

# 성문특성 측정을 통한 유/무성음 결정에 관한 연구

민소연\*, 강은영, 신동성, 배명진  
송실대학교 \*전자공학과, 정보통신공학과  
전화: (02)824-0906, 팩스: (02)820-0018

## A Study on Decision of Voiced/Unvoiced Region through Measuring the Vocal Cord Property

SoYeon MIN\*, EunYoung KANG, DongSung SHIN, MyungJin BAE  
Dept. of \*Electronics, Information and Telecommunication Engr, Soongsil University  
E-mail : pasternak@hanmail.net

### Abstract

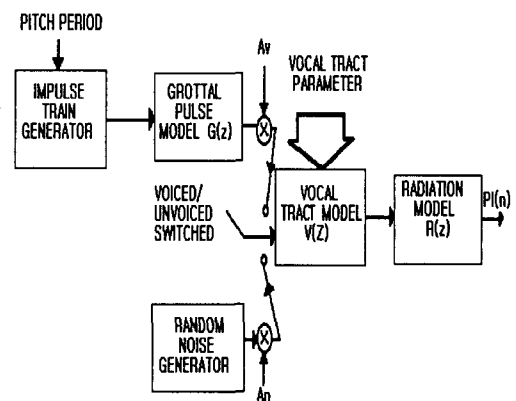
Speech is classified into voiced signal and unvoiced signal. Since the amplitude of voiced fall off at about  $-20\text{dB/decade}$ , dynamic range is often compressed prior to spectral analysis so that details at weak, high frequencies may be visible[5][6]. There is a distinct difference in spectrum slope between voiced signal and unvoiced signal. In this paper, we got the slope of each frame by using autocorrelation method, and determined voiced/unvoiced region. Also, we used energy to decide region of silence. To show experimental results, we allot to 1 value in voiced region,  $-1$  value in unvoiced region and 0 value in silence region.

### 1. 서론

음성 시스템을 여기의 형태로 구분을 하면 크게 유성음과 무성음으로 구분되어진다. 유성음은 성문, 즉 성대근육(vocal fold) 사이의 열린 부분을 통해 공기를 억지로 밀어 넣을때 생긴다. 성대(vocal cords)의 장력은 그것이 진동의 형태로 떨리도록 조절된다. 성문 아래에서 나오는 공기의 흐름이 주기적으로 막히게 되면 공기가 준 주기적으로 뿜어져 나와 성도를 여기 시킨다. 이렇게 후두에서 생기는 소리가 유성음이다. 무성음은 성도를 수축시켜, 공기를 이 수축된 부위에 밀어

넣어, 공기의 무질서한 흐름을 만들어냄으로써 생긴다. 본 논문에서는 유/무성음의 스펙트럼상의 포먼트 특성, 즉 기울기를 측정하여 유/무성음을 분리하였다. 기울기 측정을 위해 자기상관법을 사용하였고 목음구간 판단을 위해서는 에너지를 이용하였다. 본 논문의 구성은 다음과 같다. 2장에서는 일반적인 음성생성 모델과 유/무성음의 스펙트럼상 특징에 대하여 설명하고, 3장에서는 주파수 영역의 포락선 기울기 측정에 사용한 방법을 간단한 수식을 통해서 살펴본다. 4장에서는 실험 결과, 5장에서는 결론을 맺는다.

그림 1. 일반적인 음성생성에 대한 이산시간 모델



## II. 음성신호의 생성모델

### 2-1. 일반적인 음성생성모델

음성생성에 대한 선형모델은 50년대 후반 Fant에 의해 개발되었다. 여기서는 음성출력을 음원이 여파기를 통과하여 나오는 신호로 가정하고, 음원과 성도의 각 부분을 독립적인 것으로 간주하는 선형예측모델을 제시하였다. 음원에 대한 모델로 유성음의 음원은 준주기적인 펄스, 무성음의 음원은 백색잡음을 사용하였고, 성대에서 성문이 음원에 미치는 영향은 다음과 같은 성문모델로 모델링 하였다[1][2].

$$G(z) = \frac{1}{(1 - e^{-cT}z^{-1})^2} \quad (1)$$

여기서, T는 준주기이고, cT는 감소인자이며 1 보다 충분히 작다.

