

유·무성음 척도를 포함한 재구성 특징 파라미터의 음성인식 성능평가

이광석* · 한학용** · 고시영*** · 허강인**

Performance Evaluation of Speech Recognition Using the Reconstructed Feature Parameter with Voiced-Unvoiced Measure

Gwang-Seok Lee* · Hag-Yong Han** · Si-Young Koh*** · Kang-In Hur**

요 약

본 연구는 유사음에 강인한 음성인식을 위하여 음성의 유·무성음 척도를 특징 파라미터에 추가 구성하여 음절과 음소단위의 음성인식을 행하였다. 이를 위하여 피치검출에 이용되는 알고리즘인 HPS(Harmonic Product Spectrum)의 스펙트럼 정보를 이용하여 유·무성음의 정도를 나타내는 척도를 제안한다. 제안된 척도는 HPS의 첨도와 피크의 개수 그리고 높이척도이다. 이들 척도 값을 포함하여 특징 파라미터를 재구성하고 제안된 특징의 유효성을 검증하기 위하여 CVC형 유사 음절 DB하에서 기존 특징 파라미터와 비교하여 음성인식 실험을 행하였다.

ABSTRACT

In this study, we research the robust speech recognition for the syllables and phoneme units with the feature parameter including the voiced-unvoiced measures for the confusable words. In order to make it possible, we propose the measure representing the voiced-unvoiced degree by using the HPS(Harmonic Product Spectrum) information, used on pitch detection. We proposed this measures with the sharpnes, peak count and height measure of HPS. We reconstructed the feature parameter including this measures, then we performs the speech recognition experiments and compared with the typical feature parameters under the CVC type confusable syllables DB.

키워드

Speech Recognition, HPS, Feature Parameter, Voiced, Unvoiced

1. 서 론

현재 대부분의 음성인식 시스템에서는 음성의 스펙트럴 특징에 인간의 청각 특성을 반영한 특징들이 널리 사용되고 있다. 그러나 실제 음성인식 시스템에서 이 특징을 사용할 경우, 주변 환경이나 화자의 변이 등과 같은 음향학적 요소들이 대부분 오인식의

원인이 되고 있다.

본 논문에서는 기존 음성인식에서 많이 사용하고 있는 특징 파라미터인 MFCC에 음성의 유·무성음 정도를 나타내는 척도 정보를 추가하여 새로운 특징 파라미터를 재구성하고 음성인식에 적용하여 실험을 행하였다.

이를 위하여 구현하기가 비교적 용이하고 잡음환

*전주산업대학교 전자공학부 부교수

**경일대학교 전자정보공학부 교수

**동아대학교 전기·전자·컴퓨터공학부 교수

접수일자 : 2003. 3. 20

경 하에서 비교적 강인한 피치검출 알고리즘인 HPS(Harmonic Product Spectrum)의 스펙트럼 정보를 이용하여 일차적으로 음성의 유·무성음의 정도를 추출한다. 얻어진 스펙트럼 정보를 음성 특징 파라미터화 하기 위하여 적절한 척도를 사용하여 유·무성음의 정도를 얻어내고 이를 기존의 특징 파라미터에 추가하여 재구성한다. 이러한 유·무성음 정보는 음성 분석적인 측면에서 뿐 만 아니라 그 정확한 구간이 결정될 경우 음성인식의 전처리 단계에서도 이용될 수 있다.

한편, 음성에 대한 연구가 발전함에 따라 분할과 라벨링된 음성 DB에 대한 필요가 증가하고 있다. 음성을 일정 세그먼트의 연결로 가정하고 처리되는 음성인식 시스템의 경우, 인식단위인 음소, 다이폰 그리고 트라이폰 단위로 정확하게 분할되고 라벨링 된 음성 DB는 인식기의 성능에 결정적인 영향을 미치게 된다. 현재, 음소 단위의 음성인식이 음성의 모델의 수가 적어 많은 이점이 있는데도 불구하고 이러한 분할구간의 자동결정과 자음의 인식률의 저하로 실제 적용에서 문제점으로 지적되고 있다. 본 논문에서는 제한한 재구성 파라미터를 이용한 음소단위의 음성인식기의 가능성을 확인하기 위하여 음성인식의 단위를 음절단위와 음소단위로 나누어 실험을 각각 행하였다.

