

## 인지 가중 필터를 이용한 화자 인식의 성능 향상에 관한 연구

<sup>o</sup>배재욱, 오세영, 이향원, 배명진  
숭실대학교 정보통신공학과

### On a Study of the Improvement of Speaker Recognition with Perceptual Weighting Filter

Jaeok Bae, Seyoung Oh, Hyangwon Lee and Myungjin Bae

Dept. of Information and Telecommunication Engr., Soongsil University

#### 요 약

화자 인식의 방법에서 사용되고 있는 특징 파라미터들은 음성 인식에서 사용되고 있는 특징 파라미터를 그대로 사용하고 있다. 따라서 이를 화자 인식에 적용할 때 화자의 특성을 효과적으로 반영할 수 있어야 한다. 일반적인 화자의 특징이 고주파수 위주로 분포되어 있기 때문에 전체 스펙트럼의 고주파 영역을 강조시킬 수 있고, 또한 인간의 청각특성이 공진 주파수에 기반하여 이루어진다는 사실에 기반을 두어서 공진 주파수 위주로 강조시키는 인지 가중 필터를 인식단의 전처리로 사용하는 방법에 관한 것이다. 본 논문을 실험한 결과 전체 인식율에서는 기존의 방법보다 3.89%까지 인식율의 향상을 얻을 수 있었다. 또한 사칭자 수리율은 2.5%의 저하를 얻을 수 있었다.

#### 1. 서 론

화자 인식은 음성의 특성을 이용해서 화자의 신원을 확인하는 기술이다. 이러한 기술은 등록된 화자 집단중 화자를 식별하는 화자 식별(speaker identification)과 지금 요청한 화자를 확인하는 화자 확인(speaker verification)이 있다. 이러한 방법들의 가장 확실한 차이는 화자 확인에서 사용되는 문턱값의 차이이다. 그러나 실제 경우에 있어서 높은 화자 식별율이 높은 화자 확인율을 얻게 하고, 그 반대의 경우도 마찬가지이기 때문에 이러한 차이점은 모호

해진다. 이러한 화자 인식은 음성에 내재되어 있는 화자정보를 추출하여 개인을 확인하는 기술로 전화망을 통한 서비스가 확산되어 가고 있는 현대사회에 가장 효과적인 기술중 하나이다. 그러나 현재 국내의 화자 인식에 대한 연구는 외국에 비해 거의 진무한 상태이며, 아직까지 연구의 단계에서 실용화 단계에 못미치고 있다.

본 논문은 기존의 화자 인식보다 인식율을 향상시키는 방법에 관한 것으로 다음과 같은 특성을 이용한다. 일반적으로 화자의 특성은 저주파 특성보다 고주파수 특성을 더 많이 가지고 있기 때문에 고주파수 특성을 강조시켜서 화자를 인식해야 한다. 그러나 고주파에는 화자의 특성뿐만 아니라 일반 배경잡음 성분도 많이 분포되어 있기 때문에 고주파수의 모든 영역을 강조시키는 것보다 사람의 청각특성에 민감한 공진 봉우리 위주로 강조하고 시키는 방법이 효과적이다. 따라서 본 논문에서는 저주파수 영역의 공진 봉우리는 감쇄시키고 상대적으로 고주파 영역의 공진 봉우리는 강조시켜서 전체 스펙트럼의 포락선이 고주파 영역 위주로 강조해 주는 인지 가중 필터를 화자 인식단의 전처리로써 사용하는 것에 관한 논문이다.

#### 2. 프리엠퍼시스 필터

음성신호에서는 음성의 의미를 내포하고 있는 음향학적 특성은 저주파수 영역에 나타나고 있지만 화자의 특징은 고주파수 영역에 많이 분포되어 있다. 따라서 화자 인식의 전처리단에서 고주파수 영역을

강조시키기 위해 사용되는 프리엠퍼시스 필터는 다음과 같은 특성을 가지고 있다[1].

$$H(x) = 1 - ax^{-1}, \quad 0.9 \leq a \leq 1 \quad (2.1)$$

그림 2-1은  $a$  값에 따른 프리엠퍼시스 필터의 전달함수의 특성곡선을 보여주고 있다. 그림 2-1에서 보이는 것처럼 이 필터는 스펙트럼적으로 저주파수 영역을 감쇄시키고 고주파수 영역을 강조시킴으로써 신호를 평탄화시키는 특성을 가지고 있다.

