Proceedings of the Institute of Acoustics

ACOUSTIC CUES TO BREATHINESS : A TRUE MARKER OF GENDER?

.... Gavin J Dempster ....

Department of Computer Science, University of Sheffield, Sheffield, UK

1. INTRODUCTION

A number of speech characteristics have been described as being perceptually-distinctive markers of speaker gender (e.g. pitch, formant frequencies). This paper considers breathiness in the voice.

Several acoustic cues to the perception of breathiness have been posited, one of which is the difference in amplitude between the first and second harmonics. Studies involving speakers from the USA (Klatt & Klatt 1990, Nittrouer *et al.* 1990), the UK (Henton & Bladon 1985) and the Netherlands (Gunzburger 1991) all conclude that, on average, this difference in harmonics is more positive in women than in men. Thus it would appear that breathiness, as signalled by this harmonic amplitude difference, represents a consistent, cross-cultural cue to speaker gender.

However, the authors' conclusions are based on relatively small sample sets and disguise significant inter- and intra-speaker variations. Thus it was felt appropriate to investigate these claims for a larger number of speakers.

1.1 Physiology of Breathy Voice

During 'normal' or 'modal' phonation, the vocal folds tend to close simultaneously along their length, causing an abrupt cessation of the airflow from the lungs. The effect on the glottal volume velocity waveform is that the closing phase is more rapid than the opening phase; and subsequently the higher harmonics receive a relatively strong excitation (Klatt & Klatt 1990:822).

In contrast, during the production of breathy voice the arytenoid cartilages at the rear of the larynx hold the vocal folds apart, ensuring a constant airflow during phonation (although there is sufficient tension in the vocal folds for vibration of the folds to occur). This posterior glottal opening has been observed in a comparison of breathy and clear vowels by Fischer-Jorgenson (1967, cited Nittrouer *et al.* 1990:763). The increase in airflow through the glottis over modal phonation can be as much as 60% (Fischer-Jorgenson 1967, cited Klatt & Klatt 1990:823), resulting in a degree of aspiration noise accompanying voicing.

As a direct consequence of the action of the arytenoid cartilages during breathy phonation, the vocal folds close first at the front of the glottis, closure being propagated back along their length (Klatt & Klatt 1990:822). Thus the glottal volume velocity waveform is more sinusoidal in nature. This in turn has the effect of producing a strong fundamental frequency component in the frequency domain, while the higher harmonics are substantially attenuated.

Many of the attributes of breathy phonation have been observed in women's vowel production. For example, the opening in the posterior glottis has been found in a greater number of women than men (Bless *et al.* 1986, cited Klatt & Klatt 1990:826; Bless *et al.* 1989, cited Nittrouer *et al.* 1990:763).

1.2 Perceptual correlates of breathiness

Klatt & Klatt (1990:852) posited a number of acoustic cues to breathy voice. These include aspiration noise observable at higher frequencies (perceivable due to the attenuation of the higher harmonics), an increase in the amplitude of the first harmonic and a reduction in higher harmonic amplitudes, and an increase in the bandwidth of F_1. Gunzburger (1991) noted a slight lowering of F_0 when speakers of both sexes were asked to read in a 'sexy' voice, which were judged breathier than their normal reading voices. This was probably due to the slackening of the vocal folds (p65).

In a review of studies comparing the perception of breathy and nonbreathy vowels from male speakers, both synthesised and natural, Klatt & Klatt (1990:824-5) summarised the conclusions as follows: "breathy phonation is characterised by a glottal source with (1) an increased open quotient, resulting in an increased relative amplitude of the fundamental component in the spectrum and (2) a tendency for higher harmonics to be replaced by aspiration noise". Further, the studies suggest that the relative amplitude of the fundamental component is the most important cue. Using natural vowels from speakers of both sexes, Klatt & Klatt (1990:835-7) found that both of the factors listed above correlated well with breathiness perception.

However, in subjective listening tests using synthesised speech to which various cues to breathiness were added individually (see pp849-52), they found the sample perceived as most breathy was the one in which all the cues were present: "It is as if the perceiver is aware of all of the systematic changes that go into breathy phonation and uses these expectations during perception in such a way that no single cue is as effective as all in combination" (853).

Finally, the only study this author could find of perceptual listening tests comparing breathiness in female and male speakers was by Klatt & Klatt (1985:835-7). Klatt & Klatt found that, while on average women were judged slightly more breathy than men, individual breathiness was highly varied, with some males breathier than females. While this study involved the vowels of only ten female and six male speakers, it casts doubt on the claims that breathiness is a marker of speaker sex.

