



Effects of native language on perception of voice quality

Jody Kreiman*, Bruce R. Gerratt, Sameer ud Dowla Khan

UCLA School of Medicine, Division of Head and Neck Surgery, 31-24 Rehab Center, Los Angeles, CA 90095-1794, USA

ARTICLE INFO

Article history:

Received 3 March 2010
Received in revised form
6 August 2010
Accepted 9 August 2010

ABSTRACT

Little is known about how listeners judge phonemic versus allophonic (or freely varying) versus post-lexical variations in voice quality, or about which acoustic attributes serve as perceptual cues in specific contexts. To address this issue, native speakers of Gujarati, Thai, and English discriminated among pairs of voices that differed only in the relative amplitudes of the first versus second harmonics (H1–H2). Results indicate that speakers of Gujarati (which contrasts H1–H2 phonemically) were more sensitive to changes than are speakers of Thai or English. Further, sensitivity was not affected by the overall source spectral slope for Gujarati speakers, unlike Thai and English speakers, who were most sensitive when the spectrum fell away steeply. In combination with previous findings from Mandarin speakers, these results suggest a continuum of sensitivity to H1–H2. In Gujarati, the independence of sensitivity and spectral context is consistent with use of H1–H2 as a cue to the language's phonemic phonation contrast. Speakers of Mandarin, in which creaky phonation occurs in conjunction with the low-dipping Tone 3, apparently also learn to hear these contrasts, but sensitivity is conditioned by spectral context. Finally, for Thai and English speakers, who vary phonation only post-lexically, sensitivity is both lower and contextually determined, reflecting the smaller role of H1–H2 in these languages.

© 2010 Elsevier Ltd. All rights reserved.

1. Introduction

The relative amplitude of the first versus the second harmonic of the voice source (H1–H2) has been implicated as an important cue to voice quality contrasts, particularly along a continuum from creakiness (smaller H1–H2 values) to breathiness (larger H1–H2 values). For example, in White Hmong H1–H2 values are consistently (and significantly) higher in phonemically breathy phonation than in modal phonation (Huffman, 1987); H1–H2 values distinguish tense and non-tense phonation in Chong (Di Canio, 2009); and in Gujarati, average H1–H2 for phonemically breathy vowels exceeds that for modal vowels by 4.4 dB in careful speech (Fischer-Jørgensen, 1967) and by 2.6 dB in connected speech (Keating, Esposito, Garellek, Khan, & Kuang, 2010; Khan, 2010; although differences could also be due to use of more accurate pitch-synchronous analyses in the recent work). (See also Bickley, 1982, on Gujarati; Blankenship, 2002, on Chong and Mpi; Esposito, 2004, on Zapotec, and Esposito, 2010, on Gujarati, Spanish, and English; Ladefoged & Antoñanzas-Barroso, 1985, on !Xóó; Watkins, 2002, on Wa; or Wayland & Jongman, 2003, on Khmer, for more examples.)

Less is known about how listeners perceive this attribute. A number of studies have demonstrated correlations between

H1–H2 and perceived breathy voice quality (e.g., Fischer-Jørgensen, 1967; Klatt & Klatt, 1990; Wayland & Jongman, 2003). Esposito (2010) showed that speakers of Gujarati were more accurate and more consistent in distinguishing breathy from modal phonation than were speakers of English or Spanish (which lack phonemic voice quality contrasts), and that their judgments were strongly predicted by H1–H2. She speculates that the linguistic status of the phonation contrast is at the heart of this difference, but her experimental design did not permit conclusions about causation. Similarly, Kreiman and Gerratt (in press) found that Mandarin listeners were significantly more sensitive to changes in H1–H2 than were English listeners (just-noticeable difference = 2.72 dB, vs. 3.61 dB). They proposed two possible explanations for this finding. First, because pitch and loudness interact perceptually (e.g., Melara & Marks, 1990), increased attention to the frequency of H1 might have the side effect of increasing awareness of its amplitude. In this case, the observed difference between listener groups could be the result of increased attention to the fundamental frequency (f_0) in speakers of tone languages (because $f_0 = H1$; Krishnan & Gandour, 2009; Krishnan, Xu, Gandour, & Cariani, 2005) compared to speakers of languages like English where tone (and thus H1) is not manipulated on a lexical level. Alternatively, the low-dipping Tone 3 in Mandarin is often produced with creaky voice (Davison, 1991), and Mandarin listeners recognize Tone 3 faster when it is produced with creak than when produced without creak (Belotel-Grenié & Grenié, 1997, 2004). This suggests that creak is used in tone perception in Mandarin, and thus that these listeners could have developed

* Corresponding author. Tel.: +1 310 825 0736.

