THE ACOUSTIC CORRELATES OF HEIGHTENED EMOTION: THE MAKING OF MARRIAGE VOWS

L. Penny and M. Carmody

Department of Speech Pathology
Flinders University of South Australia

ABSTRACT The speech of people under conditions of high emotional arousal, making their marriage vows, is compared on a number of variables with their speech recorded in ordinary relaxed circumstances. The findings relating to speaking fundamental frequency (but not its variability) and rate of utterance are in broad agreement with the reported influence of heightened emotional arousal on the speech signal. Also reported are changes to formant bandwidths, but the interpretation of these findings is problematical.

INTRODUCTION

In a recent review article, Murray and Arnott (1993) summarise present understanding of the acoustic correlates of emotional expression in speech. This understanding has accrued from a number of studies using different methodologies and investigating different emotional states. Generally there are three different approaches. One uses field recordings of people in naturally occurring situations where strong emotions are likely to be, or have been, aroused, a second uses actors (professional or amateur) to simulate emotion, and the third synthesises speech according to the properties identified by the first two methods, and seeks to confirm the validity of these descriptions in perceptual judgment studies. The first two methods suffer from what might be thought of as complementary disadvantages. Field studies are opportunistic and rarely are recording conditions ideal. There can be no control over the content of speech, and it is not always possible to make comparisons with the speaker's voice under ordinary conditions, or while expressing other emotions. One cannot be certain exactly what emotion is being experienced - though this problem applies also to the simulation approach, where actors may have different understandings of particular emotions. The problem is that neither descriptions of emotions as internal feelings nor descriptions of the observable states of another are available in rigorous form. There are also culturally prescribed modes of behaviour which may modify or mask the quality and intensity of emotional expression, as, for example, almost certainly happened during the broadcast of the Space Shuttle launch which ended in disaster. There is no sign, (Penny and Barrett 1993), either perceptually or acoustically in the voice of the broadcaster that the disaster had happened. Presumably he was conscious of the watching audience of family and friends, and exercised control over his voice. This point nicely illustrates some of the difficulties - what does one conclude when the expression and the circumstances are apparently at variance? But, despite this, and the other difficulties described, real-life studies of emotion have a power that makes them an essential part of the general approach to the topic.

Simulation studies have the advantage of control over recording conditions and content of message. The same actors may be recorded performing a variety of tasks. But these studies rest on the assumption that emotions can indeed be accurately and completely simulated. This seems a reasonable assumption, and Williams and Stevens (1972) produce some evidence for it, in that an actor working from the transcript of the Hindenburg disaster modelled quite closely features of the original broadcast. But in fact they compare only short stretches of the records, on a rather restricted range of measures and they admit that the simulation was a somewhat attenuated version of the original.

Despite the problems inherent in each approach, there is at least a general agreement emerging on the significant acoustic characteristics of speech and voice in a number of emotional states. These findings have been well summarised by Murray and Arnott (1993) and also by Couper-Kuhlen (1986). In general, heightened emotional arousal, whether of anger, fear or joy, is accompanied by a marked increase in fundamental frequency and its range. The extent of the increase seems highly correlated with the intensity of the emotion experienced. Rate of utterance often increases, more in fear than in anger or joy. Anger may be marked by abrupt pitch changes and an increase in vocal intensity. Joy is marked by smooth, rising intonation contours, and also by an increase in vocal intensity. Fear has apparently normal intonation and intensity levels, but voicing may be irregular. There is a good deal of over-lap in these descriptions, perhaps surprising as the emotions are subjectively so different. But the psychological literature has long recognised the similarity across emotions when expression is intense, and has noted that the accurate recognition of emotion is much aided by knowledge of the

situation, not just the appearance of the person. By contrast, sadness is marked by reduced speech rate, lower fundamental frequency, reduced variability in pitch, and flattened intonation contours marked by downward terminal inflections. Intense emotion may also cause speech dysfluencies and word-finding difficulties. This general description of the vocal qualities of emotion has been largely confirmed (except for joy) by Penny and Barrett (1993) who examined the productions of the same speaker, the radio journalist Murray Nicol, who continued to broadcast his experiences over several hours during the horrific bushfires of the Adelaide hills in 1983.

Nicol was able to indicate on the transcript the state of his feelings at different times during the fires, and these ranged from alarm to fear, to terror (when he thought that he and those sheltering with him would die as the fire swept over them) and finally to despair as he sat in the middle of the road and watched his house burn down. Recordings of Nicol's voice under ordinary conditions were readily available for comparison purposes. Fundamental frequency and its range correlated strongly with intensity of emotion.

