韓國 OA學會 論文誌
第7卷 第3號, 2002. 9.

2002-7-3-2-3

퍼지 이론을 이용한 한국어 및 영어 화자 인식에 관한 연구

김 연 숙* 김 희 주** 김 경 재***

A Study on Korean and English Speaker Recognitions using the Fuzzy Theory

Yeoun-sook Kim* Hee-joo Kim** Kyoung-jae Kim***

요 약

본 논문에서는 피치 파라미터와 퍼지를 포함한 화자 인식 알고리즘을 제안한다. 음의 시간적인 특징을 이용하여 시간 영역에서 분해력을 높이고 주파수 영역에서 잡음에 강인함을 갖는 국부 봉우리와 골에 의한 피치 검출법을 제안하여 피치를 검출한다. 또한 화자 인식에서 음성 신호의 애매성을 보완할 수 있는 퍼지의 소속함수를 이용하여 표준 패턴을 작성하고 퍼지 패턴 매칭을 이용하여 인식을 수행한다.

Abstract

This paper proposes speaker recognition algorithm which includes both the pitch parameter and the fuzzy. This study proposes a pitch detection method for the peak and valley pitch detection function by means of comparing spectra which utilizes the transform characteristics between time and frequency. It measures the similarity to the original spectrum while arbitrarily varying the period in the time domain. It heavily weights the error due to the changing characteristics of the phonemes, while it is strong against noise.

In this paper, makes reference pattern using membership function and performs vocal track recognition of common character using fuzzy pattern matching in oder to include time variation width for non-linear utterance time.

* 건국대학교 전자공학과
** 강원관광대학
*** 홍익대학교

논문접수 : 2002. 6. 20
심사완료 : 2002. 9. 14

Ⅰ. 서론

오늘날 세계 각국은 지식기반사회에서 과학기술의 발달로 인해 엄청난 속도로 서로 가까워지고 있다. 특히 세계 공용어로 사용되고 있는 영어는 전 세계 각국이 정보 교환을 하기 위해 사용하고 있어 우리도 세계적인 최신 정보를 획득하기 위해 영어를 익힐 필요성이 있다.

한국어에 있는 모음과 자음과 같이 영어에도 모음과 자음이 있다. 모음과 자음의 차이는 시간 파형에서 쉽게 관찰할 수 있다. 화자 인식 연구는 1963년 Bell 연구소의 Pruzansky가 화자 식별 실험을 하였고, 1974년 Bell 연구소의 Atal은 화자 식별 및 화자 확인 실험을 하였으며, 1981년 Bell 연구소의 Furui는 텍스트 의존 화자 인식 실험을 하였다. 1981년 Bell 연구소의 Atal은 화자 식별 및 화자 확인 실험을 하였다. 국내에서는 1989년 이혁재가 결정 함수 개념을 도입하여 화자 인식 실험을 하였으며, 1991년 권석규는 이혁재의 연구 결과에 DSP 칩을 사용하여 H/W 설계를 하였다.

음성 인식에 사용하는 발음은 단어의 수에 제한을 받지 않아야 실용성과 일반성을 가질 수 있다.[1][2][3]

이러한 문제는 화자 인식에 사용되는 파라미터들을 통계적으로 추출, 적용함으로써 해결할 수 있다. 따라서 화자의 개성을 대변해 줄 수 있는 새로운 파라미터의 제안은 음성 인식에서 해결되어야 할 중요한 과제이다.[1]

본 논문에서는 시간 영역에서 검출 알고리즘에 대해 분해력을 높이고 주파수 영역에서 잡음에 강인한 국부 봉우리와 골에 의한 피치 검출법을 제안하여 피치를 검출하고 시간 변동의 폭을 모두 포함할 수 있도록 음성 신호의 애매성을 보완할 수 있는 퍼지의 소속 함수를 이용하여 표준 패턴을 작성하고 퍼지 패턴 매칭을 이용하여 인식을 수행한다.

본 논문은 다음과 같이 구성되어 있다. 제 2장에서는 음성학적 분석과 전 처리 과정 및 분석을 설명하고, 제 3장에서는 국부 봉우리와 골에 의한 피치 검출법에 대해 살펴보기로 한다. 제 4장에서는 피치 검출과 퍼지 이론을 이용한 화자 인식을 살펴보기로 하고 제 5장과 제 6장에서는 본 논문에서 제안한 내용과 실험을 평가하고 결론을 맺는다.