E-mail addresses: jkreiman@ucla.edu (J. Kreiman), bgerratt@ucla.edu (B.R. Gerratt), sameeruddowlakhan@gmail.com (S.D. Khan).

increased sensitivity to changes in harmonic amplitudes due to their perceptual experience with the phonation differences associated with contrastive tone in that language.

To clarify this matter, in this experiment we compared just-noticeable differences (JNDs) in H1–H2 for speakers of Gujarati, Thai, and English. Gujarati phonemically contrasts breathy and modal phonation (but not tone) on both consonants and vowels (Fischer-Jørgensen, 1967). Although one report (Nair, 1979) suggested that some varieties of Gujarati are losing this distinction, data from recent fieldwork indicate that production differences between breathy and modal phonation remain robust (Khan, 2010). Thai contrasts tones, but not phonation types, and there is little evidence of any systematic variation in phonation type. Some prosodic and sociolinguistic variation in phonation does occur (e.g., Abramson, 1962; Potsiuk, Harper, & Gandour, 1999; Thepboriruk, 2009): laryngeally coarticulated vowels occur in free variation with unlaryngealized vowels in some contexts (Abramson, 1979; Hudak, 1990, 2008), and phonemically voiced stops may have leading creakiness in some contexts (see Esling, Frasier, & Harris, 2005, for review). Finally, English contrasts neither tone nor phonation, and is characterized only by consonantly conditioned, prosodic, and sociolinguistic variations in voice quality similar to those in Thai (e.g., Cutler, Dahan, & van Donselaar, 1997; Dilley, Shattuck-Hufnagel, & Ostendorf, 1996; Epstein, 2002). These observations lead to the following predictions. If increased sensitivity to H1–H2 differences is the result of increased attention to f_0/H_1 , then JNDs for both Gujarati and Thai listeners should be smaller (i.e., listeners should be able to hear smaller changes) than those for English listeners and closer to those for the Mandarin listeners in our previous study. However, if increased sensitivity to H1–H2 differences is related to the role of H1–H2 as a cue to a lexically relevant phonation contrast, then JNDs for Gujarati listeners should be smaller than those for English and Thai listeners.

2. Materials and methods

2.1. Stimuli

Stimuli were a subset of those used in our previous study (Kreiman & Gerratt, in press), and were created using the UCLA voice synthesizer (Kreiman, Gerratt, & Antoñanzas-Barroso, in press). Synthetic stimuli were modeled on eight natural voice samples produced by speakers of American English (four males and four females). These natural samples were selected to represent all combinations of two noise-to-harmonics ratios (NHR; low and high) and two rates of source spectral slope roll-off (relatively quick – a quasi-sinusoidal source – and relatively

slow; Table 1, Fig. 1). H1–H2 values for these talkers roughly spanned the range of values reported for normal speakers of English (about 2 dB to about 19 dB; Kreiman, Gerratt, & Antoñanzas-Barroso, 2007). Systematic differences across languages in average H1–H2 values have not been reported, to our knowledge. A one-second sample of the vowel /a/ produced by each speaker was copied using the synthesizer to match the original voice. Spectral slopes and NHR values for the synthetic voice samples were then manipulated slightly to improve the contrast between talker types with respect to these independent variables.

Each of the eight synthetic voice samples (“standards”) served as the basis for creating a series of stimuli in which H1 varied in 15 steps of 0.5 dB relative to the standard value (Table 1). H1 values were manipulated as follows. Using the synthesizer, the source spectrum was obtained by performing a pitch-synchronous Fourier transform on the time-domain synthetic source waveform (Antoñanzas-Barroso, Kreiman, & Gerratt, 2008). The first two harmonics were selected in this spectrum, as shown in Fig. 2A. The amplitude of the first harmonic was then altered by increasing or decreasing the slope of the line segment connecting these harmonics (as shown by the arrow) 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 (i.e., f_0 , formant frequencies, pitch and amplitude contours, NHR) held constant. Fig. 2B shows the resulting voice spectrum, with the change in the amplitude of the first harmonic indicated with an arrow.

