Perceptual sensitivity to first harmonic amplitude in the voice source[a)]

Jody Kreiman[b)] and Bruce R. Gerratt

Division of Head and Neck Surgery, School of Medicine, University of California, Los Angeles, 31-24 Rehab Center, Los Angeles, California 90095-1794

(Received 12 January 2010; revised 18 June 2010; accepted 19 July 2010)

Little is known about the perceptual importance of changes in the shape of the source spectrum, although many measures have been proposed and correlations with different vocal qualities (breathiness, roughness, nasality, strain…) have frequently been reported. This study investigated just-noticeable differences in the relative amplitudes of the first two harmonics (H1–H2) for speakers of Mandarin and English. Listeners heard pairs of vowels that differed only in the amplitude of the first harmonic and judged whether or not the voice tokens were identical in voice quality. Across voices and listeners, just-noticeable-differences averaged 3.18 dB. This value is small relative to the range of values across voices, indicating that H1–H2 is a perceptually valid acoustic measure of vocal quality. For both groups of listeners, differences in the amplitude of the first harmonic were easier to detect when the source spectral slope was steeply falling so that F0 dominated the spectrum. Mandarin speakers were significantly more sensitive (by about 1 dB) to differences in first harmonic amplitudes than were English speakers. Two explanations for these results are possible: Mandarin speakers may have learned to hear changes in harmonic amplitudes due to changes in voice quality that are correlated with the tones of Mandarin; or Mandarin speakers' experience with tonal contrasts may increase their sensitivity to small differences in the amplitude of F0 (which is also the first harmonic).
 [DOI: 10.1121/1.3478784]

PACS number(s): 43.71.Bp [AJ] Pages: 2085–2089

I. INTRODUCTION

Little is known about the perceptual importance of changes in the shape of the source spectrum, although many measures have been proposed (see Kreiman *et al.*, 2007a) and correlations with different vocal qualities (breathiness, roughness, nasality, strain…) have frequently been reported (e.g., de Krom, 1995; Lee *et al.*, 2009; Hartl *et al.*, 2001; Klich, 1982; Hammarberg *et al.*, 1980). The relative amplitudes of the first two harmonics (H1–H2) in particular have often been associated correlationally with both breathy voice quality (e.g., Hillenbrand *et al.*, 1994; Klatt and Klatt, 1990; Sundberg and Gauffin, 1979) and with phonemically meaningful contrasts among different phonation types. For example, phonemically breathy phonation in White Hmong is characterized by consistently and significantly higher H1–H2 values (and thus more energy in the fundamental than in the second harmonic) than is modal phonation (an average of 9.48 dB versus 2.0 dB; Huffman, 1987). Measures of H1–H2 (or $H1^*-H2^*$)[1] also distinguish significantly between tense and non-tense phonation in Chong (DiCanio, 2009), between breathy and non-breathy phonation in Green Mong (Andruski and Ratliff, 2000), between creaky and modal phonation in Coatzospan Mixtec (Gerfen and Baker, 2005), and between breathy and clear vowels in Khmer (Wayland and Jongman, 2003) (Table I).

Despite this abundance of correlational evidence, no study to our knowledge has demonstrated causation between changes in H1–H2 (or any other measure of source spectral slope) and perceived voice quality. It is reasonable to hypothesize that listeners are sensitive to changes in H1–H2, in part because of the well-established pattern of correlation but also because H1–H2 accounts for substantial variance (about 30%) in source spectral shapes across voices (Kreiman *et al.*, 2007a), and thus may also account for source-related changes in voice quality. This study tested this hypothesis by measuring just-noticeable differences in H1–H2. If these differences are small relative to the range of H1–H2 across voices, we may reasonably conclude that the parameter is perceptually important.