1.3 Previous experimental findings

Four previous studies attempting to measure perceptual correlates of breathiness are examined here, namely Henton & Bladon (1985), Klatt & Klatt (1990), Nittrouer *et al.*

(1990) and Gunzburger (1991). First, we will consider the choice of measure used in each study; this is followed by a look at their results and conclusions.

1.3.1 Choice of measure Henton & Bladon (1985:222) adopt an argument of convenience to justify the choice of an acoustic measure. As it requires no special processing (unlike, for example, measuring the amount of aspiration noise energy), they used the amplitude difference between the first two harmonics of the vowel (H1-H2). Klatt & Klatt (1990:828-9) examined this and other measures of relative amplitude (H1 relative to F_1, and H1 relative to rms amplitude) and found that, for group measures, there is little to choose between them. Because of the difficulties of measuring the others (for example, accurately estimating the amplitude of F_1) they too chose to use H1-H2. Both Nittrouer *et al.* (1990) and Gunzburger (1991) used the H1-H2 measure because Henton & Bladon and Klatt & Klatt used it.

It therefore seemed pertinent, for comparison purposes, to follow their lead and use the amplitude difference between the first two harmonic components for this large-scale study.

1.3.2 Results The results from the above-mentioned studies are summarised in Table 1. All four studies show considerable differences between female and male speakers for the measure H1-H2 (see the 'Differences' column).

Source	Language Spoken	Speaker Sex	No. of Speakers	Mean (SD) (dB)	Difference (dB)
Henton & Bladon (1985)	MD English	f m	12 13	7.6 (-) 1.6 (-)	6.0
Henton & Bladon (1985)	RP English	f m	20 16	6.4 (-) 0.8 (-)	5.6
Klatt & Klatt (1990)	US English	f m	10 6	1.9 (2.3) -3.8 (1.9)	5.7
Nittrouer et al. (1990)	US English	f m	4 4	-0.2 (3.5) -4.8 (2.1)	4.6
Gunzburger (1991)	RP Dutch	f m	13 13	3.9 (5.2) -0.6 (3.9)	4.5

Table 1 Survey of previous studies comparing measures of H1-H2 for female and male speakers. Note: MD=Modified Northern; RP=Received Pronunciation; f=female; m=male. Also, the 'Difference' column gives the difference between the female and male H1-H2 scores.

However, these results should be treated with caution. Even for speakers in the same gender group there are large differences in H1-H2 between speakers. In the Klatt & Klatt (1990) study, the mean H1-H2 scores for each female speaker range from -1.6dB to 7.1dB, and the males from -5.4dB to -0.3dB (see Table III, p829). Similarly, the female speakers in Gunzburger (1991)'s study range from -2.5dB to 14.5dB, and the males from -5.0dB to 5.0dB (estimated from Figure 2, p64). In addition to the between speaker differences, the data from Klatt & Klatt also show that while most speakers appear to have reasonably consistent levels for H1-H2, some speakers exhibit considerable variation in different phonetic contexts.

This led Klatt & Klatt to conclude that, "within each gender there is much ... variation in acoustic manifestations of breathiness, with some males being more breathy than many females. In addition, it is likely that any individual is capable of adopting a fairly wide range of speaking styles that differ in degree of breathiness [see Gunzburger (1991), who found that speakers adopting a breathier than normal, 'sexy' voice increased (made more positive) the H1-H2 measure]. Thus it is unwise to make sweeping generalisations with regard to sex typing, as well as the behaviour of particular individuals" (p852).

It is also worth noting the relatively few number of speakers used in each of these studies (see column 4 in Table 1). If such variation can be observed in population samples of this size, then an investigation of a much larger sample should enable rather more concrete conclusions to be drawn.

2. METHOD

2.1 Speakers

The speech data used for this study came from the National Institute of Standards and Technology (NIST) production of the DARPA TIMIT Acoustic-Phonetic Speech Database Training Set, held on CD-ROM. The database holds a total of 130 female and 290 male speakers from all over the USA. The speakers are divided into eight dialect regions: New England (dr1), Northern (dr2), North Midland (dr3), South Midland (dr4), Southern (dr5), New York City (dr6), Western (dr7), and those who have moved around (dr8, known as Army Brat!).