2장에서 적용된 HPS알고리즘과 사용한 척도를 소개하고 3장에서는 실험 및 결과 그리고 4장에서 결론을 맺는다.

II. HPS 알고리즘

1. Harmonic Product Spectrum

음성의 유·무성음 판별은 기본적으로 유성음은 음원으로 준 주기적인 임펄스 열이 존재하는 반면, 무성음의 경우 랜덤한 잡음을 음원으로 하고 있다는 가정 하에 처리가 이루어진다. HPS 알고리즘은 음성의 피치를 찾기 위한 방법중의 하나로 NOLL이 제안하였다^[1]. 이 알고리즘은 유성음의 진폭 스펙트럼은 기본주파수의 정수배에 해당하는 배음이 존재하고 주기신호에서 정수 배로 주파수 스케일로 압축하게 되면 배음 성분들이 기본주파수와 그 근방에서

일치하게 된다는 사실에 기반한 알고리즘이다^[2].

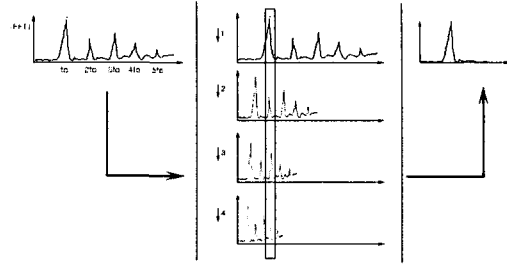


그림 1. HPS 알고리즘의 개관

Fig. 1 Overview of HPS algorithm

또한, 주기신호의 진폭 스펙트럼은 배음간에는 영의 값을 가지므로 압축된 진폭 스펙트럼의 곱은 기본주파수를 제외한 나머지 배음에 대해서는 제거되는 효과를 나타내게 된다. 완전한 배음을 가진 이상적인 경우에는 HPS가 그림 1과 같이 기본주파수에서만 높은 피크치를 나타내게 된다. HPS를 이용한 피치 추출은 이러한 성질을 이용한 것으로 HPS $P(\cdot)$ 는 다음의 식(1)과 같이 정의된다.

$$P(n) = R \sqrt{\prod_{r=1}^R |X_n(e^{jn\Delta\omega})|^2} \quad (1)$$

그러나, 음성은 단구간 푸리에 변환에 근거하고 유성음이 준주기적인 성질을 나타내므로 HPS의 기본주파수를 기준한 배음 성분간에 반드시 영의 값을 가지지 않는다는 특징이 있으나 의사적인 피치 추출은 가능하며 실제 음성에 적용된 예는 그림 2에 나타내었다.

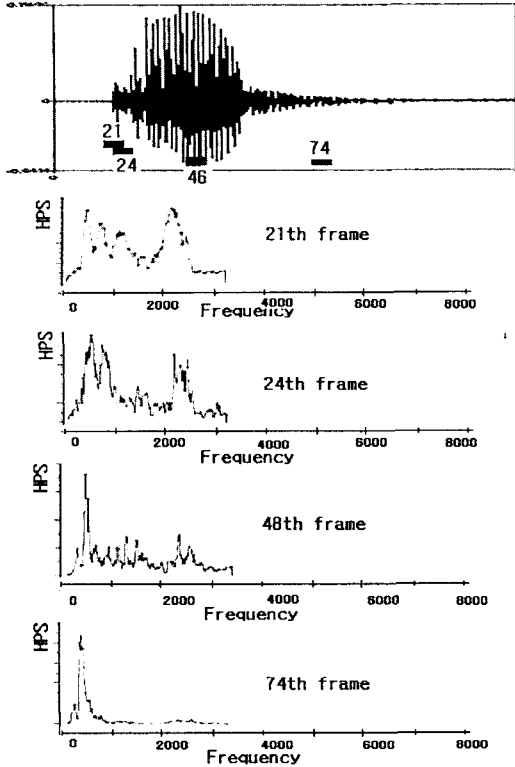


그림 2. 음성 “간”에 적용되어 단구간 분석된 HPS
Fig. 2 Short time analyzed HPS applying speech “간”

