

# 발성유형지수 $k$

박 한 상  
서울대학교 언어학과

## Phonation Type Index $k$

Hansang Park  
Department of Linguistics, Seoul National University  
E-mail : hotphans@hotmail.com

### Abstract

This study proposes phonation type index  $k$  as a descriptor of the overall spectral tilt, which is free from the effects of fundamental frequency and vowel quality. The newly proposed phonation type index  $k$  presents a simple and single measure of the overall spectral tilt. Phonation type index  $k$  can be applied to speech technology. It can also be used in diagnosing patients voice qualities in speech pathology. The distribution of phonation type index  $k$ , which is speaker-dependent, may be useful in forensic phonetics and voice recognition as an indicator of speaker identity.

### I. 서론

#### 1.1 연구의 목적

발성유형의 한 특성인 스펙트럼의 기울기는  $H1^*-A1$  [2],  $H1-F2$  [3][4], and  $H1^*-A3^*$  [2] 등으로 측정되어 왔다. 그러나 기존의 측정단위들은 기본주파수와 모음의 영향을 전부 혹은 일부 고려하지 않았다. 본 연구는 기존의 측정단위에서 고려되지 않았던 기본주파수와 모음의 영향을 모두 배제한 발성유형지수  $k$  (phonation type index  $k$ , 줄여서 PTI  $k$ )를 새로이 제시한다.

본 연구는 문두의 CV와 어중의 VCV 연쇄체에서 한국어의 치조파열음과 마찰음, 그리고 경구개 파찰음

다음에 따라오는 모음 /a/에서 실현되는 발성유형 (phonation type)의 특성에 관한 정보를 제공한다는 점에서 의의가 있다. PTI  $k$ 는 음성학 및 음성공학 그리고 음성의학 등에 두루 쓰일 것으로 기대된다. 특히 PTI  $k$ 는 화자에 따라 그 값의 통계적 지표들이 뚜렷이 달라지는 점을 고려할 때 범음성학이나 화자인식과 같은 분야에서 화자의 특성을 보여주는 지표로 쓰일 수 있을 것으로 기대된다.

#### 1.2 이론적 배경

음원필터이론(Source Filter Theory)에 따르면 음압 (sound pressure)의 주파수반응(frequency response)은 다음의 식 (1)과 같이 음원특성과 방출특성, 그리고 필터특성의 곱으로 나타낼 수 있다. [1]

$$|P(f)| = |U(f)||R(f)||H(f)| \quad (1)$$

위의 식 (1)에서 좌변은 음압의 주파수반응을 우변은 왼쪽에서부터 순서대로 음원특성, 방출특성, 그리고 성도전달함수를 각각 나타낸다. 그 중 음원특성과 방출특성의 곱은 다음의 식 (2)와 같이 나타낼 수 있다. [1]

$$|U(f)||R(f)| = P_k \frac{(f/100)}{1+(f/100)^2} \quad (2)$$

위의 식 (2)에서  $P_k$ 는 특정의 음압수준을 결정하는 상

수를 나타낸다. 위의 식 (2)의 결과 나타나는 스펙트럼은 대략 6 dB/octave의 비율로 감소한다. 방출특성은 대략 주파수에 비례하고 소스특성은 주파수의 곱의 역수에 비례한다. [1]

위의 식 (2)에서 음원특성과 방출특성의 곱은 주파수의 함수로 나타나 있는데 모음과 같은 소리에서 배음의 주파수는 기본주파수의 정수배임을 감안하면 위의 식 (2)는 다음의 식 (3)과 같이 기본주파수의 함수로 변환할 수 있다.

$$|U(nf_0)||R(nf_0)| P_k \frac{(nf_0/100)}{1+(nf_0/100)^2} \quad (3)$$

나아가 위의 식 (3)은 아래의 식 (4)와 (5)와 같이 음원특성과 방출특성으로 각각 분리될 수 있다.

$$U(nf_0) P_k \frac{(f_0/100)}{1+(f_0/100)^2} \quad (4)$$

$$R(nf_0) n \quad (5)$$

위의 식 (4)와 (5)를 곱하면 (3)이 된다. 음원특성은 아래의 식 (6)과 (7)과 같이 다시 기본음(fundamental component)과 배음(harmonic)간 크기의 비율로 나눌 수 있다.

$$U(f_0) P_k \frac{(f_0/100)}{1+(f_0/100)^2} \quad (6)$$

$$Source\ Ratio \frac{1+(f_0/100)^2}{1+(nf_0/100)^2} \quad (7)$$

(6)과 (7)을 곱하면 다시 (5)가 된다.