II. METHODS

A. Stimuli

Stimuli were created using the UCLA voice synthesizer (Kreiman *et al.*, 2010). Eight natural target voices (four males and four females) were selected from a library of samples. Because previous studies (Klatt and Klatt, 1990; Kreiman and Gerratt, 2005; Shrivastav and Sapienza, 2006) indicate that source spectral slope interacts with spectral noise levels in voice quality perception, voices were selected to represent all combinations of two noise-to-signal ratios (NSR; low and high) and two rates of source spectral slope roll-off (relatively quick—a quasi-sinusoidal source—and relatively slow; Table II). A one-second sample of the vowel /a/ produced by each speaker was copied with the synthesizer such that the synthetic vowels formed good acoustic and

[a)]Preliminary reports of this research were presented at the 6th International Conference on Voice Physiology and Biomechanics in Tampere, Finland, and at the 154th meeting of the Acoustical Society in New Orleans, LA.

[b)]Author to whom correspondence should be addressed. Electronic mail: jkreiman@ucla.edu

TABLE I. Linguistically-contrasting values of H1–H2 for selected languages.

Study	Language	Average H1–H2 or H1*–H2* values (dB)
Andruski and Ratliff, 2000	Green Mong	Breathy: 7.00 Nonbreathy: −0.54
Huffman, 1987	Hmong	Breathy: 9.48 Modal: 2.00
Gerfen and Baker, 2005	Coatzospan Mixtec	Modal: −3.13 Creaky: −9.27
Wayland and Jongman, 2003	Khmer	Breathy: 0.66 Clear: −2.21

perceptual matches to the original voices. Spectral slopes and NSR values were subsequently manipulated slightly to increase the orthogonality between conditions with respect to the independent variables.

Each of these eight synthetic voices ("standards") was then used as the basis for creating two series of stimuli, one in which H1 increased in amplitude relative to the standard value, and one in which H1 decreased in amplitude. The amplitude of the first harmonic was manipulated in 15 steps of 0.5 dB up or down from the original value, as follows. To avoid distorting effects of spectral leakage on harmonic amplitudes, the source spectrum was obtained by performing a pitch synchronous Fourier transform. The first two harmonics were selected in this spectrum, as shown in Fig. 1(a). The slope of the line segment connecting these harmonics in the figure was then increased or decreased by altering the amplitude of the first harmonic while leaving all other harmonics unchanged. The new time-domain source waveform was generated by inverse Fourier transform, after which the voice was resynthesized with the new source but with all other parameters held constant [Fig. 1(b)] (see Kreiman *et al.*, 2010, for more details regarding this process). In this manner, we created two sets of stimuli for each target voice: one in which H1 amplitude increased in 15 steps of 0.5 dB, and one in which it decreased in 15 steps of 0.5 dB.

TABLE II. Characteristics of the standard voice stimuli. The first value in each cell represents the female voice sample; the second was used for male stimuli. Values in the first, second, and third rows (labeled 'A') were used in Experiment 1; values in the second, third, and fourth rows (labeled 'B') were used in Experiment 2.

	F0 (Hz)	Noise-to-harmonics ratio (dB)	Source spectral slope (dB/octave)
Low noise/ normal source (A)	222 114	−37.5 −40.8	−9.6 −7.6
Low noise/ sinusoidal source (A and B)	238 103	−40.8 −42.4	−20.6 −16.2
High noise/ normal source (A and B)	175 102	−23.0 −24.8	−9.9 −8.2
High noise/ sinusoidal source (B)	192 196	−23.8 −20.5	−20.7 −19.4

