Buletinul Ştiinţific al Universităţii "Politehnica" din Timişoara

Seria ELECTRONICĂ şi TELECOMUNICAŢII
TRANSACTIONS on ELECTRONICS and COMMUNICATIONS

Tom 49(63), Fascicola 1, 2004

Comparing Various Automatic Speaker Recognition Approaches

Tudor Barbu[1], Mihaela Costin[2]

Abstract – Connectionist methods in speaker recognition give promising results on the condition of selecting a well balanced feature vector. Special attention has to be paid to the classifier procedure applied on the presented templates. We are discussing here several different speaker recognition methods, focusing on two main approaches: a MFCC supervised approach using a Hausdorff-based metric and an AR – ART2 structure (autoregressive features classified with an Adaptive Resonance Theory type 2 Network). Then, we report them to our previous results. These methods can function in a single or parallel structure, needing a weighted inference to accomplish the recognition task. Respecting the well known a-priories in signal treatment, and carefully selecting the frequency bands to be "examined", our trials gave good results, with a mean detection error reduced with about 3% compared to our previous essays. Our studies concentrate on tests made with discrete-words, in real, noisy environment and real-time / off-line speaker verification.

Keywords: speaker recognition, speech feature vectors, mel-frequency cepstral coefficients, autoregressive (AR) coefficients, ART 2 (Adaptive Resonance Theory analog model).

I. INTRODUCTION

Speaker recognition [1,2], usually classified into identification, respectively verification processes, implies to automatically detect the speaking person on the basis of individual information intricate in speech waves [3], selected and formalized in templates.

In speaker verification, the task is to use a speech sample to test whether a person who claims to have produced the utterance has really done it, while, in speaker identification, the task is to use a speech sample to select the identity of the person that produced the speech from among a population of speakers, or to cluster it in the most resembling group.

The most significant factor affecting automatic speaker recognition performance is variation in the signal characteristics from trial to trial (intersession variability, variability over time).

Variations arise from different sources:

- background noise channel distortions;
- differences in recording and transmission conditions;
- wide variety of accents;
- people's voices change over time;
- emotional, stress, fatigue states;
- speakers cannot repeat an utterance precisely the same way from trial to trial.

These reasons make difficult both the establishment of reference examples and the selection of the good pattern matching method.

It is well known that samples of the same utterance recorded in one session are much more highly correlated than samples recorded in separate sessions.

Technically, the recognition systems have to treat these variations. Two types of normalization techniques have been tried: one in the parameter domain, and the other in the distance/similarity domain. In our previously presented methods [4-10], linear prediction coding (LPC), Mel-frequency cepstral coefficients (MFCC) and autoregression (AR) coefficients classified with MLP (multi layer perceptron) and RBF (radial basis functions) recognition, were compared with those obtained by technique of Hausdorff-based distance measure (HDM) and vowel detection. The HDM gave good results, together with AR in a general computing schema we proposed (to be patented). The need of a special emphasized research resulted on two stages: limits of sub-bands to be explored and a more specialized classifier applied to the feature vector.

II. A SUPERVISED TEXT-DEPENDENT SPEAKER RECOGNITION APPROACH

Let us propose now an automatic supervised text-dependent speaker recognition method, which uses a special nonlinear metric in the classification stage.

We consider the following speaker recognition task.

[1] Computer Science Institute, IIT, Romanian Academy, Iaşi, Dd. Carol I, nr. 11A, e-mail: tudbar@iit.tuiasi.ro,
[2] IIT, Romanian Academy, associated UAIC, Iaşi, e-mail: mccostin@iit.tuiasi.ro, cstf2004@yahoo.com

Let us assume that we have a sequence of spoken utterances, all of them having the same transcript.

The recognition task consists of identifying the speaker for each utterance, assuming that the available speakers are known and a training set is also available.

So, let $s_1,...,s_n$ be the audio signals of the input utterances and N be the number of speakers. Each of these speakers provides several vocal recordings, by pronouncing the text of s_i a number of times. Thus, for each speaker we get a set of prototype vocal signals. The training set will contain all the obtained same-transcript prototypes, therefore having the form $\{S_1^1,...,S_{n_1}^1,...,S_1^N,...,S_{n_N}^N\}$, where S_j^i represents the j^{th} vocal signal of i^{th} speaker.

