Buletinul Ştiinţific al Universităţii "Politehnica" din Timişoara

Seria ELECTRONICĂ şi TELECOMUNICAŢII
TRANSACTIONS on ELECTRONICS and COMMUNICATIONS

Tom 51(65), Fascicola 2, 2006

# Speech and Speaker Recognition Application on the TMS320C541 board

Eugen Lupu[1], Petre G. Pop[1], Radu Arsinte[1]

**Abstract – The paper presents a speech and speaker recognition application developed on the EVM C541 board using the CCS®. The application represents the implementation of the TESPAR coding method on a DSP support. The TESPAR alphabet for the coding process was obtained formerly. The speech/speaker information contained in the utterances is extracted by TESPAR coder and provides the TESPAR A matrices. For the recognition decision, the distances among the TESPAR A test matrix and the TESPAR A reference matrices are computed. The results of the experiments prove the high capabilities of the TESPAR method in the classification tasks.**

## I. INTRODUCTION

TESPAR *(Time Encoded Signal Processing and Recognition)* coding is a method based on the approximations to the locations of the 2TW (where W is the signal bandwidth and T the signal length) real and complex zeros, derived from an analysis of a band-limited signal under examination. Numerical descriptors of the signal waveform may be obtained via the classical 2TW samples ("Shannon numbers") derived from the analysis. The key features of the TESPAR coding in the speech-processing field are the following:

-the capability to separate and classify many signals that cannot be separated in the frequency domain

-an ability to code the time varying speech waveforms into optimum configurations for processing with Neural Networks

-the ability to deploy economically, parallel architectures for productive data fusion [2].

The key in the interpretation of the TESPAR coding possibilities consists in the complex zeros concept. The band-limited signals generated by natural information sources include complex zeros that are not physically detectable. The real zeros of a function (representing the zero crossing of the function) and some complex zeros can be detected by visual inspection, but the detection of all zeros (real and complex) is not a trivial problem. To locate all complex zeros involves the numerical factorization of a $2TW^{th}$-order polynomial. A signal waveform of bandwidth W and duration T, contains 2TW zeros; usually 2TW exceeds several thousand. The numerical factorization of a $2TW^{th}$-order polynomial is computationally infeasible for real time. This fact had represented a serious impediment in the exploitation of this model. The key to exceed this deterrent and use the formal zeros-based mathematical analysis is to introduce an approximation in the complex zeros location [5].

Instead of detecting all zeros of the function the following procedure may be used:

- The waveform is segmented between successive real zeros and
- This duration information is combined with simple approximations of the wave shape between these two locations.

These approximations detect only the complex zeros that can be identified directly from the waveform.

In this transformation of signals, from time-domain in the zero-domain:

- The real zeros, in the time-domain, are identical to the locations of the real zeros in the zero-domain, and
- The complex zeros occur in conjugate pairs and these are associated with features (minima, maxima, points of inflexion etc.) that appear in the wave shape between the real zeros [3][4].

In this way examining the features of the wave, shape between its successive real zeros may identify an important subset of complex zeros.

In the simplest implementation of the TESPAR method [1], two descriptors are associated with every segment or epoch of the waveform.

These two descriptors are:

- The *duration (D), in number of samples,* between successive real zeros, which defines an e*poch*
- The shape *(S),* the *number of minima* between two successive real zeros.

The TESPAR coding process is made by using an alphabet (symbol table) to map the duration/shape (D/S) attributes of each epoch to a single descriptor or symbol [5]. The mode to get the TESPAR alphabet is well presented in [6][9].

The TESPAR symbols string may be converted into a variety of fixed-dimension matrices. For example, the

[1] Facultatea de Electronică Telecomunicaţii şi Tehnologia Informaţiei, Catedra Comunicaţii, str. Bariţiu 26-28, 400027 Cluj-Napoca, Eugen.Lupu@com.utcluj.ro

S-matrix is a single dimension 1xN (N- number of symbols of the alphabet) vector, fig.1. which contains the histogram of symbols that appear in the data stream (Nr. App). Another option is the A-matrix, which is a two dimensional NxN matrix that contains the number of times each pair of symbols appears at a "*lag*" distance of n symbols (fig. 2) [1][2]. The "*lag"* parameter provide the information on the short-term evolution of the analyzed waveform if its value is less than 10 or on the long-term evolution if its value is higher than 10. This bidimensional matrix assures a greater discriminatory power.

