

한국어의 경음에 대한 분석

임 수 호, 김 주 곤, 김 병 국*, 정 호 열, 정 현 열
 영남대학교 전자정보공학부
 *대구과학대학 영상통신미디어과

Analysis of Unaspirated sound for Korean

Soo-Ho Lim, Joo-Gon Kim, Bum-Guk Kim*, Ho-Youl Jung, Hyun-Yeol Chung
 Dept. of Information and Communication Eng., Yeungnam University
 *Dept. of Image Communication Media, Taegu Science College
 E-mail: shlim@yumail.ac.kr

요 약

본 논문에서는 한국어에만 나타나는 경음에 대하여 음운학적, 음향학적 특성을 고찰하고 이를 기반으로 음성인식 실험을 수행한 후 그 결과를 분석하였다. 음성인식 실험을 위하여 입력 음성을 48개의 유사음소단위 (PLU; Phoneme Likely Unit)로 레이블링을 한 후 각각의 음소군에 대하여 LPC (Liner Predictive Coding) 분해능을 증가시키면서 음소인식 및 단어인식 실험을 수행하였다. 그 결과, 음소 인식 실험에서 경음군의 인식률이 가장 낮게 나타나 경음에 대한 분석이 보다 많이 필요함을 알 수 있었다. 또한 LPC의 분해 차원이 23차 일 때 경음과 전체 음소 인식률이 각각 34.11%, 46.1%로 나타나 가장 양호함을 알 수 있었으며 단어인식 실험에서도 LPC 23차와 25차 일 때 81.68%, 81.87%로 인식률이 가장 좋음을 알 수 있었다. 이상의 실험 결과에서 한국어의 경음은 전체 시스템의 인식 성능과 밀접한 관계가 있음을 알 수 있었다.

1. 서 론

최근 음성인식 기술의 발전으로 소규모 고립단어 인식에서 벗어나 불특정 화자가 제한 없이 자연스럽게 발성한 음성을 인식하는 대어휘 연속음성 (Large Vocabulary Continuous Speech)의 인식 시스템 개발을 위한 활발한 연구가 진행 되고 있다. 이러한 대어휘 연속음성인식 시스템을 구현하기 위해서는 음성 특징의 정확한 분석과 화자의 개인성, 발성의 종류, 어휘수, 언어의 복잡성, 환경적인 요인, 인식의 단위등에 의한 여러 가지 문제

점을 극복하기 위한 연구가 아직도 많이 필요하다. 특히, 음소를 기반으로 하는 음성인식 시스템에서는 한국어의 각 음소에 대한 정확한 분석이 필요하다.

한국어의 음소에 대하여 살펴보면 크게 자음과 모음으로 분류되며 자음은 조음 방식에 의해 폐쇄음, 마찰음, 파찰음, 비음, 설측음의 다섯 가지로 구분 된다. 이 중에서 폐쇄음과 파찰음은 기식(aspiration)의 유무와 긴장에 따라 경음, 기음, 평음의 3분류[1][2][3]로 나누어지고 특히 경음은 일본어, 영어, 독일어(이분 대립)와 달리 한국어(삼분 대립)에만 나타나는 특성이 있다. 조음 방식에 따른 한국어의 자음 체계를 표 1에 나타내었다.

표 1. 한국어의 자음 체계

| | | 양순음 | 치정음 | 치경정구 개음 | 연구개음 | 상분음 |
|-----|----|-----|-----|------------|------|-----|
| 폐쇄음 | 평음 | ㅂ | ㄷ | | ㄱ | |
| | 기음 | ㅃ | ㄸ | | ㅋ | |
| | 경음 | ㅍ | ㅌ | | ㆁ | |
| 마찰음 | 평음 | | ㅅ | | | ㅎ |
| | 경음 | | ㅆ | | | |
| 파찰음 | 평음 | | | ㅈ | | |
| | 기음 | | | ㅊ | | |
| | 경음 | | | ㅉ | | |
| 비음 | | ㅁ | ㄴ | | ㅇ | |
| 설측음 | | | ㄹ | | | |

본 논문에서는 한국어 음소의 음성학적, 음향학적 특성을 분석하고 음성인식 실험을 통하여 한국어 음소에 적합한 특징파라미터의 차원을 찾아 음성인식 시스템의 성능을 높이고자 한다.

