

Microphone 거리에 따른 Glottal Spectrum 성분 분석에 관한 연구

박현영, 장경아, 배명진
숭실대학교 정보통신공학과

A Study on Glottal Spectrum Analysis According to the Distance between the Microphone and the lips

Hyunyoung Park, Kyunga Jang, Myungjin Bae
Dept. of Information and Telecommunication, Soongsil University
E-mail : phy717@hotmail.com

요약문

현재 음성인식기는 다 채널의 음성입력방식을 사용하고 있는 추세이다. 이런 방법으로 음성인식기를 사용할 때 자동적으로 음성을 검출하는 음성입력 방식은 발성자와 마이크간의 거리에 따라 Glottal Spectrum 성분이 변하는 특성을 가지고 있다. 이러한 Glottal Spectrum 성분은 $a=R1/R0$ (LPC 포락선의 기울기)로 나타낼 수 있다. 본 논문에서는 발성자와 마이크 거리에 따른 Glottal Spectrum 성분을 비교·분석 하고자 한다.

1. 서 론

음성에 의한 의사전달은 말하는 이가 전달하고자 하는 생각에서 출발하여 말하는 이의 생각을 일련의 신경과정과 근육의 움직임을 통해 음향 압력파로 듣는 이의 청각 시스템에 나뉜다. 이러한 인간의 의사소통 수단인 음성을 기계가 인지할 수 있게 하는 것을 음성인식 시스템이라 하고 이러한 음성 인식 알고리즘 개발은 현재 활발히 진행중이다. 개인이나 특정 단체의 정보의 보안을 위해서는 사용자의 확인 과정이 필요하다. 이때 확인 절차는 사용자에게 사용이 용이해야 하며 확인 내용은 정확해야 한다. 이러한 점을 고려하여 근래에 들어 사용자의 음성특성을 이용한 사용자 확인 방법이 고안되었다. 즉, 사용자가 특정 패스워드 또는 임의의 음성을 마이크를 통하여 발성한 뒤 발성된 음성을 바탕으로 사용자를 확인하는 방법이다. 이러한 방법에는 화자가 마이크를 통하여 발성한 음성으로부터 스펙트럼의 특성을 나타내는 특징벡터를 추출하여 저장된 각각의 기준 패턴과 패턴매칭을 통해 화자를 인식하는 방법이 있다.

이러한 패턴매칭 과정에서 어느정도로 고차 포먼트 성분을 보상해 주느냐에 따라 인식률이 좌우될 수 있다. 고차 포먼트 성분의 보상을 위해서는 현재 프리엠퍼시스 필터를 사용하고 있다. 본 논문에서는 이 필터를 이용해 보다 적응적으로 보상하기 위하여 KQ음성 인식 과정시 차이가 날수 있는 마이크 거리에 따른 Glottal Spectrum의 변화도를 구하여 프리엠퍼시스 필터에 사용되는 계수 a 값을 비교·분석하였다[9].

2. 음성생성모델

2.1. 음성신호의 생성

음성신호의 구조에 대한 연구는 음성정보를 추출하거나 강조할 수 있다. 따라서 음성신호의 생성에 대한 수학적 모델은 음성을 처리하는데 있어서 매우 중요한 영역이다. 성도(vocal tract)는 성대(vocal cord)와 입술 끝까지를 말한다. 따라서 성도는 인두와 입 또는 구강으로 구성된다. 남성의 성도 길이는 17cm정도이다. 성도의 단면적은 혀, 입술, 턱 그리고 0cm(완전히 닫혔을 때)에서 약 20cm까지 변화하는 연구개의 위치에 의해 결정된다. 비도는 연구개에서 시작하여 콧구멍에서 끝난다. 연구개가 낮아질 때 비도는 비음을 생성하기 위해 음향학적으로 성도에 연결된다. 음성생성의 과정을 연구함에 있어서 수학적 모델로서 물리적 시스템을 표현하는 것은 매우 중요하다. 그림 2-1은 성문의 구조적 그림을 보여주고 있다. 볼록도는 혀, 기관지와 호흡기관으로 구성된 하부-성문 시스템을 표현한다. 이 하부 성문 시스템은 음성을 생성하는 에너지원이다. 음성은 간단히 공기가 혀로부터 방출되고 결과적으로 성도에 있는 협착점에 의해 공기가 동요될 때 이 시스템으로부터

방사되는 음향학적 파형이다.

