THE OVERVIEW OF ON-LINE SPEECH PROCESSING ALGORITHMS FOR SPEAKER IDENTIFICATION SYSTEMS

Korshakov A. V. Russian Science Center "Kurchatovsky Institute", Moscow, Russia

Abstract. This paper describes main trends in speech recognition and speaker identification/verification.

Speech recognition systems crucial for modern information society and can be applied in a various fields. Direct speech recognition can be used in voice identification systems, which in turn can be used in various types of security and access control systems. Algorithms, considered in this paper also can be a new way to communicate with a surrounding world for the people with some physical disability. It also can ease some time consuming work such as typing. Applications in robotics are also possible.

Such conveniences and advantages on a full scale can only be achieved if voice recognition function, speech interpretation or answer for voice directives will have an acceptable level of errors and simplicity in usage in real world applications. It is also necessary to operate in "real-time" for such systems.

Systems of voice recognition and voice typing in a present day are not far from the reality, unlike for the voice recognition systems, which in conditions of real world applications are rather close to science fiction.

Nevertheless, there are various facts exists, confirms the possibility of creation of reliable voice recognition system in the near future [1].

It is necessary to notice a tremendous progress had been made in the field from the beginning of the works on artificial intelligence.

The base of first research was a classical linguistic. The attempt of direct approach to voice interface was fruitless because of insufficient of first computer's calculation power, as well as because not complete understanding of such phenomena as "speech", which have some mysterious aspects even for today.

In present day insufficient of computational power (which still preserves) and completion of hypothesis on natural language phenomena try to compensate in several ways. As for example in [2] suggests "intermodel" approach which implies idea of use all known models of natural language (describing language in several aspects) and models of artificial intelligence language perception, performing a bridge between determinacy of mathematics and natural fuzziness of language. This is only one of examples.

Later the field of "near-language" computer science was enriched by theory of artificial neural networks, neuroscience and general theory of information processing. Noticeable amount of data was provided by hypothesis on human brain functioning during speech processing and direct MR-observations of people during processing of some linguistic tasks [3].

It is necessary to mark several levels of human speech recognition. Firstly, on acoustic level or phoneme level the separation of incoming speech into independent language-specific phonetic units takes place. Secondly, algorithms of morphological level or level of words interprets the words (not yet the meaning of words) with usage of information on phoneme sequence provided by phonetic level. Next sequence of words passed to semantic level, where (at least theoretically) system deals with interpretation of meaning of the words. At the final stage at sintagmatic or syntactic level (they are usually merges in speech and voice recognition systems) the processing of complete logical notion or sentence takes place. Two last levels are closely related and sometimes practically merge with semantic level. Interpretation of sentences as a whole might be the most difficult work for computer program and cannot be implemented at acceptable level in modern speech processing systems. This fact also concerns such speech phenomena as prosodic variation (intonation in particular, which in turn influences essentially on phonetic level). Sarcasm is a good example.

Because of fuzziness of natural human language the borders between several levels are blurred and it is difficult to completely distinguish one level from another. For instance, it is common for information from phoneme level proves to be ambiguous and it is necessary to involve data from one or more upper processing levels for adequate identification and interpretation of particular sound.

Besides, there are unique problems at every level with no optimal solution at the present time.

Let us introduce a short classification of existing natural human speech and voice processing algorithms, suitable for usage man-machine interfaces.

First of all every methods can be divided at two big classes – the algorithms depending on speaker and speaker independent algorithms. First class must be trained (in terms of artificial neural networks, though the usage artificial neural networks not necessarily implies) before they can be used mainly by only one user with quite precise results in comparison with systems of second class. The systems and algorithms can be successfully used for speaker recognition or user identification via voice.

Algorithms of second class has an advantage of universality and capable of speech processing for every user indiscriminately. However, this advantage naturally brings many complications in structure and implementation of such algorithms as well as in preliminary tuning. This in turn cause to rising of general error level. Nevertheless this type of methods in case of achieving acceptable levels of simplicity, convenience and error tolerance is a type of choice for everyday live.

This is not only concerns speech recognition, but also speaker identification which in general means distinguishing deferent people and/or discriminate speakers for several simple classes such as “friend/foe”. One must also separate identification problem and verification problem which implies prove that the speaker is an individual he or she claims to be.

Phonetic level is absolutely essential for performing such tasks and its information can only be supplemented by interrelation with any upper level of speech processing.

