The Relationship between the Source Characteristic and Frequency

HANSANG PARK
University of Texas at Austin

0. Introduction

The Source Filter Theory assumes that the amplitude level of the source characteristics of the harmonics falls off at the approximate rate of -12 dB/octave in the source spectrum.

> The combined source and radiation characteristics constitute a spectrum that falls off at the approximate rate of 6 dB/octave. $|R(f)|$ is proportional to frequency, f, and it will be assumed that $|U(f)|$ is approximately proportional to $1/f^2$ above cutoff frequency of 100 c/s. This relation can thus be written
>
> (1) The combined source and radiation characteristics
>
> $$|U(f)||R(f)| = P_k \frac{(f/100)}{1+(f/100)^2}, \qquad (1.3\text{-}2)$$
>
> where brackets indicate absolute values and (f) function of frequency, $|U(f)|$ symbolizes the amplitude-versus-frequency characteristics of the source, $|R(f)|$ the frequency characteristic of radiation, and P_k is a constant determining the particular sound pressure level. (Fant 1960:49)

This paper will make it clear that the source rate is not fixed to -12 dB/octave but varies with frequency, such that the source rate is not a constant but a variable. Specifically, this paper will first derive the ratio between the source characteristics of the first two harmonics.[1] Second, this paper will calculate the rate in dB/octave from the ratio between the source characteristics of the first two harmonics at certain frequencies and illustrate the relationship between the source rate and frequency. Finally, this paper will reconsider the existing metrics for phonation type that employ the first two harmonics of the source spectrum and, furthermore, propose a new metric for phonation type: *Phonation Type Index K*.

[1] The discussion of the ratio will be restricted to the first two harmonics for the time being, so that only f_0 and $2f_0$ will appear in the discussion.

The finding that the source rate is not fixed to -12 dB/octave but varies with frequency is important in two respects. First, the ratio and the rate can be applied to the whole range of frequencies. It is not necessary to fix a cutoff frequency. The only limitation is the range of frequency within which human beings can produce speech sounds (humans are generally not thought to produce speech sounds with a fundamental frequency below 50 Hz or above 2000 Hz).

Second, it is necessary to reconsider the measures of phonation type, such as H1-H2, H1*-H2*, and Cor(H1-H2), considering the contribution of the varying rate to the amplitude of the harmonics. The difference in amplitude between the first two harmonics has been used as a metric of phonation type. H1-H2 was an acceptable metric of the phonation type for low vowels minimally affected by the first formant, while it was not appropriate for high vowels, since F1 boosts the amplitude level of the first two harmonics. H1*-H2* was a corrected metric of the phonation types (Stevens and Hanson 1995), but details of the metric are not available. Cor(H1-H2) was recently designed by Ahn (1999) based on the Source Filter Theory. It has two advantages over H1-H2. One is that the contributions of filter and radiation characteristics can be removed. The other is that it is possible to compare different phonation types in terms of the relative difference between the observed H1-H2 and the expected H1-H2, even though the absolute values of the source characteristics of the harmonics are not available. However, these measures did not clarify the contribution of the varying rate that may lead to a significantly different result. If we take the varying rate into consideration, we can provide more reliable data with which we can determine the significance of the difference.

1. The Source Characteristic and Frequency

1.1. The Ratio between $|U(f)|$ and $|U(2f)|$

This section shows how the ratio between the source characteristics of the first two harmonics was derived. The derivation of the ratio between $|U(f)|$ and $|U(2f)|$ is shown in (2). It should be noted that the ratio between the source characteristics of the first two harmonics was derived from the formula in (1).

The first two rows in (2-1) and (2-2) represent the combined source and radiation characteristics at the fundamental frequency f_0 and at double the fundamental frequency $2f_0$. The second row is obtained by substituting f_0 with $2f_0$.

