International Journal of Smart Home
Vol. 10, No. 9 (2016), pp. 197-206
http://dx.doi.org/10.14257/ijsh.2016.10.9.18

Speech Recognition Algorithm in Complex Noisy Environments Based on Multi-Space Compensation★

Yuanchang Zhong[1*], Wenjin Xie[1], Donghai Xiao[2] and Zunzhao Wang[1]

[1]*College of Automation, Chongqing University, Chongqing, 400044, China*
[2] *Mianyang lingtong Telecom Equipment Co. Mianyang, 621000, China*

Abstract

Speech recognition rate drops significantly when interfered by noise in complex environment. In order to improve the accuracy and the robustness of the speech recognition in adverse acoustical environments, this paper reviewed the main problems of noise robust speech recognition, proposed a multi-space compensation algorithm which from signal-space, feature-space and model-space based on wiener filter, histogram equalization and vector Taylor series. Theory analyses and experiment results show that the proposed method can overcome the defect of the sharp descent of speech recognition rate of existing speech recognition algorithm interfered by environmental noise and improve the accuracy and the robustness in adverse acoustical environments. The algorithm provides the theoretical support for the speech recognition in airport, station, wharf and other complex noise environment.

Keywords*: Robust speech recognition, Noise suppression, Feature compensation, Model compensation*

1. Introduction

There exist various kinds of noise in the airport, station, wharf and other complex noise environment, resulting in the mismatch of training and application, and then causing extreme decline in the recognition performance of the system. Therefore, how to improve the robustness of the system has become the key problem in practical application of speech recognition system.

Research shows that, most speech recognition system's performance will greatly decrease greatly decrease when applied in practical noise environment, due to the mismatch of training and testing environment caused by the environmental noise [1]. The research purpose of the noise robust speech recognition algorithm is to eliminate or to reduce this mismatch, and to make the recognition system performance under noise environment as close as possible to the performance in quiet environment.

We can analyze the effect of mismatch of training and testing environment on the performance of the speech recognition system from signal space, feature space and model space. The robust speech recognition algorithm from these 3 levels is used to eliminate the mismatch impact through the method of speech compensation. At present, each space robust speech recognition algorithm has been very mature. In signal space, the Wiener filtering method of (Wiener filter) is a traditional-classical one, and the European Telecommunications Standards Institute released the noise robust algorithm for distributed speech recognition based on it in October, 2002. Michael Tinston and Yariv Ephraim proposed the multistage Wiener filter for speech enhancement in signal

★Project supported by the Science and Technology Innovation Foundation of the National Department of Science and Technology, China (No. 10C26215113031) and the Chongqing scientific and technological project (No.CSTC2012gg-yyjs40010).

[*] Corresponding author: *Email address:* zyc@cqu.edu.cn (Yuanchang Zhong).

ISSN: 1975-4094 IJSH

space to eliminate the effects of additive noise on the speech, and achieved good experimental results [2]. In feature space, in order to reduce the mismatch of training and testing data, the feature warping algorithm is commonly used. Histogram equalization (HEQ), proposed by De La Torre Ángel and others (Reference[14]), is a cumulative histogram normalization algorithm using feature parameters, has transformed probability density distribution of the noisy speech into that of the clean speech (generally mean is 0 and variance is 1), and has achieved a good effect. In model space, Pedro J Moreno first proposed a vector Taylor series (VTS) algorithm. Then, the application of VTS in the model space of the hidden Markov model (HMM) parameter adaptive adjustment algorithm is improved by Alex Acero, Li Deng and other scientists, achieving excellent results.

In order to further improve the accuracy and robustness of speech recognition under noise environment, and to solve the problem of the descent of speech recognition rate under noise interference in complex environment, we put forward a speech recognition algorithm based on multi-space compensation. The algorithm, integrated speech enhancement, noisy speech regular and adaptive parameter adjustment algorithm, optimizes from the signal space, feature space and model space.

2. Ambient Noise Models

Research shows that, the actual environmental noise is the real reason of mismatch between training and testing. Among them, the additive noise and channel influence are the two most common factors. Therefore, we pay more attention to the additive noise and convolution noise.