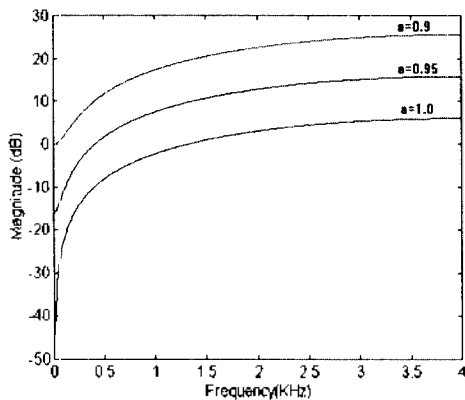


그림 2-1.  $a$  값에 따른 프리엠퍼시스 필터의 시스템 전달 함수

프리엠퍼시스 필터는 고주파수 영역의 신호성분을 모두 강조시키기 때문에 인지에 중요한 특성을 보이고 있는 공진 붐우리 성분뿐만 아니라 인식에 불필요한 성분들도 강조하게 되는 결과를 보이게 된다. 또한 프리엠퍼시스 필터에 통과된 음성 스펙트럼이 완전히 평탄화 되지 못하기 때문에 인식시 저주파수의 영향이 많이 지배하는 특징을 가지고 있다.

따라서 본 논문에서는 음성 신호의 스펙트럼 성분을 고주파수 성분을 더욱 강조시키기 위해서 인지 가중 필터를 인식 시스템의 선처리단으로 구성시키는 방법에 관한 것이다.

### 3. 인지 가중 필터

인지 가중 필터의 전달특성은 다음과 같다[2].

$$W(z) = \frac{A(z)}{A(\gamma z)} = \frac{1 - \sum_{j=1}^n a_j \beta^j z^{-j}}{1 - \sum_{j=1}^n a_j \alpha^j z^{-j}}, \quad 0 \leq \gamma \leq 1 \quad (3.1)$$

그림 3-1에서는 원래 음성 파형의 스펙트럼과 인지 가중 필터의 스펙트럼을 보여주고 있다. 이 때 사용된  $\alpha$ 는 0.2,  $\beta$ 는 0.5이다.

그림 3-1에서 보이듯이 인지 가중 필터의 모양을 원 스펙트럼에서  $x$ 축으로 대칭된 형태를 가지고 있고 주파수 측면에서 보면 저주파수 영역에서 공진 주파수 위주로 감쇄시키는 구조를 갖지만 고주파수 영역에서는 강조시키는 특성을 가지고 있다.

본 논문에서는 프리엠퍼시스 필터에 통과된 음성 신호에서 인지 가중 필터를 거쳐서 얻은 특징 벡터를 화자 인식에 사용하고자 한다.

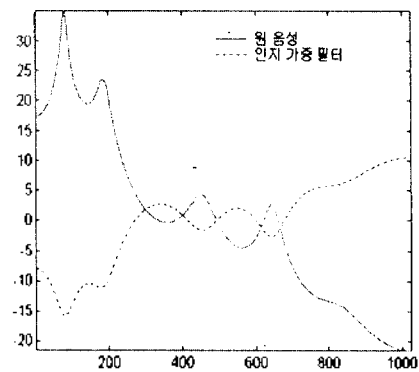
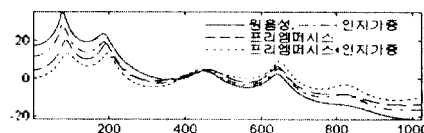
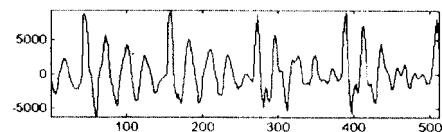
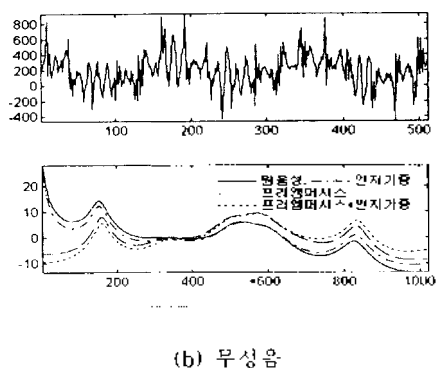


그림 3-1. 원 음성의 스펙트럼과 인지 가중 필터의 스펙트럼 비교

이러한 인지 가중 필터의 효과를 알아보기 위해서 남성화자와 여성화자의 음성에서 유성음과 무성음을 인지 가중 필터와 프리엠퍼시스 필터에 통과시켰다. 이 때 사용한  $\alpha$ 는 0.2,  $\beta$ 는 0.5이다. 그림 3-2는 남성 화자의 유성음과 무성음 파형에서 인지 가중 필터의 특성을 보여주고 있고, 그림 3-3은 여성 화자의 유성음, 무성음에서 인지 가중 필터의 특성을 보여주고 있다.