Problems in a correct and errorless phonetic speech recognition and/or speech processing are usually places by poor acoustic environment in which the reception take place. This primarily concerns awareness, focus of a perception process, orientation on speech source and influence of noise.

Speech perception in a poor acoustic environment may be considered as a separate difficult task. A good example is a cocktail party problem, which in order to solve, a methods of independent component analysis (ICA) group may need to be applied. Generally this class of problems is subclass of computational auditory scene analysis (CASA) problem [4].

Speech signal is a quasi periodic process which time recordings including some periods of silence and some periods of spiking activity which can be refer to some “events” or some meaningful sound. Those periods are carry main meaning load of speech, and treated as phoneme. Phonemes by definition are minimum meaningful unit of phonation with influence of what is to be said. Thus identification of events noted above in signal is main problem for speech perception. Usually this problem solves with various methods of spectral and cepstral analysis using several time-frequency tags for separation of areas with signal events. Usage of spectrographic representation of signal is also common. The particular way to solve this varies from paper to paper.

Large fraction of phonemes has unique for each one time and spectral signature, yet not unique enough to distinguish them definitely. This signature is varies depending on speaker and even considering only one speaker can shift depends on health condition and mood of a person. Nevertheless, identification of phonemes and verifications of phonemes pronouncing way can be carried out with some level of success. Same stands for identification for classification purposes (considering several classes). Classification and clusterization problems are crucial for computational phonetic analysis and implementation of speech interface. Effectiveness of classification/clusterization algorithm is a foundation of successful speech stream separation for basic phoneme sequence. Next classification problem arise during building groups of phonemes consisting in words. The solution of this problem strongly depends on the success of solution of previous one. It seems to be achieving of 100% efficiency is not possible. There are plenty reasons for this. For instance many sounds usable in “every day speech” have a tendency to combine with each other (and

begin to represents mean between 2 or 3 of its neighbors). It is also exists tendency for phonemes to have a fuzzy boarders. All this complicates classification and separation problem sufficiently. Fricative class of phonemes plays a necessary role in several natural languages (for example in Russian) have time-spectral characteristics almost the same as such for simple noise.

There are many classification methods and algorithms. Some of them are use unique properties of particular "target" natural human language. Second group of methods based on theory of hidden Markov models. In every algorithm it is necessary to use some sort of classification quality or compare criterion. And almost in all algorithms there is some form of correlation dependencies analysis between standard "events" stored in thesaurus and "events" from incoming phonetic sequence takes place.

The essential part in solution of problems noted, plays algorithms build on artificial neural networks and/or closely related "new" mathematical methods such as fuzzy logic and genetic algorithms. It is popular to call this group of methods with nickname "intellectual". And implementation of such methods are quite successful. Besides theoretically based speech processing method (such as recognition and/or verification of user via speech) clearly must contain some of fundamental principles of speech generation by human (at the phonetic level as well as at semantic level).

Seems to be trying to understand algorithm of speech signal analysis in time-frequency domain in a human brain and "in a way" try to mimic it is a most fruitful approach.

Word builds by combination of phoneme. Phoneme sequence has a tone and intonation as property of particular speaker. Nevertheless almost every man capable of generating recognizable speech and listener in turn almost in every cases by the sound of voice clearly understand which man he speaks to.

Speech recognition methods may also be based on comparison of some parts and events of a target phonetic sequence and sequences of imitating speech produce by some models of human voicing tract. Here different sounds represents by superposition of sources of periodic and noises signals. Before comparison sounds from both sides pass through filter cascade.

All considered methods and algorithms concerns in this paper are sensitive to speed of data processing and capacity of data storing units. Commonly this became a reason for compromises in relation cost/quality. This is possible because all algorithms have a large number of adjusting parameters with known variation corridor.

It must be noted, however, that the idea of definite and sole identification some time sequence by meaningful tag is very old and facing old difficulties in a way of its implementation. Besides the prosodic cause words variability which exists almost in all natural languages also plays part in this problem. It is rather difficult (if possible at all) adequately converge phonetic sequence to some sort of universal notion for example for every context where one can meet numerical and noun in some case. Also it is possible for neighbor words of varying word to vary them self.

All recognition algorithms works based on vocabulary, consists of some number of words, for some numbers of pronunciation. This number drastically defers for different languages and recognition modes.