Assume: the clean speech is $x[m]$, convolution noise is $h[m]$, additive noise is $n[m]$, noisy speech is $y[m]$. Then, the noise environment model of speech recognition can be established, as shown in Figure 1 [3-6].

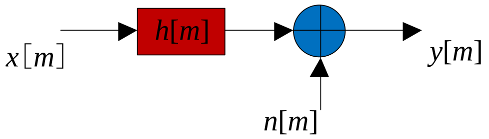

Figure 1. The Noise Environment Model

From Figure 1, the time domain model [5] of noise environment can be build:

$$y[m]=x[m]*h[m]+n[m] \tag{1}$$

And then get the frequency domain model from formula (1):

$$Y(w)=X(w)\left|H(w)\right|^2+N(w) \tag{2}$$

In formula (2), $X(w)$ and $Y(w)$ respectively denote the power spectrum of clean speech and noisy speech, $H(w)$ is the the frequency response of channel distortion filter, and $N(w)$ is the power spectrum of additive noise.

Now, we analyze the problem of mismatch between training and testing environment caused by noise from signal, feature, and model space.

Assume: the speech data in training environment is S; feature extraction from speech data in training environment is X; and speech model obtained from training data is Λx; T, Y and Λy respectively represent speech, characteristics and speech model in training environment. When the training and testing environment mismatch, noise makes T, Y and Λy distorted, and then $D_1(\cdot)$, $D_2(\cdot)$ and $D_3(\cdot)$ respectively represent the distortion function.

Thus, we can build the mismatch model of training and testing, as Figure 2 shows.

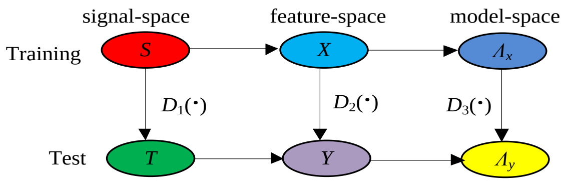

Figure 2. The Training and Testing Mismatch Model

Now we will discuss how to eliminate the influence of different training and testing environment according to Figure 2 from the three levels of signal space, feature space and model space.

3. Speech Recognition Algorithm based on Multi Space Compensation

In ambient noise model shown in Figure 1, first, Wiener filtering is used to restrain additive noise in signal space firstly, then, HEQ acts on each section of cepstrum domain feature vector in feature space, and finally, VTS is used in model space to adjust HMM model's parameters.

3.1. Signal space

In formula (1), supposing that the effect of convolution noise caused by the channel is neglected and only additive noise is considered, we need a linear filter $h[m]$ for recovering the clean speech from noisy speech interfered by additive noise, to make the signal operated by filter $\hat{x}[m] = y[m] * h[m]$ reach the smallest expectation value of $(\hat{x}[m] - x[m])^2$.

On the premise of that both $x[m]$ and $n[m]$ are irrelevant stationary signal, we can use suppression filter to represent the frequency domain:

$$H_s(w) = \frac{X(w)}{X(w) + N(w)} \tag{3}$$

Formula (3) is the Wiener filter [7,8]. When signal $x[m]$ and $n[m]$ satisfy the above assumptions, the Wiener filter can achieve suppression of noise, and will not cause great target estimation distortion and background remain noise [9-11]. The desired power spectrum $X(w)$ and $N(w)$ can be respectively obtained from the time series of $x[m]$ and $n[m]$ through multi-frame mean. However, in practice, the speech signal and background noise are non-stationary, that is, the power spectrum changes with time, which can be represented by $X(m, w)$ and $N(m, w)$. Therefore, we focus on the research of STFT (Short Time Fourier Transform, STFT) for each frame signal through different wiener filters.

The time-varying Wiener filter is:

$$H_s(pL, w) = \frac{\hat{X}(pL, w)}{\hat{X}(pL, w) + \hat{N}(pL, w)} \tag{4}$$

In formula (4), $\hat{X}(pL, w)$ is the estimation of the time-varying power spectrum $X(m,w)$ of $x[m]$, and $\hat{N}(pL, w)$ is that of power spectrum $N(m, w)$ of stationary noise.

