

단어인식을 위한 음소의 동적특징에 관한 검토

김주곤*, 김범국*, 정현열*

*영남대학교 전기전자공학부

A Study on The Dynamic Feature of Phoneme for Word Recognition

Ju-Gon Kim*, Bum-Koog Kim*, and Hyun-Yeol Chung*

*School of Electrical and Electronic Engineering, Yeungnam University

요약

본 연구에서는 음소를 인식의 기본단위로 하는 한국어 단어인식 시스템의 인식 정도를 개선하기 위해 각 음소의 시간방향의 정보를 포함하고 있는 동적특징인 회귀계수와 K-L(Karhunen-Loeve) 변환으로 얻은 특징파라미터(이하 K-L계수라 함)를 이용하여 음소인식과 단어인식 실험을 수행한 결과 그 유효성을 확인하였다.

이를 위해 먼저 파열음을 대상으로 정적 특징파라미터인 멜-켄스트럼(Mel-Cepstrum)과 동적 특징파라미터인 회귀계수(Regressive Coefficient)와 K-L계수(Karhunen-Loeve Coefficient)를 추출하여 음소 인식 실험을 수행하였다. 그 결과 멜-켄스트럼을 사용한 경우 39.84%, 회귀계수를 사용한 경우 48.52%, K-L계수를 사용한 경우 52.40%의 인식률을 얻었다. 이를 참고로 각각의 특징파라미터를 결합하여 인식 실험한 결과 멜-켄스트럼과 K-L계수를 사용한 경우 47.17%, 멜-켄스트럼과 회귀계수의 경우 60.11%, K-L계수와 회귀계수의 경우 60.35%, 멜-켄스트럼과 K-L계수, 회귀계수를 사용한 경우 58.13%를 인식률을 얻어 동적특징인 K-L계수와 회귀계수를 사용한 경우와 멜-켄스트럼과 회귀계수를 사용한 경우가 높은 인식률을 보였으며 이를 단어로 확장하여 인식실험을 수행한 결과 기존의 특징파라미터를 이용한 경우보다 높은 인식률을 얻어 동적 특징파라미터의 유효성을 확인하였다.

1. 서론

음성인식 기술은 아직도 많은 해결해야 할 과제가 남아 있지만 1980년대로 돌아서면서부터 HMM과 같은 현재 가장 널리 사용되어지고 있는 음성 처리기술이 확산되면서 획기적인 발전을 거듭해왔

으며 현재는 소규모 고립단어 인식에서 벗어나 불특정 화자가 제한 없이 자연스럽게 발성한 음성을 정도 높게 인식하는 대어휘(Large Vocabulary), 연속음성(Continuous Speech)의 인식 및 이해 시스템 개발을 위한 활발한 연구가 진행되고 있으나 실용화를 위해서는 화자, 어휘수, 발성 내용, 발성 방법, 발성 환경 등에 대한 조건 대한 많은 연구가 선행되어야 한다.

따라서 저자들의 연구실에서는 그 동안 음소 인식의 기본단위로 한 단어인식 시스템, 연속음성 인식 시스템 등[1,2,3,4]을 개발하였으며 현재 이들 시스템의 실용화를 목표로 채널왜곡과 환경잡음, 인식속도 개선, 인식률의 정도 향상을 위한 연구를 진행하고 있으며 이들 연구의 일환으로서 본 연구에서는 인식 시스템의 성능 향상을 위해서 기존의 특징으로 많이 사용하고 있는 정적특징인 멜-켄스트럼과 음소의 시간방향의 정보를 포함하고 있는 동적특징인 K-L계수와 회귀계수를 이용하여 화자독립 음소인식 실험을 수행하여 동적 특징파라미터의 유효성을 검토하고, 이를 참고로 멜-켄스트럼과 K-L계수, 멜-켄스트럼과 회귀계수, K-L계수와 회귀계수, 멜-켄스트럼과 K-L계수, 회귀계수 등과 같이 특징파라미터를 결합하여 인식 실험을 수행하고 동적특징을 이용한 경우가 우수할 경우 이를 단어 인식으로 확장하여 동적특징의 유효성을 검토하고자한다.

이때 인식의 기본단위는 CHMM(Continuous Hidden Markov Model) 음소모델로 하고 단어인식을 위해서는 탐색 시간 면에서 좋은 성능을 보이면서, 상대적으로 적은 기억장소를 이용하는 OPDP(One Pass Dynamic Programming)법을 이용한다. 본 논문의 구성은 다음과 같다. II장에서 음성 자료와 분석방법에 대해서, III장에서 OPDP법에

의한 인식방법에 대해서 서술한다. IV장에서는 인식실험 및 고찰을 통하여 동적특징의 유효성을 확인하고 마지막 V장에서 결론을 맺는다.