2.2. Listeners

Three groups of listeners participated in this study. The first included native ($n=13$) and heritage ($n=3$) speakers of Gujarati; the second included native ($n=11$) and heritage ($n=7$) speakers of Thai; and the third included native English speakers ($n=15$). Heritage speakers all reported native or near-native comprehension of the target language, and all reported using the target language with relatives on a regular basis. All listeners reported normal hearing. They were paid \$20 for their time.

2.3. Procedure

All testing took place in the Los Angeles area, in a quiet room or a double-walled sound suite. Trials were blocked by talker, with block order randomized for each listener. Listeners heard stimuli in pairs over Etymotic ER-1 insert earphones at a comfortable listening level. Voices within a pair were separated by 250 ms; listeners could hear each order (i.e., A–B and B–A) once prior to responding. For each pair, listeners judged whether the stimuli

Table 1

Stimulus characteristics. The first value in each cell represents a female voice (F); the second value represents a male voice (M).

Stimulus type	NHR	Source spectral slope	Standard H1–H2 value
Flat slope/low noise	–37.5 dB (F)	–9.6 dB/octave (F)	5.41 dB (F)
	–40.8 dB (M)	–7.6 dB/octave (M)	3.80 dB (M)
Flat slope/high noise	–23.0 dB (F)	–9.9 dB/octave (F)	6.20 dB (F)
	–24.8 dB (M)	–8.21 dB/octave (M)	6.16 dB (M)
Steep slope/low noise	–40.8 dB (F)	–20.6 dB/octave (F)	16.25 dB (F)
	–43.4 dB (M)	–16.2 dB/octave (M)	8.08 dB (M)
Steep slope/high noise	–23.8 dB (F)	–20.7 dB/octave (F)	14.71 dB (F)
	–20.5 dB (M)	–19.4 dB/octave (M)	13.45 dB (M)