3.2. Feature Space

HEQ is usually applied to digital image processing [12,13], aiming to provide a transformation x=F(y) to transform original variable probability density function(PDF) $p_y(y)$ into the reference probability density function $p_{ref}(x)$ [14]. Under certain conditions, the relationship between the cumulative density function of x and y (cumulative density functions, CDF) respectively is:

$$C_Y(y) = C_{ref}(F(y)) \tag{5}$$

$$F(y) = C_{ref}^{-1}(C_Y(y)) \tag{6}$$

In practice, we can only obtain limited data. Therefore, we can use the cumulative histogram to replace the CDF [15].

In this paper, we use HEQ in cepstrum domain. Each cepstral coefficients are respectively independent equilibrium as a reference probability density function. The specific process is as follows: each coefficient was estimated as a cumulative histogram; unify each sentence by considering 100 intervals between (μ_y-4σ_y) and (μ_y+4σ_y); μ_y and σ_y are the mean and standard deviation of the original value. This reference cumulative histogram can be obtained from normal distribution with zero mean and unit variance. In formula (6), we can find each interval center point, which can be used as the two recent columns linear interpolation points to compensate parameters.

3.3. Model Space

In model space, we choose first-order VTS (Vector Taylor Series). First of all, on the premise of knowing the additive noise and convolution noise, we use first-order VTS to get the mean value and variance of HMM in the MFCC cestrum domain. Then, we can re-estimate the influence of additive noise and channel on the actual speech model by using EM algorithm. And the accurate band parameters in noisy speech model are obtained.

The specific algorithm is: transform formula (2) and make it into the log-filter-bank domain, and then the equation multiplied by the non-square discrete cosine transform (DTC) matrix. The effective non-linear distortion model is obtained:

$$y = x + h + C\ln(1 + e^{C^{-1}(n-x-h)}) \tag{7}$$

In formula (7), C^{-1} is the inverse transformation of DCT matrix. y, x, n and h are the MFCC cestrum domain vector value of the noisy speech, clean speech, additive noise, and the convolution noise respectively.

We can get the following formula by using VTS [16,17]:

$$\begin{aligned}\mu_y &= \mu_x + \mu_h + C\log(1 + \exp\frac{\mu_n - \mu_x - \mu_h}{C}) \\ &= \mu_x + \mu_h + g(\mu_x, \mu_h, \mu_n)\end{aligned} \tag{8}$$

$$g(\mu_x, \mu_h, \mu_n) = C\log(1 + \exp\frac{\mu_n - \mu_x - \mu_h}{C}) \tag{9}$$

where μ_y、 μ_x、 μ_h and μ_n are the mean vector of y, x, h and n respectively.

After differential processing of formula (8), we can get:

$$\frac{\partial \mu_y}{\partial \mu_h} = C \cdot diag\left(\frac{1}{1+\exp\frac{\mu_n-\mu_x-\mu_h}{C}} \right) \cdot \frac{1}{C} = G \tag{10}$$

$$\frac{\partial \mu_y}{\partial \mu_n} = I - G \tag{11}$$

In formula (10), *diag*(.) represents the diagonal matrix. On the premise of knowing μ_h and μ_n , *G*(.) depends on the mean vector μ_x. In particular, for the *k*-th Gauss based on state *j*, the corresponding element of matrix *G*(.) is:

$$G(j,k) = C \cdot diag\left(\frac{1}{1+\exp\frac{\mu_n-\mu_x-\mu_h}{C}} \right) \cdot \frac{1}{\mathrm{C}} \tag{12}$$

Hence, we can get the relation between noisy speech adaptive HMM and the *k-th* Gauss mean vector in state *j* of clean speech by using the first-order VTS:

$$\begin{aligned}\mu_{y,jk} &= \mu_{x,jk} + \mu_h + g(\mu_{x,jk}, \mu_h, \mu_n) \\ &\approx \mu_{x,jk} + \mu_{h,0} + g(\mu_{x,jk}, \mu_{h,0}, \mu_{n,0}) + G(j,k)(\mu_h - \mu_{h,0}) + (I - G(j,k))(\mu_n - \mu_{n,0})\end{aligned} \tag{13}$$

where $\mu_{n,0}$ and $\mu_{h,0}$ respectively represent the VTS expansion point of μ_n and μ_h.

