Original Paper

Phonetica 1997;54:76–93

Received: March 4, 1996
Accepted: November 7, 1996

Effect of Voice Quality on Perceived Height of English Vowels

A.J. Lotto, L.L. Holt, K.R. Kluender

University of Wisconsin, Madison, Wisc., USA

Abstract

Across a variety of languages, phonation type and vocal-tract shape systematically covary in vowel production. Breathy phonation tends to accompany vowels produced with a raised tongue body and/or advanced tongue root. A potential explanation for this regularity, based on a hypothesized interaction between the acoustic effects of vocal-tract shape and phonation type, is evaluated. It is suggested that increased spectral tilt and first-harmonic amplitude resulting from breathy phonation interact with the lower-frequency first formant resulting from a raised tongue body to produce a perceptually 'higher' vowel. To test this hypothesis, breathy and modal versions of vowel series modelled after male and female productions of English vowel pairs /i/ and /ɪ/, /u/ and /ʊ/, and /ʌ/ and /a/ were synthesized. Results indicate that for most cases, breathy voice quality led to more tokens being identified as the higher vowel (i.e. /i/, /u/, /ʌ/). In addition, the effect of voice quality is greater for vowels modelled after female productions. These results are consistent with a hypothesized perceptual explanation for the covariation of phonation type and tongue-root advancement in West African languages. The findings may also be relevant to gender differences in phonation type.

In a number of vowel systems, there exists a production covariation between vocal tract shape and phonation type. Breathy phonation tends to be associated with a raised tongue body or an advanced tongue root [Denning, 1980]. For example, in a variety of East and West African languages, an advanced tongue root ([+ATR]) vowel is produced with breathy phonation (sometimes called 'lax' voice) whereas a modal or even creaky voice quality is used with non-ATR vowels [Berry, 1955; Stewart, 1967; Jacobson, 1980]. In fact, for some vowel harmony systems such as Akan, breathiness has been referred to as 'the main auditory correlate of root advancing' [Stewart, 1967]. Denning [1989] suggests, based on a study of around 50 languages, that, when there is a covariation between phonation type and vocal tract shape, it is invariably the case that breathy phonation is associated with 'higher' vowels.

Covariations between presumably independent articulatory variables often suggest that the variables operate synergistically either in terms of articulatory ease and/or perceptual robustness. A potential explanation for the covariation of phonation type

KARGER

E-Mail karger@karger.ch
Fax+41 61 306 12 34
http://www.karger.ch

Andrew J. Lotto
Department of Psychology
1202 West Johnson Street, Madison, WI 53706 (USA)
Tel. (608) 262-6110
E-Mail: ajlotto@facstaff.wisc.edu

and vocal-tract shape present in languages such as Akan may be based on hypothesized perceptual advantages from the interaction between acoustic effects of source and transfer function. If breathy phonation and an advanced tongue root or raised tongue body conspire to enhance those spectral features which distinguish the resultant vowel from other vowels in the language system, then one would expect that breathiness and these vocal tract shapes would be used in a correlated fashion both phonemically and subphonemically. The increase in communicative robustness obtained from this kind of articulatory covariation is presumably important to a communication system frequently challenged by unpredictable and noisy environments. If this covariation *does* result in a more distinctive phonetic inventory, then it would lengthen the list of speech inventory regularities that may exploit general auditory predispositions of listeners [e.g. Liljencrants and Lindblom, 1972; Diehl et al., 1990]. To evaluate this hypothesis, one needs to be acquainted with the joint acoustic/auditory effects of the phonation types and varying vocal tract shapes.

Acoustic effects of breathy phonation are essentially twofold. Vowels produced with breathy phonation are characterized by an increased open quotient in the vocal fold vibratory cycle [Holmberg et al., 1988; Price, 1988]. That is, for breathy phonation, vocal folds are adducted for a relatively short time in relation to the entire vibratory cycle. Acoustically, an increased open quotient generates an amplified first harmonic (fundamental) and a more radical energy decline in the higher-frequency region of the spectrum, both of these resulting from the more nearly sinusoidal glottal waveform [Hanson, 1995; Klatt and Klatt, 1990]. Together, these acoustic consequences contribute to relatively greater energy in the low frequencies of the resulting vowel. Figure 1 displays FFTs of two versions of the same synthesized vowel, one using parameters appropriate for breathy voice quality and the other using parameters appropriate for modal voice.

The primary acoustic effect of a raised tongue body and/or an advanced tongue root is the lowering of the first formant (F_1) frequency. The articulations are, to some extent, similar. Both articulations are usually accompanied by a lowering of the larynx, which results in a widening of the pharynx [Perkell, 1969; Denning, 1989]. Thus, extralaryngeal articulations lead to an increased prominence in the lower frequencies (as the result of lower F_1 frequency).

Because source and transfer function as described here both effectively enhance lower-frequency energy, it is possible that their joint acoustic consequences interact to enhance synergistically the physical/perceptual distinctiveness of resultant vowels. Due to these acoustic consequences, breathy phonation could be useful in signalling those vowels that are distinguished by low-frequency prominences, such as [+high] vowels or [+ATR] vowels. It seems reasonable to postulate that speakers (and, by consequence, language systems) will tend to favor the covariation of breathy phonation with ‘higher’ vowels, because the acoustic consequences of breathy phonation complement the acoustic ‘signature’ of ‘high’ vowels, i.e. a low-frequency F_1[1]. This prediction is, in spirit, similar to those arising from the Auditory Enhancement Hypothesis

[1] There is some evidence that vocal-tract shape and phonation type are allowed to vary independently. It has been reported that the Nilotic language Kalenjin covaries breathy phonation with non-ATR vowels [Local, 1995]. It should be noted, however, that independence of articulation is not a requisite proposition of the argument being made here. Even if breathy phonation and ATR are articulatorily interdependent, the resulting speech sounds are very distinct from non-ATR modal vowels and they will, thus, tend to be included in vowel systems.

