論文 88-25-7-1

不特定 話者의 音聲 認識을 위한 標準音 設定 方法에 관한 硏究

(A Study on the Creation Rule of Reference Templates to Recognize Speech for Speaker-independent)

金 桂 國,* 安 泰 玉,* 金 淳 協,* 李 鍾 岳**

(Kye Kook Kim, Tae Ock Ann, Soon Hyub Kim and Jong Arc Lee)

要 約

불특정 화자의 音聲을 認識하기 위해서는 각 화자의 성대 변화를 모두 수렴할 수 있는 標準音을 設定하는 일이 참으로 중요하다.

이를 위해서 集團化 방법을 도입하고 있으며 集團의 중심이 될 수 있는 標準音을 設定하는 방법이 문제의 핵심이 되고 있다.

本 論文에서는 기존의 minimax방법과 MMS (minimum of mutual sum) 방법을 사용하여 標準音을 設定하였다.

또한 標準音은 각 단어당 최고 3개까지 허용하여 인식 결과를 비교하였다. 동일음 11개에 대하여 3개의 標準音을 허용했을 경우 기존의 minimax 방법이 91.6%, MMS 방법이 95.8%의 인식률을 얻었다.

Abstract

It is very important that we create reference templates to recognize speech of speeker-independent as convergence as possible vocal tract variation of each speaker. We used to clustering technique for this and creation rule of reference templates to be cluster centers is key point of thema. In this paper, we created reference templates using the minimax for existance and MMS technique suggested in this study. Also, we created reference template until top 3 and compared to recognition result. When we create 3 reference templates recognition rate is 91.6% for minimax and recognition rate is 95.8% for MMS.

*正會員, 光云大學校 電子計算機工學科
(Dept. of Computer Science, Kwangwoon Univ.)
**正會員, 建國大學校 電子工學科
(Dept. of Elec. Eng., Konkuk Univ.)
接受日字 : 1987年 4月 20日

序 論

최근의 연구 보고에 의하면 不特定 話者의 음성을 認識시키기 위하여 모든 音聲을 集團化 하여 標準音을 設定하고 있다.

本 研究에서는 사전에 발음 연습을 시키지않은 3名의 話者(男性 2名 女性 1名)가 발음한 13개의 地下鐵名을 集團化하여 認識시켰다.

本 研究에서 認識 대상으로한 地下鐵名은 발음상 유사한 이대, 이촌, 이수, 옥수, 수유, 신촌, 신사, 신당, 사당, 대림, 신림, 신대방, 대방 등이다.

本 研究에서는 集團化하여 設定된 標準音을 토대로하여 그 認識 결과를 서술하고 기존의 方法과는 달리 標準音을 設定하는 새로운 方案을 제시하는데 그 目的이 있다.

그림 1은 本 研究의 인식실험 과정을 나타내고 있다.

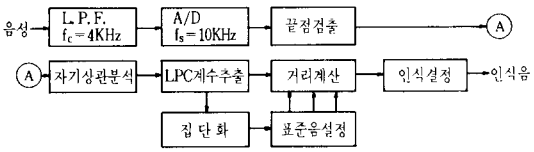

그림 1. 단독음 인식 시스템 계통도
Fig. 1. Block diagram of the isolated word recognition system.

ZCR(zero crossing rate)과 LE(log energy)를 이용하여 음성의 끝점을 검출하고 거리를 계산하기 위하여 12차 자기상관 계수와 12차 LPC 계수를 추출하였다.

이들 특징 파라미터를 이용하여 各 단어를 集團化하고 各 集團을 대표할 수 있는 標準音들을 設定 하였다.

이들 標準音과 비교 대상어들(test words)간의 유사도를 비교하기 위하여 거리를 계산하고 認識 결정법에 따라 認識하였다.

Ⅱ. 集團化 알고리즘

音聲 데이타의 특징 파라미터가 N개로 이루어질 때 이것을 식(1)과 같이 表現할 수 있다.

$$S=\{x_1, x_2, x_3, \cdots\cdots, x_N\} \qquad (1)$$

여기서 x_i는 서로 다른 話者가 반복 발음한 N개의 동일 음성(이하 토큰이라 한다)을 나타내고 있다.