배음들의 진폭은 기본음에 바탕을 두고 있어서 특정 배음의 음원특성의 기여분은 음원비율에 의해 결정된다. 위의 식 (7)에서 알 수 있듯이 음원비율은 상수가 아니라 기본주파수의 함수이다. 다시 말하면 기본음과 배음의 음원비율은 기본주파수에 따라 변한다는 것을 의미한다.

비록 음원비율이 기본주파수에 따라 변한다고 하더라도 발생유형에 따라 달라지는 스펙트럼기울기의 변화는 여전히 설명하지 못한다. 기본주파수가 동일할 경우 다른 발생유형의 음원스펙트럼 기울기가 -12 dB/octave보다 클 수도 있고 작을 수도 있다는 것을 의미한다. 스펙트럼기울기의 변화를 설명하기 위해서는 스펙트럼기울기를 통제하는 요소의 설정이 필요하

다. 본 연구에서는 다음의 식 (8)로 표현되는 요소를 설정한다.

$$\left(\frac{1}{n}\right)^k \quad (8)$$

위의 식 (8)에서  $n$ 은 배음번호이며 지수  $k$ 는 “발성유형지수”라고 명명하였다.

PTI  $k$ 는 기본주파수와는 무관하지만 기본음에 바탕을 두고 있다. PTI  $k$ 는 일종의 분산으로 간주할 수 있다. modal voice를 가장 빈번히 나타나는 발생유형이라고 정의한다면, PTI  $k$ 의 분포에서 modal voice는 mode를 이루는 값을 보여 줄 것이고 creaky voice와 breathy voice가 각각의 끝을 차지하는 값을 보여 줄 것이다.

지금까지의 논의를 바탕으로 음원특성과 방출특성의 곱을 다음의 식 (9)로 다시 쓸 수 있다.

$$|U(nf_0)||R(nf_0)| P_k \frac{(f_0/100)}{1+(f_0/100)^2} \frac{1+(f_0/100)^2}{1+(nf_0/100)^2} \left(\frac{1}{n}\right)^k \quad (9)$$

그리고 배음번호  $n$ 인 배음의 진폭은 다음의 식 (10)과 같이 나타낼 수 있다.

$$H_n = 20\log_{10} P_k \frac{(f_0/100)}{1+(f_0/100)^2} \frac{1+(f_0/100)^2}{1+(nf_0/100)^2} \left(\frac{1}{n}\right)^k n |H(nf_0)| \quad (10)$$

위의 식 (10)에서 알 수 있듯이 어떤 배음의 진폭은 음원, 방출, 그리고 필터특성의 기여분으로 이루어져 있고 음원특성의 기여분은 다시 기본음, 음원비율, 그리고 스펙트럼기울기요소로 구성되어 있다.

PTI  $k$ 는 선형회귀분석법을 이용하여 구할 수 있다. 위의 식 (10)에서 성도전달함수 및 방출특성 그리고 음원비율은 측정된 여러 가지 매개변수들의 값을 대입하면 상수가 된다. 그 결과 기본음과 스펙트럼기울기요소를 미지수로 가지는 다음의 식 (11)과 같은 선형 방정식을 얻을 수 있다.

$$H_n - FILTER - RADIATION - SR = FC + k20\log_{10}\left(\frac{1}{n}\right) \quad (11)$$

위의 식 (11)에서 암시하듯이 성도전달함수(FILTER), 방출특성(RADIATION), 그리고 음원비율(SR)을 배음

의 진폭으로부터 뺀 값으로 구성된 스펙트럼으로부터 기울기의 추이를 보여주는 직선을 구할 수 있다. 그 경우 그 최적선의 y절편은 기본음(FC)이 되고 기울기는 PTI  $k$ 가 된다. 즉 그 선형방정식의 계수  $k$ 를 취함으로써 PTI  $k$ 를 구할 수 있는 것이다.

