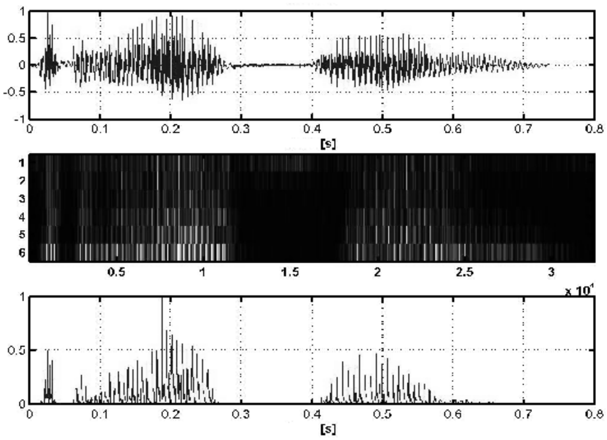

Fig. 10. Word "Rafał" as a signal, its discrete wavelet transform and power waveform

Each speech recognition system must be equipped with a dictionary which consists of possible words. Especially its size has a great impact on the efficiency of the system. Words in the dictionary should be proper for needs of customers. They should give them the comfort of natural speech in their subjects. However, the dictionary should be as small as possible to increase the speed of processing. Moreover, too many words may cause errors.

Speech recognition consists of a number of elementary steps: voice capturing, noise compensation, segmentation, transformation, feature extraction, classification, lexical correction and grammar correction. Automated speech recognition remains an open problem, since there are only few working and efficient solutions (e.g. Dragon). Speech recognition is useful and demanded by many people. It would considerably increase the speed of cooperation with the computer.

7 Speech synthesizer

Speech synthesizer is a computer-based system that should be able to read any text aloud, whether it was directly introduced in the computer by an operator or scanned and submitted to an Optical Character Recognition (OCR) system. It is a much easier problem than speech recognition and there are already very good commercial (i.e. AT&T Labs, Bell Labs) and freeware (i.e. FreeTNT) solutions for many languages. Anyway speech synthesis is still an interesting area for researches. They are focused on modelling emotions for speech synthesis [12]. Speech synthesizers are commonly known as TTS (Text-To-Speech). Synthesizers typically consist of two main blocks. A Natural Language Processing module (NLP) is capable of producing a phonetic transcription of the text read with the desired intonation and rhythm. A Digital Signal Processing module (DSP) transforms the symbolic information it receives into speech.

8 Compression

Different types of phone calls demand different reactions. Let us compare two examples. The features of voice are sometimes very important for some speaking people. In that case the exact words are less important. The speech can be noised and even some words can be missed but features of the voice and emotions have to be transferred. Another case is a phone call between employee A and employee B of the same company who do not know each other. The employee A wants to present the production factors of his unit. The emotions and features of voice are not important and sometimes even inadvisable. The task is to transfer digital voice signal without any emotions. In the second case it could be much more effective to send a text instead of voice. For the comfort of employees, a communication system can be equipped with speech recognition and speech synthesis systems on both sides. In that way employees can speak and hear, each other, while their voices are transferred by network as a text.

Bit rate for uncompressed speech signal is 88.2 [kb/s] at 11025 [Hz] sampling rate and 8 bit representation. The compressed bit stream for a signal in a 2-G wireless system is 13 [kb/s] only. It is possible to decrease considerably this bit stream if text is sent to receiver instead of voice. Average phoneme lasts about 50 [ms] in speech waveform. Hoffman compression enables to code written text using 4 bits for one symbol in average. It means that voice presented as text needs a bit stream about 0.08 [kb/s] only. If we compare these results with the classical GSM mobile phone system we obtain the compression ratio 160.

Another example is presented in Tab.1. There are two compared representations of the Hamlet monolog. The first one presents the file size where the voice was recorded with 8 bits per sample at 11025 [Hz] sampling rate. The second one is a text representation (transcription). The compression gain for this example is 895!

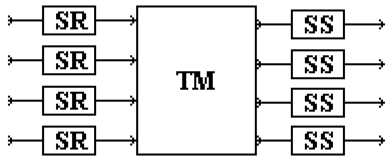

Fig. 11. Voice communication system which is based on the text data format

where:
SR - Speech Recognition,
TM - Transmultiplexer telecommunication system,
SS - Speech Synthesis.

Table 1. Sizes of text and audio files

File type	Size [kB]
Audio (.wav) 8 bit	1203244
Text (.txt)	1344

9 Examples

Let us consider the specific case when filters are FIR type. Let the orders of all combining filters be equal to K which depends on the number of channels in the following way $K \le M - 1$. Let the order of all separation filters be $L \le M$ and moreover $h_i^s(0) = 0$ for all $1 \le i \le M$. Let us introduce two matrices

$$G^c = \begin{bmatrix} h_1^c(M-1) & h_1^c(M-2) & \cdots & h_1^c(0) \\ h_2^c(M-1) & h_2^c(M-2) & \cdots & h_2^c(0) \\ \cdots & \cdots & \cdots & \cdots \\ h_M^c(M-1) & h_M^c(M-2) & \cdots & h_M^c(0) \end{bmatrix} \quad (7)$$

$$G^s = \begin{bmatrix} h_1^s(1) & h_2^s(1) & \cdots & h_M^s(1) \\ h_1^s(2) & h_2^s(2) & \cdots & h_M^s(2) \\ \cdots & \cdots & \cdots & \cdots \\ h_1^s(M) & h_2^s(M) & \cdots & h_M^s(M) \end{bmatrix} \quad (8)$$

which consist of combining and separation filter coefficients, respectively. Both matrices are square and their dimensions depend on number of channels. Under these assumptions, the perfect reconstruction conditions (1) can be written in a simple form

$$G^c G^s = E,$$

where:
E - unitary matrix.