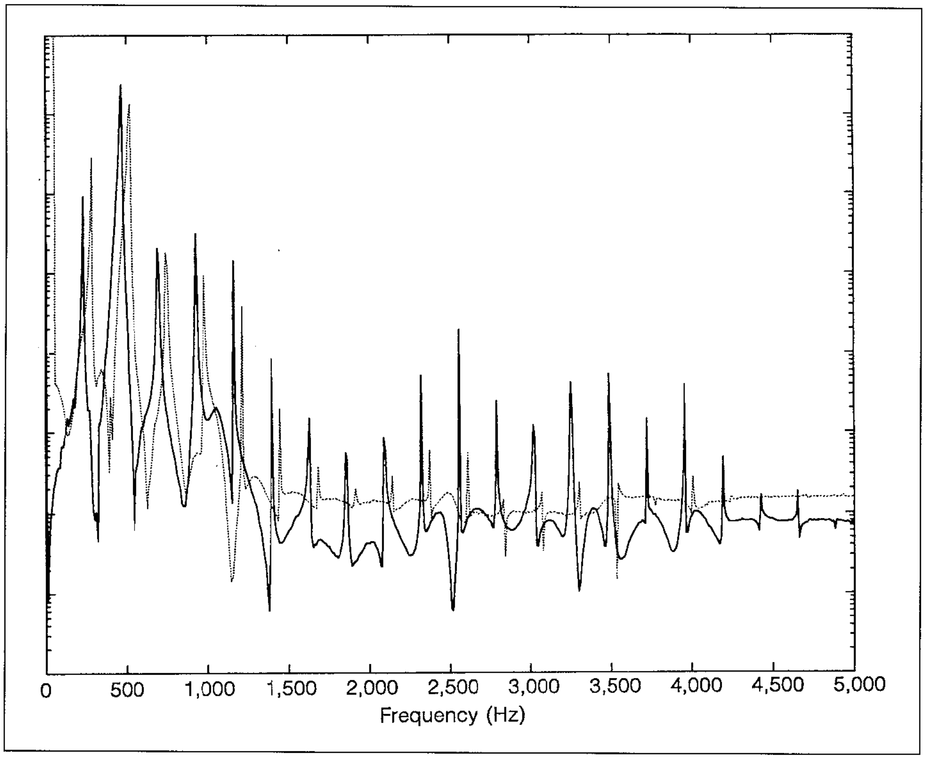


Fig. 1. FFTs of an intermediate member of the 'female' /u/ series from experiment 1. The solid line is for a vowel synthesized with parameters appropriate for a modal voice (OQ=50, TL=0) and the dashed lines, which are shifted right (+50 Hz), are for a breathy-voice (OQ=72, TL=17) [u]. Note that the amplitude of the first harmonic is enhanced in the breathy vowel.

[e.g. Diehl and Kluender, 1989; Diehl, 1991], which states that 'the phonetic features of vowels and consonants covary as they do largely because language communities tend to select features that have mutually enhancing auditory effects' [Diehl, 1991, p. 124]. The interaction between low-frequency F_1 ([+high] or [+ATR]) and the acoustic effects of breathiness are potential candidates as 'mutually enhancing auditory effects' and would be predicted to covary in language systems.

Of course, the preceding argument is critically dependent on the perceptual relevance of any interaction between the acoustic effects of breathy phonation and raised tongue body/advanced tongue root. It must be shown that these acoustic interactions affect the perceptual behavior of the listener.

A test of the perceptual interaction of F_1 frequency and voice quality (In this paper 'breathy phonation' will refer to the articulatory act of an increased open quotient, while 'breathy voice quality' will refer to the acoustic effects of this articulatory variable and to these acoustic attributes in synthesized vowels.) was reported by Thorburn et al. [1994]. Using the Garner [1974] paradigm, Thorburn et al. [1994] found that in CVC context, synthetic vowels were more easily discriminated when vowels with low-frequency F_1s were synthesized with parameters of a breathy voice quality and vowels with high F_1 frequencies were synthesized with a modal voice as contrasted with cases for which voice quality and F_1 frequency were matched in the complementary pattern. They proposed that voice quality and F_1 frequency were perceptually *integrated.*

Of course, speech perception consists of more than simple discrimination of speech sounds. At some point the listener has to identify the vowel as the member of some functional equivalence class, e.g. identify the sound as an exemplar of a particular phoneme. The question remains whether the interaction described above and in Thorburn et al. [1994] affects the identification of vowels. Is the pairing of breathy phonation with 'higher' vowels described by Denning [1989] due to an effect of breathy voice quality on perceived 'height'? Experiments described in this report were designed to test whether the interaction between voice quality and F_1 frequency affects the labelling of synthesized vowels in a manner that is consistent with the regularities reported by Denning [1989].

The first experiment involved the tense/lax contrast for [+high] English vowels (/i/ vs. /ɪ/ and /u/ vs. /ʊ/. This contrast is informative because the distinction is signalled, in part, by the lower-frequency F_1 of tense vowels. If these vowels are synthesized with a breathy voice quality, it is possible that the increased amplitude of the first harmonic along with the increased spectral tilt will result in a lower effective F_1 frequency or a 'higher' perceived vowel. This would result in more 'tense' identifications (i.e. /i/ and /u/). The following experiments tested this possibility by obtaining identifications from native English speakers for tense and lax high vowels varying in voice quality.

Experiment 1

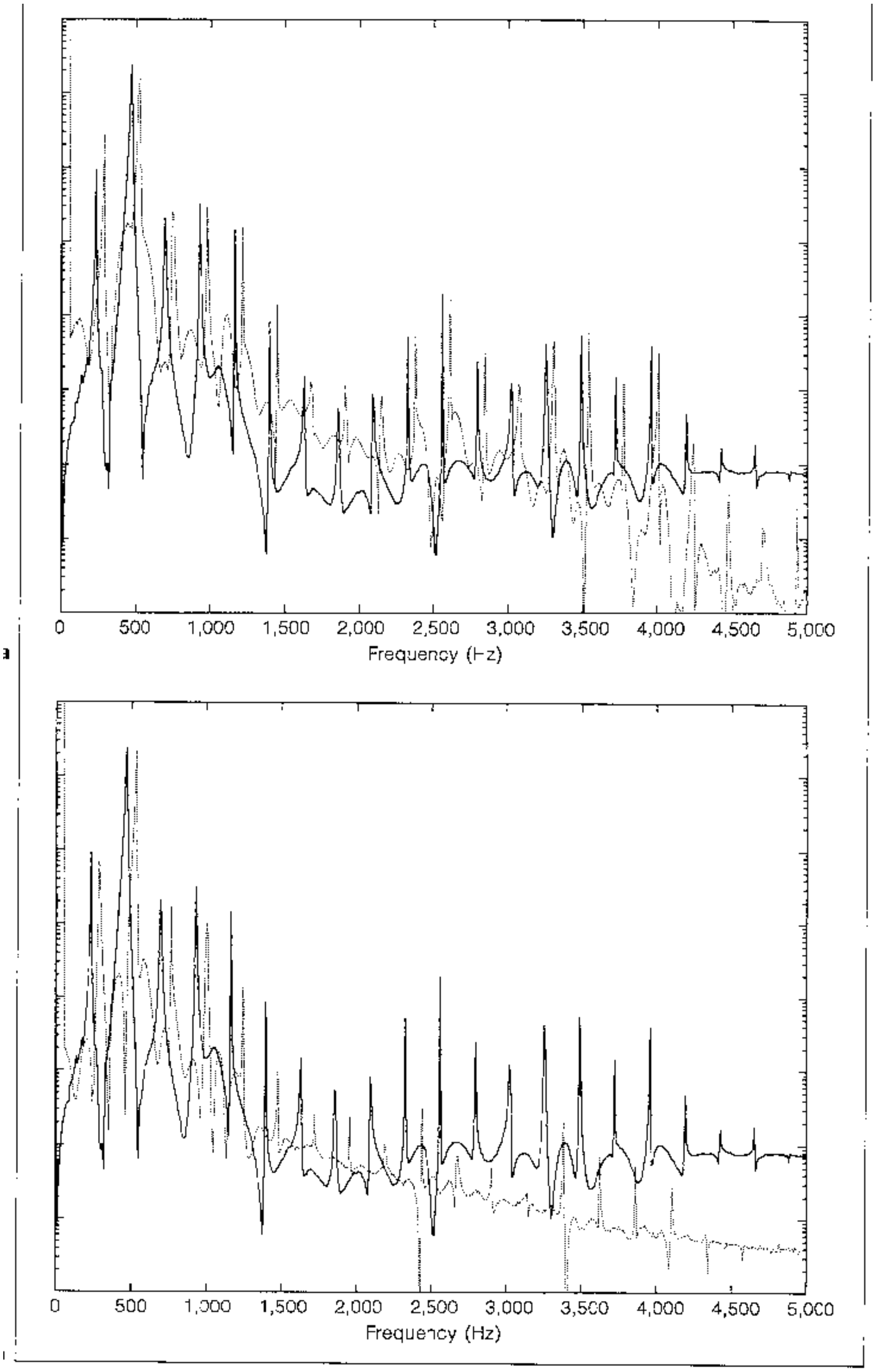
Subjects

Twenty-two college-age adults, all of whom learned English as their first language, served as listeners. All reported normal hearing. Subjects received Introductory Psychology course credit for their participation.

