집단화를 이용한 음성의 표준패턴설정에 관한 연구

*김계국 **유형근 ***김순협 *윤재광
*숭전대 ** OPC ***광운대

A study on creating Reference Pattern of speech by using the cluster

* K. K. KIM ** H. K. RYU *** S.H. KIM * J. K. YWOON
* Soong Jun Univ. ** OPC *** Kwang Woon Univ.

요 약

불 특정 화자의 음성인식을 위해 150 숫자음에 대하여 10개의 표준패턴을 설정하는데 목적을 두고 기술했다.
남성화자 3인이 각숫자음(0-9)를 5번씩 반복 발음한 150음을 집단화하여 숫자음의 표준 패턴을 설정하였다.
특징 파라미터는 포르만트 주파수를 이용하였고 유크리드거리 측정법을 유사도 비교에 사용하였다.
실험결과 85.3%의 인식률을 얻었다.

ABSTRACT

This paper is of the pupose in the creation of ten reference patterns for 150 digits to recognize a isolated digits(0-9) which three speaker-independents pronounce.
The digit reference patterns are created from a statistical clustering analysis of speech date base bonsisted of 150 replicotions of each korean isolated digits.
They repeated five times by each of 3 male talkers.
Euclidean distance measure method is used for the comparison of pattern similarity.
The recognition accuracies for three male talkers is about 85.3 percent.

I. 서 론

과학의 성장발전으로 기술혁명의 산업사회에서 통신매체의 역할은 지대한것으로 생각된다. 그러므로 한 국가가 부강하기 위해서는 컴퓨터 시스템에 의한 정보의 전달은 분명히 연구되어야 할 것이다. 그러한 견지에서 21세기 확률이라고 일컬을 만큼 우리의 음성 인식에대한 관심도는 참으로 높은 것으로 알고 있다.
본 연구의 목적은 한국어 음성을 인식하기위하여 여러가지 집단화 알고리즘을 성립하고 집단화에 의하여 표준패턴을 설정하고 보다높은 인식율을 얻고자 하는데 있다. 그림1 에서 음성인식의 기본 과정을 살펴보면, 우선 전처리된 음성데이타를 가지고 특징 파라미터를 추출하고 표준패턴을 설정하여 유사성을 비교하고 인식하게 된다.

본 연구에서는 포르만트 주파수를 특징 파라미터로 이용하였으며 이것은 ARMA (Autoregressive Moving Average) 모델의 POD (Pole Spectrum Density) 추정 알고리즘을 이용하여 추출한 제1, 제2, 제3 포르만트로, 기존의 음성데이타를 이용하여 표준 패턴설정을 시도하였다.

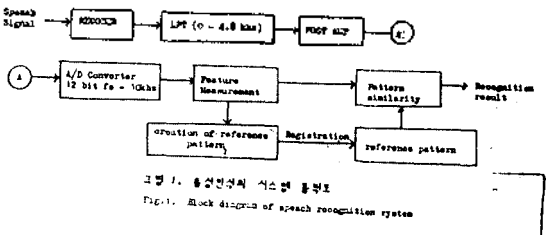

그림 1. 음성인식의 시스템 블럭도
Fig.1. Block diagram of speech recognition system

Ⅱ. 본 론

1. 표준 패턴 설정이론

본 연구에서는 한국어 단독 숫자음(0-9)을 3인의 남성화자가 각 숫자음에 대해서 3번씩 발음한 150음성을 집단화에 이용하였다.

표준 패턴은 각 집단을 대표할 수 있는 단어로서 15개의 단어들중 어떠한 단어에 대해서도 최대거리가 최소가 될수 있도록 최적 기준을 만족할 수 있는 단어가 결정되어야 한다.

그림2는 집단화의 일반적인 예를 보이고 있는데 이 그림에서 각음성 데이타의 특징파라미터집합을 X (token 이라칭함)로 나타내고 화자들간의 성대의 변화 특성 때문에 토큰들이 서로 다른 영역에 분포되어있다.