As any pattern recognition approach, our speaker recognition method consists of two parts: the speech feature extraction and the feature vector classification. We shall describe each of them in the next subsections.

A. *Feature vector computation approaches*

In this first stage we perform vector feature extraction operations for both the vocal signals to be recognized and the prototype signals. The Mel Frequency Cepstral Coefficients (MFCC) [1-4,11] are the dominant features used for speech [3] and speaker recognition [1,2,4-10], so we propose a melodic cepstral analysis for speech feature extraction.

The MFCC speech feature vector extraction approach uses a short-time analysis of the involved audio signals. So, each signal is divided in overlapping segments of length 256 samples with overlaps of 128 samples. Then, each resulted segment is windowed, by multiplying it with a Hamming window of length 256. The spectrum of each windowed sequence is then computed, by applying DFT (Discrete Fourier Transform) to it. Mel spectrum is obtained by converting spectrum on the melodic scale [1], that is described as:

$$mel(f) = 2595 \cdot \log_{10}(1 + f/700), \quad (1)$$

where f represents the physical frequencies and $mel(f)$ the mel frequencies. Further, the mel cepstral acoustic vector is computed by applying first the logarithm, then the DCT (Discrete Cosinus Transform) to the mel spectrum.

Therefore, for each vocal signal s_i or S_j^i, a set of MFCC speech feature vectors is obtained. Each MFCC acoustic vector has 256 samples and the number of the acoustic vectors depends on the signal length.

Using these acoustic vectors, new speech feature vectors can be obtained.

One of them is the truncated MFCC acoustic matrix. This is obtained by truncating the acoustic vectors at the first 12 MFCC coefficients and positioning them as columns. Other possible speech feature vectors result from DDMFCC analysis. Delta mel cepstral coefficients (DMFCC) are computed as the first order derivatives of MFCC, and the delta delta mel frequency cepstral coefficients (DDMFCC), as the second order derivatives of MFCC. Thus, the DDMFCC acoustic vectors, and then the truncated DDMFCC acoustic matrix, result as speech feature vectors [12].

Another important speech vector is the vocal sound *pitch*, which can be computed using MFCC acoustic matrix. From each column of this matrix, therefore from each MFCC acoustic vector of the initial signal, the maximum value is extracted. A small enough threshold value can also be used, the maximum value being further considered only if it is above that threshold. The pitch vector is obtained from all these extracted values. From our tests it results that the pitch is not satisfactory enough as a feature vector. The speakers cannot always be discriminated by their pitch vectors only, but the range of its variation could be taken in consideration.

So, we propose a more powerful vocal feature vector which is based on pitch. From every MFCC acoustic vector of the processed sound, we extract not only one value, but a sequence of values. Thus, the greatest m MFCC coefficients of each acoustic vector are determined, where m is a chosen small enough integer. Therefore, for each vocal signal s, a feature vector $V(s)$ is obtained as a matrix having m rows and a column number depending on s length. The resulted vector features vocal sound, much better than pitch.

Thus, we get a speech feature set, $\{V(s_1),...,V(s_n)\}$, and a training feature set, $\{V(S_1^1),...,V(S_{n_1}^1),...,V(S_1^N),...,V(S_{n_N}^N)\}$, which contains all training vectors.

The next step of the recognition process consists of classification of the speech feature vectors using the training feature vectors.

B. *Speech feature vector classification*

We propose an extended minimum-distance classifier for performing the feature vector classification. The classical variant of this classifier consists of a set of prototypes and an appropriated metric [12]. The pattern to be recognized is inserted in the class corresponding to the closest prototype. The extended variant uses a sequence of prototypes for each class and the mean distance value instead of the distance value [4,13].

We know that each speech signal s_k, $k \le n$, is produced by one of the N speakers but we do not know which one of them.

Each speaker is related to a class and to a training subset. For example $\{S_1^i,...,S_{n_i}^i\}$ corresponds to the i^{th} speaker.

