Phonation Type Index *k*: A New Measure of Spectral Tilt

Hansang Park †
† Kyungsan University, Korea
E-mail: hotphans@hotmail.com

ABSTRACT

Measures of spectral tilt, such as H1*-A1, H1-A2, and H1*-A3*, have not considered the effects of both fundamental frequency and vowel quality. These measures represent specific amplitude differences at particular instances of formant frequency, not as a simple and unified value of the overall spectral tilt. I propose phonation type index *k*, a new measure of spectral tilt which is free from the effects of fundamental frequency and vowel quality, by extending the Source Filter Theory with the addition of a component controlling spectral tilt. The results showed that phonation type index *k* is useful in describing the overall spectral tilt. In particular, the distribution of phonation type index *k* was close to normal and varied with speaker.

1. INTRODUCTION

Spectral tilt, one of the phonetic characteristics of phonation types, has been investigated in terms of acoustic parameters including H1*-A1[1][2], H1-F2[3][4], and H1*-A3*[1][2]. Examination of those measures of spectral tilt reveals that the effects of both fundamental frequency and vowel quality have not been considered. To cope with this problem, 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. In addition, those measures of spectral tilt present a specific value at a specific frequency, not a single and simple value of the overall spectral tilt.

The present study proposes phonation type index *k* as a simple and single measure of the overall spectral tilt. The present study is important in that the newly proposed measures of phonation type, i.e. phonation type index *k* is free from the effects of both fundamental frequency and vowel quality, and in that phonation type index *k* is a unified and simplified measure of the overall spectral tilt. In particular, the distribution of phonation type index *k*, which is speaker-dependent, may be useful in speaker recognition as an indicator of speaker identity.

2. THEORETICAL BACKGROUND

Phonation type index *k* was derived through a reconsideration and extension of the Source Filter Theory. This section proposes an acoustic theory of vocal source on which the methodology of the present study is based. In particular, a new component was added that controls spectral tilt variation, which the Source Filter Theory did not take into consideration.

According to the Source Filter Theory, the combined source and radiation characteristics constitute a spectrum that falls off at the approximate rate of 6 *dB/octave*. "$|R(f)|$ is approximately proportional to frequency f and $|U(f)|$ is proportional to $1/f^2$. This relation can thus be written

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

where P_k is a constant determining the particular sound pressure level." [5] In equation (1), the combined source and radiation characteristics are represented as a function of frequency. As the frequency of a harmonic is a multiple of the frequency of the fundamental component, the combined source and radiation characteristics can be converted from a function of a continuous variable frequency into a function of a discrete variable fundamental frequency. Substitution of f with f_0 in equation (1) leads to the following expression:

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

The source characteristic can be separated from the radiation characteristic, since the radiation characteristic can be considered to be a function of a positive integer such that the contribution from the radiation characteristic increases at the approximate rate of 6 *dB/octave*. The source characteristic and the radiation characteristic are given in (3) and (4), respectively.

$$U(nf_0) = P_k \frac{(f_0/100)}{1+(nf_0/100)^2} \tag{3}$$

$$R(nf_0) = n \tag{4}$$

The source characteristic can be further divided into the fundamental component and "source ratio", which is defined as the ratio of the source characteristic of the harmonics to that of the fundamental component. If the ratio is transformed on the dB scale, it can be called "source rate". The fundamental component and the source ratio are provided in (5) and (6), respectively.

ISBN 1-876346-48-5

$$U(f_0) = P_k \frac{(f_0/100)}{1+(f_0/100)^2} \quad (5)$$

$$Source \quad Ratio = \frac{1+(f_0/100)^2}{1+(nf_0/100)^2} \quad (6)$$

The amplitude level of the harmonics are anchored to the fundamental component, so that the contribution from the source characteristic of a particular harmonic can be determined by applying the source rate to the contribution from the source characteristic of the fundamental component. As is shown in equation (6), the source ratio is not a constant but a function of fundamental frequency. This means that the ratio of the source characteristic of the overtones to the source characteristic of the fundamental component varies with fundamental frequency.

