DOCUMENT RESUME

ED 399 785 FL 024 110

AUTHOR	Temple, Rosalind A. M.
TITLE	Voice Source Characteristics of Male and Female Speakers of French.
PUB DATE	Mar 96
NOTE	46p.; For complete volume, see FL 024 097.
PUB TYPE	Reports - Evaluative/Feasibility (142) -- Journal Articles (080)
JOURNAL CIT	York Papers in Linguistics; v17 p397-440 Mar 1996
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	*Articulation (Speech); Comparative Analysis; Consonants; Foreign Countries; *French; Language Research; Linguistic Theory; Native Speakers; *Phonology; Pronunciation; *Sex Differences; *Speech Habits

ABSTRACT

A study investigated the realization of voicing contrasts ("breathiness") in plosive consonants produced by young French adults, particularly as they differ in males and females. Data came from acoustic analysis of recordings of nine informants reading lists of monosyllabic words with initial plosive consonants in isolation and in the content, "Jean avait dit..." The six plosive phonemes of French occurred several times before each of three vowels, but only tokens with the vowel /a/ were measured for this purpose. Results show consistent differences between males and females in the closure period of prevoiced stops. Methodological issues raised in this analysis were then examined in light of subsequent research, including measurement of spectral tilt, statistical comparability, and interrater reliability on perceptual experiments. Contains 24 references. (MSE)

VOICE SOURCE CHARACTERISTICS OF MALE AND FEMALE SPEAKERS OF FRENCH.*

Rosalind A. M. Temple

VOICE SOURCE CHARACTERISTICS OF MALE AND FEMALE SPEAKERS OF FRENCH.*

Rosalind A. M. Temple

University of York

1. Introduction

'Breathy Voice' is a phonation-type label used in phonology, in experimental phonetics and in speech pathology. 'Breathiness' is also a quality sometimes associated with females and with onsets and offsets of voiceless consonants. It is far from clear, however, what exactly are the acoustic characteristics of breathy voice, nor whether all the uses of the terms can properly be said to refer to the same phenomenon.

My purpose in the present article is to give a detailed account of part of an investigation into the realisation of the voicing contrast in plosive consonants produced by young French adults (Temple 1988a, b), which raised several questions which it was not possible to answer within the scope of that study, and to review the questions which arose at that time, in the light of subsequent literature.

2. Background to 1988 Study

2.1 The nature of 'breathiness'.

One physiological correlate of breathy voice quality is the vocal folds being held in the position for voiceless consonants, but the airflow rate is higher than normal and they vibrate loosely, 'so they appear to be simply flapping in the airstream' (Ladefoged 1982: 128), producing the breathy-voiced sound [h̤]. This occurs during the pronunciation of English intervocalic /h/, as in *ahead*. Another, more deliberate strategy is used in languages such as Gujarati, where there are phonemically contrastive breathy vowels, during which the vocal folds are held closely enough together at the front for voicing to occur, but apart at the back so that a large volume of air passes out through the glottis producing turbulence.

York Papers in Linguistics 17 (1996) 397-440
© Rosalind A. M. Temple

Bickley (1982) examined the vowels of Gujarati and !Xhóõ to determine acoustically and perceptually robust cues to the breathy-voice : modal-voice contrast. From the physiological description given in the previous paragraph one would expect an important cue to be the presence of high-amplitude inter-harmonic noise[1], and this is indeed found in the spectra of breathy sounds. However, following Ladefoged (1981) and other studies of Gujarati, Bickley wanted to investigate a cue at the other end of the spectrum, that of the relative amplitudes of the fundamental and the first harmonic above it[2]. She reanalysed Ladefoged's recordings of !Xhóõ and compared them with her own recordings of four native speakers of Gujarati. The measurements of the amplitude of the first two harmonics for the !Xhóõ speakers and one Gujarati speaker (*op. cit.:* 73-74) are reproduced as Tables 1 and 2 below. The figures show clearly that the fundamental (henceforth 'F0') is consistently higher in amplitude than the first harmonic above it (henceforth 'H2') in breathy vowels and not in clear vowels. To test the perceptual relevance of the cue, informal judgements were elicited from a native English speaker and a native Gujarati speaker, both trained in phonetics. The average amplitude differences for vowels judged to be in four categories of breathiness were as follows (the Gujarati speaker's judgements are given first): 'Very breathy' - 12.5dB, 10dB; 'Breathy' - 8.3dB, 11dB; 'Slightly breathy' - 6.7dB, 5.3dB; 'Not breathy' - 0dB, 0dB. Bickley synthesised /a/, /i/ and /u/ vowels with independent manipulation of the amplitude of the fundamental and the amount of aspiration noise, and the vowels were played to four Gujarati speakers . She found no correlation between the noise level and the degree of breathy percept, but the vowels with the highest amplitude F0 were consistently identified as breathy. Given the greater amount of noise passing through the glottis in breathy, as opposed to modal, phonation, it is surprising that the noise level did

---

[1] Noise is the acoustic consequence of the turbulent airflow which would here be escaping between the parts of the vocal folds which are not fully adducted.

[2] The relative strength of the fundamental is known to increase as open quotient (the proportion of the vibratory cycle during which the vocal folds are open) increases. Increased open quotient is a known articulatory correlate of breathy voice quality.

not have a greater effect on the breathy percept, but this may be because of problems with synthesis.

	Difference (in dB)	
	Breathy	Clear
Speaker 1	13	0
Speaker 2	-4	-3
Speaker 3	2	-3
Speaker 4	5	-4
Speaker 5	5	-9
Speaker 6	4	-8
Speaker 7	11	0
Speaker 8	9	-2
Speaker 9	15	-2
Speaker 10	10	2

Table 1. Difference between amplitudes of first and second harmonics for breathy and clear vowels in !Xhóõ. After Bickley 1982: 73)

	Amplitudes in dB		
	first harmonic	second harmonic	difference
bar	44	42	2
maro	46	42	4
wali	47	43	4
bar	42	44	-2
maro	43	43	0
wali	38	44	-6

Table 2. Relative amplitudes (in dB) of first and second harmonics for breathy (top) and clear (bottom) vowels in Gujarati. (After Bickley 1982: 74)

"Breathiness" has also been much studied in a clinical context, sometimes being explicitly compared to the quality which is given the same label in other contexts. Hammarberg quotes a famous line of

Ladefoged's: '... what is a pathological voice in one language may be phonologically contrastive in another.' (Ladefoged 1983) and extends it to: 'What is evaluated as an abnormal voice quality in one language or dialect community may be a socially acceptable voice quality in another.' (Hammarberg, *op. cit.* 27) A particular spectral shape which is entirely attributable to physiological problems could thus be interpreted by speakers to convey a sociolinguistic message. Laver (1980) has exemplified how modes of phonation can be 'signals of emotional status' (Hammarberg, *op. cit.* 27) and Hammarberg's example is particularly pertinent to the present study, as we shall see in 2.2 below:

> 'For instance, breathiness is said to be a common female vocal attribute in many social communities, whereas creakiness often is a male characteristic.' (*ibid.* 27)

Hammarberg (1986) brings together a series of studies where pathological voices were judged by pathologists and phoniatrists against a series of voice quality parameters. The voices judged as breathy were all from patients with unilateral vocal-fold paralysis[3]. Acoustic analyses were made using long-term average spectra, and the typical long-term spectral characteristics of these voices were the high level of the fundamental, a low spectral level in the F1 region (400 to 600 Hz) and a high level of amplitude in the highest frequency band (5 to 10 kHz).

2.2 Female-male voice source differences

2.2.1 Acoustic evidence

The vocal folds of mature males are on average fifty per cent longer than those of females, and are thicker and greater in mass (Ohala, 1983). One natural result of this is that male fundamental frequency (F0) is lower than that of females[4]. As well as causing the perceived pitch of the

---

[3] Unilateral paralysis, and other deformations of the vocal folds, such as nodules, can impede complete closure during phonation, producing the same effect as in the normal speakers' production of breathy voice described above.

[4] Average values given by Fant (1956: 11, cited in Laver, 1983: 15) are 120 Hz for males and 220 Hz for females.

male and female voices to be different, this difference in F0 means that the harmonics are more widely spaced and interact in a different way with vocal tract resonances[5]. Moreover, the shape of the female source waveform is more symmetrical than for males, and this is reflected in the amplitudes of equivalent harmonics, which decline more steeply in the case of the females. Monsen and Engebretson (1977) asked subjects to phonate into a long, reflectionless metal tube, which significantly reduced the resonances of the vocal tract and enabled them to analyse the glottal waveform. The waveform shape was found to be much more symmetrical for females than for males, with the opening and closing phases occupying almost equal proportions of the period. The male waveform had a characteristic 'hump' in the opening phase with the closing phase taking only twenty to forty per cent of the total period. These differences are reflected in the spectra with the slope in dB per octave between the harmonics being much steeper in the female glottal wave. The characteristics are not entirely surprising when one considers the physiology of the vocal folds: because of their greater mass, the males' vocal folds are drawn together faster than the females' by the Bernoulli effect, giving a sharper closure onset. Their larger size also results in the upper and lower parts being somewhat out of phase, which would create an effectively longer closure period. The waveform produced would thus be irregular in shape with enhanced harmonics above the fundamental. The female vocal folds, on the other hand, are drawn together less sharply, but with a smoother motion, and acting more as a single mass, which would produce a smoother, more sinusoidal waveform with the fundamental much stronger than the rest of the harmonics. Monsen and Engebretson's harmonic-by-harmonic comparison of glottal spectra in normal phonation (cf. Figure 1) reveals this difference in slope, but when the spectra are plotted un-normalised on the same frequency and amplitude axes, i.e. with the female signal about an octave higher in F0 than the male signal and with an overall intensity level -4 to -6 dB lower, the actual spectral envelopes are seen to be almost identical (cf. Fig 1b). There thus appears to be some sort of built-in normalisation factor for this particular spectral effect.

