Problems and solutions for noisy speech recognition

J.-P. Haton

► To cite this version:

J.-P. Haton. Problems and solutions for noisy speech recognition. Journal de Physique IV Proceedings, 1994, 04 (C5), pp.C5-439-C5-448. 10.1051/jp4:1994592 . jpa-00253087

HAL Id: jpa-00253087

<https://hal.science/jpa-00253087>

Submitted on 1 Jan 1994

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Problems and solutions for noisy speech recognition

J.-P. HATON

CRIN/INRIA-Nancy, BP. 239, 54506 Vandœuvre-les-Nancy, France

Abstract

Automatic speech recognition has reached high level performances but it usually fails in coping with real-life, noisy environments. An essential reason is the mismatch between the conditions in which a system is trained and used. A large number of solutions have been proposed in order to solve this problem. Those solutions can be classified into two main, non exclusive categories. Firstly, signal processing and parametrization techniques can be used as a preprocessing step in order to enhance the SNR of the corrupted speech signal. Secondly, the different steps of the pattern matching process can be modified in order to account for the effects of noise.

This paper presents a brief survey of the noisy speech recognition field. We first summarize the major difficulties that are encountered in the development of a system, and we then introduce three main categories of solutions dealing with acoustical preprocessing and parametrization of the speech signal, statistical modelling, and recognition techniques.

1. INTRODUCTION

Speech recognition in adverse conditions has recently received increased attention since noise resistance has become one of the major bottlenecks for practical use of speech recognizers in real life. The methods proposed so far can be classified into four, non exclusive, categories:

- . signal acquisition,
- . acoustic analysis and parametrization,
- . reference modeling,
- . recognition techniques.

Most of the techniques developed so far at the analysis and parametrization levels are aimed at noise reduction in the presence of stationary noise, and they often fail in the case of non-stationary noise. It is thus important to develop robust techniques capable of dealing with non-stationary noise processes that are quite common in a number of applications areas (military environments, industry, etc.).

This paper presents a brief survey of the noisy speech recognition field. We first summarize the major difficulties that are encountered in the development of a system, and we then introduce three main

categories of solutions dealing with acoustical preprocessing and parametrization of the speech signal, statistical modelling, and recognition techniques.

2. OVERVIEW OF PROBLEMS AND SOLUTIONS

The difficulties of noisy speech recognition come from the various effects of noise on speech that can be summarized as addition of ambient noise to the speech signal, distortion of the signal and variations in articulation. They result in two main phenomena:

- the degradation of the performances of a speech recognition system if the conditions for learning and for test are different. For example, it has been found a degradation of an order of magnitude in the error rate for an isolated word recognizer when trained with clean speech and tested with noisy utterances at a signal-to-noise ratio of 18 dB [2]. That makes it necessary to increase the overall robustness of speech recognition since the simple solution which consists in having the same conditions for training and testing is seldom affordable and usually not realistic ;
- the modification of sounds pronounced in a noisy environment (the Lombard effect [1]). This effect is highly dependent upon the speaker, the context and the level of noise [3]. It is thus very difficult to quantify and to modelize.

In conclusion, the mismatch between training and testing patterns is the major problem to deal with in adverse, real life conditions of speech recognition.

A large number of methods have been proposed so far to cope with these problems, even though none is totally satisfactory. All these methods intervene at five main steps in the basic sequence of speech recognition processing:

- . speech signal acquisition: use of noise-cancelling microphones or of microphone arrays, adaptive or active noise cancellation techniques,
- . acoustic analysis and parametrization,
- . segmentation and speech-non speech detection: robust techniques have to be designed since a number of noisy speech recognition errors originate from a wrong determination of utterance boundaries,
- . reference patterns modeling,
- . recognition algorithms and distance measures.

Of course, the different methods are not exclusive, and can be combined in order to obtain satisfactory performances.

3. ACOUSTICAL PROCESSING AND PARAMETRIZATION

The overall performances of a speech recognizer are highly dependent upon the quality and robustness of the acoustic and sometimes phonetic features extracted from the speech wave as a front end of the recognition algorithms. Therefore a major effort has been devoted to this problem.