A particular limitation of field studies to date has been the relatively restricted range of emotions examined with emotions of positive tone largely omitted. In part this is due to the opportunistic nature of the data and the fact that joyful occasions are generally either intensely private, or public and accompanied by a great deal of extraneous noise, as in sporting victories. Noise levels in field recordings present a problem in that it is difficult to be sure about the cause of increases in vocal intensity. This impacts on fundamental frequency since vocal effort and speaking fundamental frequency are not, in practice, independent. There are also very few reports on emotion in women's voices.

The now popular practice among couples getting married to have their weddings video-taped has provided us with an opportunity to study heightened emotion in quiet conditions in both men and women, and furthermore, in an emotion which is not unpleasant in its tone. It is predicted that fundamental frequency and its range, and rate of utterance, will all increase while in the heightened emotional state of making marriage vows, even though increased vocal effort is not necessary. Dysfluencies may increase. It is also predicted that changes to formant bandwidths will occur since the damping properties of the vocal tract will change as a consequence of the physiological concomitants of emotion -tensing and drying of the tract.

METHOD

Subjects and Materials.

Six young couples, all native-born speakers of Australian English, and all in good health, and all with their own teeth, made available to us the video recordings of their weddings, and provided us with readings of a standard passage and a minute of free speech. The maximum time interval between the marriage and making the standard reading was three years. At the time of this recording the subjects ranged in age from 22 to 35 years, with a mean of 27. Transcriptions of the vows, the standard passage and the free speech were prepared. The wedding vows were dubbed on to audio-tape using a Panasonic Video Cassette recorder NV640 HQ and a Marantz CF230 Stereo Cassette recorder. Detail of the equipment used for the original recordings is not available. The standard passage (the Swagman) and free speech were recorded on a Sony TCD-D10 Digital Audio Tape-Corder using a Sony ECM 959 DT Microphone. A Selby 449833 stopwatch was used in the calculation of speech rate. Fundamental frequency and its range were obtained using the Kay DSP 5500 Sonograph employing programme #03, and generally from the 4th harmonic. The formant bandwidths were extracted using the Soundscope programme on the Centris 650.

RESULTS

Speech measures.

Two measures of speech production were examined, namely rate of utterance and dysfluencies. Rate is reported in Table 1.

Table 1 Rate of utterance in syllables per second

Subject	Vows	Swagman	Free speech
1 Male	5.36	4.17	4.17
1 Female	5.40	4.17	3.79
2 Male	5.08	3.84	3.80
2 Female	4.79	3.69	2.13
3 Male	3.24	2.51	2.01
3 Female	4.89	4.00	1.80
4 Male	5.00	4.17	3.75
4 Female	5.08	4.26	3.81
5 Male	6.17	5.29	2.52
5 Female	5.15	5.29	3.41
6 Male	6.70	3.56	2.92
6 Female	6.41	4.17	3.84

A 2-factor analysis of variance with repeated measures showed that there were no sex differences, and no interaction effects of sex with utterance type, but that there were differences between utterance types, F-test=30.16, df=2, p=0.001, which paired t-tests showed to be significant as follows:-
between vows and the standard passage, t=4.98, one-tail, df=10, p=0.0002;
between vows and free speech, t=7.4, one-tail, df=10, p=0.0001;
between the standard passage and free speech, t=3.5, one-tail, df=10, p=0.003.

Dysfluencies.

No subjects were dysfluent when making their wedding vows in tnat no hesitations, fillers or repetitions occurred. Dysfluencies did occur in the other conditions but as these are not the focus of comparison no further analysis is reported.

Voice Quality

Modal speaking fundamental frequencies in three conditions, making the vows, in reading a standard passage and in free speech, are reported in Table 2.

Table 2. Modal fundamental frequency in Hertz in three utterance types.

Subject	Vows	Swagman	Free speech
1 Male	190	120	130
1 Female	230	170	195
2 Male	100	120	125
2 Female	220	192	207
3 Male	110	125	130
3 Female	220	210	240
4 Male	140	120	120
4 Female	240	205	215
5 Male	120	105	130
5 Female	250	220	235
6 Male	110	115	125
6 Female	230	212	220

A 2-factor analysis of variance with repeated measures produced the following:-
a significant difference between the sexes, F-test=168.7, df=1, p=0.0001,
a significant difference across utterance types, F-test=5.27, df=2, p=0.014.
Paired t-tests yielded:
between vows and the standard passage, t=2.6 one-tail, df=10, p=0.041;
between vows and free speech, t=0.97, n.s.;
between the standard passage and free speech, t=-4.95, two-tail, df=10, p=0.01.