The formula (13) can only be applied to the static part of the MFCC vector. We can get $\Sigma_{y,jk}$ of the adaptive HMM from $\Sigma_{x,jk}$ and Σ_n:

$$\Sigma_{y,jk} \approx G(j,k)\Sigma_{x,jk} G(j,k)^T + (I - G(j,k))\Sigma_n (I - G(j,k))^T \tag{14}$$

In formula (14), $\Sigma_{x,jk}$ is the covariance matrix of HMM, Σ_n is the covariance matrix of noise. Here, we regard the channel as a fixed and known voice so that we do not consider the channel difference.

For the parts of *delta* and *delta/delta* of the MFCC vector, we can get the adaptive formula of the mean vector and covariance matrix:

$$\mu_{\Delta y,jk} \approx G(j,k)\mu_{\Delta x,jk} \tag{15}$$

$$\mu_{\Delta^2 y,jk} \approx G(j,k)\mu_{\Delta^2 x,jk} \tag{16}$$

$$\Sigma_{\Delta y,jk} \approx G(j,k)\Sigma_{\Delta x,jk} G(j,k)^T + (I - G(j,k))\Sigma_{\Delta n} (I - G(j,k))^T \tag{17}$$

$$\Sigma_{\Delta^2 y,jk} \approx G(j,k)\Sigma_{\Delta^2 x,jk} G(j,k)^T + (I - G(j,k))\Sigma_{\Delta^2 n} (I - G(j,k))^T \tag{18}$$

Considering we have already known the additive noise and channel (convolution) noise in advance, we use a first-order VTS to get the approximation of the mean and variance in the MFCC cepstrum domain. However, in practical situations, the effect of additive noise and channel noise is complex and changeable. Thus we propose that we should re-estimate the additive noise and channel (convolution) noise by using EM algorithm [18,19]. The specific process is as follows:

Ω_s represents the state sets; Ω_m represents the Gauss set of each state, θ_t and ε_t respectively represent the state value and Gauss value for the time frame at *t* time, λ is the mean parameter collection of new noise and channel, $\bar{\lambda}$ is the mean parameter collection of old noise and channel. The likelihood function of a speech can be expressed as:

$$F(\lambda \mid \bar{\lambda}) = \sum_{t} \sum_{j \in \Omega_s} \sum_{k \in \Omega_m} p(\theta_t = j, \varepsilon_t = k \mid Y, \bar{\lambda}) \cdot \log p(y_t \mid \theta_t = j, \varepsilon_t = k, \lambda) \qquad (19)$$

In formula (19), $p(y_t|\theta_t=j,\ \varepsilon_t=k,\ \lambda) \sim N(y_t;\ \mu_{y,jk},\ \Sigma_{y,jk})$ is the Gauss distribution whose mean vector is $\mu_{y,jk}$ and covariance is $\Sigma_{y,jk}$.

In order to simplify the formula, we use $\gamma_t(j,k)$ to represent the posterior probability of the *k-th* Gauss under state *j* of HMM:

$$\gamma_t(j,k) = p(\theta_t = j, \varepsilon_t = k \mid Y, \bar{\lambda}) \qquad (20)$$

In order to get the maximum value of the likelihood function, we need to obtain the differential *F* under μ_n and μ_h, let this differential expression equal to zero, as follows:

$$\sum_{t} \sum_{j \in \Omega_s} \sum_{k \in \Omega_m} \gamma_t(j,k)(I - G(j,k))^T \Sigma_{y,jk}^{-1} [y_t - \mu_{y,jk}] = 0 \qquad (21)$$

$$\sum_{t} \sum_{j \in \Omega_s} \sum_{k \in \Omega_m} \gamma_t(j,k) G(j,k)^T \Sigma_{y,jk}^{-1} [y_t - \mu_{y,jk}] = 0 \qquad (22)$$