3.1 Speech Enhancement

As a first step in the recognition process, speech enhancement techniques tend to suppress the noise which corrupts the speech signal. Besides the methods using several microphones, many different types of speech enhancement systems using a single microphone have been proposed and tested [4]. All these systems are based on techniques which intend to recover the clean speech signal by enhancing the signal-to-noise ratio. The performances depend upon the type of noise which corrupts speech and the information they require about noise. It should be noted at this point that the increase of SNR will improve the quality of the speech signal but not always its intelligibility. Therefore, as far as speech recognition is concerned, a trade-off has to be found between SNR improvement and recognition accuracy.

Four main types of methods are used for speech enhancement:

- *noise subtraction*: this is a very common method which assumes that noise and speech are uncorrelated and additive. In the spectral subtraction approach, the power spectrum of cleaned speech is obtained by subtracting the noise power spectrum from the spectrum of noisy speech [22]. The method assumes that the noise varies slowly so that the noise estimation obtained during a pause can be used for suppression. Obtaining a good estimate of the noise spectrum is obviously the most difficult part of the method.

Non-linear spectral subtraction has been proposed in order to overcome some limitations of basic subtraction. It basically consists in overestimating the noise spectrum, either in a uniform way [5], or else based on the perceptual evidence that the ear is more sensitive to the peaks of a power spectrum than to the valleys and that noise in the frequency regions of the valleys contributes the most to perceptual distortions [6];

- *filtering*: traditional adaptive filtering techniques like Wiener or Kalman filtering have been used for speech enhancement, but more for speech transmission than for recognition purposes [7]. The Wiener filter yields an optimal solution to the adaptive filtering problem in the sense of least mean square error. It necessitates the estimation of some parameters of the noise. Unless the noise is stationary and perfectly known that must usually be done iteratively. A recursive optimal estimation can be obtained with a Kalman filter [8]. Kalman filtering can also be used with a colored noise assumption instead of white noise. Experimental results show an improvement in intelligibility even though no automatic recognition was carried out;

- *use of Markov models*: hidden Markov models (HMM) decomposition is a method which makes it possible to separate speech from additive noise [9]. It is assumed that speech and noise are both modeled by separate HMMs and that noisy speech corresponds to a composite model combining these two models.

The recognition of a noisy utterance can therefore be carried out by extending the classical Viterbi decoding algorithm to a search in the state space defined by the two models. This method is rather computationally demanding, but it has been demonstrated to perform satisfactorily even in bad SNR conditions;

- *space mapping*: speech enhancement can be viewed as the process of transforming noisy speech into clean speech by some kind of mapping. For instance, spectral mapping has been implemented by a set of rules obtained by vector quantization techniques [10]. It is also possible to implement arbitrarily complex space transformations thanks to connectionist neural networks. Even simple models such as multi-layer

perceptrons have been trained on learning samples to realize a mapping of noisy signals to noise-free speech which has been tested with success in an auditory preference test with human listeners [11].

3.2 Auditory models

Since the human auditory performs remarkably well in noisy conditions (even if speech variation is important) it seems interesting to rely on models of human hearing for designing speech recognizer front-ends [12].

Several experiments have been carried out in the use of auditory models in speech recognition. Some of them consist in incorporating physiological data in existing analysis methods (like PLP, SMC, etc.). Other tend to implement a complete auditory model. Most of those models have been mainly concerned so far with periphery audition, like those of Lyon [13], Seneff [14] or Ghitza [15]. Lateral inhibition has also been used for speech enhancement [16]. Lateral inhibition is a common phenomenon in sensory reception of biological systems, the function of which is to sharpen spatial input patterns.

Some recent advances have also been made toward the modeling of the central auditory system. The central auditory system further processes the spatio-temporal firing pattern on the auditory nerve at the output of the periphery auditory model [17]. The model includes both spatial and temporal processing of signals in the central auditory system and an efferent mechanism modeling the effect induced on the cochlea and auditory nerves by the brain. Experimental results on spoken letters show a good robustness to low SNRs in the recognition of plosive consonants.

3.3 Improvement of Speech Analysis Methods

The following methods have produced substantial improvements in recognition accuracy:

- *non-parametric representations*: the use of dynamic and acceleration (second derivative) MFCC and energy features makes it possible to enhance the recognition performances for noisy and/or Lombard speech.