The 130 female speakers have a mean age of 29.9 years (s.d. 9.62), ranging from 21 to 85 years, and a mean height of 5ft 5.2in (s.d. 2.50), ranging from 5ft to 6ft. The 290 male speakers have a mean age of 29.6 years (s.d. 7.28), ranging from 20 to 85 years, and a mean height of 5ft 10.8in (s.d. 2.77), ranging from 5ft 2in to 6ft 8in.

2.2 Stimuli

The annotation files in the database detail phonetic transcriptions of the sentences spoken by each speaker, including the sample numbers delimiting the phonemes. Thus it was possible to extract all the target vowels. In this study, only /aa/ (TIMIT phonetic label, as in 'heart') was used. The reason for this is that only in an open vowel such as /aa/ is the first formant high enough to avoid interference with the second harmonic (Henton & Bladon 1985:223). In addition, this was the vowel used in each of the four previous studies considered above.

Two categories of stimuli were analysed for this study:

(1) All the /aa/ vowels on the database. Stimulus (1) consisted of 824 female tokens and 1785 male tokens.

(2) The vowel /aa/ in the phonetic context /d aa r/. This was taken from the sentence labelled `sa1', whose orthographic transcription is `She had your DARk suit in greasy wash water all year', where the capital letters indicate the position of the stimulus. Due to differences in pronunciation, not every speaker produced the vowel or the phonetic context, and so stimulus (2) consisted of 120 female tokens and 274 male tokens.

2.3 Acoustic analysis

These were then analysed using a discrete Fourier transform to determine the amplitudes of the first two harmonics (H1, H2), and so find the harmonic amplitude difference (H1-H2).

Only the middle portion of each vowel, amounting to approximately half the vowel's duration, was used to determine the values of the various parameters. A number of procedures were then employed to ensure that only vowels of an acceptable quality, relatively uncontaminated by the surrounding phonemes, were analysed. Cepstral analyses of the stimuli were performed to give estimates of fundamental frequency. This enabled the identification of any stimulus whose F_0 was untypical of the speaker's sex (given that women's and men's fundamental frequencies are generally found between 190-230Hz and 90-130Hz respectively), and any stimulus whose F_0 varied wildly during the duration of the vowel. In a similar way, the values of H1, H2 and H1-H2 for each stimulus were checked to see if they were untypical or varied wildly. Any unusual stimuli were examined more closely. In many cases it was deemed necessary to reduce the vowel length reported in the annotation files, and occasionally to reject stimuli as being unsuitable.

3. RESULTS

The results for all female and male speakers are presented in Tables 2 and 3.They show the means (and s.d.s) of H1, H2 and H1-H2 for stimuli (1) and (2) respectively.

Considering stimulus (1) (all the /aa/ tokens), there is a clear difference between the female and male means for H1-H2, although this is far less than the results quoted in previous studies (see Table 1). The female-male difference here is 1.2dB. A one-way analysis of variance indicates this result is highly significant ($p<0.1$).

However, the overlap between the female and male datasets for stimulus (1) is substantial. Figure 1 is a histogram showing the spread of the H1-H2 scores at 1dB intervals - the overlap is clearly visible, showing that for a great many female speakers H1-H2 is less (i.e. less positive) than for males.

Variable	FEMALE	MALE
H1	90.79 (4.52)	85.66 (4.74)
H2	95.65 (5.26)	91.77 (5.14)
H1-H2	-4.86 (3.19)	-6.10 (2.53)

Table 2 Overall means (s.d.) of 1st harmonic amplitude (H1), 2nd harmonic amplitude (H2), and harmonic amplitude difference (H1-H2) (all in dB) for both sexes for stimulus (1).

Variable	FEMALE	MALE
H1	92.48 (3.63)	87.53 (3.94)
H2	99.43 (4.09)	94.31 (4.29)
H1-H2	-6.94 (2.05)	-6.78 (2.09)

Table 3 Overall means (s.d.) of 1st harmonic amplitude (H1), 2nd harmonic amplitude (H2), and harmonic amplitude difference (H1-H2) (all in dB) for both sexes for stimulus (2).

In contrast to the results for all the tokens, the results for the /d aa r/ context (stimulus (2)) in Table 3 show very little difference. A one-way ANOVA test on this data indicates there is no significance in this result. Thus it would appear that phonetic context will affect the breathiness of a particular vowel. In addition, it shows here that, when considering the population as a whole, the effect of gender is negated.