For the purpose of fine tuning on voice of certain user size of necessary vocabulary is enormous consisting of several thousand words and phonation variants which must be recognized in case of continuous pronouncing. This recognition mode used in dictating systems, capable on voice commands reception and even performing voice-to-text conversion for mail dictation, in case only one registered system user. But, for the systems of general use it is crucial to operate without voice tuning. This type of software/equipment also works with vocabulary, but in this case (in existing applications) it consists of much less meaning words, plus, almost all variation of phonation of words by variety of speakers. In a way corpus of words and its phonation can be considered as a signature of speaker for such corpus included specific features of some "key" words phonation.

Vocabularies for speaker independent voice-to-text dictation systems bases on speech samples, acquired from representative set of native speakers. Commonly averaging of fragments, corresponding to particular words on all speakers is applied.

For present time, typically speaker independent vocabularies and analyzers based on them could maintain appropriate recognition of only several words, such as simple numericals from zero to nine, simple commands (for instance “yes”, “no”), in case of its separate and accurate pronouncing. Another variant is just recognition of all alphabet letters. Needless to say that commands recognition is more complicated task.

The theoretical research in the field is conducted by many research groups all over the world. The large companies must be noticed in the first place. Among them are IBM, Intel, Microsoft, AT&T. This companies studies speech recognition for about ten years.

In Russian Federation and former USSR several laboratories also conducted research on the topic. For example in laboratory of automated mass service systems, institute of control science RAS works started more than 30 years ago. Main theoretical and practical direction of research in a present time is an application on speech recognition of continuous speech in public service systems. System includes recognition of Russian and other language as well [5]. Different mathematical models describe speech recognition process were developed in recent years. Institute of system analysis RAS conducts task-oriented research on speech recognition [6]. Task-oriented includes usage of several theoretical approaches, development and implementation of real-time analysis speech perception and recognition methods, speech generation and encoding algorithms. Novelty of decisions offered consists in using “partial” artificial neural networks analysis of speech signal in couple with allocation of constant signal features and applying phonological and some engineering knowledge of fine structure of speech signals.

“Istra-soft” company’s research on speech technology spreads to several areas such as speech files compression, speech recognition, text-to-speech conversion and person or speaker verification and identification algorithms [7]. Among other achievements must be noted algorithm of real-time phoneme separation form the continuous speech. This method consists of adaptive sound signals analysis and it’s parameters for purpose of features identification caused by the form of vocal tract in the moment of phonation act. Identification of vocal tract parameters leads directly to identification of phoneme pronounced.

Since year 1996 “STEL – Computer Systems” in cooperation with leading specialists of Lomonosov Moscow State University philological faculty, RAS Computing Centre and others organizations working on project of prototyping speaker independent speech recognition on Russian language [8]. From methodological point of view project based on usage of modern speech signal processing methods and Markov model mathematical apparatus for describing phonetic, semantic and syntactic laws of Russian language.

It is obvious by now that’s speech is strongly variable and for correct identification of its features for purpose of person verification and/or speech recognition it self a non trivial works is needed to be done. This point proves by the fact the newest papers on the subject includes more and more complicated methods of signal processing and adding new elements to series of features speech can be processed upon. For instance in [9] described speech/speaker recognition based on representation of speech information as stream 2D time-frequency vectors. Classification handled by neural network, which accepts as an input low-frequency 2D wavelet transformation of some spectrogram areas. Source sound representation is a sonogram or a representation based on Hermit - like transformation. The comparison of the results produced by those signal representation for a speaker independent speech recognition and context independent speaker identification problems are given.

There is a wide class of works with “build-up” algorithms. This involves usage of one simple or classic in area of speech processing transformations (mel-scale cepstrum transformation) and then usage of some additional non-classic or author-invented methods classification/clusterization. It is also usual that the results of such algorithms verifies by the results on some simple methods such as separation by classes with minimum distance principle or with the empirical separation i.e “by hearing”. The last type of verification may also carry some statistical calculations. It must be noted that the “by hearing” criterion, even it is not quantitative in a full scale, in relation to discussing problem often means the best and reliable solution to achieve best recognition results brings by

algorithm. This fact makes it usable for final and fine tuning of speech recognition systems.

The central place in signal analysis and in speech signal analysis in particular, presently occupied by independent component analysis methods. The usage of this group of mathematical methods leads to appropriate solution for “cocktail party problem”. The problem statement means “blind” separation of speech stream consisting of voices of several speakers mixed up, to independent signal components (BSS – Blind Source Separation and ICA – Independent Component Analysis).

The primary purpose for ICA in case of speech processing is estimation of selected components of speech featured to uniquely identify speaker, sex of a speaker, spatial dimension between speakers and time synchronism of speech.