---

[5] The vocal tract resonances themselves are also different.

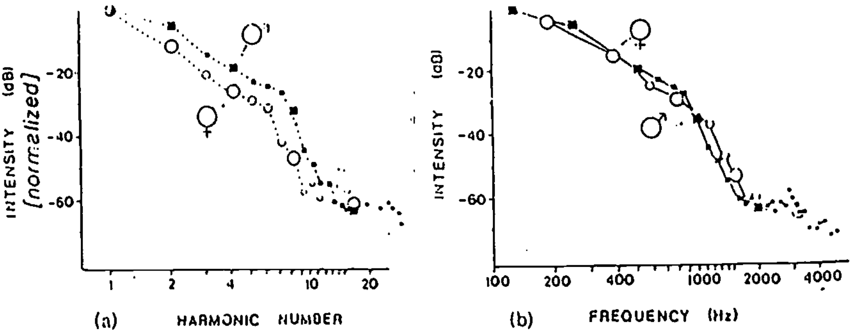

Figure 1. Average glottal spectra for male versus female normal voice phonation: (a) spectra normalised for both frequency and intensity; (b) non-normalised spectra. Male subjects, solid squares; female subjects, solid circles. (From Monsen and Engebretson 1977: 987)

It is interesting to note that when Bickley subjected steady-state vowels to inverse filtering to remove the effects of sound radiation and vocal tract filtering from the signal, her observations of the glottal waveforms produced in breathy and modal vowels corresponded closely to those observed by Monsen and Engebretson for female and male glottal waveforms respectively: 'The glottal waveforms of the clear vowels exhibited slower opening than closing phases, abrupt closure, and a closed phase that occupied approximately one third of the period of vibration.... The glottal waveforms for breathy vowels exhibited similar opening and closing phases, resulting in a more symmetrical shape. Closure was less abrupt and the closed interval was shorter.' (Bickley 1982: 76-77)

Other studies by those concerned with the synthesis of female-sounding speech confirm Monsen and Engebretson's findings concerning source differences. Klatt (1986) analysed the speech of a single female speaker with a 'pleasant voice quality'. He found considerable random breathiness noise above 2kHz over parts of many utterances and a variable degree of general tilt of the spectrum (i.e. over a larger frequency range than the F0-H2 measure) and of the strength of the fundamental. He attributes this variation to the presumed degree to which the larynx is spread or constricted.

2.2.2 Perceptual evidence

Barry (1986) reviewed some of the literature on male-female voice source differences and also concluded that they had much to do with physiology. His own study sought to make good synthetic copies of a male and female voice, to derive from these a set of tables that would reproduce the voice quality (using a rule-synthesis algorithm on the parallel-formant synthesiser developed by Holmes), and then to establish transformations which could be applied to one set of tables to derive the other. The acoustic features modified were F0, formant frequencies, spectral tilt and noise. In manipulating spectral tilt, Barry found that the best match was obtained by reducing the amplitude of the second formant (A2) by 6dB relative to the male A2, and of the third and fourth formants by 8dB. The male voice was generated without aperiodicity in the source signal (although there had been some present in the human subject) and this did not seem to make it sound unnatural. A 'good match' female voice included 25% noise. A discrimination test was carried out where listeners were played pairs of utterances and asked to select the one which sounded more like an adult female. The utterances most consistently judged as female were those where the formant frequencies <u>and</u> amplitudes and the spectral noise level of the original 'male' synthetic voice had been modified. It proved impossible to adjudicate between the relative importance of formant amplitude (and hence spectral tilt) adjustments and the degree of spectral noise. Thus, Barry's perceptual findings confirmed the importance of the production phenomena discussed in 2.2.1 above in the perception of a voice as "female".

2.2.3 Sociolinguistic claims

It would seem from the evidence just reviewed that the common claim that breathiness is a female attribute is predictable on the grounds that the physiology of female vocal folds gives rise to acoustic structures which are known to cue both a breathy and a "female" percept. However, the variability in degree of tilt found by Klatt suggests that although physiology (a constant for a given speaker) plays a significant role, voice source characteristics can be varied by manipulating the larynx constriction.[6] It is known from investigations into other acoustic phenomena that physiologically-predictable characteristics can be endowed with sociolinguistic significance by speakers and exaggerated or compensated for. For example, Mattingly (1966) tested the hypothesis that formant frequency differences between speakers of the same dialect were chiefly due to variations in the vocal tract size of the speakers, using data from Peterson and Barney's seminal 1952 study of vowels[7]. If the hypothesis were correct, Mattingly argued, there should be high correlation scores between the distributions of values for F1, F2 and F3 for the three classes of speaker (men, women and children). What he found in all but a few subsets of the data was that the correlation scores were in fact very low, and that the separation between male and female distributions of formants for some vowels was far sharper than could be explained by vocal tract size variation. He concluded:

> '... the difference between male and female formant values, though doubtless related to typical male and female vocal tract size, is probably a linguistic convention.'

Further evidence for the linguistic conventionalisation of cues to speaker identity which originated as physiological differences comes from work on children's speech before the development at puberty of physical vocal tract differences, since at the earlier stages there would be no physiological reason to account for sex-specific differences. Sachs (1975), for example, played children's' productions of /a/, /i/ and /u/ vowels to a panel of listeners, and asked them to identify the sex of the speaker. She obtained a statistically significant correct response rate of 66%, which suggests that the children (who were aged between 4 and 12) were beginning to produce sex-specific formant patterns despite the fact that the boys' and girls' vocal tracts would still be similar in size.

---

[6] If this were not possible, it would not be possible for female speakers of Gujarati and other languages, where breathy voice is used distinctively, to make the necessary distinctions. We shall return to this issue below.

[7] Peterson, G. E. & H. L. Barney (1952) Control methods used in a study of the vowels. *Journal of the Acoustical Society of America* 24: 175-184

Vowel	/a/	/ɑ/	/ʌ/	/ɒ/
Females	8.4	6.4	6.2	3.3
Males	0.98	0.77	0.16	0.39
Difference (F-M)	7.42	5.63	6.04	2.91

Table 3. Average differences in amplitude in dB between the first and second harmonics in male and female speakers of Received Pronunciation. (After Henton and Bladon 1985: 224)

Henton and Bladon (1985) did not consider the physiological basis of source spectrum differences corresponding to breathiness, but they did examine the male-female differences as a sociolinguistically determined sex-specific marker. They followed Bickley (1982)[8] and measured the amplitude of F0 and H2 in the steady-state portions of open vowels produced by male and female RP and 'Modified Northern' speakers. Their results for the RP speakers are reproduced in Table 3. The male-female differences were significant according to a *t*-test ($p<0.01$) and the difference across all the vowels (mean of means) was 5.5dB. As Henton and Bladon point out (*op. cit.* 225), the differences 'would be sufficient to carry the perceptual contrast between breathy and modal vowels' for Bickley's Gujarati speakers; however, when their measurements are compared with the values of the synthetic vowels played in Bickley's perceptual experiment, it would appear that only /a/ would be considered as more than 'slightly breathy' by either of Bickley's phoneticians (compare Table 3 with the values given on p.2 above).

Interestingly, when Watson (1987) asked colleagues to listen to his child-subject's voices, they did not perceive them as breathy until the possibility was pointed out to them:

> 'It may be that we accept as normal in children what would be 'breathy' in adults, until we are specifically

---

[8] It should perhaps be noted that speaker sex was not specified by Bickley, but it is assumed, because of the consistency of her results, that her speakers were all male.

> called on as phoneticians to attend to phonation type.' (*ibid.* 21)

The comment could easily be applied *mutatis mutandis* to sex-specific differences in breathiness: might it not be the case that breathiness is a comparative measure to be assessed against the cultural norm for modal voice, and therefore cannot be measured in universal terms? Alternatively, it could be that although we are dealing with measures along the same acoustic continuum, it is unjustified to speak of what is being labelled as breathiness as being classifiable as exactly the same phenomenon in both the case of females (and children) and that of a linguistic phonation type. If there were no difference, Gujarati women would have great difficulty in producing phonologically breathy sounds which were sufficiently different from sounds phonated with their modal voice.

