숫자음 인식을 위한 K-L 동적 특징파라미터의 확장

김 주 곤*, 김 범 국**, 황 철 준*, 정 현 열*
*영남대학교 정보통신공학과
**대구과학대학 전자과

Extension of K-L Dynamic Parameter for Connected Digit Recognition

Joo Kon Kim*, Bum Koog Kim**, Cheol Jun Hwang* and Hyun Yeol Chung*
*Department of Information & Communication Eng., Yeungnam University
**Department of Electronics, Taegu Science College

E-mail : { kjk, kbg, hcj, chy }@speech.yeungnam.ac.kr

요 약

본 논문에서는 일반적으로 인식률이 저조한 연속 숫자음의 인식 정도 향상을 위해서 K-L (Karhanen-Loeve) 동적특징의 확장에 대해서 검토한다.

이 검토결과를 4연속 숫자음을 대상으로 하는 인식 실험을 수행하여 숫자음 인식에 있어서 확장된 K-L 동적특징의 유효성을 확인하고자 한다.

이를 위하여 음성자료는 국어공학센터(KLE)에서 제록한 4연속 숫자음을 사용하며, 확장한 K-L 동적 특징의 유효성을 확인하기 위해서는 단일 특징파라미터로서 멜-캡스트럼과 회귀계수, K-L 동적계수 등과 이들 특징파라미터를 결합한 경우에 대해서 특징파라미터를 확장하여 K-L 동적 특징을 추출하고, 4연속 숫자음인식 실험을 수행하였다.

이때 인식의 기본 단위로는 48개의 유사음소단위(Phoneme Like Units ; PLUs)를 음소 모델로 사용하였으며, 인식실험에 있어서는 유한 상태 오토마타(Finite State Automata ; FSA)에 의한 구문제어를 통한 OPDP (One Pass Dynamic Programming)법을 이용하였다.

인식 실험 결과, 단일 특징파라미터로서 멜-캡스트럼을 사용한 경우 67.5%, 이를 확장한 K-L 동적계수를 사용한 경우 78.2%를 보였다.

또한, 결합한 특징파라미터에 있어서는 멜-캡스트럼과 회귀계수를 사용한 경우 78.4%의 인식률을 보였으며, 이를 K-L 동적계수로 확장한 경우 82.3%의 인식률을 얻어 확장한 K-L 동적 특징파라미터의 유효성을 확인 하였다.

Ⅰ. 서 론

음성은 인간에게 가장 자연스럽고 효과적인 정보교류 수단이라고 말할 수 있으며, 이에 관한 연구는 1950년대부터 음성의 발성과 이해에 관해 많은 기초적 연구가 수행되어 최근에는 개인용 컴퓨터의 보급의 가속화, 컴퓨터에 의한 신호처리기술과 정보처리기술의 급속한 발전과 더불어 음성을 통한 인간과 기계와의 효율적인 통신을 위한 Man-machine Interface의 중요성이 강조되어 현재 국외의 경우 대어휘 연속음성인식 시스템에 대한 연구가 활발하게 수행되어 한정된 태스크를 대상으로한 실용화 시스템이 개발되고 있다.

또한, 국내의 경우 단어와 연속음성을 대상으로한 인식시스템에 있어서는 비교적 높은 인식률을 얻고 있으나 한국어 특성을 고려할 때 연속숫자음의 경우에 있어서는 아직까지 인식률이 비교적 저조한 실정이며 심도있는 연구가 요구되어지고 있다.

특히, 이러한 인식 시스템의 실용화를 위해서는 음성특징의 정확한 분석과 발성화자의 개인성, 발성의 종류, 언어의 복잡성과 환경 잡음 등의 문제점에 대한 연구와 더불어 음성의 최소 인식단위와 그 특징파라미터에 대한 연구가 이루어져야 하지만 현재 대부분의 인식 시스템에서는 이에 대한 충분한 검토 없이 정적 특징과 동적

특징을 결합하여 이용하고 있다.

따라서 본 연구에서는 음성인식에 있어서 일반적으로 인식률이 저조한 연속 숫자음의 인식률 향상을 위해서 K-L(Karhanen-Loeve) 동적특징의 확장에 대해서 검토하고 4연속 숫자음을 대상으로 인식 실험을 수행하여 확장한 K-L 동적계수의 유효성을 확인하고자 한다.