Although the source rate varies with fundamental frequency, the source rate does not account for variation in spectral tilt which originates from variation in phonation type. In order to account for variation in spectral tilt, I establish a component for controlling spectral tilt:

$$(\frac{1}{n})^k \quad (7)$$

where *n* is a harmonic number. The power *k* in (7) is labeled "phonation type index *k*"(hereafter PTI *k*), because *k* is a factor determining spectral tilt, the variation of which is an important attribute of phonation type.

If the component controlling spectral tilt is incorporated into the source ratio, the source ratio can thus be written

$$Source \quad Ratio = \frac{1+(f_0/100)^2}{1+(nf_0/100)^2}(\frac{1}{n})^k \quad (9)$$

Phonation type index *k* (PTI *k*) is independent of the fundamental frequency but anchored to the fundamental component. All harmonics are affected by PTI *k* in such a systematic way that all harmonics contain some contribution from PTI *k*.

PTI *k* can be considered to be a kind of deviation score. The distribution of PTI *k* may show characteristic differences across speakers and contexts. In particular, the distribution of PTI *k* would demonstrate "mode(s)" and, furthermore, clarify the concept of "modal" voice, when modal voice is defined as the voice that occurs most frequently, i.e. at the mode in the distribution of PTI *k*. An extreme creaky voice may be located at the negative end and an extreme breathy voice at the positive end of the distribution of PTI *k*.

To sum up, the amplitude of a harmonic is composed of the contributions from three characteristics: the source characteristic, the radiation characteristic, and the vocal tract transfer function. The contributions from the source characteristic include three components: the fundamental component, the source rate, and the spectral tilt controller.

3. METHODS

Three male native speakers of Standard Korean in their mid-twenties participated in the experiment. None of the subjects had any history of speech disorders. The material used in the present study was a Korean vowel /a/ immediately following 8 coronal obstruents in the following tokens:

[**t**^{h}ada] "to ride"	[**t**ada] "to be 'ta'"
[**t***ada] "to pick"	[**tʃ**^{h}ada] "to kick"
[**tʃ**ada] "to sleep"	[**tʃ***ada] "to squeeze"
[**s**^{h}ada] "to buy"	[**s***ada] "to wrap"

All tokens are infinitival forms of real Korean verbs except 'tata'. Tokens were embedded in two positions in a carrier sentence /____ ka anila ____ / "It is not '____ ' but '____.'" Subjects were asked to read a list of randomly ordered sentences. Each sentence was repeated 10 times. Recording was conducted in the Phonetics Lab in the Department of Linguistics at the University of Texas at Austin. Speech data were recorded onto analogue tapes via a Marantz SuperScope. The analogue signals were digitized on SoundScope at a sampling rate of 22,050 Hz.

Fundamental frequency (F0), the frequencies, bandwidths, and amplitudes of the first four formants (F1, F2, F3, F4, B1, B2, B3, B4), and the frequencies and amplitudes of the first 20 harmonics were measured. All of the measurements are necessary for the computation of PTI *k*. The frequencies and bandwidths of the first four formants were measured from LPC spectra in SoundScope. The frequencies and amplitudes of individual harmonics were taken from the narrowband FFT spectra in SoundScope. The data from the FFT and LPC spectra were measured in 20 *ms* Hamming windows. The center of the first window was located 10 *ms* from the voice onset in the post-release phase of CV sequences. Measurements were taken from the windows at intervals of 10 *ms* with each window overlapping the preceding window by 10 *ms*.

To obtain PTI *k*, the contributions of the vocal tract transfer function were removed by subtracting the contribution of the vocal tract transfer function from the raw FFT data. The contributions of the radiation characteristic and the source rate were also removed from the amplitude of the harmonics by subtracting the contributions of the radiation characteristic and the source rate from the raw FFT data. A best-fit line was drawn for the remaining data. The contribution from the source characteristic of the fundamental component constitutes *y*-intercept, and *k* forms the slope of the best-fit line for the given data. PTI *k* can be obtained by taking the coefficient of the best-fit line.