Ⅱ. 음성학적 분석과 전 처리 과정 분석

1. 음성학적 분석

그림 1은 음성이 생성되는 모형도로 음성을 생성하는 에너지원인 허파, 기관지와 호흡 기관부터 음성이 방사되기까지를 나타낸 것이다. 음성은 간단히 공기가 허파로부터 방출되고 결과적으로 성도에 있는 협착점에 의해 공기가 동요될 때 이 시스템으로부터 방사되는 음향학적 파형이다.

시스템 모델 링의 관점에서 볼 때 조음기관들이 음성 시스템 필터의 성질을 결정한다. 이러한 공진이 전체 스펙트럼 모양을 만드므로 음성 학자들은 이를 포만트라 부른다. 원칙적으로 주어진 음에는 무한개의 포만트가 있으나, 실제로 샘플링(sampling) 후에 Nyquist 대역(일반적으로 10㎑로 샘플링하여 5㎑로까지 나타냄)에서 3~5개를 발견할 수 있다.

2. 전 처리 과정

시간 영역 측정법에 의하여 음성신호를 표현하는 방법에는 평균 영 교차율, 자기 상관관계 함수 등이 있다.

음성신호의 짧은 구간을 분리해서 마치 정체된 특징을 갖는 연속적인 소리의 짧은 구간으로 처리하는 단시간 처리 기법은 식 1과 같이 나타낼 수 있다.

$$Q_n = \sum_{m=-\infty}^{\infty} T[x(m)]w(n-m) \quad (1)$$

음성신호(필요한 주파수 대역을 추출하기 위해 선형 필터를 통과시킨 후)는 선형 또는 비선형이 되는 전달함수 T[ ]를 필요로 하고 어떤 조정 가능한 변수 또는 변수 쌍에 의존하게 된다. Q_n값은 시컨스 T[x(m)]의 국부적으로 가중치가 적용된 평균값 시컨스이다.

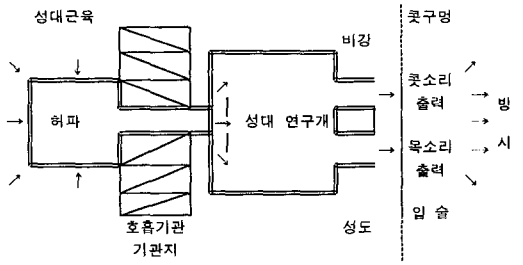


그림 1. 음성 생성 시스템의 모형도
Fig. 1 Schematized diagram of the vocal system

이산 시간 신호에서는 연속적인 샘플의 부호가 서로 다를 때에 영교차가 발생한다. 예를 들어, Fs 비율로 샘플링된 F0 주파수의 정현 신호는 정현파의 사이클당 Fs/F0 샘플을 갖는다. 각 사이클은 2개의 영교차를 갖고 있기 때문에 긴 시간의 평균 영 교차율은 식 2와 같다.

$$Z = 2F_0 / F_s \quad \text{〔교차/샘플〕} (2)$$

주파수가 높다는 것은 영 교차율이 높다는 것을 의미하고 주파수가 낮다는 것은 영 교차율이 낮다는 것을 나타내기 때문에 영 교차율과 주파수에 따른 에너지 분포 사이에는 강한 상관관계가 존재한다.

이산 시간 결정론에서 신호가 불규칙하거나 주기적이면 자기 상관관계 함수는 식 3과 같이 정의된다.

$$\phi(k) = \lim_{N\to\infty} \frac{1}{2N+1} \sum_{m=-N}^{N} x(m)x(m+k) \quad (3)$$

자기 상관관계 함수의 특성은 다음과 같다.

1. 우함수 이다. 즉, $\phi(k) = \phi(-k)$
2. k = 0에서 최대가 된다. 즉, $|\phi(k)| \le \phi(0)$
3. $\phi(0)$값은 결정론 신호의 에너지 또는 불규칙하거나 주기적인 신호에 대한 평균 에너지와 같다.[4]

Ⅲ. 국부 봉우리와 골에 의한 피치 검출법

본 논문에서는 음의 시간적인 특징에 대해 국부적으로 봉우리와 골을 이룬다는 것을 이용하여 계산량이 적고 잡음에 강인한 피치 검출법을 제안하였다.

피치를 정확히 검출할 수 있다면 음성 인식에 있어서 화자에 따른 영향을 줄일 수 있기 때문에 인식의 정확도를 높일 수 있고, 음성 합성시 자연성과 개성을 쉽게 변경하거나 유지할 수 있다.[5]

제안한 국부 봉우리와 골에 의한 피치 검출법의 구성도를 그림 2에 나타내었다.

국부 봉우리와 골에 의한 피치 검출의 결과는 그림 3에서 처럼 얻을 수 있다.

