주성분 해석기법에 의한 음성 특징벡터 추출 및 인식 성능 평가

이광석*, 김현주**

요 약

본 연구에서는 기존 방법과는 달리, 음성의 특징추출의 한 방법으로 음성의 통계적인 특성들을 고려하여 입력 공간 내에서 변동량이 가장 많은 방향으로 주축을 발견한 후, 그 정보를 이용하여 데이터의 중복성을 제거하는 주성분 해석기법을 사용하여 음성의 특징을 추출하는 방법을 제안한다. 본 연구에서의 숫자음 인식실험 결과가 기존 멜-켑스트럼을 사용하였을 때와 0.5%의 인식률 차의 유사 인식률을 나타내었다. 한편, 데이터의 통계적 특성을 이용한 최적의 기저벡터를 이용하여 단어나 문장 인식에 적용한다면 보다 더 개선된 인식률을 얻으리라 기대하고 있다.

Performance Evaluation for Speech Recognition and Extraction of the Speech Characteristic Vector by using the Principal Component Analysis Method

Gwang-Seok Lee*, Hyun-Ju Kim**

ABSTRACT

The new method of characteristic extraction is proposed, considering the statistic characteristic of human speech, unlike the conventional methods of the traditional speech characteristic extraction. PCA(Principal Component Analysis) is applied to new this method. Then, the new method is applied to real speech recognition to assess performance. When results of the number recognition in this research and the conventional Mel-cepstrum of speech characteristic parameter are compared, there is 0.5% difference of recognition rate. Better recognition rate is expected than word or sentence recognition in that less convergence time than the conventional method in extracting speech characteristic. Also, Better recognition rate is expected when the optimum vector is used by statistic characteristic of data.

Key Words : Speech Signal Processing, Principal Component Analysis, Feature Extraction, Speech Recognition

* 진주산업대학교 전자공학과(✉kslee@jinju.ac.kr)

** 진주산업대학교 컴퓨터공학과

· 제1저자(First Author) : 이광석 · 교신저자(Correspondent Author) : 김현주

· 접수일(2010년 10월 22일), 수정일(1차 : 2010년 11월 19일), 게재확정일(2010년 11월 23일)

I. INTRODUCTION

Human Speech is the natural communication means and human-machine interface by the speech is fast thing without any special training. Also, speech recognition technique is the main research topic due to computer technique and information telecommunication technique, and much research is being carried out to increase performance of speech recognition up to date. The part of the speech recognition shows the good performance of speech recognition up to date especially. But, pattern recognition and ability of pattern clustering are limited. The most important topic to increase performance of the speech recognition is to select the speech characteristic of input speech of speech recognition unit efficiently, finally.

That is, it has the good performance of speech recognition as using input speech recognition unit as extracting characteristic reflecting each patterns characteristic. It is essential to recognize speech, not to accept the effect of the speech compression, because the speech characteristic parameter obtained from the above method shows characteristic of speech signal.

In this research, we propose to extract speech characteristic using Principal Component Analysis method to removes the repeating of data with the above data after finding the axis direction which has the greatest variance in input dimension considering characteristic of the speech statistics[1]. Then, the new method applied to real speech recognition, and assessed performance of the speech recognition. We focused on whether Principal Component Analysis method is actually available for applying to the method of the characteristic extraction in speech recognition. After finding the basis vector using the Principal Component Analysis method to the input speech the flow of the characteristic extraction algorithm is as follows. It is to extract speech characteristic by the correlation of the basis vector and input speech signal considering the basis vector coefficients.

The configuration of this study is as follows. Principal Component Analysis Method was mentioned in section II. In section III, we simulated with an input, the produced basis vector using Principal Component Analysis and examined performance of speech recognition. and discuss result of the speech recognition simulation in section Ⅳ[2][3][4][5]. Finally, we conclude this research.