II. 음성자료의 분석

2.1 특징파라미터 추출

음성자료의 특징파라미터 추출 과정을 그림 1에 나타내었다. 음성자료는 7Khz의 LPF를 거쳐 16비트로 양자화된 ETRI 445 단어를 사용한다. 음성자료는 Pre-emphasis 필터를 통과시켜 고역 강조된 후 16ms(256points)의 길이의 해밍 윈도우를 사용하여 5ms(80points)씩 쉬프트 시키면서 분석된다.

이로부터 20차의 LPC 계수를 구하고 14차 LPC cepstrum 계수를 구한 후, 10차의 LPC 멜-cepstrum 계수를 구하여 정적 특징파라미터로 사용한다. 그리고 멜-cepstrum으로부터 동적 특징파라미터인 10차의 회귀계수와 10, 20차의 K-L계수를 구하여 음소인식과 단어인식에 이용한다. 표 1에 음성자료의 분석 조건을 나타내었다.

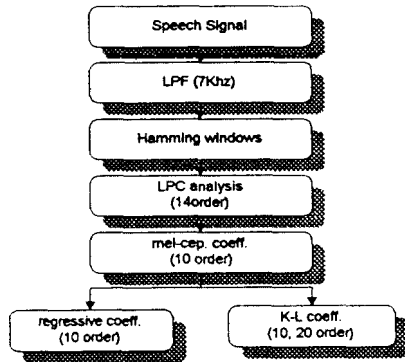


그림 1. 특징 파라미터의 추출.

표 1. 음성자료의 분석조건.

Speech Data	ETRI 445
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.

2.2 동적 특징파라미터 추출

2.2.1 회귀계수

본 연구에서는 음성신호 스펙트럼 내에서의 순시적인 변화를 나타내는 동적 특징파라미터로서 회귀계수를 이용한다[4,7]. 회귀계수는 2.1절에서 구한 10차의 LPC 멜-cepstrum 계수로부터 각 프레임마다 추출하는데, 이때 회귀계수 $R_m(t)$ 는 시간 t 를 중심으로 $2\delta+1$ 폭 만큼의 단위로 다음의 (1)식으로부터 구할 수 있다.

$$R_m(t) = \frac{\sum_{n=-\delta}^{\delta} n C_m(t+n)}{\sum_{n=-\delta}^{\delta} n^2} \quad (1)$$

여기서, $C_m(t)$ 는 발성 음의 t 번째 프레임의 m 번째 계수 값이고 $R_m(t)$ 은 여기에 해당하는 회귀계수의 값을 의미한다.

2.2.2 K-L 계수

여러개의 양적 변수들 사이의 관계를 분석하여 이 변수들의 선형결합으로 표시되는 주성분을 찾고 이 중에서 중요한 몇 개의 주성분으로 전체의 변동을 설명하고자 하는 다변량 분석법이 K-L 변환법으로 자료의 요약이나 선형관계식을 통하여 차원을 감소시켜 분석을 용이하게 하는데 목적이 있다. 즉, 다차원 공간에 대하여 관측 벡터 분포의 불균일성을 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-L변환법[7]에 대해서 간략하게 서술한다.

n 차원의 관측벡터 X 의 공분산 행렬을 S 라 하면 다음과 같다.

$$S = \frac{1}{N-1} \sum_{i=1}^N (X_i - \bar{X})(X_i - \bar{X})' \quad (2)$$

여기서, X_i 는 i 번째 관측벡터, \bar{X} 는 전체 관측벡터의 평균벡터, N 는 전체의 샘플수이다.

K-L변환에서 선형변환행렬 A 는 다음의 특징평가함수 $J(a)$ 의 최대화 조건을 만족하는 벡터 a_j ($j=1, 2, 3, \dots, m$)로 구성된다.

$$J(a) = \frac{a' \cdot S \cdot a}{a' \cdot a} \quad (3)$$

여기서, a 는 다음의 고유치 문제를 풀어서 얻어진 고유벡터로 된다.

$$S \cdot a - \lambda \cdot a = 0 \quad (4)$$

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

$$A = [a_1, a_2, \dots, a_m] \quad (5)$$

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

III. OPDP법에 의한 인식

인식알고리즘은 탐색 시간 면에서 좋은 성능을 보이면서, 상대적으로 적은 기억장소를 이용하는 OPDP(One Pass Dynamic Programming)법을 사용한다. OPDP법은 일반적으로 음절이나 단어 단위로 작성된 표준패턴을 사용하여 연속음성 인식에 사용되어진다.