Presently there have been build many methods of ICA class. Most of them can be applied not only for speech signal processing. Several ICA methods can be programmed in terms of neural networks [10]. Some of those methods are closely connected with the nature of signal, which makes them not completely “blind”. There are several fields, where ICA methods finds application most often. Those are for example brain study problems like EEG-analysis [11] and MEG, denoising and signal processing in MRI and fMRI images and studies [12]. There is also a probability for successful application in the field of economic time series analysis. Some of the ICA algorithms can function without any preliminary information on nature of the signal and its properties, the others does not. Nevertheless, all of them capable to function only with limited precision, marked out not a speech stream of person of interest, but only a set of components, including parasitic ones (mainly random noise). The usage of useful components only in a case of reverse procedure provides us with the almost “clean” signal. The bypass product of ICA decomposition is parts of initial signal, brought by algorithm to a certain speech channel (one per each speaker), but sometimes those parts are “not-speech”. This part of a speech channel might be a channel of prosodic features or “noise” channel. First ones are useful for person identification means. The precision of described process are not always goes at an equal degree. Sounds produced by several speakers’ overlaps not only in several time intervals, but also in the frequency domain. This may cause loss of some fragment with necessity of its following reconstruction.

Considerable quantity of ICA algorithms have a build in property to sort output components by increasing/decreasing of main separation criterion. As criterion usually consider such measures of similarity/distinction of two signals like statistical likehood of hypothesis of ”signal similarity”, mutual entropy, mutual information, negentropy, curtosis e.t.c. However, during sorting procedure the signal’s component of interest (indiscriminately information one or prosodic/noise one) does not necessarily occupies at fixed position, which arises problem of its identification.

There are many papers released later on subject of blind source separation and independent component analysis in particular. For instance [13] describes elegant ICA-methods for cocktail party problem solving based on “missing feature” technique. Information criterion of signal separation used there was “uncertainty information” of sound signal. The paper stated efficiency of method even in strongly reverberating acoustic environments.

This is only one of many examples. Another one is [14]. In this work independence criteria used maximum of speech time-series negentropy. The authors provide reader with result’s comparison between described method and other methods of high order statistic. A new criterion function is described build on approximation of speech time series by power functions. This method also gives good results in reverberating acoustic environments.

The next wide and impotent class if ICA/BSS methods aim to the features belongs uniquely to a certain signal. In [15] ICA method applied to Fourier power coefficients of speech signal frames of limited length. The results components treated as vectors suitable for signal identification and consequently for building training set for some classification algorithms. Some subsets of those vectors correspond to noise influence from estimated frequency diapason. Excluding of all members of this subset (setting to zero its mixing coefficients) can cause the nullifying corresponding signal distortion. Paper results compared to results given by widely used speech recognition methods such as Mel-frequency cepstrum transformation.

In [16] “algebraic” ICA (AICA) algorithm is described. Algorithm using purely algebraic

separation criterion based only on calculation vector distances. The usage of such simple criterion gives a considerable decrease of computational costs for estimation mixing matrix and increase of precision. Previously this research group developed “geometric” ICA-algorithm or “geo-ICA”. But most importantly the algorithm mentioned capable to solve problem of every reasonable dimension. This means problem can be solved for any reasonable quantity of speakers. And increasing of dimensionality leads only to linear increasing of computational costs.

It must be noted that the target noise filtration of time series and speech signals in particular one of the most promising fields of using ICA/BSS methods. For example in [17] described an interesting method of adaptive detection an exclusion of noise implements ICA, wavelet analysis and spectral analysis of noise speech signal. Firstly input signal with usage of wavelets an entropy criterion separates to two components. First one consists on noise and speech. Second one noise only. On the second step for each part ICA decomposition were performed for more precise noise components estimation and detection, the “noise part” acquired at the first step used as an “etalon” of in-signal noise. Finally filtering method calculates spectral signature of noise and subtracts it in the frequency domain.

Alternative approach to decompose signal to components, in which it is easer to interpret their importance is empirical mode decomposition or EMD. This algorithm allows to extract all oscillating modes (in many practical problems they may considered as components) composing incoming signal. Method can provide user with reasonable result whether signal represents a linear process or not, whether it represents stationary process or not.

Empirical mode decomposition algorithms decompose signal to a set of intrinsic mode functions (IMF) possess such properties as symmetry, unique of local frequency, and non equality of frequencies for different IMF-functions in the same time.