ISBN 1-876346-48-5

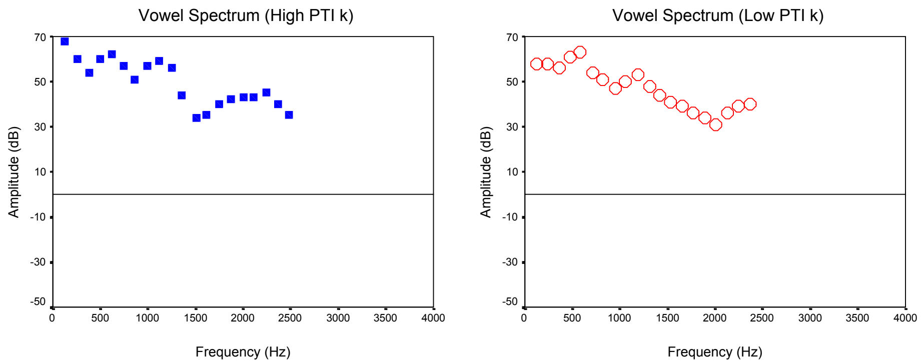

Figure 1 The vowel spectra of two speech samples. The vowel spectrum of the sample with a high PTI *k* is displayed on the left and that with a low PTI *k* on the right. Each data point accounts for the amplitude of the harmonics.

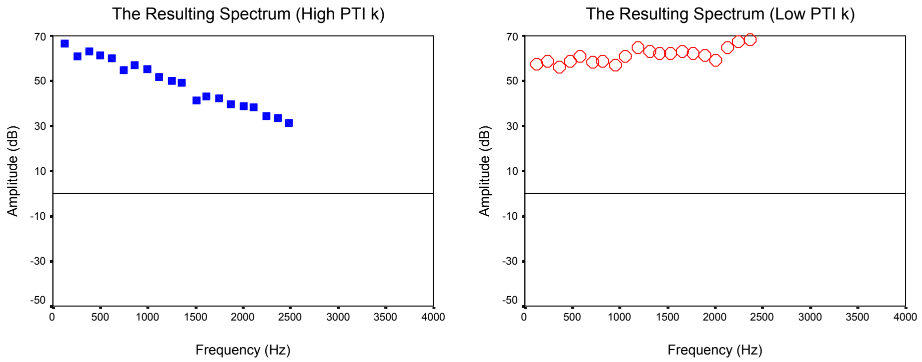

Figure 2 The resulting spectra of the two samples

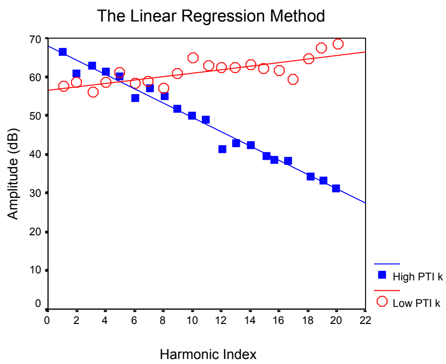

Figure 3 The best-fit lines for the two samples

Examples of obtaining PTI *k* are demonstrated with concrete data above. The spectra of two speech samples, one of which has a high PTI *k* in the final results and the other a low PTI *k*, are illustrated in **Figure 1.**

The contributions of the vocal tract transfer function, the radiation characteristic, and the source rate were subtracted from the vowel spectra provided in **Figure 1**. The resulting spectra in **Figure 2** show that there are apparently linear trends. The best-fit lines drawn from the resulting spectra are presented in **Figure 3**.

Data points were tightly clustered around the best-fit lines. The statistics of the linear regression method are provided in **Table I**.

	Slope	*y*-intercept	R^2	PTI *k*
High PTI *k*	-1.843	67.913	.977	1.843
Low PTI *k*	.449	56.547	.614	-.449

Table I Statistics of the linear regression method

4. RESULTS

The distribution of PTI *k* for each speaker was first examined. The distributions of each speaker were close to normal, as illustrated in **Figure 4**.