이를 위하여 음성자료는 국어공학연구소(KLE)에서 채록한 4연속 숫자음(KLE 숫자음)을 사용하였으며, K-L 동적계수의 유효성을 확인하기 위해 정적 특징으로서 멜-캡스트럼과 동적 특징으로서 K-L 동적계수 및 회귀계수를 추출한 후 연속 숫자음 인식실험을 수행한다.

이때, 인식의 기본 단위로는 48개의 유사음소단위(PLUs)를 음소모델로 사용하며, 연속 숫자음 인식을 위해서는 유한상태 오토마타(FSA)에 의한 구문제어를 통한 OPDP법[2,6,7]을 이용한다.

본 논문의 구성은 다음과 같다. Ⅱ장에서 K-L 변환법에 대해서, Ⅲ장에서 음성자료 분석 및 인식방법에 대해서 서술하고, Ⅳ장에서 동적 특징파라미터의 확장에 대해서, Ⅴ장에서는 인식 실험 및 고찰을 통하여 동적 특징의 유효성을 확인하고 마지막 Ⅵ장에서 결론을 맺는다.

Ⅱ. K-L 변환법

여러 개의 양적 변수들 사이의 관계를 분석하여 이 변수들의 선형결합으로 표시되는 주성분을 찾고 이 중에서 중요한 몇 개의 주성분으로 전체의 변동을 설명하고자 하는 다변량분석법이 K-L변환법[1]으로 자료의 요약이나 선형관계식을 통하여 차원을 감소시켜 분석을 용의 하게 하는데 목적이 있다. 즉, 다차원 공간에 대하여 관측 벡터 분포의 불 균일성을 이용하고 통계적으로 최적인 차원으로 감소시키는 방법이다.

따라서, 이러한 K-L변환의 성질을 이용하여 본 연구에서는 각 음소의 시간방향 정보에 대한 동적 특징을 추출하여 이를 특징파라미터로 사용하였다. 이하에 K-L 변환법에 대해서 간략하게 서술한다.

n 차원의 관측벡터 X의 공분산 행렬을 S 라 하면 다음과 같다.

$$S=\frac{1}{N-1}\sum_{i=1}^{N}(X_i-\overline{X})(X_i-\overline{X})^t \quad (1)$$

여기서, X_i 는 i 번째 관측벡터, $\overline{X}$는 전체 관측벡터의 평균벡터, N 는 전체의 샘플수 이다. K-L변환에서 선형변환행렬 A는 다음의 특징평가함수 $J(a)$의 최대화 조건을 만족하는 벡터 a_j $(j=1,2,3,\ldots,m)$로 구성된다.

$$J(a)=\frac{a^t\cdot S\cdot a}{a^t\cdot a} \quad (2)$$

여기서, a는 다음의 고유치 문제를 풀어서 얻어진 고유벡터로 된다.

$$S\cdot a-\lambda\cdot a=0 \quad (3)$$

고유치 $\lambda_1\geq\lambda_2\geq\ldots\geq\lambda_m\geq\ldots\lambda_n\geq0$에 대응하는 고유벡터 $a_1,a_2,\ldots,a_m$의 벡터에 의해 A를 구성하면 다음과 같다.

$$A=[a_1,a_2,\ldots,a_m] \quad (4)$$

즉, 고유벡터 간에는 직교하고 얻어진 특징 벡터의 각 요소간에는 무상관이 된다.

Ⅲ. 음성자료의 분석 및 인식 방법

3.1 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.

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 캡스트럼 계수를 구하고, 10차의 LPC 멜-캡스트럼을 구하여 정적 특징파라미터로 사용한다. 또한 이로부터 10차의 회귀계수와 10, 20차 등의 K-L 동적계수를 동적 특징파라미터로 사용한다. 그리고, 멜-캡스트럼과 회귀계수를 결합하여 하나의 특징파라미터로 사용하고, 이로부터 15, 20차 등의 K-L 동적계수를 추출하여 4연속 숫자음 인식에 이용한다.

표 1. 음성자료의 분석조건.

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.

3.2 음소 모델

HMM은 출력확률의 분포에 따라 크게 이산분포 HMM(Discrete HMM)과 연속분포 HMM(Continuous HMM)으로 분류한다. DHMM에서는 추출된 음성 특징 파라메터들의 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[4,5,6]. 따라서 본 연구에서는 CHMM을 이용하여 초기 음소모델을 작성하여 인식에 이용한다. 이때 CHMM 음소모델의 구조는 4상태 1혼합을 사용한다. 그림 1에 본 연구에서 사용한 연속분포 HMM 모델의 구성을 나타내었다.