(2) Derivation of the ratio between $|U(f)|$ and $|U(2f)|$

$$(2-1): |U(f_0)||R(f_0)| = P_k \frac{100(f_0)}{10{,}000 + (f_0)^2}$$

$$(2-2): |U(2f_0)||R(2f_0)| = P_k \frac{200(f_0)}{10{,}000 + 4(f_0)^2}$$

$$(2-3): \frac{|U(2f_0)||2R(f_0)|}{|U(f_0)||R(f_0)|} = \frac{P_k \dfrac{200(f_0)}{10{,}000 + 4(f_0)^2}}{P_k \dfrac{100(f_0)}{10{,}000 + (f_0)^2}}$$

$$(2-4): \frac{|U(2f_0)||2R(f_0)|}{|U(f_0)||R(f_0)|} = \frac{\dfrac{200(f_0)}{10{,}000 + 4(f_0)^2}}{\dfrac{100(f_0)}{10{,}000 + (f_0)^2}}$$

$$(2-5): \frac{|U(2f_0)||R(2f_0)|}{|U(f_0)||R(f_0)|} = \frac{\dfrac{2}{10{,}000 + 4(f_0)^2}}{\dfrac{1}{10{,}000 + (f_0)^2}}$$

$$(2-6): \frac{|U(2f_0)||R(2f_0)|}{|U(f_0)||R(f_0)|} = \frac{2(10{,}000 + (f_0)^2)}{10{,}000 + 4(f_0)^2}$$

$$(2-7): \frac{|U(f_0)|}{|U(2f_0)|}\frac{2}{1} = \frac{2(10{,}000 + (f_0)^2)}{10{,}000 + 4(f_0)^2}$$

$$(2-8): \frac{|U(2f_0)|}{|U(f_0)|} = \frac{10{,}000 + (f_0)^2}{10{,}000 + 4(f_0)^2}$$

The third row in (2-3) represents the ratio between the combined source and radiation characteristics at f_0 and at $2f_0$. The third row in (2-3) is obtained by dividing the combined source and radiation characteristics at $2f_0$ by those at the fundamental frequency f_0. The fourth and fifth rows in (2-4) and (2-5) show the arrangement of the right-hand side by canceling P_k and 100(f_0), which both the numerator and denominator have in common. The row in (2-6) is the result of the computation. It is necessary to remove the ratio between |R(2f_0)| and |R(f_0)|, since we are interested in the ratio between |U(2f_0)| and |U(f_0)|. The ratio between |R(2f_0)| and |R(f_0)| is 2. The row in (2-7) represents the substitution of the ratio of the radiation characteristics by 2. The row in (2-8) shows the ratio between the source characteristics of the first two harmonics obtained by canceling the 2 in both sides. It is apparent from the derivation in (2) that the ratio is a function of frequency.

1.2. The Source Rate

It is necessary to examine the source rates to see how the source rate varies with frequency. The source rate is calculated by taking the logarithm with base 10 of the ratio between the source characteristics of the first two harmonics and multiplying 20 to the logarithm. For example, the source rate in dB/octave equals -7.96, as seen in (3).

(3) The source rate in dB/octave at 100 Hz

$$\frac{|U(200)|}{|U(100)|} = \frac{10{,}000 + 100^2}{10{,}000 + 4(100)^2} = \frac{2}{5} = .4$$

$$20\log_{10}\frac{|U(200)|}{|U(100)|} = -7.96$$

As another example, the source rate when the frequency approaches infinity is -12.04 dB/octave, as seen in (4).

(4) The source rate in dB/octave when frequency approaches infinity

$$\lim_{f\to\infty}\frac{|U(2f)|}{|U(f)|} = \lim_{f\to\infty}\frac{(10{,}000 + f^2)}{(10{,}000 + 4f^2)} = \frac{1}{4} = .25$$

$$20\log_{10}\lim_{f\to\infty}\frac{|U(2f)|}{|U(f)|} = 20\log_{10}\lim_{f\to\infty}\frac{(10{,}000 + f^2)}{(10{,}000 + 4f^2)}$$

$$= -12.04$$

In the same way, we can obtain the source rates at given frequencies. The source rates in dB/octave at some frequency values are given in (5).