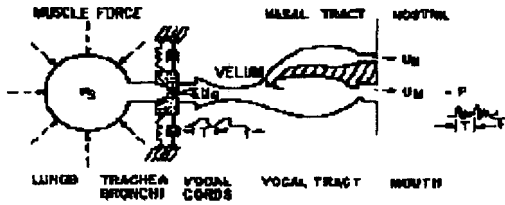


그림 2-1. 음성발생기관의 블록도

유성음은 조정된 성대의 팽창과 함께 성문을 통한 공기의 힘에 의해 생성되어 감쇄 진동하여 성도를 자극하는 공기의 준 주기적인 펄스를 만든다. 마찰음 또는 무성음은 성도에 있는 어떤 점에서 협착을 형성하고, 동요를 만들기 위해 고속으로 협착점을 통과하는 공기의 힘에 의해 발생된다. 파열음은 완전히 입을 폐쇄하고, 이 폐쇄 뒤에서 압력을 만들어 갑자기 느슨하게 함으로써 생성된다.

3. 화자인식 시스템

3.1 화자 인식의 분류

일반적으로 화자 인식은 크게 두 가지로 나누어 처리되고 있다. 첫째로 화자식별은 등록된 화자집단에 지금 요청중인 화자의 발성이 등록되어 있는지를 결정하는 과정이다. 둘째로 화자확인(Verification)은 지금 발성중인 화자가 인식시스템이 요청한 그 사람인지 아닌지를 결정하는 과정이다.

또한 화자인식은 인식 방법에 따라 4가지로 구분할 수 있다. 첫째로 패턴 정합법에 의한 동적 정합은 입력패턴을 미리 정해진 기준 패턴과 비교하여 최적화된 유사성을 판단하는 방법이다. 둘째로 신경회로망을 이용한 방법은 각 화자별로 신경회로망을 구성하고 화자간의 변별력을 갖도록 학습을 수행하는 인식 방법이다. 그러나 이 방법은 새로운 화자의 추가시 인식 시스템을 다시 학습시켜야 하고, 고도의 병렬계산 능력이 요구되기 때문에 실제 응용 시에는 적합하지 않다는 단점이 있다. 세 번째 방법인 벡터양자화 방법은 입력 패턴과 양자화 코드북 사이의 거리로 유사성을 판단하는 방법이지만 많은 학습자료가 필요하고 화자간의 동적인 변화 특성을 이용하지 못하기 때문에 인식률에 한계가 있다.

마지막으로 은닉마코프 모델은 학습기능을 이용하여 화자내의 변이를 흡수 할 수 있으며, 입력패턴의 비선형 정합을 수행하는 특성이 있다. 화자인식 시스템은 인식에 사용하는 문장의 종속여부에 따라 정해지지 않는 어휘로 인식을 수행하는 텍스트 독립형과 정해진 어휘만을 발생해야 하는 텍스트 종속형으로

나눌 수 있다.

3.2 화자 인식 과정

일반적으로 패턴매칭을 이용한 화자 인식 과정은 다음과 같다. 먼저 발생된 음성신호로부터 음성구간을 검출한다. 검출된 음성신호를 창함수를 이용하여 단구간으로 나눈다. 이렇게 단구간으로 나누어진 음성 데이터에서 화자의 특징벡터를 추출하여 기준패턴으로 사용한다.

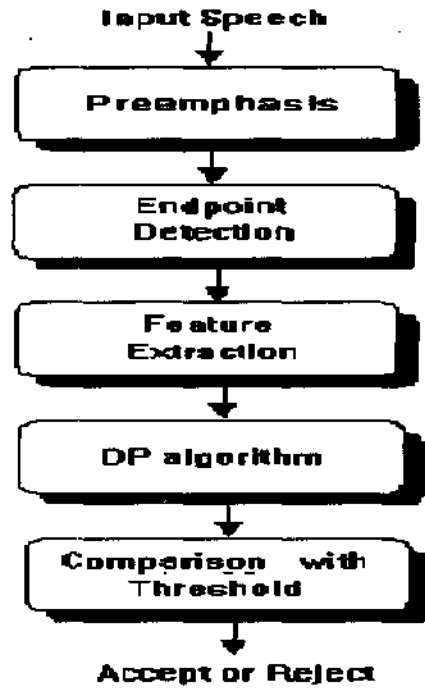


그림 3-1. 일반적인 화자인식 과정

이러한 방법으로 저장된 기준패턴들과 음성입력단에서 들어온 비교패턴을 DTW 방법을 이용하여 화자인식을 수행한다[2][3].