ISBN 1-876346-48-5 © 2003 UAB

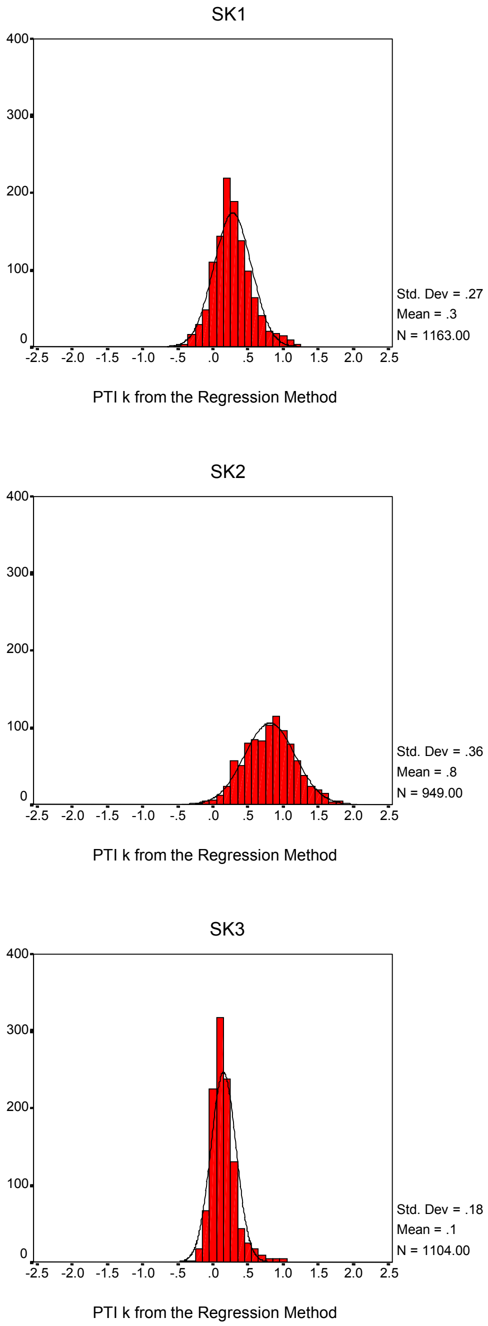

Figure 4 The distributions of PTI *k* for each speaker

Some descriptive statistics are given in **Table II**:

	SK1	SK2	SK3
Mean	0.28	0.81	0.15
Std. Deviation	0.27	0.36	0.18
Range	1.70	2.05	1.49
Kurtosis	0.69	-0.22	3.53

Table II Descriptive statistics of PTI *k*

As shown in **Table II**, there was a substantial difference in specific statistics across speakers. For instance, one speaker had a much higher mean than the other speakers. These descriptive statistics of PTI *k* suggest that modal voice should be speaker-dependent, in the sense that "modal" means "most frequently occurring." The inter-speaker differences in the descriptive statistics also strongly suggest that the specific or combined statistical moments of central tendencies, dispersion, and distribution can be employed as an indicator of speaker-identity.

5. CONCLUSIONS

This study has proposed PTI *k*, a new measure of the overall spectral tilt which is free from the effects of fundamental frequency and vowel quality. Unlike the other measures of spectral tilt, PTI *k* presents a single value for the overall spectral tilt. PTI *k* is useful to measure the overall spectral tilt.

PTI *k* should be validated with more empirical data including a wider range of frequencies in the vowel spectra and a wider range of fundamental frequencies. This study was conducted with only one vowel, /a/. Other vowels, particularly the high vowels /i, u/ which would affect the amplitude of the harmonics in a very different fashion, should be included to properly evaluate the usefulness of PTI *k*. In addition, speech samples showing a greater variation in fundamental frequency should be investigated to observe the effects of fundamental frequency, to properly evaluate the usefulness of PTI *k*, and to investigate the interaction between the fundamental frequency and phonation types.

REFERENCES

[1] K. Stevens and H. Hanson, "Classification of glottal vibration from acoustic measurements", in *Vocal Fold Physiology*: *Voice Quality Control*, O. Fujimura and M. Hirano, Ed., pp. 145-170. San Diego: Singular Publishing Group Inc., 1995.

[2] H. Hanson, "Glottal characteristics of female speakers: acoustic correlates," *Journal of the Acoustical Society of America*, vol. 101(1), pp. 466-481, 1997.

[3] P. Ladefoged, "The linguistic use of different phonation types," in *Vocal Fold Physiology*: *Contemporary Research and Clinical Issues*, D. Bless and J. Abbs, Ed., pp. 351-360. San Diego: College Hill Press, 1983.

[4] P. Kirk, P. Ladefoged, and J. Ladefoged, "Linguistic use of different phonation types," *UCLA Working Papers in Phonetics*, vol. 59, 102-113, 1984.

[5] G. Fant, *Acoustic Theory of Speech Production*, The Hague: Mouton, 1960.

ISBN 1-876346-48-5