III. 유·무성음 정도 측정을 위한 척도

단구간마다 관측되는 HPS의 스펙트럼 분포에서 유·무성음을 판별하고 유·무성음의 정도를 알아내기 위한 척도를 사용한다. 사용된 척도는 Zolnay가 제안한 높이척도^[3]와 본 연구에서 제안하는 척도와 피크의 개수 척도이다.

3.1. 첨도(Sharpness)

첨도는 자료집단의 분포에서 가운데 봉오리 부분이 얼마나 뾰족한가를 측정하는 단위이다. 자료 값이 평균값을 중심으로 부드럽게 나타나는가, 아니면 첨예하게 나타나는가를 표현하는 단위로 식(2)과 같이 정의되며 정규분포의 경우의 값은 3이 된다.

$$\text{Sharpness} = \frac{E[(X - \bar{X})^4]}{\sigma^4} = \frac{\left[\frac{1}{N-1} \sum_{\phi=1}^N (x_{\phi} - \bar{x})^4 \right]}{\sigma^4} \quad (2)$$

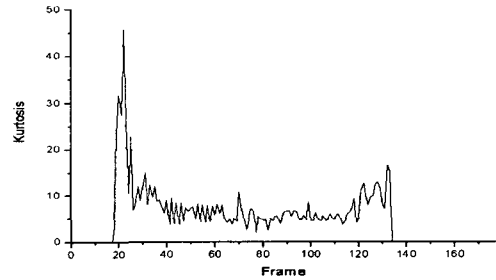


그림 3 HPS 첨도 척도값 분포(“간”)
Fig. 3 Distribution of HPS sharpness measure(“간”)

3.2. 피크의 수

HPS의 스펙트럼의 피크의 개수는 음성의 유·무성음에 따라 많은 차이가 생기며 무성음의 경우, 많은 피크의 개수가 존재하는 반면, 유성음의 경우, 피크의 수는 상대적으로 적게 된다.

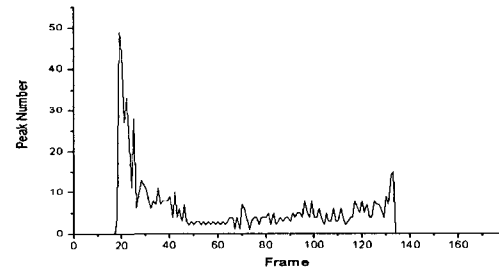


그림 4. HPS 피크 개수 척도값 분포(“간”)
Fig. 4 Distribution of HPS peak number measure(“간”)

3.3. 높이 척도[3]

높이 척도는 HPS의 피크의 진폭만을 고려한 것으로 주파수 n_{max} 에서의 최대 진폭 값과 그 주위의 진폭 값의 기하학적 평균과의 비율로 식(3)과 같이 정의한다.

$$v_{height} = \frac{P(n_{max})}{2w \sqrt{\prod_n P(n)}} \quad (3)$$

여기서, n_{max} 는 최대진폭의 위치이고 n_{max} 를 포함하여 $n_{max} - W$ 에서부터 $n_{max} + W$ 까지에 해당되는 n_{max} 의 근방에 있는 n 개의 값들의 곱을 구한다. 근방의 진폭 값의 개수는 첫 번째 배음의 피크치가 평균에 포함되어지는 것을 피하기 위하여 적절한 값이 선택되어진다. 본 연구에서는 W 의 값은 10으로 설정하였으며 무성음에서 이 척도 값이 두드러지게 하기 위하여 높이척도의 역수를 취하여 이용하였다.