Henton and Bladon would presumably not consider these questions to be problematic, as they see the spectral tilt[9] characteristics as being produced deliberately by the British female speakers, rather than as being a result of physiology, and would presumably hypothesise that female modal voice would not have the same culturally determined properties in Gujarati. On the premise that breathy voice is used to convey intimacy in English (Laver 1980: 135) they suggest that the RP. speakers are trying to sound 'sexy' [*sic*]:

> 'At an ethological level, breathy voice may be seen as part of the courtship display ritual, as important as bodily adornment and gesture. A breathy woman can be regarded as using her paralinguistic tools to maximise the chances of her achieving her goals, linguistic or otherwise.' (*op. cit.* 226).

---

[9] Hitherto the term 'tilt' has been used in its generally accepted designation of the rate of decrease in amplitude across the whole source spectrum; I shall also be using the term in this article to refer to the difference in amplitude between F0 and H2. I make no claims as to the equivalence of these two measures, using the term in refer to this amplitude difference.

The claim that the female RP voice has the distinctive spectral characteristics described solely with the paralinguistic aim of aiding the speaker to attract a mate seems rather exaggerated, especially in the light of the other papers discussed above which hold that the female source spectrum would tend towards the 'breathiness' pattern anyway for physiological reasons. However, this does not rule out the role of other sociolinguistic forces which could cause female speakers to move nearer to or further away from the physiologically determined female 'norm', which is the implication of the findings of Mattingly, cited above. It should also be pointed out, of course, that males may well be modifying their voice quality for similar reasons.

2.3 Breathiness and the Voicing Contrast

As is well-known, French, like English, has a two-way 'voicing contrast' between cognate pairs of obstruents, but as far as plosives are concerned, the labels *'Voiced'* and *'Voiceless'*[10] correspond to different phonetic patterns of realisation in the two languages, most obviously in the timing of vocal-fold vibration relative to the release of the consonant when in absolute initial position. The *Voiced* plosives of French are canonically voiced throughout the closure and release period, usually with no break (though see Temple 1988a, b); *Voiceless* plosives have no prevoicing and little or no aspiration. English *Voiced* stops are phonetically voiceless unaspirated, while the *Voiceless* ones are voiceless and with longer aspiration following release. In addition to the timing of voicing relative to the release of the consonant, there are many other phonetic correlates to the voicing contrast in French and English plosives which are well-documented elsewhere and which it is not necessary to review here (see Temple 1988a for references). One correlate which has been less thoroughly documented, although it is taken to be a well-known fact about at least English plosives, is that *Voiceless* plosives tend to have breathy voice at vowel onset, due to the

---

[10] The labels *Voiced* and *Voiceless*, in italics and with initial capital letters will be used throughout this paper to refer to phonological categories. The same words in non-italic script, and entirely in lower-case will be used to refer to the phonetic distinction between stops with prevoicing and those without. Henceforth no citation marks will be used.

vocal folds' beginning to vibrate before being fully adducted for the vowel. Ní Chasaide and Gobl (1988) reported an analogous process during the pre-aspiration of plosives in Swedish. Laryngographic traces showed vibration of the vocal folds as they opened for the *Voiceless* plosive, and this was accompanied by an increase in spectral tilt. However, they also found that the onset of voicing in post-consonantal vowels was much less 'clean' than the breathy offset of the pre-consonantal vowel.

The evidence reviewed thus far shows that F0-H2 differences have been found to correlate with perceived "breathiness" in languages where this quality plays a phonological role. The same measure has been found to differentiate male and female voice sources, and this is to some extent predictable from male-female physiological differences. Moreover, it has been suggested that variability in this measure could have a sociolinguistic value. Temple 1988a and 1988b thus attempted to draw together whether degree of breathiness, measured by the F0-H2 difference, was yet another marker of the voicing contrast in initial position, and whether there were differences between male and female French speakers, and if so, whether there was interaction between sex-specific and voicing-specific effects.

3. The 1988 Study

3.1 Methodology

Seven speakers were recorded in their study bedrooms at the École Normale Supérieure in Paris, and two at Oxford University Phonetics Laboratory (O.U.P.L.), reading lists of monosyllabic words with initial plosive consonants in isolation and in the frame, 'Jean avait dit ____'. The stimuli were presented individually on cards to minimise listing effects, and the first element of each list was discounted. The six plosive phonemes of French occurred several times before each of three vowels, /i/, /a/ and /u/. Only tokens with the vowel /a/ were measured for this part of the experiment because it is in here that the lower harmonics are least likely to be affected by the first formant, either in transition or in steady-state. The data were analysed using the Signal File Manager of O.U.P.L.'s New England Digital microcomputer (see Clark 1986 for details). Windows were positioned at the points indicated by the letters A to E and V in Figure 2, that is, in the relatively steady-state parts of the pre-voicing and the vowel, over the release itself and over the pitch periods closest to the release. The two frames which fall into this latter category were at varying distances in milliseconds from the release: B covered the last three pitch periods of prevoicing for females and the last two for males, including cases of *Voiced* stops which were partially devoiced (i.e. where voicing ceased before release); and D covered the first three and first two periods after release in both *Voiced* and *Voiceless* stops, the latter having varying Voice Onset Times. The frame lengths of 20ms and 16ms for males and females respectively were chosen after experimenting to find settings which would give the best resolution of harmonics whilst maintaining comparable lengths in both time and number of periods. For each frame, frequency in Hz and amplitude in decibels (dB) of F0 and H2 were noted.[11]

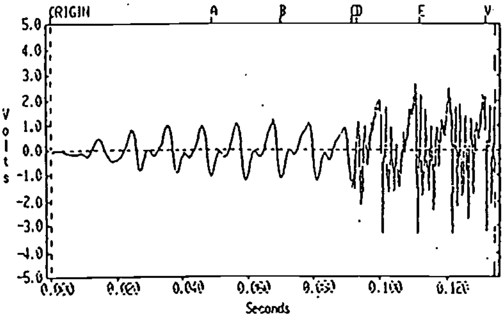

Figure 2. Positions of start of spectral windows for utterance "bac", by speaker PIG (male)

---

[11] For more details on the analytical procedure followed, see Temple 1988a: 57-70.

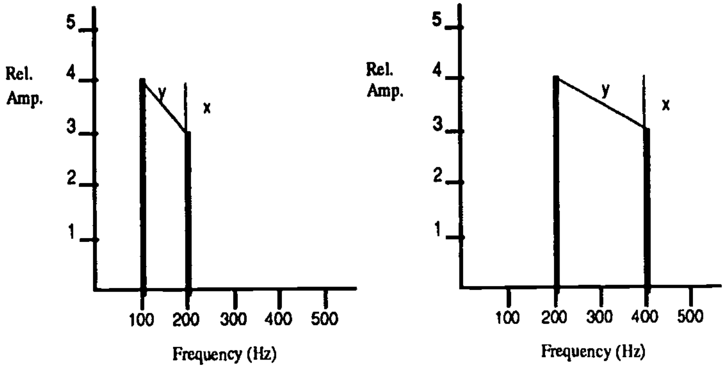

Figure 3. Schematic representation of the effects of fundamental frequency on the relations between harmonics in the spectrum.

A technical problem arises here in the question of how to compute what we have been referring to 'spectral tilt'. Both Bickley and Henton and Bladon calculated the straightforward difference between the amplitude measurements of the harmonics. Assuming that all Bickley's subjects were male, it is unimportant whether the measure is computed in this way or whether a true slope is calculated in amplitude loss per frequency unit (difference in dB 'over' difference in Hz). However, as soon as speakers with notably different F0 are to be compared, the choice of calculation method becomes important, since a higher F0 means a greater distance in Hz between F0 and H2, which would have a significant effect on the calculation of the slope. A schematic example is given in Figure 3 to illustrate this effect. The horizontal axis represents frequency in Hz, the vertical axis a hypothetical amplitude range. The solid vertical lines correspond to idealised harmonics for a male versus a female speaker. The difference in amplitude between F0 and H2 in both pseudo-spectra is 1. However, if the slopes are calculated in Amplitude/Frequency the results are 1/100 = 0.01 'A'/Hz for spectrum M, but 1/200 = 0.005 'A'/Hz for spectrum F. As well as having implications for comparisons across studies, this has implications for comparisons within a single study wherever speakers have significantly different fundamental frequencies. Indeed, spectra with a different amplitude difference could actually have the same slope gradient: if the difference in 'A' in spectrum M were 10, and in spectrum F 20, the gradients would be 10/100 = 0.1 'A'/Hz, and 20/200 = 0.1 'A'/Hz respectively. The question of which is the best way of measuring spectral 'tilt' is evidently potentially important and we shall return to it below. For the purposes of the experiment being described here it was decided to compute the measure both in terms of amplitude differences and in terms of dB/Hz slope.

Statistical analysis of the measurements was carried out using S.A.S.[12] Institute package implemented on the VAX mainframe computer at Oxford University Computing Service. The data were subjected to a 'General Linear Models' (G.L.M.) procedure, which allows Analysis of Variance to be carried out on 'unbalanced' models, because the numbers of tokens analysable for each speaker were not the same, principally because of the hazards of making recordings outside the recording booth.