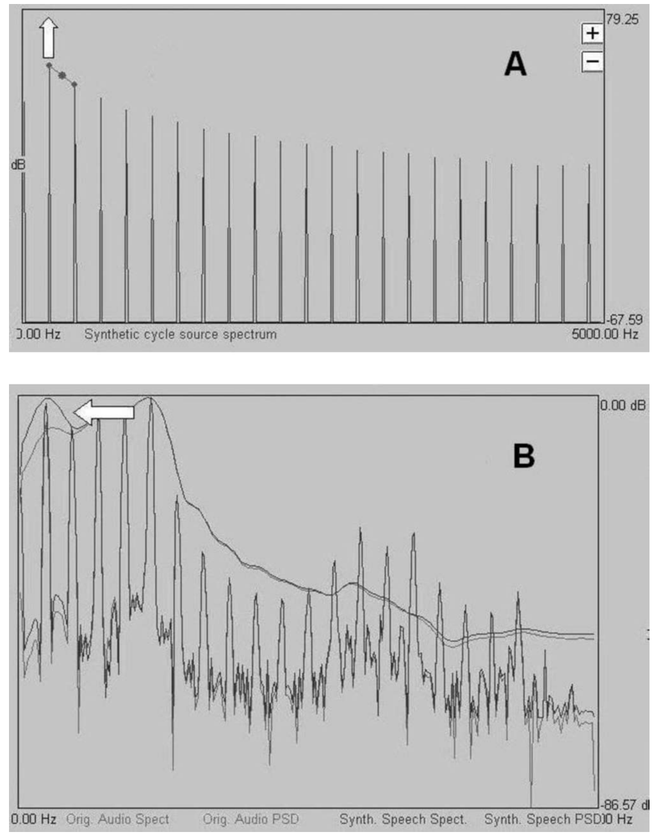

FIG. 1. Manipulating the acoustic voice source in the spectral domain. (A) The first two harmonics in the synthesizer display have been selected, as indicated by the line segment, and the amplitude of the first harmonic has been increased, as shown by the arrow. The resulting increase in harmonic amplitude is indicated by a second arrow in panel (B).

Pilot studies indicated that it was not possible to include all 16 blocks of stimuli (8 voices by two directions of H1 manipulation) in a single experiment without unduly taxing listeners' attention. For this reason, blocks were apportioned to two experiments, each consisting of 12 blocks of trials. Listeners in Experiment 1 heard both the steeply and slowly falling harmonic sources combined with low amounts of spectral noise, plus the slowly-falling source combined with high levels of spectral noise. Listeners in Experiment 2 heard the steeply-falling harmonic source combined with high and low levels of noise, plus the slowly-falling harmonic source combined with high levels of noise, as shown in Table II.

B. Listeners

Twenty listeners (17 female, 3 male) participated in Experiment 1, and 21 listeners (17 female, 4 male) participated in Experiment 2. They ranged in age from 19–53, with a mean age of 24.8 years (sd=7.85 years). Eleven of the listeners in Experiment 1 were native speakers of English, and nine were native speakers of Mandarin or bilingual speakers of Mandarin and another Chinese language (Cantonese or Shanghaiese) with English as a third language. Nine participants in Experiment 2 were native speakers of English, and

TABLE III. Average just-noticeable differences in H1–H2 for different voice types and listener groups. Standard deviations are given parenthetically. A: Experiment 1; B: Experiment 2.

	Normal source/ low noise (A)	Sinusoidal source/ low noise (A/B)	Normal source/ high noise (A/B)	Sinusoidal source/ high noise (B)
English-speaking listeners	3.68 (1.87)	2.71 (1.51)/2.77 (1.35)	4.26 (1.77)/ 4.59 (1.90)	3.73 (1.82)
Mandarin-speaking listeners	2.82 (2.14)	1.80 (1.53)/2.21 (1.58)	3.24 (2.14)/3.54 (1.78)	2.63 (1.37)

12 were native speakers of Mandarin or bilingual speakers of Mandarin and Cantonese with English as a third language. All listeners reported normal hearing.

C. Procedure

Listeners were tested individually in a sound-treated room. Stimuli were presented at a comfortable listening level over Etymotic ER-1 headphones (Etymotic Research, Inc., Elk Grove Village, IL). Trials were blocked by talker and by the direction of change in H1–H2 relative to the standard stimulus. Blocks were presented to each listener in a new random order. For each block, listeners heard a series of pairs of voices and were asked to judge whether the voices in each pair were the same or different (an AX procedure). One voice in each pair was always the standard stimulus, and the other was a test stimulus that differed from the standard only in H1–H2. Voices within a pair were separated by 100 ms. Listeners could play the pair once only in each order (AB and BA) before making their decision. For the first trial in a block, H1–H2 for the test stimulus differed from the standard by 2 dB. This amount was modified for each trial based on the listener's responses to the two previous trials. If the listener correctly distinguished the stimuli in both of the previous two trials, then the difference was decreased by 0.5 dB; but if the listener incorrectly responded "same" to either of the two previous trials, then the difference between stimuli was increased by 0.5 dB. The test proceeded until 12 reversals were obtained, and the just-noticeable difference for that listener and block was calculated by averaging the difference between the standard and test stimuli in H1–H2 at the last eight reversals. This procedure identified the H1–H2 value for which a listener can correctly distinguish the target and test stimuli 70.7% of the time (see Levitt, 1971, for theoretical justification and mathematical derivation).

