K-L 전개를 이용한 연속 숫자음 인식에 관한 연구
A Study on Connected Digits Recognition Using the K-L Expansion

김 주 곤, 오 세 진, 황 철 준, 김 범 국, 정 현 열
Joo-Gon Kim, Se-Jin Oh, Chul-Joon Hwang, Bum-Koog Kim, Hyun-Yeol Chung

요 약

K-L 전개 방법은 특징의 차원을 효과적으로 압축하므로 인식 처리에서 계산량을 줄일 수 있는 방법으로 잘 알려져 있다. 본 논문에서는 한국어 인식 시스템의 인식 정도를 개선하기 위해, 음성의 특징 파라미터에 대하여 효과적으로 K-L전개를 적용하는 방법(K-L 계수)을 제안한다. 그리고 제안한 방법으로 얻어진 새로운 음성 특징 파라미터를 이용하여 화자 독립 연속 숫자음 인식실험을 수행하고, 기존의 Mel-cepstrum과 회귀계수의 인식 결과와 비교, 분석하였다. 인식 실험 결과, 제안한 K-L 계수를 이용한 방법이 기존의 방법보다 높은 인식률을 얻어 제안한 방법의 유효성을 확인할 수 있었다.

ABSTRACT

The K-L expansion is a method for compressing dimensions of features and thus reduces computational cost in recognition process. Also This is well known that features can be extracted without much loss of information in the statistical pattern recognition. In this paper, the method that effectively applies K-L(Karhunen-Loeve) expansion to feature parameters of speech is proposed to improve the recognition accuracy of the Korean speech recognition system. The recognition performance of a novel feature parameters obtained by the proposed method(K-L coefficients) is compared with those of conventional Mel-cepstrum and regressive coefficients through speaker independent connected digits recognition experiments. Experimental results showed that average recognition rates using the K-L coefficients with regression coefficients obtained higher accuracy than conventional Mel-cepstrum with their regression coefficients.

Keywords : speech recognition, K-L Expansion, feature parameters

I. 서 론

인간의 의사전달은 주로 음성을 통하여 이루어지고 있다. 음성에 의해 표현되는 말은 인간과 인간사이의 의사소통의 수단으로서 뿐만 아니라 논리적으로 사물을 생각하는 경우에 있어서도 중요한 역할을 한다. 이 음성을 통하여 인간과 기계와의 통신, 즉 음성으로 기계의 여러 가지 제어 장치들을 다룰 수 있게 하기 위해서는 선행되어야 될 것이 음성인식에 관한 연구이다.

음성인식 기술은 1960년대부터 기초적 연구가 수행되어 현재까지 계속 발전되어 오고 있다. 현재 음성인식 시스템은 마이크를 통한 음성입력을 전처리 단계를 거쳐 특징 파라미터로 변환된 후 미리 만들어 둔 표준 패턴과의 정합을 통해 인식이 이루어지는 패턴인식의 한 분야이다[2]. 음성인식에서 가장 기본이 되는 것이 음성의 특징 추출 부분이라 할 수 있으며, 서로 다른 음성 data를 잘 구분해 주는 특징 파라미터는 음성인식 시스템의 인식률과 매우 밀접한 관계에 있다. 따라서 특징 파라미터에 관한 심도 깊은 연구가 필요하다. 음성의 특징 파라미터로는 에너지, ZCR(Zero-Crossing Rate), pitch period, formant, short-time spectrum, filter-bank 출력, LPC(Linear Predictive Coding) 계수, cepstrum 계수 등이 있으며, 이들의 개선된 형태의 파라미터들과 새롭게 제안된 많은 파라미터들이 있다.

본 논문에서는 인식 시스템의 성능 향상을 위해서, 음성의 특징 파라미터에 대하여 효과적으로 K-L(Karhunen-Loeve)전개를 적용하는 방법(K-L 계수)을 제안한다. 제안된 K-L 계수는 음소의 시간방향의 정보를 포함하고

있는 동적 특징 파라미터로 사용된다.

국내의 경우 단어와 연속음성을 대상으로 한 인식시스템에 있어서는 비교적 높은 인식률을 얻고 있으나 한국어 특성을 고려할 때 연속 숫자음의 경우에 있어서는 아직까지 인식률이 비교적 저조한 실정이며 보다 많은 연구가 요구된다.

본 논문에서는 음성인식에 있어서 일반적으로 인식률이 저조한 연속 숫자음의 인식률 향상을 위해서 4연속 숫자음을 대상으로 인식 실험을 수행하여 제안한 K-L 계수의 유효성을 확인하고자 한다. 이를 위하여 음성자료는 국어공학연구소(KLE)에서 채록한 4연속 숫자음을 사용하고, 인식의 기본 단위로는 48개의 유사음소단위(PLUs)를 음소모델로 사용하며, 연속 숫자음 인식을 위해서는 유한상태 오토마타(FSA)에 의한 구문제어를 통한 OPDP법[5]을 이용한다.