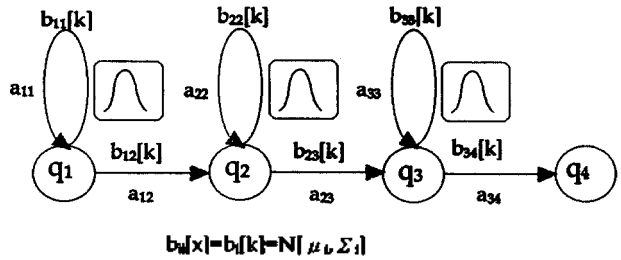


그림 1. 연속분포 HMM의 구성.(4상태 1혼합)

3.3 인식 시스템

인식시스템은 표준패턴을 작성하기 위한 학습 단계와 표준패턴과 입력패턴과의 유사도를 측정하여 최적의 상태열을 찾는 인식 단계로 구성되며 그림 2에 4연속 숫자음 인식을 위한 인식 시스템의 전체 구성도를 나타내었다.

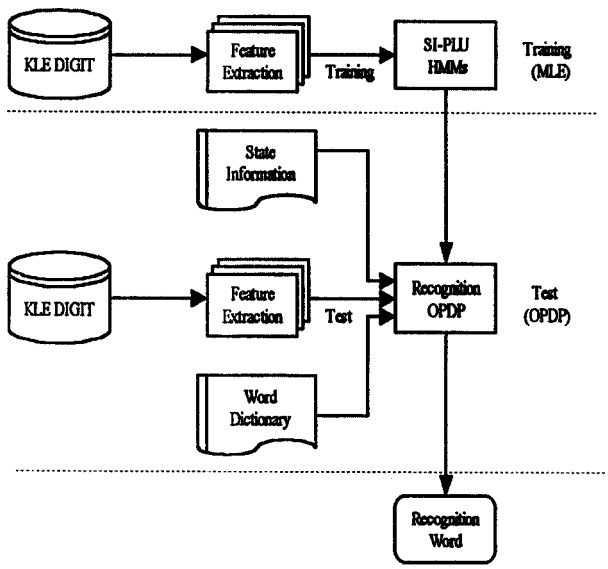


그림 2. 연속 숫자음 인식 시스템의 흐름도.

이때, 학습 단계에서 CHMM을 이용하여 음소 표준패턴을 작성하고, 인식단계에서 미리 작성한 단어사전과 유한 상태 오토마타(FSA)에 의한 구문제어를 통하여 OPDP법으로 인식을 수행한다.

3.4 숫자음 인식을 위한 FSN의 구성

표준패턴과 입력패턴 사이의 유사도를 측정하기 위한 일반적인 방법으로는 예측되어진 전체 표준패턴과 입력패턴을 정합시키는 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 정합시키는 방법이 널리 사용되고 있다. 그림 3은 4연속 숫자음에 대한 유한 상태 오토마톤의 예를 나타낸다.

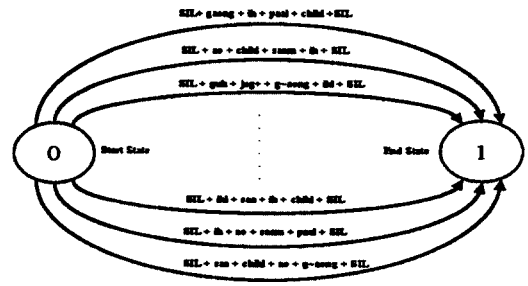


그림 3. 한정된 연속 숫자음 인식을 위한 FSN의 구성 예

그림 3의 경우에있어서는 한정된 4연속 숫자음에 대해서는 효율적이지만 가능한 모든 4연속 숫자음을 대상으로한 인식을 고려할 경우 확장성 등에 문제점이 있다.

따라서 연속 숫자음을 대상으로한 실용화 시스템을 구성하기 위해서는 그림 4와 같이 유한 상태 오토마톤을 작성하는 것이 매우 유리하다.

본 연구에서는 가능한 모든 4연 숫자음을 인식을 고려하여 그림 4와 같이 유한 상태 오토마톤을 구성하여 인식 실험을 수행한다.