Prior to the beginning of the test, listeners heard training stimuli (one male and one female voice) to familiarize them with the contrast being tested (which was not linguistically contrastive for either language group). Three tokens were contrasted for each voice: the standard stimulus and two additional stimuli whose H1–H2 values differed from the standard by ± 6.5 dB. Listeners first heard the two extreme stimuli (which differed in H1–H2 by 13 dB) several times, until they were confident they could distinguish them. They then heard each extreme stimulus paired with the standard. Training lasted no more than 5 min, after which the experimental trials began immediately. Total testing time for the twelve blocks of stimuli averaged about one hour.

III. RESULTS

One-way analyses of variance (ANOVAs) showed that the direction of H1–H2 change (increased versus decreased relative to the standard) had no effect on just-noticeable differences in either experiment [Experiment 1: $F(1,238) = 1.97$, $p = 0.16$; Experiment 2: $F(1,250) = 2.35$, $p = 0.13$]. As a result, data were combined for these two conditions.

Mean just-noticeable differences for the four experimental conditions and two listener groups (English speakers/ Mandarin speakers) are given in Table III. Listeners were quite sensitive overall to changes in H1–H2. Just-noticeable differences across voices, listeners, and experiments averaged 2.72 dB for Mandarin speakers, and 3.61 dB for English speakers. Differences between language groups were statistically significant in both experiments [two-way repeated measures ANOVAs; between-subjects factor=native language; within-subjects factor=type of target voice; Experiment 1: $F(1,62) = 43.15$, $p < 0.01$; Experiment 2: $F(1,82) = 10.73$, $p < 0.01$], with the Mandarin-speaking listeners showing consistently greater sensitivity to changes in H1–H2 than English-speaking listeners. Just-noticeable differences were nearly identical for the two conditions that were included in both experiments [two sample t-test; $t(326) = -0.584$, $p = 0.56$; see Table III]. No significant effect of block presentation order was observed in either experiment [Experiment 1: $F(1,238) = 0.168$, $p = 0.68$; Experiment 2: $F(1,250) = 0.08$, $p = 0.77$], indicating that short-term learning did not affect measured sensitivity.

The overall slope of the voice source spectrum also significantly impacted listeners' sensitivity in both experiments [Experiment 1: $F(2,124) = 18.16$, $p < 0.01$; Experiment 2: $F(2,164) = 31.22$, $p < 0.01$]: H1–H2 differences were easier to detect when the voicing source was quasi-sinusoidal (so that H1 dominated the spectrum) than when it was normal (flatter spectral slope; Scheffé post-hoc comparisons, $p < 0.01$). In contrast, post-hoc comparisons revealed no significant effect of NSR on listener sensitivity in either experiment [Experiment 1: $F(1,61) = 1.44$, $p = 0.24$; Experiment 2: $F(1,81) = 0.55$, $p = 0.46$]. No interaction between voice type and listener group was observed in either experiment [Experiment 1: $F(2,124) = 0.03$, $p = 0.97$; Experiment 2: $F(2,164) = 1.10$, $p = 0.34$].