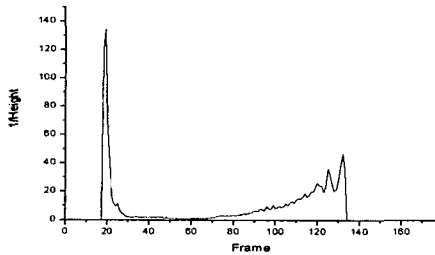


그림 5 HPS 높이 척도값 분포(“간”)
Fig. 5 Distribution of HPS height measure(“간”)

IV. 실험 및 결과

4.1. 음성분석조건 및 DB

제한한 재구성 특징 파라미터의 유효성을 검증하기 위한 실험을 위하여 CVC형 108음절 DB를 구성하였다. 본 DB는 우리말 음성의 CVC형 음절로 초성(비, 디, 기, 피, 티, 키), 중성(나, 니, 니, 키, 디, 리), 종성(ㄴ, 르, 모)으로 이루어진 108개의 유사 음절로 구성되어 있으며 5명의 화자가 5회 발성하여 3회분은 학습용으로 나머지 2회분은 평가용으로 이용하였다. 목적에 의하여 초성, 중성, 종성을 분리한 음소 DB를 별도 구성하여 음절과 음소단위의 인식 실험을 행하였다. 음성분석조건은 표 1과 같다.

표 1. 음성분석조건

Table 1. Speech data analysis conditions

항 목	분석 조건
A/D	16 [kHz], 16 bit
Filtering	LPF, 7 [kHz]
Step Size	60 points
Window Length	256 points

4.2. 인식실험 및 결과

인식실험은 연속출력분포 HMM의 상태수를 5로 하고 혼합수 1로 한 모델에서 실험을 행하였다. 사용된 특징 파라미터는 표 2와 같다. 모든 특징의 차원은 10차원이며 이들 중 인식률이 상대적으로 높은 MFCC의 경우는 10차원과 14차원에서의 실험과 유·무성음 척도인 첨도와 피크의 개수, 높이척도를 각각 추가하여 11차원과 15차원으로 재구성하여 실험을 행하였다.

표 2. 특징 파라미터

Table. 1 Applied characteristic parameters

Feature	Description
LPC	Linear Prediction Coefficients
LPCC	LPC Cepstrum
LPCMC	LPC Mel Cepstrum
PLP	Perceptual Linear Prediction
LFCC	Linear Frequency Cepstrum Coefficient
MFCC	Mel Frequency Cepstrum Coefficient
MFCC+K	MFCC + HPS's Kutosis
MFCC+P	MFCC + HPS's Peak Count
MFCC+H	MFCC + HPS's Height Measure

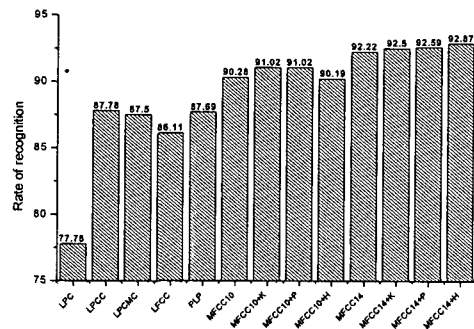


그림 6. CVC형 108음절 인식실험결과(평가용)
Fig. 6 Recognition results of CVC type's 108 syllable

V. 결론

본 논문에서는 HPS를 이용하여 음성의 유·무성음 정도를 나타내는 몇 가지 척도 값을 기존의 특징 파라미터에 포함하여 재구성된 특징 파라미터로 인식실험을 행하였다. 제안된 유·무성음 판별 방법은 음성인식의 전처리로도 이용할 수 있는 기법으로 정량적인 임계치를 결정하기는 어려우나 상대적인 유·무성음의 정도는 비교적 정확하게 나타냄을 확인하였다. 그리고 이를 이용한 재구성된 특징 파라미터가 다른 파라미터와 비교할 때 상대적으로 인식률이 향상되었다.