이들 N개의 토큰이 K개의 集團을 형성한다고 하면 식(2)와 같이 表現할 수 있다.

$$S=\bigcup_{i=1}^{k} W_i \qquad (2)$$

本 研究에서는 LPC계수와 자기 상관계수를 특징 파라미터로 이용하였으므로 Itakura가 제안한 LPC 거리 측정법을 사용하여 유사도를 비교하였다. 이를 式으로 表現하면

$$d_{ij}=\delta(x_i, x_j)=\frac{1}{n_i}\sum_{k=1}^{n_i} d(k, w(k)) \qquad (3)$$

여기서 n_i는 標準音 x_i의 프레임수이며 $d(k, w(k))$는 x_i의 k번째 프레임과 x_j의 w(k)번째 프레임 사이의 LPC거리를 나타낸다.

$$d(k, w(k))=\log\left[\frac{\vec{a}_{w(k)}\, V\, \vec{a}_{w(k)}^{T}}{\vec{a}_k\, V\, \vec{a}_k^{T}}\right] \qquad (4)$$

$\vec{a}_k$는 비교 대상음 x_i의 k번째 프레임에서 추출한 LPC계수 벡터, $\vec{a}_{w(k)}$는 標準音x_j의 w(k)번째 프레임에서 추출한 LPC계수 벡터, v는 비교 대상음의 k번째 프레임에서 계산한 자기 상관계수 행렬, w(k)는 d_{ij}를 최소로 할 수 있도록한 워핑 함수라고 한다.

T는 전치 벡터를 뜻한다.

그림 2는 集團化 알고리즘에 대한 계통도로써 다음과 같은 과정을 통해서 이루어져 있다.

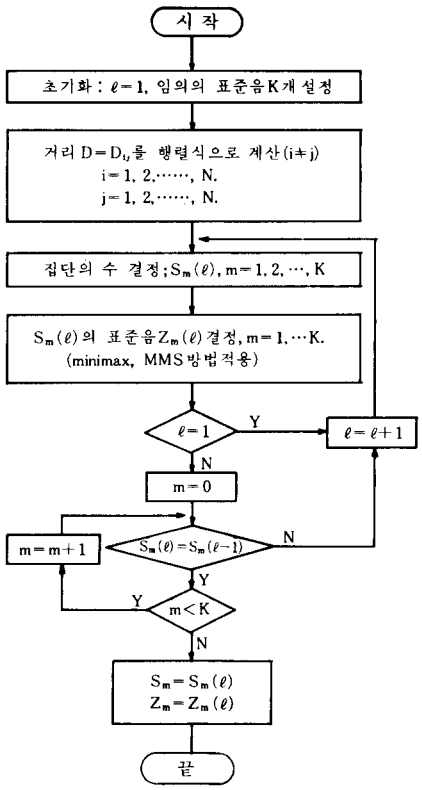

그림 2. 집단화 흐름도
Fig. 2. Flow diagram of the clustering procedure.

첫째 N개의 토큰들의 集合에서 임의로 k개의 토큰 $Z_1(1), Z_2(1), Z_3(1), \cdots, Z_k(1)$를 선택하여 이것을 標準音으로 초기화한다.

둘째, 集團化하기 위해서 토큰들간의 거리를 계산한다.

세째, $i \neq m$인 모든 토큰들에 대해서 式(5)을 적용하여 K개의 集團에 토큰 X를 소속케한다. 이를 위하여 ℓ번 반복 수행한다.

$$X \in S_m(\ell) \text{ if } |X - Z_m(\ell)| \leq |X - Z_i(\ell)| \quad (5)$$

여기서 $S_m(\ell)$은 標準音 $Z_m(\ell)$이 소속된 集團을 말한다.

네째, 이상과 같이 결정된 集團에서 설정해야 하는 標準音은 minimax방법과 MMS 방법에 의하여 설정하였다. Minimax 방법에 의한 標準音은

$$\max_k \{\delta(x_m, x_k) + \delta(x_k, x_m)\} \quad (6)$$

이 最小가 되는 토큰 x_m이 標準音이 되어 $Z_m(\ell) = x_m$의 관계가 성립한다.

다섯째, 만약 $m = 1, 2, 3, \cdots, k$에 대하여 $Z_m(\ell) = Z_m(\ell-1)$이 될때 모든 수행이 끝나고 그렇지 않으면 다시 세째 과정부터 다시 반복 수행한다.

Ⅲ. MMS방법에 의한 標準音 設定

minimax방법은 알고리즘 자체는 단순하지만 인식율이 대체로 저조하다.