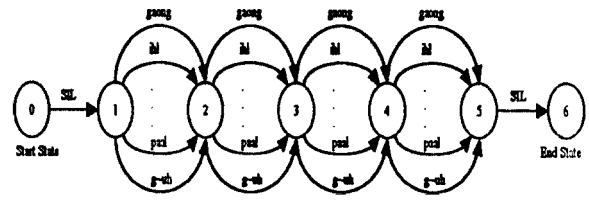


그림 4. 가능한 모든 연속 숫자음 인식을 위한 FSN의 구성 예.

Ⅳ. 동적 특징파라미터의 확장

단일 특징파라미터로 가장 많이 사용되고 있는 멜-켑스트럼은 보통 10차에서 16차원으로 추출하여 실험에 많이 이용하고 있다.

이를 본 연구에서는 그림 5과 같이 4프레임 구간에 대해서 1프레임씩 시점을 이동시키면서 K-L 전개를 수행하여 동적특징이 포함된 10, 15, 20차 등의 K-L 동적 특징 파라미터를 추출한다.

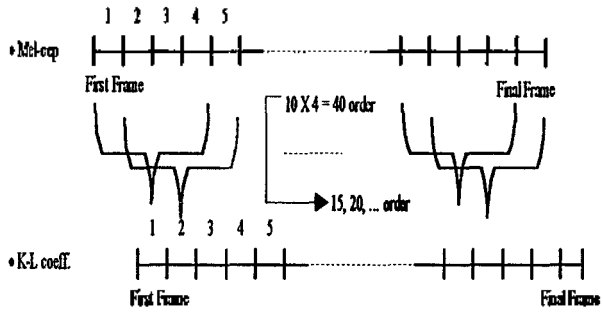


그림 5. K-L 동적계수 추출 방법.(단일)

또한, 멜-켑스트럼과 동적 특징으로 많이 이용하고 있는 회귀계수를 결합한 경우에 대해서 그림 6과 같이 2프레임 구간에 대해서 1프레임씩 시점을 이동시키면서 K-L 전개를 수행한 후 확장한 15, 20, 25차 등의 K-L 동적 계수를 추출하여 특징파라미터로 사용한다.

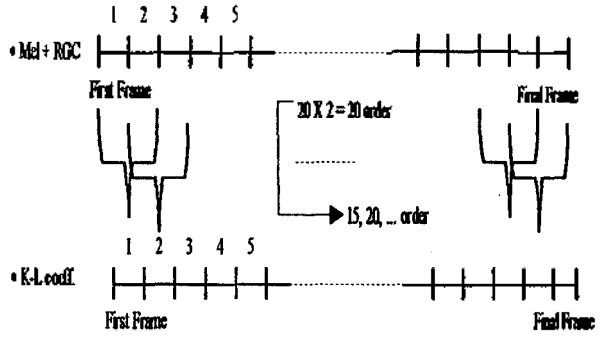


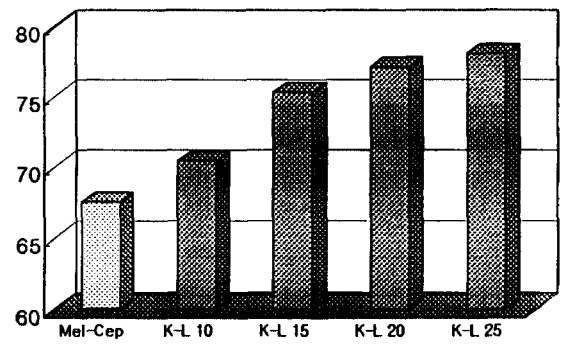
그림 6. K-L 동적계수의 추출 방법.(결합)

Ⅴ. 인식 실험 및 고찰

연속 숫자음에 강건한 모델 구성과 동적특징인 K-L 동적계수의 유효성을 확인하기 위해 인식 실험을 위한 음성자료로는 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 인식 실험을 수행하였다.

5.1 단일 특징을 이용한 인식실험

단일 특징파라미터로 많이 사용되고 있는 멜-켑스트럼과 이를 K-L 전개를 통해 추출한 K-L 동적 특징 파라미터를 사용하여 인식실험을 수행하였다. 그 결과를 그림 7에 나타내었다.