본 논문의 구성은 다음과 같다. Ⅱ장에서 기존의 음성 특징 파라미터에 대해 기술하고, Ⅲ장에서는 본 논문에서 제안한 K-L 전개에 의한 동적 계수 추출 방법에 대해 설명한다. Ⅳ장에서 음성자료의 분석 및 인식 방법에 대하여 기술하고, Ⅴ장에서는 제안한 K-L 계수를 이용한 화자 독립 4연속 숫자음 인식실험을 실시한 후 그 결과를 검토한다. 마지막으로 Ⅵ장에서 결론과 음성인식 기술의 실용화를 위한 향후 연구 방향에 대해 서술한다.

Ⅱ. 음성의 특징 파라미터

음성인식 시스템의 입력은 인간이 발성한 음성이 마이크 등을 통하여 전기적 신호로 바뀌어진 연속파형이다. 이러한 연속파형을 음성인식 시스템에 사용하기 위해서는 적합하고 유용한 형태로 바꾸어 주는 전처리 과정(Preprocessing)을 거치게 된다. 음성신호를 LPF(Low-Pass Filtering), sampling, A/D 변환 등의 처리를 수행하고 음성신호를 묵음 구간으로부터 분리해 내는 음성구간 검출과정 등이 전처리 과정에 속한다[6]. 음성구간 검출에 의해 얻은 음성신호를 적합한 파라미터로 변화시키는 과정을 특징 추출과정(Feature extraction)이라 한다.

현재까지 사용되고 있는 특징 파라미터들을 살펴보면 1960년대 초기에는 대역필터뱅크(Band Pass Filter Bank)를 사용하였으나 스펙트럼 포락을 잘 표현하지 못하는 단점을 가지고 있으며 All Pole Model로 표현하여 최적 모델파라미터를 비교적 적은 계산량으로 안정하게 추출하는 선형예측분석(Linear Predictive Coding; LPC)법이 많이 사용되었다.

이후 스펙트럼 포락과 피치(Pitch; 기본주파수)를 분리하여 추출하는 켑스트럼(Cepstrum)법이 개발되었고 현재에는 인간의 청각 특성을 고려한 멜-켑스트럼과 동적 특징을 표현한 회귀계수(Regressive Coefficients)가 널리 사용되고 있다. 다음은 이들 특징 파라미터의 추출 방법에 대해 간략히 기술한다.

2.1 Mel-Cepstrum 계수 추출

인간의 청각능력은 음의 크기에 대하여 근사적으로 대수적인 특성을 나타내며, 주파수 분해능은 1kHz이하의 낮은 주파수영역에서는 선형적이고 그 이상의 주파수영역에서는 대수적인 멜 척도(Mel-scale) 특성을 가진다. 앞서 구한 선형예측계수를 멜 척도에 의해 비선형변환을 수행한 것을 멜-켑스트럼이라고 부르고 음성인식에 유효하다고 널리 알려져 있다.

멜-켑스트럼 계수 $\{Mc_m\}$는 LPC 켑스트럼 계수 $\{c_m\}$으로부터 근사적으로 식 (1)과 같은 전역통과필터의 위상특성을 이용하여 Bilinear 주파수변환으로 나타낼 수 있는 데 식 (2)와 같은 근사식으로부터 도출이 가능하다.

$$H_{BT}(Z)=\frac{(Z^{-1}-\alpha)}{(1-\alpha Z^{-1})} \qquad 0<\alpha<1 \tag{1}$$

$$Mc_k(m) = \tag{2}$$

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

$$m=\cdots,-2,-1,0$$

파라미터 α가 양수이면 낮은 주파수에 대한 해상도(Resolution)를 더 높일 수 있고, $0.4<\alpha<0.8$ 이면 멜 척도는 Bark scale로 변환할 수 있다.

여기서 샘플링주파수가 6.67kHz, 8kHz, 10kHz, 16kHz일 경우, α를 각각 0.28, 0.31, 0.35, 0.45로 두면, 쉽게 멜-켑스트럼을 근사적으로 구할 수 있다[5].