Stimuli

Four vowel series were synthesized with eight endpoints modelled 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 mainpulating the nominal center frequencies of the first three formants. In particular, frequency of F_1 varied linearly from 210 to 330 Hz for the male front series and from 330 to 440 Hz for the male back series. For the series based on female productions, F_1 varied in equal steps from 310 to 430 Hz for front vowels and from 370 to 480 Hz for back vowels. These F_1 frequency values are lower than those used in Thorburn et al. [1994], which varied between 450 and 544 Hz. Nominal formant bandwidths were equal for all stimuli. The values for the first three formants were 60, 90, and 150 Hz for F_1, F_2, and F_3, respectively. Fundamental frequency was constant at 135 Hz for the series modelled after male productions and 233 Hz for the female series. All stimuli were 120 ms in duration. A fuller description of the endpoint synthesizer parameters is given in table 1. Use of both male and female versions of the stimuli allowed a more complete test of the validity of the hypothesis than has been available in past studies of language regularities which focused solely on the perception of vowels modelled after male productions.

Breathy versions of each seven-step series were created by increasing the amplitude of the first harmonic and increasing spectral tilt utilizing the software synthesizer described in Klatt and Klatt [1990]. The amplitude of the first harmonic was controlled by the synthesizer parameter OQ (open


0
500
1,000
1,500
2,000
2,500
3,000
3,500
4,000
4,500
5,000
Frequency (Hz)
a
0
500
1,000
1,500
2,000
2,500
3,000
3,500
4,000
4,500
5,000
Frequency (Hz)

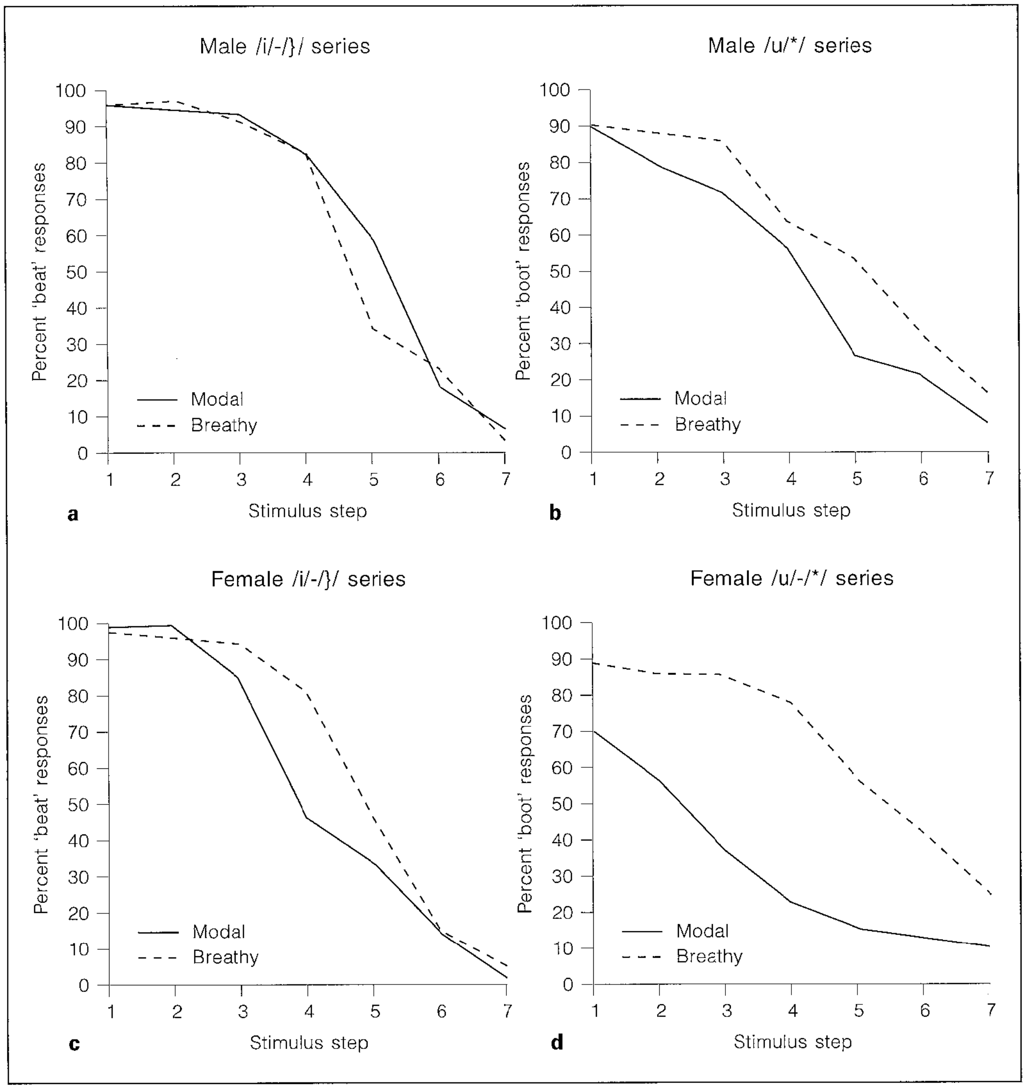


Fig. 3. Identification functions for experiment 1. Stimulus step 1 is the 'tense' endpoint of the series and 7 is the 'lax' endpoint. **a** 'Male' front series. **b** 'Male' back series. **c** 'Female' front series. **d** 'Female' back series.

Fig. 2. FFTs of an intermediate member of the 'female' /u/ series. For both figures the solid line is from the modal vowel series (OQ=50, TL=17). The dashed line, which is shifted right, is for the modal vowel with OQ=72 (**a**) the modal vowel with TL=17 (**b**).

Table 1. Synthesizer parameter values for the endpoint stimuli of series from experiment 1

	Frequency of endpoint stimuli, Hz							
	front series				back series			
	male /i/	male /ɪ/	female /i/	female /ɪ/	male /u/	male /ʊ/	female /u/	female /ʊ/
F_1	210	330	310	430	330	440	370	480
F_2	1,830	1,530	2,630	2,195	810	1,020	900	1,130
F_3	2,750	2,290	3,310	2,850	2,240	2,240	2,490	2,490

Breathy stimuli: OQ=72, TL=17; modal stimuli: OQ=50, TL=0.

quotient). As can be seen in figure 2a, the primary effect of an increase in OQ is an increase in the first harmonic amplitude with little effect on higher frequencies. The values of OQ differed for modal (OQ=50) and breathy (OQ=72) vowels. Increased spectral tilt, which is a signature of breathy phonation, was replicated using the TL (tilt) parameter. This variable establishes the additional downward tilt of the spetrum at 3 kHz. Figure 2b demonstrates the difference between TL=0 (modal) and TL=17 (breathy).