IV. DISCUSSION

The issue of how to determine the perceptual validity of an acoustic measure has received little attention in the literature on voice perception. Listeners are highly flexible in the perceptual strategies they apply when listening to voices (e.g., Kreiman *et al.*, 1992; Van Lancker *et al.*, 1985), and no matter how prominent a parameter, they always have the option of adopting non-optimal or idiosyncratic strategies, even if these lead to perceptual mistakes. However, because a cue cannot be perceptually valid if listeners' sensitivity is poor relative to the cue's variability across voices, we can assess the *potential* perceptual usefulness of a given acoustic parameter. That is, if listeners require relatively large differences in some parameter before they can distinguish two voices, that parameter is probably not perceptually useful. We propose the ratio of sensitivity relative to variability across voices as a measure of this relationship, and as an index of the potential viability of a measure as a perceptual attribute. In the present case, just-noticeable differences average 3.18 dB across voices, listeners, and experiments. The range of H1–H2 values observed in our previous study of 70 pathological and normal voices equaled about 24 dB (Kreiman *et al.*, 2007b), for a ratio of 0.13. Differences between breathy and modal phonation or modal phonation and creak reported for natural languages with phonemic contrasts in phonation type (Table I) also exceed just-noticeable differences in H1–H2, in this case by a ratio of about 2:1. Because the amount of change listeners can hear is small relative to the variability of the parameter across speakers and the amount of difference that is linguistically meaningful, we conclude that H1–H2 is likely perceptually meaningful to listeners, and is thus a valid acoustic measure of voice quality.

Consistent differences have been previously reported between speakers of different languages in the manner in which they utilize a fixed set of acoustic cues during speech perception (e.g., Terbeek, 1977; Harnsberger, 2001; Lively *et al.*, 1993), but almost no evidence has appeared regarding such differences in perception of voice quality (see Esposito, 2010, for a recent exception). Because tones in Mandarin are correlated with changes in phonation type (Davison, 1991; Belotel-Grenié and Grenié, 2004; Liu and Samuel, 2004), it is not possible to determine the precise origin of the Mandarin listeners' increased sensitivity from these data. Two explanations for this finding suggest themselves. First, the low-dipping third tone in Mandarin is often produced with creaky voice (Davison, 1991; Liu and Samuel, 2004; Belotel-Grenié and Grenié, 2004), and it is possible that Mandarin speakers have learned to hear H1–H2 contrasts in part through attention to this allophonic cue to tone identity. Alternatively, perception of tone in languages like Mandarin requires attention to details of F0 contours, including direction and magnitude of change over time. Focusing attention on such details could provide the ancillary benefit of increased sensitivity to H1 amplitude as well, because H1 is identical to F0. Consistent with this explanation, differences in H1–H2 were easier for both English and Mandarin listeners to hear when the spectral slope fell away steeply, so that H1 dominated the spectrum in a kind of "auditory spotlight" (cf. Bregman, 1990). Recent neuropsychological and neurophysiological evidence showing that speakers of tone languages process F0 information differently than do speakers of non-tone languages is also consistent with this explanation. At the cortical level, these differences are reflected by left-hemispheric involvement in the perception of tone, versus right-hemisphere involvement in perception of F0 in speakers of non-tone languages (e.g., Wang *et al.*, 2004; see also Van Lancker and Fromkin, 1973). Additional evidence indicates that speaking a tone language also causes neuroplastic effects in auditory processing as early as the level of the brainstem: native speakers of Mandarin showed more robust and more accurate pitch tracking (measured by the frequency following response generated by the inferior colliculus) than did native speakers of English, suggesting that long-term learning sharpens the tuning characteristics of brainstem neurons to increase early sensitivity to linguistically relevant information (Krishnan *et al.*, 2005; Krishnan and Gandour, 2009). (Recall that no evidence of short-term learning appeared in the present experiments.) This explanation is also consistent with evidence that changes in F0 interfere with listeners' ability to judge spectral slope, although changes in vocal tract resonant frequencies produce no such effect (Li and Pastore, 1995). This finding suggests that spectral slope is integral with F0, and that listeners have difficulty separating different source attributes (for example, H1–H2 and F0) during perception. Studies comparing sensitivity in speakers of non-tonal languages (e.g., Gujarati) with phonation contrasts (Fischer-Jørgensen, 1967) with that of speakers of a tone language without a phonation contrast (e.g., Thai) can distinguish between these two explanations, and are underway (Kreiman *et al.*, 2009).

In conclusion, changes in the amplitude of H1 relative to the rest of the spectrum are easy for listeners to hear, regardless of language background, but they are particularly salient to speakers of Mandarin. Although it is not obvious how to incorporate long-term learning effects and principled differences among listeners into models of voice perception, these data provide additional evidence that interlistener variability in quality perception is not necessarily a result of random measurement errors, as has been claimed (Shrivastav *et al.*, 2005; see also Kreiman *et al.*, 2007b for additional counter-evidence to this claim). In any case, examining listeners' perceptual sensitivity to the acoustic concomitants of perceived differences among voices is an essential first step toward a comprehensive psychoacoustic model of voice quality.