2.2 회귀 계수 추출

음성을 스펙트럼 영역에서 분석할 때 스펙트럼 내에서의 순시적인 변화는 음성의 중요한 정보를 가지고 있으며 스펙트럼 기울기의 변화는 스펙트럼 정보를 읽는데 중요한 단서가 된다. 일반적으로 화자가 달라지면 포만트의 절대적 위치는 변화하지만 포만트의 기울기는 상대적으로 변화하지 않는다. 이러한 특징은 인식률에 큰 영향을 미치기 때문에 음성의 특징으로서 스펙트럼 내의 순시적인 변화를 나타내는 동적 특징 파라미터로서 회귀계수가 사용된다. 회귀계수 추출은 음성의 정적 특징량 벡터의 각 차원에 대해 식 (3)을 이용하여 계산하고 이를

또 다른 특징 파라미터로 사용한다. 회귀계수 $R_m(t)$는 시간 t를 중심으로 $2\delta+1$ 폭 만큼의 단위로 계산하여 구해진다[5].

$$R_m(t) = \frac{\sum_{n=-\delta}^{\delta} nC_m(t+n)}{\sum_{n=-\delta}^{\delta} n^2} \tag{3}$$

여기서 $C_m(t)$는 t 번째 프레임의 m번째 정적 파라미터의 계수값이고 $R_m(t)$는 여기에 해당하는 회귀계수 값을 의미한다.

III. 특징 파라미터에 대한 K-L 전개

3.1 K-L(Karhunen-Loeve) 전개

음성의 특징 파라미터의 차원수가 높아질수록 계산량도 따라서 증가하게 된다. 이러한 특징 파라미터의 차원수를 효과적으로 줄이면서 인식 성능을 저하시키지 않는 방법으로 K-L 전개가 있다. 이 방법은 여러 개의 양적 변수들 사이의 관계를 분석하여 이 변수들의 선형결합으로 표시되는 주성분을 찾아내고 그 중에서 중요한 몇 개의 주성분으로 전체의 변동을 표현하고자 하는 다변량 분석법으로서, 자료의 요약이나 선형관계식을 통하여 차원을 감소시켜 분석을 용이하게 하는데 그 목적이 있다. 즉, 다차원 공간에 대하여 관측 벡터 분포의 불균일성을 이용하고 통계적으로 최적인 차원으로 감소시키는 방법이다[3].

따라서 본 논문에서는 이러한 K-L 전개의 특징을 이용하여 각 음소의 시간방향 정보에 대한 동적 특징을 추출하여 이를 특징파라미터로 사용한다. 이하에 K-L 전개에 대해 간략하게 서술한다.

n차원벡터 $X=[x_1, x_2, \cdots, x_i]$, $x_i=[x_i^1, x_i^2, \cdots, x_i^n]^T$의 공분산 행렬을 $S=[s_{jk}]$라 하면 다음과 같이 나타낼 수 있다.

$$s_{jk} = \frac{1}{I} \sum_{i=1}^{I} (x_i^j - \bar{x}^j)(x_i^k - \bar{x}^k) \tag{4}$$

여기서, x_i는 i번째 관측벡터, x_i^j는 x_i의 j차원의 관측벡터이고, $\bar{x}^{[j,k]}$는 전체 관측벡터 중에서 j,k차원까지의 평균벡터, I는 전체의 샘플 수이다.

따라서 전체 공분산 행렬 S는 다음과 같다.

$$S = \begin{bmatrix} s_{11} & s_{12} & \cdots & s_{1n} \\ s_{21} & s_{22} & \cdots & s_{2n} \\ & \cdots & \cdots & \\ s_{n1} & s_{n2} & \cdots & s_{nn} \end{bmatrix} \tag{5}$$

위의 전체 공분산 행렬에서 고유치 $\{\lambda_j\}$와 고유벡터 a_j를 다음의 고유치 문제를 풀어서 구할 수 있다.

$$\begin{aligned} S \cdot a_j - \lambda_j \cdot a_j &= 0 \\ S \cdot a_j &= \lambda_j \cdot a_j \end{aligned} \tag{6}$$

고유치 $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_m \geq \ldots \lambda_n \geq 0$에 대응하는 고유벡터 $[a_n]$중에서 관측벡터 X의 분산을 잘 나타낼 수 있는 m개까지의 고유벡터($a_1, a_2, \ldots, a_m$)에 의해 고유벡터 A를 구성하면 다음과 같다.

$$A=[a_1, a_2, \ldots, a_m]^T \tag{7}$$

이것은 n차원 공간 속 점들의 분산을 분석하여 서로 직교하는 주축으로 변환하고, 분산이 큰 축의 성분부터 차례로 나열해서 m차원으로 자른 것이 된다. 이 고유벡터 A를 사용하여 압축후의 특징 파라미터를 계산한다.

$$x_i' = A \cdot x_i \tag{8}$$

위의 식(8)을 사용하여 음성의 일반적인 특징 파라미터인 멜-켑스트럼의 차원을 압축하여 사용한다.