성문을 지난 신호는 성도를 거치면서 성도의 형태에 따라 몇 개의 공명주파수를 갖게 되는데 이 공명주파수와 대역폭을 2-극(pole) 여파기로 나타내면 성도에 대한 모델은 다음과 같이 구할 수 있다.

$$V(z) = \frac{1}{\prod_k [1 - 2e^{-\alpha_k T} \cos(B_k T)z^{-1} + e^{-2\alpha_k T}z^{-2}]} \quad (2)$$

여기서, k는 포만트의 개수이고,  $\alpha_k$ 와  $B_k$ 는 포만트 주파수와 대역폭을 결정하는 값으로 대역폭의 좁다는 가정에서 실제 포만트의 공명주파수와 대역폭은 각각  $B_k/2\pi$ ,  $\alpha_k/2\pi$ 로 결정된다.

성도를 통과한 신호는 마지막으로 입술을 통과하는데 입술에서의 방사특성에 대한 모델은 다음과 같다.

$$L(z) = 1 - z^{-1} \quad (3)$$

음원에 대한 모델을 E(z)로 하면 음성출력 X(z)는 다음과 같다[1][2][8].

$$X(z) = E(z)G(z)V(z)L(z) \quad (4)$$

이를 진극(all-pole)형 합성모델로 다시 정의하면,

$$X(z) = E(z) \frac{1}{A(z)} \quad (5)$$

여기서,

$$A(z) = \frac{1}{G(z)V(z)L(z)} \quad (6)$$

이 A(z)는 다음 관계식에 의해 음성으로부터 성도특성을 제거하고 음원을 이끌어 내는 가역 여파기 입을 알 수 있다.

$$E(z) = X(z)A(z) \quad (7)$$

이상에서 살펴본 것과 같이 음성을 음원과 그 음원이 통과하는 성도여파기로 모델링 하고 각각을 독립적으로 모델링 함으로써 수학적으로 선형방정식의 해를 구할 수 있다[1][2][8].

### 2-2. 유/무성음의 일반적 특징

정상상태의 성도 모양에 의해 발음되는 소리가 연속음이고 성도 모양이 변화하면서 나는 소리는 불연속음이다. 즉, 모음, 마찰음, 파찰음, 비음 등은 연속음 계열이고 이중모음, 유음, 과도음, 파열음 등은 불연속음 계열이다. 모음은 성대 떨림(유성음) 수반하며 보통 큰 진폭을 갖는다. 지속 시간은 40-400ms 정도로 넓은 구간에서 변할 수 있으며 성도를 따라 형성되는 단면적의 변화에 의해 포만트가 결정된다. 그림 2에서는 모음의 포만트 특성을 보인다[1][2][3].