ACKNOWLEDGMENTS

This research was supported by Grant No. DC01797 from the National Institute on Deafness and other Communication Disorders and by NSF grant BCS-0720304. Synthesizer software is available for download as open source freeware at http://www.surgery.medsch.ucla.edu/glottalaffairs/.

[1]When measured directly from the voicing source, this measure is written H1–H2. H1–H2 is also commonly estimated from the speech signal as recorded at the mouth, which requires correction for the influence of vocal tract resonances on harmonic amplitudes (Hanson, 1997; Iseli and Alwan, 2004). Measures made in this manner are written $H1^*-H2^*$.

Andruski, J., and Ratliff, M. (**2000**). "Phonation types in production of phonological tone: The case of Green Mong," J. Int. Phonetic Assoc. **30**, 37–61.

Belotel-Grenié, A., and Grenié, M. (**2004**). "The creaky voice phonation and the organization of Chinese discourse," in International Symposium on

Tonal Aspects of Languages: With Emphasis on Tone Languages, Beijing, China, pp. 5–8.
Bregman, A. S., (**1990**). *Auditory Scene Analysis: The Perceptual Organization of Sound* (MIT, Cambridge, MA).
Davison, D. S. (**1991**). "An acoustic study of so-called creaky voice in Tianjin Mandarin," UCLA Working Papers in Phonetics, Vol. **78**, pp. 50–57.
de Krom, G. (**1995**). "Some spectral correlates of pathological breathy and rough voice quality for different types of vowel fragments," J. Speech Hear. Res. **38**, 794–811.
DiCanio, C. T. (**2009**). "The phonetics of register in Takhian Thong Chong," J. Int. Phonetic Assoc. **39**, 162–188.
Esposito, C. M. (**2010**). "The effects of linguistic experience on the perception of phonation," J. Phonetics **38**, 306–316.
Fischer-Jørgensen, E. (**1967**). "Phonetic analysis of breathy (murmured) vowels in Gujarati," Indian Linguist. **28**, 71–139.
Gerfen, C., and Baker, K. (**2005**). "The production and perception of laryngealized vowels in Coatzospan Mixtec," J. Phonetics **33**, 311–334.
Hammarberg, B., Fritzell, B., Gauffin, J., Sundberg, J., and Wedin, L. (**1980**). "Perceptual and acoustic correlates of abnormal voice qualities," Acta Oto-Laryngol. **90**, 441–451.
Hanson, H. M. (**1997**). "Glottal characteristics of female speakers: Acoustic correlates," J. Acoust. Soc. Am. **101**, 466–481.
Harnsberger, J. D. (**2001**). "On the relationship between identification and discrimination of non-native nasal consonants," J. Acoust. Soc. Am. **110**, 489–503.
Hartl, D. M., Hans, S., Vaissière, J., Riquet, M., and Brasnu, D. F. (**2001**). "Objective voice quality analysis before and after onset of unilateral vocal fold paralysis," J. Voice **15**, 351–361.
Hillenbrand, J., Cleveland, R. A., and Erickson, R. L. (**1994**). "Acoustic correlates of breathy vocal quality," J. Speech Hear. Res. **37**, 769–778.
Huffman, M. K. (**1987**). "Measures of phonation type in Hmong," J. Acoust. Soc. Am. **81**, 495–504.
Iseli, M., and Alwan, A. (**2004**). "An improved correction formula for the estimation of harmonic magnitudes and its application to open quotient estimation," in Proceedings of ICASSP, pp. 669–672.
Klatt, D. H., and Klatt, L. C. (**1990**). "Analysis, synthesis, and perception of voice quality variations among female and male talkers," J. Acoust. Soc. Am. **87**, 820–857.
Klich, R. (**1982**). "Relationships of vowel characteristics to listener ratings of breathiness," J. Speech Hear. Res. **25**, 574–580.