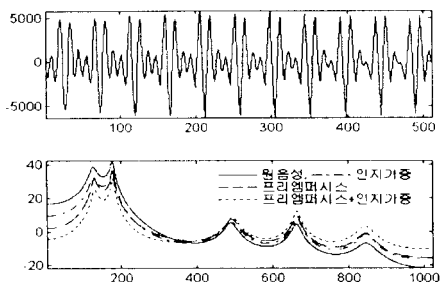


(a) 유성음

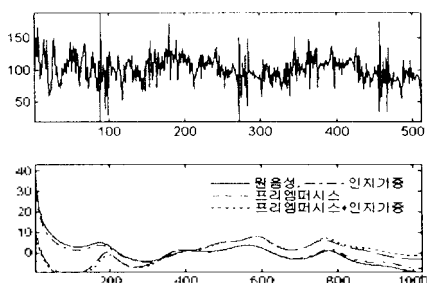


(b) 무성음

그림 3-2. 남성 화자의 음성 에 대한 프리엠퍼시스 필터와 인자 가중 필터 처리 에



(a) 유성음



(b) 무성음

그림 3-3. 여성 화자의 음성 에 대한 프리엠퍼시스와 인자 가중 필터 처리 에

#### 4. 인자 가중 필터를 이용한 개선된 화자 인식 시스템

그림 4-1은 본 논문에서 제안한 텍스트 종속 화자 인식 시스템이다. 우선 입력으로 들어오는 신호는 등록자(customer)의 등록된 음성과 사칭자(imposter)

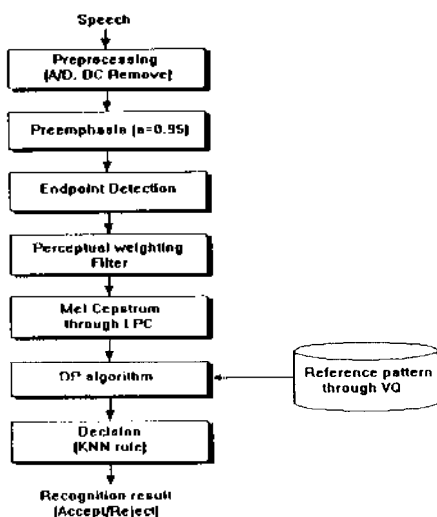


그림 4-1. 본 논문에서 사용한 텍스트 종속 화자 인식의 구성도

의 음성으로, 등록자는 자신이 원하는 고립단어를 미리 발성하게 하여 벡터 양자화를 이용하여 학습시키고, 사칭자는 등록자가 사용한 텍스트를 발성하는 방식이다. 이 때 사용한 패턴 정합법으로는 DP 알고리즘을 사용하였는데[1][3], 그 이유는 본 논문에서 사용할 화자의 수가 10명이기 때문에 그 비교수가 적다는 점에 있고, 등록자가 사용하고 있는 텍스트의 시간적인 음운의 변화 특성이 중요하기 때문이다.

### 5. 실험 및 결과

#### 5-1. 실험 환경 및 데이터 베이스

본 논문에서 사용한 실험 환경은 다음과 같다.

- 일반 IBM-PC/MMX 200MHz
- A/D 변환기 : 일반 PC에 장착된 사운드 카드
- 샘플링 : 11.025KHz
- 양자화 : 16 Bit
- 윈도우 종류 : Hamming Window
- 프레임 크기 : 30msec
- 프레임 오버랩 : 15msec
- 멜 캡스트럼 계수의 차수 : 14

본 논문에서 사용되는 데이터 베이스는 다음과 같다.