본 논문의 구성은 다음과 같다. 2장에서 한국어의 경음 특성에 대해서 살펴보고, 3장에서는 음성인식 실험을 통하여 음소 군별로 분석하고, 경음의 변화를 비교 분석하였으며, 이를 확장하여 단어 실험을 하였다. 마지막으로 4장에서 결론을 맺는다.

2. 한국어의 경음 특성

한국어 자음의 분류에서 기음, 평음, 경음은 다음과 같은 세 가지 척도로 구분 된다.

1. 성대 진동 시작 시간 (VOT: Voice Onset Time)
2. 폐쇄 지속 시간 (Closure Duration)
3. 기식 (aspiration)

이 중에서 VOT는 폐쇄(closure)의 개방(release)에서 성대 진동의 시작 시점 사이의 시간이다. 즉, 파열음이 개방한 후에 후행 모음을 위해 성대가 진동하기 시작하는 시점까지 걸린 시간을 말한다. 성대 진동 시작 시간은 음향학적으로 기식의 정도에 비례한다. 기식성이 큰 파열음일수록 폐쇄기간 동안 성대가 멀리 떨어져 있는 상태로 조음되므로 폐쇄의 개방 이후 후행하는 모음을 위해 성대가 진동하기 시작할 때까지 더 많은 시간이 걸리기 때문이다[1][2][4].

경음, 평음, 기음의 성대 진동 시작 시간(VOT)을 그림 1에 나타내었다.

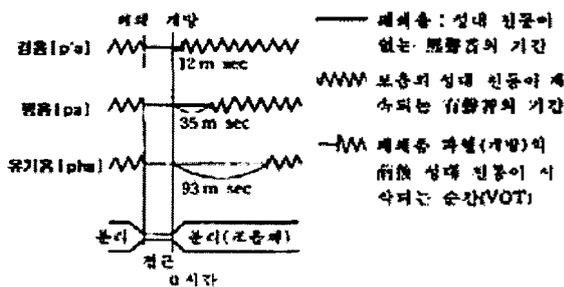


그림 1 경음, 평음, 기음의 VOT 비교

표 2. 국어 자음의 VOT 기각:

| | 음소 | VOT |
|-----|------------|------|
| 경음 | ㄱ, ㄷ, ㅂ, ㅈ | 12ms |
| 평음 | ㄴ, ㄷ, ㄹ, ㅅ | 35ms |
| 유기음 | ㅋ, ㆁ, ㅍ, ㅊ | 93ms |

그림 1.과 표 2에서 보는 바와 같이 VOT의 시간이 '유기음 > 평음 > 경음'의 순서로 지연된다. 이 이유는 경음을 발성 할 때 성문이 좁게 열리므로 VOT가 가장 짧게 나타난다[3][8].

현재까지 경음을 포함한 파열음을 대상으로 한 연구의 예를 살펴보면 서민경[4]의 "한국어 파열음의 VOT에 관한 실험 음성학적 연구 -환경에 따른 VOT 변이를 중심으로-", 배재연 등[5]의 "음성 환경에 따른 한국어 폐쇄음의 음향적 특성: 시간적 특성을 중심으로" 등의 연구들이 있으나 이들은 음성학, 음운론적인 접근 방법 이었다. 음소를 기반으로 하는 음성인식 시스템에서도 경음을 포함한 음소모델의 정확한 특징 분석과 검토 등에 대한 연구가 심도 있게 진행되어야 한다.

3. 실험 결과 및 분석

본 논문에서는 한국 전자통신 연구소(ETRI)에서 구축한 한국인 남성 22인의 2회 발성의 445단어(ETRI445)를 사용하며 이 중에서 5명이 발성한 단어로부터 추출한 음소를 모델학습에 사용하였으며, 학습에 참여 하지 않은 3명의 화자를 음소실험 및 단어 인식에 사용하였다. 음성인식 알고리즘은 탐색 시간 면에서 좋은 성능을 보이면서 상대적으로 적은 기억장소를 이용하는 OPDP(One Pass Dynamic Programming)법을 사용하였다[6][7][9].

음성자료는 Pre-emphasis 필터를 통과 시켜 고역 강조된 후 16ms(256points) 길이의 해밍 윈도우를 사용하여 5ms (80points) 씩 쉬프트 시키면서 분석된다. 이로부터 다양한 차원으로 LPC 캡스트럼 계수를 구한 후, 10차의 LPC 멜-캡스트럼을 정적 특징 파라미터로 사용하여 음소 인식과 단어 인식에 이용한다. 표 3에 음성자료의 분석 조건을 나타내었다[7][9][10].