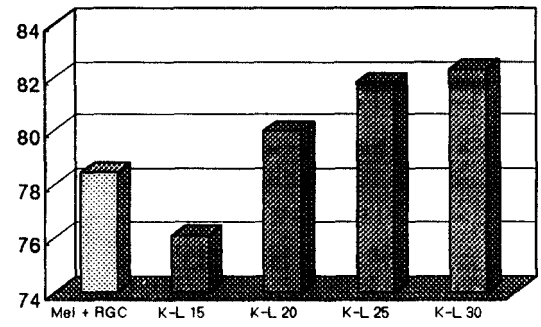
*Mel-Cep : 멜 켑스트럼, Rgc : 회귀 계수, K-L : K-L 계수

그림 7. 단일 특징파라미터의 비교

이상의 실험결과로부터 단일 특징파라미터로서 멜-켑스트럼을 사용한 경우 67.5%의 인식률을 보였으며, K-L 동적계수를 사용한 경우는 78.2%의 인식률을 나타내어 K-L 동적계수의 유효성을 확인할 수 있었다.

5.2 확장한 특징을 이용한 인식실험

확장한 K-L 동적특징의 유효성을 확인하기 위해 결합한 특징파라미터로서 멜-켑스트럼과 회귀계수를 사용한 경우와 이를 확장한 K-L 동적계수를 특징파라미터로 사용한 경우에 대해서 각각 인식실험을 수행하고, 그 결과를 그림 8에 나타내었다.



*Mel-Cep : 멜 켑스트럼, Rgc : 회귀 계수, K-L : K-L 계수

그림 8. 결합 특징파라미터의 비교

실험 결과로부터 멜-켑스트럼과 회귀계수를 결합하여 특징파라미터로 사용한 경우 78.4%의 인식률을 보였다. 또한, 이로부터 확장한 K-L 동적계수를 추출하여

특징파라미터로 이용한 경우 82.3%의 높은 인식률을 얻었다.

전체적으로 볼 때 단일 특징파라미터를 사용한 경우에 있어서는 기존의 멜-켑스트럼보다 K-L 동적계수를 사용한 경우가 10.7%의 매우 향상된 인식률을 보였으며, 특징파라미터를 결합한 경우에 있어서도 멜-켑스트럼과 회귀계수를 사용한 경우보다 멜-켑스트럼과 회귀계수로부터 확장한 K-L 동적 특징 계수를 사용한 경우가 비교적 높은 4%정도의 인식률의 향상을 보여 확장한 K-L 동적계수의 유효성을 확인하였다.

Ⅵ. 결 론

본 논문에서는 일반적으로 인식률이 저조한 연속 숫자음의 인식 정도 향상을 위해서 K-L (Karhanen-Loeve) 동적특징의 확장에 대해서 검토하고 4연속 숫자음을 대상으로 인식 실험을 수행하여 확장한 K-L 동적특징의 유효성을 확인하였다.

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

또한, 결합한 특징파라미터로서 멜-켑스트럼과 회귀계수를 사용한 경우 78.4%의 인식률을 보였으며, 이를 확장한 K-L 동적계수를 사용한 경우 82.3%로 4% 정도의 인식률을 향상을 보여 확장한 K-L 동적계수의 유효성을 확인하였다.

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

※ 본 연구에서 사용한 단어데이터베이스는 국어공학센터에서 구축한 4연속 숫자음 음성데이터베이스를 사용하였습니다.

참고 문헌

[1] Kazumasa Yamamoto and Seiichi Nakagkwa, " Comparative Evaluation of Segmental Unit Input HMM and Conditional Density HMM", ESCA.EUROSPEECH '95.4

[2] J.H.Lee, B.K.Kim and H.Y.Chung," Environmental Adaptation Using Maximum A Posteriori Estimation for Korean Word Recognition", Proceeding of IEEE Invited Workshop on Pattern Recognition for Multimedia Techniques, 1996.

[3] B.K.Kim, H.Y.Chung, "Typical FrameExtraction for Korean Phoneme Recognition", IEEE, APWT '95, 72-75, 1995.

[4] 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.

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

[6] 越川忠,"連續音聲認識システムにおけるHMMの話者適應化に關する研究",修士學位論文, 1993.

[7] 中川聖一, 甲裴充彦 "文脈自由文法制御によるOnePass型HMM音聲認識法," 信學論誌 D-II, Vol. J76-D-II, No.7. pp. 1337-1345, 1993.