4. 제안한 알고리즘

본 논문에서는 프리엠퍼시스 필터계수를 성문 특성에 따라 적응적으로 구하기 위하여 음성 입력을 받을 때 동일한 3개의 마이크로 0cm, 5cm, 10cm 씩 차이를 두어 Glottal Spectrum 성분을 비교하였다. 필터계수를 구하는 방법은 다음과 같다[2][5]. 단구간 자기상관 함수는 식(4.1)로 표현 가능하다.

$$\phi_n(i, j) = \sum_{m=0}^{N-1-i} s_n(m) s_n(m+i-j), 1 \leq i \leq p, 0 \leq j \leq p \quad \text{식(4.1)}$$

$$\text{여기서 } R_n(j) = \sum_{m=0}^{N-1-j} s_n(m)s_n(m+j) \quad \text{식(4.2)}$$

$$\sum_{j=1}^p a_j \phi_n(i,j) = \phi_n(i,0), \text{ for } i=1, \dots, p \quad \text{식(4.3)}$$

자기상관법을 이용하여 식(4.3)를 풀면 다음과 같이 표현된다.

$$\begin{bmatrix} R_n(0) & R_n(1) & \dots & R_n(p-1) \\ R_n(1) & \dots & \dots & R_n(p-2) \\ \vdots & \vdots & \ddots & \vdots \\ R_n(p-1) & \dots & \dots & R_n(0) \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix} = \begin{bmatrix} R_n(1) \\ R_n(2) \\ \vdots \\ R_n(p) \end{bmatrix} \quad \text{식(4.4)}$$

$p=1$ 에 대하여 위의 식을 정리하면 다음과 같은 식으로 표현 가능하다.

$$a_1 = \frac{R_n(1)}{R_n(0)} \quad \text{식(4.5)}$$

그림 4-1, 그림 4-2는 음성신호와 자기상관법을 이용하여 측정된 음성신호의 기울기를 나타낸 것이다.

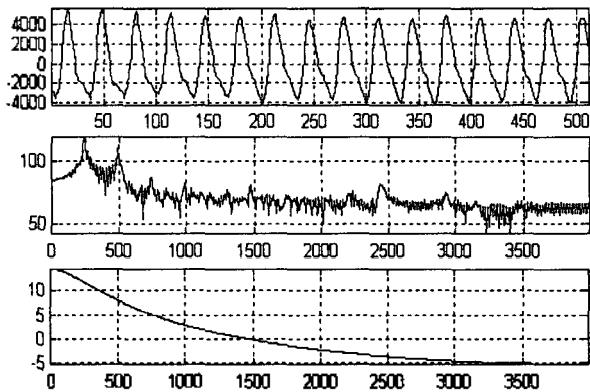


그림 4-1. 유성음 (a) 음성신호, (b) Glottal Spectrum, (c) 측정된 기울기

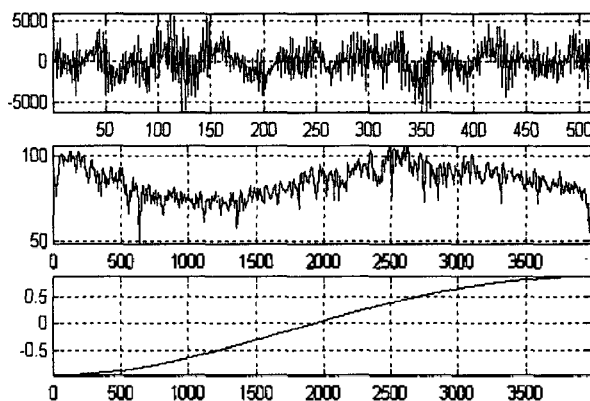


그림 4-2. 무성음 (a) 음성신호, (b) Glottal Spectrum, (c) 측정된 기울기

위의 특성은 유성음과 무성음간의 평탄화를 위한 필터 계수의 기울기를 보인 것이고 이러한 특성의 차이

를 마이크와 발생자의 거리에 따라 어떻게 변하는가를 실험 한다.