그림 3 a)는 한국어 모음의 한 프레임이고, b)는 저역통과 여파기를 통과한 결과이며 c)는 1차 봉우리와 골을 검출한 결과이고, d)는 2차 데시메이션을 통해서 걸려진

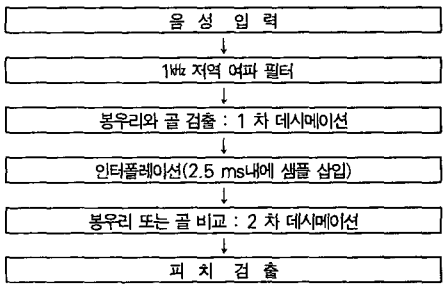


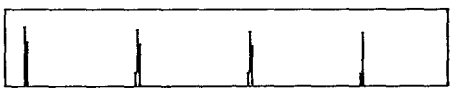
그림 2. 제안한 국부 봉우리와 골에 의한 피치 검출법의 구성도
Fig. 2 Block diagram of proposed pitch detection method

a) 원 신호

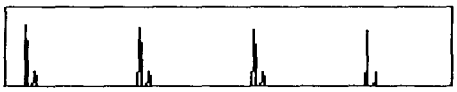
b) 저역 통과 여파기를 통과한 신호

c) 검출된 봉우리와 골

d) 클리프된 신호



e) 인터폴레이트된 신호



f) 검출된 피치

그림 3. 국부 봉우리와 골에 의한 피치 검출 알고리즘의 처리 과정
Fig. 3 Disposal processing of pitch detection algorithm

봉우리와 골에 대한 결과이다. 그리고 e)는 1차 인터폴레이션을 수행한 것이고, f)는 이것을 다시 결정 논리를 통해서 피치를 결정한 결과이다.

Ⅳ. 피치 검출과 퍼지 이론을 이용한 화자 인식

음성신호의 피치 검출을 위해서는 퍼지 추론의 생성 규칙을 이용하여 특징 파라미터에 대한 소속도 함수를 구하고 퍼지 집합을 생성시켜야 한다.

IF 음성 신호내 모음을
  선형 필터된 후의 예비 피치 성분의 주파수 에너지 Ei가 퍼지값 X_{Ei}를 갖고, 예비 피치 index Pi가 퍼지 값 Xpi를 갖는다면
THEN
  음성신호 "/아/, /에/, /오/, /우/, /이/ "이다.

각 화자가 발성한 모음에 대해서 국부 봉우리와 골을 이용한 피치 검출법에서 얻은 피치 주파수를 사용하여 퍼지 집합을 형성한다. 이것을 예비 피치(pre-pitch)라고 정의한다.

표 1은 예비 피치 주파수 특징량을 나타낸 것이다. 여기서 주파수 값을 100개의 퍼지 값으로 나타낸 것은 실제 주파수 스펙트럼 상에 존재 가능한 영역을 주파수와 개수로 나누어 사용한 것이다.

표 1. 예비 피치 주파수 특징량
TABLE. 1 A fuzzified features of pre-pitch frequencies

예비 피치 주파수	퍼지 값
1 - 10	0
11 - 20	1
:	:
991 - 1000	99

퍼지 이론에 의한 화자 인식은 화자가 발성한 음성에 대해 FFT를 수행한 후 행렬 양자화 인덱스와 각 주파수의 스펙트럼 양자화는 주파수를 채널로 나누어 각각의 중심 주파수에 해당하는 에너지에 대해 0.1dB 마다 퍼지 값을 주어 대응시킨다. 스펙트럼 에너지는 입력 음성에 대해 시간 영역에서 주파수 영역으로 변환하기 위해서 FFT를 수행하고, 예비 피치에서 얻은 값을 사용하여 주파수 저역 여파를 수행한다.

예비 피치에 대한 예비 피치 주파수 특징량의 퍼지화를 표 2에 나타내었다.

표 2. 예비 피치 에너지 특징량의 퍼지화
TABLE. 2 A fuzzified features of pre-pitch energy

퍼지 값 / 주파수(Hz)	표준 패턴		시험 패턴	
	퍼지값	에너지(dB)	퍼지값	에너지(dB)
대역 1 : 33	32	0.825	31	0.775
대역 2 : 66	30	0.750	29	0.750
:	:	:	:	:
대역 30 : 1000	12	0.300	11	0.275

확신도를 구하기 위하여 표준 패턴과 시험 패턴의 스펙트럼 양자화 값에 대한 퍼지 값의 소속도 함수 값을 구해야 한다. 여기서 두 패턴 사이의 확신도 Se(i)는 본 연구의 확신도 계산에 적합하도록 max-min 연산에 의해 구한다.

$$S_e(i) = \vee(\mu_{ei}\,\mathrm{ref} \wedge \mu_{ei}\,\mathrm{test}) \quad (4)$$

단, i = 1, 2, ... , N (i : i번째 채널)

표 3은 대역 1의 확신도를 나타낸 것이다.