본 연구에서는 이를 개량하여 유사음소단위로 표준패턴을 작성하고 단어사전과 구문제어를 통하여 인식을 수행한다. 단어의 누적거리 계산은 입력의 첫 번째 프레임과 각 유사음소단위 표준패턴과의 Viterbi Score를 프레임에 동기시켜 가면서 계산한 후 저장한다[1,6]. 그리고 이를 단어사전중의 각 단어의 마지막 프레임까지 확장해서 총 누적거리 값이 최소가 되는 단어의 인덱스를 인식결과로 출력하게 하였다. 그림 2에 단어 인식시스템의 전체 흐름도를 나타낸다.

로서 마지막 상태를 제외한 모든 상태에서 자기 천이를 가지는 순수한 Left-to-Right 모델을 사용하였으며 이 모델로써 각 음소별 특징파라미터를 학습한 후 음소와 단어 인식모델로 이용하였다. 그림 3에 본 연구에서 사용한 연속분포 HMM 모델의 구성을 나타내었다.

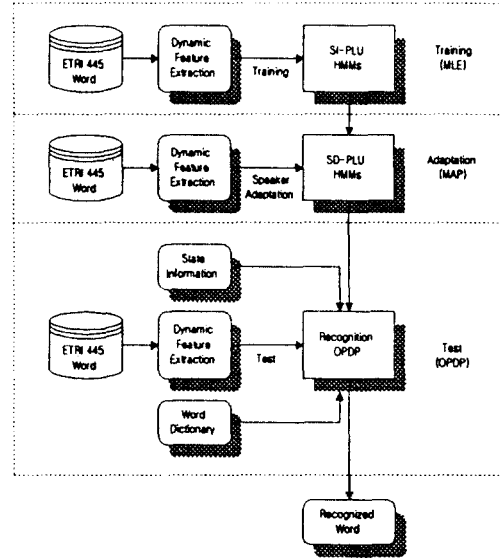


그림 2. 단어인식 시스템의 흐름도.

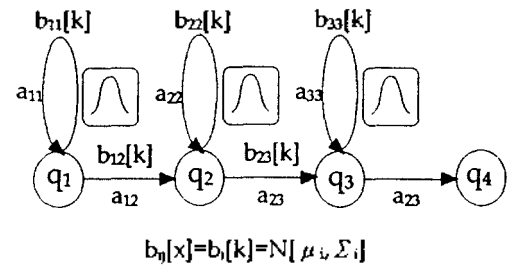


그림 3. 연속분포 HMM의 구성.
(4상태 1혼합)

IV. 실험 및 고찰

4.1 음성자료 및 인식모델

본 연구에서는 한국 전자통신 연구소(ETRI)에서 구축한 한국인 남성 22인의 2회 발성의 445단어(ETRI445)를 사용하며 이 중에서 5인이 발성한 단어로부터 추출한 음소를 모델학습에 사용하였으며 3인의 화자가 발성한 100단어를 인식에 사용하였다.

초기 HMM 음소모델의 구조는 4상태 1 혼합으

4.2 인식 실험 및 고찰

4.2.1 음소 인식 실험

인식실험에 있어서는 먼저 각 음소로부터 추출한 동적 특징파라미터의 유효성을 확인하기 위해 정적특징인 맨-캡스트럼과 동적특징인 K-L계수, 회귀계수를 특징파라미터로해서 파일음을 대상으로 화자독립 인식실험을 수행하였다. 또, 맨-캡스트럼과 K-L계수, 맨-캡스트럼과 회귀계수, K-L계

수와 회귀계수, 맨-캡스트럼과 K-L계수, 회귀계수 등과 같이 특징파라미터를 결합하여 음소인식 실험을 수행하였다. 표 2에 특징파라미터에 따른 인식결과를 나타내었다.

표 2. 특징파라미터 따른 화자독립 음소 인식률.

특징파라미터	차원수	인식률(%)
①Mel-Cep	10차	39.8
②Rgc	10차	48.5
③K-L	10차	48.0
	20차	52.4
④Mel-Cep+Rgc	10차+10차	60.1
⑤Mel-Cep+K-L	10차+10차	47.2
⑥K-L+Rgc	10차+10차	57.9
	20차+10차	60.4
⑦Mel-Cep+K-L+Rgc	10차+10차+10차	58.1

표 2의 인식실험 결과로부터 ①, ②, ③의 경우에 있어서는 ③의 경우가 비교적 높은 인식률을 보였으며 특징파라미터를 ④, ⑤, ⑥, ⑦의 경우와 같이 결합한 경우에 있어서는 동적 특징파라미터인 K-L계수와 회귀계수가 추가된 ④, ⑥의 경우가 높은 인식률을 나타내어 동적특징의 유효성을 확인하였다.