3.2 K-L 계수 추출

음성 특징파라미터로 많이 사용되고 있는 멜-켑스트럼은 정적 파라미터로서 음성의 동적 정보를 포함하지 못하는 단점이 있다. 멜-켑스트럼은 음성 파형의 매우 짧은 구간에 대하여 1프레임씩 추출하기 때문에 음성 파형의 보다 넓은 구간에 대한 정보가 미흡하다. 이를 보상하기 위하여 회귀계수는 2.2절에서 설명한 식(3)을 이용하여 멜-켑스트럼의 여러 프레임에 대한 음성 파형의 동적 흐름에 대한 특징을 추출한다. 회귀계수의 1프레임은 식(3)의 δ를 4로 두면, 결과적으로 멜-켑스트럼의 $2\delta+1=9$프레임의 정보를 얻을 수 있다. 따라서 회귀계수는 음성 파형의 동적 특징을 잘 나타내는 파라미터로 널리 사용되고 있다.

이 절에서는 음성의 동적 특징을 추출하기 위해 멜-켑스트럼의 여러 프레임에 대해 K-L 전개를 효과적으로 적용하여 새로운 파라미터를 추출하는 방법에 대해 기술한다. 음성 파라미터에 대한 K-L 전개는 그림 1과 같이

10차의 멜-켑스트럼을 4프레임 구간을 한 단위로 하여 1 프레임씩 시점을 이동시키면서 K-L 계수를 추출한다. 그러나 이렇게 되면 한 프레임의 차원은 40차원이 되어 계산량이 증가하게 되고, 같은 차원의 개수가 4개가 되어 다변량의 문제가 발생한다. 이를 해결하는 방법으로 3.1절에서 설명한 K-L 전개를 사용한다[1,4].

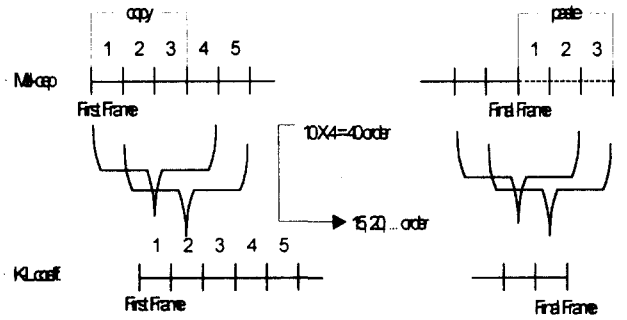


그림 1. K-L 전개를 이용한 동적 특징 추출
Fig 1. Dynamic feature extraction using K-L expansion

그림 1에서, 멜-켑스트럼의 마지막 프레임에 대해 4프레임 폭의 K-L 계수를 추출하기 위해 앞의 3프레임을 마지막 프레임에 추가하였다. 따라서, K-L 계수의 1프레임은 멜-켑스트럼의 4프레임에 대한 정보를 가지므로 음성 파형의 동적 정보를 포함할 수가 있다. 본 연구실에서는 K-L 계수를 추출할 때, 멜-켑스트럼에 대한 프레임 폭과 마지막 프레임에서 4프레임 폭을 만들 때 어떤 부분을 복사하는 가에 대한 다양한 예비 실험을 수행하였다. 그림 1은 예비 실험을 통한 최적의 방법을 나타내었다.

본 논문에서는 멜-켑스트럼과 동적 특징으로 많이 이용하고 있는 회귀계수를 결합한 경우에 대해서 K-L 전개를 수행한 후 15, 20, 25차등의 K-L 계수를 추출하여 특징파라미터로 사용하는 방법을 제안한다. 이는 동적 정보를 포함한 회귀계수에 대해 K-L 전개를 수행하면 보다 많은 동적 정보를 가진 새로운 파라미터를 추출할 수 있기 때문이다. 회귀계수만 가지고 K-L 전개를 적용하여 예비 실험을 수행했을 때, 회귀계수가 음성 파형의 본래의 성질을 많이 포함하지 않기 때문에 오히려 인식률의 저하를 가져왔다. 그리고 멜-켑스트럼과 이를 K-L 전개를 수행한 K-L 계수를 이용한 예비 실험에서도 멜-켑스트럼의 고유성분을 K-L 계수가 가지고 있기 때문에 인식률은 많이 향상되지 않았다. 따라서 멜-켑스트럼과 회귀계수를 결합한 경우에 대해서 인식률 변화를 고려하였다. 다양한 예비 실험을 통하여 너무 많은 동적 정보는 오히려 인식률을 저하시키고 K-L 계수의 차원 수를 증가시켜 인식기의 성능을 저하시켰다.

따라서, 본 논문에서는 실험을 통하여 그림 2와 같이 2프레임의 구간에 대해서 1프레임씩 시점을 이동시키면서 K-L 전개를 수행한다. 여기서 멜-켑스트럼의 값들은 회귀계수의 값들보다 큰 값으로 구성되어있어 멜-켑스트럼의 마지막 차원 뒤에 회귀계수를 추가하여 총 20차원의 특징 파라미터로 구성하고 이를 K-L 전개를 통하여 K-L 계수를 추출한다.