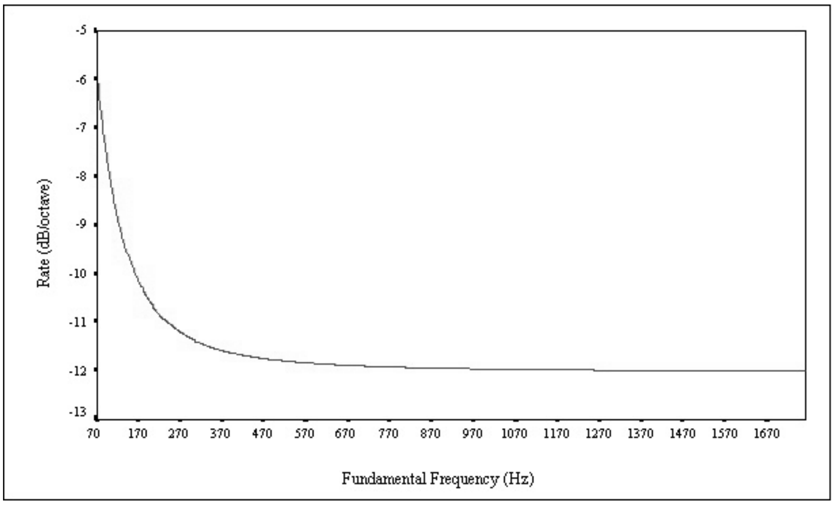
(5) The source rates in dB/octave at given frequencies

Frequency in Hz	Ratio	Rate in dB/octave	Remarks
1	1.00	0	Lowest positive integer
20	0.90	-0.95	Lower threshold of audibility
70	0.63	-5.96	Lowest F0 of human voice
100	0.40	-7.96	
150	0.33	-9.79	Mean of male voice
225	0.29	-10.89	Mean of female voice
250	0.28	-11.09	Mean of children's voice
20000	0.25	-12.04	Higher threshold of audibility
∞	0.25	-12.04	Infinity

As seen in (5), the source rate in dB/octave at mean fundamental frequencies of male, female, and children's voices, which are estimated to occur around 150, 225, and 250 Hz, respectively, are -9.79 for male voices, -10.89 for female voices, and -11.09 for children's voices, respectively. On the other hand, the source rates in dB/octave at given extreme values of frequencies are also available. The source rates in dB/octave at the thresholds of audibility, that is 20 Hz and 20000 Hz, are -0.95 and -12.04, respectively. The source rate in dB/octave at the lowest integer, or in the case where the frequency equals 1, is 0.

It is apparent that the rate is not fixed to -12 dB/octave but varies with frequency; the rate is not a constant but a variable. The relationship between the source rate and frequency is illustrated in (6).

(6) The source rate in dB/octave over fundamental frequency



In (6), the x-axis represents frequency, while the y-axis represents the source rate in dB/octave. The range of frequency is 1700 Hz, since the minimum fundamental frequency human beings can produce is estimated to be 70 Hz and the maximum fundamental frequency is estimated to be 1770 Hz. It should be noted that ordinary speech sounds seldom exceed 1000 Hz. We can see that the source rate varies dramatically in lower frequencies. The value of -12 dB/octave is never applicable to lower frequencies below 200 Hz, rather only to frequencies higher than 500. The fact that male voices seldom reach that frequency suggests that male voices are seriously influenced by the varying source rate. The lower the frequency, the more significant the effect of the rate is.

2. Reconsideration of the Metrics of Phonation Type

Since the source rate is not fixed to -12 dB/octave but varies with frequency, the source rate is not a constant but a variable. This fact leads us to reconsider the metrics of phonation type, since the single most widely used metric has been difference in amplitude between the first two harmonics—that is, H1-H2—and the varying source rate may significantly affect the values of H1-H2. In this section, such metrics of phonation type [2] as H1-H2 (Huffman 1987) and Cor(H1-H2) will be examined.

2.1. H1-H2

Phonation type has been measured in terms of difference in amplitude between the first two harmonics—that is, H1-H2 (Huffman 1987). H1 and H2 can be represented as in (7), respectively.

(7) Representation of H1 and H2

$$HARMONIC_1 = H_1 = 20\log_{10}|Q(f_0)U(f_0)||R(f_0)||H_n(f_0)|$$
$$HARMONIC_2 = H_2 = 20\log_{10}|Q(2f_0)U(2f_0)||R(2f_0)||H_n(2f_0)|$$

This representation is based on the Source Filter Theory where sound pressure is the product of source, radiation, and filter characteristics. Terms for phonation type, $Q(f_0)$ and $Q(2f_0)$, were inserted into the source characteristic, since phonation type was basically considered to be a factor of the source characteristic.

H1-H2 can be calculated by subtracting H2 from H1. H1-H2 is represented in (8).