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 DDT-100) at 75 dB SPL.

Procedure

Subjects participated in a two-response forced-choice identification task. In each experimental session, 1–3 subjects were tested concurrently in single-subject sound-attenuated booths (Suttle Equipment). Subjects heard each of four stimulus blocks (male/female x front/back vowels) separately and presentation of blocks was counterbalanced across subjects. Within each block, vowels were presented 10 times each in random order. Subjects identified each vowel by pressing either of two labelled response box buttons which corresponded to /i/ and /ɪ/ or to /u/ and /ʊ/ depending on the stimulus block being presented. To aid subjects, sample words 'beat' and 'bit' or 'boot' and 'book' were written next to the response buttons. In all, subjects responded to 560 stimuli at a presentation rate of approximately 1 stimulus every 3s. Each experimental session took approximately 30 min dependent on subjects' speed of response.

Results

Data from 6 subjects were withheld from analysis because they failed to reach 80% correct average performance across the endpoints. A 2×7 (Voice Quality × Stimulus Series Step) within-subjects analysis of variance (ANOVA) was performed separately for each of the four stimulus blocks (male/female×front/back) on the percent of 'tense' (i.e., /i/ and /u/) responses. Identification functions for the breathy and modal versions of each gender×vowel series are displayed in figure 3.

For both vowel pairs modelled after female productions, there were significant main effects of voice quality ($F_{(1,15)}=27.08$, $p<0.0005$, for the front series; $F_{(1,15)}=29.88$, $p<0.0005$, for the back series). In both cases, stimuli synthesized to mimic breathy phonation were judged more often to be 'tense'. There was also a significant main effect of voice quality for the male back series ($F_{(1,15)}=4.75$, $p<0.05$), with 'breathy' stimuli being labelled more often as 'tense'. In contrast, for the male

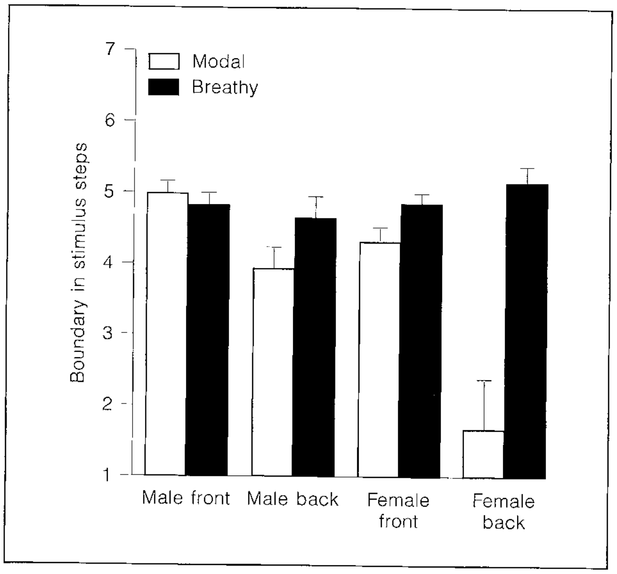


Fig. 4. Probit boundary values for all of the series from experiment 1 with attendant standard errors.

front series the effect of voice quality exhibited the opposite pattern. In this case, subjects more often identified vowels using the 'tense' labels when voice quality was modal. This effect was small, but consistent and statistically significant ($F_{1,15)}=7.57$, $p<0.05$).

In addition to overall percentage of 'tense' judgments, identification boundary estimates were determined for the breathy and modal versions of each series using probit analysis. Mean boundaries for each of the eight series are displayed in figure 4 in terms of the stimulus steps (one='lax' endpoint; seven='tense' endpoint).

Despite the singular effect of voice quality on the male front series, the three largest boundary shifts indicate that the purported acoustic interaction between voice quality and F_1 frequency does have an effect on English vowel identification. In general, breathy voice quality gives rise to more 'tense' identifications, particularly for back vowel series. Given these results, one may suppose that a breathy [u] is more distinct from a modal [ʊ] than a modal [u] is distinct from a breathy [ʊ]. A phonetic inventory that included breathy [u] and modal [ʊ] would benefit in terms of perceptual distinctiveness. (This is assuming that the pattern of voice quality does not reduce the distinctiveness of these two segments from the rest of the phonetic inventory. Of course, the extreme positions of /i/ and /u/ in articulatory and acoustic space virtually assures that breathy productions of these segments will be beneficial.) Thus, the covariation of vowel height and production type described by Denning [1989] is very possibly an example of a combination of articulatory factors which is favored across languages because the acoustic consequences of these factors are mutually enhancing in a manner that is perceptually relevant.

Of course, this putative explanation may be weakened by the anomalous effect for the male front block. In this case, the coupled effects of an increased amplitude first harmonic and a rapid decline of energy for higher frequencies resulted in fewer 'tense' responses. Walsh et al. [1995] report that there are conditions in the Thorburn et al. [1994] study in which a reverse shift also was obtained, i.e. there was greater discriminability when high frequency F_1 was associated with breathy voice quality. This occurred for intermediate levels of breathiness and not for the higher level of breathiness (i.e. higher amplitude of first harmonic and greater tilt) used in the present experiment. It is not clear whether the increase in spectral tilt or the increase in the first harmonic is culpable for this reverse. Experiment 2 was an attempt to clarify the effects of each of these acoustic manipulations on subjects' responses.

Experiment 2

One possible explanation for the results of experiment 1 is that increased spectral tilt of the series modelled after breathy productions degraded higher frequencies to the point where subjects were encouraged or forced to rely solely on the first formant for identification. This is especially a possibility for the back vowel series because a reasonable sounding [u] can be constructed from a single low-frequency formant [Chistovich and Lublinskaya, 1979]. Experiment 2 was designed to parcel effects of exaggerated spectral tilt and increased amplitude of the fundamental harmonic.

Subjects

Subjects were 11 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

Stimuli were identical to those used in experiment 1 except that for the breathy series only the amplitude of the first harmonic (OQ=72) was increased. Spectral tilt remained at the modal values from experiment 1 for all series.

Procedure

Stimulus presentation and equipment was identical to that used in experiment 1. Subjects identified vowels as those in the words 'beat' and 'bit' or 'boot' and 'book' via a computer-controlled response box.

Results

As in experiment 1, percent 'tense' responses for each series were subjected to a 2×7 (Voice Quality×Stimulus Series Step) within-subject ANOVA. Identification functions are displayed in figure 5.

The ANOVA revealed a significant effect of first harmonic amplitude for both the female front series ($F_{(1,10)}=48.65$, $p<0.0001$) and the female back series ($F_{(1,10)}=10.69$, $p<0.01$). On the other hand, there was no significant effect of increased first harmonic amplitude on the two male series (front: $F_{(1,10)}=0.45$, $p=0.52$; back: $F_{(1,10)}=2.14$, $p=0.17$). Results are redescribed in the probit boundaries displayed in figure 6.