II. PROPOSED ANALYSIS METHOD

2.1 Extraction of the Mel-cepstrum Coefficients

Human ability of auditory sense is represented the approximate logarithmic characteristics to the speech magnitude, Frequency resolution ability is represented linear in the low frequency under 1kHz, and the logarithmic Mel-scaled characteristic over 1kHz. Mel-cepstrum coefficients $\{Mc_m\}$ is approximately represented by bilinear frequency transform using phase characteristics of the whole pass filter of equation (1) from LPC cepstrum coefficients $\{c_m\}$, and it could be derived from approximation formula of equation (2).

$$H_{BT}(Z) = (Z^{-1} - a)/(1 - aZ^{-1}) \quad (1)$$

$$Mc_k(m)=\begin{cases} c_{-m}+a\cdot Mc_0(m-1), k=0 \\ (1-a^2)\cdot Mc_0(m-1)+a\cdot Mc_1(m-1), k=1 \\ Mc_{k-1}(m-1)+a(Mc_k(m-1)-Mc_{k-1}(m)), k>1 \\ m=\cdots,-2,-1,0 \end{cases} \quad (2)$$

If parameter a is positive, it could raise the resolution of low frequency, when $0.4 < a < 0.8$, it could transform Mel-scale into Bark scale. Where, in case of sampling frequency is 6.67kHz, 8kHz, 10kHz, 16kHz, putting a as 0.28, 0.31, 0.35, 0.45 individually, it could derive Mel-cepstrum approximately[2].

2.2 Principal Component Analysis Method

The most important topic is to extract characteristic considering the speech characteristic. This characteristic extraction problem is the main one in real speech data The Principal Component Analysis method is one of decreasing dimension and extracting characteristic methods, and is called "Karhunen-Loeve" transform in pattern recognition fields [2][3]. If the unit vector to the n dimension signal x with zero-mean characteristic is v_i, n available projection p_i with $i=1,2,\cdots,n$ is shown equation (3).

$$p_i = x\cdot v_i = x^T\cdot v_i = v_i^T\cdot x \quad (3)$$

It is shown equation (4) if scalar p_i represent characteristics to each unit vector as vector, as p_i is project x onto v_i area of unit vector dimension.

$$P=\{p_i, p_i,\cdots, p_n\}=[x^Tv_1, x^Tv_2,\cdots, x^Tv_n]^T = V^Tx \quad (4)$$

The method to recover input x using vector p unit vector v_i in equation (4) is shown equation (5), input signal can find with unit vector and projection value to that.

$$x = V^Tp = \sum_{i=1}^{n} p_iv_i \quad (5)$$

Input signal x represented by unit vector v_i has only coordinate transformation without reduction of the dimension on the above. Next,it explain approximation method of x as more small order with reduced dimension of v_i and p. When x reduced from n to $m\,(m<n)$ in i of equation (5) is ĕ, it shows as equation (6).

$$\check{e} = \sum_{i=1}^{m} a_iu_i \quad (6)$$

Where, a_i and u_i are eigen value and part of the eigen vector of covariance matrix R of x. The method to find R, eigen value and eigen vector is shown equation (7) and equation (8) respectively. It is a that get higher m among descending order eigen vector λ of R of n in equation (8). The u_i represent as $\overline{u}_i$ is shown in equation (9).

$$R=\frac{1}{n}\sum_{i=1}^{n}(xx^T) \quad (7)$$

$$Ru_i = \lambda u_i, i = 1,2,\cdots,n \quad (8)$$

$$R\bar{u}_i = a\bar{u}_i, i = 1,2,\cdots,m \quad (9)$$

When x and error of approximated ĕ is e, that's magnitude is shown equation (10).

$$e = \sum_{m+1=n}^{n} a_i \bar{u}_i \quad (10)$$

III. SIMULATION OF THE SPEECH RECOGNITION

3.1 Speech Data and Conditions of the Analysis

We simulated for speech recognition with ten speech data "gong", "ihl",…, "paal", "gu~h" among the "ETRI samdori" speech database in this research. We used 600 data that twenty adults spoken ten number of digits three times to find the basis vector using PCA and used 200 data for test speech recognition unit among total 800 data that twenty adults spoken ten number of digits four times.

Speech signal quantized as 16bits in 16kHz of sampling frequency. The portion of speech data has 60 samples, 3.75msec of time intervals[6][7][8]. We used HMM model of the continuous output distribution, left-to-right model and the model of each digit represented by 16 states.