3.2 Results and discussion.

3.2.1 Waveforms

No procedures were used to derive the source waveform from the vowel signal, but the waveforms during the closure period of prevoiced stops did appear consistently differently in male versus female subjects. Generally the waveform shapes in the speakers considered here seemed to be as predicted by Monsen and Engebretson, that is with a near-sinusoidal appearance for females, but with a 'hump' in the opening phase and a sharper closing phase for males (compare Figures 2 and 4).

---

[12] Statistical Analysis System.

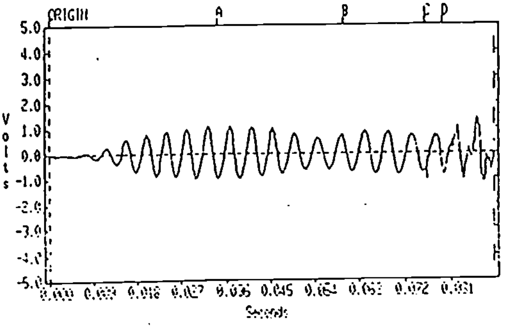

Figure 4. Waveform in prevoicing of "bac" by speaker ISR (female).

3.2.2 Relationship between F0 and H2: male versus female speakers

Position Sex		A	B	D	E	V
Males	dB/ Hz	-0.0378	-0.0813	-0.0262	-0.0093	-0.0183
Females		-0.0491$	-0.104$	-0.0492$	-0.0398$	-0.0396$
Males	dB	-5.026	-6.213	-3.758	-1.346	-0.404
Females		-15.853$	-18.330$	-10.642$	-8.920$	-9.504$

Table 4. Mean F0-H2 differences for frames positioned at A, B, D, E & V by male and female speakers expressed in terms of slope (dB/Hz) and amplitude (dB)

Mean values for the differences between F0 and H2 at the different positions in the word are given in Table 4 and Figure 5 in terms both of the dB/Hz slope and of amplitude comparisons in dB. A negative number indicates that the value for the fundamental is higher than for the second harmonic, and a positive number represents a lower value for F0. Another convention adopted has been to indicate the steeper gradient slope or greater amplitude difference in a particular two-way comparison with a superscript dollar sign ($^{\$}$). All the values in the table are higher for females than for males, as predicted from the evidence discussed hitherto, and the male-female contrast is high

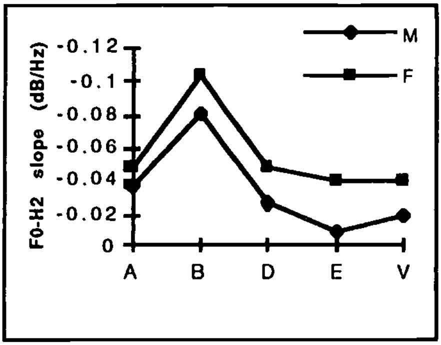

Figure 5a. Mean F0-H2 slope (dB/Hz) across positions of all tokens for males and females.

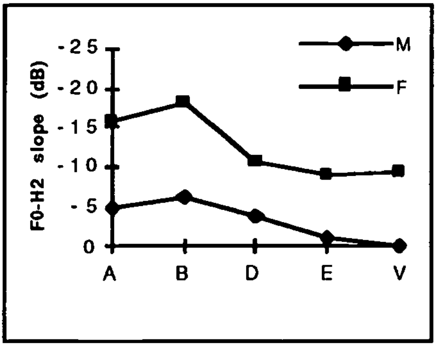

Figure 5b. Mean F0-H2 differences of amplitude across positions of all tokens for males and females.

significant according to a t-test ($p<0.001$) in all cases except V for the dB/Hz measure, which fails to reach significance even at the 5% level. It is clear from Figure 5a that the male and female trends in terms of slope stay firmly apart but follow much the same pattern with a sharp rise in steepness at B, that is as the release approaches or the prevoicing is about to cease. However, this effect is apparently reduced dramatically, particularly for females, in Figure 5b, where both curves are much smoother, showing only a slight rise in the dB difference at B. Also apparent in this Figure is the reflection of how the male-female difference at V 'becomes' statistically significant when calculated in terms of amplitude.

These findings are interesting for two particular reasons. Firstly, the only position where a significant difference was not found is the only one where measurements were taken in the other experiments reported, i.e. the relatively steady-state portion of the vowel. Secondly, they seem to confirm that changing the method of calculating the 'value' of the harmonic difference does have a significant effect on the apparent relationships between the sets of production data, which in turn suggests it could be relevant perceptually. Moreover, the measure which fails to reach statistical significance in this position is not the one used in the papers cited above, which begs the question 'how would those results look when calculated in these terms?'

3.2.3 Possible influence of consonant place of articulation

The steady-state part of the vowel was chosen by the other researchers referred to in order to avoid the possible effects of the F1 transition from the preceding or following consonant, which could enhance the amplitude of F0 or H2 and thereby distort the results. However, because the focus of this study was on the voicing contrast in consonants, these transition sections were precisely the parts of the signal in which we were interested. The only way to counteract the influence of formants would have been inverse filtering, which it was not possible to carry out at the time. Instead statistics were used to compare the effects of the different places of articulation of the consonants on the spectral values. Of course, the use of statistics cannot be seen as a replacement of inverse filtering by an equivalent measure, but we can hope that it would at least make us aware of any significant effect of components which would have been filtered out by that process. The slope values obtained for males and females are given in Table 5, and the amplitude-difference values in Table 6. Values are given for each position for each phoneme, and accompanying each value, an indication of those phonemes which are significantly different from the one in question, at the 5% level (t-test).

Position		A		B	
Consonant		*Mean*	*Diff From*	*Mean*	*Diff From*
/b/	*m*	-0.04348	g	-0.05464	
	f	-0.09521$^{\$}$	g	-0.12022$^{\$}$	g
	bth	-0.06430	g	-0.08147	g
/d/	*m*	-0.04975	g	-0.05778	
	f	-0.09092$^{\$}$	g	-0.10448$^{\$}$	
	bth	-0.06606	g	-0.07637	g
/g/	*m*	-0.01971	b d	-0.03248	
	f	-0.06282$^{\$}$	b d	-0.08860	b
	bth	-0.03781	b d	-0.05479	
/p/	*m*				
	f				
	bth				
/t/	*m*				
	f				
	bth				
/k/	*m*				
	f				
	bth				

Table 5(a). Mean slope differences (dB/Hz) across place of articulation for the different sexes with indications of pair-wise contrasts significant at the 5% level (t- test). Positions A and B as in Figure 2.

Position		**D**		**E**	
Consonant		*Mean*	*Diff From*	*Mean*	*Diff From*
/b/	*m*	-0.01503	d	-0.01150	t
	f	-0.03099$^{\$}$	g k	-0.04675$^{\$}$	
	bth	-0.02134	d g k	-0.02545	
/d/	*m*	-0.04283	b g t	-0.01776	t
	f	-0.05104$^{\$}$		-0.03575$^{\$}$	
	bth	-0.04614	b t	-0.02502	
/g/	*m*	-0.02125	d	-0.01418	t
	f	-0.06695$^{\$}$	b	-0.03492$^{\$}$	
	bth	-0.03871	b	-0.02210	
/p/	*m*	-0.02941		-0.01638	t
	f	-0.04457$^{\$}$		-0.03897$^{\$}$	
	bth	-0.03611	b	-0.02637	
/t/	*m*	-0.01626	d	-0.00897	p b d g
	f	-0.04713$^{\$}$		-0.04004$^{\$}$	
	bth	-0.03056	d	-0.01375	
/k/	*m*	-0.03092		-0.00690	
	f	-0.05593$^{\$}$	b	-0.04233	
	bth	-0.04182	b	-0.02234	

Table 5(b). Mean slope differences (dB/Hz) across place of articulation for the different sexes with indications of pair-wise contrasts significant at the 5% level (t- test). Positions D and E as in Figure 2.

Position		V	
Consonant		*Mean*	*Diff From*
/b/	*m*	+0.01744	p
	f	-0.04594$^{\$}$	
	bth	-0.00763	
/d/	*m*	-0.00461	
	f	-0.03259$^{\$}$	
	bth	-0.01591	
/g/	*m*	-0.05037$^{\$}$	
	f	-0.02796	
	bth	-0.04181	
/p/	*m*	-0.07412$^{\$}$	b
	f	-0.03895	
	bth	-0.05857	
/t/	*m*	-0.00814	
	f	-0.04275$^{\$}$	
	bth	-0.01544	
/k/	*m*	-0.01151	
	f	-0.04672	g
	bth	-0.02685	

Table 5(c). Mean slope differences (dB/Hz) across place of articulation for the different sexes with indications of pair-wise contrasts significant at the 5% level (t- test). Position V as in Figure 2.

Position		A		B	
Consonant		*Mean*	*Diff From*	*Mean*	*Diff From*
/b/	*m*	-5.546	g	-7.044	
	f	-17.160$^{\$}$	g	-20.989$^{\$}$	g
	bth	-10.218	g	-12.749	g
/d/	*m*	-6.328	g	-7.268	g
	f	-16.435$^{\$}$	g	-18.754$^{\$}$	g
	bth	-10.331	g	-11.840	g
/g/	*m*	-2.762	b d	-4.040	d
	f	-11.682$^{\$}$	b d	-14.903$^{\$}$	b d
	bth	-6.506	b d	-8.359	b d
/p/	*m*				
	f				
	bth				
/t/	*m*				
	f				
	bth				
/k/					

Table 6(a). Mean amplitude differences (dB) across place of articulation for the different sexes with indications of pair-wise contrasts significant at the 5% level (t- test). Positions A and B as in Figure 2.