To identify the right speaker for each utterance, s_k must be inserted in the class corresponding to the smallest mean distance between speech feature vectors. This means that distances between different sized vectors, like $V(s_k)$ and $V(S_j^i)$, must be computed. Linear metrics, like Euclidean distance, cannot be used in this case, so we create a new metric. We propose a special nonlinear metric which is able to compare matrices having a single common dimension, like the matrices representing our speech feature vectors. It derives from the Hausdorff metric for sets [13, 14], described as:

$$h(A,B) = \max_{a\in A}\{\min_{b\in B}\{dist(a,b)\}\}, \quad (2)$$

where *dist* is any metric between the points of sets A and B. We obtain the following Hausdorff-based distance:

$$d(A,B) = \max\left\{\sup_{1\le k\le p}\inf_{1\le j\le m}\sup_{1\le i\le n}\left|b_{ik}-a_{ij}\right|, \sup_{1\le j\le m}\inf_{1\le k\le p}\sup_{1\le i\le n}\left|b_{ik}-a_{ij}\right|\right\} \quad (3)$$

where $A=(a_{ij})_{n\times m}$, $B=(b_{ij})_{n\times p}$, and n represents here the row number. The metric given by (3) constitutes a satisfactory discriminator between sound feature vectors in the classification process [13]. Thus, each speech signal s_k must be inserted in a class indicated by:

$$id = \arg\min_i \frac{\sum_{j=1}^{n_i} d(V(s_k),V(S_j^i))}{n_i} \quad (4)$$

where d represents the newly created distance. This means that the vocal utterance corresponding to s_k was generated by the id^{th} speaker.

III. NUMERICAL EXPERIMENTS

In this section we present some numerical results of our experiments. So, let us consider the following data as an example: the speaker number $N = 3$, the vocal utterances number $n = 6$, $m = 3$ and the transcript of the utterances is the word *start*. We create also a training set containing two vocal prototype signals for each speaker. The six signals of the training set, $\{S_1^1,S_2^1,S_1^2,S_2^2,S_1^3,S_2^3\}$, are represented in Fig. 1.

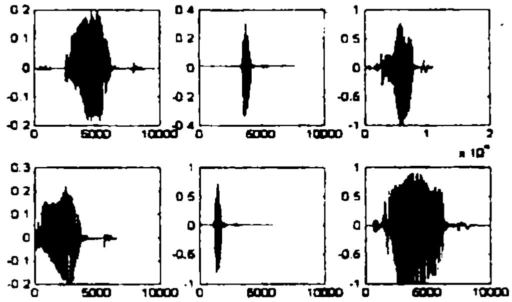

Fig.1. Training set

Each column from the figure above corresponds to a speaker. For each $i\le 3$ and $j\le 2$, the training feature vector $V(S_j^i)$ is computed as a 3 row matrix with a column number depending on S_j^i length. The six signals of the vocal utterances are displayed in the next figure. Their speech feature vectors, $\{V(s_1),...,V(s_6)\}$, are also computed.

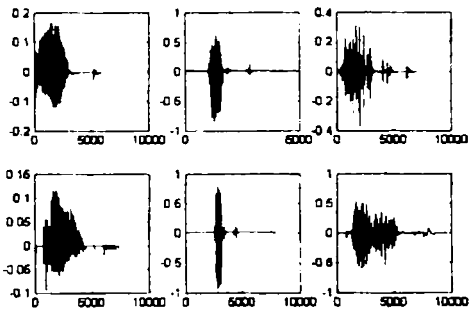

Fig.2. Initial speech signals

The distances, given by relation (3), between the training vectors $V(S_j^i)$ and the feature vectors $V(s_k)$ are then computed. The computing results are displayed in the next table. In the column corresponding to each s_k there are registered the mean distance values $\frac{\sum_{j=1}^{n_i} d(V(s_k),V(S_j^i))}{n_i}$ to each class (speaker), referred by 1, 2 or 3. The minimum distance value must be located on the row corresponding to speaker which produced s_k.

Table 1

	s_1	s_2	s_3	s_4	s_5	s_6
1	0.85	1.05	1.92	1.93	0.84	1.33
2	1.89	1.55	1.34	1.31	1.65	2.33
3	1.85	2.27	3.15	3.39	0.92	1.09

Knowing that s_1 and s_2 are produced by the first speaker, s_3 and s_4 by the second, and s_5 and s_6 by

the third. from Table 1 it results that a single classification error is produced. In the column of s_5 the lowest value is wrong placed on the first row instead of the third.
We may perform the same experiment for other transcripts.
If we use the spoken word *stop* instead of *start*, we get the numerical results registered in Table 2. From this table it results that each registered value is placed in the right place.