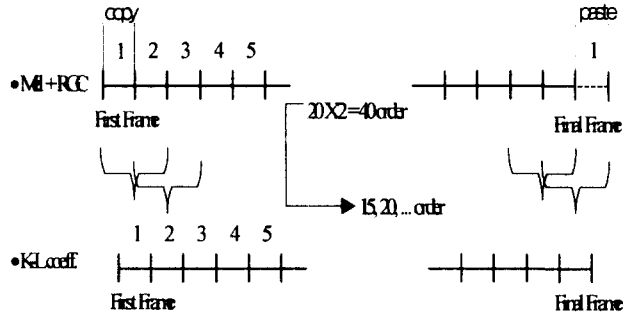


그림 2. 제안한 K-L 계수의 추출 방법
Fig 2. The proposed method of K-L coefficient extraction

그림 2와 같이 추출한 K-L 계수는 멜-켑스트럼과 회귀계수의 고유성분을 모두 포함하고 있고, 보다 넓은 음성 파형의 동적 성분을 표현하면서 1프레임의 차원 수를 줄일 수가 있다.

Ⅳ. 음성자료의 분석 및 인식 방법

4.1 음성자료의 분석

연속 숫자음의 인식 실험을 위한 음성자료는 국어공학센터(KLE)에서 구축한 한국인 남·여 72인의 4회 발성한 4연속 숫자음 중에서 남성 20인이 발성한 4연속 숫자음을 모델학습에 사용하고, 학습에 참여하지 않은 남성 10인의 화자가 발성한 4연속 숫자음을 평가용 자료로 사용한다.

표 1. 음성자료의 분석조건
Table 1. Analysis conditions of speech DB

Speech Data	KLE 4연속 숫자음
Sampling frequency	16kHz
Filtering	LPF, 7kHz
Resolution	16bits
Hamming window	16ms (256points)
Frame rate	5ms (80points)
Analysis	14order LPC analysis
Static Feature parameters	10order Mel-Cep. coeff.
Dynamic Feature parameters	10order Regressive coeff. 10, 20order K-L coeff.

음성자료의 분석은 표 1과 같이 음성 데이터를 7kHz의 LPF를 통과시킨 후 샘플링 주파수 16kHz, 양자화 정도 16Bits A/D 변환기를 통해 이산데이터로 변환되고 Preemphasis 필터를 통과한 후 16ms(256 points) 길이의 해밍 윈도우를 사용하여 5ms(80points)씩 쉬프트 시키면서 분석된다. 이로부터 14차 LPC 켑스트럼 계수를 구하고, 10차의 LPC 멜-켑스트럼을 구하여 정적 특징파라미터로 사용한다. 또한 이로부터 10차의 회귀계수와 10, 20차 등의 K-L 계수를 동적 특징파라미터로 사용한다. 그리고, 제안된 멜-켑스트럼과 회귀계수를 결합한 경우에 대하여 15, 20차 등의 K-L 계수를 추출하여 4연속 숫자음 인식에 이용한다. 그림 3은 음성 특징 파라미터의 추출 과정을 나타낸다.

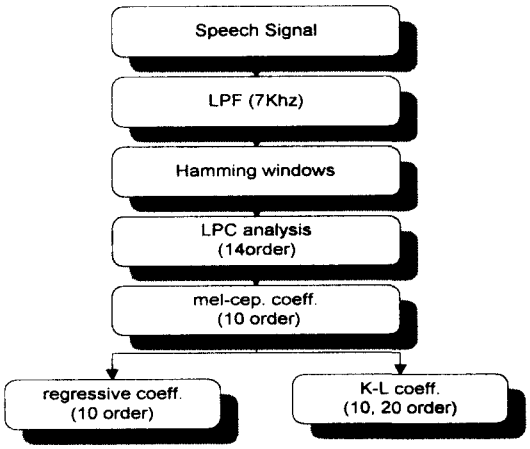


그림 3. 음성 특징 파라미터의 추출
Fig 3. Extraction of speech feature parameters

4.2 음소 모델

HMM은 출력확률의 분포에 따라 크게 이산분포 HMM(Discrete HMM)과 연속분포 HMM(Continuous HMM)으로 분류한다.

DHMM에서는 추출된 음성 특징 파라미터들의 출력확률분포가 벡터양자화에 의해 코드북내의 코드워드로 매핑되므로 벡터 양자화에 따르는 양자화 오차가 발생한다. 그러나, CHMM에서는 출력확률분포를 Gauss분포나 Cauchy분포로 직접 모델링 함으로써 양자화 오차를 막을 수 있다[5,6,7].