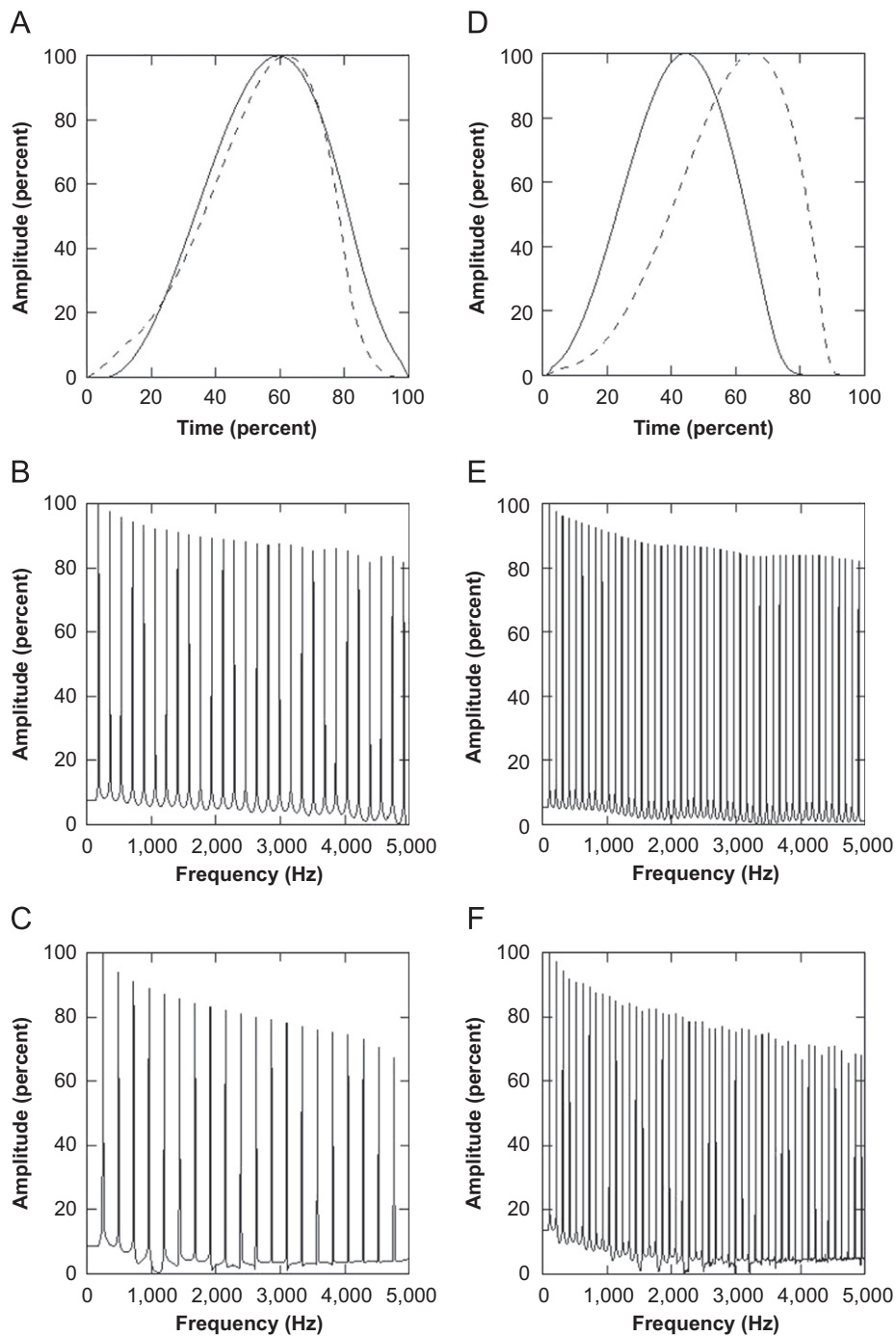


Fig. 1. Representative source pulses and corresponding spectra for the synthetic stimuli. (A) Female source pulses corresponding to the relatively flat spectrum shown in (B) (plotted with a dashed line) and to the relatively steeply falling spectrum shown in panel (C) (plotted with a solid line). (D) Male source pulses corresponding to the relatively flat spectrum shown in (E) (plotted with a dashed line) and to the relatively steeply falling spectrum shown in panel (F) (plotted with a solid line). Source pulses have been time and amplitude normalized, and spectra have been normalized to peak amplitudes.

were the same or different (an AX task) in a “one-up, two-down” adaptive paradigm (Levitt, 1971; Shrivastav & Sapienza, 2006). In the first trial of each block, H1–H2 for the two stimuli differed by 2 dB. This difference was adjusted by 0.5 dB in each successive trial, based on the listener’s responses to the two previous trials. The difference between stimuli was increased if one or both of the previous two trials were answered incorrectly, and the difference was decreased when both were answered correctly. Testing for each block proceeded until 12 reversals were obtained, after which the JND for that listener and block was calculated by

averaging the difference between stimuli in H1–H2 at the last eight reversals. (The first four reversals are considered practice and are discarded.) 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. Three tokens were contrasted for each voice: the standard stimulus and two additional stimuli whose