5. 실험 및 결과

본 논문의 실험을 하기 위해 동일한 종류의 Dynamic Microphone 3개와 1.1Ghz Pentium3 노트북 3대를 사용했다. 실험에 사용된 음성 시료는 /인수네 꼬마는 천재 소년을 좋아한다/ 라는 문장이다. 이 음성 시료는 8khz로 샘플링하고 16비트로 양자화하여 사용하였고 한 프레임의 길이는 512샘플이다. 프로그램은 MATLAB 및 VISUAL C++을 사용하였다.

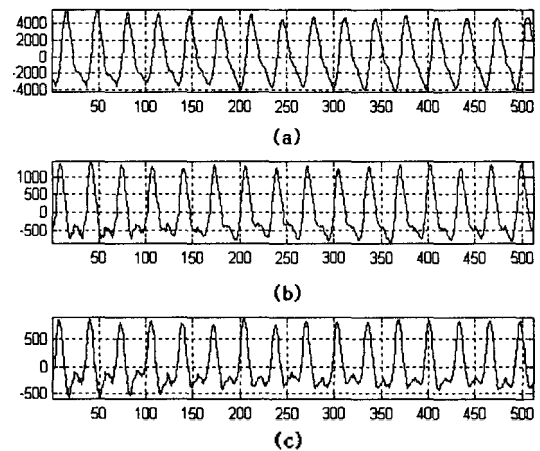


그림 5-1. 마이크 거리에 따른 음성파형(유성음) (a) 0cm (b) 5cm (c) 10cm

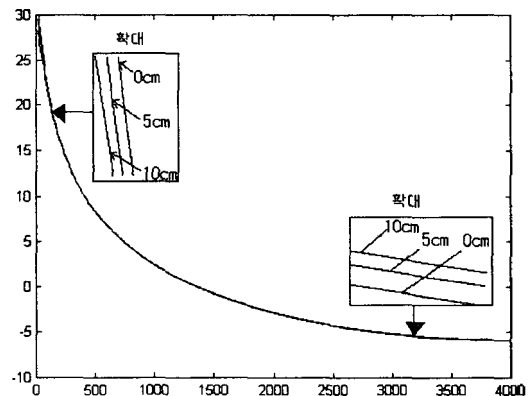


그림 5-2. 유성음의 기울기 (1프레임)

그림 5-1은 동일한 음성파형중 유성음을 마이크의 거리에 따라 각각의 한프레임 파형을 보여준다. 그림 5-2는 그림 5-1의 유성음 파형의 기울기를 보여준다. 마이크와의 거리가 가까운 음성일수록 기울기가 크고 먼 음성일수록 기울기가 작아짐을 알 수 있다.

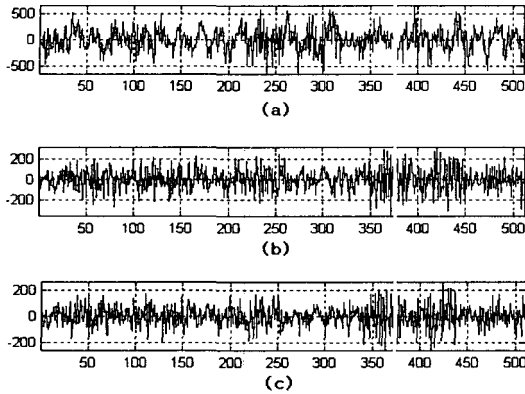


그림 5-3. 마이크 거리에 따른 음성파형(무성음)
(a) 0cm (b) 5cm (c) 10cm

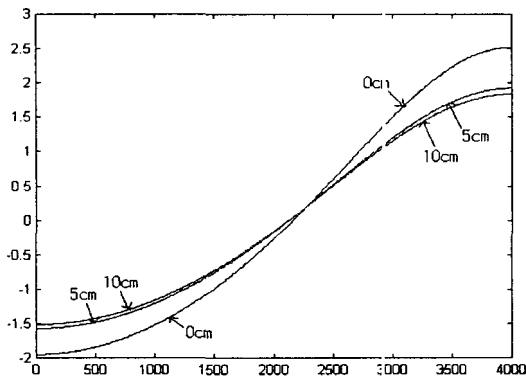


그림 5-4. 무성음의 기울기 (1프레임)