따라서 본 논문에서는 CHMM을 이용하여 초기 음소 모델을 작성하여 인식에 이용한다. 음소 모델을 작성하기 위해 레이블링된 음소의 길이를 조사하여 가장 짧게 나타난 파찰음과 파열음 등의 최소 4프레임을 고려하였다. 이를 토대로 HMM 모델을 4상태로 구성하고 여러 가지 천이에 대한 예비 실험을 수행한 후, 최종 CHMM 음소 모델의 구조는 그림 4와 같은 형태의 4상태 1혼합을 사용하였다.

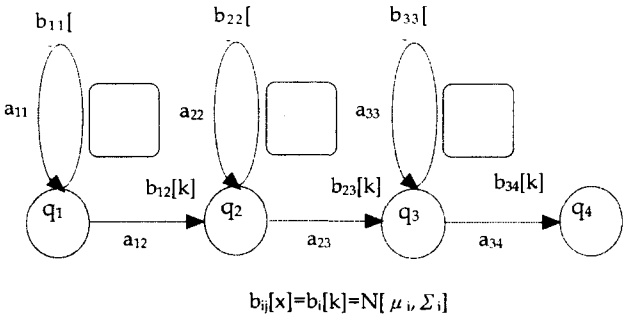


그림 4. 연속분포 HMM의 예
Fig 4. An example of CHMM

4.3 인식 시스템

인식시스템은 표준패턴을 작성하기 위한 학습 단계와 표준패턴과 입력패턴과의 유사도를 측정하여 최적의 상태열을 찾는 인식 단계로 구성되며 그림 5에 4연속 숫자음 인식을 위한 인식 시스템의 전체 구성도를 나타내었다.

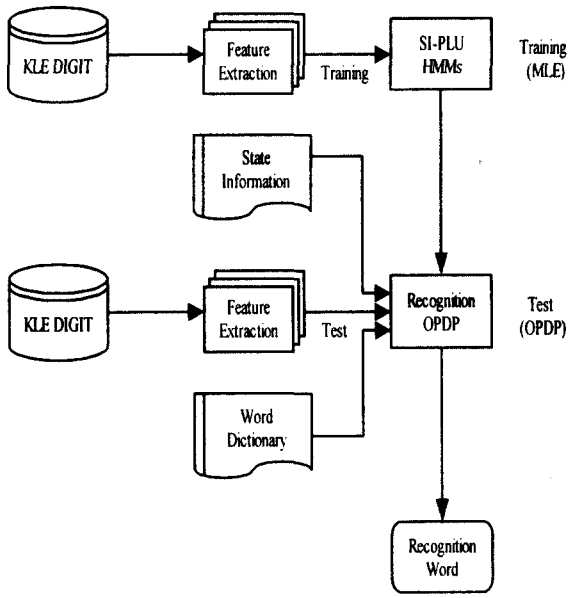


그림 5. 연속 숫자음 인식 시스템의 전체 구성도
Fig 5. Overall diagram of connected digits recognition system

이때, 학습 단계에서 CHMM을 이용하여 음소 표준패턴을 작성하고, 인식단계에서 미리 작성한 단어사전과 유한 상태 오토마타(FSA)에 의한 구문제어를 통하여 OPDP법으로 인식을 수행한다. 단어의 누적거리 계산은 입력의 첫 번째 프레임과 각 유사음소단위 표준패턴과의 Viterbi Score를 프레임에 동기 시켜 가면서 계산한 후 저장한다[5,7]. 그리고 이를 단어사전중의 각 단어의 마지막 프레임까지 확장해서 총 누적거리 값이 최소가 되는

단어의 인덱스를 인식결과로 출력된다.

4.4 숫자음 인식을 위한 FSN의 구성

표준패턴과 입력패턴 사이의 유사도를 측정하기 위한 일반적인 방법으로는 예측되어진 전체 표준패턴과 입력패턴을 정합 시키는 방법이다. 그러나 이 방법은 인식하고자 하는 카테고리가 증가하고 인식 알고리즘이 복잡해짐에 따라 많은 시간과 문법적인 제약에 영향을 받는다.

따라서 유한상태 오토마톤에 의한 구문제어를 통해 효율적으로 입력음성을 정합 시키는 방법이 널리 사용되고 있다.

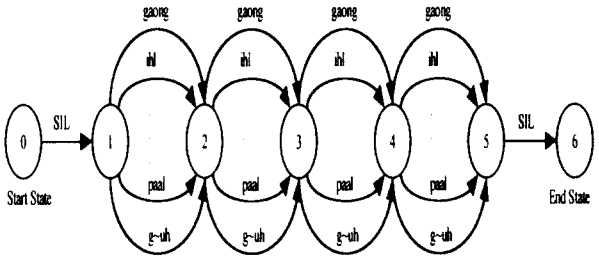


그림 6. 4연속 숫자음에 대한 FSN의 구성
Fig 6. FSN for 4 connected digits

본 논문에서는 가능한 모든 4연속 숫자음을 인식을 고려하여 그림 6과 같이 유한 상태 오토마톤을 구성하여 인식 실험을 수행한다.