이러한 점을 개선하여 인식 효과를 높이기 위해 거리를 상호 합산하여 式(7)과 같이 表現할 수 있다.

$$Z = \hat{X} = \min_{1<i<n} \sum_{i=1}^{n} [\delta(\hat{X}, X_i) + \delta(X_i, \hat{X})] \quad (7)$$

여기서 $\delta(\hat{X}, X_i)$은 $\hat{X}$를 標準音으로 하고 X_i를 비교 대상음(test word)으로 설정했을 때 이들간의 거리를 나타낸다.

따라서 式(7)을 利用하여 토큰 $\hat{X}$을 중심으로 X_i과의 거리를 계산하고 또 X_i을 중심으로 $\hat{X}$과의 거리를 계산하여 이들 거리의 총합의 결과중에서 그 값이 가장 최소가 되는 토큰 Z을 標準音으로 설정할 수 있다.

파라미터의 차수가 높아지고 標準音과 비교 대상음들간의 프레임수가 다를때에 $\delta(\hat{X}, X_i)$와 $\delta(X_i, \hat{X})$의 값이 상당히 큰 차를 나타낼 수 있다.

따라서 이러한 차때문에 발생되는 불합리성을 MMS 관계식에 의하여 개선시킬 수 있다.

Ⅳ. 實驗結果 考察

本 硏究에서는 韓國語 地下鐵名 13개를 대상으로 하여 成人 男性 話者 2인과 女性 話者 1인이 발음한 143토큰을 集團化하여 認識하였다. 13개 地下鐵名에 대한 이들 143토큰은 男性 화자 2인이 각각 5번, 4번씩 발음하고 女性 화자 1인은 2번씩 반복 발음한 음성 데이타이다.

本 硏究에서의 標準音은 minimax방법과 MMS방법에 대하여 각각 1개, 2개, 3개의 標準音을 設定하고 이들을 중심으로 認識 실험하였다.

Minimax방법과 MMS 방법에 따라 設定한 標準音을 중심으로 認識한 결과는 상당한 차를 보였다. 標準音을 1개만 설정했을 경우와 2개, 3개를 설정했을 경우 인식 결과는 물론이고 標準音의 변화도 크다.

표 1은 2가지 방법에 의하여 설정한 標準音을 나타내었고, 표 2와 3에서는 認識결과를 나타내고 있다.

각 단어에 대하여 1개의 標準音을 設定했을 경우 minimax 방법을 사용하면 59.4%, MMS방법을 사용했을 경우는 57.3%로 인식 결과가 minimax방법이 더 높다. 그러나 각 단어에 대하여 2개의 標準音을 설정했을 경우 minimax 방법을 사용하면 80.4%, MMS 방법을 사용하면 88.1%로 MMS에 의한 인식율이 크게 향상됨을 알 수 있다.

각 단어당 3개의 표준음을 設定했을 경우는 그 인식율이 minimax방법 사용시 91.6%, MMS방법 사용시 95.8%의 인식율을 얻었다. 그러므로 MMS 방법이 標準音 설정에 더욱 효율적인 것으로 실험을 통하여 입증되었다.

Ⅴ. 結 論

사람은 누구나 저마다 고유한 음색을 가지고 있기 때문에 모든 사람의 음성을 수렴할 수 있는 標準音 設定은 참으로 중요하다.

本 硏究에서는 기존의 minimax방법과 새로 제안한 MMS방법에 의하여 標準音을 設定하였다. 그 결과 1개의 표준음을 설정한 경우 minimax가 우수하지만 2개 이상을 설정할 경우 MMS가 minimax보다 인식률이 상당히 좋다는 것을 입증할 수 있었다. 그 이유는 MMS방법을 적용하면 거리값이 유사한 토큰들이 집중된 경우 이들 중에서 標準音이 설정된다.

그러나 minimax는 거리값이 큰 토큰들까지도 수용할 수 있는 標準音을 설정하다 보면 이때문에 밀집도가 높은 유사토큰들 중에서 標準音이 설정되지 못하기 때문이다.

앞으로의 연구 과제는 보다 더 효율적인 標準音 設定 알고리즘 개발과 韓國語 특성에 맞는 특징 파라미터를 추출할 수 있는 새로운 방법 개발도 중요한 것으로 생각된다.

표 1. 標準音設定
Table 1. Creation of reference templates.

