COMPARING THE ACOUSTIC PROPERTIES OF NORMAL AND SHOUTED SPEECH: A STUDY IN FORENSIC PHONETICS.

Jennifer Elliott
Phonetics Laboratory, Linguistics Program
School of Language Studies, Faculty of Arts
The Australian National University

ABSTRACT
Forensic phoneticians are able to exercise little control over the data they are required to examine and compare. When two speech samples, one from a criminal and one from a suspect, are provided for forensic analysis, it is quite possible that one sample may contain shouted speech, while the other will contain normally spoken speech. Analysing these dissimilar speech samples requires an understanding of how the acoustic properties of shouted speech differ from normal speech. This paper reports the findings of a pilot study which investigates the similarities and differences between the acoustic properties of natural speech in both normal and shouted modes. Results of the experiment indicate that F_0 and F_1 may be significantly higher in shouted speech, but there is no evidence for a significant difference in F-pattern for the other formants.

1. BACKGROUND

Forensic phonetics is concerned with comparing two or more samples of speech in order to determine the likelihood that they were spoken by the same person, or, conversely, to determine the likelihood that they were spoken by different people. The process of speaker identification for forensic purposes is based on the premise that there are more differences in speech between speakers than there are differences within the speech of one person. Therefore a sample of speech from a suspect of a crime (recorded, for example, during a police interview) is likely to contain more similarities to the voice of the criminal recorded at the scene of the crime if they are the same person. On the other hand, it is likely that the speech samples will contain more distinctive differences if they were not uttered by the same person. Whilst there is mounting evidence to support this thesis, there is only limited knowledge as to precisely what the parameters of similarity and difference are.

Speaker identification for forensic purposes must take into account a complex of variables. Hollien (1990: 190-191) for example, notes that these can include the non-contemporaneous nature of the recordings, distortions due to the recording systems used, and variations which occur within the speech of any individual due to a multiplicity of factors such as changes in emotional state or health, or even deliberate disguising of the voice. Nolan (1983) proposes a model of the factors that contribute to and effect variation within a speaker's speech signal output. These include a speaker's linguistic and vocal mechanisms, their communicative intent, as well as the other indexical factors which a speaker reveals non-volitionally, such as sex, social background and psychological state. This model suggests, therefore, that one can expect the communicative intent underlying normal speech to be different from that of shouted speech, with concomitant changes to the linguistic mechanism, particularly in terms of tone of voice. In turn, this will be constrained by other indexical factors, such as physique, sex, state of health, and so on. Before the forensic phonetician can distinguish between two different speech samples spoken with differing communicative intent, the limits of within-speaker variation under the presenting conditions must be known. In the case of normal speech versus shouting, this requires a knowledge of the acoustic differences between these modes of speech. Previously little research has been undertaken on the acoustic differences between normal and shouted speech, and this paper offers one attempt to address this deficiency from a forensic point of view, by presenting the findings of a pilot study on within- and between-speaker differences in normal and shouted speech of the high front vowel /i:/.

Other than observing an increase in fundamental frequency (F_0), the focus of previous research on shouting has been largely directed towards speech production and articulatory phonetics, rather than the acoustic properties of shouting. For example Laver (1980: 148-150) discusses 'loud' voice in terms of tension settings of the vocal tract, while Hacki (1996) talks about shouting in terms of vocal intensity and relative pitch range resulting from an increase in the sound

pressure level. One previous study which does focus on an acoustic analysis of loud speech was a small-scale study by Harris & Weiss (1964) which examined the effects of loud speech on pitch, using a group of 14 subjects in a laboratory setting. This same study also examined briefly F-pattern variations between normal and loud speech for one speaker.

The present study was commenced with two principal hypotheses: (H1) that there is a difference between normal speech and shouted speech in a person's average fundamental frequency; and (H2) there is a difference between normal and shouted speech in a person's average F-pattern. To maintain a forensic focus on between-speaker differences, two further hypotheses were considered: (H3) that there is a difference in the F_0 between speakers, regardless of whether they speak normally or shout; and (H4) there is a difference in F-pattern between speakers, regardless of whether they speak normally or shout.

