1998년도 한국음향학회 학술발표대회 논문집 제 17 권 제 1(s) 호

K-L 동적 계수를 이용한 단어 인식

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

Word Recognition Using K-L Dynamic Coefficients

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

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

요 약

본 논문에서는 음성인식 시스템의 인식 정도의 향상을 위해서 동적 특징으로서 K-L(Karhanen-Loeve)계수를 이용하여 음소모델을 구성하는 방법을 제안하고, 음소, 단어, 숫자음 인식 실험을 통하여 그 유효성을 검토하였다.

인식 실험을 위한 음성자료는 한국 전자통신 연구소에서 채록한 445단어와 국어정보공학연구소에서 채록한 4연속 숫자음을 사용하였으며, K-L계수 동적 특징의 유효성을 확인하기 위해 정적 특징으로서 멜-켑스트럼과 동적 특징으로서 K-L계수 및 회귀계수를 추출한 후 음소, 단어, 숫자음 인식 실험을 수행하였다. 인식의 기본 단위로는 48개의 유사음소단위(Phoneme Likely Units ; PLUs)를 음소모델로 사용하였으며, 단어와 숫자음 인식을 위해서는 유한상태 오토마타(Finite State Automata; FSA)에 의한 구문제어를 통한 OPDP(One Pass Dynamic Programming)법을 이용하였다.

인식 실험 결과, 음소인식에 있어서는 정적특징인 멜-켑스트럼을 사용한 경우 39.8%, K-L 동적 계수를 사용한 경우가 52.4%로 12.6%의 향상된 인식률을 얻었다. 또한, 멜-켑스트럼과 회귀계수를 사용한 경우 60.1%, K-L계수와 회귀계수를 결합한 경우에 있어서도 60.4%로 높은 인식률을 얻었다.

이 결과를 단어인식에 확장하여 인식 실험을 수행한 결과, 기존의 멜-켑스트럼 계수를 사용한 경우 65.5%, K-L계수를 사용한 경우 75.8%로 10.3% 향상된 인식률을 얻었으며, 멜-켑스트럼과 회귀계수를 결합한 경우 91.2%, K-L계수와 회귀계수를 결합한 경우 91.4%의 높은 인식률을 보였다. 또한, 4연속 숫자음에 적용한 경우에 있어서도 멜-켑스트럼을 사용한 경우 67.5%, K-L계수를 사용한 경우 75.3%로 7.8%의 향상된 인식률을 보였으며 K-L계수와 회귀계수를 결합한 경우에서도 비교적 높은 인식률을 보여 숫자음에 대해서도 K-L계수의 유효성을 확인할 수 있었다.

I. 서 론

최근 정보통신 기술의 비약적인 발전과 더불어 멀티미디어 통신을 위한 휴먼 인터페이스에 관한 관심이 증가하고 있다. 이를 구현하기 위한 음성인식 기술의 발전은 반도체 메모리와 컴퓨터의 처리능력의 급속한 발달로 인하여 일부 제한된 부분에서 실용화가 이루어지고 있으나 실제로 이용 가능한 시스템 구현을 위해서는 아직 해결되어야 할 문제가 많은 실정이다.

현재 국내외 여러 연구 기관(한국통신, AT&T Bell Lab. ATR, CMU 등)에서 대어휘 연속음성인식 시스템의 상용화를 위한 연구가 활발하게 수행되고 있지만 아직까지 실용화 시스템을 성공적으로 개발한 예는 찾아보기 어렵다.

이러한 자동 음성인식 시스템의 실용화를 위해서는 음성 특징의 정확한 분석과 발성화자의 개인성, 발성의 종류, 어휘수, 언어의 복잡성, 환경 잡음 등과 같은 발성 환경적 요인에 의해 발생되는 여러 가지 문제점과 인식의 기본 단위에 대한 많은

연구가 요구되고 있다.

특히, 실용화를 위한 단어, 또는 대어휘 연속음성인식 시스템의 인식 정도의 향상을 위해서는 음성의 최소 인식단위와 그 특징파라미터에 대한 연구가 심도 있게 이루어져야 하지만 현재 대부분의 인식 시스템에서는 이에 대한 충분한 검토 없이 정적 특징과 동적 특징을 결합하여 이용하고 있다.

따라서 본 연구에서는 음성인식 시스템의 인식 정도의 향상을 위해서 일반적으로 이용되고 있는 정적특징과 더불어 동적 특징으로서 K-L 계수를 이용하여 음소 모델을 구성하는 방법을 제안하고, 음소, 단어, 숫자음 인식 실험을 통하여 그 유효성을 검토하고자 한다.

이를 위하여 음성자료는 한국 전자통신 연구소(ETRI)에서 채록한 445단어(ETRI 445)와 국어정보공학연구소(KLE)에서 채록한 4연속 숫자음(KLE 숫자음)을 사용하였으며, K-L 동적 계수의 유효성을 확인하기 위해 정적 특징으로서 멜-켑스트럼과 동적 특징으로서 K-L계수 및 회귀계수를 추출한 후 음소, 단어, 숫자음 인식 실험을 수행하여 K-L 동적 계수의 유효성을 확인하고자 한다.