標準音의 數 / 設定方法 / 地下鐵名	3		2		1	
	minimax	MMS	minimax	MMS	minimax	MMS
이 대	W1 M24 M15	M21 M24 M12	W1 M21	M21 M12	W1	M22
이 촌	W2 M14 M21	W2 M14 M21	W1 M21	M21 M11	W1	M24
이 수	W2 M13 M23	W2 M13 M23	M11 M22	M23 M11	W2	W2
옥 수	W2 M15 M24	W2 M15 M24	M11 M22	M24 M12	W2	W2
수 유	W1 M11 M24	M21 M11 M15	W1 M24	M21 M14	W2	W2
신 촌	M13 M12 M21	M11 M13 M22	M13 M22	M22 M11	W1	M13
신 사	M11 W1 M21	M11 M24 M12	M14 W1	M14 M24	W1	M12
신 당	M11 M23 M13	M11 M23 M12	M14 W1	M11 M12	W1	M12
사 당	W1 M14 M12	M14 M12 M23	W1 M14	M11 M14	W1	M11
대 림	W1 M14 M23	M14 M12 M23	M12 M23	M12 M23	W1	M13
신 림	W1 M13 M22	M11 M12 M22	W1 M13	M24 M12	W1	M11
신 대 방	W1 M14 M13	M23 M14 M13	W1 M14	M23 M14	W1	M14
대 방	W1 M11 M13	M22 M12 M14	W1 M11	M11 M12	W1	M14

기호표시 : M_{1i}는 첫번째 남성화자가 i번째 발음한 지하철 이름
M_{2j}는 두번째 남성화자가 j번째 발음한 지하철 이름
W_k는 여성화자가 k번째 발음한 지하철이름

표 2. Minimax 방법에 의한 인식결과
(a) 1 개의 표준음 설정시 인식결과
(b) 2 개의 표준음 설정시 인식결과
(c) 3 개의 표준음 설정시 인식결과

Table 2. Recognition result by minimax technique.
(a) Recognition result on creating 1 reference template. 85/143 = 59.4%
(b) Recognition result on creating 2 reference templates. 115/143 = 80.4%
(c) Recognition result on creating 3 reference templates. 131/143 = 91.6%

입력 / 출력	이 대	이 촌	이 수	옥 수	수 유	신 촌	신 사	신 당	사 당	대 림	신 림	신대방	대 방
이 대	5	1		2			1						
이 촌		7											
이 수	2		5	1				1					
옥 수		2		8	1	2				1			2
수 유					6			1				1	
신 촌	2		2		2	8		1		1	2	2	
신 사							6	1	1				
신 당								5					
사 당					2			1	10	2	2		
대 림	2		3							6	2		
신 림							2				5		
신대방								1				5	
대 방		1	1			1	2			1		3	9

(a)

입력 / 출력	이 대	이 촌	이 수	옥 수	수 유	신 촌	신 사	사 당	신 당	대 림	신 림	신대방	대 방
이 대	7										2		
이 촌		8											
이 수	2	2	11	2									
옥 수				9	1								
수 유					8								
신 촌	1	1			2	10	1				1		
신 사							8		1				
사 당	1					1		10					
신 당									10				
대 림										8			2
신 림							2			2	8	1	
신대방								1				9	
대 방										1		1	9

(b)

출력 \ 입력	이 대	이 촌	이 수	옥 수	수 유	신 촌	신 사	신 당	사 당	대 림	신 림	신대방	대 방
이 대	10												
이 촌		11				1							
이 수			11										
옥 수				11									
수 유					9								
신 촌					2	9	2						
신 사							9	1					
신 당								10				1	
사 당									10				
대 림	1									11			2
신 림						1					11		
신대방									1			10	
대 방													9

(c)

표 3. MMS 기법에 의한 인식결과
(a) 1 개의 표준음 설정시 인식결과
(b) 2 개의 표준음 설정시 인식결과
(c) 3 개의 표준음 설정시 인식결과

Table 3. Recognition result by MMS technique.
(a) Recognition result on creating 1 reference template. 82/143 = 57.3%
(b) recognition result on creating 2 reference templates. 126/143 = 88.1%
(c) Recognition result on creating 3 reference templates. 137/143 = 95.8%

출력 \ 입력	이 대	이 촌	이 수	옥 수	수 유	신 촌	신 사	신 당	사 당	대 림	신 림	신대방	대 방
이 대	6				1					3	3	1	
이 촌		5	3		1								
이 수	1	1	5	3	1	2	2	3	1			1	
옥 수				8				1					
수 유	1				6		3	1		1	2	2	
신 촌		2				6						1	2
신 사	1	1	2			1	6		1	1			1
신 당		2	1					6					1
사 당									9				
대 림					1					6			
신 림	2				1						6		
신대방												6	
대 방													7