4.2.2 단어 인식 실험

이상의 인식실험 결과를 참고로 음소 인식률이 높게 나타난 K-L계수와 회귀계수를 사용한 경우와 맨-캡스트럼과 회귀계수를 사용한 경우에 대해서 ETRI 445 단어중에 100단어를 대상으로 단어 인식 실험을 수행하고 기존의 특징파라미터를 이용한 경우와 비교 검토하였다.

그 결과 맨-캡스트럼을 특징파라미터로 이용한 경우보다 K-L계수를 이용한 경우가 높은 인식률을 얻었으며 전체적으로는 맨-캡스트럼과 회귀계수, K-L계수와 회귀계수를 이용한 경우가 비교적 높은 단어 인식률을 보여 음소인식의 기본단위로한 단어인식 시스템에 동적특징을 적용함으로써 인식률 향상을 기대 할 수 있었다.

V. 결론

본 연구에서는 음소인식의 기본단위로 하는 한국어 단어인식 시스템의 인식 정도를 개선하기 위해 각 음소의 시간방향의 정보를 포함하고 있는 동적특징인 회귀계수와 K-L계수를 이용하여 음소인식과 단어인식 실험을 수행한 결과 그 유효성을 확인하였다.

이를 위해 먼저 정적 특성파라미터인 맨-캡스트럼과 동적 특성파라미터인 회귀계수와 K-L계수를 사용하여 파일음을 대상으로 음소인식 실험을 수행하였다. 그 결과 맨-캡스트럼을 사용한 경우 39.84%, 회귀계수를 사용한 경우 48.52%, K-L계수를 사용한 경우 52.40%의 인식률을 얻었다. 이를 참고로 각각의 특징파라미터를 결합하여 인식 실험한 결과 맨-캡스트럼과 K-L계수를 사용한 경우 47.17%, 맨-캡스트럼과 회귀계수의 경우 60.11%, K-L계수와 회귀계수의 경우 60.35%, 맨-캡스트럼과 K-L계수, 회귀계수를 사용한 경우 58.13%를 인식률을 얻어 동적특징인 K-L계수와 회귀계수를 사용한 경우와 맨-캡스트럼과 회귀계수를 사용한 경우가 높은 인식률을 보였으며 이를 단어로 확장하여 인식실험을 수행한 결과 기존의 맨-캡스트럼을 이용한 경우보다 K-L계수를 이용한 경우가 높은 인식률을 나타내었으며 전체적으로는 동적특징인 맨-캡스트럼과 회귀계수, K-L계수와 회귀계수를 이용한 경우가 비교적 높은 단어 인식률을 보여 음소인식의 기본 단위로한 단어인식 시스템에 동적특징을 적용함으로써 인식률 향상을 기대 할 수 있었다.

향후 이상의 결과를 바탕으로 학습과 평가용 음성데이터를 증가시키면 인식실험을 수행하여 단어인식기, 연속음성 인식시스템의 인식률 향상을 위한 연구를 계속하고자한다.

※ 본 연구에서 사용한 단어데이터베이스는 한국통신이 출연하여 한국전자통신연구소가 구축한 445단어 음성데이터베이스를 사용하였습니다.

참고 문헌

- [1] 이 정훈, 정 현열, "단어인식을 위한 환경적응화에 관한 연구", 제13회 음성통신 및 신호처리 워크샵 논문집, pp. 50-54, 한국음향학회, 1996.8.
- [2] 이 시욱, 정 현열, "음성인식 기능을 가진 주소 입력검색 시스템" 제9회 신호처리 합동학술대회 논문집, pp. 611-614, 한국음향학회, 1996.10.
- [3] 이 우형, 정 현열, "환경잡음에 강한 음성인식기의 FRONT-END" 제13회 음성통신 및 신호처리 워크샵 논문집, pp. 356-360, 한국음향학회, 1996.8.
- [4] J.R.Choi, B.K.Kim, H.Y.Chung and Seiichi Nakagawa, "A Korean Flight Reservation System Using Continuous Speech Recognition", The Journal of the Acoustical Society of Korea, Vol.

15. No. 3E, 1996.9.
- [5]B.K.Kim, H.Y.Chung, "Typical Frame Extraction for Korean Phoneme Recognition", IEEE, APWT '95, 72-75, 1995.12.
- [6]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, (1996.10).
- [7]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.
- [8]越川忠, "連続音聲認識システムにおけるHMMの話者適應化に関する研究", 修士學位論文, 1993.
- [9]Rabiner, Juang, "Fundamentals of Speech Recognition", Prentice-Hall International, Inc, 1993.
- [10]甲斐充彦, "自然發話のための音聲認識システムに関する研究", 1996.
- [11]Peter F.Brown, "Speech Recognition By Statistical Methods", IBM 1985.11.