- 녹음 환경 : 잡음이 존재하는 사무실 환경

- 등록자의 수 : 10명
- 사칭자의 수 : 4명
- 사용되는 텍스트 : 2~3음절의 등록된 사용자의 이름
- 등록자 데이터 베이스 : 1주일 동안 5번씩 총 4회 발생 (총 200개)
- 사칭자 데이터 베이스 : 1주일 동안 등록자의 이름을 2번씩 총 2회 (총 160개)

이와 같이 구축되는 음성 신호는 다음과 같은 처리를 통해서 인식에 사용된다. 우선 입력된 음성신호에서 DC 바이아스를 제거하기 위해서 식 5.1과 같은 관계식을 사용해서 데이터를 처리하였다.

$$\hat{s}(n) = s(n) - \sum_{l=1}^7 s(l), \quad 1 \leq n \leq 7 \quad (5.1)$$

이와 같이 얻어진 음성 신호는 프리엠퍼시스 필터에 여과된 후 에너지와 영교차율을 이용하여 분할 검출 과정을 수행하였다. 이러한 음성 구간에 인식 가능 필터를 여과시킨 후 LPC 캡스트럼을 추출하여 맨 캡스트럼을 추출하였다. 이와 같이 추출된 맨 캡스트럼을 이용해서 이중 분할 벡터 양자화법을 사용하여 각 화자마다의 64 크기를 갖는 코덱북을 만든 후에 기준 패턴으로 작성하였다. 그리고 인식시 사용한 패턴 저장법으로는 DP 알고리즘을 사용하였다. 이때 사용하는 거리 측정법은 가공화된 캡스트럼 거리 측정을 사용하였고, 이 때 사용한 기중치는 식 5.3과 같다[1].

$$d_c^{(l)} = \sum_{n=1}^N (w(n)c_n - w(n)c'_n)^2 \quad (5.2)$$

$$w(n) = (1 + 7 * \sin(\frac{\pi n}{14})) \quad (5.3)$$

## 5. 2. 실험 결과

인식 실험으로 얻어진 결과는 표 5-1과 같다. 우선 기존의 방법은 프리엠퍼시스 필터와 인지가중 필터를 전처리단계에서 사용하지 않은 상태에서 인식 실험을 한 결과이고, 프리엠퍼시스 필터와 인지가중 필터를 독립적으로 사용한 경우, 마지막으로 두 가지 필터를 모두 사용하였다. 인식 결과는 통해 기존의 방법보다 프리엠퍼시스 필터와 인지가중 필터를 독립적으로 사용한 방법에서 인식이 향상됨을 알 수 있는데, 있지만 두 방법들 사이에서는 인식율의 차이가 없었다. 그리고 프리엠퍼시스 필터가 인식 가능 필터보다

인식율의 측면에서는 조금 더 높은 것을 알 수 있었다. 이는 프리엠퍼시스 필터가 고주파항을 더 많이 강조시키기 때문에 나타나는 화자의 특성에 의한 영향이라 생각된다. 두 가지 필터를 모두 사용하는 경우에는 3가지 방법보다 높은 인식율을 얻을 수 있었는데, 이는 고주파 항을 가장 많이 강조 시켰기 때문이다.

표 5-1. 인식 실험 결과

	기존의 방법	프리엠퍼시스 필터	인지가중 필터	프리엠퍼시스 필터와 인지가중 필터
True Accept	4.17%	3.05%	2.50%	1.67%
True Reject	3.33%	3.05%	4.17%	1.94%
인식율	92.5%	93.89%	93.33%	96.39%

## 6. 결론 및 향후 연구 과제

현재의 화자 인식 기술은 음성 인식 기술에 기반해서 크게 벗어나고 있기 못하기 때문에 아직까지 연구 단계에서 실용화 단계로 넘어가지 못하고 있는 실정이다. 본 논문에서는 이러한 문제점을 해결하기 위해서 화자의 특성이 많이 분포되어 있는 고주파수 영역을 강조시키는 방법에 관한 것이다. 본 논문에 제시한 방법을 모의 실험한 결과 기존의 방법보다 3.89%까지 인식율의 향상을 얻을 수 있었다.

### [참고 문헌]

- [1] L. R. Rabiner and B. H. Juang, *Fundamentals Of Speech Recognition*, Prentice Hall, ATR&T, USA, 1993
- [2] A.M. Kondiz, *Digital Speech*, John Wiley & Sons, 1994
- [3] Hiroaki Sakoe & Seibi Chiba, "Dynamic Programming Algorithm Optimization for Spoken Word Recognition", IEEE Trans. on ASSP, vol.26, No.1, pp.43-49, Feb.1978.
- [4] B. H. Juang and L. R. Rabiner, "On the Use of Bandpass Liftering in Speech Recognition", IEEE Trans on ASSP, vol.35, No.7, pp.947-954, July 1987.