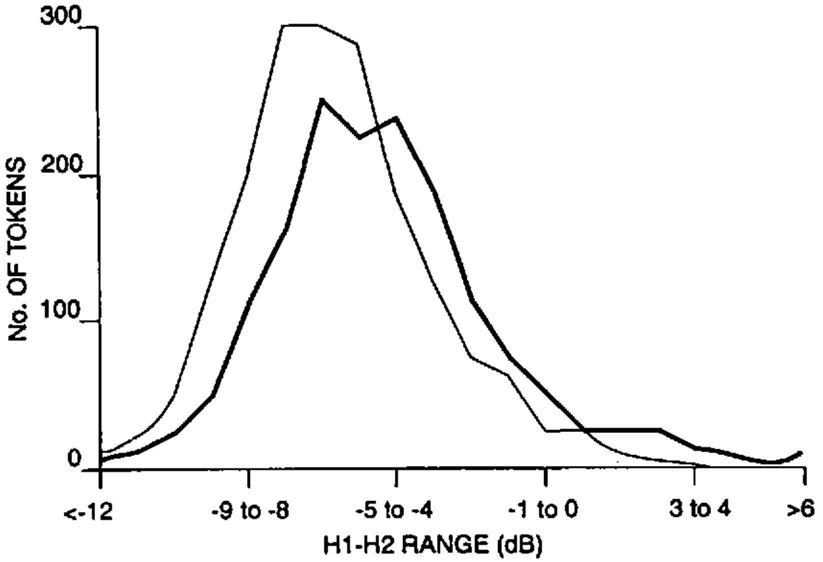

Figure 1 Histogram showing spread of data of H1-H2 scores for female (thick line) and male (thin line) speakers. (Female scores normalised against male total.)

4. CONCLUSIONS

In seeking to draw conclusions about a speaker categorisation as large as gender, it would appear only natural to select the population sample to be as diverse as possible. Thus of most importance here are the results for stimulus (1), which represents tokens from a spread of different speakers exhibiting varied speaker characteristics and in a variety of phonetic contexts. These data show that when a sample representative of the population as a whole is considered, what may appear to be a significant difference between the sexes in a small data set, is in fact merely a general tendency.

Statements to the effect that "Female speakers produce significantly more breathiness in comparison with male speakers" (Gunzburger 1991:65) carry little weight, especially when the available data (e.g. Gunzburger 1991, Klatt & Klatt 1990) clearly show wide variations in H1-H2 between speakers of a different sex. Correlation analyses from previous studies (see the review in Klatt & Klatt 1990:823-5) show that the relative amplitude of H1 is an acoustic correlate of breathiness. If then a relatively high H1 amplitude can be considered as a sufficient indicator of breathiness, the results presented here show that, while on average female speakers are slightly more breathy, a great many female speakers are in fact less breathy than male speakers. This forces the conclusion that breathiness is not a reliable marker of gender, and, as Klatt & Klatt (1990:852) say, "it is unwise to make sweeping generalisations with regard to sex typing".

It is worth noting that both the differences between the sexes for H1 and H2 are highly significant (more so than for H1-H2) for both stimulus (1) and stimulus (2). Again, these differences are highly variable (especially for H2), although an informal examination appears to indicate that H1 especially may prove to be a more reliable marker of gender.

5. REFERENCES

BLESS DM, BIEVER DM, SHAIKH A (1986) 'Comparisons of vibratory characteristics of young adult males and females.' Proceedings of International Conference on Voice, Kurume, Japan, Vol. 2, 46-54.

BLESS DM, BIEVER DM, CAMPOS G, GLAZE GE, PEPPARD R (1989) 'Videostroboscopic, acoustic, and aerodynamic analysis of voice production in normal adults.' Paper presented at the Vocal Fold Physiology Conference, Stockholm.

FISCHER-JORGENSON E (1967) 'Phonetic analysis of breathy (murmured) vowels in Gujarati.' Indian Linguistics 28, 71-139.

GUNZBURGER D (1991) 'Breathiness in male and female speakers.' 12th International Conference of Phonetic Sciences (Aix-en-Provence), Vol. 3, 62-5.

HENTON C, BLADON A (1985) 'Breathiness in normal female speech: inefficiency vs desirability.' Language & Communication 5, 221-7.

KLATT D, KLATT L (1990) 'Analysis, synthesis and perception of voice quality variations among female and male talkers.' Journal of the Acoustical Society of America 87, 820-57.

NITTROUER S, MCGOWAN R, MILENKOVIC P, BEEHLER D (1990) 'Acoustic measurements of men's and women's voices: a study of context effects and covariation.' Journal Speech and Hearing Research 33, 761-75.