표 3. 음성 자료의 분석조건

| Speech Date | ETRI 445 |
|---------------------------|-------------------------|
| Sampling frequency | 16kHz |
| Filtering | LPF |
| Resolution | 16bits |
| Hamming window | 16ms(256points) |
| Frame rate | 5ms(80points) |
| Analysis | LPC analysis |
| Static Feature parameters | 10order Mel-Cep. coeff. |

또한 음성자료는 핸드레이블링 된 정보에 의해 각각의 음소군으로 분류되었다. 그림 2는 한국 전자통신 연구소(ETRI)에서

구축한 한국인 남성 22인의 2회 발성한 445단어의 음소 군별 분포 비율을 나타내고 있다.

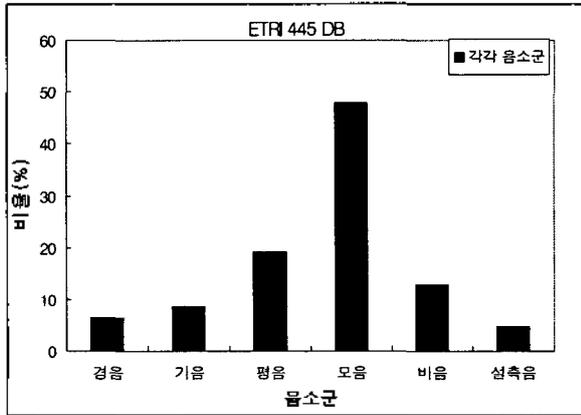


그림 2. 각 음소군별 분포

3.1 음소 인식 결과 및 분석

먼저, 각각의 음소들의 인식성능을 알아보기 위한 화자 독립 음소 인식 실험에서는 묵음을 제외한 47개의 유사 음소에 대하여 4상태 3출력 Left-to-Right 모델을 구성하였다. 음소의 지속시간을 고려하여 LPC 차원을 11차부터 26차 까지 증가시키면서 입력 음성을 분석하였다.

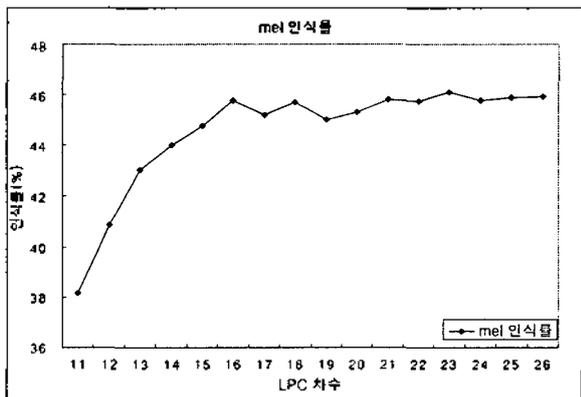


그림 3. 음소 인식률

그림 3는 LPC 차원을 11차부터 26차 까지 증가 시켜 입력음성을 분석하면서 전체 음소에 대하여 인식실험을 한 결과이다. 그 결과, LPC 차원이 23차 일 때 가장 좋은 인식률, 46.1%를 나타내었다.

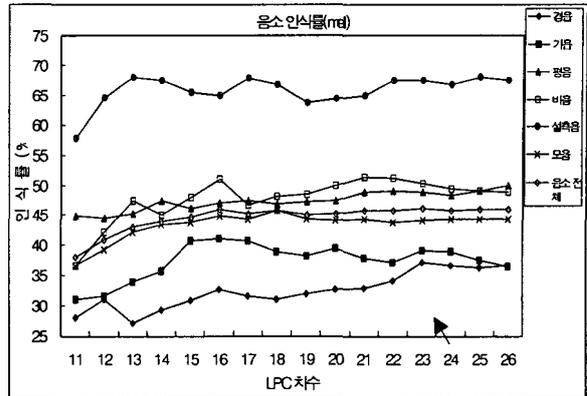


그림 4. 음소군별 인식률(mel)

음소 인식 실험결과 바탕으로 각각의 음소 군별로 인식률을 분석하여 그림 4에 나타내었다. 이 분석 결과에서 경음의 인식성능이 가장 저조함을 알 수 있었다. 기음과 비음은 각각 LPC 16차 일 때 가장 좋은 인식률을 보였으며, 모음, 설측음, 평음은 인식률이 LPC 분해능에 민감하지 않았다. 이에 반해 경음은 LPC 차원의 증가에 따라 인식률도 증가하여 경음에 대해 보다 많은 LPC 분해능이 필요함을 알 수 있었다. 또한 경음의 인식률이 가장 좋은 LPC 23차 일 때, 전체 음소 인식률도 가장 좋은 성능을 나타내었다. 이상으로부터 경음은 전체 음소 인식률에 상당한 영향이 있음을 알 수 있었다.

