EFFECT OF VOICE QUALITY ON THE TENSE/LAX DISTINCTION FOR ENGLISH VOWELS

Keith R. Kluender, Andrew J. Lotto, and Lori L. Holt
Dept. of Psychology, University of Wisconsin-Madison, USA

ABSTRACT

For a variety of East and West African languages, voice quality covaries with tongue root advancement. This regularity may be due to the mutually enhancing auditory effects of breathy voice and a low first formant frequency. Evidence is adduced for this explanation in results from perceptual categorization experiments in which voice quality and formant values were independently manipulated.

I. INTRODUCTION

In a number of vowel systems, there exists an articulatory covariation between advanced tongue root as a distinctive feature and voice quality. For a variety of East and West African languages, an advanced tongue root vowel is produced with a breathy phonation (sometimes called 'lax' voice); whereas, a modal or even creaky voice quality is used with non-advanced tongue root [1,2,3]. In fact, for some vowel harmony systems such as Akan, breathy voice has been referred to as, "the main auditory correlate of root advancing" [3].

Covariations between presumably distinctive articulatory variables often suggests that the variables are mutually enhancing either in terms of articulatory ease or perceptual distinctiveness. The experiments reported here evaluate a potential explanation for this regularity based on hypothesized perceptual advantages from the interaction between acoustic effects of vocal-tract shape and voice quality.

The acoustic effects of breathy phonation on the resulting vowel are essentially twofold. Due to the longer open quotient of the glottis, which characterizes breathy phonation, the resultant waveform has relatively greater energy at low frequencies. The amplitude for the fundamental component (H1) increases and the falloff across higher frequencies is steeper [4]. Similarly, the relatively low first formant of [+advanced tongue root] high vowels contributes to greater energy at low frequencies. The hypothesis being considered is that the joint acoustic consequence of vocal-tract shape and voice quality interact in a perceptually enhancing manner by jointly contributing to a low frequency prominence.

Because it has been suggested that the advanced tongue root contrast is similar to the tense/lax distinction in English [5], perception of English vowels may provide a reasonable experimental test of this hypothesis. The contrast between tense and lax high vowels in English (e.g. /i/ vs. /ɪ/) is signalled, in part, by the lower frequency of the first formant (F1) in tense vowels. If these vowels are produced with a breathy voice, it is possible that the increased H1 and steeper dropoff will lead to a lower perceived F1 and hence to a more 'tense' vowel. The following experiments tested this possibility by obtaining categorizations from native English speakers for tense and lax high vowels varying in voice quality.

II. EXPERIMENT 1

Subjects.

Fifteen college-age adults, all of whom learned English as their first language, served as listeners. All reported normal hearing. Subjects received Introductory Psychology course results of Experiment 1 is that the increased spectral tilt of the series modeled after breathy productions degraded the higher frequencies to the point where subjects were forced to rely solely on the first formant for identification. This is especially a possibility for the back vowel series, since a reasonable sounding /u/ can be constructed from a single low formant. Experiment 2 was designed to parcel out the effects of exaggerated spectral tilt and the increasing of H1 amplitude.

Subjects.

Subjects were eleven college-age students, none of whom had participated in Experiment 1. All subjects reported normal hearing and learned English as their first language. Course credit was awarded for participation.

Stimuli.

The stimuli were identical to those used in Experiment 1 except that for the "breathy" series only the amplitude of H1 (OQ = 72) was increased. Spectral tilt remained the same for all series.

Procedure.

The procedure was identical to that used in Experiment 1.

Results.

Again, identification boundaries were calculated using probit analysis. The mean boundary values are displayed in Figure 2. As in Experiment 1, breathy phonation led to more vowels being identified as tense for series modelled after female productions. However, in the case of vowels modelled after male productions, breathy phonation, synthesized solely through manipulation of H1 amplitude, did not lead to more vowels being identified as tense.

These results suggest that the amplitude of the first harmonic is largely responsible for the identification boundary shifts in vowels modelled after female productions, whereas overall spectral tilt seems to determine the effect produced by male series.