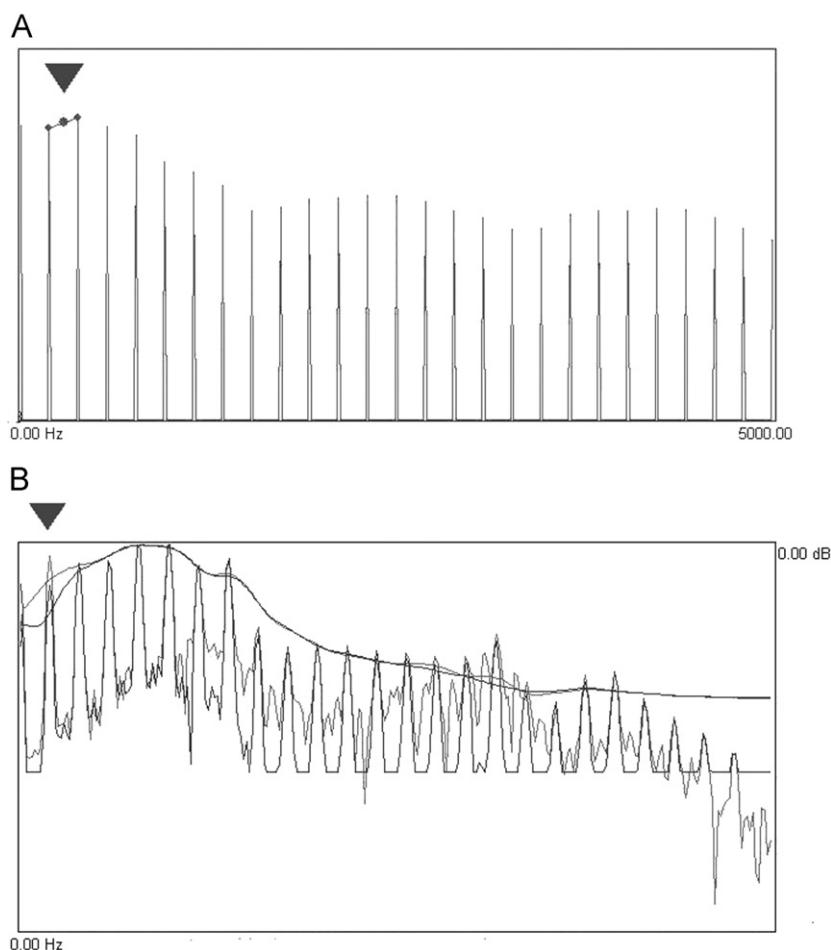


Fig. 2. Adjusting harmonic amplitudes. (A) The user selects the first two harmonics by clicking to create a line segment, indicated by an arrow. The height of the first harmonic is altered by entering the desired slope of this line segment into a box (not shown). The user then clicks on the line segment to adjust the slope to the desired value. (B) The effect of the change in H1 amplitude is indicated with an arrow. In this case, H1 amplitude was decreased by 8.2 dB, from an original value of 4.2 dB above H2 to a new value of 4 dB below H2 (representing a somewhat strained voice quality). Note that the remainder of the spectral envelopes remain unchanged. (Apparent mismatches in interharmonic noise amplitude is an artifact of display normalization to peak amplitude.)

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 five minutes, after which the experimental trials began immediately. Total testing time for the eight blocks of trials averaged 45 minutes to one hour.

3. Results

Preliminary one-way ANOVAs showed no significant effects on JNDs of presentation order or f_0 (male versus female speakers) for any listener group, and no effect of status as native versus heritage speaker of Thai or Gujarati (Table 2). As a result, data were merged across these categories for subsequent analyses.

Mean JNDs for the three groups are shown in Table 3 for the different spectral slope and NHR conditions. Across conditions, JNDs for Gujarati listeners averaged 2.60 dB ($sd=1.80$ dB), JNDs for Thai listeners averaged 3.35 dB ($sd=1.87$ dB), and JNDs for English listeners averaged 3.67 dB ($sd=1.83$ dB). A three-way ANOVA showed significant main effects of native language and spectral slope on JND, plus a significant interaction between native language and spectral slope (Fig. 3; Native language:

Table 2

Statistical results for effects on JNDs of presentation order, f_0 , and listener status as a native or heritage speaker. $p > 0.01$ in all cases.

Variable	Listener group	Statistic
Presentation order	Gujarati	$F(7, 120)=0.44$
	Thai	$F(7, 136)=0.37$
	English	$F(7, 112)=0.21$
f_0	Gujarati	$F(1, 126)=0.31$
	Thai	$F(1, 142)=1.38$
	English	$F(1, 118)=0.12$
Native/heritage speaker status	Gujarati	$F(1, 126)=6.19$
	Thai	$F(1, 142)=3.81$

$F(2, 380)=13.18$, $p < 0.01$; Spectral slope: $F(1, 380)=28.13$, $p < 0.01$; Interaction: $F(2, 380)=4.38$, $p < 0.01$). No significant effect of NHR was observed ($F(1, 380)=5.71$, $p > 0.01$), and NHR did not interact significantly with other variables. Post-hoc Tukey's HSD paired comparisons with criterion $p=0.01$ showed no differences between Thai and English listeners in any condition. When spectral slopes fell gradually, Gujarati listeners performed better than did either English or Thai speakers (Fig. 3), but no group differences occurred when the source spectrum fell

Table 3
Mean just-noticeable differences for Gujarati-, Thai-, and English-speaking listeners in the different listening conditions. Standard deviations are given in parentheses.

	Gujarati listeners	Thai listeners	English listeners
Spectral slope			
Slowly falling	2.73 dB (1.85 dB)	4.43 dB (2.03 dB)	3.87 dB (2.04 dB)
Steeply falling	2.47 dB (1.76 dB)	2.91 dB (1.20 dB)	2.83 dB (1.54 dB)
NHR			
High (+noisy)	2.97 dB (1.92 dB)	3.80 dB (1.82 dB)	3.50 dB (1.91 dB)
Low (–noisy)	2.23 dB (1.60 dB)	3.55 dB (1.84 dB)	3.20 dB (1.84 dB)