Data analysis techniques have been used in the IMELDA system [18] in order to obtain a robust representation of noisy speech (but also of clean speech, as further experiments have demonstrated). IMELDA carries out a linear transformation based on discriminant analysis with minimization of within-class differences and maximization of inter-class differences. The result is a low dimensionality representation space in which recognition algorithms perform well. The combination of IMELDA with a non linear spectral subtraction method has also been shown as giving improved performances in the recognition of speech in a car [19];

- *parametric representations*: several improvements of LPC analysis in the presence of noise have been tested with some success. Some of these methods are based upon an alternative solution to the speech deconvolution problem. The classical solution consists in identifying the impulse response of the vocal tract by AR or ARMA modeling. The other solution is to map the time signal space into a linear structure by using a homomorphic transformation, that corresponds to a filtering in the cepstral domain. The logarithm homomorphic deconvolution can be further generalized to a spectral root deconvolution [20].

The short-time modified coherence (SMC) representation of speech [21] computes coefficients which are LPC coefficients obtained by taking into account the characteristics of the autocorrelation domain. It can be best described as the estimate of an AR model in the autocorrelation domain, and it exploits the fact that an autocorrelation sequence is more resistant to noise than the original signal. SMC analysis made it possible to increase the SNR by 10 to 12 when it was initially 0 to 20 dB. It is worth noticing that the amount of improvement is essentially data dependent.

PLP (Perceptual Linear Prediction) [22] differs from LPC by the use of three concepts derived from the study of human hearing, i.e. critical band spectral resolution, pre-emphasis with equal-loudness curve, and spectral compression according to a intensity-loudness power law, as illustrated in figure 10. A comparative study has shown the superiority of PLP over LPC for noisy speech recognition [23], especially in conjunction with a lifted cepstral distance.

The RASTA (Relative Spectral) approach can be considered as another improvement of basic LPC [24]. The method consists in operating in the log power spectral domain. That makes it possible to remove, or at least efficiently reduce, by filtering techniques slow-varying communication noise, which is additive in the log domain.

3.4 Noise Masking

In the presence of noise, certain low energy regions of the speech frequency spectrum will be more heavily corrupted by noise than others, and thus cause distortions in the computation of a distance between spectra during the recognition phase [25]. A solution was proposed by D. Klatt, based on a speech masking method in which only those frequency regions of the spectrum with energy level higher than the masking level are used in the distance computation.

Klatt's initial method was further improved in order to overcome its limitations, especially for the comparison of two speech patterns with very different noise levels [26]. Noise masking has been demonstrated as giving particularly robust recognition performances down to a very low 3dB SNR. The method was also extended for incorporation in a stochastic recognition framework. It should be noticed that the masking relation between speech and noise can be considered as a generalization of the HMM decomposition method presented above in which the relation is simply an addition.

4. STATISTICAL MODELING

Stochastic methods, particularly hidden Markov models, have been used to model some aspects of noise and variability. We have already mentioned the technique of noise compensation by HMM decomposition for speech enhancement. The basic idea is to take into account the presence of noise in speech by separately modeling an independent noise source as well as a speech source, assuming some relationship between the two sources.

Besides the already mentioned HMM decomposition technique, a possibility consists in exploiting the capability of HMM to segment a speech utterance into quasi-stationary segments corresponding to the states of the HMM. That has been used for iteratively cleaning the speech wave [27]. This approach

capitalizes on statistical modeling of both clean speech and noise. From the estimated statistics of the two processes, a maximum a posteriori (MAP) estimation was carried out and implemented using the Estimate-Maximize (EM) algorithm. Experimental results are especially good for SNRs above 10 dB.

The EM algorithm was designed for maximizing complicated likelihoods like the ones encountered in statistical speech processing, and it is widely used in noisy speech processing [28]. As an alternative to the MAP algorithm a minimum mean square error (MMSE) algorithm has also been proposed in order to obtain estimators of the clean signal [29]. This algorithm has been experimentally demonstrated to perform better than MAP for noisy speech enhancement [30]. However, in case of complex stochastic models for speech and noise the MAP estimator remains necessary.

The noise robustness of HMMs has also been studied as a function of the training mode. The use of a discriminant minimum error classification (MEC) optimization criterion [31] instead of the conventional maximum likelihood (ML) approach increases the robustness of HMMs in all types of environments. In MEC optimization, it is necessary to define a loss function that will be minimized by a non-linear multidimensional minimization technique, based either on a simple descent minimization, or a on some more sophisticated algorithm.

5. RECOGNITION TECHNIQUES

Rather than trying only to remove noise from speech, it is also possible to develop robust recognition models and techniques capable of coping with noise.