Position		**D**		**E**	
Consonant		*Mean*	*Diff From*	*Mean*	*Diff From*
/b/	*m*	-2.362	d	-1.444	
	f	-7.236$^{\$}$	k	-9.158$^{\$}$	
	bth	-4.290	ptkd	-4.496	
/d/	*m*	-5.048	b	-2.466	t
	f	-10.343$^{\$}$		-7.309$^{\$}$	
	bth	-7.185	b	-4.421	
/g/	*m*	-2.896		-1.084	
	f	-11.085$^{\$}$		-7.255$^{\$}$	
	bth	-6.025		-3.442	
/p/	*m*	-4.586		-1.703	
	f	-10.339$^{\$}$		-9.467$^{\$}$	
	bth	-7.131	b	-5.137	
/t/	*m*	-2.846		-0.220	d
	f	-11.377$^{\$}$		-9.674$^{\$}$	
	bth	-6.799	b	-4.601	
/k/		-4.682		-1.168	
		-12.750$^{\$}$	b	-10.077$^{\$}$	
		-8.197	b	-5.050	

Table 6(b). Mean amplitude differences (dB) across place of articulation for the different sexes with indications of pair-wise contrasts significant at the 5% level (t- test). Positions D and E as in Figure 2.

Position		**V**	
Consonant		*Mean*	*Diff From*
/b/	*m*	-0.093	
	f	$-10.317^{\$}$	
	bth	-4.137	
/d/	*m*	-0.797	
	f	$-7.573^{\$}$	
	bth	-3.532	k
/g/	*m*	-0.680	
	f	$-6.797^{\$}$	k
	bth	-3.017	k
/p/	*m*	+0.579	
	f	$-9.561^{\$}$	
	bth	-3.906	
/t/	*m*	-0.117	
	f	$-10.426^{\$}$	
	bth	-4.769	
/k/		-1.593	
		$-11.607^{\$}$	
		-5.955	

Table 6(c). Mean amplitude differences (dB) across place of articulation for the different sexes with indications of pair-wise contrasts significant at the 5% level (t- test). Position V as in Figure 2.

measure, but we can hope that it would at least make us aware of any significant effect of components which would have been filtered out by that process. The slope values obtained for males and females are given in Table 5, and the amplitude-difference values in Table 6. Values are given for each position for each phoneme, and accompanying each value, an indication of those phonemes which are significantly different from the one in question, at the 5% level (t-test).

Again, the dB/Hz slopes for females are consistently steeper than the males' slopes across all positions except at V for /g/ and /p/. The picture becomes more interesting when these values are compared with the dB values. For /p/, the male H2 is seen to be higher than F0. For /g/, both the measures show F0 generally higher than H2, but whereas the dB difference is greater for females than for males, with the other measure the result is the opposite. An extension of the hypothetical example above shows that this is mathematically unsurprising: with differences in 'A' of 10 in spectrum M, and of 20 in spectrum F, we saw that the gradients would be the same; however, a reduction of just one 'A' unit would give an apparently steeper slope for spectrum M, even though the amplitude difference would still be greater in spectrum F: 10/100 = 0.1 'A'/Hz; 19/200 = 0.095 'A'/Hz. Moreover, bringing the amplitude difference in spectrum F down to, say, 13 would still leave it greater than the difference for M, but in the slope would be 0.06 'A'/Hz, only just over half as steep as the male counterpart.

There are further differences between the two tables in terms of which pair-wise contrasts between phonemes show a significant difference. To take the values for the prevoicing first, although the 'Diff From' columns for measurements at position A are identical, there are discrepancies in the same column for position B, where, for example, /d/ enters into no significant contrasts for the dB/Hz measure, but contrasts with /g/ for all groups of speakers for the dB measure. With or without these discrepancies, these pair-wise contrasts also indicate that a caveat needs to be added to our suggestion above that the waveform of the prevoicing was the closest we were likely to get to the glottal source waveform. They show (not surprisingly) that the supralaryngeal characteristics of the consonants do affect the pre-voicing F0-H2 tilt. There are still large differences between males and females, but it could be argued that since place of articulation obviously does have an effect on the slope, the differences in the lower spectral components could be accounted for by supra-glottal differences, rather than differences generated by the vocal folds themselves. In view of the findings of the literature reviewed earlier, it is improbable that the male-female spectral differences found can be entirely ascribed to supra-glottal effects, but there was no possibility of testing the extent of those effects within the framework of this study.

In the post-release positions, the numbers of pairs of phonemes with significant differences between them decreases in both tables from D through E to V, but again different pair-wise contrasts were found to be significant in the different tables. It is clear too that the formant transitions do have an effect on the slope, and one is again forced to question whether the highly significant male-female differences found at D and E (as opposed to the failure to attain significance at V in the dB/Hz measure) were not at least enhanced by supraglottal resonance differences between the males and females. The effect of F1 would be reduced by the time it had passed through the frequency band where it would affect H2, hence the reduced inter-phoneme differences through E to V. If H2 is being enhanced, that would reduce the difference between it and F0, thus masking the characteristics of the 'breathy' spectrum. That there still is at least some male-female difference at V is encouraging for our original hypothesis that there is an effect independent of formant differences. However, this should be confirmed by examining the possible influence of the different F1 values of the vowels themselves. Actual measurements of the formant frequencies were not carried out, but a statistical analysis of possible vowel effects was done.

3.2.4 Possible effect of following vowel

Henton and Bladon (*op. cit.*) restricted their study to the English vowels /a/, /ɑ/, /ʌ/ and /ɒ/ in order to try and minimise the interference of F1 (which is relatively high in these vowels) with F0 or H2. The results comparing vowel-contexts for the present data in dB/Hz are given in Table 7 and Figure 6. Unfortunately the full set of statistics for the dB measure is not available, so in the light of the differences noted in the previous paragraph, the following comments, which are based on the dB/Hz values, should be taken with a note of caution.

Position		A		B	
Vowel		*Mean*	*Diff From*	*Mean*	*Diff From*
/i/	*m*	-0.04733	–	-0.04956	–
	f	-0.19102$^{\$}$	–	-0.10871$^{\$}$	–
	bth	-0.06517	e	-0.07338	e
/e/	*m*	-0.00096	–	-0.00485	–
	f	-0.06823$^{\$}$	–	-0.07630$^{\$}$	–
	bth	-0.02787	i u	-0.03036	i a u
/a/	*m*	-0.04061	–	-0.04887	–
	f	-0.07567$^{\$}$	–	-0.09836$^{\$}$	–
	bth	-0.05492		-0.06985	e
/u/	*m*	-0.03751	–	-0.05564	–
	f	-0.08800$^{\$}$	–	-0.11258$^{\$}$	–
	bth	-0.05738	e	-0.07758	e

Table 7(a). Mean slope values (dB/Hz) showing effects of different following vowels at positions A, and B across the sexes and indications of pair-wise contrasts significant at the 5% level (t-test figures for both groups only).

Position		**D**		**E**	
Vowel		*Mean*	*Diff From*	*Mean*	*Diff From*
/i/	*m*	-0.03399	–	-0.00966	–
	f	-0.08181$^{\$}$	–	-0.06914$^{\$}$	–
	bth	-0.05418	e a	-0.03477	e a u
/e/	*m*	+0.00739	–	+0.01403	–
	f	-0.07696$^{\$}$	–	-0.00605$^{\$}$	–
	bth	-0.02424	i u	+0.00650	i u
/a/	*m*	-0.01902$^{\$}$	–	-0.01444$^{\$}$	–
	f	+0.00132	–	-0.00339	–
	bth	-0.01028	i u	-0.00969	i u
/u/	*m*	-0.03485	–	-0.00576	–
	f	-0.07782$^{\$}$	–	-0.05574$^{\$}$	–
	bth	-0.05290	e a	-0.02675	i e a

Table 7(b). Mean slope values (dB/Hz) showing effects of different following vowels at positions D, and E across the sexes and indications of pair-wise contrasts significant at the 5% level (t-test figures for both groups only).

Position		V	
Vowel		*Mean*	*Diff From*
/i/	*m*	-0.03901	–
	f	-0.06499$^{\$}$	–
	bth	-0.04997	a
/e/	*m*	+0.00408$^{\$}$	–
	f	+0.02486	–
	bth	+0.01187	
/a/	*m*	-0.00409	–
	f	-0.01133$^{\$}$	–
	bth	-0.00766	
/u/	*m*	-0.02060	–
	f	-0.05256	–
	bth	-0.03402	a

Table 7(c). Mean slope values (dB/Hz) showing effects of different following vowels at position V across the sexes and indications of pair-wise contrasts significant at the 5% level (t-test figures for both groups only).