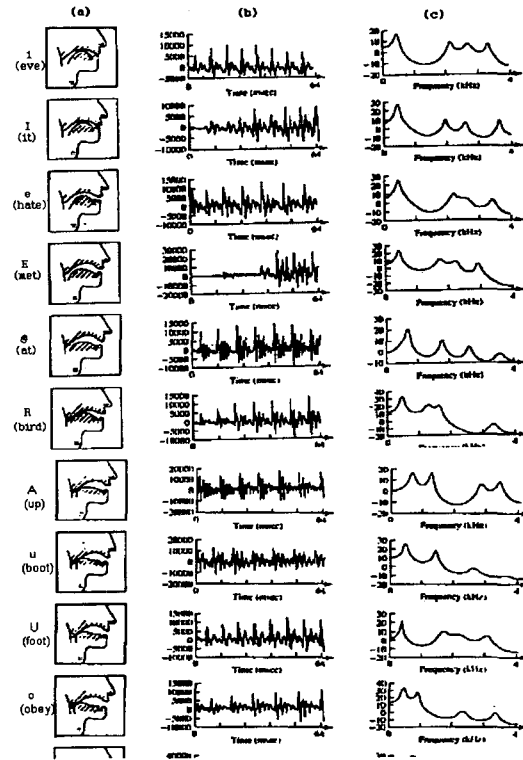


그림 2. 영어 모음의 특징도  
(a) 성도모양 (b) 음성파형 (c)스펙트럼

자음은 모음의 경우보다 큰 수축에 의해 형성된 성도에서 나오는 소리를 자음이다. 여기에는 마찰음과 파찰음, 파열음, 비음 등이 속한다. 마찰음에는 무성 마찰음과 유성 마찰음이 있다. 무성마찰음에는 /t/, /tʰ/,

/s/, /S/, /h/ 등이 있다. 성도의 주파수 특성을 보면 대개 저주파 부분이 적고 고주파 부분이 많다. 유성 마찰음에는 /v/, /D/, /z/, /Z/ 등이 있다. 유성 마찰음은 수축 부위에서 마찰 잡음을 일으키면서 동시에 성도를 여기 시키는 주기적인 성문 펄스를 발생시킨다. 그림 3은 무성 마찰음의 포먼트 특성을 나타낸다.

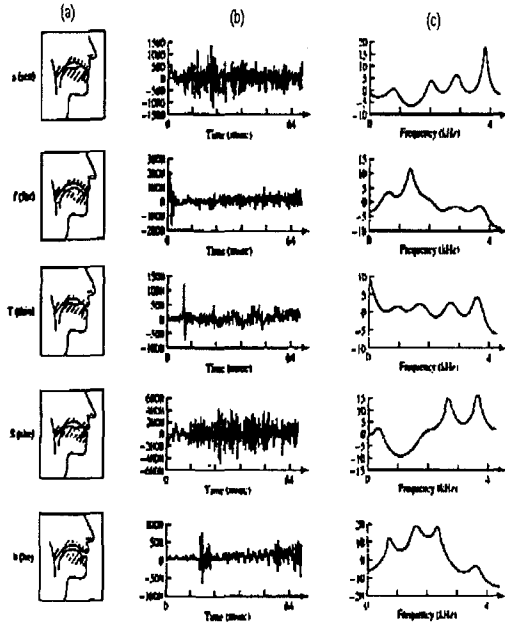


그림 3. 무성마찰음의 특징도  
(a)성도모양 (b)음성파형 (c)스펙트럼

파찰음은 파열에서 마찰로의 전이에 의해 형성된다. 예를 들어 change의 /c/가 무성 파찰음, jam의 /j/가 유성 파찰음으로 존재한다. /c/는 무성 파열음 /t/에서 무성 마찰음 /S/로 가는 전이에 의해 생성되고 /j/는 유성 파열음 /d/에서 유성 마찰음 /Z/로의 전이에 의해 생성된다. 파열음 /b, d, g, p, t, k/는 전이 특성과 불연속음 특성을 갖는다. 파열음은 성도 어느 부분의 수축으로 압력을 지니고 있다가 이것을 갑자기 터뜨림으로 생성된다. 무성파열음은 성도 폐쇄 부위에 압력을 만든 후 이것을 갑자기 터뜨리게 된다. 유성파열음은 무성파열음과 대부분 비슷하고 전체 발음 동안 계속되거나 터뜨림 후에 시작되는 성대 떨림만 다를 뿐이다. 또한, 구강을 닫고 비강을 열게 하는 성문파에 의해 생기는 유성음이 비음으로 /m, n, ŋ/ 등이 있다. 파형은 모음과 비슷하지만 에너지가 모음보다 작은데 이것은 비강이 소리를 발산시키는 능력이 제한되어 있기 때문이다[1][2][8].