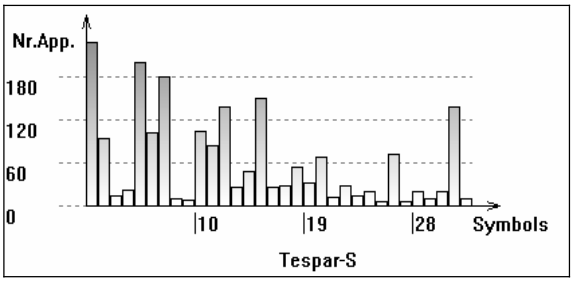


Fig. 1. TESPAR S-matrix

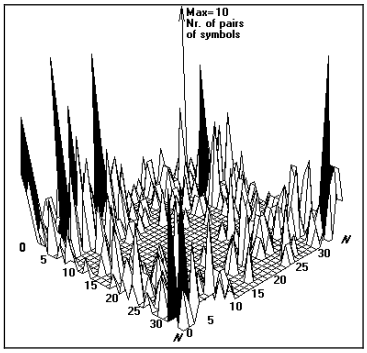


Fig. 2. TESPAR A-matrix

These matrices are ideal to be used as fixed-sized training and interrogation vectors for the MLP neural-networks. There are two main methods of classifying using TESPAR:

- Classifying using archetypes
- Classifying with neuronal networks [1][5].

This paper deals with the first method that was implemented on the system. An archetype is obtained by averaging several TESPAR A-matrices obtained from different versions of the same utterance. Such archetypes tend to outline the basic mutual characteristics and dim the particular cases that might appear in different utterances of the same word, for example.

The created archetype may be loaded in the database and then used. In the classification process, a new matrix might be created and then compared to the archetype. Many different forms of correlation can be used to achieve the classification. A threshold is required to establish whether the archetype and the new matrix are sufficiently alike; the archetype with the highest ratings is chosen after it has been compared to a threshold.

## II. RECOGNITION SYSTEM OVERVIEW

Fig.3 shows the block diagram of the application, this being shared between PC and the DSP board. In order to run the application we have to load the ***VoiceR/ SpeakeR*** program on the EVM DSP board and to run the program A-Matrix Tools (on PC) if some reference TESPAR A-matrices are to be loaded by the DSP program.

The applications on the DSP board are built up using the CCS® (*Code Compose Studio*) environment that allows the fast application development using its own resources: C compiler, linker, debugger, simulator, RTDX (Real time Data Exchange) and DSP/BIOS components [7][8].

