

Time Course of the First Formant Bandwidth

HANSANG PARK

The University of Texas at Austin

0. Introduction

Phonetic characteristics of modal or non-modal voice qualities, such as open quotient, spectral tilt, the first formant bandwidth, and the degree of noise, have been measured in terms of the relative amplitude difference between the first two harmonics, the relative amplitude difference between the first harmonic and formant peaks, the first formant bandwidth, and harmonic-to-noise ratio, or subjective noise ratings. It was observed that one of the characteristics of breathy voice is an increased first formant bandwidth (Klatt & Klatt 1990) and that breathiness varies over time within a vowel (Silverman *et al.* 1995). It is predicted from these observations that the first formant bandwidth varies over time as well as across phonation types. This paper investigates the time course of the first formant bandwidth of the vowel /a/ in tokens of type /Ca/ and /VCa/, where C is 8 Korean coronal obstruents /t, t^h, t*, c, c^h, c*, s^h, s*/, to see the behavior of the first formant bandwidth in those contexts.

0.1. Theoretical background

Formant bandwidth, which is measured by taking the width of the band forming 3 dB down from the resonance peak, is a measure of resonance. Resonance, or formant, is described in terms of the peak frequency and its bandwidth. As is well known, the vocal tract transfer function can be represented as a function of formant frequency and bandwidth (Fant 1960:53).

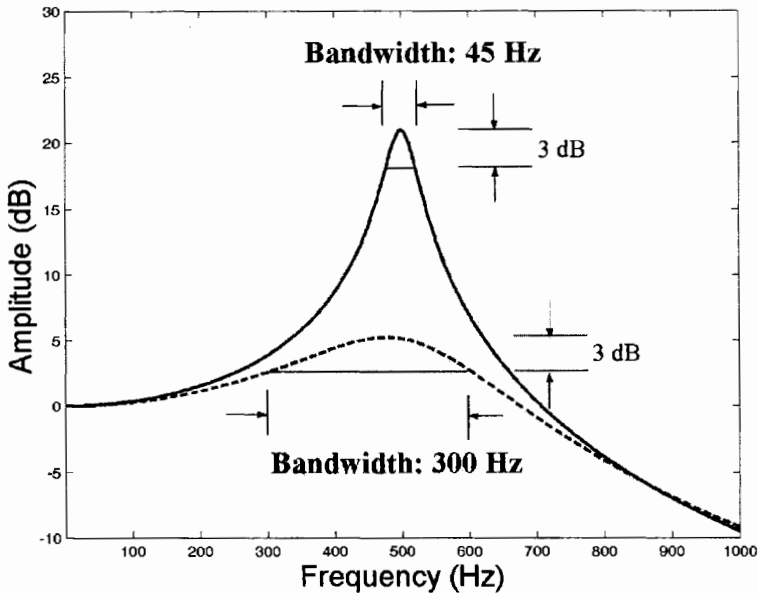
(1) Vocal tract transfer function

$$H_n(f) = \frac{F_n^2 + (B_n/2)^2}{\sqrt{(f - F_n)^2 + (B_n/2)^2} \sqrt{(f + F_n)^2 + (B_n/2)^2}},$$

where F_n is the frequency of the n^{th} formant and B_n the bandwidth of the n^{th} formant.

Formant bandwidth is associated with the amplitude of the formants such that the wider the bandwidth, the lower the peak amplitude, hence the amplitude level of the harmonics. The relationship between formant bandwidth and amplitude is illustrated in (2) below.

(2) Bandwidth and amplitude (adapted from (1))



The abscissa represents frequency and the ordinate amplitude. Two bandwidths, 45 Hz (solid line) and 300 Hz (dashed line), are illustrated with the first formant frequency set at 500 Hz. Other things being equal, the first formant peak decreases at the approximate rate of 6 dB/octave as the first formant bandwidth increases.