표 3. 대역 1의 예비 피치 에너지에 대한 확신도 결과
TABLE. 3 Result of certainty factors of 1st band pre-pitch energy

확신도 / 퍼지값	표준 패턴의 소속도 함수	시험 패턴의 소속도 함수	확신도(1)
1	0.0	0.0	0.0
:	:	:	:
40	0.0	0.0	0.0

각 화자가 10번씩 발음한 한국어 단모음 /아/, /에/,

/오/, /우/, /이/ 를 사용해서 FFT 512 샘플을 구한다.

그림 4는 예비 피치를 사용하여 선형 디지털 필터 링을 수행한 후에 원래 스펙트럼과 LPF가 수행된 신호에 대한 스펙트럼의 차 신호를 구한 값이다. 소속도 함수 값은 퍼지화 패턴의 의미를 어느 정도 포함하고 있는가를 나타내는 것으로, 1.0에서 0.0사이의 값으로 표현한다. 입력된 음성에 대해서 한 프레임을 512단위로 피치를 구해 평균 피치를 구한 다음 전체 구간에 대해서 피치 패턴을 구한다. 또한 프레임별로 에너지를 구해 평균 에너지를 구한 다음 전체 에너지 패턴을 구한다.

피치를 검출하는 구성도를 그림 5에 나타내었다.

퍼지 이론을 사용할 표준 패턴과 시험 패턴의 주파수에 대한 퍼지 값의 소속도 함수를 구해야 하는데, 두 패턴 사이의 확신도 $S^c(i)$는 퍼지 추론의 합성 규칙을 적용

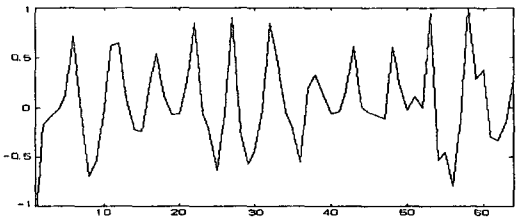


그림 4. 한국어 /아/의 예비 피치를 사용한 퍼지 집합 검출 과정
Fig. 4. Fuzzy detection using pre-pitch for Korean /a/

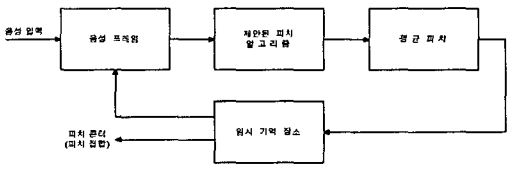


그림 5. 피치 콘터 검출의 블럭도
Fig. 5 Block diagram of pitch contour detection

하여 식 5와 같이 표현한다.

$$S^c(i) = \vee(\mu^{ci}\text{ref} \wedge \mu^{ci}\text{test}) \quad (5)$$

단, i = 1, 2, ... , n (i : 프레임 번호)

여기서,

$\mu ci\ ref$: i 번째 프레임 코드 북 인덱스에 대한 표준 패턴의 소속도 함수

$\mu ci\ test$: i 번째 프레임 코드 북 인덱스에 대한 시험 패턴의 소속도 함수

$S^c(i)$: i 번째 프레임 코드 북 인덱스에 대한 확신도 값

생성 규칙의 전제가 어느 정도 만족하는가를 추론하기 위해서 확신도 값을 모두 더하여 그 음성신호의 확신도를 사용하게 되면 식 6처럼 표현된다.

$$S^c{}_{TOTAL} = \sum_{i=1}^{n} S^c(i)$$

$$(i = 1, 2, 3, \cdots, n) \quad (6)$$

V. 실험 및 고찰

표 4는 본 논문에서 사용된 시료를 나타낸 것이다.

표 4. 한국어 및 영어 시료
TABLE. 4 Korean and English data

한국어 발음	아	에	오	우	이
한국어	아	에	오	우	이
영 어	a	e	o	u	i

표 5, 표 6은 LPC 켑스트럼을 사용했을 때의 인식율과 본 논문에서 제안한 국부 봉우리와 골에 의한 피치 검출법과 퍼지 추론을 사용했을 때의 인식율을 나타낸 것이다.