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

본 논문의 구성은 다음과 같다. Ⅱ장에서 음성자료 및 분석에 대해서, Ⅲ장에서 인식방법에 대해서 서술한다. Ⅳ장에서는 인식 실험 및 고찰을 통하여 동적특징의 유효성을 확인하고 마지막 Ⅴ장에서 결론을 맺는다.

Ⅱ. 음성자료 및 분석

2.1 음성자료

음소와 단어의 인식 실험을 위한 음성자료는 한국 전자통신 연구소(ETRI)에서 구축한 한국인 남성 22인의 2회 발성한 445단어(ETRI 445)데이터 중에서 5인이 발성한 단어로부터 추출한 음소로 표준 패턴을 구성하고 학습에 참여하지 않은 3인의 화자가 발성한 단어를 인식 실험에 사용한다.

또한 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 인식 실험에 사용한다.

2.2 분석 방법

음성자료의 분석은 표 1에서와 같이 7Khz의 LPF를 통과한 후 샘플링 주파수 16KHz, 양자화 정도 16Bits A/D 변환기를 통해 이산데이터로 변환되고 Preempha-sis 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.

또, 멜-켑스트럼으로부터 동적 특징파라미터인 10차의 회귀계수와 10, 20차의 K-L계수를 구하여 음소, 단어, 숫자음 인식에 이용한다.

표 1. 음성자료의 분석조건.

Speech Data	ETRI 445 단어 / 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.

2.3 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 용의하게 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.

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

고유치 $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_m \geq \ldots \lambda_n \geq 0$에 대응하는 고유벡터 $a_1, a_2, \ldots, a_m$의 벡터에 의해 A를 구성하면 다음과 같다.

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

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

Ⅲ. 인식 방법

3.1 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.

3.2 인식 시스템

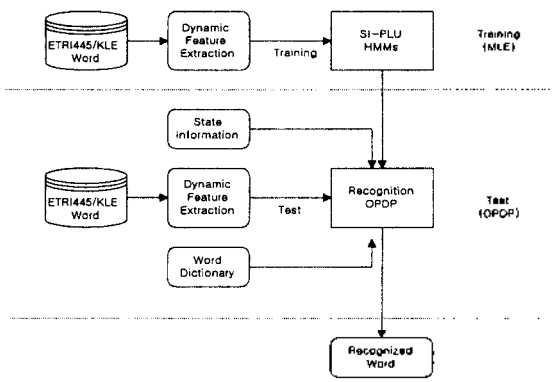

그림 1. 단어/숫자음 인식 시스템의 흐름도.

인식시스템은 표준패턴을 작성하기 위한 학습 단계와 작성된 표준패턴과 입력된 테스트패턴과의 유사도를 측정하여 최적의 상태열을 찾는 인식 단계로 구성된다.

학습 단계에서 CHMM을 이용하여 음소 표준패턴을 작성하여 음소인식 실험을 수행하고, 단어와 숫자음 인식시에는 인식단계에서 미리 작성한 단어사전과 유한상태 오토마타(FSA)에 의한 구문제어를 통하여 OPDP법으로 인식을 수행한다.

그림 1에 단어와 숫자음 인식을 위한 인식 시스템의 전체 구성도를 나타내었다.

Ⅳ. 인식 실험 및 고찰

4.1 음소 인식 실험

인식 실험에 있어서는 2.1절에 설명한 음성자료를 이용하여 추출한 음소로부터 동적특징으로서 K-L(Karhan-en Loeve)계수의 유효성을 확인하기 위해 정적특징인 멜-캡스트럼과 동적특징인 K-L계수를 이용하여 화자독립 인식 실험을 수행하였다. 또, 멜-캡스트럼과 회귀계수, K-L계수와 회귀계수와 같이 특징파라미터를 결합하여 음소 인식 실험을 수행하였다. 표 2에 각 특징파라미터에 대한 인식 실험 결과를 나타내었다.

표 2. 특징파라미터에 따른 화자독립 음소인식률.

특징파라미터	차원수	인식률(%)
① Mel-Cep.	10차	39.8
② K-L	10차	48.0
	20차	52.4
③ Mel-Cep+Rgc	10차+10차	60.1
④ Mel-Cep+K-L	10차+10차	47.2
⑤ K-L+Rgc	10차+10차	57.9
	20차+10차	60.4

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

표 2의 인식 실험 결과로부터 ①의 경우는 39.8%, ②의 경우는 52.4%로 12.6%의 높은 인식률을 얻었다. 또한, 특징파라미터를 결합한 ③, ④, ⑤와 같이 결합한 경우에 있어서도 K-L계수와 회귀계수가 포함된 경우가 비교적 높은 인식률을 나타내어 음소인식에 있어서 K-L계수의 유효성을 확인할 수 있었다.