Formant bandwidth is determined by acoustic energy losses in the vocal tract. In general, variation in bandwidth results from the power dissipated at frequency f . The general scheme of formant bandwidth variation is represented in (3). (Adler 1957, Stevens 1999)

(3) The general scheme of formant bandwidth variation

$$B = \frac{\text{power dissipated at frequency } f}{2\pi \times \text{total stored energy at frequency } f}$$

Time Course of the First Formant Bandwidth

There are a number of factors affecting the power dissipation, such as radiation, rigidity of the vocal tract walls, heat conduction, and viscosity, airflow at a constriction, and glottal opening. The contributions from these factors account for the numerator in the general scheme.

The contribution to the formant bandwidth due to the radiation impedance depends on the configuration of the vocal tract. If the length of a uniform tube is 17.7 cm and the cross-sectional area is 3 cm^2 , then for the formants at 500, 1500, 2500, and 3500 Hz, the contribution of the radiation loss to the bandwidth of a formant is about 2, 16, 50, and 90 Hz (Stevens 1999). The contribution of the radiation resistance to the bandwidths of the front cavity resonances is particularly important. The contribution of the radiation loss to bandwidth is mainly dependent on the natural frequency of the front cavity resonance, particularly at higher frequencies.

The resistive component of the wall impedance causes energy loss in the vocal tract and an increase in the bandwidths of the formants, particularly the first formant. The bandwidth contribution of the wall resistance for the first formant of a circular uniform tube of length 17.7 cm and diameter 2.0 cm is roughly equal to 11 Hz (Stevens 1999). When the vocal tract configuration is a uniform tube, the reactive component of the wall impedance has a negligible effect on the frequencies of the formants. The contribution to the first formant bandwidth of the resistive component of the wall impedance for a Helmholtz resonator is a function of the first-formant frequency. This frequency is manipulated by changing the size of the opening in the front part of the tube. The bandwidth component decreases with increasing frequency.

The propagation of sound in a tube is accompanied by two kinds of losses at the walls of the tube, in addition to the losses due to the yielding walls (Fant 1960, Flanagan 1972). One type of loss is caused by viscous friction at the wall of the tube. Another source of loss arises because energy is drawn from the acoustic wave by heat conduction at the walls of the tube. The contributions from both viscosity and heat conduction are proportional to the square root of frequency. When the lowest natural frequency of a circular cylindrical tube of cross-sectional area 3 cm^2 is 500 Hz, the contributions from viscosity and heat conduction are 5 Hz and 3 Hz, respectively (Stevens 1999). The contribution from both viscosity and heat conduction become larger as the cross sectional area decreases, assuming the shape of the cross section of the tube does not change as the cross-sectional area changes. These components also increase as the surface area of the tube increases, for example, if the surface contains undulations or irregularities. The contributions to the formant bandwidths arising from viscosity and heat conduction are much less than those from other sources.

In summary, the contributions of viscosity, heat conduction, radiation, and vocal tract walls to the bandwidth depend on frequency and the dimensions of the resonator. Radiation contributes the most to the formant bandwidths at high frequencies, whereas the other three components are more important at low

frequencies. Each of these loss mechanisms depends on frequency, with the wall losses tending to dominate at low frequencies and radiation loss dominating at high frequencies.

Airflow through a constriction can introduce acoustic loss in the form of a nonlinear resistance. This resistance can have a significant effect on the bandwidths of vocal tract resonances and hence on the overall shape of the spectrum of the sound produced by various sources and configurations. The contribution of the constriction to the lowest formant bandwidth is inversely proportional to the cross sectional area of the constriction. The resistance at a short and narrow constriction can have a significant effect on the bandwidths of resonances of cavities anterior or posterior to the constriction in addition to that of the low-frequency Helmholtz resonance. The effect of the constriction on the bandwidth is greater than that of the radiation resistance at low frequencies and less at high frequencies.