Table 2

	s_1	s_2	s_3	s_4	s_5	s_6
1	1.26	1.31	1.84	2.91	1.67	1.61
2	1.56	2.51	0.81	0.93	2.10	1.70
3	1.93	2.03	2.59	3.19	1.43	1.03

By applying this procedure on a number of 300 utterances, 3 speakers, we obtained satisfactory speaker recognition results. Using the proposed automatic text-dependent method, the recognition rate of our approach is about 80-90%.

IV. OTHER RECOGNITION METHODS

Generally, temporal processing and filtering in speech feature extraction are used with the purpose to increase the performance and robustness of the automatic speech recognition process. Conventional speech analysis techniques for automatic recognition (such as Mel Cepstrum or Perceptual Linear Prediction – PLP - values) in combination with dynamic features (estimates of derivatives of cepstral trajectories - „*speed*" and „*acceleration*" coefficients) proved to be sub-optimal [15], [16]. In [17], H. Hermansky argued against "blind implementing of scattered accidental knowledge" when neurological studies come to teach that human auditory perception has the "ability to suppress some parts of information in the speech message" to understand. Thus, selective use of auditory knowledge, optimized on real speech data should be used. (Side comment, a child recognizes mother's voice by only some sounds, no need of words!). "Providing speech evolved to be heard, properties of hearing should be reflected in the speech signal"[18]. From music we know that cochlea does a kind of „octave" spectral analysis of the sounds it receives. If the initial speech representation could be the speech waveform, the short term power spectrum could constitute criteria in the initial speech representation to derive features for speech recognition. In [18], [19] the optimization is done using LDA (linear discriminant analysis). LDA of the short-time Fourier spectrum in speech yields spectral basis functions which provide comparatively lower resolution to the high-Frequency region of spectrum. This is consistent with critical-band resolution and is shown to be caused by the spectral properties of vowel sounds, for speech recognition. LDA can yield either spectral basis functions (if feature vectors are spectral vectors) or FIR RASTA filters [20], if the feature vectors are cut out of time trajectories of spectral energies. But this technique considers the impulse response of FIR filters, implied by delta and double-delta features, too short and their frequency response, too selective. The lengths of impulse responses of discriminant RASTA [21] filters, derived by LDA are of the order 200ms and the frequency response is a BP with pass-band of about 1-15Hz.
What we propose is a method, that partially neglect the first lower bands, very high in amplitude and therefore in energy, taking only 2 or 3 bands up to 1500Hz, to select features on them, and concentrating more in exploring higher bands, around the high formantic frequencies of the vowels, where seems that refined characteristics came to make a difference between speakers. In [22] and [4] we stated that it is possible to disregard spectral correlations altogether and to have a very good indication on recognizing the speaker even from rather vowel spectral structure selected from temporal patterns of spectral energies.

IV. AUTOREGRESSIVE PARAMETERS IN SPEAKER RECOGNITION METHODS

In scientific research, autoregression (AR), time series, prediction, are important tools in estimating insight features [23], [24].
Using the AR coefficients with and an ART classifier the speaker recognition coefficient rose up 92%.
The main steps of an autoregressive schema we designed as automatic speaker recognition tool [5], presented in Appendix 1, are as follows: compute the signal spectrum, by a classical Fourier Transform (FFT), then plot the temporal evolution of spectral energy, detect the maxima and the inflection points by inverse descent procedure: select a number of 4-7 frequency bands, having the limits bounded by the inflexion points. On each band is practiced a zero phase filtering [5], [25]. In fact the selected bands correspond this way to the formantic areas of main vowels most often presented in a certain language, and the segments that interest us more are situated in high frequencies, that proved to be more selective in speaker recognition.
The AR coefficients are computed on each band, their values constituting the training set for a clustering structure. To them, we added, in more trials different complementary information on the respective band, as we previously discussed, with various results.
When the classification results from all available bands are combined, the performance is comparable to conventional recognizers. We obtained results ranging from 80 up to 90%, depending on the recognition structure and on the number of coefficients used as features on each channel. MLP and RBF, compared gave slightly different results (some percents) in favor of the second one, having as

disadvantage a complicate computing structure. Actual results obtained with ART-2 structure raised up to 92%, with essays made on different resolution on sub-band selection, centered more in higher frequency range, the centralization made - coordinated to the high formantic principal vowels structure. A comment, here, has to be done: the published figures were obtained on carefully selected training sets, with a preparing stage that excepted from the training group all the individuals that could give aberrant scores.

