LPC에 의한 화자 식별

*조병모, 송봉석, 황영수, *박종철, 차일환, 윤대희

연세대학교 전자공학과 음향 음성 및 신호 처리 연구실

국립과학 수사연구소 음성연구실

On using the LPC parameter for Speaker Identification

*B.M. Cho, B.S. Song, Y.S. Hwang, *J.C. Bag, I.W. Cha, D.H. Youn

Acoustics, Speech and Signal Processing Laboratory

Dept of Electronic Engineering Yonsei University, Seoul, Korea

* Acoustic Phonetics section

The national institute of scientific investigation

ABSTRACT

Preliminary results of using the LPC parameter for text-independent speaker identification problem are presented. The identification process includes log likelihood ratio for distance measure and dynamic programming for time normalization. To generate the data base for experiments, ten speakers were asked to utter the same sentance ten times. Experimental results show 99.4% of identification accuracy. incorrect identification were made when the speaker uses a dialect.

1. 서론

화자 식별에 대한 연구는 청각에 의한 식별, 시각에 의한 식별 그리고 기계적 수단에 의한 식별이 있으며, 기계적 수단에 의한 화자 식별으로는 filter bank등 특수한 hardware와 전자 계산기를 이용하여 음성에 포함되어 있는 개인성 정보를 추출하고 이 추출한 parameter와 미리 등록되어 있는 parameter와의 pattern matching 수법을 이용하여 화자를 식별한다. 1945년 미국의 Bell 연구소에서 sound spectrograph가 발명되어 이를 이용하여 성문(voice print)을 추출할 수 있어 음성으로 화자를 식별할 수 있게 되었다. 이후 1963년 Pruzansky가 17차 filter bank를 이용하여 87%의 인식율을 얻었으며 [1] 일본의 Furui가 Linear prediction으로 12차 PARCOR 계수와 pitch를 이용하여 99%의 인식율을 얻었다[2]. 그 외 Lummis, Doddington, Atal, Hughes, Su 등이 98 % ~ 99 %의 인식율을 얻었다[3 ~ 6]. 그리고 1976년 Sambur가 선형예측 parameter (LPC, PARCOR)를 이용한 text-independent 화자 식별을 실험하여 94 %의 인식율을 얻었으며 음운성 정보와 개인성 정보를 분리할 수 있다고 설명하고 있다 [7]. 본 논문에서는 text-independent 화자 식별의 기초 연구로서 LPC를 이용한 text-dependent 화자 식별을 행하려 한다.

2. 연구 방법

음성에 포함되어 있는 개인성 정보를 정확히 추출하여 화자가 누구인가를 식별하는데는 두가지가 있다. 하나는 표준 패턴 (reference pattern)과 시험 패턴 (test pattern)의 내용이 같은 text-dependent 화자 식별과 내용이 같지 않은 text-independent 화자 식별이 있다. 화자 식별에 이용되는 개인성 정보에는 성도 길이(vocal tract length), 성대(vocal cord) 등 선천적인 발성 기관의 개인차에 기인한 것과 액센트, 방언등 후천적인 발성 습관에 기인한 것이 있다. 성도 길이나 성대는 formant 주파수의 고저, 대역폭의 대소, 평균 기본 주파수 등으로 판정할 수 있으며 액센트, 방언등은 에너지, pitch의 시간적인 변화로 판정할 수 있지만,

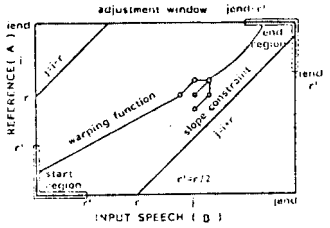

그림 4.Warping function by dynamic programming technique

이때 A,B 사이의 matching 값은 다음과 같이 정의된다.

$$D(A,B) = \min [\, g(i,j) / (i + j) \,]$$

3. 실험 방법 및 결과 고찰

한 문장(지금 몇시입니까?)를 한사람이 10번 반복 발성한 문장을 녹음하여 차단 주파수가 4.5KHZ인 Low pass filter를 통한 후 10 KHZ (16 bit resolution)로 sampling 하였다. 분석 구간은 25.6 ms, 분석 차수는 14차로 하였다. 화자에 따른 단어의 지속시간 차를 흡수하기 위하여 DP matching에 사용된 window length는 10 frame 으로 하였으며 표준 패턴과 시험 패턴의 관계를 나타내면 표1과 같다.