During vocal fold vibration, the acoustic impedance of the glottis varies with time. The glottal impedance has the greatest influence on the bandwidth of the first formant. The contribution to the bandwidth of a formant due to losses at glottis for a uniform tube of length 17.7 cm, a cross-sectional area of 3.0 cm^2 , and a subglottal pressure of 8 cm H₂O is approximately proportional to cross sectional area of the glottis (Stevens 1999). During normal voicing a typical value for the average glottal opening during the open phase is about 0.06 cm^2 , and this opening would contribute about 120 Hz to the average first formant bandwidth in this part of the glottal cycle. When the average glottal opening is greater, as it might be during an aspirated consonant, the first formant bandwidth is considerably greater. The bandwidth may be so great that the perturbation method would probably lead to a bandwidth that is comparable to the frequency of the first formant. A wide average bandwidth would also be expected during breathy voicing, when the average glottal area over the entire cycle of the glottal vibration may be 0.06 cm^2 or more.

The effect of glottal impedance on the bandwidth of a formant depends to some extent on the configuration of the vocal tract above the glottis. The effect can be particularly large when the vocal tract is constricted in the region immediately above the glottis (Fant 1979). These contributions to formant bandwidth depend on the glottal configuration used by the speaker and on the subglottal pressure, and may vary somewhat from speaker to speaker. The reduction in cross-sectional area in the pharyngeal region can influence the bandwidth or the degree of spectral prominence of particular formants. For a low vowel with a backed tongue body, the first formant is often associated primarily with the back cavity, particularly for adult male speakers, since most of the acoustic energy at the frequency of the first formant is stored in this pharyngeal region. The contribution of these losses due to pharyngeal constriction to the formant bandwidths is roughly the same for different speakers, except for differences in scale. (Stevens 1999)

0.2. Goals of the present study

Klatt and Klatt (1990: 825) stated, "Additional characteristics of a breathy vowel include the possibility of increased first-formant bandwidth and/or the appearance of tracheal poles and zeros in the vocal tract transfer function due to the greater glottal opening." In other words, a breathy voice is accompanied by a wider glottal opening that causes an increased first formant bandwidth as well as a coupling of the supraglottal and subglottal cavities. This statement suggests that the breathier the vowel, the wider the first formant bandwidth.

Silverman *et al.* (1995) showed that in Jalapa Mazatec, which has modal, breathy, and laryngeal voices, contrastive breathiness lasts for only 43 % of the vowel duration, giving way to modal vibration for the remainder. It is predicted from this observation that the first formant bandwidth varies over time within the breathy vowel itself. It is evident that phonetic characteristics of different phonation types vary over time. Blankenship (1997) showed that phonation type varies over time in terms of H1-H2, H1-F2, and cepstral peak prominence in Tagalog, Mazatec, Chong, Mpi, and Navajo. Ahn (1999) also indicated that voice qualities are different at different time points within a vowel as well as across phonation types in Korean.

Stevens and Hanson (1995) and Hanson (1997) employed first formant bandwidth as well as H1*-H2*, H1*-A1, H1*-A3* in their studies of voice qualities of female speakers. They estimated the first formant bandwidth by measuring the decay rate of the first formant waveform during the early part of the glottal period, where the glottal area is expected to be smallest, under the assumption that F1 oscillation is a damped sinusoid in the form $e^{-\alpha t} \cos 2\pi f t$ where f is the first formant frequency and α is related to the first formant bandwidth by the equation $B1 = \alpha/\pi$.

First of all, they selected 8 consecutive pitch periods from a relatively stable portion of the vowel, generally at the middle. Then, they isolated F1 oscillation by applying a bandpass filter of 600 Hz centered at the first formant frequency. They determined the rate of decay from the change in the peak-to-peak amplitude in the first two cycles of the F1 oscillation. Finally, they estimated the decay rate for 8 consecutive pitch periods and then average 40 estimates to obtain a mean value for each speaker.