(a)

입력 / 출력	이 대	이 촌	이 수	옥 수	수 유	신 촌	신 사	신 당	사 당	대 림	신 림	신대방	대 방
이 대	10												
이 촌		11	1										
이 수	1		9			1							
옥 수			1	11					1				
수 유					10								
신 촌						10						1	
신 사					1		11	2	1	1	1	1	2
신 당								9					
사 당									9				
대 림										10			2
신 림											10		
신대방												9	
대 방													7

(b)

입력 / 출력	이 대	이 촌	이 수	옥 수	수 유	신 촌	신 사	신 당	사 당	대 림	신 림	신대방	대 방
이 대	11												
이 촌		11	1										
이 수			10			1							
옥 수				11									
수 유					11								
신 촌						10							
신 사							10		2				1
신 당								11					
사 당									9				
대 림										11			
신 림											11		
신대방							1					11	
대 방													10

(c)

參 考 文 獻

[1] L.R. Rabiner and R.W. Schafer, *"Digital Processing of Speech Signals,"* Prentice Hall, Inc, 1978.

[2] 김계국, 고덕영, 이종악, "한국어 단독음 인식을 위한 표준패턴설정," 한국음향학회논문집 vol. 6. no. 1. 1987. 3.

[3] 안태옥, 변용규, 김순협, "구문분석과 level building을 이용한 한국어 연속음 인식," 한국음향학회 논문집 vol. 5. no. 4, 1986. 12.
[4] R.A. Jorvis "Clustering using a similarity measure based on shared Nearest Neighbors," *IEEE*, vol. C-22, no. 11. Nov. 1973.
[5] L.R. Rabiner, "On creating reference templates for speaker-Independent recognition of lsolated word," *IEEE* vol. ASSP-26, no. 1, Feb. 1978.
[6] Helmuth Spath, Ellis Horwood Limited "Cluster Analysis algorithm for DATA reduction and classification of objects,"
[7] Fumitada Itakura "Minimum prediction residual principle applied to speech recognition," *IEEE*, vol. ASSP-23. no. 1. Feb. 1976.
[8] L.R. Rabiner and J.G. Willpon "Application of clustering techniques to speaker-traind isolated word recognition," Bell Lab. vol. 58. no. 10. Dec. 1979.
[9] Peter U. de Souza "Statistical test and distance Measures for LPC coefficients," *IEEE*, vol. ASSP-25 no. 6. Dec. 1977.
[10] Ihroaki Sakoe, Seihi Chiba. "Dynamic programming algorithm optimization for spoken word recognition," *IEEE Trans.* vol. ASSP-26. no. 1, Feb. 1978. *

著 者 紹 介

金 桂 國(正會員)
1954年 8月 24日生. 1983年 2月 원광대학교 전자공학과 졸업. 1985年 2月 숭실대학교 대학원 전자공학과 졸업. 현재 건국대학교 대학원 전자공학과 박사과정. 주관심분야는 음성인식, 마이크로파등임.

金 淳 協(正會員)
1947年 12月 28日生. 1974年 2月 울산공과대학 전기공학과 졸업(공학사). 1976年 2月 연세대학교 대학원 전자공학과 공학석사 학위취득. 1983年 2月 연세대학교 대학원 전자공학과 공학박사 학위취득. 1986年 8月～1987年 8月 The University of Texas Austin 전기 및 전자계산기 공학과 객원교수. 현재 광운대학교 전자계산기공학과 부교수. 주관심분야는 음성인식, 마이크로파 등임.

安 泰 玉(正會員)
1953年 6月 24日生. 1981月 2月 울산공과대학 재료공학과 졸업. 1987年 2月 광운대학교 전자계산기공학과 공학석사 학위취득. 현재 광운대학교 대학원 전자계산기 공학과 박사과정. 주관심분야는 음성인식, 마이크로파 등임.

李 鍾 岳(正會員)
1966年 2月 한양대학교 전기공학과 공학사. 1970年 2月 연세대학교 대학원 전기공학과 석사학위취득. 1974年 2月 연세대학교 대학원 전기공학과 공학박사 학위취득. 1974年 4月～1975年 3月 일본 경도대학 전기공학과 연구원. 1979年 8月～1980年 7月 프랑스 Lyon 제일대학 물리학과 연구원. 현재 건국대학교 전자공학과 교수. 주관심분야는 마이크로파임.