최적경로와 가중직교인자를 이용한 화자인식

Speaker Recognition Using Optimal Path and Weighted Orthogonal Parameters

박 승 규* 배 철 수**

(Seung Kyu Park* Chul Soo Bai**)

요 약

최근, 많은 연구자들이 KLT를 이용한 통계적 처리방법으로 화자인식을 수행하고 있으나, 통계적 처리방법의 개인성 포함정도와 음성의 동적인 발성속도는 화자인식율의 저하요인이 되고 있다.

본연구에서는 각 화자의 직교인자에 개인성을 강조하기 위하여 화자의 고유치를 가중치로 한 가중직교인자와 음성의 동적인 시간특성을 정규화하는 DTW의 최적경로를 이용한 화자인식방법을 연구하였다.

이방법을 확인하기 위하여 종래의 통계적 처리에 의한 화자인식, 최적경로와 최적경로와 가중직교인자를 이용한 화자인식의 결과를 비교한 결과, 종래의 방법보다 우수한 화자인식율을 얻어 그 유효성을 확인하였다.

ABSTRACT

Recently, many researchers have studied the speaker recognition through the statistical processing method using Karhunen-Loeve Transform. However, the content of speaker's identity and the vocalization speed cause speaker recognition rate to be lowered.

This parer studies the speaker recognition method using weighted orthogonal parameters which are weighted with eigen-values of speech so as to emphasize the speaker's identity, and optimal path which is made by DWP so as to normalize dynamic time feature of speech.

To confirm this method, we compare the speaker recognition rate from this proposed method with that from the conventional statistical processing method. As a result, it is shown that this method is more excellent in speaker recognition rate than conventional method.

I. 서 론

화자인식이란 음성의 특징을 토대로 사람들을 구별해내는 작업을 말한다. 1945년 미국의 Bell 연구소에서 관찰에 의한 화자인식을 위하여 성문(voice-print)을 자동적으로 추출하는 sound spectrograph가 발명되면서부터 보다 폭 넓은 자동화자인식에 대한 연구가 시작 되었다.

1963년 Pruzansky[1]는 17차 필터군을 이용하여 화

* 전북산업대학 전자계산학과
** 관동대학교 전자통신공학과

자인식을 하였으며, Furui[2]는 12차 RARCOR 계수와 Pitch를 이용하여 화자인식을 하였다. 또한, 1976년 Sambur[3]는 각각 LPC, PARCOR및 Log Area Ratio의 계수로부터 통계적으로 추출한 직교인자들의 거리를 구하여 화자인식의 실험을 수행하여 직교인자가 음운성보다 개인성을 더 많이 포함하고 있음을 확인하였으며, 1977년 Markel[4]은 평균값을 이용하여 긴 기간동안의 화자인식을 시작하였다. 그러나, 동일한 화자에 대해서도 발성되는 음성은 시간에따라 동적으로 변화하기 때문에, 음성신호의 동적인 특성은 화자인식에 있어서 간과할 수 없는 하나의 중요한 문제[5][6]라고 할 수 있다. 그리고, 화자인식율을 향상시키는 문제에 있어서는 비교해야 할 계수의 분산을 가중치[7]로 고려하여 줌으로써, 인식율을 더욱 향상 시킨 바 있다.

본 연구에서는 음성의 시간에 따른 동적인 변화를 정규화하기 위하여 비선형 시간축 정규화법인 DTW 방식으로 추적한 최적경로와, 화자인식율을 향상시키기 위하여 직교인자의 고유치를 가중치로 고려한 가중직교인자를 이용하여 화자인식을 수행하였다. 또한 이방법의 우수성을 확인하기 위하여 종래의 통계적 처리방법과 비교 검토하였다.

Ⅱ. 통계적 처리에 의한 화자인식

음성을 차단주파수가 3.4 KHz인 저주파 필터를 통한후, 10KHz(16 bit resolution)로 샘플링 하였다. 이 샘플링된 음성데이타의 에너지와 영교차율을 이용하여 무음구간을 제외하고 음성구간만을 추출하였으며, 이 추출된 데이타를 autocorrelation방법으로 10차의 PARCOR계수를 구하였다.

이상의 PARCOR 계수로 부터 Karhunen Loeve 변환을 통하여 직교인자를 구하는 과정은 다음과 같다.