For the two series modelled after male productions, amplitude of the first harmonic seems to play a negligible role in the height of the perceived vowel. Thus, the spectral tilt that was present for 'male' vowels in experiment 1 likely caused the shifts in the identification functions. It is possible that increased tilt resulted in a perceptual lowering of the second formant (F_2) frequency. For the harmonics which define F_2 in the signal, there is a change in relative amplitude when spectral tilt is increased.

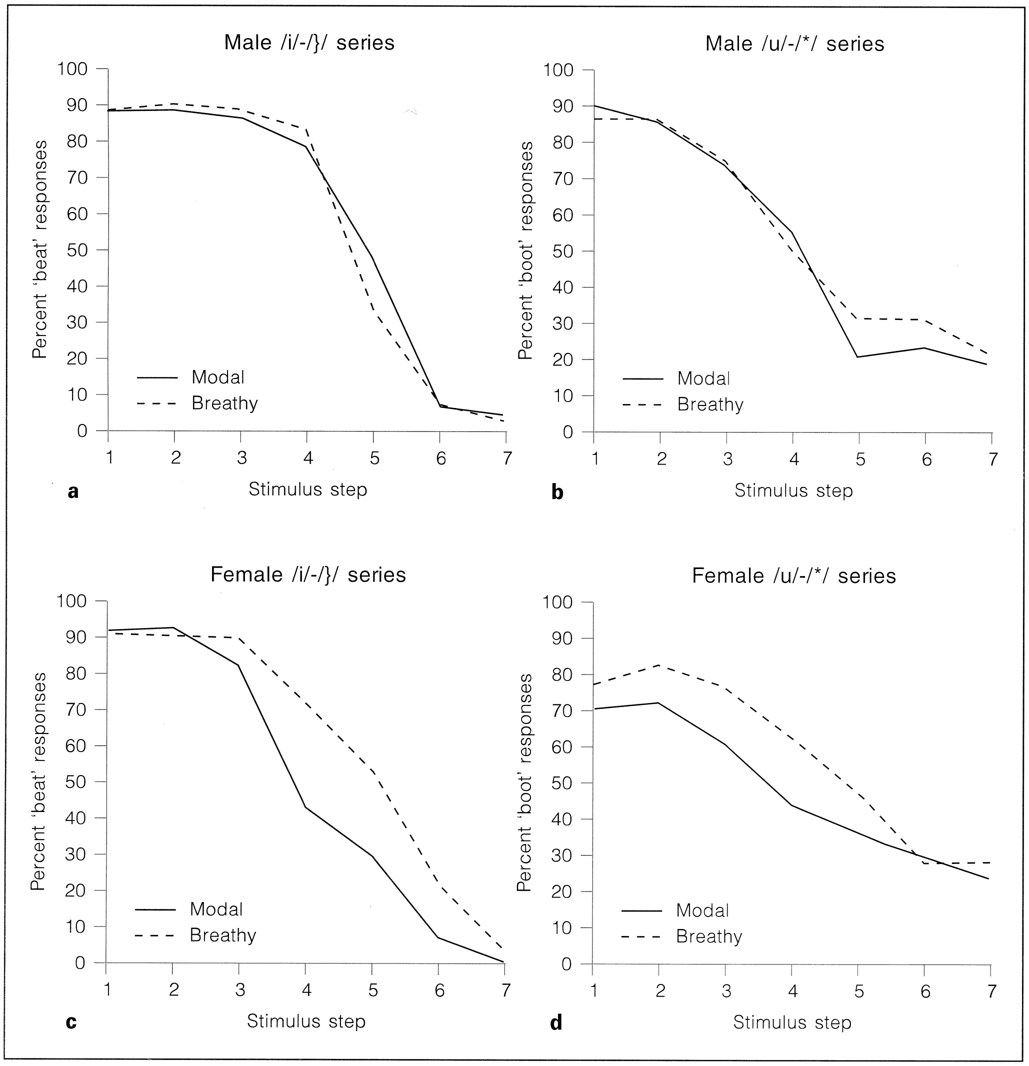


Fig. 5. Identification functions for experiment 2. Stimulus step 1 is the 'tense' endpoint of the series and 7 is the 'lax' endpoint. **a** 'Male' front series. **b** 'Male' back series. **c** 'Female' front series. **d** 'Female' back series.

Higher-frequency harmonics are attenuated in relation to lower frequency harmonics. Thus the center of the spectral peak of F_2 is shifted to a lower frequency. This state of affairs is represented schematically in figure 7. If this acoustic change affected the perceived frequency of F_2, then one would expect more /u/ responses for the back series and more /ɪ/ responses for the front series. (It is not necessary for this explanation that formant frequency information is extracted and used for vowel identification. A more holistic model of vowel identification, e.g. template matching, would presumably pre-

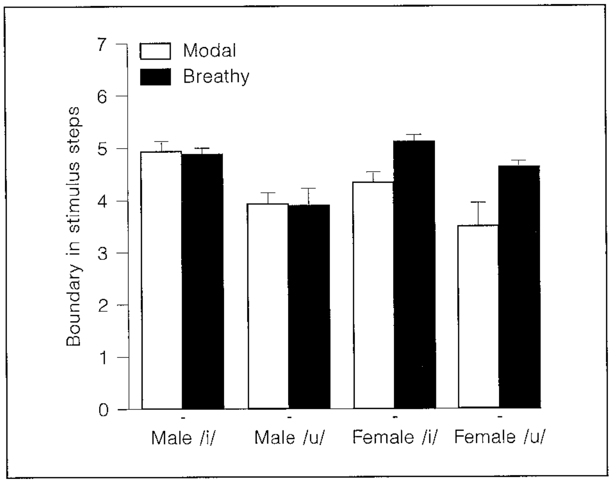


Fig. 6. Probit boundary values for all of the series from experiment 2 with attendant standard errors.

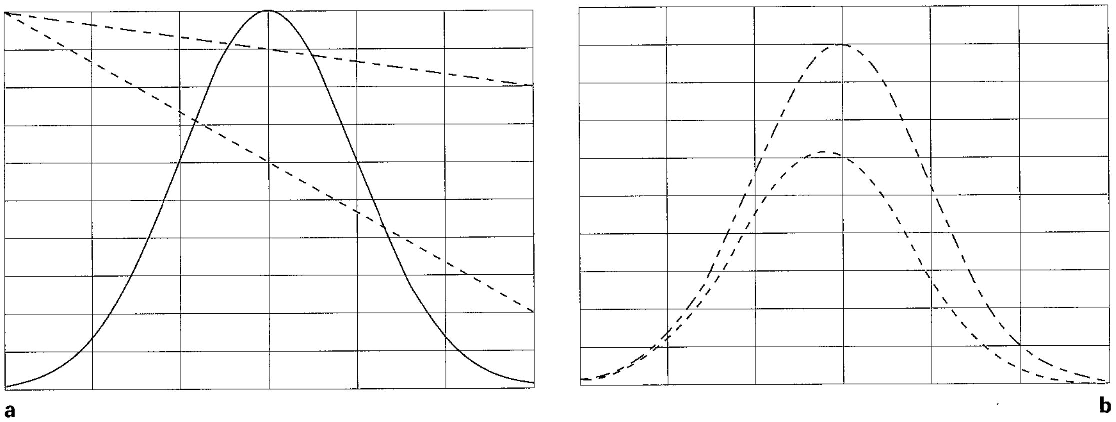


Fig. 7. Schematic of a formant filter function with two vectors representing two different slopes of spectral tilt. Breathy voice quality corresponds to the vector with the greater slope. **b** Results of a multiplication of the filter and each vector. Note that there is a shift in the peak of this function to the left (lower frequency) for the 'breathy' case.

dict this change in responses given the change in relative spectral energy.) These were the obtained results in experiment 1.