They restricted estimation of the first formant bandwidth to a relatively stable portion of a vowel /æ/, providing a caveat that there must be, in order to obtain accurate estimates, a high enough first formant frequency and a long enough pitch period to get at least two oscillations during the closed part of the cycle. This caveat suggests that the estimation method is not adequate for speech signals with a lower F1 or a short pitch period where two oscillations are not available during the closed part of the cycle. This means that the closed part of the glottal cycle, $(1 - OQ)T_0$, where OQ is open quotient and T_0 is a fundamental period, must be greater than double the period of F1 oscillation, T_{F1} . If T_{F1} is considered to be the period of the harmonic closest to the F1, and therefore, the frequency of the F1

oscillation $1/T_{F1}$ can be represented as a multiple of the fundamental frequency, this relation can be roughly expressed as in (4).

(4) Condition for B1 estimation

$$2T_{F1} < T_0(1 - OQ), \text{ where } nT_{F1} = T_0$$

$$0 < (1 - OQ) - 2/n$$

where, n is the index of the harmonic closest to the first formant frequency. According to this condition, their method is applicable to rather a restricted range of speech samples. It is hard to apply this method to speech samples with a low harmonic index due to a high fundamental frequency, with a low harmonic index due to a high vowel with a low first formant frequency, or to breathy voices with a high OQ. It is, in reality, not easy to obtain accurate estimates from female voices which usually have a higher fundamental frequency, hence a low harmonic index, or from breathy voices which have a higher open quotient, since it is difficult to obtain at least two oscillations during the closed part of the fundamental period. As a corollary, they did not estimate B1 for the marginal part of the vowel where quite different voice qualities are likely to be observed.

The present study does not apply the method provided in Stevens and Hanson (1995) and Hanson (1997), since the present study is interested in quite different voices qualities, particularly breathy voice that is highly likely to be observed at the onset of the vowel after an aspirated consonant. This study, instead, simply measures first formant bandwidths from LPC spectra. Even though the LPC algorithm may have its inherent problems, it could be a reasonable way to investigate variation in the first formant bandwidth if the first formant bandwidth is consistently measured for all tokens within a given setting. The present study investigates time course of the first formant bandwidth of a Korean vowel /a/ from the onset of the vowel to the offset, based on the observation that the earlier part of the vowels after different consonant types in Korean have different voice qualities (Ahn 1999). This study may provide yet another profile of phonation type variation.

1. Method

1.1. Subjects

Three male native speakers of Standard Korean in their mid-twenties participated in the experiment. All subjects were students at the University of Texas at Austin who had been in the United States for less than 6 months. Two of them were native Seoulites born to parents who were also native Seoulites while one was born and raised in Kyunggi province, an area surrounding Seoul, to parents who had also been born and raised in the same province. None of the subjects had any history of speech disorders.

1.2. Materials

The material used in the present study is represented in (5).

(5) The material used in the present study

a. Tokens

[t ^h ada]	'to ride'	[tada]	'to be <i>ta</i> '
[t [*] ada]	'to pick'	[tj ^h ada]	'to kick'
[tʃada]	'to sleep'	[tj [*] ada]	'to squeeze'
[s ^h ada]	'to buy'	[s [*] ada]	'to wrap'

b. Carrier Sentence

/σ₁σ₂ ka anila σ₂σ₁ta/ 'It is not σ₁σ₂ but σ₂σ₁'

c. Sample Sentence

[t^hadagaanira t^hada]

All tokens provided in (5a) are infinitival forms of real Korean verbs except *tata*. However, *tata* is also often spoken in real speech, meaning '(The answer) is *ta*'. The vowels in all tokens are phonologically short. Tokens are different from each other only in their initial consonants. Two-syllable words were embedded in a carrier sentence, as is shown in (5b). Utterance-initial CV sequences were available in the first slot of the carrier sentence. The two syllables in the second slot were derived by a metathesis of the token in the first slot of the sentence to obtain VCV sequences. The reason why the present study investigates both utterance-initial CV sequences and non-utterance-initial VCV sequences, even though the latter properly includes the former, is that the present study was interested in positional variation of the first formant bandwidth. An example of the tokens embedded in a carrier sentence is presented in (5c).