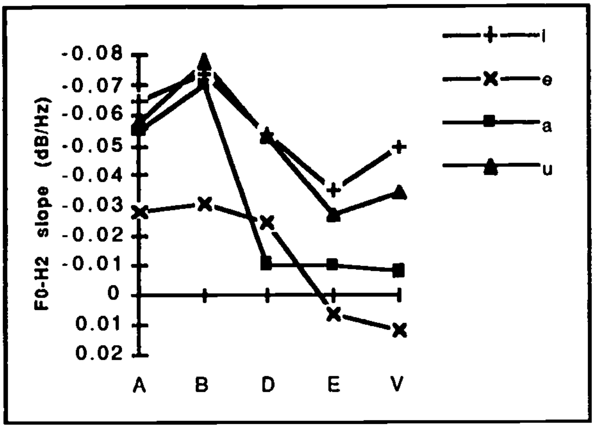

Figure 6. Slope differences as a function of following vowel. All speakers.

In Figure 6 the patterns for the four vowels when all speakers are taken together have a somewhat similar trajectory. Apart from /e/, there is a striking degree of similarity before the release, suggesting relatively little coarticulatory effect on this part of the spectrum in prevoicing. The atypical pattern for /e/ can be explained by the lack of tokens following either /b/ or /d/. There are large post-release differences and an inspection of the values for males and females separately (cf Table 7) shows that there is a complex effect, which is not surprising when one considers the complex sex-specific differences found in the acoustic structure of vowels. The female slope is again generally steeper. However, in /a/, where following previous studies we had expected to see the hypothesis confirmed most firmly, the male-female position is reversed after the release through D and E, and the only mean value for females to be a positive value (indicating H2 higher than F0) is at D for /a/ (although the male-female difference fails to reach significance at either D or E). At V there is a return to the more common pattern of females having the steeper mean slope, although this difference fails to reach significance by a long way (p>0.05). Clearly more detailed analysis of the interaction of slope and formant frequency is needed.

3.2.5. The Voicing contrast

It was suggested above that the F0-H2 difference may be found to vary following voiced versus voiceless consonants as an indicator of increased breathiness in the voiceless case. Values for the Voiced versus Voiceless classes as wholes are given in Table 8. None of the differences in slope between Voiced and Voiceless reaches significance. The greatest differences tend to occur in the vocalic portion, which is again where we should least expect to find them. The cross-phoneme comparisons shown in Tables 5 and 6 above revealed hardly any significant differences between cognate pairs, so these values are not surprising and no positive conclusions can be drawn from then concerning the discrimination of phonological classes.

	Position	**D**	**E**	**V**
Sex	**Voicing**			
m	*Voiced*	$-0.02731^{\$}$	$-0.01466^{\$}$	-0.01206
	Vless	-0.02509	-0.00415	$-0.04039^{\$}$
f	*Voiced*	$-0.04945^{\$}$	-0.03898	-0.03542
	Vless	-0.03579	$-0.02040^{\$}$	$-0.03263^{\$}$

Table 8. Mean values for F0-H2 slope (in dB/Hz) across Voicing categories for males and females at post-release positions.

If, as suggested above, this is not an effect manipulated by speakers but one due more to the physical effects of the gradual adduction of the vocal folds, we should expect the de-voiced tokens to follow the pattern of the *Voiceless* ones. Means were therefore computed across phonetic voicing type and are presented in Figures 7 to 9 and Table 9. Two graphs are given for the data for the male speakers and for the data for all speakers considered together because of the drastic effect of the mean V value for the O-PREV tokens. The categories represented are fully-voiced tokens (FVOICED); *Voiceless* tokens (PHON VLESS); *Voiced* tokens where prevoicing ceased at some time at or before release (DEVOICED); *Voiced* tokens with no actual prevoicing (0 PREV).

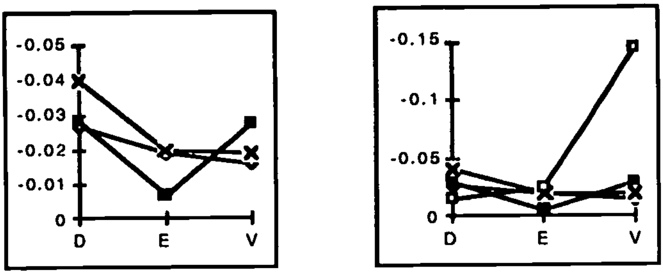

Figure 7. Slope values across positions for voicing type. Male speakers. Including (b), and not including (a), 0-prevoiced Voiced *tokens.*

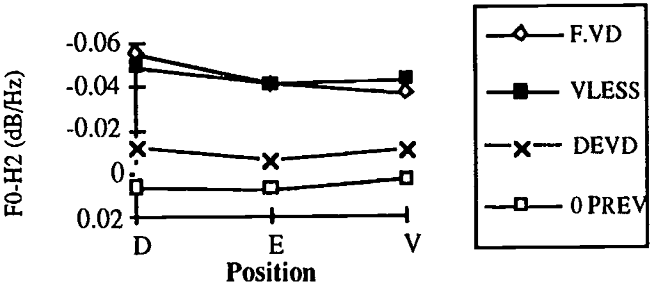

Figure 8. Slope values across all positions for voicing type. Female speakers.

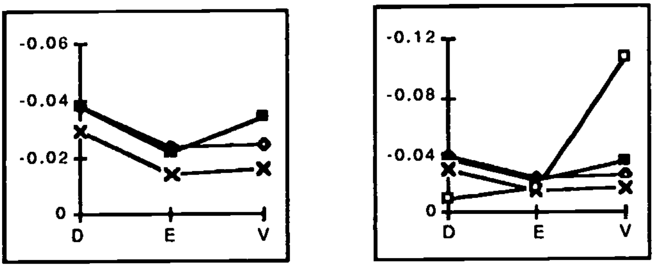

Figure 9. Slope values across positions for voicing type. Male speakers. Including (b), and not including (a), 0-prevoiced Voiced *tokens.*

Position		D	E	V
Sex	Voicing type			
	voiced	-0.02697	-0.01186	-0.01625
m	*Voiceless*	-0.02829	-0.00615	-0.02821
	devoiced	-0.03975	-0.01986	-0.01898
	Vd -- no prev	-0.01354	-0.02455	-0.14358
	voiced	-0.05509	-0.04092	-0.03706
f	*Voiceless*	-0.04896	-0.04039	-0.04275
	devoiced	-0.01121	-0.00546	-0.01030
	Vd -- no prev	+0.00628	+0.00687	-0.00290
	voiced	-0.03826	-0.02353	-0.02461
both	*Voiceless*	-0.03755	-0.02150	-0.03473
	devoiced	-0.02965	-0.01478	-0.01592
	Vd -- no prev	-0.00858	-0.01670	-0.10695

Table 9. Mean values for F0-H2 slope (in dB/Hz) across voicing categories for males and females.

When the effect of the male 0-PREV tokens is disregarded, the patterns for the different voicing types across the spectral window positions are very similar. There are no significant differences between types for males or for the group as a whole, but for females the FVOICED and the VLESS are significantly different from the DEVOICED and 0-PREV types, as reflected in Figure 9. With regard to the voicing contrast, therefore, there seems to be no phonetic or phonological grouping for which this measure of breathiness is a robust acoustic correlate.

4. Studies published since 1988

A good deal of work has been published since 1988 on the nature of voice source characteristics. We shall restrict ourselves here to a description of just a small number of important studies.

The most substantial single study is that of Klatt and Klatt (1990) on the analysis, synthesis and perception of voice quality variation. Klatt and Klatt analysed recordings of ten female and six male speakers uttering two 'real' sentences and reiterant imitations of those sentences using [ʔɑ] and [hɑ] syllables and measured the relative strength of the first harmonic, the presence of noise in the F3 region and above, and the presence of extra poles and zeros in the vowel spectrum, mid-way through the vowel. They found an average male-female difference of about 5.7dB in F0-H2 difference, but there was considerable subject-to-subject variability within each group, with average F0-H2 across sentences ranging from 8.4 to 17.1dB in females, and from 4.6 to 9.7 in males. Periodicity versus noise excitation of F3 was measured for the reiterant sentences with [hɑ], on a subjective five-point scale and noise was found to be commonly present for both sexes with on average more noise in female than male subjects, but again considerable within-group variation. Both reiterant imitations of one of the original sentences pronounced by all subjects were then played to a panel of eight listeners, who were asked to judge the vowels on a seven-point scale from 'not breathy' to 'strongly breathy'. On average, females were perceived to be slightly more breathy than males, and sentences consisting of [hɑ] syllables were generally perceived as considerably more breathy than those with [ʔɑ]. Correlations of breathiness ratings with acoustic measures suggested that both the F0-H2 measure and the presence of noise were important. Finally, pairs of synthetic 'female' vowels (the first of each pair being a constant reference vowel) were played to a panel of five listeners who were asked to judge the relative breathiness of the second, its naturalness and its nasality. The results suggested that noise amplitude was more important than F0-H2 difference in giving a breathy percept; the latter cue was insufficient on its own to induce a breathy percept and often contributed to a perceived increase in nasality. The tentative conclusion of the authors is that,

> '... either breathiness is signalled differently for men and women, or that the increases in the first harmonic observed in production data from women must be accompanied by other cues to be interpreted by the listener as cues to breathiness.' (851)

Ní Chasaide and Gobl have published several papers developing the theme of the 1988 presentation mentioned above, among them one in *Speech Communication* (Gobl and Ní Chasaide 1992) where they analysed repetitions of a prose passage read with a range of voice qualities by a male phonetician who is a native speaker of British English. The data were subjected to manual interactive inverse filtering and analysed using the four-parameter LF-model of differentiated glottal flow developed by Gunnar Fant. Correlates of breathy voice were found to be high values for the parameters RA (corresponding to attenuation of higher frequencies), RK (corresponding to a more symmetrical pulse shape) and OQ (Open Quotient, thus also suggesting a more symmetrical pulse). Gobl and Ní Chasaide also used data from frequency domain analysis of the speech waveform to measure the levels of F1 and F2 relative to the first harmonic (our F0) and their Figure 5 (487) shows marked attenuation of both in the breathy data. An important feature to note about both sets of measurements is that they vary over time, and in their conclusion the authors emphasise the point that, 'a switch between voice qualities may not necessarily involve a single transformation which remains uniform throughout an utterance.'