Outcomes are quite different for the two series modelled after female productions. More 'tense' responses were obtained for stimuli with an increased-amplitude first harmonic for both series. This is probably due to the fact that, because of the high f_0 of the female stimuli, the first harmonic was the peak harmonic for the first formant for most of the stimuli, including stimuli near the boundaries. The consequence of increasing the amplitude of this harmonic is a lower perceived F_1 frequency [Assmann and Nearey, 1987]. It also appears that spectral tilt is an important determinant of perceived height for the female back series, since the boundary shift for this series was

Table 2. Synthesizer parameter values for the endpoint stimuli of series from experiment 3

	Frequency of endpoint stimuli, Hz			
	male series		female series	
	/ʌ/	/a/	/ʌ/	/a/
F_1	645	465	700	1,000
F_2	1,191	1,089	1,300	1,400
F_3	2,388	2,442	2,550	2,700

Breathy stimuli: OQ=72, TL=17; modal stimuli: OQ=50, TL=0.

much smaller in experiment 2 than that from experiment 1 (1.05 steps to 4.02 steps, respectively). Thus, it appears that the increased spectral tilt which accompanies breathy phonation affects the perception of vowels modelled after both male and female speakers.

From this analysis, it appears that the perceptual effect of breathy voice quality is lowering of the effective frequency of F_1 for female productions and the effective F_2 frequency for male productions. Except for the case of male [i], the perceptual effect of breathiness is an enhanced 'tenseness' for high vowels. But, how general is this effect across the vowel space? In particular, do these findings extend to low vowels?

Experiment 3

To answer questions concerning generality, a low-vowel contrast was synthesized with breathy and modal parameters. Subjects were asked to classify stimuli from [ʌ]-[a] series based on male and female productions. If the acoustic effects of breathy phonation lead to greater perceived height, then more /ʌ/ responses should be elicited with stimuli from the breathy series. However, it is possible that there will be no effect of voice quality because the higher F_1 frequency of these low vowels presumably will not interact with the manipulation of the amplitude of the first harmonic.

Subjects

Twenty-three college-age adults, all of whom reported to be native English speakers and to have normal hearing, served as listeners. Subjects received course credit for their participation.

Stimuli

Four series were synthesized with endpoints modelled after male and female productions /ʌ/ and /a/. Five intermediate steps were created by manipulating the center frequencies of the first three formants. Parameter values for the endpoints are displayed in table 2. As in experiments 1 and 2, all stimuli were 120 ms in duration. Breathy versions of both the male and female series were created by increasing the amplitude of the first harmonic and increasing spectral tilt (as in experiment 1).

Procedure

The forced-choice identification task was identical to that used in experiments 1 and 2, except that the buttons were labelled 'hut' and 'hot'. Subjects heard four randomized blocks in a session. Two of the blocks contained stimuli modelled after male productions and two of the blocks contained female stimuli. Subjects heard all of the four continua with order of gender counterbalanced. Thus, subjects responded to each stimulus 20 times.

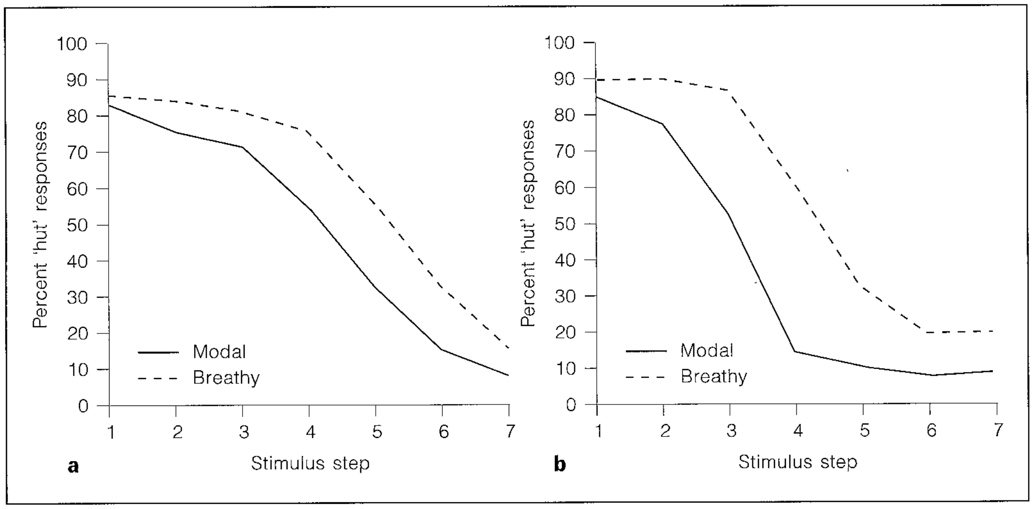


Fig. 8. Identification functions for experiment 3. Stimulus step 1 is the /ʌ/ endpoint of the series and 7 is the /a/ endpoint. **a** 'Male' series. **b** 'Female' series.

Results

Data from 5 subjects were not included in the analyses because the subjects correctly identified the endpoint stimuli less than 80% of the time. This cutoff was considered reasonable because the authors judged endpoints to be very good exemplars of the intended vowels and because most subjects were well over 90% correct on these endpoints. Data from the remaining 18 subjects were subjected to two separate 2×7 (Voice Quality×Stimulus Series Step) ANOVAs, one for each gender condition. In both cases there was a highly significant effect of voice quality (male: $F_{(1,17)}=44.77$, $p<0.0001$; female: $F_{(1,17)}=78.91$, $p<0.0001$). Identification functions are presented in figure 8. For both gender series, a vowel synthesized with a breathy source led to an increased percentage of identifications corresponding to the 'higher' (lower F_1) vowel, [ʌ].

Given the higher F_1 of these stimuli (645–765 Hz for male series and 700–1,000 Hz for female series), this result may be rather surprising. It is difficult to imagine that the increased amplitude of the first harmonic would affect the perceived frequency of the first formant, since the peak harmonic for F_1 is the third and fourth for female stimuli and the fifth for male stimuli. It should be noted, however, that Thorburn et al. [1994] found perceptual effects of breathy voice quality for stimuli with higher F_1s (450–544 Hz) than were used in experiment 1. Experiment 4 was designed to uncover the manipulation – spectral tilt or increased amplitude of the first harmonic – that was responsible for this effect.

Experiment 4

Subjects

Twenty-two native-English speaking college-age adults with normal hearing participated as listeners for course credit.