The C5 structure is a very useful tool in this sense clustering and in the main time selecting data that have characteristics out of the normal range of the classification [26].

Special studies dedicated to variances obtained in recognition values, due to the classifier structure, in order to find the optimal-result, are further conducted.

V. ADAPTIVE RESONANCE THEORY FOR CLASSIFICATION

ART [27], [28], [29] - neural unsupervised, feedback models that realize and develop a theory called adaptive resonance theory, introduced by Grossberg in 1976. The term "resonance" refers to the so called *resonant* state of the network in which a category prototype vector matches the current input vector close enough so the orienting subsystem will not generate a reset signal to the F2 layer. In this case, the activity pattern in the F1 layer causes the same F2 node to be selected, which in turn sends the same prototype vector down to the F1 layer, which again matches the current input close enough, and so on. The network learns only in its resonant state, associative learning of many-to-one and one-to-many maps, where many-to-one learning includes both categorization and associative prediction. ART is capable of developing stable clustering of arbitrary sequences of input patterns by self-organization. More varieties of ART networks exist: [illegible] which is able to cluster binary input vectors), ART-2 [31] (cluster real-valued input vectors), ART-2A (fast version of the ART2 learning algorithm), or more sophisticated - MART [32] (multi-channel ART, for adaptive classification of patterns through multiple input channels) fuzzyART [33], ARTMAP [30] (supervised version of ART learning arbitrary mappings of binary patterns). LAPART (an ART-based neural architecture for pattern sequence verification through inferencing) [34], etc.

We used an ART2 structure [27]. We tried it for 8 and respectively for 16 [illegible] speakers [illegible] with AR features selected on 7 bands, on 20 different command words, each presented word repeated 10 times. We obtained 92% coefficient of recognition.

VII. CONCLUSIONS

Speaker recognition has constantly grown, as research domain, during the two last decades. The struggle to improve the recognition coefficients previously obtained, pass by different attempts and trials, for each gained percent. The recognition rate was around 90% in both methods we presented (slightly different in AR-ART), but in order to obtain a good reliability (similar with the human cortex process of decision) in a speaker recognition system, results, obtained by one procedure have to be at least doubled by a reserve system. Computing a weighted aggregation [35], [36], by using confidence coefficients, reinforce recognition probity.

Smart appliances have to perform an inference on more parallel methods, simultaneously run, in order to aggregate results, or to be able to organize hierarchic structures, in an automatic attempt of clustering (as further project).

VIII. ACKNOWLEDGEMENTS

We would like to express our particular gratitude for the suggestions made by Mr. *Anthony J. Grichnik* (Strategy & Technology Manager, Technical Services Division, Caterpillar Inc. Peoria, Illinois).