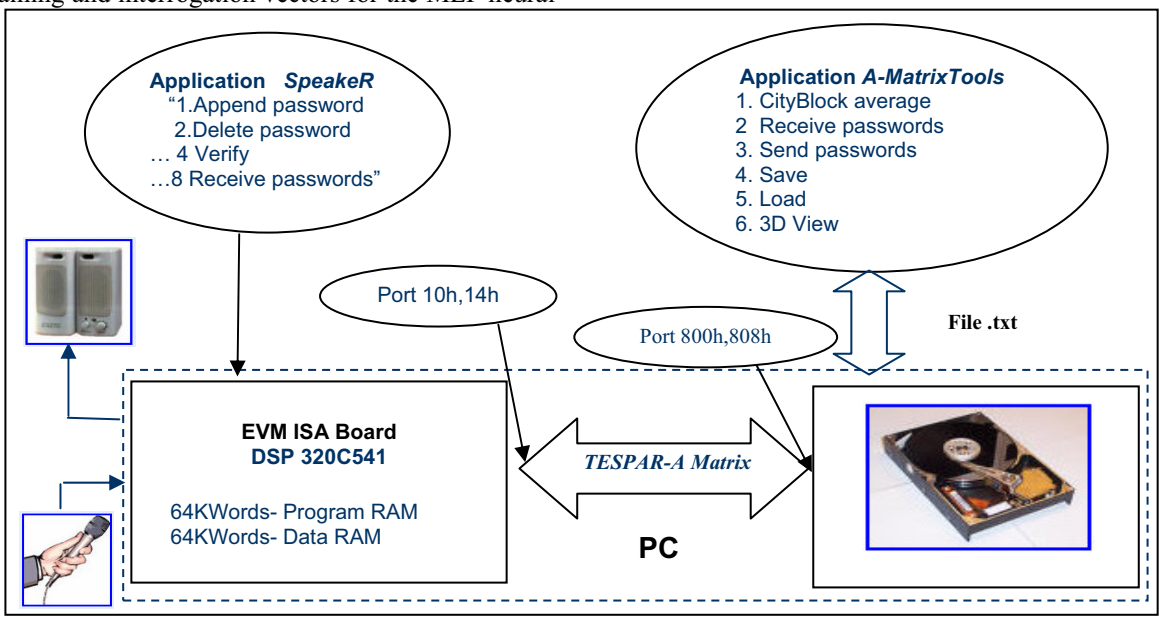


Fig. 3. The block diagram of the SpeakeR application

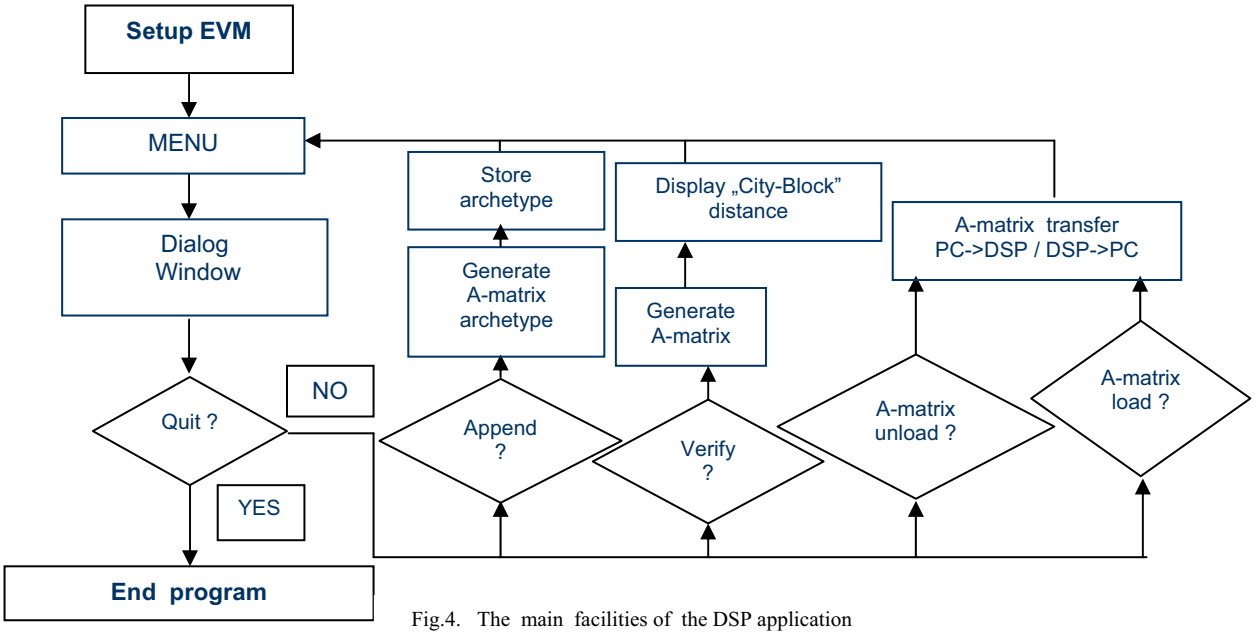


Fig.4. The main facilities of the DSP application

The main facilities of the ***SpeakeR*** application can be remark in the flow diagram, fig. 4.

The A-Matrix Tools program allows the TESPAR A-matrices transfer between the host PC and the EVM C541 board and offer facilities to extend the SpeakeR program operation. The tasks of this program are the following:

- TESPAR A-matrices collections transfer between host PC and EVM board
- TESPAR A-matrix transfer between EVM board and host PC
- save matrix/matrices collections to host PC hard disk in text files
- load matrix/matrices collections from host PC hard disk to EVM board
- 3D TESPAR A diagrams visualization City-block distance computation between two TESPAR-A matrices
- Archetype generation for TESPAR A-matrices collections.