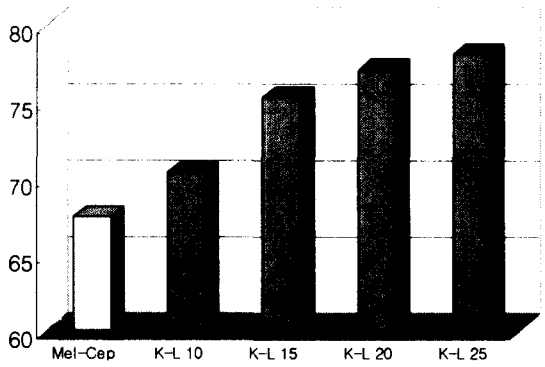
V. 인식 실험 및 고찰

연속 숫자음에 강건한 모델 구성과 동적 특징인 K-L 계수의 유효성을 확인하기 위해, 인식 실험을 위한 음성 자료로는 국어공학센터(KLE)에서 구축한 한국인 남·여 72인의 4회 발성한 4연속 숫자음 중에서 남성 20인이 발성한 4연속 숫자음으로 표준 패턴을 작성하고, 학습에 참여하지 않은 남성 10인의 화자가 발성한 4연속 숫자음을 평가용 자료로 사용하여 인식실험을 수행하였다.

5.1 K-L 계수를 이용한 인식실험

음성 특징파라미터로 많이 사용되고 있는 멜-켑스트럼과 K-L 전개를 통해 추출한 K-L 계수를 사용하여 화자 독립 4연속 숫자음 인식실험을 수행하였다. 4 연속 숫자음 인식을 위해서 그림 6의 유한상태 오토마타(FSA)에 의한 구문제어를 통한 OPDP법을 이용하여 효과적으로 인식실험을 수행한다. 앞 절에서 기술한 음소 모델과 인식 시스템으로 많은 예비 실험을 통하여 K-L 계수를 추출할 때 적절한 프레임 폭과 K-L 계수의 차수를 선정하였다. 프레임 폭은 4프레임을 기준으로 인식률의 변화가 거의 없었다.

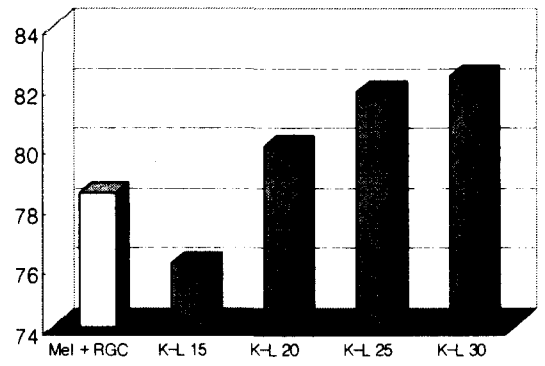
그림 7의 인식실험 결과에서, 멜-켑스트럼을 사용한 경우 67.5%의 인식률을 보였으며, 4프레임 구간에 대하여 K-L 전개를 수행한 K-L 계수의 차수를 증가할수록 인식률도 증가함을 알 수 있었다. 25차원의 K-L 계수를 사용한 경우는 78.2%의 인식률을 보여 기존의 멜-켑스트럼보다 약 8.7%의 인식률 증가를 나타내었다.



*Mel-Cep : 멜 켑스트럼, Rgc : 회귀 계수, K-L : K-L 계수
그림 7. K-L 계수를 이용한 인식실험 결과
Fig 7. Results of recognition experiment using K-L coefficients

5.2 제안된 K-L 계수를 이용한 인식실험

음성 파형의 동적 정보를 보다 많이 포함하기 위해 제안된 K-L 계수의 유효성을 확인하기 위해 멜-켑스트럼과 회귀계수를 사용한 경우와 이를 확장한 K-L 계수를 특징파라미터로 사용한 경우에 대해서 각각 인식실험을 수행하고, 그 결과를 비교하였다.



*Mel-Cep : 멜 켑스트럼, Rgc : 회귀 계수, K-L : K-L 계수
그림 8. 제안된 K-L 계수를 이용한 인식실험 결과
Fig 8. Results of recognition experiment using the proposed K-L coefficients

인식실험 결과에서, 멜-켑스트럼과 회귀계수를 결합

하여 사용한 경우 78.4%의 인식률을 보였다. 또한, 이로부터 확장한 K-L 계수를 특징파라미터로 이용한 경우에서도 차원수의 증가에 따라 인식률도 개선되었다. 30차원의 K-L 계수에 대해 82.3%의 높은 인식률을 얻었다.