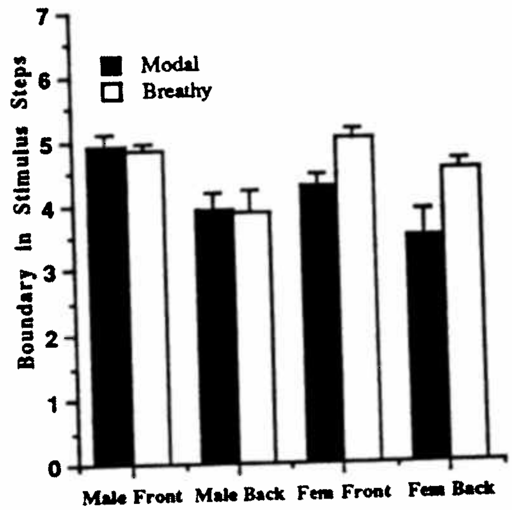

Figure 2. Probit identification boundary values (in stimulus steps) for the eight series in Experiment 2.

IV. DISCUSSION

The purpose of this investigation was to determine the effect of a potential perceptual interaction between voice quality and first formant frequency on the categorization of [+high] English vowels. The data indicate that the quality of voice used to produce a high vowel will affect whether the vowel is perceived as tense or lax.

These results suggest that the covariation in African languages may be consistent with general principles of Auditory Enhancement [6] and Adaptive Dispersion [7]. These theoretical frameworks predict that articulatory factors which tend to enhance the perceptual discriminability of a resultant speech sound will be favored in phonetic inventories. Breathy phonation enhances the low frequency prominence of tense high vowels and of [+advanced tongue root] high vowels. According to the present results, if a language included breathy production of lax vowels then the distinctiveness of the tense/lax contrast would be reduced. However, it credit for their participation.

Stimuli.

Four vowel series were synthesized with eight endpoints modeled after male and female productions of /i/ and /ɪ/ (front), and /u/ and /ʊ/ (back). Five intermediate vowels between tense (/i/, /u/) and lax (/ɪ/, /ʊ/) were synthesized by manipulating the nominal center frequencies of the first three formants. In particular, the frequency of F1 varied linearly from 210 Hz to 330 Hz for the male front series and from 330 Hz to 440 Hz for the male back series. For the series based on female productions, F1 varied in equal steps from 310 Hz to 430 Hz for the front vowels and from 370 to 480 Hz for the back vowels. Fundamental frequency was 135 Hz for series modeled after the male productions and 233 Hz for the female series and all stimuli were 120 msec in duration. Breathy versions of each seven-step series were created by increasing the amplitude of the first harmonic (open quotient; OQ = 72) and increasing spectral tilt (TL = 17) using the Klatt software synthesizer [4].

Stimuli were synthesized with 12-bit resolution at a 10-kHz sampling rate and stored on computer disk. Stimulus presentation was under control of a microcomputer. Following D/A conversion (Ariel DSP-16), stimuli were low-pass filtered (Frequency Devices 677, cutoff frequency 4.8 kHz) prior to being amplified (Stewart HDA4), and played over headphones (Beyer DT-100) at 75 dB SPL.

Procedure.

Subjects participated in a two-response forced choice identification task arranged in randomized blocks with ten presentations per stimulus. Each block contained the breathy and modal versions of each gender x place of articulation series, for a total of four blocks (male/female x front/back). Stimuli were presented at a rate of approximately one stimulus every 3 seconds. Subjects identified each vowel by pressing the appropriate button on a response box connected to a microcomputer. The buttons were labeled 'beat' and 'bit' or 'boot' and 'book'.

Results.

For each of the eight series, the identification boundary value was determined for each subject using probit analysis. The mean boundaries for each of the eight series are displayed in Figure 1 in terms of the stimulus steps (one = 'lax' endpoint; seven = 'tense' endpoint).