1)PARCOR 계수들의 covariance 행열을 계산한다.

$$C_{ij}=1/(NF-1)\sum_{k=1}^{NF}(X_{ik}-\bar{X}_i)(X_{jk}-\bar{X}_j) \quad (1)$$

$$\bar{X}_i=1/NF\sum_{k=1}^{NF}X_{ik} \quad (2)$$

X_{ik} : k번째 프레임의 i차 계수
NF : 음성구간내 프레임 수
$\bar{X}_i$: i차 계수의 평균값

2)covariance 행열의 고유치와 고유 벡터를 구한다.

3)고유 벡터를 단위 길이로 정규화 시킨다.

4)고유 벡터를 이용하여 직교 인자를 구한다.

$$\phi_{ik}=\sum_{i=1}^{p}T_{ki}X_{ii} \quad (3)$$

ϕ_{ik} : i번째 프레임의 k차 직교 인자
T_{ki} : k차 정규화된 고유벡터
p : 차수

이상에서, 고유치를 이용한 통계적 처리 방법에 의한 거리는 다음과 같다.

$$D1=1/2\,J_m\sum_{i=1}^{p}[(V_i-\lambda_{im})/\lambda_{im}]^2 \quad (4)$$

여기서, J_m : m번째 표준 패턴 구성시 사용된 평균 프레임수
λ_{im} : m번째 표준패턴의 i번째 고유치
V_i : 시험패턴의 i번째 고유치
이다.

Ⅲ. 최적경로와 가중직교인자를 이용한 화자인식

(1)최적경로

본 연구에서는 화자 인식을 직교인자를 DTW 방식과 결합시킨 것으로써, log likelihood ratio 거리 측정 방법으로 DTW 방식의 최적 경로를 구한후, 그 경로에 따라 직교인자를 대입하여 거리를 구하는 것이다. 그 구성도는 다음과 같다.

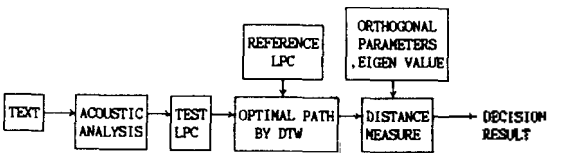


그림 1. 화자인식의 구성도
Fig. 1. Block diagram of speaker recognition.

음성의 동적인 시간특성을 정규화하기 위하여 DTW에 의해 최적경로를 구하기 위한 각 프레임 단위의 거리는 log likelihood ratio로 다음과 같다.

$$d(n,m)=\ln\left|\frac{a_m V_n a_m}{a_n V_n a_n}\right| \quad (5)$$

a_m : 표준패턴 m 프레임의 선형 예측 계수 벡터
V_n : 시험패턴 n 프레임의 자동 상관 함수 벡터
a_n : 시험패턴 n 프레임의 선형 예측 계수 벡터

이상에서 구해진 두가지 형태의 프레임 단위의 거리를 수행하는 DTW 알고리즘은 기울기는 1이고 대칭인 조건을 가지고 있으며, 초기치 설정과 반복 계산하는 누적 거리는 다음과 같다.

(초기치 설정)

$D(1,1)=2d(1,1)$
$D(1,m)=D(1,m-1)+d(1,m)$: $2<m<r$
$D(n,1)=D(n-1,1)+d(n,1)$: $2<n<r$
$D(1,m)=$: $m>r$
$D(n,1)=$: $n>r$ (6)

여기에서 r은 adjustment window이다.

(누적거리)

$$D(n,m)=\min \begin{matrix} D(n-1,m-2)+2d(n,m-1)+d(n,m) \\ D(n-1,m-1)+2d(n,m) \\ D(n-2,m-1)+2d(n-1,m)+d(n,m) \end{matrix} \quad (7)$$

(2)직교인자를 이용한 거리측정

식 (5)와 같이 주어진 log likelihood ratio를 DTW 알고리즘에 적용하여 시간축 정규화 거리를 구할 때 얻는 최적 경로를 구한 후, 다음과 같은 최종적인 거리를 구함으로써 화자인식을 한다.