그림 5-3은 동일한 음성파형중 무성음을 마이크의 거리에 따라 각각의 한프레임 파형을 보여준다. 그림 5-4는 그림 5-3의 무성음 파형의 기울기 보여준다. 유성음과 마찬가지로 마이크와의 거리가 가까운 음성일수록 기울기의 크고 먼 음성일수록 기울기가 작아짐을 알 수 있다.

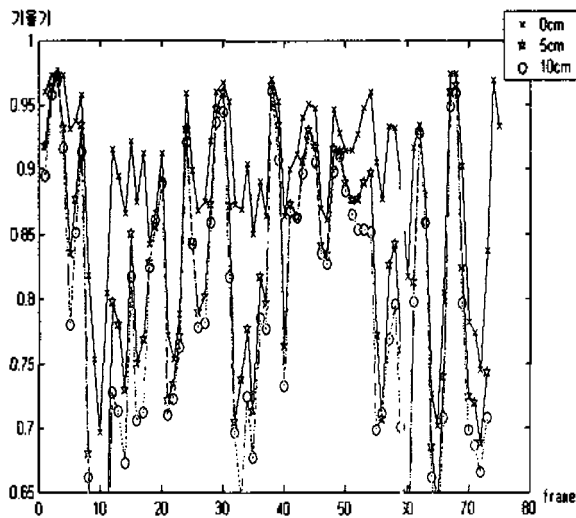


그림 5-5. 전체 음성의 기울기 변화도

그림 5-5는 마이크 거리에 따른 전체음성 /인수내 꼬마는 천개 소년을 좋아한다/의 기울기 변화도이다. 마이크와의 거리가 가까울수록 기울기가 크고 멀어질수록 기울기가 작아짐을 알 수 있다.

6. 결론

개인이나 특정 단체의 정보의 보안을 위해서는 사용자의 확인 과정이 필요하다. 이때 확인 절차는 사용자에게 사용이 용이해야 하며 확인 내용은 정확해야 한다. 이러한 점을 고려하여 근래에 들어 사용자의 음성특성을 이용한 사용자 확인 방법이 고안되었다. 본 논문에서는 프리엠퍼시스 필터 계수를 적응적으로 구하기 위하여 성문특성을 마이크 거리에 따라 실험하였고 그 결과 마이크 거리가 멀어짐에 따라서 그 기울기가 작아짐을 알 수 있었다.

7. 참고 문헌

- [1] L. R. Rabiner & Biing-Hwang Juang, *Fundamentals Of Speech Recognition*, Prentice-Hall AT&T, U.S.A., 1993
- [2] L. R. Rabiner & R.W.Schafer, *Digital Processing of Speech Signal*, Prentice-Hall, Englewood Cliffs, N.J., U.S.A., 1978
- [3] A.M. Kondoz, *Digital Speech*, Jhon wiley & Sons, 1994
- [4] 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.
- [5] L. R. Rabiner, R.W. Schafer, " Digital Processing of Speech Signal", Prentice Hall, 1978.
- [6] 배명진, "디지털 음성분석", 동영출판사, 1998. 4.
- [7] Oppenheim, Schafer, "Discrete Time Signal Processing", Prentice Hall, 1989.
- [8] Emanuel C. Ifeachor, "Discrete Time Signal Processing", Addison Wesley, 1993.
- [9] 오영환, "음성언어정보처리", 홍릉과학출판사, 1998.
- [10] A. M. Kondoz, "Digital Speech", John Wiley & Sons Ltd, 1994.
- [11] 민소연, 강은영, 배명진, "성문특성이 제거된 성도특성에 관한 연구", 대한전자공학회, 추계 종합학술대회, 2000년 11월 25일.