We can get the mean vector for posterior probability of additive noise μ_h by taking the VTS approximation formula (13) into (21):

$$\mu_n = \mu_{n,0} + \left\{ \begin{array}{l} \sum_{t} \sum_{j \in \Omega_s} \sum_{k \in \Omega_m} \gamma_t(j,k)(I - G(j,k))^T \\ \cdot \Sigma_{y,jk}^{-1} (I - G(j,k)) \end{array} \right\}^{-1} \cdot \left\{ \begin{array}{l} \sum_{t} \sum_{j \in \Omega_s} \sum_{k \in \Omega_m} \gamma_t(j,k)(I - G(j,k))^T \\ \cdot \Sigma_{y,jk}^{-1} \left[y_t - \mu_{x,jk} - \mu_{h,0} - g(\mu_{x,jk}, \mu_{h,0}, \mu_{n,0}) \right] \end{array} \right\} \qquad (23)$$

Similarly, take the equation (12) into (22), and make μ_h=$\mu_{n,0}$, then the channel mean vector can be estimated as:

$$\mu_h = \mu_{h,0} + \left\{ \sum_{t} \sum_{j \in \Omega_s} \sum_{k \in \Omega_m} \gamma_t(j,k) G(j,k)^T \Sigma_{y,jk}^{-1} G(j,k) \right\}^{-1} \times \left\{ \begin{array}{l} \sum_{t} \sum_{j \in \Omega_s} \sum_{k \in \Omega_m} \gamma_t(j,k) G(j,k)^T \\ \cdot \Sigma_{y,jk}^{-1} \left[y_t - \mu_{x,jk} - \mu_{h,0} - g(\mu_{x,jk}, \mu_{h,0}, \mu_{n,0}) \right] \end{array} \right\} \qquad (24)$$

Formula (23) and (24) constitute the iteration process of EM algorithm. We can getthe accurate value of μ_h and μ_n which shown in Figure 3 by using EM algorithm.

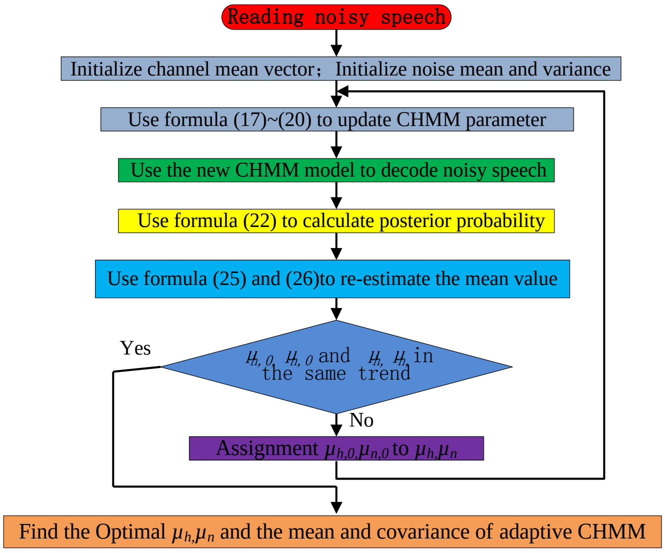
Reading noisy speech
Initialize channel mean vector; Initialize noise mean and variance
Use formula (17)~(20) to update CHMM parameter
Use the new CHMM model to decode noisy speech
Use formula (22) to calculate posterior probability
Use formula (25) and (26)to re-estimate the mean value
Yes
$\mu_{h,0}$, $\mu_{h,0}$ and μ_h, μ_n in the same trend
No
Assignment $\mu_{h,0}$,$\mu_{n,0}$ to μ_h,μ_n
Find the Optimal μ_h,μ_n and the mean and covariance of adaptive CHMM

Figure 3. The Model Reference Adaptive Flow Chart

4. Experimental Results and Analysis

The algorithm is based on AURORA II database, which is a subset database based on TI-DIGITS and is continuous digital speech data recorded in clean environment. 8 kinds of additive noise (including the MTR, gurgling sound, automobile, Exhibition halls, restaurants, streets, airports and train stations) and a channel interference (convolutional noise) can be added into the database according to the different SNR (such as 20dB, 15dB, 10dB, 5dB, 0dB, -5 dB).