1.3. Recording and Digitization

Subjects were asked to read a list of randomly ordered sentences to the effect that each sentence was repeated 10 times. Subjects were given the instruction that each sentence should be separated by a long pause. 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.

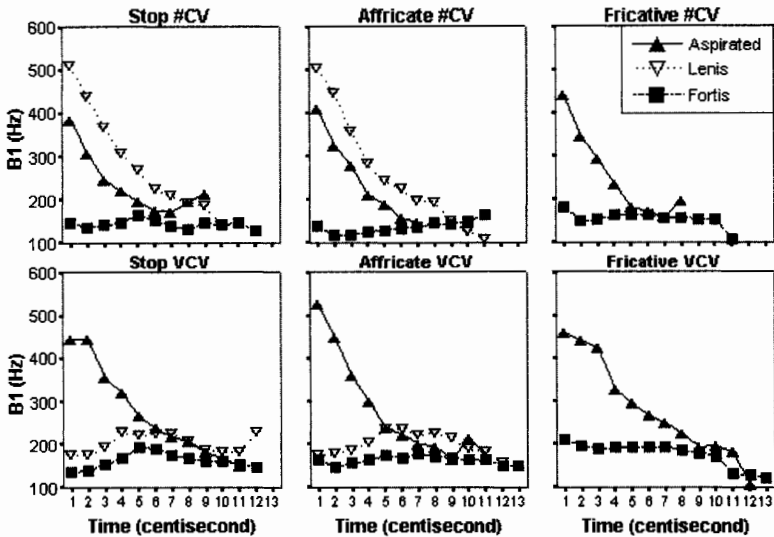
1.4. Measurements

First formant bandwidth was measured from pre-emphasized LPC spectra of 20 ms Hamming windows. The first window was placed so that the center of the window was set at 10 ms from the onset. Measurements were taken from the windows at the interval of 10 ms, overlapping 10 ms with the preceding window. Measurement was not performed for any window whose size was less than 20 ms around the offset of the vowel.

2. Results

Time course of the first formant bandwidth by manner of articulation and position is provided in (6) below.

(6) Time course of the first formant bandwidth by manner and position



In (6), the upper panels indicate the utterance-initial CV sequence (#CV) while the lower panels the intervocalic VCV sequence (VCV). In each position, panels indicate manners of articulation from the left: Stop, Affricate, and Fricative. The abscissa represents time in centisecond from the onset of the vowel, while the ordinate the first formant bandwidth (B1). In each panel, filled triangles stand for the vowel after aspirated obstruents (Aspirated), empty upside-down triangles for the vowel after lenis obstruents (Lenis), and rectangles for the vowel after fortis obstruents (Fortis). It should be noted that Fricative has only two series in Korean.

At word onset, B1 was greatest in Lenis tokens (around 500 Hz) in Stops and Affricates, intermediate in Aspirated tokens (around 400 Hz), and least in Fortis tokens (around 150 Hz). Intervocalically, B1 was greatest in Aspirated tokens (around 500 Hz), intermediate in Lenis tokens (around 200 Hz), and least in Fortis tokens (around 150 Hz).

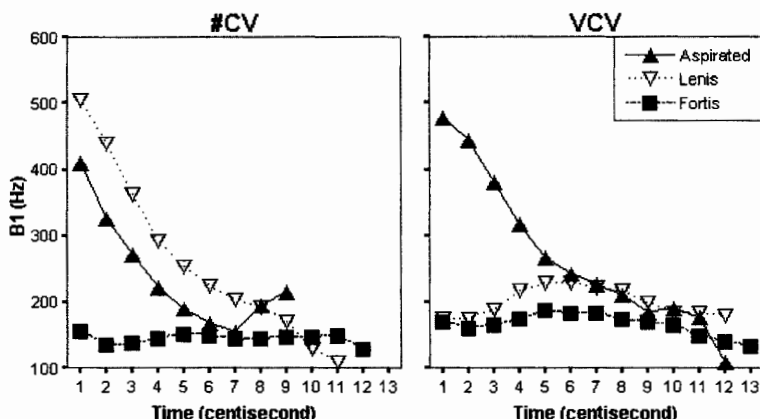
A rapid decrease was observed over time for both Aspirated and Lenis tokens in #CV, while only for Aspirated ones in VCV. There was no substantial change in B1 for Fortis tokens in either position, nor for Lenis tokens in VCV. With a