$$D2=\sum_{F}\sum_{k=1}^{P}|\phi_{nk}-\phi_{mk}| \quad (8)$$

F : 최적 경로 warping funtion
n,m : F로 결정된 시험 패턴과 표준 패턴의 mapping 프레임

(3)가중직교인자를 이용한 거리측정

본 논문에서 사용한 가중치는 음성의 고유치로써 직교인자의 분산이며, 직교인자로 구한 거리에 가중치를 곱하는 형태와 나누는 형태의 두가지를 사용 하였다.

$$D21=\sum_{F}\sum_{k=1}^{P}|\phi_{nk}-\phi_{mk}|\times w_k^{1/2} \quad (9)$$

$$D22=\sum_{F}\sum_{k=1}^{P}|\phi_{nk}-\phi_{mk}|\times w_k^{-1/2} \quad (10)$$

w_k : 표준 패턴의 k차 직교인자의 분산

Ⅳ. 실험및 결과 고찰

본논문에서 사용한 음성은, "아가야 아가야 이리 오너라"로서 가능한한 유성음이 많이 포함된 문장을 택하였다. 이 문장을 일곱 사람이 네 기간에 걸쳐 각각 10번씩 자연스러운 상태에서, 미리 훈련을 받지 안고 반복 녹음 하였다.
또한 통계적 처리에 의한 거리측정인 식(4)에서 사용한 표준패턴은 네달전 같은 사람이 한달을 주기로 세 기간에 걸쳐서 10번씩 녹음한 문장으로부터 구한 각 고유치의 평균값으로 하였고, 직교인자를 DTW 방식과 결합시킨 방법에서는 표준거리 측정의 표준 패턴에 사용된 각 화자의 30개의 문장중에서 MKM 방법(modified K-means method)을 이용하여 표준 패턴으로 선택하였다.

표1, 표2, 표3, 표4는 각각 식 (4),(8),(9),(10)에 따른 결과를 나타내고 있다.

표 1. 통계적 처리에 의한 인식결과
Table 1. Recognition result using the statistical processing.

TEST / REF.	A	B	C	D	E	F	G
A	5				6		
B	3	10					
C			9				
D			1	10		10	
E	2				4		
F							
G							10

표 2. 최적경로와 직교인자를 이용한 인식결과

Table 2. Recognition result using optimal path and orthogonal parameters.

TEST REF.	A	B	C	D	E	F	G
A							
B		10					
C			10				
D				10			
E					10		
F						10	
G	10						10

표 3. 최적경로와 가중직교인자 식(9)를 이용한 인식결과

Table 3. Recognition result using optimal path and weighted orthogonal parameters Eq.(9)

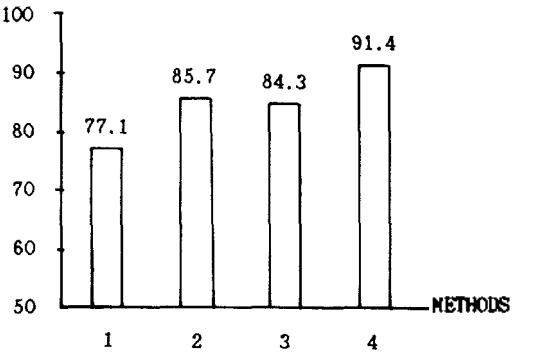
TEST REF.	A	B	C	D	E	F	G
A							
B		9					
C			10				
D				10			
E					10		
F						10	
G	10	1					10

표 4. 최적경로와 가중직교인자 식(10)를 이용한 인식결과

Table 4. Recognition result using optimal path and weighted orthogonal parameters.

TEST REF.	A	B	C	D	E	F	G
A	5						
B	4	10					
C			10			1	
D				10			
E					10		
F						9	
G	1						10

표1은 통계적 처리에 의한 화자인식의 결과로써 77.1%의 인식율을 보이고 있으며, 표2는 최적경로와 직교인자를 이용한 화자인식의 결과로써 85.7%의 인식율을 보이고 있다. 여기서 표2의 인식율이 통계적 처리방법보다 향상된 것은 음성의 동적인 시간 특성을 정규화 해 주는 DTW의 최적 경로를 이용했기 때문이라 생각된다. 또한, 표3에서는 DTW의 최적경로와 식(9)와 같이 고유치를 곱하는 형태의 가중직교인자를 이용한 인식결과이고, 표4는 최적경로와 식(10)과같이 고유치를 나누어 주는 형태의 가중직교인자를 이용한 인식결과이다. 이들로 부터 각각 84.3%, 91.4%의 인식율을 얻었다. 여기서 식(10)형태의 가중직교인자를 이용하여 더 높은 인식율을 얻은 것은 직교인자를 고유치로 나누어 주므로써 차수가 높아져도 그 크기가 작아지지 않도록하여 직교인자의 차수별로 갖는 음성의 개인성이 강조 되었기 때문이라 생각된다.