[2] Stevens and Hanson (1995) also proposed H1*-H2*. However, details of the metric are not available.

(8) Representation of H1-H2

$$H_1 - H_2 = A + B + C + D$$
$$A = DIFF(PHONATIONTYPE) = 20\log_{10}|Q(f_0)| - 20\log_{10}|Q(2f_0)|$$
$$B = DIFF(SOURCERATE) = 20\log_{10}|U(f_0)| - 20\log_{10}|U(2f_0)|$$
$$C = DIFF(RADIATION) = 20\log_{10}|R(f_0)| - 20\log_{10}|R(2f_0)| = 6dB/octave$$
$$D = DIFF(TRANSFER) = 20\log_{10}|H_n(f_0)| - 20\log_{10}|H_n(2f_0)|$$

As seen in (8), H1-H2 can be represented as the sum of four differences: differences in phonation type (A), the source rate (B), the radiation characteristic (C), and vocal tract transfer function (D). The difference in the radiation characteristic is fixed at 6 dB/octave. The source rate (B) varies with F0 and the vocal tract transfer function (D) also varies with F0 as well as formant frequencies, which result from a difference in vowel quality. However, H1-H2 did not remove the contributions of the radiation characteristic, the vocal tract transfer function, or the varying source rate. H1-H2 is acceptable only if B and D stay the same across the speech samples.

2.2. Cor(H1-H2)

Phonation type has also been measured in terms of relative difference between the observed difference between the first two harmonics (Obs(H1-H2)) and the expected difference between the first two harmonics (Exp(H1-H2)) (Ahn 1999). Exp(H1-H2) was subtracted from Obs(H1-H2) to remove the contributions from the radiation and the vocal tract transfer function. Observed H1 and H2 can be represented as in (9).

(9) Representation of Obs(H1) and Obs(H2)

$$Obs(H_1) = 20\log_{10}\left|Q(f_0)P_k \frac{(f_0/100)}{1+(f_0/100)^2}\right| |R(f_0)| |H(f_0)|$$
$$Obs(H_2) = 20\log_{10}\left|Q(2f_0)P_k \frac{(f_0/100)}{1+(2f_0/100)^2}\right| |R(2f_0)| |H(2f_0)|$$

Obs(H1) and Obs(H2) are the same as H1 and H2 in (9), respectively, except that the source characteristics are divided into the constant determining the particular sound pressure level (P_k) and frequency response of the source characteristic. On the other hand, Exp(H1) and Exp(H2) are represented in (10).

(10) Representation of Exp(H1) and Exp(H2)

$$Exp(H_1) = 20\log_{10}\left|Q'(f_0)P_k \frac{(f_0/100)}{1+(f_0/100)^2}\right|\left|R(f_0)\right|\left|H(f_0)\right|$$

$$Exp(H_2) = 20\log_{10}\left|Q'(2f_0)P_k \frac{(f_0/100)}{1+(2f_0/100)^2}\right|\left|R(2f_0)\right|\left|H(2f_0)\right|$$

Ahn (1999) assumed that the vocal tract transfer functions of Exp(H1) and Exp(H2) are the same as those of Obs(H1) and Obs(H2). He also assumed that difference in phonation type equals 0—that is, that $Q'(f_0)$ equals $Q'(2f_0)$. In addition, he fixed the source rate to -12 dB/octave and set P_k to 1.

Based on the representations given in (9) and (10), Obs(H1-H2) and Exp(H1-H2) can be represented as in (11) and (12), respectively.

(11) Obs(H1-H2)

$$Obs(H_1 - H_2) = A + B + C + D$$

$$A = DIFF(PHONATIONTYPE) = 20\log_{10}\left|Q(f_0)\right| - 20\log_{10}\left|Q(2f_0)\right|$$

$$B = b1 + b2$$

$$b1 = DIFF(PRESSURE) = 20\log_{10} P_k - 20\log_{10} P_k$$

$$b2 = DIFF(SOURCERATE) = -20\log_{10}\left|\frac{1+(f_0/100)^2}{1+(2f_0/100)^2}\right|$$

$$C = DIFF(RADIATION) = 20\log_{10}\left|R(f_0)\right| - 20\log_{10}\left|R(2f_0)\right| = 6dB/octave$$

$$D = DIFF(TRANSFER) = 20\log_{10}\left|H(f_0)\right| - 20\log_{10}\left|H(2f_0)\right|$$