If we assume that both matrices G^c and G^s are nonsingular, then we obtain a simple algorithm for filter designing:

- choose an arbitrary matrix (7) (i.e. coefficients of composition filters)
- compute the coefficients of separation filters

$$G^s = (G^c)^{-1}. \tag{9}$$

It is possible to provide all calculations using the integer numbers only. For this case it is convenient to use such filters that

$$\det G^c = \det G^s = 1 .$$

Let us consider an example of a transmultiplexer which consists of four channels and let the following coefficients

$$h_1^c = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}, \; h_2^c = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}, \; h_3^c = \begin{bmatrix} 0 \\ 1 \\ 1 \\ 1 \end{bmatrix}, \; h_4^c = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \tag{10}$$

for FIR combining filters be assumed. Simple computations (9) provide the coefficients of separation filters

$$h_1^s = \begin{bmatrix} 0 \\ 1 \\ -1 \\ 0 \\ 0 \end{bmatrix}, \; h_2^s = \begin{bmatrix} 0 \\ 0 \\ 1 \\ -1 \\ 0 \end{bmatrix}, \; h_3^s = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \\ -1 \end{bmatrix}, \; h_4^s = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}. \tag{11}$$

This example depicts the extraordinary simplicity of filters which can be obtained while using algorithm described in [8].

10 Conclusions

Transmultiplexers allow transmitting the contour signals by a single transmission line assigned to transmit acoustic signals. Theoretically, transmultiplexing does not distort contour signals because digital filters satisfy the requirements of a perfect reconstruction. Some small distortions caused by the lossy compression presented above are almost invisible by the human eye.

The presented method for speech compression enables to transmit a large number of speech signals through a single channel. It can be obtained due to strong signal compression. Such system needs to apply the speech recognition and synthesis software. Some disadvantages result from imperfect properties of speech-to-text and text-to-speech converters, especially the recognition system brings noticeable harmful effects. The large variability in the signal makes the speech recognition difficult.

If integer filters are incorporated in filter banks then not only theoretically but also in practice the perfect reconstruction conditions are fulfilled. The usefulness of integer filters enables transmission of text, software files or coded multimedia data like MPEG files.

Acknowledgments

This work was supported by MNiI under grant number 4 T11D 005 23.

References

[1] C.S. Burrus, R.A. Gopinath, H. Guo, *"Introduction to Wavelets and Wavelet Transform"*. Prentice Hall, New Jersey, 1998.

[2] M. Vetterli, J. Kovacevic, *"Wavelets and Subband Coding"*. Prentice Hall PTR, New Jersey, 1997.

[3] H. Freeman, "Computer Processing of Line-Drawing Images," in *ACM Computing Surveys (CSUR)*, 2, 1974, pp. 57-97.

[4] A. K. Jain, *"Fundamentals of Digital Image Processing",* New Jersey: Prentice Hall International, 1989.

[5] R. Baran, A. Dziech, "Tangent Method and The Other Efficient Methods of Contour Compression", *WSEAS Transactions on Computers*, Issue 7*,* Vol. 4, July 2005, pp. 805-813.

[6] M. Vetterli, A theory of multirate filter banks, *IEEE Trans. Accoust. Speech Signal Process*, Vol. 35, 1987, pp. 356-372.

[7] A.N. Akansu, P. Duhamel, X. Lin, and M. Courville, Orthogonal Transmultiplexers, *IEEE Trans. on Signal Processing*, Vol. 46, No. 4, 2001, pp. 979-995.

[8] B.Ziółko, M.Ziółko and M.Nowak, Design of Integer Filters for Transmultiplexer Perfect Reconstruction, *Proceedings of XIII European Signal Processing Conference EUSIPCO-2005*, Antalya. (in appear)

[9] M.Ziółko, M. Kępiński, J. Gałka, Wavelet-Fourier Analysis of Speech Signal, *Proceedings of the Workshop on Multimedia Communications and Services*, Kielce, 2003.

[10] O. Segawa, K. Takeda, F. Itakura, Continuous speech recognition without end-point detection, *Transactions of the Institute of Electrical Engineers of Japan*, Part C, Vol.124-C, No.5, 2004, pp. 1121-7.

[11] O. Farooq, S. Datta, Wavelet based robust sub-band features for phoneme recognition, *IEE Proceedings: Vision, Image & Signal Processing*, Vol.151, No.3, 2004 pp. 187-93.

[12] M. Schroder, Dimensional emotion representation as a basis for speech synthesis with non-extreme emotions, *Affective Dialogue Systems. Tutorial and Research Workshop, ADS 2004*. Proceedings (Lecture Notes in Artificial Intelligence Vol. 3068), Springer-Verlag, 2004.