표 1, 표준 패턴과 시험패턴 관계

표준 패턴	시험 패턴									
A1	A2	B2	C2	D2	E2	F2	G2	H2	I2	J2
A2	A3	B3	C3	D3	E3	F3	G3	H3	I3	J3
:					:					
J10	A1	B1	C1	D1	E1	F1	G1	H1	I1	J1

이와 같이 반복하여 인식한 결과 전체 99.4%의 인식율을 얻었으며 인식 결과는 표2 와 같다.

표 2. 인식 결과

R \ T	A	B	C	D	E	F	G	H	I	J
A	100	0	0	0	0	0	0	0	0	0
B	0	100	0	0	0	0	0	0	0	0
C	0	0	100	0	0	0	0	0	0	0
D	0	0	0	100	0	0	0	0	0	0
E	0	0	0	0	98	0	2	0	0	0
F	0	0	0	0	0	99	1	0	0	0
G	0	0	0	0	2	1	97	0	0	0
H	0	0	0	0	0	0	0	100	0	0
I	0	0	0	0	0	0	0	0	100	0
J	0	0	0	0	0	0	0	0	0	100

4. 결론

본 연구에서는 text-independent 화자 인식의 기초 연구로써 LPC 계수에 의한 text-dependent 화자 인식을 행하였다. 인식 결과 99.4% 의 인식율을 얻었으며 G, B 가 인식율이 떨어진 원인은 방언이 혼성된 것으로 고려되며 이러한 문제점을 개선하려면 화자 특징만을 추출하는 algorithm 개발이 요구된다.

본 논문에서는 표준 패턴을 임시적으로 수집하였지만 실제적인 이용면을 고려하면 장시간에 걸쳐 표준패턴을 주기적으로 작성할 필요가 있으며, 이를 기본으로하여 발성 내용과는 관계없는 text - independent 화자 인식을 실험 중이다.

참고문헌

1] S. Pruzasky : Pattern-matching Procedure for automatic talker recognition, J. Acoust. Soc. Am. , 35 3, 1963.

2] Furui : 單語の統計的パラメータによる話者認識, 信學会論文誌, 55A-10, 1972.

3] R.C.Lummis : Speaker Verification by Computer using speech intensity for temporal registration,

이 두가지를 명확히 분리하여 추출하는 것은 곤란하다 [8]. 이 때문에 대부분의 화자 식별 실험은 이 두가지 특징이 동시에 포함되어 있는 특징 parameter를 이용한다. 일반적인 화자 식별 블럭도를 살펴보면 그림 1과 같다.

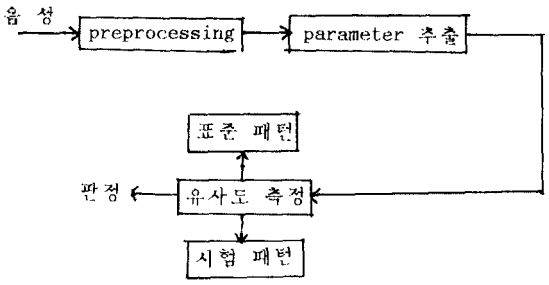

그림 1. 화자 식별 블럭도

실제의 음성 신호와 예측 신호의 차 e(n)은

$$e(n) = S(n) + \sum_{k=1}^{p} a(k)\, S(n-k) \quad \cdots\cdots\cdots\cdots (1)$$

$$E(Z) = S(Z) + \sum_{k=1}^{p} a(k)\, Z^{-k} S(Z)$$

$$= \left(1 + \sum_{k=1}^{p} a(k)\, Z^{-k}\right) S(Z)$$

$$= (1 + P(Z))\, S(Z) \quad \cdots\cdots\cdots\cdots (2)$$

따라서 음성 신호 S(n)과 잔차 신호 e(n)의 관계를 그림 2 와 같이 나타낼 수 있다.