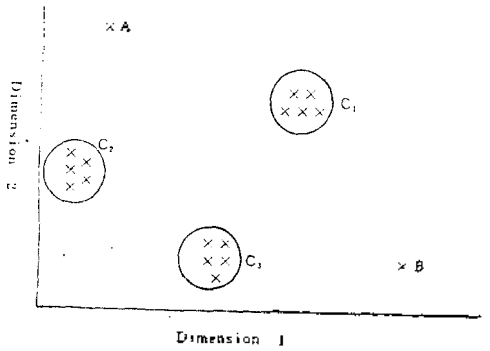

그림2. 집단화의 일반예
Fig. 2 Example showing clustering of tokens

그림2에서는 3개의 집단이 이루어져있기 때문에 적어도 3개의 표준 패턴이 설정되어야 하며 3개의 집단과 멀리떨어져 있는 두토큰A와B를 분리물이라고 하고 이 분리물은 3개의 집단중 어느것과도 한정거리를 이루지 못하고 있다.

음성데이타를 나타내는 J개의 토큰을 집합으로 표현하면 식(1)과 같다.

$$\Omega = \{x_1, x_2, \cdots x_J\} \quad (1)$$

이들 토큰들중 임의의 x_i와 x_j 토큰간의 차는 거리 d_{ij}로 측정하고 식(2)와 같이 표현된다.

$$d_{ij} = \delta(x_i, x_j) = \frac{1}{K}\sum_{k=1}^{K} d(k, \omega(k)) \quad (2)$$

여기서 K는 토큰 x_i의 프레임수를 나타내며 $d(k,\omega(k))$는 Itakura가 제안한것으로 x_i의 k번째 프레임과 x_j의 $\omega(k)$번째 프레임들간의 거리를 의미한다.

또 $\omega(k)$는 x_i와 x_j의 DTW 정합으로 얻어진 워핑함수라고 한다.

식(1)의 Ω는 M개의 집단으로 이루어진 여러번 반복발음한 똑같은 음의 단어를 나타내고있으며 집단의 수 M는 실험과정에서 적절히 고려되어야 하고 일반적으로 식(3)과 같이 나타내고 있다.

$$\Omega = \bigcup_{i=1}^{M} \omega_i \quad (3)$$

여기서 ω_i는 i번째 집단을 의미하며 ω_i에 속해있는 토큰의 갯수를 m_i로 결정하고 ω_i를 대표할수 있는 중심점 (cluster Centerpoint)을 $x_p^{(i)}$로 정의하고 $x_p^{(i)}$는 반드시 i번째 집단의 토큰들중 하나가 결정되어야 한다.

$$x_p^{(i)} \in \omega_i \quad (4)$$

표준 패턴설정에 이용되는 일반적인 집단화 알고리즘에는 Chainmap, SNN (Shared nearest neigbour), K-means iteration, ISODATA (Iterative Self Organizing DATA Analysis Technique A) 등이 있다.

본 연구에 사용된 알고리즘 흐름도는 그림3과 같다.
집단화에 의한 표준 패턴 설정과정을 살펴보면 처음에 동일음의 모든 토큰들을 집단화에 반복 수행하였다.
동일음 15개의 토큰을 평균하여 평균 길이에 가장 가까운 토큰을 첫번째 표준 패턴으로 선택하였다.
다시 결정 되어야 하는 표준 패턴은 타임 워핑된 포르만트 주파수를 평균하여 얻었으며 이 표준 패턴과 앞에서 선택된 표준 패턴과의 거리를 계산하고 계산된 거리가 임계치 이하가 될 경우 최종 표준 패턴과 표준 패턴에 이용된 모든 토큰들과의 거리를 계산하여 이들의 각 거리가 집단의 임계치를 초과한 토큰들은 첫번째 집단에서 제외 시키고 임계치를 초과하지 않는 토큰들만을 대상으로 집단화 하고 표준 패턴을 설정하였다.

첫번째 집단화에서 제외된 토큰들은 두번째 집단화에 반복 이용 하였으며 이러한 작업은 동일음 15개의 모든 토큰들이 몇개의 집단에 완전히 소속될때 까지 수행시킬수 있다.
일단, 동일음 15개 토큰에 대하여 확고한 표준 패턴들이 완전히 설정 되면 그 다음 처리되지 못한 다른음의 단어 들에 대해서 표준 패턴 설정을 시도 하였다.
본 연구에서는 동일음 15개 단어에 대하여 단 하나의 표준 패턴만을 이용하였고 하나의 집단에 소속되지 못한 음성들은 분리음로 처리하였다.