Since the best performances for a system are obtained when the training and testing conditions are similar, a first idea that has been investigated consists in training a recognizer with a multi-style training procedure. The training data are made up of speech signals produced in different talking styles and noise conditions, thus resulting in a multi-reference recognition system. Although this solution was demonstrated as feasible, it is not easy to implement in practice, and does not really satisfy the problem of robustness in noisy speech recognition.

5.1 Robust Distance Measures

The definition of an appropriate speech representation must be complemented by an adapted distance measure so that the recognition algorithm can take full advantage of the robustness of the representation. Amongst the various distance measures used in pattern recognition, the following ones have been specifically adapted to the problem of noisy speech recognition:

- weighted spectral measures [32]: this distance has been shown robust against white noise by weighting spectral peak regions that are less affected by noise. A similar weighted measure including a noise-adaptive mechanism, and similar to a frequency-weighted Itakura distance, has also been proposed [33];

- cepstral distances have given good results in speech recognition. However, their performances degrade in case of varying environments or speakers. A solution to this problem consists in defining weighted (liftered) cepstral distances which are usually represented as:

$$d = \sum_n w(n) [C_n(f) - C_n(g)]^2$$

where C_n are the cepstral coefficients of the two spectra f and g to be compared, and $w(n)$ represents the weighting factors [34].

Root-Power Sums (RPS) are computed from the cepstral distance by weighting the cepstral coefficients by their indices, i.e. $w(n) = n$. They are equivalent to a distance between spectral slopes. RPS have been used with PLP analysis for noisy speech recognition [35]. Other types of weighted cepstral distances with different weighting factors have also been considered ;

- cepstral projection has also been proven to be effective in coping with mismatched noise conditions [36]. The principle is based on the observation that additive white noise causes a shrinkage of the cepstral vector norm as a function of the noise level, and that the vector orientation is less affected. This shrinkage obviously affects a traditional distance calculation. It has been suggested to use a projection operation to formulate a new family of distortion measures. A good choice appears to be:

$$d(C_r, C_t) = |C_t|^a (1 - \cos \beta)$$

where C_r and C_t are respectively the reference and test cepstral vector.

$a = 0, 1$ or 2 is a compensation factor in order to take into account the reduction of cepstral energy.

$$\cos \beta = (C_r^T \cdot C_t) / (|C_r| |C_t|).$$

In a speaker dependent isolated word recognition experiment, a projection measure was shown to be the best distortion measure.

5.2 Noise Adaptation Methods

Noise adaptation methods consist in adapting the reference in order to make its parameters as close as possible to the unknown pattern. Such an adaptation process can be carried out continuously during the recognition, as in [37] where the original reference patterns are updated frame by frame after every successful dynamic matching.

The space matching techniques presented above can be used to some extent for adaptation. The main techniques of spectral adaptation proposed so far were initially developed for speaker adaptation purposes, and they have been extended to the problem of adaptation to a new environment. These techniques fall into two categories:

- codebook mapping: in the histogram method [38], the adaptation between the codebooks of two speakers (or two different environments) is carried out by a mapped codebook which is determined by dynamic time warping alignment from a small sample of speech from the new speaker. An application of this method to noisy speech recognition is reported in [39].

The technique of hierarchical spectral clustering has also been applied to the spectral normalization of noisy speech [40]. This is a non-supervised technique which builds up the adaptation codebook iteratively;

- spectral transformations of parameter spaces: such transformations can be either linear or non-linear. In both cases, the transformation is determined by multiple linear regression. Non-linear transformations can also be implemented by neural networks [41].

5.3 Noise Contamination of Reference Patterns

A technique for avoiding the mismatch between training and testing conditions consists in adding estimated noise to the reference patterns instead of trying to clean up the observed speech signal. This technique is quite easy to implement and has sometimes given better results than those obtained with more sophisticated speech enhancement techniques. Several methods have been proposed, including the noise contamination of phoneme templates [42] and noise immunity learning. However those methods obviously do not cope with Lombard effect, and they may also be sensitive to variations of noise level.

6. CONCLUSION

The problem of speech recognition in noisy conditions is very difficult and diverse. It constitutes a major bottleneck for the practical use of speech recognizers in real conditions since no completely satisfactory solution has been proposed so far. This paper has briefly reviewed the major categories of methods, belonging to the two approaches of signal acquisition and preprocessing, and adapted recognition algorithms. The two approaches are not exclusive, and they can be combined in a clever way for obtaining better performances.