### III. 주파수영역에서의 기울기 측정

단구간 자기상관 함수는 식 (8)로 표현 가능하다[1][8]

$$\Phi_n(i, j) = \sum_{m=0}^{N-1-(i-j)} s_n(m)s_n(m+i-j), 1 \leq i \leq p, 0 \leq j \leq p \quad (8)$$

여기서,  $R_n(j) = \sum_{m=0}^{N-1-j} s_n(m)s_n(m+j) \quad (9)$

$$\sum_{j=1}^p a_j \Phi_n(i, j) = \Phi_n(i, 0), \text{ for } i=1, \dots, p \quad (10)$$

자기상관법(Auto-correlation Method)을 이용하여 식 (10)을 풀면 다음과 같이 표현된다.

$$\begin{vmatrix} R_n(0) & R_n(1) & \dots & R_n(p-1) \\ R_n(1) & \dots & \dots & R_n(p-2) \\ \vdots & \vdots & \vdots & \vdots \\ R_n(p-1) & \dots & \dots & R_n(0) \end{vmatrix} \begin{vmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{vmatrix} = \begin{vmatrix} R_n(1) \\ R_n(2) \\ \vdots \\ R_n(p) \end{vmatrix} \quad (11)$$

p=1에 대하여 위의 식을 정리하면 다음과 같은 식으로 표현 가능하다[1][2][6][7][8].

$$a_1 = \frac{R_n(1)}{R_n(0)} \quad (12)$$

### IV. 실험 및 결과

실험을 위해 IBM-PC(333 MHz)에 마이크 입력이 가능한 A/D 변환기를 인터페이스 하였다. 음성시료는 다음과 같으며 8kHz로 표본화하고 16bit로 양자화하여 사용하였다.

- 발성1: /speech/
- 발성2: /What is this/
- 발성3: /Nice meet you/

그림 4, 그림 5는 음성신호와 자기상관법을 이용하여 측정된 음성신호의 기울기를 나타낸 것이다.

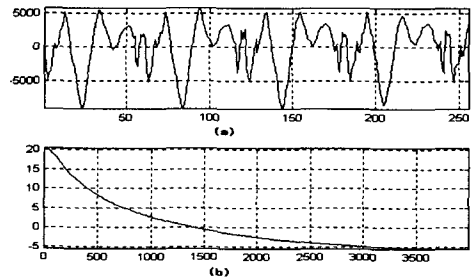


그림 4. (a)음성신호 (b)측정된 기울기

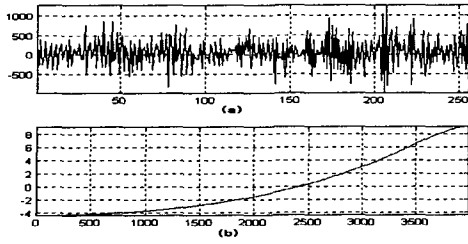


그림 5. (a)음성신호 (b)추정된 기울기

각 프레임별로 기울기를 측정하여 유/무성음을 판단한 결과를 그림 6~그림 8에서 나타낸다. 묵음구간 판별을 위해서 에너지를 이용하였다. 즉, 각 프레임의 기울기와 에너지를 이용하여 유성음구간이라고 판단된 구간에는 1의 값을 무성음 구간이라고 여겨진 부분에서는 -1의 값을 할당하였고 묵음구간에서는 0의 값을 주어서 출력형태를 관찰한 결과이다.