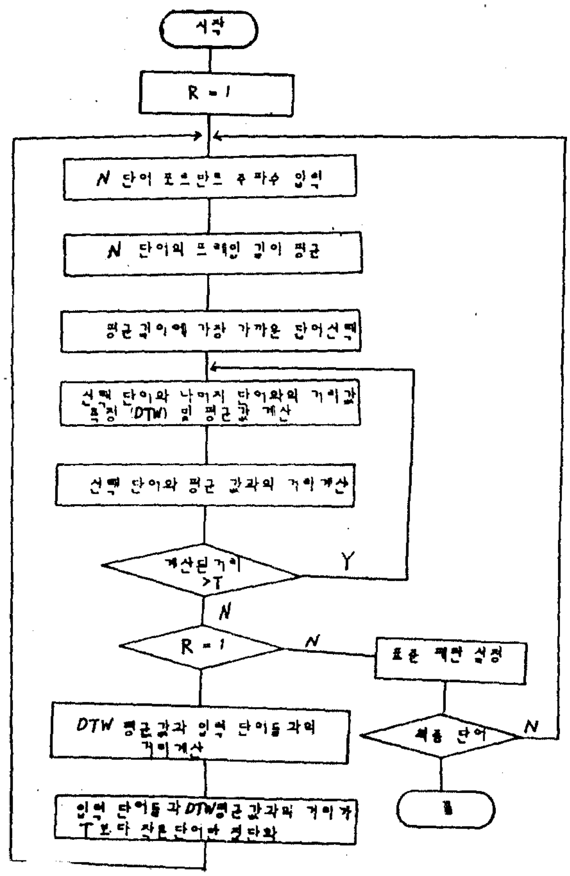

그림 3. 집단화에 의한 표준패턴 설정에 이용된 알고리듬의 흐름도
Fig. 3. Flow chart of algorithm used to reference pattern cre tion into clusters

2. DTW알고리즘 원리

인간의 음성은 피치 검출, 에너지 측정, 그리고 선형 예측 계수, 포르만트 주파수 분석등의 특징 추출로 특징 벡터를 표현할수 있다.

포준 패턴과 시험 패턴 의 특징 벡터는 식(8), (9)와 같다.

$$R(m) = [R_1, R_2, \cdots\cdots, R_j \cdots\cdots, R_N] \quad (8)$$

$$T(n) = [T_1, T_2, \cdots\cdots, T_i \cdots\cdots, T_n] \quad (9)$$

본 연구에서는 선형 예측법에 의한MA형으로 추출한 제1, 제2, 제3 포르만트 주파수를 포준패턴과 시험 패턴의 특징 벡터로 이용하였다.

1) 거리 측정법

시험패턴과 포준패턴간의 두 특징 벡터의 차는 두 벡터간의 거리 측정으로 구할 수 있다.

$$d(c) = d(i(k), j(k)) = \| T_i - R_j \| \quad (10)$$

최적 워핑 경로를 구하기 위하여, 이용되는 전체거리는 다음과 같다.

$$D(i(k), j(k)) = \frac{\sum_{k=1}^{K} d(i(k), j(k)) \cdot w(k)}{N(w)} \quad (11)$$

$w(k)$는 k번째 소구간 경로의 웨이팅 함수를 뜻하며 $N(w)$는 웨이팅 함수(w)의 함수인 정규화 요소이며 최적 경로는 전체거리를 최소로 하는 경우로서 식(12)와 같이 구할 수 있다.

$$D_T = \min_{(k), i(k), j(k))} [D(i(k), j(k))] \quad (12)$$

본 연구에서는f_1, f_2, f_3의 포르만트를 특징 벡터로 이용하고 식(13)의 Euclidean 거리 측정법을 도입하였다.

$$d = \left[\sum_{p=1}^{3} \left(\frac{f_{i,p}^{(T)} - f_{i,p}^{(R)}}{\Pi_K} \right)^2 \right]^{\frac{1}{2}} \quad (13)$$

단, $\Pi_K = \begin{cases} 500 ; & \text{for } K=1 \\ 1500 ; & \text{for } K=2 \\ 2500 ; & \text{for } K=3 \end{cases}$

$f_{i,p}^{(T)}$는 시험 패턴의 특징 벡터이며 $f_{j,p}^{(R)}$는 포준 패턴의 특징 벡터를 말하며, p는 3개의 포르만트를 말한다.