Despite significant results, the problem is not yet completely solved. An important effort has still to be done in order to comparatively evaluate the different existing methods, and to improve them. That necessitates an important, and on-going, effort of noisy speech data bases recording and assessment method design.

It can be expected that substantial improvements in the robustness of speech recognition systems will then be obtained through present worldwide effort on this topic.

References

1. Lombard, E.: Le signe de l'élévation de la voix, *Ann. Maladies Oreille, Larynx, Nez, Pharynx*, 37, 101-119, 1911.
2. Dautrich, B.A., Rabiner, L.R. and Martin, T.B.: On the Effects of Varying Filter Bank Parameters on Isolated Word Recognition, *IEEE Trans. ASSP*, 31(4), 793-806, 1983.
3. Junqua, J.C.: The Variability of Speech Produced in Noise, *Proc. ESCA Workshop on Speech Processing in Adverse Conditions*, 43-52, Cannes-Mandelieu, 1992.
4. Van Compernelle, D.: DSP Techniques for Speech Enhancement, *Proc. ESCA Workshop on Speech Processing in Adverse Conditions*, 21-30, Cannes-Mandelieu, 1992.
5. Boll, S.F.: Suppression of Acoustic Noise in Speech Using Spectral Subtraction, *IEEE Trans. ASSP*, vol. 27, 113-120, 1979.
6. Lockwood, P., Boudy, J.: Experiments with a Non-linear Spectral Subtractor (NSS), Hidden Markov Models and the Projection, for Robust Speech Recognition in Cars, *Proc. EUROSPEECH-91*, 79-82, Genova, 1991.
7. Widrow, B. et al.: Adaptive Noise Cancelling: Principles and Applications, *Proc. IEEE*, 63(12), 1692-1716, 1975.

8. Koo, B., Gibson, J.D., Gray, S.D.: Filtering of Colored Noise for Speech Enhancement and Coding, Proc. ICASSP-89, 349-352, Glasgow, 1989.
9. Varga, A.P., Moore, R.K.: Hidden Markov Model Decomposition of Speech and Noise, Proc. ICASSP-90, 845-848, Albuquerque, 1990.
10. Juang, B.H., Rabiner, L.R.: Signal Restoration by Spectral Mapping, Proc. ICASSP-87, 2368-2371, Dallas, 1987.
11. Tamura, S., Waibel, A.: Noise Reduction using Connectionist Models, Proc. ICASSP-88, 553-556, New York, 1988.
12. Theme issue: Representation of Speech in the Auditory Periphery, J. Phonetics, vol. 16, n° 1, 1-149, 1988.
13. Lyon, R.F.: A Computational Model of Filtering, Detection, and Compression in the Cochlea, Proc. ICASSP-82, 1282-1285, Paris, 1982.
14. Seneff, S.: Pitch and Spectral Estimation of Speech Based on Auditory Synchrony Model, Proc. ICASSP-84, 36.2.1-36.2.4, San Diego, 1984.
15. Ghitza, O.: Robustness Against Noise: the Role of Timing-Synchrony Measurement, Proc. ICASSP-87, 2372-2375, Dallas, 1987.
16. Cheng, Y.M., O'Shaughnessy, D.: Speech Enhancement Based Conceptually on Auditory Evidence, Proc. ICASSP-91, 961-964, Toronto, 1991.
17. Gao, Y., Huang, T., Haton, J.P.: Central Auditory Model for Spectral Processing, Proc. ICASSP-93, 704-707, Minneapolis, 1993.
18. Hunt, M.J., Lefebvre, C.: A Comparison of Several Acoustic Representations for Speech with Degraded and Un-degraded Speech, Proc. ICASSP-89, Glasgow, 1989.
19. Sorensen, H.B.D., Hartmann, U.: Robust Speaker-independent Speech Recognition using Non-linear Spectral Subtraction based IMELDA, Proc. EUROSPEECH-93, 235-238, Berlin, 1993.
20. Lim, J.S.: Spectral Root Homomorphic Deconvolution System, IEEE Trans. ASSP, 27(3), 223-233, 1979.
21. Mansour, D., Juang, B.H.: The Short-time Modified Coherence Representation and its Application for Noisy Speech Recognition, Proc. ICASSP-88, 525-528, New York, 1988.
22. Hermansky, H.: An Efficient Speaker-independent Automatic Speech Recognition by Simulation of Some Properties of Human Auditory Perception, Proc. ICASSP-87, 1159-1162, Dallas, 1987.
23. Junqua, J.C., Wakita, H.: A Comparative Study of Cepstral Lifters and Distance Measures for All Pole Models of Speech in Noise, Proc. ICASSP-89, 476-479 Glasgow, 1989.
24. Hermansky, H., Morgan, N., Bayya, A., Kohn, P.: Compensation for the Effect of Communication Channel in Auditory-like Analysis of Speech (RASTA-PLP), Proc. EUROSPEECH-91, 1367-1370, Genova, 1991.
25. Klatt, D.H.: A Digital Filter Bank for Spectral Matching, Proc. ICASSP-76, 573-576, Philadelphia, 1976.
26. Holmes, J.N., Sedgwick, N.C.: Noise Compensation for Speech Recognition Using Probabilistic Models, Proc. ICASSP-86, 741-744, Tokyo, 1986.
27. Ephraim, Y., Malah, D., Juang, B.H.: On the Application of Hidden Markov Models for Enhancing Noisy Speech, Proc. ICASSP-88, 533-536, New York, 1988.