Ní Chasaide and Gobl (1993) investigated voice quality in the vicinity of *Voiced* and *Voiceless* stop consonants spoken by male and female speakers in different languages. They found considerable cross-linguistic differences, but the effects were not grouped according to language-family as they had expected. Thus Swedish and, to a somewhat lesser extent, Italian /pː/ was preceded by a markedly higher RA than /b/, whereas, although the values were occasionally slightly higher in French and German (suggesting a slight tendency to relax the vocal folds in anticipation of the following *Voiceless* stop), the effect was not found to be consistent. The English speakers produced both patterns, but information is not given as to whether the division corresponds to the speaker's sex. RK values also rose in Swedish in anticipation of /pː/. Spectral measurements on the whole confirmed these findings, with the voicing category of the following consonant having little differential effect on F1 (their L1) relative to F0 in French and German, but showing a marked relative decline in F1 before the Swedish /pː/ with a rather lesser effect in the same direction in Italian. The English subjects fell into two groups, as for the source parameter measures. It is noticeable that for both sets of measures, the Figures show some marked differences between the languages, even within one of the two groupings (i.e. those with a /_p/ - /_b/ difference and those without).

In postconsonantal vowels, little categorial effect was found in the source parameters in French and Italian, but German RA was much higher at vowel onset following /p/ than /b/, and declined less rapidly. The authors infer that this is the result of incomplete glottal closure with the vocal folds vibrating in breathy mode following the aspirated stop. However, the difference between voicing categories is less marked in Swedish and English, despite the fact that these languages also have a voiceless unaspirated vs. voiceless aspirated phonetic contrast. The spectral data show less similarity between Swedish and the two Romance languages, with a lower F1 in Swedish post /p/ onset than following /b/, but no consistent effect in French or Italian. German follows a similar pattern to Swedish, but with an even greater relative lowering of F1. Data for English are not given. In the light of these findings, it is perhaps not surprising that no difference was found in the study reported above for vowels following voice versus voiceless stops in French.

A smaller-scale study is currently being carried out by Scobbie (1995 and personal communication), in which he found a marked difference between F0-H2 measures in vowel onset following /t/ vs. /d/, and to a lesser extent /p/ vs. /b/ in four-year-old speech-disordered child speakers of Edinburgh English.

5. Discussion

The 1988 study reported above raised several issues, to which we shall now return in the light of the subsequent work reported above.

5.1 Methodology

There are various methodological questions raised by a comparison of the studies mentioned, principal among which are how the oft-referred-to, but ill-defined feature 'spectral tilt' or 'spectral slope' is measured, and how measurements are analysed.

5.1.1 The measurement of spectral tilt.

The studies take one of two approaches to gaining access to an accurate measure of the voice source. Some invoke some procedure for negating the effects of the supra-glottal filter. Thus, Fant and Ní Chasaide and

Gobl used inverse filtering techniques, whereas Monsen and Engebretson had their subjects phonate down a reflectionless tube to reduce the resonances of the vocal tract. Bickley also used inverse filtering when she was looking at waveforms. The rest rely for the most part on analysing vowels with a relatively high F1 to minimise its effect on the lower harmonics, and/or on averaging large amounts of data to derive an accurate picture of the shape of the source spectrum. Henton and Bladon and Temple use statistical tests, while Hammarberg uses Long-Term Average Spectra (LTAS). Of course, with either approach it is impossible to be absolutely sure that a true picture of the glottal wave has been revealed, although inverse filtering techniques have improved greatly over recent years. The second type of approach seems the less satisfactory one, particularly for the purposes of comparing across studies, or even comparing different groups of speakers within studies: it is well-known that vowel qualities differ somewhat across languages (thus /a/ could represent something different in Gujarati from French), and across sex groups (and that the degree of sex-specific variation varies from language to language - see Bladon *et al* 1984)[13]. The fact that the trajectory for /a/ from position D to V in Figure 5 (above) is different from those of the other three vowels does suggest that we might be able to claim that the F1 transition is not affecting H2 in this case, but the uncomfortable fact remains that it is only this vowel which shows the unexpectedly steeper male slope in two positions. Moreover, Table 7 shows that only in a few measurements were the slope measurements for /a/ seen to be significantly different from those for the other vowels, where F1 is likely to have had an effect.

The actual measure of spectral tilt also differed from study to study. Fant and Ní Chasaide and Gobl used the LF model of glottal flow developed by the former, and measured parameters assumed to correspond to characteristics of the glottal wave. Because Hammarberg used LTAS, she was unable to make detailed measurements of spectral features, and instead identified breathy voice quality with relatively low energy in the F1 region (400-600Hz) and high levels in the highest

---

[13] It could also be the case that /a̤/ and /a/ in Gujarati do not have the same formant values.

frequency band (5-10kHz). Monsen and Engebretson measured slope in the first two octaves of their spectra in terms of dB fall-off per octave. Others measure formants, but in different ways: Barry compared amplitude levels for the same formant in his female and male subjects, while Gobl and Ní Chasaide measured F1 and F2 relative to F0. The rest of the studies measured harmonics, and I shall return to them in the next paragraph. The point needs to be made, however, that while these different measures allow generalised comparisons to be made of greater or less spectral tilt, the kind of detailed comparisons made, for example, between Henton and Bladon's data and that of Bickley is not possible.

The studies using F0-H2 all measured the difference in amplitude between the two harmonics in dB. As we have seen, comparison using this measure between speakers with the same F0 is unproblematic (which is not to say that the interpretation of comparisons is without problems), but as soon as speakers with different F0 are compared, the analyst is faced with a choice which has implications for the results and can affect their statistical significance. Tables 10 and 11 present recalculations of Bickley's and Henton and Bladon's figures to see how this might affect the comparison between their sets of data.

	Difference (in dB/Hz)	
	Breathy	Clear
Speaker 1	0.1182	0
Speaker 2	-0.0364	-0.0273
Speaker 3	0.0182	-0.0273
Speaker 4	0.0455	-0.0364
Speaker 5	0.0455	-0.0818
Speaker 6	0.0364	-0.0727
Speaker 7	0.1	0
Speaker 8	0.0818	-0.0182
Speaker 9	0.1364	-0.0182
Speaker 10	0.0909	0.0182

Table 10. Slope between first and second harmonics for breathy and clear vowels (in dB/Hz) in !Xhóõ. Calculated from figures given in Table 1 above, assuming F0 to be 110 Hz.

Since the frequency data were not available, hypothetical values of 110 Hz for male speakers and 220 Hz for females were assumed. Moreover, only mean amplitude differences are available for Henton and Bladon's data.The Tables are intended to give an idea of how a different method of calculation might affect the comparison between them, rather than a mathematically precise reformulation of the data.

Vowel	/a/	/ɑ/	/ʌ/	//
Females	0.0382	0.0291	0.0282	0.0150
Males	0.0089	0.0070	0.0015	0.0036

Table 11. Average slope (in dB/Hz) between the first and second harmonics in male and female speakers of Received Pronunciation. Calculated from figures given in Table 3 above, assuming F0 to be 220 Hz for female speakers and 110 Hz for male

Table 11 shows a clear difference still between the male and female RP speakers and the female slopes are still steeper than the !Xhóõ clear vowels. However, whereas the F0-H2 amplitude difference for the RP females' /a/, /ɑ/ and /ʌ/ was greater than for six of the !Xhóõ breathy vowels, it is only greater than two in the dB/Hz measure (with /a/ alone being greater than one other in addition). Moreover, if the RP female /a/ measurement is compared with, for example, !Xhóõ speaker 10, the ratio is 0.84 on the dB measure, but only 0.42 on the slope measure. More significantly, the recalculation changes the relationship of the measurements of the RP speakers with the evaluations of Bickley's phoneticians. The recalculated average amplitude differences for vowels judged to be in the four categories of breathiness (see p.4 above for dB figures) are as follows: 'Very breathy' - 0.1136 dB/Hz, 0.0909 dB/Hz; 'Breathy' - 0.0755 dB/Hz, 0.1 dB/Hz; 'Slightly breathy' - 0.0609 dB/Hz, 0.0482; 'Not breathy' 0dB/Hz, 0 dB/hZ. When these values are compared with the RP females, the latter are seen not even to reach the 'Slightly breathy' level. It is the case that many of the Gujarati and !Xhóõ vowels also do not reach that level in either measure, and it must be remembered that the phoneticians were asked to judge degree of breathiness rather than whether the vowels were breathy or not, and that these are average values. Nevertheless, these calculations show that there are potential problems for comparative statements which remain to be resolved.