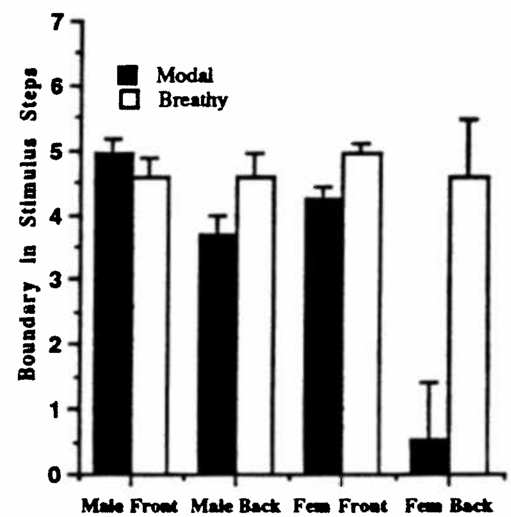

Figure 1. Probit identification boundary values (in stimulus steps) for the eight series of Experiment 1.

The results indicate that, in general, breathy phonation led to more vowels being identified as tense, particularly for male and female /u/-/ʊ/ series. Except for the male /i/-/ɪ/ series, the boundaries for the breathy series were significantly greater (more 'tense') than the boundaries for the modal series.

These results suggest that there exists a perceptual interaction between tenseness/laxness and phonation type.

III. EXPERIMENT 2

One possible explanation for the

appears that languages tend to contain the more distinctive pairing of breathy phonation with advanced tongue root and creaky voice (which would de-emphasize the low frequencies) with unadvanced tongue root.

There is also an interesting divergence in the data based on whether the stimuli were modeled on male or female productions. The effect of breathiness on tense/lax categorizations was much more robust for the female series. Even when spectral tilt was held constant in Experiment 2, subjects labeled the female vowels differentially when the amplitude of the first harmonic was changed.

It has been widely reported that females use breathy phonation more often than men in a variety of languages, including English [4,8]. If breathy productions lead to a more 'tense' high vowel, then it would be advantageous for English speaking females to use breathiness distinctively, i.e. use a breathy voice for tense high vowels and a modal voice for lax high vowels. Since the effect of voice quality is less robust in males, it is less probable that they would develop such a "strategy" for productions.

There is a dearth of relevant data concerning females' voice quality across the vowel space. It *has* been reported that females' produce a larger vowel space than males [9]. Even when differences in vocal tract size are accounted for, the variation in F1 across the space is larger for females. If females *are* producing high tense vowels with a breathy phonation, then the increased amplitude of the first harmonic could be effectively lowering F1 causing a non-linear stretch of the F1 space.

Along with the regularity among African languages, this could be another example of the premium placed on the perceptual distinctiveness of speech sounds.

This work was supported by NIDCD Grant DC-00719 and NSF Young Investigator Award DBS-9258482 to the first author.

IV. REFERENCES

[1] Berry, J. (1955), "Some notes on the phonology of the Nzema and Ahanta dialects", *Bulletin of the School of Oriental and African Studies*, vol. 17, pp. 160-165.

[2] Jacobson, L.C. (1980), "Vowel-quality harmony in Western Nilotic languages", in R.M. Vago (ed.) *Issues in Vowel Harmony*, John Benjamins: Amsterdam, pp. 183-200.

[3] Stewart, J.M. (1967), "Tongue root position in Akan vowel harmony", *Phonetica*, vol. 16, pp. 185-204.

[4] 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*, vol. 87, pp. 820-857.

[5] Halle, M., & Stevens, K. (1969), "On the feature 'Advanced Tongue Root'", *MIT Research Laboratory of Electronics Quarterly Progress Report*, vol. 94, pp. 209-215.

[6] Diehl, R.L., & Kluender, K.R. (1989), "On the objects of speech perception", *Ecological Psychology*, vol. 1, pp. 121-144.

[7] Liljencrants, J., & Lindblom, B. (1972), "Numerical simulation of vowel quality systems: The role of perceptual contrast", *Language*, vol. 48, pp. 839-862.

[8] Henton, C.G., & Bladon, R.A.W. (1985), "Breathiness in normal female speech: Inefficiency versus desirability", *Language & Communication*, vol. 5, pp. 221-227.

[9] Fant, G. (1975), "Non-uniform vowel normalization", *Quarterly Progress Status Report*, RIT Stockholm, vol 2/3, pp. 1-19.