28. Wu, C.S., Nguyen, V.V., Sabrin, H., Kushner, W., Damoulakis, J.: Fast Self-Adapting Broadband Noise Removal in the Cepstral Domain, Proc. ICASSP-91, 957-960, Toronto, 1991.
29. Porter, J.E., Boll, S.F.: Optimal Estimators for Spectral Restoration of Noisy Speech, Proc. ICASSP-84, 18A.2.1-18A.2.4, San Diego, 1984.
30. Ephraim, Y.: A Bayesian Estimation Approach for Speech Enhancement Using Hidden Markov Models, IEEE Trans. ASSP, 40(4), 725-735, April 1992.
31. Chou, W., Juang, B.H., Lee, C.H.: Segmental GPD Training of HMM Based Speech Recognizer, Proc. ICASSP-92, 473-476, San Francisco, 1992.
32. Matsumoto, H. and Imai, H. : Comparative Study of Various Spectrum Matching Measures on Noise Robustness, Proc. ICASSP-86, 769-772, Tokyo, 1986.
33. Soong, F.K., Sondhi, M.M.: A Frequency-weighted Itakura Spectral Distortion Measure and its Application to Speech Recognition in Noise, IEEE Trans. ASSP, 36(1), 41-48, 1988.
34. Hanson, B. and Wakita, H. : Spectral Slope Distance Measures with Linear Prediction Analysis for Word Recognition in Noise, IEEE Trans. ASSP, 35, 968-973, 1987.
35. Hansen, J.H., Clements, M.A.: Spectral Slope Based Distortion Measures for All-pole Models of Speech, Proc. ICASSP-86, 757-760, Tokyo, 1986.
36. Mansour D., Juang, B.H.: A Family of Distortion Measures Based upon Projection Operation for Robust Speech Recognition, IEEE Trans. ASSP, 37(1), 1659-1671, November 1989.
37. Dvorak, S., Hörmann, T.: High-Performance Speech Recognition in Noise by Continuously Updated Reference Templates, Proc. EUROSPEECH-91, 1375-1378, Genova, 1991.
38. Shikano, K., Lee, K.F., Reddy, D.R.: Speaker Adaptation through Vector Quantization, Proc. ICASSP-86, 2643-2646, Tokyo, 1986.
39. Ohkura, K., Sugiyama, M.: Speech Recognition in a Noisy Environment Using a Noise Reduction Neural network and a Codebook Mapping Technique, Proc. ICASSP-91, 929-932, Toronto, 1991.
40. Cung, H.M., Normandin, Y.: Noise Adaptation Algorithms for Robust Speech Recognition, Speech Communication, 12(3), 267-276, 1993
41. Barbier, L., Chollet, G.: Robust Speech Parameters Extraction for Word Recognition in Noise Using Neural Networks, Proc. ICASSP-91, 145-148, Toronto, 1991.
42. Morii, S., Morii, T., Hoshimi, M., Hiraoka, S., Watanabe, T., Niyada, K.: Noise Robustness in Speaker Independent Speech Recognition, Proc. ICSLP-90, 1145-1148, Kobe, 1990.