PTI  $k$  산출의 구체적인 예를 다음의 그림으로 보여 줄 수 있다. 높고 낮은 PTI  $k$  값을 가진 두 음성샘플의 스펙트럼이 그림 1에 나타나 있다.

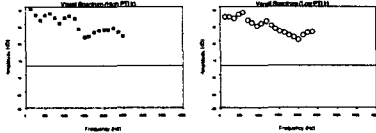


그림 1. 두 음성샘플의 스펙트럼. 왼쪽의 그림은 높은 PTI  $k$ 를, 오른쪽의 그림은 낮은 PTI  $k$ 를 가진 샘플이다. 각각의 점은 배음의 진폭을 나타낸다.

먼저 성도전달함수의 기여분을 계산하면 그림 2와 같이 나타낼 수 있다.

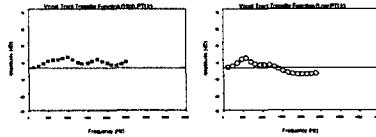


그림 2. 두 음성샘플의 성도전달함수 기여분

다음으로 방출특성의 기여분을 계산하면 그림 3과 같다.

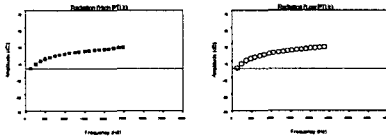


그림 3. 두 음성샘플의 방출특성 기여분

마지막으로 음원비율을 계산하면 그림 4와 같이 나타낼 수 있다.

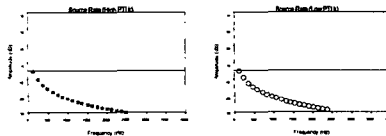


그림 4. 두 음성샘플의 음원비율 기여분

그림 1에서 보여준 모음 스펙트럼의 값에서 그림 2, 3, 4에서 보여준 성도전달함수, 방출특성, 그리고 음원비율의 기여분을 빼면 그림 5와 같은 스펙트럼을 얻을 수 있다.

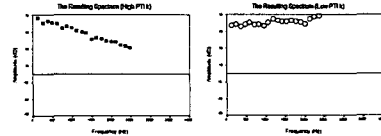


그림 5. 두 음성샘플의 성도전달함수, 방출특성, 그리고 음원비율의 기여분을 뺀 스펙트럼

그림 5에서 보는 바와 같이 두 음성샘플의 성도전달함수, 방출특성, 그리고 음원비율의 기여분을 뺀 결과 나타나는 스펙트럼은 선형추이(linear trend)를 분명히 보여주고 있다. 각각의 샘플에 대하여 최적선(best-fit line) 그리면 그림 6과 같다.

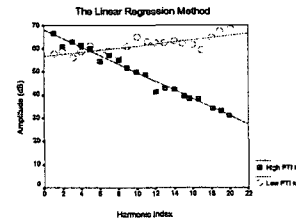


그림 6. 성도전달함수, 방출특성, 그리고 음원비율의 기여분을 뺀 스펙트럼의 선형 추이와 최적선

## II. 연구 방법

본 연구에서 사용된 음성자료는 한국어 치조파열음  $[t^h, t^*, t]$  및 마찰음  $[s^h, s^*]$  그리고 경구개 파찰음  $[t^h, t^*, t]$  다음에 따라오는 모음  $[a]$ 를 포함한 2음절의 동사원형, 예를 들면  $[t^hada]$ , 이다. 각 토큰들은 “\_\_가 아니라 \_\_다”라는 프레임에 넣어서 발음하였다. 프레임의 두 번째 자리에는 모음간 환경에서 발성유형의 특성을 살펴 보기 위하여 토큰의 두 음절을 서로 바꾸어 넣었다.

위에서 제시한 토큰들을 포함한 문장을 3명의 화자가 10번씩 반복하여 240개의 실험문장을 녹음하였다. 화자는 모두 서울 출신의 남자들이었다. 녹음은 UT Austin의 언어학과에 있는 음성실험실에서 이루어졌다. Marantz SuperScope를 이용하여 아날로그 테이프에 녹음을 한 뒤 SoundScope를 이용하여 디지털 신호로 옮겼다. Sampling rate는 22,050이었다.