(12) Exp(H1-H2)

$$Exp(H_1 - H_2) = A + B + C + D$$

$$A = DIFF(PHONATIONTYPE) = 20\log_{10}\left|Q'(f_0)\right| - 20\log_{10}\left|Q'(2f_0)\right| = 0$$

$$B = b1 + b2$$

$$b1 = DIFF(PRESSURE) = 20\log_{10} 1 - 20\log_{10} 1 = 0$$

$$b2 = DIFF(SOURCERATE) = -20\log_{10}\left|\frac{1+(f_0/100)^2}{1+(2f_0/100)^2}\right| = -12dB/octave$$

$$C = DIFF(RADIATION) = 20\log_{10}\left|R(f_0)\right| - 20\log_{10}\left|R(2f_0)\right| = 6dB/octave$$

$$D = DIFF(TRANSFER) = 20\log_{10}\left|H(f_0)\right| - 20\log_{10}\left|H(2f_0)\right|$$

As seen in $b1$ in (12), it is unnecessary to set P_k to 1, since P_k is cancelled by subtracting H2 from H1. It is improper to set P_k to 1, since that means that there is no audible pressure. It is also improper to set the source rate to -12 dB/octave, since it varies with frequency.

On the other hand, Cor(H1-H2) can be obtained by subtracting Exp(H1-H2) from Obs(H1-H2), which is shown in (13).

(13) Cor(H1-H2)

$$\begin{aligned} &Cor(H1-H2) \\ &= Obs(H_1-H_2) - Exp(H_1-H_2) \\ &= DIF(A) + DIF(B) + DIF(C) + DIF(D) \end{aligned}$$

$$\begin{aligned} &DIF(A) = (20\log_{10}|Q(f_0)| - 20\log_{10}|Q(2f_0)|) - 0 \\ &= 20\log_{10}|Q(f_0)| - 20\log_{10}|Q(2f_0)| \\ &DIF(B) = DIF(b1) + DIF(b2) \\ &DIF(b1) = 0 - 0 = 0 \\ &DIF(b2) = -20\log_{10}\left|\frac{1+(f_0/100)^2}{1+(2f_0/100)^2}\right| + 12 \\ &DIF(C) = 6 - 6 = 0 \\ &DIF(D) = 0 \end{aligned}$$

As seen in (13), Cor(H1-H2) is the sum of the differences in phonation type and the source rate between Obs(H1-H2) and Exp(H1-H2). The differences in phonation type and the source rate vary with fundamental frequency, while all other differences equal 0. It was claimed that Exp(H1-H2) was subtracted from Obs(H1-H2) to remove the contribution of the vocal tract transfer function. However, it seems that the vocal tract transfer function in Obs(H1-H2) was substituted with the calculated values of the vocal tract transfer function in Exp(H1-H2). It was stated that the source rate was fixed to -12 dB/octave, but in reality, the source rate was not fixed to -12 dB/octave, since Cor(H1-H2) was calculated by subtracting Exp(H1-H2) from Obs(H1-H2) without fixing the source rate to -12 dB/octave. Ahn (1999) seems to have been unaware of the varying source rate. However, Cor(H1-H2) is well designed in the sense that it tried to remove the contribution of the filter characteristic. Cor(H1-H2) can be a good metric of phonation type for speech samples with varying F0 and formant frequencies only if Cor(H1-H2) incorporates the varying source rate into the metric.

3. Phonation Type Index *K*

It was noted that Cor(H1-H2) could serve as a metric of phonation type if it incorporates the varying source rate into the metric without fixing the source rate to -12 dB/octave. However, it only describes how speech samples are different with respect to phonation type. It does not explain why spectral tilt varies with phonation types. In this section, a general term for harmonics will be presented to explain why phonation type varies across speech samples. A general term for harmonics is given in (14).