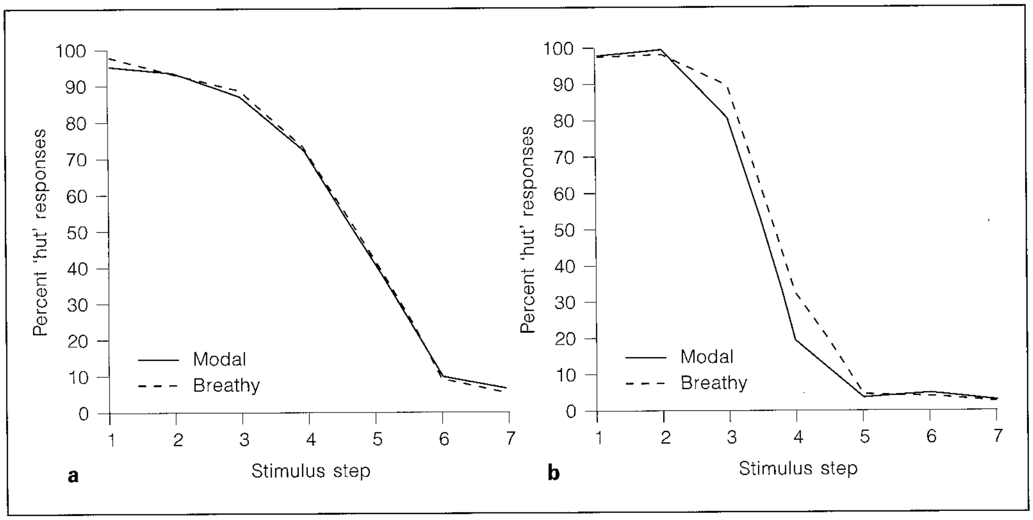


Fig. 9. Identification functions for experiment 4. Stimulus step 1 is the /ʌ/ endpoint of the series and 7 is the /a/ endpoint. **a** 'Male' series. **b** 'Female' series.

Stimuli

Stimuli for experiment 4 were identical to those in experiment 3; except spectral tilt was held constant between the modal and breathy series. Thus, only the amplitude of the first harmonic (OQ; OQ=72 for breathy stimuli and OQ=50 for modal series) was varied.

Procedure

The procedure was identical to experiment 3.

Results

Data from 2 subjects were withheld from analysis because they failed to reach 80% correct performance on endpoints. Data from the remaining 20 subjets were subjected to two within-subjects ANOVAs. Identification functions are presented in figure 9.

The ANOVA for the male stimuli revealed what clearly can be ascertained visually from figure 9a: there was no effect of increased first harmonic on the identification of the vowels ($F_{(1,19)}=0.57$, $p=0.46$). One may suppose that spectral tilt lowers the perceived frequency of the first formant in the manner described for the lowering of F_2 in experiment 2. That is, higher frequency harmonics are compromised relative to lower frequency harmonics resulting in a shift of the center of spectral energy to a lower frequency. Follow-up studies support this view. When subjects are played low vowels varying in tilt alone, one sees a shift in identifications similar to that seen in experiment 3[a].

[a] Likewise, one can replicate the effects for males in experiment 1 by simply manipulating spectral tilt. For high vowels modelled after female stimuli, spectral tilt results in more lax identifications for front vowels and more tense identifications for back vowels. This is in line with the hypothesis that spectral tilt lowers the perceived frequency of F_2. These experiments were all conducted using the same procedures as the experiments reported in this paper. The details have not been included here because there does not appear to be an interaction between spectral tilt and amplitude of F_1 and, thus the dissociations in experiments 2 and 4 are fully informative.

The effect of voice quality (OQ) was also attenuated for the female-modelled stimuli as compared to experiment 3. However, a small, but significant, effect of source characteristics on identification remained ($F_{(1,19)}=19.24$, $p<0.0005$). Subjects identified vowels near the boundary as /ʌ/ more often when the amplitude of the first harmonic was increased. This correponds to an increase in perceived height. (The authors are agnostic concerning the psychological reality of phonetic labels such as 'height'. To be more accurate: there was a mapping from stimulus step to the 'hut' response more often in the breathy condition.) It appears that amplitude of the first harmonic may play a role in perceived frequency of F_1 even when the peak harmonic for the formant is the third as is the case for the female low vowels. Also, FFTs of the female stimuli exhibit some attenuation of higher frequency harmonics – a result of RMS matching stimuli after energy is added to the first harmonic.

The effects of spectral tilt and first harmonic amplitude on vowel identification are very interesting from the standpoint of a general theory of vowel perception and identification. Whether one postulates a model of formant value extraction or a more holistic spectral pattern recognition, these data need be addressed. If the perceptual system extracts formants, then the effect of the first harmonic amplitude on female low vowels suggests that the spectral information used to determine formant frequencies may be gathered over a wide frequency range and that this range may vary with f_0 or speaker given the lack of an effect for male vowels. Likewise, a more holistic pattern recognizer for speech may have difficulty with perceptual effects of lowering the center of spectral gravity for female-produced speech, but not for male-produced speech.

General Discussion

The purpose of this investigation was to test a hypothesis concerning a potential interaction between voice quality and F_1 frequency on the identification of English vowels. The data indicate that the quality of voice used to produce a vowel affects the height of the perceived vowel. The five largest shifts in identification across two experiments were toward the higher vowel when the stimulus was synthesized with parameters appropriate for breathy phonation.

These results suggest that the covariation in African languages between phonation type and tongue root advancement may be consistent with the general principles of Auditory Enhancement [Diehl and Kluender, 1989]. This theoretical framework predicts that articulatory factors which tend to enhance the perceptual distinctiveness of resultant speech sounds will be favored in phonetic inventories. Breathy phonation enhances the low frequency prominence of tense high vowels and, by extension, ATR high vowels. Consistent with the present results, if a language contained the complementary pattern of breathy production for lax high vowels, then the distinctiveness of the tense/lax contrast would be reduced. However, it appears that languages tend to employ the more distinctive pairing of breathy phonation with advanced tongue root and creaky voice (which would deemphasize the low frequencies) with nonadvanced tongue root [Denning, 1989].

It should be noted that the explanation offered here for the effects of voice quality on identification is acoustic, as opposed to auditory, in nature. The interaction of breathy voice quality and F_1 frequency occurs in the physical acoustics of the signal.

The manipulation of spectral tilt changes the relative amplitude of the harmonics. The manipulation of F_1 frequency as a synthesizer parameter also changes the relative amplitude of the harmonics. For stimuli with a high f_0, such as stimuli modelled after female productions, the first harmonic falls into the region affected by the F_1 transfer function. Therefore, increasing the amplitude of this harmonic will interact with the effects of the F_1 filter. Kingston et al. [1995] describe the discrimination results from Thorburn et al. [1994] as evidence for a 'perceptual integration' of the acoustic correlates of voice quality and tongue root advancement. According to Kingston et al. [1995], covariation in articulations leads to an integrated perceptual property called 'flatness'. (This should not be confused with the Jakobsonian feature 'flatness' [Jakobson et al., 1953]). Given the acoustic explanation offered here, it would be imprecise to label this interaction 'perceptual'. Because acoustic effects of voice quality and F_1 frequency interact in the physical waveform, it should not be unexpected that there is an interaction in perceptual responses when these parameters are crossed. In this case, postulation of an integrated perceptual property seems unnecessary. The covariation of voice quality and tongue root advancement appears to be an example of articulations which together result in speech sounds which are physically distinctive and, as an important result, perceptually distinctive.

Gender of Speaker Differences

There are interesting divergences in the data based on whether stimuli were modelled on male or female productions. The effect of breathiness on identification was much more robust for female series. Even when spectral tilt was held constant in experiments 2 and 4, subjects labelled female vowels differentially when the amplitude of the first harmonic was changed, and in every comparison the effects of voice quality were larger for female series.