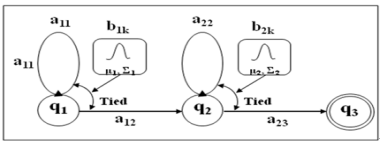


그림 1. 연속출력분포의 HMM 모델

Fig. 1. HMM Model of the Continuous Output Distribution

It produces 60×60 of basis vector when it applied PCA algorithm to all of the input speech signals containing 60 samples in the simulation. This produced basis vectors are arranged in the importance according to the basis vector coefficients. The basis vector coefficients represented by importance to each basis vector is shown as Fig.2.

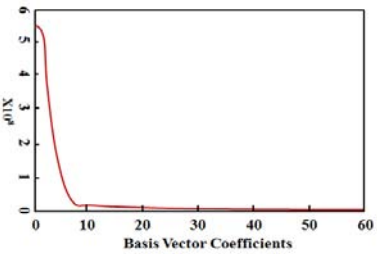
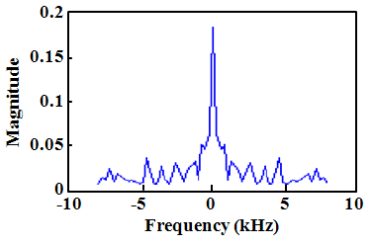


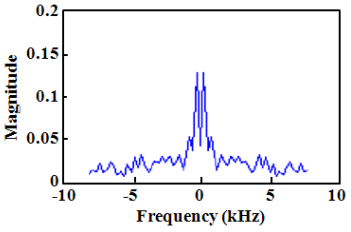
그림 2. 기저 벡터 계수

Fig. 2. Coefficients of the Basis Vector

We acknowledged that basis vector coefficients falling rapidly in Fig.2. Also, it shows that frequency characteristic to the 1st, 2nd, 10th, 20th basis vectors in Fig.3.



(a)



(b)

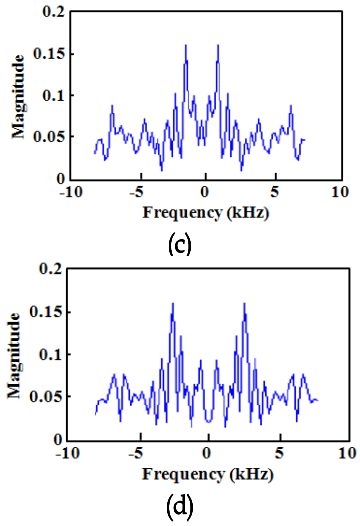


그림 3. 기저벡터의 주파수 특성
Fig. 3. Frequency Characteristics to the Basis Vector

Fig.4. is shown center frequency of 60 basis vectors got by PCA algorithm arranged in order. It can find that center frequencies of the basis vectors in range of 250Hz~8kHz has linear distribution. Also, we acknowledged that it arranged from lower frequency component to higher frequency component. This is attributed to that human speech energy has more energy in range of lower frequency component relatively. From the above facts, we can choose the important characteristic using PCA algorithm to apply speech recognition.

PCA algorithm method can find independent elements corresponding input and it may occur unnecessary elements or overlap elements. We choose dimension of the basis vector considering variance in weights of the basis vector coefficients to reduce in this unnecessary and overlap. This chosen basis vector decrease in dimension with correlation of the input speech signal, it translate into the input characteristic parameter.

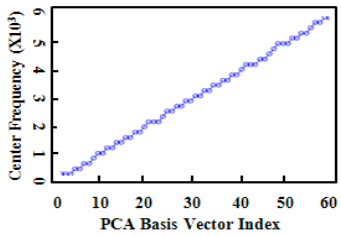


그림 4. 기저벡터의 중심 주파수
Fig. 4. Central frequency of the basis vector

IV. SIMULATION RESULT AND CONSIDERATION

Table 1 is shows that speech characteristic of Mel-cepstrum according to the number of states of HMM, and it shows that compare speech characteristic using proposed method with simulated recognition results that varying selected basis vector. When it selected number of 15 basis vector according to the weights of basis vector, recognition rates proposed with speech characteristic was 98.5%.