For the polling communication between the host PC and the DSP board the following ports are employed; for data the port 800h (PC) and 10H (DSP) and for control 808h (PC) and 14h (DSP)[8].

## III. EXPERIMENTS, RESULTS AND CONCLUSIONS

The applications facilitate to perform "on-line" speech/speaker recognition experiments. In the classification process, the distance calculation between the TESPAR A-matrices archetypes and the test matrices or parallel MLP neural networks may be employed. In this paper, the experiments focus on the use of "city-block"distance calculation between the A-matrices archetypes and test matrices in the classification task. The EVM board resources limit the number of enrolled speakers to 10 and the dictionary dimension for the speech recognition experiments.

### A. *Speech recognition experiments*

Two types of experiments were been made, one using the ten digits as utterances and the other using different commands (left, right, up…). In the first experiment, seven speakers were enrolled for the system training and 10 speakers for the test. Each of the enrolled speakers uttered three times every digit for the training and ten times for the recognition. The results of this experiment are presented in fig.5. In this case an average recognition rate of 92% was obtained that we find to be good in the condition of using for test also utterances of not enrolled speakers.

For the other type of experiments, we used 10 commands words. In the "Test2", experiment the training was made by using the three utterances from every speaker to build its own archetype for every command. For the "Test3" experiment the archetype were been built by using three utterances from two enrolled speakers.

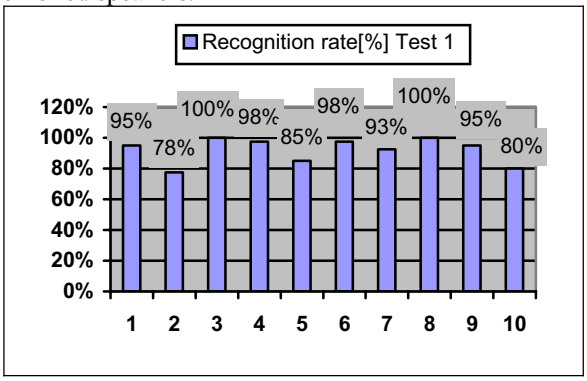


Fig.5. Digits recognition rate for the "Test 1" experiment

In every experiment to test the system all the speakers has uttered 10 times every command [10]. The results are presented in table 1. For the "test2" experiment, a

98% average recognition rate was provided by the system and for the "test 3" slightly lower than 97%.
The results of the experiments prove the high capabilities of the TESPAR method in the classification tasks, noticed also in [1][3]. The results generally are better than 90% and the DSP resources are not very highly used. In order to improve the recognition rate the employment of MLP neural networks for classification will be employed.

Table 1

| Word | Recognition Rate [%] Test 2 | Recognition Rate [%] Test 3 |
|---|---|---|
| **Up** | 100 | 100 |
| **Down** | 95 | 100 |
| **Left** | 100 | 100 |
| **Right** | 100 | 100 |
| **Enter** | 95 | 95 |
| **Cancel** | 100 | 95 |
| **Abort** | 100 | 85 |
| **Ok** | 100 | 100 |
| **Back** | 90 | 95 |
| **Forward** | 100 | 100 |

B. *Speaker recognition experiments*

Two types of speaker identification experiments were been made, one using different passwords for each enrolled speaker (his name) and the other using the same password. For the first experiment, eight speakers were enrolled. Each of them uttered three times the same password to provide the TESPAR-A matrix archetype. For the identification experiment, 10 sessions of attempts were been made during a week, each speaker uttering its own password 10 times in every session. The results of this experiment are presented in fig.6. In this case the system provide an average recognition rate of 96.1*%.*