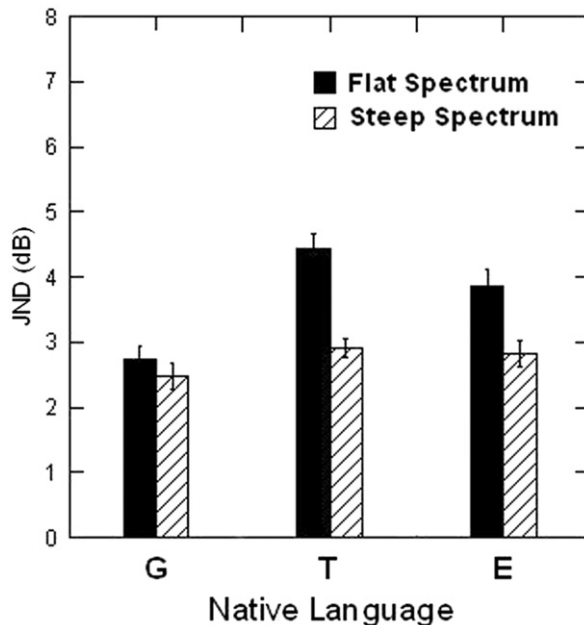


Fig. 3. The interaction between native language and spectral slope in determining JNDs for H1–H2. JNDs for relatively flat spectra are shown by the solid bars, and for relatively steeply falling spectra by striped bars. Smaller values (smaller JNDs) indicate greater sensitivity. G=Gujarati listeners; T=Thai listeners; E=English listeners. Differences between Thai and English speakers were not statistically significant in either condition.

steeply. Gujarati listeners performed equally well in both spectral slope conditions, but both English and Thai speakers performed better when the source spectrum fell steeply.

4. Discussion

Average JNDs in this experiment compare well to those from our previous study (Kreiman & Gerratt, *in press*), in which mean JNDs for H1–H2 for English-speaking listeners equaled 3.61 dB, compared to 3.67 dB in the present study. JNDs for Mandarin listeners in the previous experiment were similar to those for the Gujarati listeners in the present study (mean JNDs for Mandarin listeners=2.72 dB; mean JNDs for Gujarati listeners=2.60 dB). JNDs for Gujarati listeners are also consistent with production data showing an average difference of about 4.4 dB in H1–H2 between breathy and modal phonation in careful pronunciation (Fischer-Jørgensen, 1967) and 2.61 dB in connected speech (Keating et al., 2010; Khan, 2010). (Unfortunately, comparable production data for Mandarin are not available to our knowledge.) In contrast, Thai speakers did not differ significantly from English

speakers. These findings suggest that experience with more systematic, lexical use of phonation differences, whether phonemic in its own right (as in Gujarati) or used in conjunction with phonemic tone (as in Mandarin), underlies the increase in sensitivity to H1–H2, rather than the increased attention to H1/f0 that characterizes speakers of lexical tone languages (Krishnan & Gandour, 2009; Krishnan et al., 2005).

These data show an additional effect of native language on perceptual strategy as well as on perceptual sensitivity to changes in harmonic amplitudes. Despite the overall similarity in JNDs, Mandarin listeners differed from the Gujarati listeners in the influence of spectral slope on JNDs. Like Thai and English speakers, Mandarin speakers were significantly less sensitive to changes in H1–H2 when the harmonic voice spectrum was relatively flat (source spectral slope=8–10 dB/octave), and more sensitive when the spectrum fell away relatively steeply (source spectral slope=16–20 dB/octave). Increased spectral roll-off has the effect of making H1 stand out more in the spectrum, creating a kind of “auditory spotlight” that makes the task significantly easier. In contrast, no significant effect of the harmonic spectral slope was observed for speakers of Gujarati.

These findings suggest that there exists a continuum of sensitivity to H1–H2, based on the role this feature plays in the listener’s native language. In Gujarati, the independence of sensitivity to H1–H2 and the spectral context in which these harmonics exist is consistent with use of H1–H2 as a cue to a phonemic distinction: acoustic studies of Gujarati speakers indicate that H1–H2 is a good indicator of a vowel’s underlying voice quality (Bickley, 1982; Fischer-Jørgensen, 1967; Keating et al., 2010; Khan, 2010), and perception studies report that Gujarati-speaking listeners primarily attend to H1–H2 when categorizing synthesized (Bickley, 1982; Klatt & Klatt, 1990) and foreign (Esposito, 2010) vowels. Thus, speakers of Gujarati must be able to recover spectral information in similar ways regardless of the varying vocal characteristics of the speakers they encounter. Speakers of languages with allophonic or freely varying voice quality contrasts (e.g., Mandarin) apparently learn to hear these contrasts as well, but their sensitivity is conditioned by spectral context, consistent with primary reliance on f0 as a cue to tone (and thus word) identity. Finally, for languages like Thai and English, which vary voice quality only at a post-lexical level (e.g. prosodically, sociolinguistically), and not phonemically or allophonically, sensitivity is both lower and contextually determined, reflecting the minor role H1–H2 plays in the segmental aspects of these languages.