2) 실험 결과

본 연구에서는 적어도 두개 이상의 포준을 하나의 집단으로 처리하기로 하고 실험 하였다.

표 같은 반복음 15 단어에 대하여 단 한개의 포준 패턴만 설정 해주고 인식에 사용 하였으며 다른 포준들과 함류하지 못하여 어느 집단에도 포함되지 못한 단일 포준은 분리율로 처리하였다.

표 1은 본 연구에 사용된 150 숫자음을 집단화한 결과를 제시하고있다. 일반적으로 분리율이 아닌 모든 포준들은 하나의 집단으로 묶을 수 있었으며 제일 큰 집단에는 포준 15개중 15포준이 모두 속해 있음을 알 수 있었다. 앞으로는 집단화 임계치에 대한 연구가 계속되어 좀 더 바람직한 집단화가 이루어져야 할것으로 생각된다. 실험에 이용한 150단어 중 22개의 단어가 잘못 인식되어 85.3%의 인식 결과를 얻었다.

표 1, 집단화 결과

숫자음	집단수	분리율수	포준수
영	1	1	15
일	1		15
이	1		15
삼	1	2	15
사	1		15
오	1		15
육	1		15
칠	1		15
팔	1		15
구	1	1	15

표2, 인식 결과

Spoken word \ Recognized word	영	일	이	삼	사	오	육	칠	팔	구
영	10				2			1	2	
일		15								
이	1	14								
삼				11				1	3	
사					15					
오			1			12	1			1
육	1	1	1				12			
칠		2						13		
팔									15	
구						2	1	1		11

Ⅲ. 결 론

본 연구에 제안된 알고리즘을 가지고 보존 패턴 설정에 이용한 결과는 85.3%의 인식률을 얻었다.

본 논문을 쓰면서 이를 했던 것은 한국어 음성 인식을 위해서 보존 패턴 설정에 관한 분석 자료가 없었던 것이 큰 어려움이었다. 인식 결과 22개 음성이 잘못 인식 되었으며, 이것은 150 단어에 대하여 10개의 보존 패턴만을 설정하고 집단화 한데 이유가 있다. 150당어에 대해서 20개 정도의 보존패턴을 이용하고, 집단을 적절히 분리하여 패턴을 설정하였다면 더욱 높은 인식효과를 얻을 수 있을 것으로 생각된다.

후에 앞으로는 집단화에 포함되지 못한 분리율의 처리 방안이 강구되어야 할것이며 본 연구에서는 단독 숫자음의 보존 패턴 설정에 그쳤으나 앞으로 어려가지 단어를 인식하기 위해서 한국어 특성에 맞는 효과적인 패턴 설정 알고리즘이 개발되어야 할 것이다. 그러나 여기에 제시된 알고리즘도 잘 이용만 된다면 화자수와 어휘에 관계없이 음성 인식에 기여할수 있을 것으로 본다.

참 고 문 헌

1. Warren l.g.kootz." A banch and bound clustering Algorithm." IEEE Transaction computers 1975.9.
2. Helmuth spath, Ellis Horwood Limited. "Cluster Analisis Algorithms for Data reduction and classification of objects".
3. 김순협, "한국어 음성의 분석과 자동인식에 관한 연구" 연세대학원 박사학위논문 1982. 12.
4. Ihroaki Sakoe. " Dynamic Programming Algorithm optimization for spoken word recognition". IEEE Trans. Vol. Assp-26 No.1, 1978.2.
5. L.R.Rabiner. " Speaker-Independent Recognition of Isolated Words using Clustering Techniques", IEEE Vol.ASSP-27,NO.4, 1979.8
6. R.a.Jarvis. "Clustering using a similarity Measure Based on Shared Nearest Neighbors", IEEE Vol c-22.No.11. 1973.11.
7. L.r.Rabiner. " On creating Reference templates for speaker independent recognition of isolated word", IEEE,Vol.ASSP-26, No.1,1978.2.