It has been widely reported that females use breathy phonation more often than men in a variety of languages, including English [Klatt and Klatt, 1990; Henton and Bladon, 1985]. If breathy productions lead to a 'higher' perceived vowel, then it would be advantageous for English-speaking females to use breathiness distinctively, i.e. use a breathy phonation for tense high vowels and modal phonation 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.

Unfortunately, there are limited relevant data concerning females' phonation types across the entire vowel space to support this aperçu. Bloedel [1994] measured the difference in amplitudes between the first (H_1) and second (H_2) harmonic (a common measure of breathy phonation) for male and female speakers reading a passage. She found that for females (but not males) H_1-H_2 was greater (that is, more 'breathy') for the high tense vowels /i/ and /u/ than for high lax vowels /ɪ/ and /ʊ/. This is exactly the pattern predicted by the hypothesis that breathiness is used to enhance perceptual identification. However, it is unclear at this point how the lower F_1 frequencies of the tense vowels may interact with the H_1-H_2 measurement. Production measurements which are insensitive to changes in formant parameters are necessary to test more fully the predictions listed above. Such work is currently underway in this laboratory.

If it turns out that females use breathiness distinctively as suggested by the Bloedel [1994] data, then another long-standing problem in speech production may be resolved. It has been reported that females produce a larger vowel space than males [Fant, 1966, 1975; Yang, 1990]. Even when differences in vocal tract size and proportion are accounted for, the variation in measured F_1 across the space is larger for females, due mostly to unexpectedly low F_1 frequencies for high vowels [Nordström, 1977]. If females *are* producing high tense vowels with a breathy phonation, then the increased amplitude of the first harmonic could be effectively lowering the frequency of measured F_1s causing a nonlinear stretch of the F_1-F_2 space. As described above, the interaction between voice quality and F_1 frequency occurs in the acoustic waveform and, thus, breathiness will affect the measurements of F_1 frequency from the physical waveform. Together with the regularity among African languages, gender differences in phonation may provide another example of the premium placed on the perceptual distinctiveness of speech sounds.

Acknowledgments

This work was supported by NIDCD Grant DC-00719 and NSF Young Investigator Award DBS-9258482 to the third author and by a fellowship from the Wisconsin Hilldale Undergraduate Research Program to the second author. Some of the data were presented at the fall meeting of the Acoustical Society of America, November 1994, in Austin, Tex. and at the International Congress on Phonetic Sciences, August 1995, in Stockholm. The authors would like to thank Suzi Bloedel and Beth Horner for their help with pilot studies and Ray Kent, John Kingston, Diane Bless, and Charles Read for discussions concerning breathy phonation.

References

Assmann, P.F.; Nearey, T.M.: Perception of front vowels: the role of harmonics in the first formant region. J. acoust. Soc. Am. *81:* 520–534 (1987).

Berry, J.: Some notes on the phonology of the Nzema and Ahanta dialects. Bull. Sch. Orient. Afr. Stud. *17:* 160–200 (1955).

Bloedel, S.L.: An analysis of the acoustic correlates of breathy phonation in the speech of adult men and women and pre-pubescent males; MS thesis University of Wisconsin, Madison (1994).

Chistovich, L.A.; Lublinskaya, V.V.: The 'center of gravity' effect in vowel spectra and critical distance between the formants: psychoacoustical study of the perception of vowel-like stimuli. Hear. Res. *1:* 185–195 (1979).

Denning, K.: The diachronic development of phonological voice quality; PhD diss. Stanford University (1989).

Diehl, R.L.: The role of phonetics within the study of language. Phonetica *48:* 120–134 (1991).

Diehl, R.L.: Kluender, K.R.: On the objects of speech perception. Ecol. Psychol. *1:* 121–144 (1989).

Diehl, R.L.; Kluender, K.R.; Walsh, M.A.: Some auditory bases of speech perception and production; in Ainsworth, Advances in speech, hearing and language processing (JAI Press, London 1990).

Fant, G.: A note on vocal tract size factors and non-uniform F-pattern scalings. Q. Prog. Status Rep., Speech Transm. Lab., R. Inst. Technol. Stockh., No. 4, pp. 22–30 (1966).

Fant, G.: Non-uniform vowel normalization. Q. Prog. Status Rep., Speech Transm. Lab., R. Inst. Technol., Stockh., No. 2–3, pp. 1–19 (1975).

Garner, W.R.: The processing of information and structure (Erlbaum, Potomac 1974).

Hanson, H.M.: Glottal characteristics of female speakers; PhD diss. Harvard University (1995).

Henton, C.G.; Bladon, R.A.W.: Breathiness in normal female speech: inefficiency versus desirability. Lang. Commun. *5:* 221–227 (1985).

Holmberg, E.B.; Hillman, R.E.; Perkell, J.S.: Glottal air flow and pressure measurements for soft, normal and loud voice by male and female speakers. J. acoust. Soc. Am. *84:* 511–529 (1988).

Jacobson, L.C.: Vowel-quality harmony in Western Nilotic languages; in Vago, Issues in vowel harmony (Benjamins, Amsterdam 1980).

Jakobson, R.; Fant, G.; Halle, M.: Preliminaries to speech analysis: the distinctive features and their correlates (MIT Press, Cambridge 1953).

Kingston, J.; Dickey, L.W.; Thorburn, R.; Bartels, C.; Macmillan, N.A.: Integrating voice quality and tongue root position in perceiving vowels; in Elenius, Branerud, Proc. 13th Int. Congr. on Phonet. Sci., vol. 2, pp. 514–517, Stockholm 1995.

Klatt, D.H.; Klatt, L.C.: Analysis, synthesis, and perception of voice quality variations among female and male talkers. J. acoust. Soc. Am. *87:* 820–857 (1990).

Liljencrants, J.; Lindblom, B.: Numerical simulation of vowel quality systems: the role of perceptual contrasts. Language *48:* 839–862 (1972).

Local, J.K.: Making sense of dynamic, non-segmental phonetics; in Elenius, Branerud, Proc. 13th Int. Congr. on Phonet. Sci., vol. 3, pp. 2–9), Stockholm 1995.

Nordström, P.-E.: Female and infant vocal tracts simulated from male area functions. J. Phonet. *5:* 81–92 (1977).

Perkell, J.S.: Physiology of speech production: results and implications of a quantitative cineradiographic study (MIT Press, Cambridge 1969).

Price, P.: Male and female voice source characteristics: inverse filtering results. Speech Commun. *8:* 261–267 (1988).

Stewart, J.M.: Tongue root position in Akan vowel harmony. Phonetica *16:* 185–204 (1967).

Thorburn, R.; Walsh, L.J.; Macmillan, N.A.; Kingston, J.: Components of integrality in the perception of voice quality and tongue root position (abstract). J. acoust. Soc. Am. *S95:* 2871 (1994).

Walsh, L.; Bartels, C.; Thorburn, R.; Kingston, J.; Macmillan, N.A.: Laxness integrates with F_1 (usually, but not always, negatively) (abstract). J. acoust. Soc. Am. *S97:* 3421 (1995).

Yang, B.: A comparative study of normalized English and Korean vowels; PhD diss. University of Texas at Austin (1990).