표 1. 기저벡터에 대응하는 인식률 변화
Table 1. Change of the recognition rate according to the basis vector (%)

-	12 states	15 states	16 states
mel-cepstrum	99.0		
10 basis vector	97.0	97.0	96.5
15 basis vector	98.0	98.0	98.5
20 basis vector	97.5	97.5	97.5

V. CONCLUSIONS

In this research, basis vector got by PCA algorithm applied speech recognition considering with basis vector coefficients. When results of the number recognition in this research and the conventional Mel-cepstrum of speech characteristic parameter are compared, there is 0.5% difference of recognition rates. But, because PCA algorithm utilize the statistics characteristic of the input speech data, when it use optimum basis vector, we can recognized that it is the one of the characteristic extraction methods coding human speech signal even if use parts of the basis vector efficiently.

Better recognition rate is expected than word or sentence recognition in that less convergence time than the conventional method in extracting speech characteristic. Also, Better recognition rate is expected when the optimum vector is used by statistic characteristic of data. The proposed PCA algorithm method in this research has good advantage of extracting time of characteristic parameter. Therefore, we hope that proposed PCA algorithm method has good performance to the real time processing of speech recognition.

Acknowledgement

This work was supported by Jinju National University Grant.

REFERENCES

[1] Zhao, S. J., Zhang, J. and Xu, Y. M., "Monitoring of Processes with Multiple Operating Modes through Multiple Principle Component Analysis Models", Ind. Eng. Chem. Res., 43(22), 7025-7035(2004).

[2] Tuske, Zoltan and Mihajlik, Peter and Tolbar, zoltan and Fegyo, Tibor "Robust voice activity detection based on the entropy of noise suppressed spectrum", in *Proc.* of *Interspeech*, pp.245-248, Sep. 2005.

[3] S. Rangachari and P.C. Loizou "A noise estimation algorithm for highly non-stationary environments", Speech Communication, vol 48, no 2, pp.220-231, 2006.

[4] Venkatasubramanian, V., Rengaswamy, R., Yin, K. and Kavuri, S. N., "Review of Process Fault Detection and Diagnosis Part I: Quantitative Model-based Methods", Comput. Chem. Eng., 27(3), 293-311(2003).

[5] David Kozel and Constantin Apostoaia, "Colored Noise Reduction Using Bark Scale Spectral Subtraction, Statics, and Multiple Time Frames", in *Proc. IEEE International Conference Electro/Information Technology*, pp. 416-421, May, 2007.

[6] Ahmed, B. Holmes, P.H., "A voice activity detector using the chi-sequre test", *Acoustics, Speech, and Signal Processing, 2004. Proceedings*., pp.I-625-8, R. Melbourne Inst. of Technol., Vic., Australia, May, 2004.

[7] Venkatasubramanian, V., Rengaswamy, R., Kavuri, S. N. and K., "Review of Process Fault Detection and Diagnosis Part III: Process History Based Methods", Comput. Chem. Eng., 27(3), 327-346(2003).

[8] University of Texas Dallas Speech Copus NOIZEUS,http://www.utdallas.edu/~loizou/speech/noizeus/ , 2007.

이광석(Gwang-Seok Lee)

1983년 동아대학교 전자공학과(공학사)
1985년 동아대학교 대학원 전자공학과 (공학석사)
1992년 동아대학교 대학원 전자공학과 (공학박사)
1995년~현재 국립진주산업대학교 전자공학과 교수
※ 관심분야 : 음성신호처리 및 인식, 퍼지 및 신경회로망, Biometrics, 지능화 기술

김현주(Hyun-Joo Kim)

1988년 경상대학교 컴퓨터과학과(이학사)
1990년 숭실대학교 대학원 전자계산학과 (공학석사)
2000년 경상대학교 대학원 컴퓨터과학과 (이학박사)
2002년~현재 국립진주산업대학교 컴퓨터공학과 부교수
※ 관심분야 : 정보검색, XML, 컴퓨터교육, 데이타마이닝