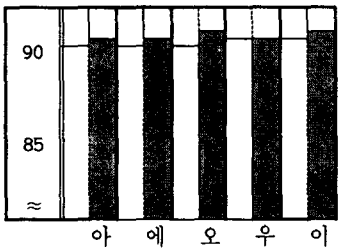
표 5. LPC 켑스트럼 방법과 제안된 방법의 한국인 화자가 발음한 모음 인식율 비교와 개선율
TABLE. 5 Recognition rate comparison and improvement rate of vowels for Korean speakers using LPC cepstrum method and proposed method

방법 / 시료		LPC 켑스트럼	제안된 방법	개선된 인식율
데이터	아	92 %	94 %	2 %
	에	93 %	95 %	2 %
	오	90 %	95 %	5 %
	우	94 %	94 %	0 %
	이	90 %	96 %	6 %
평 균		91.8 %	94.8 %	3 %

그림 6은 한국의 모음에 대해 LPC 켑스트럼 방법, 국부 봉우리와 골에 의한 피치 검출 방법과 퍼지 추론 방법을 에 대해 각각의 화자 인식율을 막대 그래프로 나타낸 것이다.

표 6. LPC 켑스트럼 방법과 제안된 방법의 미국인 화자가 발음한 모음 인식율 비교와 개선율
TABLE. 6 Recognition rate comparison and improvement rate of vowels for American speakers using LPC cepstrum method and proposed method

방법 / 시료		LPC 켑스트럼	제안된 방법	개선된 인식율
데이터	a	92 %	96 %	4 %
	e	92 %	96 %	4 %
	o	93 %	93 %	0 %
	u	94 %	94 %	0 %
	i	93 %	95 %	2 %
평 균		92.8 %	94.8 %	2 %



□ : LPC Cepstrum을 이용한 화자 인식
■ : 국부 봉우리와 골에 의한 피치 검출과 퍼지 추론에 의한 화자 인식

그림 6. 인식 방법별 한국어 모음 인식율 비교
Fig. 6 Comparison of recognition rate of Korean vowels according to methods

Ⅵ. 결론

본 논문의 특징은 인식 알고리즘으로 본 논문에서 제안한 국부 봉우리와 골에 의한 피치 검출법과 퍼지 추론을 사용하여 한국어 및 영어를 인식하는데 얼마만큼의 오인식율을 향상시킬 수 있는지에 대한 실험과 시뮬레이션을 수행하였다.

음성으로부터 특징을 추출하는 방법으로 FFT 방식, 선형 디지털 필터 방식 등을 사용하였으며, 예비 피치로부터 얻은 주파수를 사용하여 디지털 주파수 필터에 적용한다. 이렇게 주파수 필터를 사용하여 얻은 스펙트럼과 원래 신호의 차를 구하여 차 신호에 대해서 퍼지 집합을 구한다.

본 논문에서 제안한 국부 봉우리와 골에 의한 피치 검출법과 주파수 필터에 의한 퍼지 이론을 겸용하여 한국어 및 영어 화자 인식율이 개선되어 좋은 알고리즘임을 확인하였다. 앞으로 연구 과제는 실시간에서 실용화되도록 계속적인 연구가 이루어져야 할 것이다.

참고문헌

[1] Tetsunori Kobayashi and Hidetoshi Sekine, "STATISTICAL PROPERTIES OF FLUCTUATION OF PITCH INTERVALS AND ITS MODELING FOR NATURAL SYNTHETIC SPEECH", IEEE, CH2847-2/90/0000-0321, pp. 321-324, 1990.

[2] 김연숙, "피치 정보를 이용한 격리 단어 인식에 관한 연구", 한국학술진흥재단, SEPTEMBER 1995.

[3] Sungwook Chang, Y.Kwon, and Sung-il Yang,"Speech Feature Extracted from Adapted Wavelet for Speech Recognition", Electronics Letters, 1998.

[4] J. M. Baker, "A New Time-Domain Analysis of Human Speech and Other Complex Waveform", Ph. D Dissertation, Carnegie Mellon Univ., Pittsburgh, PA., 1975.

[5] Kai-Fu Lee, "AUTOMATIC SPEECH RECOGNITION", KLUWER ACADEMIC PUBLISHERS,Boston/Dordreecht/London, 1989.

저 자 소 개

김 연 숙

1983 아주대학교 일반대학원 전자공학과(공학석사)
1998 건국대학교 일반대학원 전자공학과(공학박사)
1984-1994 서울동덕여고
1994-2001 서울산업정보고
현재 상봉중학교

김 희 주

1987 연세대학교 교육대학원 (교육학석사)
1995 성신여자대학교 일반대학원 식품영양학과(이학박사)
현재 강원관광대학 교수

김 경 재

홍익대학교 일반대학원 건축학과(공학석사)
1999 홍익대학교 일반대학원 건축학과(공학박사)