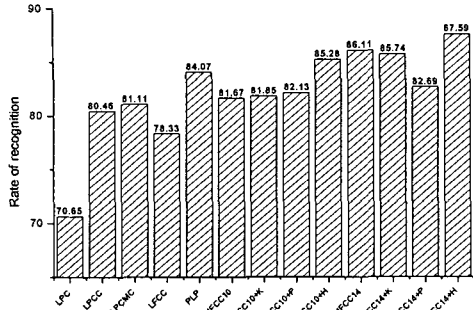


그림 7. 초성 음소 실험결과(평가용)
Fig. 7 Recognition results of an initial phoneme

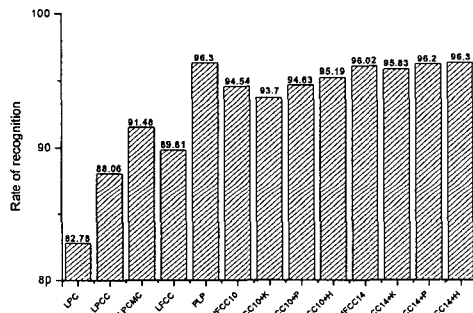


그림 8. 중성 음소 실험결과(평가용)
Fig. 8 Recognition results of the 2nd phoneme

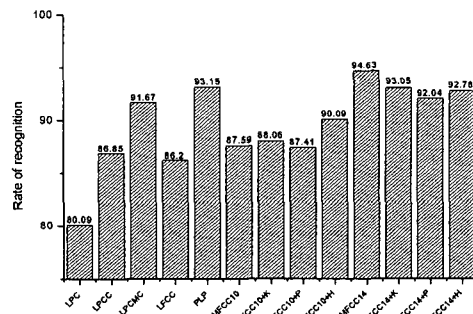


그림 9. 종성 음소 실험결과(평가용)
Fig. 9 Recognition results of the final phoneme

참고문헌

- [1] Noll, M., "Pitch determination of human speech by the harmonic product spectrum, the harmonic sum spectrum, and a maximum likelihood estimate," In Proceedings of the Symposium on Computer Processing Communications, pp. 779-797, Polytechnic Institute of Brooklyn.
- [2] L. R. Rabiner, R. W. Schafer : "Digital-processing of speech signals," Prentice Hall.
- [3] Andras Zolnay, Ralf Schluter, Hermann Ney, "Robust speech recognition a voiced-unvoiced feature" Vol.2, pp. 1065, ICSLP, 2002.
- [4] <http://www-ccrma.stanford.edu/~pdelac/PitchDetection/icmc01-pitch.pdf>.
- [5] Lawrence Rabiner, Bing-Hwang Juang, "Fundamental of speech recognition," Prentice Hall, 1993.

저자소개



이광석(Gwang-Seok Lee)

1983년 2월 동아대학교 전자공학과 공학사

1985년 2월 동아대학교 전자공학과 공학석사

1992년 2월 동아대학교 전자공

학과 공학박사

1995년 9월~현재 진주산업대학교 전자공학과 부교수

※관심분야: 음성 신호처리 및 인식, 생체 신호처리, 지능화기술



허강인(Kang-In Hur)

1980년 2월 동아대학교 전자공학과 공학사

1982년 2월 동아대학교 전자공학과 공학석사

1990년 8월 경희대학교 전자공학과 공학박사

1984년~현재 동아대학교 전기·전자·컴퓨터공학부 교수

1988년 9월~1989년 8월 일본筑波大學 객원연구원

1992년 9월~1993년 8월 일본豊橋大學 객원연구원

※관심분야: DSP, 음성인식 및 합성, 신경회로망



한학용(Hag-Yong Han)

1994년 2월 동아대학교 전자공학과 공학사

1994년 3월~1997년 5월 (주)경남에너지 근무

1998년 2월 동아대학교 전자공학과 공학석사

2001년 2월 동아대학교 전자공학과 박사과정 수료

2001년 3월~현재 (주)이지하모니 부설기술연구소장,

동명정보대학교 정보공학부 겸임교수

※관심분야: 신호처리, 패턴인식, DSP응용



고시영(Si-Young Koh)

1979년 2월 영남대학교 전자공학과 공학사

1983년 2월 영남대학교 전자공학과 공학석사

1992년 8월 동아대학교 전자공학과 공학박사

1986년~현재 경일대학교 전자정보공학과 교수

※관심분야: 음성 신호처리, 생체신호처리