2. THE EXPERIMENT

Forensic phoneticians are required to deal with data which can be thought of as "natural" speech, rather than data which is generated in the laboratory and therefore is highly controlled. For this reason, it was decided that data for this study should also reflect as far as possible the kind of speech found in natural discourse. Because this is experimental work, however, there has been an attempt to control for some of the complex variables noted in Nolan's model which can affect the speech signal, although of course this would not be possible in a 'real life' forensic context.

Since this was a pilot study, only two speakers were used. (Subsequent studies should, of course, involve a much larger sample of speakers.) Both were male speakers of General Australian English (Mitchell & Delbridge (1965), Burridge & Mulder (1998)), were between 15 and 16 years of age, came from similar socio-economic backgrounds and attended the same school. In order to minimise the possible effects of convergence in linguistic styles during the interaction (as suggested by Communication Accommodation Theory (CAT) (Giles & Coupland (1991: 60-93) and Giles Coupland & Coupland (1991)), both participants were also friends. It should be noted at this juncture that while convergence may not be relevant in the forensic context, it was considered important to the present study in order to control as far as possible any changes in the speaker's phonetic/phonemic repertoire as the result of accommodating towards the other speaker.

An additional limitation of this study is that only one vowel phoneme was measured: the high front long vowel /i/. This vowel was chosen because it characteristically shows less variation between speakers than other vowels, and therefore was considered to provide a more rigorous test. Furthermore, the F-pattern for this phoneme can be readily recognised in spectrograms, with a low F_1 and higher F_2. For a more comprehensive picture of the within- and between-speaker variation between normal and shouted speech, any future experiment should consider a range of Australian vowel phonemes.

The data for normal speech was collected with a map task. Both participants were given maps which were identical with the exception that only one map contained names for the various geographical features. The speaker with this map was asked to explain to his partner the location and names of all the geographical features which he in turn was required to mark on his map. The participants were instructed not to look at each other's map so that all information had to be communicated verbally. The map features were chosen to elicit a number of /i:/ vowel segments from generally similar phonemic environments (Table 1). At the end of the labelling task, the roles of the participants were reversed so that the participant who had been asked to label his map was

/bi:(C)/	/pi:(C)/:
/bi:/ (from 'BP')	/pi:/ (from 'Pee Wee', 'BP')
/bi:t/ (from 'beetle')	/pi:t/ (from 'Peter's', & 'peat')
/bi:tʃ/ (from 'beach')	
/bi:n/ (from 'bean')	

Table 1: Phonemic segments used in the analysis

asked to guide the other participant on a "journey" around the map. This ensured that enough tokens of the required phonetic segments were elicited from both participants to provide adequate data for analysis.

A second, similar, map task was devised using a different map for the elicitation of shouted speech, however the names used for the various geographical features were similar to those in the first map. This time each participant was also given a set of headphones to wear through which music of their choice was played. The music was played at a level which was loud enough to require the participants to shout at each other in order to be heard, but was not so loud that it interfered with their individual auditory feedback, nor was it so loud as cause discomfort.

Both interactions were recorded using a PZM microphone and a Sony TCM-5000EV cassette recorder. The data from both recordings was transcribed and the /i:/ segments required for analysis were identified. These were then digitised at a sampling rate of 16,000 Herz using CSL Model 4300B speech analysis equipment. Wideband (264 Hz) spectrograms were generated for the data under analysis, along with Fast Fourier Transform (FFT) power spectra overlaid with the Linear Prediction Coefficient (LPC) frequency response to generate graphs of the transfer function estimate at the stable point in the centre of the vowel in each spectrogram. This sampling point was chosen because the Australian vowel /i:/ characteristically begins with an onglide, and the stable point of the spectrogram was considered to more accurately represent the vowel target. The fundamental frequency (F_0) and the centre frequencies of the F-pattern (using both spectrograms and FFT and LPC functions) were then measured. The measurements were subjected to a series of independent t-tests, measuring differences between normal and shouted speech for each subject, as well as differences between each subject in both normal and shouted modes of speech. Independent t-tests were chosen over dependent t-tests because the variables "normal" and "shouted" were considered to belong to two independent groups of data, even though they were uttered by the same speakers.