Time Course of the First Formant Bandwidth

greatest difference in B1 at the onset, the lines representing different phonation types merged around 50% of the vowel duration (around 5th or 6th window). Lenis patterned with Aspirated in #CV, while with Fortis in VCV.

We can summarize the results by presenting time course of B1 by position alone, which is illustrated in (7) below.

(7) Summary of the time course of the first formant bandwidth



A striking difference was observed at the onset of the vowel. The difference was reduced drastically, so that there was no discernible distinction after 50 ms from the onset. Lenis patterned with Aspirated in #CV, while with Fortis in VCV.

3. Discussion and Conclusion

The striking difference in the first formant bandwidth across phonation types at the onset of the vowel suggests that this difference may be a function of the glottal state in the prevocalic phase. In the utterance-initial CV sequence, a wide bandwidth in Aspirated is associated with an open glottis in the prevocalic phase. In other words, vocal fold vibration is initiated with a considerable glottal opening. On the other hand, a narrow bandwidth in Fortis is related to a closed glottis in the prevocalic phase. Vocal fold vibration begins with a completely closed glottis. This closed glottis can be achieved by a voluntary adduction of vocal folds.

In the intervocalic VCV sequence, a wide bandwidth in Aspirated and a narrow bandwidth in Fortis are each associated with an open glottis and a closed glottis in the prevocalic phase, as in the utterance-initial CV sequence. However, a narrow bandwidth in Lenis in the intervocalic VCV sequence is also related to a closed glottis in the prevocalic phase. In Lenis, vocal fold vibration begins with a

completely closed glottis. This closed glottis seems to be achieved by a natural process of intervocalic voicing. In the particular case of Korean, intervocalic voicing of lenis obstruents seems to be a function of closure duration. The intervening interval between two vowels is so short that the voicing of the previous vowel is extended to the following vowel, maintaining vocal fold vibration throughout the interval. In this case, a closed glottis is ready even before the following vowel begins. This glottal state, in turn, leads to a low bandwidth at the onset of the following vowel.

This view is supported by Kagaya (1974). In his fiberoptic study of glottal width of Korean obstruents, Kagaya (1974) showed that glottal width is greatest in Aspirated, intermediate in Lenis, and least in Fortis in the prevocalic phase. This observation is comparable to Kim (1970) for stops. In addition, Kagaya (1974) provided the time point of a complete contact of the vocal folds. For the isolated /CV/ environments, a complete contact of the glottis is observed around voice onset in Aspirated and Lenis while a complete contact of the vocal processes is achieved early before the voice onset in Fortis. For the intervocalic /VCV/ environments, the same pattern occurred in both Aspirated and Fortis as in the /CV/ environments. However, a complete contact of the vocal processes is observed in Lenis all through the utterance. This observation is consistent with the generally accepted allophonic variation of this type of Korean consonants, that is, intervocalic voicing. The vocal processes contact each other again at a point from 40-100 ms before the voice onset as in the case of /CV/ environments.

A wide first formant bandwidth does not abruptly give way to modal vibration but decreases gradually over time in Aspirated in both positions and in Lenis in the utterance-initial CV sequence. The monotonic decrease in the first formant bandwidth during the earlier part of the vowel indicates a gradual yielding to modal vibration. This pattern suggests an interpolation between two target glottal states. This observation is agreement with Silverman *et al.* (1995) who, as noted earlier, observed that in Jalapa Mazatec contrastive phonation types last only for 43% of the vowel duration, giving way to modal phonation for the rest of the vowel duration.