4.2 단어 인식 실험

이상의 음소 인식 실험 결과를 참고로 ETRI 445단어로 확장하여 단어 인식 실험을 수행하고 각 특징파라미터에 따른 인식률의 변화를 표3에 나타내었다.

표 3. 특징파라미터에 따른 화자독립 단어인식률.

특징파라미터	차원수	인식률(%)
① Mel-Cep	10차	65.5
② K-L	10차	72.1
	15차	75.8
③ Mel+Rgc	10차+10차	91.2
④ K-L+Rgc	10차+10차	89.6
	15차+15차	91.4

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

이상의 결과로부터 정적 특징을 사용한 ①의 경우는 65.5%, 동적특징인 K-L계수를 이용한 ②의 경우는 75.8%로 10.3%의 향상 된 인식률을 얻었다. 또한, ③, ④와 같이 특징파리미터를 결합한 경우에 대해서도 K-L계수를 이용한 경우가 비교적 높은 인식률을 보여 음소를 인식의 기본단위로 한 단어인식 시스템에 있어서도 K-L특징의 유효성을 확인할 수 있었다.

4.3 숫자음 인식 실험

단어와 숫자음에 강건한 모델 구성과 동적특징인 K-L계수의 유효성을 확인하기 국어공학센터(KLE)에서 구축한 4연속 숫자음을 대상으로 인식 실험을 수행하였다. 그 결과를 표 4에 나타내었다.

표 4. 특징파라미터에 따른 화자독립 4연속 숫자음 인식률.

특징파라미터	차원수	인식률
① Mel-Cep	10차	67.5
② K-L	10차	70.4
	15차	75.3
③ Mel+Rgc	10차+10차	76.4
④ K-L+Rgc	10차+10차	79.5
	15차+15차	79.4

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

이상의 결과로부터 단일 특징파라미터를 사용한 경우 ①의 경우보다 K-L 계수를 사용한 ②의 경우가 7.8% 높은 인식률을 나타내었다. 또한, 특징파라미터를 결합한 ③, ④의 경우에서 있어서도 K-L계수를 이용한 경우가 3%정도의 높은 인식률을 보여 숫자음에 대해서도 K-L 계수의 유효성을 확인하였다.

전체적으로 볼 때 음소, 단어, 숫자음에서 단일 특징인 경우에 있어서는 기존의 경우보다 K-L계수를 사용한 경우가 매우 향상된 인식률을 보였으며 특징파라미터를 결합한 경우에도 비교적 안정된 인식률을 얻어 제안한 K-L 동적 계수의 유효성을 확인할 수 있었다.

V. 결 론

본 논문에서는 음성인식 시스템의 인식 정도의 향상을 위해서 동적 특징으로서 K-L계수를 이용하여 음소모델을 구성하는 방법을 제안하고, 음소, 단어, 숫자음 인식 실험을 통하여 그 유효성을 확인하였다.

인식 실험 결과, 음소인식에 있어서는 정적특징인 멜-캡스트럼을 사용한 경우 39.8%, K-L 동적 계수를 사용한 경우가 52.4%로 12.6%의 향상된 인식률을 얻었다. 또한, 멜-캡스트럼과 회귀계수를 사용한 경우 60.1%, K-L계수와 회귀계수를 결합한 경우에 있어서도 60.4%로 높은 인식률을 얻었다.

이 결과를 단어인식에 확장하여 인식 실험을 수행한 결과, 기존의 멜-캡스트럼 계수를 사용한 경우 65.5%, K-L계수를 사용한 경우 75.8%로 10.3% 향상된 인식률을 얻었으며, 멜-캡스트럼과 회귀계수를 결합한 경우 91.2%, K-L계수와 회귀계수를 결합한 경우 91.4%의 높은 인식률을 보였다. 또한, 4연속 숫자음에 적용한 경우에 있어서도 멜-캡스트럼을 사용한 경우 67.5%, K-L계수를 사용한 경우 75.3%로 7.8%의 향상된 인식률을 보였으며 K-L계수와 회귀계수를 결합한 경우에서도 비교적 높은 인식률을 보여 숫자음에 대해서도 K-L계수의 유효성을 확인할 수 있었다.

전체적으로 볼 때 음소, 단어, 숫자음에서 단일 특징인 경우에 있어서는 K-L계수를 이용한 경우가 매우 향상된 인식률을 보였으며 특징파라미터를 결합한 경우에도 비교적 안정된 인식률을 얻어 제안한 K-L 동적 계수의 유효성을 확인할 수 있었다.

향후 이상의 결과를 바탕으로 단어와 숫자음에 강건한 모델을 구성하여 대어휘 연속음성인식 시스템에 적용하고자 한다.

※ 본 연구에서 사용한 단어데이터베이스는 한국통신이 출연하여 한국전자통신연구소가 구축한 445단어 음성데이터베이스와 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

[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.