It is evident that further experiments are needed to test whether the straightforward amplitude difference between successive harmonics, or the 'slope' between them is perceptually salient. The evidence reviewed in the present article provides little basis for deciding between the measures, but Monsen and Engebretson's suggestion that there is some sort of built-in normalisation factor in the differing slopes (see Fig. 1 and comments in section 2.2 above) would imply that maybe it is the slope which is important. Figure 1(b) shows the near-identity of the spectral envelopes in un-normalised spectra: it is not the amplitude difference alone between each pair of harmonics which allows this to happen, but the combined effect of that and the distance between them in frequency.

5.1.2 The use of statistics

Many of the studies discussed, use statistical analyses of the data. This not only poses problems of comparability between studies because of the different numbers of subjects studied, but also those studies which present only statistical comparisons of groups of speakers risk masking variability within each group. Dempster (1992) illustrates this dramatically with an analysis of F0-H2 differences in two contexts in the large DARPA TIMIT Acoustic-Phonetic Speech Database Training Set, a database containing material from 420 speakers of U.S. English. Whilst one might want to take issue with aspects of Dempster's study, his evidence for the dangers of relying on statistics for drawing conclusions is salutary: he found a statistically significant difference ($p<0.1$) between male and female F0-H2 differences for the vowel /aa/[14] (measured in dB), but when the data are presented in histogram form, a very large degree of overlap is apparent.

While it is right, as Dempster says, that we should heed Klatt and Klatt's warning that, 'it is unwise to make sweeping generalisations with regard to sex typing' (*op. cit* 852), this does not invalidate or preclude further exploration of some of the questions raised in the

---

[14] TIMIT phonetic label representing the vowel in *heart* etc.

present paper concerning the undoubtedly strong sex-specific tendency found in the work reviewed.

5.1.3 Perceptual experiments

All the perceptual experiments reported involve trained phoneticians. The answers thus tell us whether phoneticians judge the voice qualities according to a linear scale of 'breathiness' which they have learned. This does not really tease out the different contributing factors or enable us to make much progress with one of the central questions, that is whether the findings discussed above are addressing something which can really be construed as the same phenomenon in the real world. For example, does F0-H2 difference contribute to the perception of [a] versus [a̤] for the ordinary, untrained speaker of Gujarati?

That the judgements elicited tend to be on a scale of breathiness is also worthy of comment. When breathiness is being examined as a possible correlate of maleness or femaleness, or of degree of severity of a pathological condition, the justification for the approach is evident, but in an investigation of the acoustic correlates of phonological categories its relevance is less clear (compare, for example, the fact that English native speakers do not tend to hear absolute initial prevoiced French stops as 'very voiced'; when students of French are asked to attend to prevoicing, they often perceive a preconsonantal nasal element.)

5.2 Are we all talking about the same thing?

Perhaps the most important question, and one which needs to be considered before further detailed investigations of some of the problems highlighted in this paper are carried out, is whether we are not being mislead by applying a single label to a variety of phenomena which are different in some respects. There is common ground between all the studies discussed, but they are looking at spectral tilt as a marker of breathiness in four different contexts:

1. as indicative of male-female physiological differences (e.g. Monsen and Engebretson);

2. as indicative of breathy voice quality for sociolinguistic or paralinguistic effect (e.g. Henton and Bladon);
3. as a characteristic of phonological categories (e.g. Bickley);
4. as indicative of a pathological problem (e.g. Hammarberg).

Is it justifiable to extend Ladefoged's 1983 statement quoted earlier to apply to the studies reviewed here? That is, is it really reasonable to claim that the 'breathiness' of pathological subjects or Gujarati speakers' [a̤] vowels, rather than a tendency for the difference between F0 and H2 to be greater, is characteristic of female speech? Barry's finding that noise in the high-frequency regions of the spectrum was as important for generating a 'good match' female voice suggests that it may be, and indeed the vibratory pattern suggested by Monsen and Engebretson for female vocal folds would predict that more noise would be generated than by males, as well as females having an enhanced fundamental. But this does not guarantee that the relative 'amounts' of noise and tilt are the same in all the cases. If, as Klatt and Klatt claim, noise is more important than tilt for giving a breathy percept, then maybe the F0-H2 differences found by Henton and Bladon are not indicative of breathiness at all.

In addition, the physiological correlates of the acoustic phenomena are reported or hypothesised to be different in the different cases: Ladefoged (see page 2 above) describes different correlates for breathiness in Gujarati vowels and English voiced /h/, the former a deliberate configuration of the vocal folds, and the latter a passive effect; Hammarberg posits incomplete abduction of the vocal folds as a result of unilateral paralysis or nodules on the folds; and Monsen and Engebretson ascribe the greater spectral tilt and noise to the different vibratory patterns of the vocal folds in males and females, which are in turn caused by differences in mass and structure. There is no reason why the relationship between production settings and acoustic structure has to be one-to-one, but it cannot be taken for granted that the different settings will necessarily produce something which can be called the same.

REFERENCES

Barry, M. 1986. Synthesising female voice quality: parameters and test methods. *Cambridge Papers in Phonetics and Experimental Linguistics* 5.

Bladon, R. A. W., Henton, C. G. and Pickering, J. B. 1984. Outline of an auditory theory of speaker normalization. In van den Broecke, M. P. R. and Cohen, M. (eds) *Proceedings of the Xth International Congress of Phonetic Sciences* 313-317.

Bickley, C. 1982. Acoustic analysis and perception of breathy vowels. *MIT Working Papers in Speech Communication* 1. 71-81.

Ní Chasaide, A. and Gobl, C. 1988. Voicing contrasts and the voice source. Paper delivered at the British Association of Academic Phoneticians Colloquium, Dublin.

Ní Chasaide, A. and Gobl, C. 1993. Contextual variation of the vowel voice source as a function of adjacent consonants. *Language and Speech* 36. 303-330.

Clark, C. J. 1986. Description of a speech analysis system. *Progress Reports from Oxford PHonetics* 1. 13-25.

Dempster, G. J. 1992. Acoustic cues to breathiness: a true marker of gender? *Proceedings of the Institute of Acoustics.* Vol. 14, Part 6. 249-256.

Gobl, C. and Ní Chasaide, A. 1992. Acoustic characteristics of voice quality. *Speech Communication* 11. 481-490.

Hammarberg, B. 1986. *Perceptual and Acoustic Analysis of Dysphonia.* Dissertation from Dept of Logopedics and Phoniatrics, Huddinge University Hospital, Sweden.

Henton, C. G. and Bladon, R. A. W. 1985. Breathiness in normal female speech: inefficiency versus desirability. *Language and Communication* 5. 221-227.

Klatt, D. H. 1986. Detailed spectral analysis of a female voice. *Journal of the Acoustical Society of America* 80, Supplement 1. S69.

Klatt, D. H. and 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.

Ladefoged, P. 1981. The relative nature of voice quality. *Journal of the Acoustical Society of America* 69, Supplement 1. DD3.

Ladefoged, P. 1982. *A Course in Phonetics*. New York: Harcourt, Brace Jovanovich.

Ladefoged, P. 1983. Linguistic uses of different phonation types. In Bless, D. M. and Abbs, J. H. (eds). *Vocal Fold Physiology*. San Diego: College Hill Press. 351-360.

Laver, J. 1980. *The Phonetic Description of Voice Quality*. Cambridge: C.U.P.

Mattingly, I. G. 1966. Speaker variation and vocal tract size. *Journal of the Acoustical Society of America* 39. 1219.

Monsen, R. and Engebretson, A. M. 1977. Study of variations in the male and female glottal wave. *Journal of the Acoustical Society of America*. 62. 981-993.

Ohala, J. J. 1983. Cross-language use of pitch: an ethological view. *Phonetica*. 40. 1-18.

Sachs, J. P. 1975. Cues to the identification of sex in children's speech. In Thorne, B. and Henley, N. (eds) *Language and Sex: difference and dominance*. 152-171.

Scobbie, J. M. 1995. Phonological and phonetic perspectives on the delayed and disordered acquisition of English initial consonant clusters. Paper presented at the Third Phonology Workshop, North-West Centre for Romance Linguistics, Manchester, May, 1995.

Temple, R. A. M. 1988a. *Sex-specific Aspects of the Voicing Contrast in French Stop Consonants*. Oxford M.Phil. thesis.

Temple, R. A. M. 1988b. In search of sex-specific differences in the voicing of French stop consonants. *Progress Reports from Oxford PHonetics* 3. 74-99.

Watson, I. M. C. 1987. Problems in quantifying the child voicing contrast. Ms, University of Oxford.

FL024097 - FL024111