3. RESULTS OF ACOUSTIC ANALYSIS

General comments

A total of 26 tokens of normal speech and 20 tokens of shouted speech were collected from speaker A, while 20 tokens of normal speech and 22 tokens of shouted speech were collected from speaker B. Table 2 sets out the mean F_0 as well as mean centre frequencies for the F-pattern of /i:/ for both speakers in both normal and shouted modes of speech.

Speaker	F_0 (Hz)		F_1 (Hz)		F_2 (Hz)		F_3 (Hz)		F_4 (Hz)	
NORMAL SPEECH (n=26 (A), 20 (B))										
	Mean	s.d.	Mean	s.d.	Mean	s.d.	Mean	s.d.	Mean	s.d.
A	135	13.4	393	49.9	2027	181.6	2469	140.5	3228	602.1
B	180	13.6	378	22.4	2102	252.0	2634	352.5	3617	543.8
SHOUTED SPEECH (n=20 (A), 22 (B))										
A	223	10.8	482	33.8	2057	144.3	2494	218.3	3165	321.2
B	261	15.6	529	37.5	2076	237.6	2583	215.7	3494	509.8

Table 2. Means and standard deviations for F_0 and F-pattern for /i:/ for both normal and shouted speech.

Table 2 shows an increase of around 39% in speaker A's F_0 for shouted speech, while speaker B showed an average increase of 31% in F_0 when he shouted. It can also be seen that F_1 increased for shouted speech in both speakers, with an average increase of 19% for speaker A and an average increase of 28% for speaker B. However the other formants (F_2 to F_4) did not show a great deal of variation in raw scores between normal and shouted speech for either participant.

Differences between 'normal' and 'shouted' speech

The data for each participant was then analysed using independent t-tests for each of F_0 and the first four formants. The results of the t-tests are set out in Table 3. These results show that for both speakers there is a highly significant difference in F_0 between normal and shouted speech. There is also a significant increase in F_1 between normal and shouted speech for both speakers.

However there is no significant difference in the other formants (F_2 to F_4) of either speaker regardless of whether an utterance was shouted or spoken at a normal level of intensity.

	Speaker A			Speaker B		
	t score	significance level	df	t score	significance level	Df
F_0	23.913	.000	44	16.975	.000	38
F_1	6.804	.000	44	15.012	.000	38
F_2	.607	.547	44	.498	.621	38
F_3	.486	.629	44	.496	.623	38
F_4	.419	.678	44	.632	.531	38

Table 3. Results of independent t-tests for difference between normal and shouted speech for each speaker

Differences between speakers

T-tests were also run between the speakers in both normal and shouted modes of speech to determine what, if any, differences there were between the speakers. The results of these t-tests, set out in Table 4, indicate that there is a highly significant difference between the F_0 of both speakers, regardless of whether or not they were shouting. The t-tests also showed there was a highly significant difference in F_1 between speakers for shouted speech but there was no significant difference between the speakers' F_1 in normal speech. There was no significant difference in F_2 between the speakers in either normal or shouted speech, and there was a significant difference between the speakers in F_3 in normal speech, but this was not the case for F_3 in shouted speech. However in F_4 in both normal and shouted speech, there was a significant difference between the speakers.

	Normal speech			Shouted speech		
	t score	significance level	df	t score	significance level	df
F_0	11.166	.000	44	9.111	.000	40
F_1	1.282	.207	44	4.295	.000	40
F_2	1.170	.248	44	.302	.764	40
F_3	2.179	.035	44	1.319	.195	40
F_4	2.274	.028	44	2.473	.018	40

Table 4. Results of t-tests for between speaker differences for both normal and shouted speech.