PTI  $k$ 를 계산하기 위해서는 많은 음향매개변수의 값을 측정해야 한다. 먼저 성도전달함수의 기여분을 계산하기 위해서는 가장 낮은 네 포먼트의 주파수와 bandwidth 값 그리고 기본주파수 값이 필요하고, 최적선 분석을 위해서는 20개의 배음의 주파수와 진폭이 필요하며, 음원비율의 계산을 위해서는 기본주파수의 값이 필요하다. 포먼트 주파수와 bandwidth 값은 SoundScope의 LPC 스펙트럼에서 구했으며 20개의 배음의 주파수와 진폭은 SoundScope의 Narrowband FFT 스펙트럼에서 구했다. 기본주파수는 Praat을 이용하여 측정하였다. LPC 스펙트럼(LPC coefficient: 25)과 FFT 스펙트럼(FFT point: 1024)은 20 ms Hamming 윈도우를 이용하여 구하였으며 매 10 ms 마다 이전의 윈도우와 10 ms 중복하여 측정하였다.

### III. 결과1)

선형회귀분석법을 이용하여 구한 PTI  $k$ 의 분포를 각각의 화자에 대하여 나타내면 그림 7과 같다.

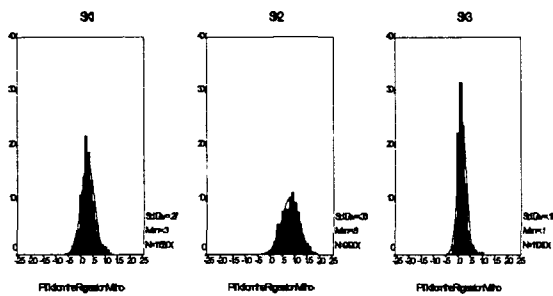


그림 7. 각 화자의 PTI  $k$ 의 분포

각 화자의 PTI  $k$ 의 분포는 정규분포에 가까움을 알 수 있다. 자세한 기술 통계를 표 1에 제시하였다.

	SK1	SK2	SK3
Mean	0.28	0.81	0.15
Median	0.26	0.83	0.12
Variance	0.07	0.13	0.03
Std. Deviation	0.27	0.36	0.18
Minimum	-0.47	-0.21	-0.45
Maximum	1.23	1.84	1.04
Range	1.70	2.05	1.49
Skewness	0.53	0.04	1.27
Kurtosis	0.69	-0.22	3.53

표 1. 각 화자의 PTI  $k$ 의 통계 지표

1 자세한 결과는 [5] 참조.

표 1에서 보여주듯이 모든 통계지표에 대하여 화자간에 상당한 차이가 있음을 알 수 있다. 이것은 modal voice는 화자에 따라 다르며 각각의 통계지표들이 화자를 나타내는 중요한 지표로 사용될 수 있음을 보여 준다.

### IV. 결론

본 연구에서는 기본주파수와 모음의 영향을 배제한 가운데 전체적인 스펙트럼의 기울기를 하나의 지표로 나타내는 PTI  $k$ 를 제시하였다. 하지만 이 연구는 보다 넓은 영역의 기본주파수 변이와 보다 다양한 모음을 가진 음성샘플을 통해서 검증해야 할 일이 남아 있다.

본 연구에서는 모음만을 대상으로 하여 PTI  $k$ 를 계산하였다. 하지만 다른 필터 특성이 있는 비음이나 유음 그리고 기식음의 연구에서는 가능한 한 정확한 음원특성을 얻기 위하여 그러한 필터 특성도 제거되어야 할 것이다.

### 참고문헌

- [1] Fant, G. (1960) *Acoustic theory of speech production*. The Hague: Mouton.
- [2] Hanson, H. M. (1997) Glottal characteristics of female speakers: acoustic correlates, *Journal of the Acoustical Society of America*, 101(1), 466-481.
- [3] Kirk, P., Ladefoged, P., and Ladefoged, J. (1984) Linguistic use of different phonation types, *UCLA Working Papers in Phonetics*, 59, 102-113.
- [4] Ladefoged, P. (1983) The linguistic use of different phonation types, In *Vocal fold physiology: contemporary research and clinical issues* (D. M. Bless and J. H. Abbs, editors), pp. 351-360. San Diego: College Hill Press.
- [5] Park, H. (2002) *Temporal and Spectral Characteristics of Korean Phonation Types*, Ph.D. Dissertation, The University of Texas at Austin.