The test has provided two training sets: (1) clean training set containing only clean speech data; (2) the noisy training set including not only the clean speech data but also the noise from subway, gurgling sound, automobile and Exhibition hall whose signal-to-noise ratio is from 20dB to 5dB. 3 test groups (A, B and C): group A contains the same noise and signal-to-noise ratio used in noisy training; group B contains the noisy data whose signal-to-noise ratio is same with group A, but the current noise is from restaurants, streets, airports and train stations; group C contains two kinds of noise of group A (subways and cars), and also includes a distortion of channel (convolutional noise). Two types of experiment are defined: clean training experiment (Clean)--use a clean training set to train acoustic models; noisy training experiment (Multi)--use noisy training set to train acoustic models. The test result is the mean value of each test group (A, B and C) on the condition of the five signal-to-noise ratios of 20dB~0dB. Both clean training experiment and noisy training experiment use HTK toolkit to train the HMM model [19,20].

We use level 23 MEL triangular filter bank, obtaining the voice signal between 64Hz~4 KHz by using convolution calculation of 100Hz to extract the basic parameter. Use DCT to transform these parameters until into cepstrum domain, and retain the cepstrum coefficient C1~C12 only. Finally, we obtain *delta* and *delta/delta* part of MFCC based on these coefficients are differential and two differential treatments.

Acoustic model assumes: each digital training of 0~9 represents a HMM mode, and "oh" is also trained as a HMM model. For the 11 digital HMM model, we suppose *N*=6 of each model, the structure is from left to right, and there is no jump between states. Each state corresponds to a 3 order mixture Gauss mixture density function. The covariance matrix of HMM is a diagonal matrix.

The recognition accuracy rate contrast of this algorithm and related algorithms is shown in table 1.

Table 1. The Contrast between Different Algorithms for the Accuracy

	Group A	Group B	Group C	average
Base-Clean	61.34	55.73	66.16	61.08
WF	73.71	71.35	70.62	71.89
WF+HEQ	82.08	81.61	81.73	81.81
WF+HEQ+VTS	90.98	89.79	90.15	90.31
Base Multi	91.06	90.47	90.78	90.77

Table 1 has listed out the result of each recognition test; the value in the table is the mean of AURORA II test array A, B and C in different signal-to-noise ratios(20dB~0dB). Among them, Base Clean represents a clean training experiment which does not use any speech compensation algorithm; WF represents a clean training experiment, which uses only the Wiener filter in the signal space; WF+HEQ represents a clean training experiment, which firstly uses Wiener filter in the signal space, then uses the SNR higher cepstrum domain histogram equalization (HEQ) treatment; WF+HEQ+VTS is the speech recognition algorithm based on spatial compensation combining the Wiener filtering, histogram equalization, adaptive feature space, VTS model space compensation under clean training experiment; Base Multi does not use any speech compensation algorithm under the noisy training experiment. Conclusion of this table: (1) the recognition results are improved in this order: Base Clean, WF, WF+HEQ, WF+HEQ+VTS, Base Multi; (2) the performance of WF+HEQ+VTS is very close to that of the noisy training test (Base Multi); (3) WF, WF+HEQ, WF+HEQ+VTS have slightly different influence on the test group A, Band C--WF has little compensation effect on group C; WF+HEQ and WF+HEQ+VTS greatly improve the accurate identification of group C; all the recognition accuracy rate of group A in WF, WF+HEQ, and WF+HEQ+VTS is higher than that of group B.