EMD algorithms are well used with other methods of speech processing.

This on a full scale applicable to a speech signals. And as well some components might include unique features of speaker voice, on which identification/verification could be handled. The feature extraction capability is actually strong enough to reveal stress condition of a person by voice analysis [18]. This is so called VSA problem (Voice Stress Analysis), closely related to speaker identification problem. Again, among all components or “intrinsic modes” (or intrinsic functions - IMF) in this case, can be estimated those which during of pronouncing of some key words are statistically more often can be attributes of one speaker than another. Physiological microthremor always existing in the voice and on some degree unique to a person can be extracted from the speech with using an EMD algorithm. It must be carry out carefully though, because of some components may involve an emotional state rather than a vocal tract geometry feature reflection. Extraction of emotional state components also useful in many ways because, for example, if we considered them as a noise, it is naturally that after extraction signal analysis will be simplified [19].

In general most of the methods of speaker identification means algorithms based primary on wavelet analysis and hidden Markov models (see [20]), as well as on pure mathematical statistic [21]. Those are classic algorithms for signal processing and speech signal analysis in particular. Next group refers to a Viterbi algorithm, allow to estimate similarity between two signals and calculates likehood of incoming and etalon signal’s equality hypothesis [22]. Implementation of discriminant analysis are also common [23]. Classical methods suffer “classical” lacks such as for example necessity to analyze an enormous databases of signals for acquire statistically accurate estimations.

However the usage only “classical” methods is not a rule.

For example in [24] deals with linear predicative spectra building (FLPCS) for speaker identification purposes. FLPCS applies at the stage of features extraction, which classification conducted with general regression neural networks. Unconditional advantage here is simplicity of implementation and calculation speed.

Hidden Markov models are main but not only direction of research. The modifications of Markov models almost any kind are in used [25].

Classification and clusterization problems are inseparable from speech recognition. In this field implementation of neural networks is usual. Among them recurrent networks, multilayered

perceptrons and radial basis networks, also described several “exotic” algorithms based on neural networks. Recently a progress has been made in field of neural networks training algorithms. A good example of utilizing neural networking concept is [26]. This paper describes Kohonen self organized map with specific training method for purpose of speaker-independent speech analysis.

All concepts discussed till now are only basic engineering instruments for speech processing. According to [27], with using standard and classical algorithms for analysis of speech signal in time and in frequency domain development of online speech recognition system base on hidden Markov models, suitable for real-world application are only possible if such system will include some element of adaptation and adjusting to speaker voice (like it already is in most existing applications). However, even in this case it is necessary for person to speak closely to microphone, in a relatively clean acoustic environment and in constant manner. For every system the degree of constant engineering control can be decreased in case for example of vocabulary reduction. Sadly to say, but the best results, archived for today for speaker recognition is only 25% as correct answer percentage. For speech recognition – up to 99,5% on a small vocabulary and 96% for speaker independent speech recognition under the severe restriction on environment quality [9].

In the conclusion it must be noted that seems to be in present time computing power of modern calculation systems is not yet reach the adequate to real-world applications direct solving threshold to online, speaker independent, continuous, large vocabulary speech recognition problem and/or speaker identification/verification problem [29, 30]. Some drastic improvement need to be done in field of software and/or hardware.

Considering efficiency and practical errorless of human natural abilities of speech recognition it is obvious necessary to conduct a deeper studies of human brain for deeper understanding of how humans recognizes speech and identifies speakers.

References:

1. Р. К. Потапова «Речевое управление роботом: лингвистика и современное автоматизированные системы» Изд. 2-е, перераб. И. доп. – М.: КомКнига, 2005. – 328 с.
2. Теленик С.Ф. and Смичик Р. В. (2004) Межмодельный подход к разработке естественно-языкового интерфейса с использованием методов нечёткой логики. In: УкрПрог, 1-3 июня 2004 г., г. Киев, Украина.
3. Michael W. L. Chee, Edsel W. L. Tan, and Thorsten Thiel. Mandarin and English Single Word Processing Studied with Functional Magnetic Resonance Imaging. The Journal of Neuroscience, April 15, 1999, 19(8):3050–3056
4. http://www.cse.ohio-state.edu/~dwang/papers/Brown-Wang05.pdf
5. http://www.ipu.ru/s_006/s_006_000_000000000000000.htm
6. http://www.isa.ru/
7. http://www.istrasoft.ru/
8. http://www.stel.ru/company/
9. http://www.keldysh.ru/papers/2001/prep87/prep2001_87.html
10. С. Осовский. «Нейронные сети для обработки информации» М.: «Финансы и статистика» 2002, 344 с.
11. Weidong Zhou, Jean Gotman Automatic removal of eye movement artifacts from the EEG using ICA and the dipole model Natural Science 19 (2009) 1165–1170
12. MRI : basic principles and applications / Mark A. Brown, Richard C. Semelka. — 3rd ed.
13. Dorothea Kolossa, Hiroshi Sawada, Ramon Fernandez Astudillo, Reinhold Orglmeister, Shoji Makino Recognition of convolutive speech mixtures by missing feature techniques for ICA 2006 Proc. Asilomar Conference on Signals, Systems and Computers.
14. Rajkishore PRASAD1, Hiroshi SARUWATARI and Kiyohiro SHIKANO Blind Separation of Speech by Fixed-Point ICA with Source Adaptive Negentropy Approximation. Special Section on Multi-channel Acoustic Signal Processing -- Papers -- Blind Source Separation.
15. KASPRZAK Wlodzimierz; OKAZAKI Adam F; KOWALSKI Adam B; ICA-based speech features in the frequency domain. Lecture notes in computer science. Congress on

Independent component analysis and blind signal separation: 6th international conference, ICA 2006, Charleston, SC, USA, March 5-8, 2006
16. K Waheed, FM Salem Algebraic independent component analysis: an approach for separation of overcomplete speech mixtures. Neural Networks, 2003. Proceedings of the International Joint Conference on, Vol. 1 (2003), pp. 775-780 vol.1.
17. REVATHI A; CHINNADURAI R; VENKATARAMANI Y; A noise reduction technique of speech signal using ICA and spectral analysis. International journal of electronics 2007, vol. 94, no11-12, pp. 1171-1179
18. http://www.ijme.us/cd_08/PDF/140_ENG101.pdf
19. Sethu, V.; Ambikairajah, E.; Epps, J. Empirical mode decomposition based weighted frequency feature for speech-based emotion classification. Acoustics, Speech and Signal Processing, 2008. ICASSP 2008. IEEE International Conference on Volume , Issue , March 31 2008-April 4 2008 Page(s):5017 - 5020
20. Bovbel, E.I.; Kheidorov, I.E.; Chaikou, Y.A. Wavelet-based speaker identification. Digital Signal Processing, 2002. DSP 2002. 2002 14th International Conference on Volume 2, Issue , 2002 Page(s): 1005 - 1008 vol.2
21. http://www.journal.au.edu/ijcem/jan98/article5.html
22. http://www.aclweb.org/anthology-new/H/H92/H92-1069.pdf
23. http://research.microsoft.com/pubs/69350/jointhmm.pdf
24. Jian-Da Wu, Jin-De Rd, Bing-Fu Lin, Jin-De Rd. Speaker identification based on the frame linear predictive coding spectrum technique. Expert Systems with Applications: An International Journal archive Volume 36, Issue 4 (May 2009)
25. V. Digalakis, M. Ostendorf, J. R. Rohlicek. Fast search algorithms for connected phone recognition using the stochastic segment model. Human Language Technology Conference archive Proceedings of the workshop on Speech and Natural Language table of contents Hidden Valley, Pennsylvania Pages: 173 – 178, 1990
26. Najet Arous, Noureddine Ellouze. Cooperative supervised and unsupervised learning algorithm for phoneme recognition in continuous speech and speaker-independent context. Neurocomputing 51 (2003) 225 – 235
27. Nelson Mogran, Hervé Bourlard, Hynek Hermansky. Automatic Speech Recognition: An Auditory Perspective. Speech Processing in the Auditory System. Springer New York
28. pp. 309-338 2006.
29. Rajeev Krishna, Scott Mahlke, Todd Austin. Architectural optimizations for low-power, real-time speech recognition. International Conference on Compilers, Architecture and Synthesis for Embedded Systems archive Proceedings of the 2003 international conference on Compilers, architecture and synthesis for embedded systems. San Jose, California, USA SESSION: Embedded applications table of contents. Pages: 220 – 231, 2003.
30. Binu Mathew, Al Davis, Zhen Fang. A low-power accelerator for the SPHINX 3 speech recognition system . International Conference on Compilers, Architecture and Synthesis for Embedded Systems archive Proceedings of the 2003 international conference on Compilers, architecture and synthesis for embedded systems. San Jose, California, USA SESSION: Embedded applications Pages: 210 – 219. 2003