(14) A general term for harmonics

$$HARMONIC_n(f_0) = SOURCE(f_0) + RADIATION + TRANSFER(f_0)$$

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

$$RADIATION = 20\log_{10} n$$

$$TRANSFER(f_0) = 20\log_{10}|H(nf_0)|$$

As seen in (14), the source characteristic is the product of the constant determining the particular sound pressure level, the frequency response at the fundamental frequency, the source rate, and the phonation type. It should be noted that the combined source and radiation characteristics was first divided into the radiation characteristic and the source characteristic which was, in turn, divided into the frequency response of the fundamental component and the source rate. The only difference between the present model and the Source Filter Theory is the term for phonation type, that is, $(1/n)^K$. This term was established under the rationale that it is responsible for spectral tilt and, therefore, varies with phonation type, and that it contributes nothing to the spectral tilt in H1 or when *K* equals 0. According to the general term for harmonics, H1 and H2 can be represented as in (15).

(15) Representation of H1 and H2

$$H_1(f_0) = 20\log_{10}\left|P_k \frac{(f_0/100)}{1+(f_0/100)^2}\right||H(f_0)|$$

$$H_2(f_0) = 20\log_{10}\left|P_k \frac{(f_0/100)}{1+(f_0/100)^2}\frac{1+(f_0/100)^2}{1+(2f_0/100)^2}(\frac{1}{2})^K\right|2|H(2f_0)|$$

Difference in phonation type can be calculated by subtracting H2 from H1. The difference in phonation type is represented in (16).

(16) Difference in phonation type

$$[H_1 - H_2](f_0) = A + B + C + D$$

$$A = DIFF(PHONATIONTYPE) = -20\log_{10}(\frac{1}{2})^K$$

$$B = DIFF(SOURCERATE) = -20\log_{10}\frac{10{,}000 + (f_0)^2}{10{,}000 + 4(f_0)^2}$$

$$C = DIFF(RADIATION) = -20\log_{10} 2$$

$$D = DIFF(TRANSFER) = 20\log_{10}|H(f_0)| - 20\log_{10}|H(2f_0)|$$

$$DPT(f_0) = A = [H1 - H2](f_0) - (B + C + D)$$

$$PTI \quad K = \frac{A}{6.02}$$

Difference in phonation type (DPT) can be obtained by subtracting the difference in the source rate (B), the radiation characteristic (C), and the filter characteristic (D) from the measured values of H1-H2. Furthermore, phonation type index *K* can be obtained by dividing the difference in phonation type (A) by 6.02. DPT is the same as Cor(H1-H2) except that phonation type systematically controls the source rate and difference in phonation type is represented by phonation type index *K*. It should be noted that fundamental frequency must always be specified, since the source rate can be calculated only when fundamental frequency is specified. It is important to take the varying source rate into account when phonation types are compared across speech samples with substantially different fundamental frequency. Taking the varying source rate into consideration, we can provide more reliable data by which the significance of difference in phonation type can be determined.

4. Conclusion

The ratio between the source characteristics of H1 and H2 was derived. It was shown that the source rate varies with frequency. This finding was significant, since the varying source rate affects the amplitude difference between the first two harmonics. Metrics of phonation type, such as H1-H2 and Cor(H1-H2), were evaluated from this perspective. It was noted that it is important to incorporate the varying source rate into the metric of phonation type. In addition, special attention needs to be paid when speech samples with a substantial difference in fundamental frequency are compared. The existing metrics of phonation type do not account for why spectral tilt varies with phonation type. A new metric of phonation type, phonation type index *K*, was designed to account for the difference in spectral tilt. It has advantages over H1-H2 or Cor(H1-H2) in that it can account for variation in spectral tilt not only at the first two harmonics but also at higher harmonics. However, much empirical data are necessary to validate DPT and phonation type index *K*. This remains to be done.

References

Ahn, Hyunkee. 1999. Post-Release Phonatory Processes in English and Korean: Acoustic Correlates and Implications for Korean Phonology. Ph.D. diss., University of Texas at Austin.

Fant, Gunnar. 1960. *Acoustic Theory of Speech Production*. The Hague: Mouton.

Huffman, Marie. 1987. Measures of Phonation Type in Hmong. *Journal of the Acoustical Society of America* 81:495-504.

Stevens, Kenneth N., and Helen Hanson. 1995. Classification of Glottal Vibration from Acoustic Measurements. In O. Fujimura and M. Hirano (eds.), *Vocal Fold Physiology*. San Diego: Singular Publishing Group, Inc.

Department of Linguistics
Calhoun Hall 501
The University of Texas at Austin
Austin, TX 78712-1196

phans@mail.utexas.edu