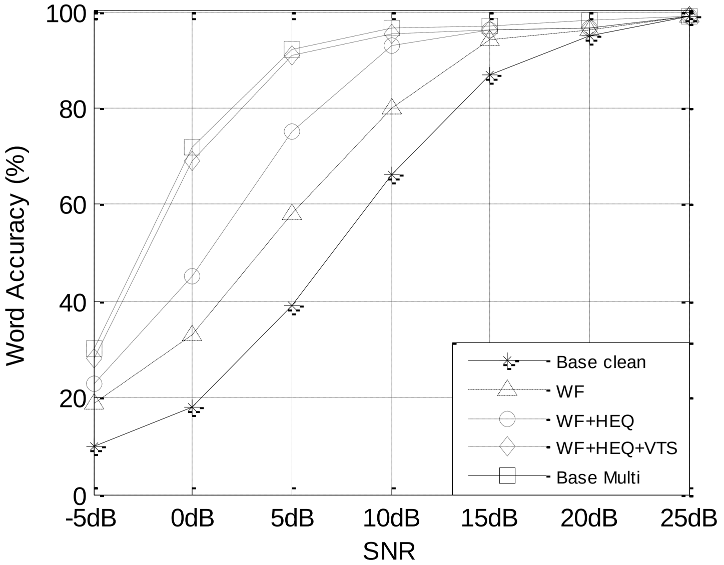

Figure 4. Different Signal to Noise Ratio of each Algorithm Recognition Results

The recognition accuracy rate of each algorithm in different SNR is shown in Figure 4. From Figure 4, we know that: (1) the recognition accuracy rate of the speech recognition algorithm based on multi-space compensation is very close to that of Base Multi in different SNR, and especially in the low SNR, the recognition performance of WF, and WF+HEQ is improved greatly. (2) The recognition accuracy rate of each noise robust recognition algorithm is low in the very low SNR (-5dB); while it is high in the high signal-to-noise ratio (above 15 dB).

The experimental results show that: the recognition accuracy rate of group A and B has been greatly improved after using the Wiener filter in the signal space, verifying that this algorithm can restrain on part of the additive noise, while the performance of group C hasn't been improved so much, the reason is that it includes channel interference (convolution noise), and Wiener filtering has little or no effect on the channel interference; the recognition accurate rate of each group becomes higher after using cepstrum domain histogram (HEQ) in high SNR cepstrum domain of feature space, because HED has suppressed the addictive noise, channel noise and nonlinear distortion comprehensively; the recognition correct rate of this algorithm in different signal-to-noise ratios is very close to that of the noisy training experiment, and especially in the low SNR, the recognition performance of WF and WF+HEQ will be improved obviously. But the accuracy rate of recognition is very low in the low SNR (less than -5dB) by using the noise robust recognition algorithm.

5. Conclusions

In view of the defect of the speech recognition algorithm based on the hidden Markov model (HMM), whose speech recognition rate decreases significantly in the noisy environment, we put forward the speech recognition algorithm based on the compensation of space from the signal space, feature space and model space, combining the environmental noise caused by the mismatches between the training and testing source.

The test results in AURORA II show that: the algorithm is effectively integrated with the advantages of Wiener filtering, histogram equalization, and vector Taylor series, and this algorithm has improved the defect of speech recognition rate of the speech recognition algorithm based on HMM dropping sharply in noisy environment, has achieved high recognition rate, and has significantly improved the accurate rate of the robustness speech recognition algorithm. This algorithm has very good application effects in the field of smart home and network appliances.

References

[1] T. Vikrant Singh, R. Richard C, "A family of discriminative manifold learning algorithms and their application to speech recognition", Speech and Language Processing, vol. 22, no. 1, **(2014)**, pp. 161-171.

[2] T. Michael, E. Yariv, "Speech Enhancement using the Multistage Wiener Filter", Proceedings of 43rd Annual Conference on Information Sciences and Systems, Baltimore, MD, United states, **(2009)**, pp. 55-60.

[3] G. Yifan, "Speech recognition in noisy environments: A survey", Speech Communication, vol. 16, no. 3, **(1995)**, pp. 261-291.

[4] A. Amrouche, A. Taleb-Ahmed, J.M. Rouvaen, and M.C.E Yagoub,. "Improvement of the speech recognition in noisy environments using a nonparametric regression", International Journal of Parallel, Emergent and Distributed Systems, vol. 24, no. 1, **(2009)**, pp. 49-67.