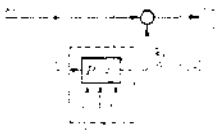

그림 2 선형 예측법의 구성

여기에서 e(n)이 최소가 되도록 하여 a(i)를 구하는 algorithm을 나타내면 그림 3 과 같다 [9].
본 논문에 사용된 LPC 거리 측정법으로는 Itakura [10] 에 의해 제안된 log likelihood ratio 이다.

$$d[a(r), a(t)] = \mathrm{Log}\left\{\frac{a(r)\, V(t)\, a'(r)}{a(t)\, V(t)\, a'(t)}\right\} \quad \cdots (3)$$

$a = (a_1, a_2, \ldots\ldots, a_p)$; 표준 패턴 분석 구간의

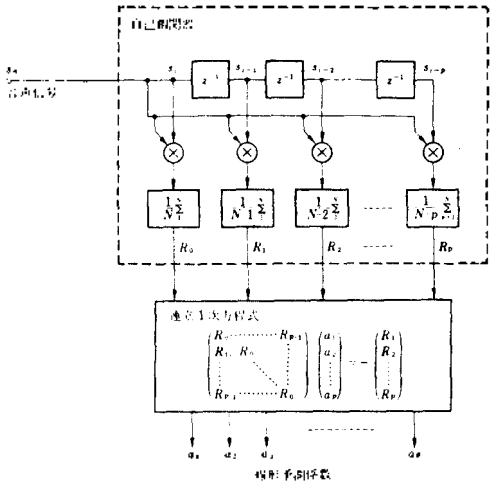

그림 3. 선형 예측 계수를 구하는 algorithm

LPC 계수.

$a = (a_1, a_2, \ldots\ldots, a_p)$; 시험 패턴 분석 구간의 LPC 계수.

V = 시험 표준 패턴 분석 구간의 자기 상관 행렬이다. 시간 축의 정규화 matching 수법으로 그림 4 에 나타낸 DP 법을 이용했다. 표준 패턴을 A, 입력 시험 패턴을 B , A 의 i 번째 frame과 B 의 j 번째 frame과의 거리를 d(i,j)라 하면 본 논문에 사용된 DP algorithm은 다음과 같다 [11].

(초기 설정)

$g(1,1) = 2\, d(1,1)$

$g(1,j) = g(1,j-1) + d(1,j)$; $2 \leqq j \leqq r'$

$g(i,1) = g(i-1,1) + d(i,1)$; $2 \leqq i \leqq r'$

$g(1,j) = \infty$; $r' \leqq j \leqq r$

$g(i,1) = \infty$; $r' \leqq i \leqq r$

(반복계산)

$$g(i,j) = \min \begin{cases} g(i-2,j-1) + 2d(i,j-1) + d(i,j) \\ g(i-1,j-1) + 2d(i,j) \\ g(i-1,j-2) + 2d(i-1,j) + d(i,j) \end{cases}$$

여기에서 $2 \leqq i \leqq i$, $2 \leqq j \leqq j$

i ; A frame 길이, j ; A frame 길이

IEEE Trans. Audio Electroacoust., AV-21-2 1973.

4] B.S.Atal : Effectiveness of linear prediction characteristics of the speech wave for automatic speaker identification and verification, J. Acoust. Soc. Am., 55-6, 1974.

5] G.W. Hughes : Talker differences as they appear in correlation matrices of continuous speech spectra, J. Acoust. Soc. Am. 55-4, 1974.

6] K.S.Su : Identification of speakers by use of nasal coarticulation, J. Acoust. Soc. Am, 56-6 1974.

7] M.R. Sambur : Speaker recognition using orthogonal linear prediction, IEEE Trans. Acoust., speech, Signal Processing, ASSP-24-4, 1976.

8] 新美康永, 音声認識, 共立出版株式会社.

9] 安居院 猛, コンピュータ音声処理 廣済堂産報出版.

10] F.I.Itakura : Minimum prediction residual principle applied to speech recognition, IEEE Trans. Acoust., Speech, Signal Processing, Vol ASSP-23, 1975.

11] M. Sugiyama : WLR measure applied to word recognition, 일본 전자 통신학회 논문지, Vol. J 66 No 4, 1984.