APPENDIX 1

GENERAL SCHEMA FOR SPEAKER RECOGNITION WITH S-TREE2 CLUSTERING

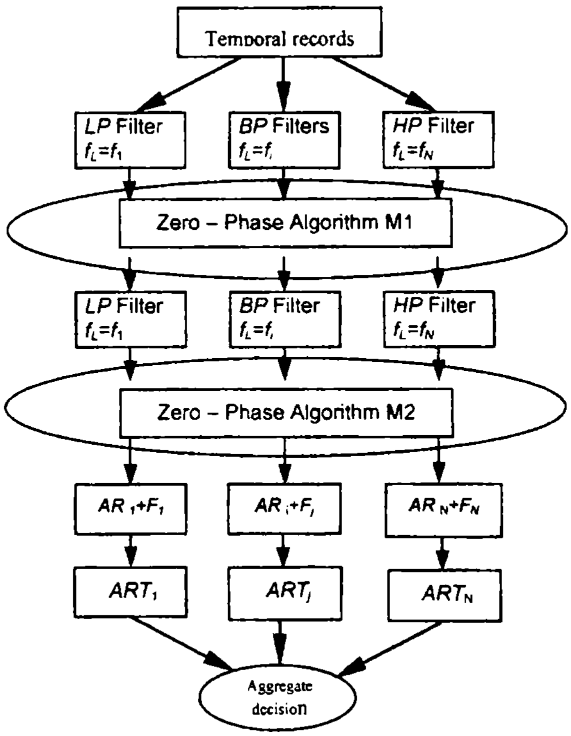

Fig.3. A general schema for AR – ART speaker recognition.

In Fig. 3. we have: low pass (*LP*) filter, band pass (*BP*) filters - for more central frequency bands, high pass filter (*HP*), the mirror 1 (*M1*) and mirror 2 (*M2*) - an algorithm in order to preserve the zero-phase filtering on each band, autoregressive coefficients (*AR*) and other features *F* added on option.

The classifier selected is an ART-2.

The inference for the decision is a weighted one, in accordance with the differentiated importance of the frequency bands, in speaker recognition process [7].

REFERENCES

[1] N. Do Minh, *"An Automatic Speaker Recognition System"*, Swiss Federal Institute of Technology, Lausanne, 2000.

[2] S. Furui, "Speaker-independent isolated word recognition using dynamic features of the speech spectrum", *IEEE Transactions on Acoustic Speech and Signal Processing*, 1986, Vol ASSP-34, No.1, pp. 52-59.

[3] L. Rabiner, B.H.Juang, *"Fundamentals of Speech Recognition"*, Englewood Cliffs, NJ: Prentice-Hall, 2000.

[4] M. Costin, T. Barbu, A. Grichnik, "Tips on Automatic Speaker Recognition, New Decision Tools", *CD - ECIT* 2004.

[5] M. Costin, A. Grichnik, M. Zbancioc, "Tips on Speaker Recognition by Autoregressive Parameters and Connectionist Methods", *International IEEE Conference SCS Proc.*, Vol. 1, pp. 169 - 172, 2003.

[6] M. Costin, M. Zbancioc, "Improving Cochlear Implant Performances by MFCC Technique", *International IEEE Conf. SCS Proc.*, Vol. 1, pp. 449-452, 2003.

[7] M. Costin, D. Galea, "Decision and Evaluation in Speaker Recognition Connectionist Attempts: New Methods and Trials", to appear in *SACCS 2004*.

[8] M. Costin, M. Zbancioc, A. Ciobanu, C. Berger Vachon, "Some Attempts in Improving Cochlear Implanted Patients Performances, Modeling and Automatic Methods", *IXth IPMU International Conf.*, Annecy France, pp 711-718, Vol. II, 2002.

[9] M. Zbancioc, M. Costin, "Using Neural Networks and LPCC to Improve Speech Recognition", *International IEEE Conference SCS Proc.*, Iaşi, Vol. 1, pp. 445–448, 2003.

[10] M. Costin, M. Zbancioc, "Hints about some basic indispensable elements in speech recognition", *Computer Science Journal of Moldova*, v.10, n.2., pp.169, 2002.

[11] Logan B., "Mel Frequency Cepstral Coefficients for Music Modelling", In *Proc Int. Symposium on Music Information Retrieval (ISIMIR)*. Plymouth, MA, 2000.

[12] R. Duda, P. Hart, D.G. Stork, *Pattern Classification*. John Wiley & Sons, 2000.

[13] T. Barbu, *"Discrete Speech Recognition using a Hausdorff-based Metric"*. Proceedings of the 1st International Conference of E-Business and Telecommunication Networks, ICETE 2004, Aug. 2004, Vol. 3, pp.363-368.

[14] N. Gregoire, M. Bouillot, "Hausdorff distance between convex polygons", Web project for the course *CS 507 Computational Geometry*, McGill University, 1998.

[15] H. Hermansky, "Perceptual linear predictive (PLP) analysis of speech". JASA 87(4), April 1990.