3.2 단어 인식 결과 및 분석

음소 인식 실험을 확장한 단어 실험에서도 LPC 차원을 11차부터 26차 까지 증가시키면서 구해진 특징파라미터를 이용하여 인식 실험을 수행한 후 그 결과를 그림 5에 나타내었다.

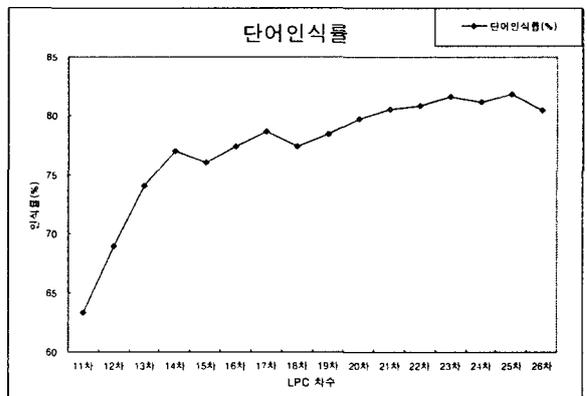


그림 5. 단어인식률(mel)

단어 인식 실험에서는 LPC 23차, 25차 일 때 81.68%, 81.87%로 인식률이 가장 좋음을 알 수 있었다.

4. 결 론

본 논문은 음소를 기반으로 하는 한국어 음성인식의 기초 연구로서 한국어에만 나타나는 경음에 대하여 음운학적, 음향학적 특성을 고찰하고 이를 기반으로 음성인식 실험을 수행한 후 그 결과를 분석하였다.

먼저, 음소 인식 실험에서 경음은 가장 저조한 인식결과를 나타내었고, LPC 차원의 증가에 따라 인식률이 증가하여 경음에 대해 보다 많은 LPC 분해능이 필요함을 알 수 있었다. 전체 음소인식률과 경음의 인식률은 각각 LPC 23차 일 때 46.1%, 34.11% 로 인식률이 가장 높게 나타났으며 단어 인식 실험에서는 LPC 23차, 25차 일 때 81.68%, 81.87%로 인식률이 가장 높게 나왔다.

이상의 음성인식 실험 결과로부터 한국어 음성인식에 있어서 한국어에만 존재하는 경음에 대한 좀 더 세밀한 특징 분석이 필요함을 알 수 있었다.

향후 이상의 결과를 바탕으로 학습과 평가용 음성 데이터를 증가시켜 연속음성인식 시스템의 성능 향상을 위한 연구를 계속 하고자 한다.

참 고 문 헌

[1] 신 지영, "말소리의 이해 -음성학음운론 인구의 기초를 위하여," 한국 문화사, 2000.
 [2] 신 지영, "말소리의 이해 -음성학음운론 인구의 기초를 위하여," 한국 문화사, 2001.
 [3] 강 옥미, "한국어 음운론," 태학사, pp49-63, 2003
 [4] 서 민경, "한국어 과열음의 VOT에 관한 실험음성학적 연구 - 환경에 따른 VOT 변이를 중심으로 -", 서울대학교 석사 학위 논문, 2002.
 [5] 배 재연, 신 지영, 고 도홍, "음성 환경에 따른 한국어 패쇄음의 음향적 특성: 시간적 특성을 중심으로" 제6회 음성 과학회 학술 발표대회, 1999.
 [6] 이 정훈, 정 현열, "단어 인식을 위한 환경 적응화에 관한 연구", 제 13회 음성통신 및 신호처리 워크샵 논문집, pp.50-54, 한국음향학회, 1996.8.
 [7] 김 주권, 김 범국, 정 호열, 정 현열, "한국어 음성 인식

시스템을 위한 MEL-LPC 분석 방법과 LPC-MEL 분석 방법의 비교," 신호처리 합동학술대회 논문집 제 14권 1호, pp 833-836, 2001.9

[8] Chin-Wu Kim "A Theory of Aspiration, *Phonetica* 21, pp 107-116, 1970.

[9] Rabiner, Juang, "Fundamentals of speech recognition" Inc, 1993

[10] 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.