4. DISCUSSION

Differences within speakers between normal and shouted speech

The above results show that there is a significant rise in F_0 in shouted speech when compared to normal speech. These findings are consistent with Harris & Weiss's results, who found an average rise in pitch (i.e. as F_0) of 34% for loud speech. This compares favourably to average rises of 31% and 39% for each of the two speakers in this study. These results also clearly support the first hypothesis: there is a significant difference in F_0 between normal and shouted speech. Furthermore, both this study and a previous study by Harris & Weiss (1964) indicate this increase can be expected to be in the order of between 30 and 40 per cent.

Although F_0 rises in shouted speech, there is not a similar rise in centre frequencies in the corresponding F-pattern. While it can be said that there is a significant rise in F_1, the results of this experiment indicate that there is no significant change in the centre frequencies of F_2, F_3 or F_4 for either speaker between normal and shouted speech. An average rise in F_1 of 10% was observed in the Harris & Weiss study. This is somewhat lower than the average increase in F_1 found in the present study, however it should be pointed out that the F_1 measurements in the Harris & Weiss study were based on one speaker only. The present study also looks at data from just two speakers. To obtain a better idea of the average increase in F_1 across the population, data from a much large sample of speakers should be obtained and analysed. It is, nevertheless, appropriate to note that the results of this study are consistent with results from the Harris & Weiss study to the extent that a small, but significant, rise in F_1 was found. It should also be noted that the Harris & Weiss study found that "no significant change could be observed for F_2 or F_3" (Harris & Weiss

1964: 936). While they did not measure F_4, the present study again reflects their findings for F_2, and F_3.

There are a number of possible explanations for the rise in F_1 in shouted speech. Firstly, since F_1 in /i:/ reflects the overall length of the vocal tract, it may be that this rise is due to the larynx being raised, with consequent shortening of the vocal tract. However if this were so, then one would also expect some corresponding increase in F_2 reflecting concomitant changes to the length of the pharynx. The results are not consistent with this as Table 2 shows, for although speaker A had a slight, although statistically non-significant increase in F_2, speaker B's F_2 was in fact lower in shouted speech. An alternative, and perhaps more plausible explanation, may be that jaw and/or lip openings may vary between normal and shouted speech, resulting in a higher F_1 in shouted speech. It is also possible that the position of the tongue body varies from normal to shouted speech. This requires further investigation.

The second hypothesis, that there is a difference between normal and shouted speech in a person's average F-pattern, has not been supported by either this study or the previous study by Harris & Weiss. Rather, the results of both studies indicate that while there is a significant difference in a person's average F_1 between normal and shouted speech, there is no significant difference in the F-pattern for formants other than F_1.

While this study was conducted under conditions which do not simulate the circumstances of forensic comparison precisely, the results nevertheless indicate that it may be feasible for the forensic phonetician to compare the F-pattern of two different speech samples when one of the samples contains shouted speech, provided increases in F_0 and F_1 between the normal and shouted samples are taken into account. Determining a reliable mean and standard deviation for these increases will depend on future studies using a larger sample population. There is evidence, however, that the F-pattern in the upper formants appears to remain relatively unchanged, regardless of mode of speech, and therefore the use of these formants in comparison of speech samples can be considered reliable.

Differences between speakers in both normal and shouted speech.
The results of the experiment clearly support the third hypothesis, that there is a difference in the average F_0 between speakers, regardless of whether they speak normally or shout. However this result is not surprising since pitch of voice was not something that was consciously controlled for in the experimental design. The average pitch of a person's voice depends very much on their physical characteristics, especially the length and thickness of their vocal folds. Furthermore, F_0 is not, by itself, necessarily regarded as a comparable measure in forensic phonetics, since two people may well manifest a similar mean F_0, particularly if this measurement occurs around the measures of central tendency for the population as a whole. In future experiments, however, it may be useful for comparing F-patterns if this variable is taken into account when selecting subjects.