Kreiman, J., Antoñanzas-Barroso, N., and Gerratt, B. R. (**2010**). "Integrated software for analysis and synthesis of voice quality," Behavior Research Methods.
Kreiman, J., and Gerratt, B. R. (**2005**). "Perception of aperiodicity in pathological voice," J. Acoust. Soc. Am. **117**, 2201–2211.
Kreiman, J., Gerratt, B. R., and Antoñanzas-Barroso, N. (**2007a**). "Measures of the glottal source spectrum," J. Speech Hear. Res. **50**, 595–610.
Kreiman, J., Gerratt, B. R., and Ito, M. (**2007b**). "When and why listeners disagree in voice quality assessment tasks," J. Acoust. Soc. Am. **122**, 2354–2364.
Kreiman, J., Gerratt, B. R., and Khan, S. D. (**2009**). "Effects of native language on perception of H1–H2," in 158th Meeting of the Acoustical Society of America, San Antonio, TX, October.
Kreiman, J., Gerratt, B. R., Precoda, K., and Berke, G. S. (**1992**). "Individual differences in voice quality perception," J. Speech Hear. Res. **35**, 512–520.
Krishnan, A., and Gandour, J. T. (**2009**). "The role of the auditory brainstem in processing linguistically-relevant pitch patterns," Brain Lang **110**, 135–148.
Krishnan, A., Xu, Y., Gandour, J., and Cariani, P. (**2005**). "Encoding of pitch in the human brainstem is sensitive to language experience," Brain Res. Cognit. Brain Res. **25**, 161–168.
Lee, G. S., Wang, C. P., and Fu, S. (**2009**). "Evaluation of hypernasality in vowels using voice low tone to high tone ratio," Cleft Palate Craniofac J. **46**, 47–52.
Levitt, H. (**1971**). "Transformed up-down methods in psychoacoustics," J. Acoust. Soc. Am. **49**, 467–477.
Li, X., and Pastore, R. E. (**1995**). "Perceptual constancy of a global spectral property: Spectral slope discrimination," J. Acoust. Soc. Am. **98**, 1956–1968.
Liu, S., and Samuel, A. G. (**2004**). "Perception of Mandarin lexical tones when F0 information is neutralized," Lang Speech **47**, 109–138.
Lively, S. E., Logan, J. S., and Pisoni, D. B. (**1993**). "Training Japanese listeners to identify English /r/ and /l/. II: The role of phonetic environment and talker variability in learning new perceptual categories," J. Acoust. Soc. Am. **94**, 1242–1255.
Shrivastav, R., and Sapienza, C. (**2006**). "Some difference limens for the perception of breathiness," J. Acoust. Soc. Am. **120**, 416–423.
Shrivastav, R., Sapienza, C., and Nandur, V. (**2005**). "Application of psychometric theory to the measurement of voice quality using rating scales," J. Speech Lang. Hear. Res. **48**, 323–335.
Sundberg, J., and Gauffin, J. (**1979**). "Waveform and spectrum of the glottal voice source," in *Frontiers of Speech Communication Research: Festschrift for Gunnar Fant*, edited by B. Lindblom and S. Ohman (Academic, London), pp. 301–320.
Terbeek, D. (**1977**). "A cross-language multidimensional scaling study of vowel perception," UCLA Working Papers in Phonetics, Vol. **37**.
Van Lancker, D., and Fromkin, V. (**1973**). "Hemispheric specialization for pitch and 'tone': Evidence from Thai," J. Phonetics **1**, 101–109.
Van Lancker, D., Kreiman, J., and Wickens, T. D. (**1985**). "Familiar voice recognition: Patterns and parameters. Part II: Recognition of rate-altered voices," J. Phonetics **13**, 39–52.
Wang, Y., Behne, D. M., Jongman, A., and Sereno, J. A. (**2004**). "The role of linguistic experience in the hemispheric processing of lexical tone," Appl. Psycholinguist. **25**, 449–466.
Wayland, R., and Jongman, A. (**2003**). "Acoustic correlates of breathy and clear vowels: The case of Khmer," J. Phonetics **31**, 181–201.