In contrast to spectral slope, the noise-to-harmonic ratio (NHR) had no significant effect on sensitivity to changes in H1–H2, for any listener group. Spectral noise levels do have a significant effect on perception of the slope of the high-frequency spectrum, and spectral slope affects perception of the NHR, presumably due to mutual masking of higher harmonics (which are typically relatively low in amplitude) and high-frequency aspiration noise (Kreiman & Gerratt, 2005; Shrivastav & Sapienza, 2006). However, perception of H1–H2 does not appear subject to such masking, presumably due to the higher amplitude of the lowest harmonics in the voice source. (The noise spectrum in most voices is relatively flat across frequencies, and thus is less likely to be the source of this effect.)

In conclusion, the present data replicate our previous finding that all listeners are sensitive to changes in H1–H2, and thus that H1–H2 is a perceptually valid acoustic measure of voice quality. Further, native language experience affects both sensitivity to source characteristics and perceptual strategy. These findings contribute to our growing understanding of the manner in which speech and voice perception interact, and of the complex manner in which humans judge vocal quality.

Acknowledgments

We thank Arthur S. Abramson, Jackson T. Gandour, Jirapat Jangjamras, Pittayawat Pittayaporn, and Ratreewayland for information about the phonetics of Thai tones, and Norma Antoñanzas-Barroso for measurement advice and programming support. Comments from two anonymous reviewers resulted in significant improvements to the manuscript. This research was supported by Grant DC01797 from the National Institute on Deafness and Other Communication Disorders, and by NSF Grant BCS-0720304. Software used to create the stimuli in this experiment is available for free download from www.surgery.medsch.ucla.edu/glottalaffairs.