The statistical results for the F-pattern present a more confusing picture, and there is by no means any clear support for hypothesis (4): there is a difference in F-pattern between speakers, regardless of whether they speak normally or shout . Clearly there is a significant difference in F_1 between the two speakers for shouted speech, however this does not appear to be the case for normal speech where there is no significant difference in F_1. For F_2 there is no significant difference between the speakers for either normal or shouted speech, and in F_3 there is a significant difference between speakers only in normal speech. F_4 is the only formant which indicates a significant difference between speakers for both normal and shouted speech.

Given that forensic phoneticians tend to favour the upper formants to distinguish between-speaker differences (Philip Rose, personal communication), one would expect differences to emerge in the data to reflect this. A number of factors may have contributed to the lack of conclusive results here. Firstly, vowel quality - and therefore formant structure – can be affected by the postvocalic consonant, which provides an articulatory target beyond the vowel itself. The number of different postvocalic contexts used in the data may therefore have led to changing targets for the vowels themselves, resulting in inconsistent F-patterns. Some evidence exists for this in the raw data. For

example F_2 tended to be higher for the vowel in /bi:t/ than it was in /bi:p/ for both normal and shouted segments for speaker B, and /bi:n/ segments had a noticeably lower F_2 than other /bi:(C)/ segments for speaker A. Future experiments should take into account the possible effects of coarticulation by using segments with more consistently similar postvocalic consonants, as well as prevocalic consonants (as was done in the experiment).

A second factor, which may have contributed to inconclusive results, is that stress patterns on segments was not always consistent: sometimes the syllable was stressed and sometimes it was unstressed. This also needs to be taken into account in future studies.

5. CONCLUSION

The pilot study reported here has provided insight into the acoustic differences between normal and shouted speech. The most useful outcome of the study is the strong indication that the F-pattern of upper formants is relatively unaffected by mode of speech, a result that is consistent with the findings of a previous study which compares the F-pattern of normal and shouted speech (Harris & Weiss, 1964). The study is therefore of importance to forensic phonetics since it demonstrates that speech samples which contain shouting may still be usefully compared for forensic purposes with samples of normal speech.

This pilot study points to useful future research on the differences between the acoustic properties of normal and shouted speech. Refinement of the method used to collect the data, particularly in terms of the segments sought for elicitation, together with the addition of several more subjects, should lead to some valuable additions to the bank of knowledge upon which the forensic phonetician relies.

ACKNOWLEDGEMENTS

I would like to thank Phil Rose for his inspiration and help in the early stages of writing this paper, and the two anonymous reviewers of the paper for their very constructive comments.

REFERENCES

Burridge, K. & J. Mulder. (1996) *English in Australia and New Zealand. An Introduction to Its History, Structure, and Use.* Melbourne: Oxford University Press.

Giles, H. & N. Coupland. (1991) *Language Contexts and Consequences.* Milton Keyes: Open University Press.

Giles, H., N. Coupland & J. Coupland (1991) "Accommodation theory: Communication, context, and consequences". Chapter 1 in Giles, H, N. Coupland and J. Coupland (eds.).

Giles, H., N. Coupland & J. Coupland (eds.) (1991) *Contexts of Accommodation. Developments in Applied Sociolinguistics.* Cambridge: Cambridge University Press.

Hacki, T. (1996) "Comparative speaking, shouting and singing voice range profile measurement: physiological and pathological aspects." *Log Phon Vocol* 21, 123-129.

Harris, C.M. & M.R. Weiss (1964) "Effects of speaking condition on pitch." *Journal of the Acoustical Society of America* 6(5), 933-936.

Hollien, H. (1990) *The Acoustics of Crime: The New Science of Forensic Phonetics.* New York: Plenum Press.

Laver, John. (1980) *The Phonetic Description of Voice Quality.* Cambridge: Cambridge University Press.

Mitchell, A.G. and A. Delbridge. (1965) *The Pronunciation of English in Australia.* Sydney: Angus and Robertson.

Nolan, Francis. (1983) *The Phonetic Bases of Speaker Recognition.* Cambridge: Cambridge University Press.