This inference seems to explain why the pattern of variation in the first formant bandwidth is consistent for Aspirated and Fortis across positions, while Lenis patterns with Aspirated in the utterance-initial CV sequence while with Fortis in the intervocalic VCV sequence. However, we did not provide an answer to why the first formant bandwidth is greater in Lenis than in Aspirated in the utterance-initial CV sequences and than in Fortis in the intervocalic VCV sequences. One possible answer may be that Lenis has a wider glottal opening than Aspirated in the utterance-initial CV sequences and than in Fortis in the intervocalic VCV sequences. This inference requires a presupposition that the first formant bandwidth is the only indicator of glottal opening and glottal opening is the only source of an increased first formant bandwidth. No evidence, however, is provided supporting this idea. In the previous section, It was noted that bandwidth is affected by many factors including radiation, vocal tract wall,

heat conduction and viscosity, airflow at a constriction, glottal opening. Among these, radiation, vocal tract wall, and heat conduction and viscosity have a minor effect on the first formant bandwidth. It seems to be clear that glottal opening is major source of an increased first formant bandwidth in the experimental setting of the present study. An alternative is that there is some other source than glottal opening that increases bandwidth. This remains to be explored.

In summary, a spontaneous transition of glottal opening from the prevocalic phase, a voluntary glottal closing, and extension of voicing through the short interval of the intervening consonant creates two possible glottal states at the onset of the vowel: a closed glottis and an open glottis. The default glottal state in the utterance initial position is open while that in the intervocalic position is either closed or open. Vocal fold vibration begins with either of the default states consistently in Aspirated and Fortis regardless of position, whereas the default state varies with position in Lenis. The interaction of these two default glottal states with the positional context results in a distinct behavior of the first formant bandwidth in Lenis.

References

- Adler, R. B., L. J. Chu, & R. M. Fano. 1957. *Electromagnetic energy transmission and radiation*. New York: Wiley.
- Ahn, Hyunkee. 1999. Post-release phonatory processes in English and Korean: Acoustic correlates and implications for Korean phonology. PhD dissertation, The University of Texas at Austin.
- Blankenship, Barbara. 1997. The time course of breathiness and laryngealization in vowels. PhD dissertation, UCLA.
- Fant, Gunnar. 1960. *Acoustic Theory of Speech Production*. The Hague: Mouton.
- Fant, Gunnar. 1979. Glottal source and excitation analysis. Speech Transmission Laboratory Quarterly Progress and Status Report 1, Royal Institute of Technology, Stockholm, 85-107.
- Flanagan, J. L. 1972. *Speech analysis, synthesis, and perception*. Berlin: Springer-Verlag.
- Hanson, Helen M. 1997. Glottal characteristics of female speakers: Acoustic correlates. *Journal of the Acoustical Society of America* 101: 466-481.
- Kagaya, Ryohei. 1974. A fiberoptic and acoustic study of the Korean stops, affricates and fricatives. *Journal of Phonetics* 2: 161-180.
- Kim, Chin-Wu. 1970. A theory of aspiration. *Phonetica* 21: 107-116.
- Klatt, Dennis H. & Laura C. Klatt. 1990. Analysis, synthesis, and perception of voice quality variations among female and male talkers. *Journal of the Acoustical Society of America* 87(2): 820-857.
- Silverman, D., B. Blankenship, P. Kirk, & P. Ladefoged. 1995. Phonetic structures in Jalapa Mazatec. *Anthropological Linguistics* 37(1): 70-88.
- Stevens, Kenneth N. 1999. *Acoustic Phonetics*. Cambridge: MIT Press.

Stevens, Kenneth N. & Helen M. Hanson. 1995. Classification of glottal vibration from acoustic measurements. In O. Fujimura & M. Hirano (eds.) *Vocal fold physiology: Voice quality control*, San Diego: Singular, 145-170.

Hansang Park
Department of Linguistics
The University of Texas at Austin
Austin, TX 78712

phans@mail.utexas.edu