[16] B. S. Atal. "Effectiveness of linear prediction characteristics of speech wave for automatic speaker identification and verification", Journal of the Acoustic Society of America, (55):1304–12, 1974.

[17] H. Hermansky, "Auditory Modeling in Automatic Recognition of Speech", *Proc. of the first European Conf. on Signal Analysis and Prediction*, pp. 7-21, Prague, Czech Rep. 1997.

[18] N. Malayah, H. Hermansky, "Data-driven spectral basis functions for automatic speech recognition", Speech Communication Volume 40 , Issue 4 (June 2003) Pages: 449 - 466 , 2003.

[19] Q. Jin, A. Waibel, "Application of LDA to Speaker Recogn.", ICARCV2000, http://www.is.cs.cmu.edu/nojs/2000.html

[20] H. Hermansky, "Exploring Temporal Domain for Robustness in Speech Recognition", Invited paper, the *15th International Congress on Acoustics*, 1995.

[21] H. Hermansky, N. Morgan, A. Bayya, P. Kohn, Rasta-PLP Speech Analysis, *ICSI Technical Report TR-91-069*, Berkeley, California, http://www.icsi.berkeley.edu/Speech/rasta.html.

[22] M. Costin, M. Zbancioc, A. Ciobanu, C. Berger Vachon, "Some Attempts in Improving Cochlear Implanted Patients Performances, Modeling and Automatic Methods", *IXth IPMU International Conf.*, Annecy France, pp 711-718, Vol. II, 2002.

[23] M. B. Priestley *"Non-linear and non-stationary time series analysis"*. Academic Press, 1988.

[24] P.J. Brockwell, R.A. Davis, *"Introduction to Time Series and Forecasting"*, Springer 1994.

[25] K. S. Mitra, *"Digital Signal Processing"*, 2nd Ed. McGraw-Hill, pag.531, 2001.

[26] R. Quinlann, C5, http://www.ph.tn.tudelft.nl/PRInfo/software, 1997.

[27] J. A. Freeman, D. M. Skapura, *"Neural Networks: Algorithms, Applications and Programming Techniques"*, Addison-Wesley, 1991.

[28] L. Heins, D. Tauritz, *"Adaptive Resonance Theory (ART): an Introduction"*, 1995, http://www.liacs.nl/~dtauritz/

[29] G. A. Carpenter and S. Grossberg. Art 2: "Self organisation of stable category recognition codes for analog input patterns". *Applied Optics*, 26:4919-4930, 1987.

[30] G. A. Carpenter, B. L Milenova, & B. W. Noeske, "Distributed ARTMAP: a neural network for fast distributed supervised learning", *Neural Networks*, 11, 793–813, 1998.

[31] G. A. Carpenter, Stephen Grossberg, "Invariant pattern recognition and recall by an attentive self-organizing art architecture in a nonstationary world", *Proceedings of the IEEE First International Conference on Neural Networks*, pages II-737-II-745, June 1987.

[32] M. Fernandez-Delgado, S. Barro. "MART: A multichannel art-based neural network", *IEEE Transactions on Neural Networks*, 9(1):139-150, January 1998.

[33] G. A Carpenter, S. Grossberg, N. Markuzon, J. H. Reynolds, & D. B. Rosen, "Fuzzy ARTMAP: a neural network architecture for incremental supervised learning of analog multidimensional maps", *IEEE Transactions on Neural Networks*, Vol. 3, No. 5, pp. 3, 698–713, 1992.

[34] K. M. Edlund, T. P. Caudell , "Realtime Visualization of the Learning Processes in the LAPART Neural Architecture as it Controls a Simulated Autonomous Vehicle", *IEEE-INNS-ENNS Internation Joint Conference on Neural Network (IJCNN'00), Vol.3*, July 24 - 27, 2000, Como, Italy.

[35] Ph. Smets, E.H. Mamdani, Didier Dubois, Henri Prade, *Logics for Automated Reasoning*, Academic Press, 1988.

[36] P. Shenoy, G. Shafer, K. Mellouli, "Propagation of Belief Functions, a Distributed Approach", *Uncertainty in Artificial Intelligence 2*, ed. J.F. Lemmer, L. N. Kanal, Elsevier Science Publisher, pp. 325-335, 1988.