[5] H. Serajul, T. Roberto, Z. Anthony, "Perceptual features for automatic speech recognition in noisy environments", Speech Communication, vol. 51, no. 1, **(2009)**, pp. 58-75.

[6] K. Do Yeong, U. Chong Kwan, K. Nam Soo. "Speech recognition in noisy environments using first-order vector Taylor series", Speech Communication, vol. 24, no. 1, **(1998)**, pp. 39-49.

[7] J. Peter; K. Münevver. "Incorporating the voicing information into HMM-based automatic speech recognition in noisy environments", Speech Communication, vol. 51, no. 5, **(2009)**, pp. 438-451.

[8] A. Alejandro, S. Richard M. "Environmental robustness in automatic speech recognition", IEEE

International Conference on Acoustics, Speech and Signal Processing - Proceedings, Albuquerque, New Mexico, USA, **(1990)**, pp. 849-852.

[9] A.V. Savchenko, Ya.I. Khokhlova, "About neural-network algorithms application in viseme classification problem with face video in audiovisual speech recognition systems", Optical Memory and Neural Networks (Information Optics), vol. 23, no. 1, **(2014)**, pp. 34-42.

[10] B. Hynek, J. Hansen, H. L. "Unsupervised equalization of lombard effect for speech recognition in noisy adverse environments", Speech and Language Processing, vol. 18, no. 6, **(2010)**, pp. 1379-1393.

[11] Y. Xie, M. Liu, Z. Yao, B. Dai, "Improved two-stage Wiener filter for robust speaker identification", Proceedings - International Conference on Pattern Recognition, Hong Kong, China , **(2006)**, pp. 310-313.

[12] A. Riadh; Sbaa, Salim; Ghendir, Said; et a1. "Novel detection algorithm of speech activity and the impact of speech codecs on remote speaker recognition system", WSEAS Transactions on Signal Processing, vol. 10, no. 1, **(2014)**, pp. 309-319.

[13] K. Tetsuo, K. Masaharu, K. Masaki. "Histogram equalization for noise-robust speech recognition using discrete-mixture HMMs", Acoustical Science and Technology, vol. 29, no. 1, **(2008)**, pp. 66-73.

[14] De La Torre Ángel, Segura José C., Benítez Carmen, Peinado Antonio M., Rubio Antonio J.. "Non-linear transformations of the feature space for robust speech recognition", IEEE International Conference on Acoustics, Speech and Signal Processing-Proceedings, Orlando, FL, United states, **(2002)**, pp. 401-404.

[15] A. Anissa Imen, D. Mohamed, A. Abderrahman. "Robust Arabic speech recognition in noisy environments using prosodic features and formant", International Journal of Speech Technology, vol. 14, no. 4, **(2011)**, pp. 351-359.

[16] S. Youngjoo, J. Mikyong, K. Hoirin. "Probabilistic class histogram equalization for robust speech recognition", IEEE Signal Processing Letters, vol. 14, no. 4, **(2007)**, pp. 287-290.

[17] P. Shing-Tai, L. Min-Lun. "An efficient hybrid learning algorithm for neural network-based speech recognition systems on FPGA chip", Neural Computing and Applications, vol. 24, no. 7-8, **(2014)**, pp. 1879-1885.

[18] K. Ozlem., S. Michael L., A. Alex.. "Noise adaptive training using a vector taylor series approach for noise robust automatic speech recognition", IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings, Taipei, Taiwan, **(2009)**, pp. 3825-3828.

[19] S. Haifeng, L. Qunxia, G. Jun, Liu Gang. "HMM parameter adaptation using the truncated first-order VTS and EM algorithm for robust speech recognition", Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), Xi'an, China, **(2005)**, pp. 979-984.

[20] M. Osama, G. Jason, M. Max. "An HTK-developed Hidden Markov Model (HMM) for a voice-controlled robotic system", 2004 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Sendai, Japan, **(2004)**, pp. 4050-4055.