이상의 결과에서, 멜-켑스트럼보다 K-L 계수를 사용한 경우가 8.7% 향상된 인식률을 보였으며, 멜-켑스트럼과 회귀계수를 사용한 경우보다 멜-켑스트럼과 회귀계수로부터 확장한 K-L 계수를 사용한 경우가 약 4%의 인식률의 향상을 보여 제안한 K-L 계수의 유효성을 확인하였다.

Ⅵ. 결 론

본 논문에서는 한국어 음성인식에서 일반적으로 인식률이 저조한 연속 숫자음의 인식률 향상을 위해서, 4연속 숫자음을 대상으로 음성 특징파라미터로 많이 사용되고 있는 멜-켑스트럼에 대하여 효과적으로 K-L 전개를 수행하고, 제안한 K-L 계수의 유효성을 확인하였다.

인식 실험 결과, 멜-켑스트럼을 사용한 경우 67.5%의 인식률을 보였으며, K-L 계수를 사용한 경우 78.2%로 8.7%의 향상된 인식률을 얻었다.

제안한 K-L 계수의 추출 방법에 대한 유효성을 확인하기 위한 실험에서, 멜-켑스트럼과 회귀계수를 사용한 경우 78.4%의 인식률을 보였으며, 이를 확장한 K-L 계수를 사용한 경우 82.3%로 약 4%의 인식률을 향상을 나타내었다.

향후 이상의 결과를 바탕으로 단어와 숫자음에 강건한 모델을 구성하고 단어 및 숫자음, 연속음성 인식에 적용하고자 한다.

※ 본 논문에서 사용한 음성데이터베이스는 국어공학센터에서 구축한 4연속 숫자음 음성데이터베이스를 사용하였다.

접수일자 : 2001. 7. 16 수정완료 : 2001. 7. 19

참고 문헌

[1] Kazumasa Yamamoto and Seiichi Nakagawa, "Comparative Evaluation of Segmental Unit Input HMM and Conditional Density HMM," pp. 1615-1618, ESCA. EUROSPEECH'95, 1994. 4.

[2] 舟久保 登 著, "パターン認識," 共立出版株式會社, 1991.

[3] 田中 豊, 脇本和昌 著, 金寬泳, 李昇洙 譯, "多變量統計解析法," 自由아카데미, 1995.

[4] 김 주성, 박 창호, 허 강인, 안 점영, "신경망의 차원 압축 능력을 이용한 음절 인식," 제8회 신호처리합동학술대회 논문집 제8권 1호, 1995.

[5] 中川 聖一, "確率モデルによる音聲認識," 電子情報通信學會編, 1989.

[6] Rabiner, Juang, "Fundamentals of Speech Recognition," Prentice-Hall International, Inc, 1993.

[7] X. D. Huang, Y. Ariki and M. A. Jack, "Hidden Markov Models for Speech Recognition," Edinburgh Univ., 1990.

김 주 곤(Joo Gon Kim)
正會員
1997년 경일대학교 전자공학과
1999년 4월-2000년 3월 일본 토요하시 기술과학대학 교환학생
2000년 영남대학교 멀티미디어통신 공학과(공학석사)
2000년-현재 영남대학교 정보통신 공학과 박사과정
관심분야 : 음성분석 및 인식

오 세 진(Se-Jin Oh)
正會員
1996년 영남대학교 전자공학과
1998년 영남대학교 전자공학과 (공학석사)
1998년 3월-현재 영남대학교 전자공학과 박사수료
관심분야 : 음성분석 및 인식, 언어처리

황 철 준(Chul-Joon Hwang)
正會員
1996년 영남대학교 전자공학과
1998년 영남대학교 전자공학과 (공학석사)
2000년 영남대학교 전자공학과 박사수료
2000년 3월-현재 대구과학대학 정보전자통신계열 전임강사
관심분야 : 음성분석 및 인식, 디지털 신호처리

김 범 국(Bum-Koog Kim)
正會員
1990년 영남대학교 수학과
1992년 영남대학교 전자공학과
(공학석사)
1998년 영남대학교 전자공학과
(공학박사)
1997년 3월-현재 대구과학대학
정보전자통신계열 조교수
관심분야 : 음성분석 및 인식, 언어처리, 멀티모달 시스템

정 현 열(Hyun-Yeol Chung)
正會員
1975년 영남대학교 전자공학과
1989년 일본 동북대학교 정보공학과
(공학박사)
1989년 3월-현재 영남대학교
전자정보공학부 교수
1992년 7월-1993년 7월 미국 CMU Robotics 연구소
객원 연구원
1994년 12월-1995년 2월 일본 토요하시기술과학대학
외국인 연구자
2000년 6월-2000년 8월 미국 Qaulcomm Inc.
수석 엔지니어
관심분야 : 음성인식, 화자인식,
음성합성 및 DSP 응용분야