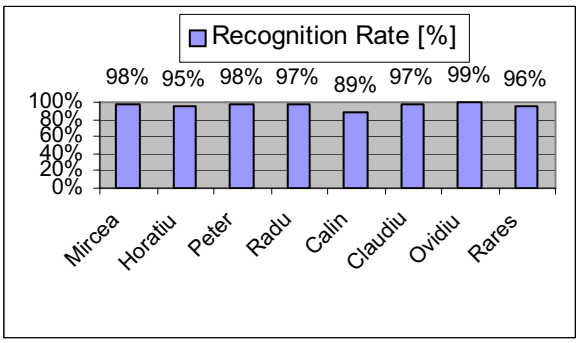


Fig. 6. The recognition rate for the experiment using different passwords

For the other experiment, the eight enrolled speakers use the same password. In this case, the training was made by using three utterances from every speaker to build its own archetype for every password.
To test the system all the speakers had uttered 10 times the password. The experiment was repeated for ten different short passwords. The recognition rates for every speaker and all passwords are presented in fig.7. An overall average recognition rate of *89.25%* was provided by the system for this experiment. The results of the experiments prove the high capabilities of the TESPAR method in the classification tasks noticed also in [1][5][9]. These results generally are better in the experiment that uses different passwords; the DSP resources are not very highly used. In order to improve the recognition rate the employment of MLP neural networks for classification is recommended. To validate the system more experiment are to be made using much amounts of utterances and different speakers are advisable to be tested. In addition, the effects of other signal processing algorithms applied before the coding process are to be studied.

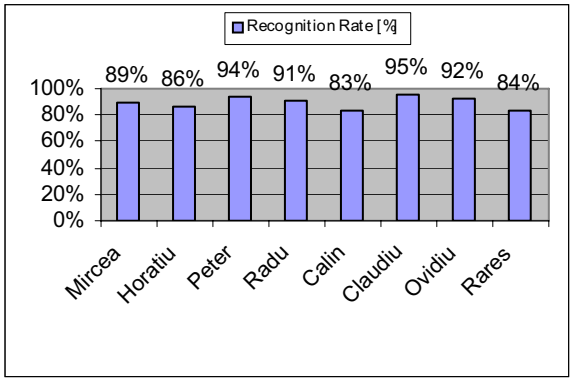


Fig. 7. The recognition rate for the experiment using the same password

## REFERENCES


[1] King, R. A., Phipps, T. C. "Shannon, TESPAR and Approximation Strategies", *ICSPAT 98,* Vol. 2, pp. 1204-1212, Toronto, Canada, September 1998.

[2] Phipps, T.C., King, R.A. "A Low-Power, Low-Complexity, Low-Cost TESPAR-based Architecture for the Real-time Classification of Speech and other Band-limited Signals" International Conference on Signal Processing Applications and Technology (ICSPAT) at DSP World, Dallas, Texas, October 2000, *www.dspworld.com/icspat/spchrec.htm.*

[3] Voelcker, H. B. "Toward A Unified Theory of Modulation Part 1: Phase-Envelope Relationships*", Proc. IEEE*, vol. 54, no. 3, pp 340-353, (March 1966).

[4] Requicha, A. A. G. "The zeros of entire functions, theory and engineering applications" Proceedings of the IEEE, vol. 68 no. 3, pp. 308-328, March 1980.

[5] Lupu, E., Feher, Z., Pop, P.G. **"**On the speaker verification using the TESPAR coding method", IEEE Proceedings of Int. Symposium on "Signals, Circuits and Systems", Iassy, Romania, 10-11 July 2003, pp.173-176, ISBN 0-7803-7979-9

[6] Lupu, E., Pop, P.G., Digital speech processing. Analysis and recognition, Cluj-Napoca, Risoprint 2004, ch.11, pp.164-175

[7]*** Texas Instruments - TMS320C54x DSP Reference Set

[8]*** Texas Instruments – TMS320C54x Evaluation Module Tehnical Reference

[9] Lupu, E., Moca,V., Pop, P.G "Environment for speaker recognition using speech coding" Proc. of Communications 2004, Bucharest, 3-5 June, Vol. 1, pp.199-204, ISBN 973-640-036-0

[10] Lupu, E., Pop, G. P., Pătraş, M." Low Complexity Speaker Recognition System Developed on the DSP TMS320C541 Board" Proceedings of the 9th International conference "Speech and Computer" SPECOM' 2004 , 20-22 sept. 2004, St. Petresburg pp. 398-402 ISBN 5-7452-0110-x