1 : 통계적 처리에 의한 인식결과
2 : 최적경로와 직교인자를 이용한 인식결과
3 : 최적경로와 가중직교인자 식(9)를 이용한 인식결과
4 : 최적경로와 가중직교인자 식(10)를 이용한 인식결과

그림 2. 화자인식율의 비교

Fig. 2. Comparison of speaker recognition rate.

V. 결 론

본 연구에서는 직교 인자를 이용하는 화자인식에서 음성신호의 시간에 따른 동적인 특성을 고려하여 주기 위하여 DTW 방식으로 거리를 측정하였으며,

화자 인식율을 보다 향상시키기 위하여 직교인자의 고유치를 가중치로 한 가중직교인자를 이용하여 거리를 측정하였다.

이 방법에 따라 실험을 수행한 결과, 통계적 처리에 의한 방법보다 최적경로와 직교인자를 이용한 방법에서 높은 인식율을 얻을 수 있었으며, 개인성을 강조하기 위해 제안한 가중직교인자를 최적경로에 적용하여 보다 높은 화자인식율을 얻으므로써 최적경로와 가중직교인자를 이용한 화자인식방법의 우수성을 확인하였다.

참 고 문 헌

1. S.Pruzansky, "Pattern-matching procedure for automatic talker recognition," J.Acoust.Soc Am., Vol.35, No.3, 1963.
2. Furui, "單語の統計的パラメータによる話者認識," 信學論 論文誌, 55A-10, 1972.
3. M.R.Sambur, "Speaker recognition using othogonal linear prediction," IEEE.ASSP-24, No.4, 1976.
4. J.D.Markel, et al., "Long term feature averaging for speaker recognition," IEEE.ASSP-25, No.4, 1977.
5. S.Furui, "Comparison of speaker recognition methods using statistical features and dynamic features," IEEE.ASSP-29, No.3, 1981.
6. J.Luck, "Automatic Speaker verification using cepstral measurements," J.Acoust.Soc.Amer., Vol.46, pp.1026-1031, 1969.
7. Y.Tohkura, "A weighted cepstral distance measure for speaker recognition," IEEE.ASSP-35, No.10, 1987.
8. F.Itakura, "Minimum Prediction residual principle applied to speech recognition," IEEE, ASSP-23, 1975.
9. 배철수, 직교인자의 동적특성을 이용한 화자인식, 명지대학교 전자공학과 1988년도 박사학위논문
10. D.O'shaughnessy, "Speaker Recognition," IEEE, ASSP magazine, 1986.
11. J.G.Wilpon, L.R.Rabiner, "A Modified K-Means Clustering Algorithm for Use in Isolated Work Recognition," IEEE Trans.Acoust.Speech. Signal Precessing, Vol.ASSP-33, No.3, 1985.
12. L.R.Rabiner, M.R.Sambur, "An algorithm for determining the endpoints of isolated utterences," Bell Syst.Tech.J., vol.54, pp.297-315, 1975.

▲朴 承 圭(Park Seung Kyu)

1976년 3월 2일~1983년 2월 : 조선대학교 전자공학과 학사
1983년 3월~1985년 2월 : 조선대학교 전자공학과 석사
1987년 9월~현재 : 조선대학교 전기공학과 전산전공 박사과정 수료
1988년 3월~현재 : 전북산업대학 전자계산학과 조교수

▲배 철 수

1950년 3월 14일생.
1979년 2월 : 명지대학교 전자공학과 졸업
1982년 2월 : 명지대학교 대학원 전자공학과 석사
1988년 8월 : 명지대학교 대학원 전자공학과 박사
1981년 3월~1990년 11월 : 관동대학교 전자공학과 부교수
1990년 12월1일~현재 : 관동대학교 전자통신과 부교수