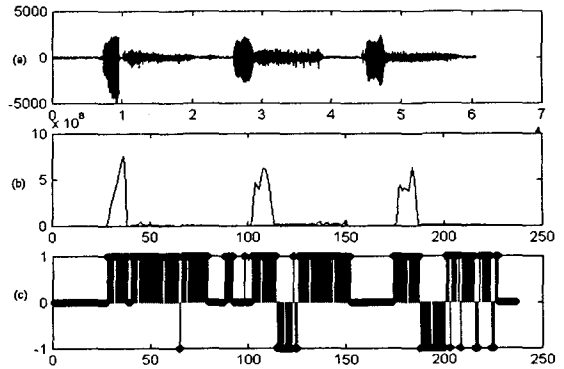


그림 8. 음성시료(3)에 대한 실험결과  
(a)음성신호 (b)에너지 (c)유/무성음 판단 결과

### V. 결론

음성 시스템을 여기의 형태로 구분을 하면 크게 유성음과 무성음으로 구분되어진다. 유성음과 무성음의 스펙트럼상 포락선의 형태를 관찰하면 차이가 있음을 알 수가 있다. 본 논문에서는 자기상관법을 이용하여 각 프레임의 기울기를 측정하여 유/무성음 분리에 대한 연구를 수행하였고, 묵음구간 판별을 위해서 에너지를 이용하였다. 각 프레임의 기울기와 에너지를 이용하여 유성음구간이라고 판단된 구간에는 1의 값을 무성음 구간이라고 여겨진 부분에서는 -1의 값을 할당하였고 묵음구간에서는 0의 값을 주어서 출력형태를 관찰하였다. 실험결과 유/무성음 구간에서는 비교적 정확한 판단을 하였으나 진이구간에서는 유성음으로 판단됨을 알 수가 있었다. 향후, 지속시간이 상대적으로 짧은 무성음 판단을 위한 연구와 한글에 대한 실험이 진행되어야겠다.

### 참고문헌

- [1] L. R. Rabiner, R.W. Schafer, "Digital Processing of Speech Signal", Prentice Hall, 1978.
- [2] 배명진, "디지털 음성분석", 동영출판사, 1998. 4.
- [3] Oppenheim, Schafer, "Discrete Time Signal Processing", Prentice Hall, 1989.
- [4] Emanuel C. Ifeachor, "Discrete Time Signal Processing", Addison Wesley, 1993.
- [5] 오영환, "음성인어정보처리", 홍릉과학출판사, 1998.
- [6] Douglas O. shaughnessy, "Speech Communication", IEEE Press, 1996.
- [7] A. M. Kondoz, "Digital Speech", John Wiley & Sons Ltd, 1994.
- [8] 배명진, "디지털 음성합성", 동영출판사, 1998. 2.
- [9] 민소연, 강은영, 배명진, "성문특성이 제거된 성도특성에 관한 연구", 대한전자공학회, 추계 종합학술대회, 2000년 11월 25일.

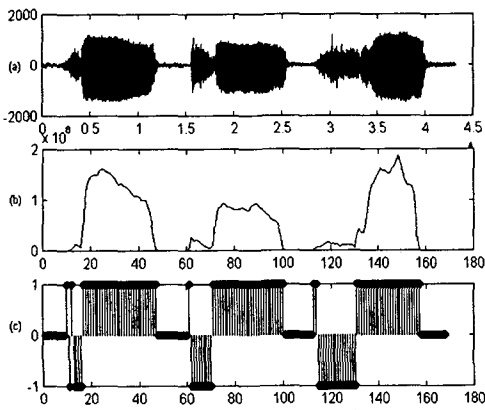


그림 6. 음성시료(1)에 대한 실험결과  
(a)음성신호 (b)에너지 (c)유/무성음 판단 결과

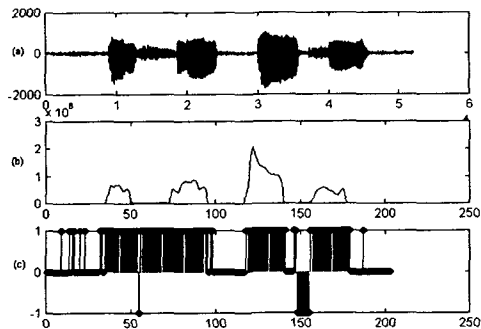


그림 7. 음성시료(2)에 대한 실험결과  
(a)음성신호 (b)에너지 (c)유/무성음 판단 결과