References

- Abramson, A. S. (1962). The vowels and tones of Standard Thai: Acoustical measurements and experiments. *International Journal of American Linguistics*, 28(2, Part III).
- Abramson, A. S. (1979). The coarticulation of tones: An acoustic study of Thai. In T. L. Thongkum, P. Kullavanijaya, V. Panupong, & T. L. Tingsabadh (Eds.), *Studies in Tai and Mon-Khmer Phonetics and Phonology, in Honour of Eugenie J. A. Henderson* (pp. 1–9). Bangkok: Chulalongkorn U.P.
- Antoñanzas-Barroso, N., Kreiman, J., & Gerratt, B. R. (2008). Recent improvements to the University of California, Los Angeles' voice synthesizer. *Proceedings of Meetings on Acoustics*, 5, 060001.
- Belotel-Grenié, A., & Grenié, M. (1997). Types de phonation et tons en chinois standard. *Cahiers de Linguistique—Asie Orientale*, 26, 249–279.
- Belotel-Grenié, A., & 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* (pp. 5–8), Beijing: ISCA.
- Bickley, C. (1982). Acoustic analysis and perception of breathy vowels. *MIT/RLE Speech Communication Group Working Papers*, 1, 71–82.
- Blankenship, B. (2002). The timing of nonmodal phonation in vowels. *Journal of Phonetics*, 30, 163–191.
- Cutler, A., Dahan, D., & van Donselaar, W. (1997). Prosody in the comprehension of spoken language: A literature review. *Language and Speech*, 40, 141–201.
- Davison, D. S. (1991). An acoustic study of so-called creaky voice in Tianjin Mandarin. *UCLA Working Papers in Phonetics*, 78, 50–57.
- Di Canio, C. T. (2009). The phonetics of register in Takhian Thong Chong. *Journal of the International Phonetic Association*, 39, 162–188.
- Dilley, L., Shattuck-Hufnagel, S., & Ostendorf, M. (1996). Glottalization of word-initial vowels as a function of prosodic structure. *Journal of Phonetics*, 24, 423–444.
- Epstein, M. (2002). *Voice quality and prosody in English*. Unpublished UCLA doctoral dissertation.
- Esling, J. H., Fräsera, K. E., & Harris, J. G. (2005). Glottal stop, glottalized resonants, and pharyngeals: A reinterpretation with evidence from a laryngoscopic study of Nuuchahnulth (Nootka). *Journal of Phonetics*, 33, 383–410.
- Espósito, C. M. (2004). Santa Ana del Valle Zapotec phonation. *UCLA Working Papers in Phonetics*, 103, 71–105.
- Espósito, C. M. (2010). The effects of linguistic experience on the perception of phonation. *Journal of Phonetics*, 38, 306–316.
- Fischer-Jørgensen, E. (1967). Phonetic analysis of breathy (murmured) vowels in Gujarati. *Indian Linguistics*, 28, 71–139.
- Hudak, T. J. (1990). Thai. In B. Comrie (Ed.), *Languages of East and South-East Asia*. London: Routledge.
- Hudak, T. J. (2008). *William J. Gedney's comparative Tai source book*. Oceanic Linguistics Special Publication No. 34. University of Hawai'i Press.
- Huffman, M. K. (1987). Measures of phonation type in Hmong. *Journal of the Acoustical Society of America*, 81, 495–504.
- Keating, P., Espósito, C. M., Garellek, M., Khan, S. D., & Kuang, J. J. (2010). Phonation contrasts across languages. *Poster presented at the 12th conference on laboratory phonology*, Albuquerque, 10 July.
- Khan, S. D. (2010). Breathiness phonation in Gujarati: an acoustic and electroglottographic study. *Poster presented at the 159th meeting of the Acoustical Society of America*, Baltimore, 23 April.
- Klatt, D. H., & Klatt, L. C. (1990). Analysis, synthesis, and perception of voice quality variations among female and male talkers. *Journal of the Acoustical Society of America*, 87, 820–857.
- Kreiman, J., & Gerratt, B. R. (2005). Perception of aperiodicity in pathological voice. *Journal of the Acoustical Society of America*, 117, 2201–2211.
- Kreiman, J., & Gerratt, B. R. Perceptual sensitivity to first harmonic amplitude in the voice source. *Journal of the Acoustical Society of America*, in press.
- Kreiman, J., Gerratt, B. R., & Antoñanzas-Barroso, N. (2007). Measures of the glottal source spectrum. *Journal of Speech, Language, and Hearing Research*, 50, 595–610.
- Kreiman, J., Gerratt, B. R., & Antoñanzas-Barroso, N. Integrated software for analysis and synthesis of voice quality. *Behavior Research Methods*, in press.
- Krishnan, A., & Gandour, J. T. (2009). The role of the auditory brainstem in processing linguistically-relevant pitch patterns. *Brain and Language*, 110, 135–148.
- Krishnan, A., Xu, Y., Gandour, J., & Cariani, P. (2005). Encoding of pitch in the human brainstem is sensitive to language experience. *Cognitive Brain Research*, 25, 161–168.
- Ladefoged, P., & Antoñanzas-Barroso, N. (1985). Computer measures of breathy phonation. *UCLA Working Papers in Phonetics*, 61, 79–86.
- Levitt, H. (1971). Transformed up-down methods in psychoacoustics. *Journal of the Acoustical Society of America*, 49, 467–478.
- Melara, R. D., & Marks, L. E. (1990). Perceptual primacy of dimensions: Support for a model of dimensional interaction. *Journal of Experimental Psychology: Human Perception and Performance*, 16, 398–414.
- Nair, U. (1979). *Gujarati phonetic reader*. Mysore: Central Institute of Indian Languages.
- Potsiuk, S., Harper, M. P., & Gandour, J. (1999). Classification of Thai tone sequences in syllable-segmented speech using the analysis-by-synthesis method. *IEEE Transactions on Speech and Audio Processing*, 7, 95–102.
- Shrivastav, R., & Sapienza, C. (2006). Some difference limens for the perception of breathiness. *Journal of the Acoustical Society of America*, 120, 416–423.
- Thepboriruk, K. (2009). Bangkok Thai tones revisited. *Working Papers in Linguistics of the University of Hawai'i at Mānoa*, 40, 1–15.
- Watkins, J. (2002). *The phonetics of Wa: Experimental phonetics, phonology, orthography, and sociolinguistics*, Pacific Linguistics, Vol. 531. Canberra: Australian National University.
- Wayland, R., & Jongman, A. (2003). Acoustic correlates of breathy and clear vowels: The case of Khmer. *Journal of Phonetics*, 31, 181–201.