Range of fundamental frequency is presented in Table 3.

Table 3. Range of fundamental frequency in Hertz over three utterance types.

Subject	Vows	Swagman	Free speech
1 Male	50	120	100
1 Female	80	180	200
2 Male	70	100	120
2 Female	90	140	200
3 Male	130	140	120
3 Female	70	140	180
4 Male	80	60	140
4 Female	120	240	260
5 Male	90	80	100
5 Female	170	160	160
6 Male	70	140	160
6 Female	70	240	180

A two-factor analysis of variance with repeated measures resulted in the following:-
A significant difference between the sexes, F-test=18.8, df=1, p=0.0015, and significance across utterance types, F-test=15.19, df=2, p=0.0001. Paired t-tests yielded:
between vows and standard passage, t=-3.24, df=10, two-tail, p=0.0079;
between vows and free speech, t=-4.6, df=10,two-tail, p=0.0008, and;
between standard passage and free speech, t=-1.4, n.s.

Bandwidths

The bandwidths of formants 1 and 2 of single target vowels were extracted from the Swagman reading and the Vows. Problems were experienced in obtaining the values in the Vows condition as it offered a limited number of examples, and the quality of the recordings was poorer than had been hoped largely because the subjects spoke very softly, especially the women. Bandwidths are reported in table 4.

Table 4 Mean bandwidths in Hertz for formants 1 and 2 for men and women speakers.

Subjects	Vows	Swagman
Male	B1=241; B2=291	B1=148; B2=214
Female	B1=215; B2=351	B1=161; B2=235

A 2-factor repeated measures analysis of variance for Bandwidth 1 showed that there was no significant sex difference, F=0.116 with 1df, and no interaction effects of sex and utterance type, F=0.675, df=1. There was a significant difference though over utterance type, F=9.24, df=1, p=0.012. No further analysis of bandwidth 2 was proceeded with for reasons discussed below.

DISCUSSION

Presumably all the subjects were emotionally stressed when saying their wedding vows. It is not quite clear what the quality of the emotion is, perhaps nervousness, but it is a situation of great significance for all the people involved, and they all also expected a celebration of rather uninhibited joy to follow. Changes to the speech signal are largely but not entirely as predicted. Rate of utterance increased significantly, and so did speaking fundamental frequency, but range of fundamental frequency decreased significantly. This pattern, of increased pitch with decreased range, is not reported for any other emotional state. There were no dysfluencies, perhaps because subjects repeated their vows in short phrases, after the officiating minister and there had been prior rehearsal. The features of the subjects' speech did not mimic the minister, however, who spoke with (perceptually) normal or above normal pitch, and slightly increased range as well as decreased rate of utterance. There were no differences between the sexes on rate of utterance, but the fundamental frequency increase was relatively greater for the women than for the men, and the flattening of frequency range was less for

men than for women. It can be concluded that the voices of women show greater effects of emotion than the voices of men.

The interpretation of the significance in the bandwidth data is difficult. Our original hypothesis was that the normal vocal tract would dampen formant energy more than the drier, more rigid tract in the emotionally tense situation, resulting in greater bandwidths in the former condition. The opposite has happened. A number of explanations are possible. The two sets of recordings from which the data were obtained were made under very different conditions, one in a large and relatively empty church and the other in the sitting-room of a private house. The subjects spoke very softly when making their vows and an unusual pattern of relative formant amplitudes characterises the vows condition, with F2 relative amplitudes exceeding F1 on many occasions. The first formant is usually the most intense in vowels. A sign test on the relative amplitudes is insignificant for this reason in the vows condition, but highly significant in the standard passage condition, with F1 relative amplitudes regularly exceeding F2. It seems therefore that we may have an effect stemming from very low but irregular vocal effort. It is of interest to note that the bandwidth values we obtained are far in excess of those suggested in the literature (Fant, 1967; Flanagan, 1972; Kent and Read, 1992). This, if it is not an artefact of our extraction method, may be due to either the fact that our measures seem to be the only ones taken from connected speech, or to the fact that our subjects were, in the main, speakers of broad Australian English and it is possible that such speakers have characteristically a vocal tract setting with higher damping effects. In any event we offer these findings tentatively, and have not included bandwidth 2 values because, although they differ statistically over the two utterance types, the difference in actual values is smaller than the 40% difference that Flanagan (1972) considers to be the limen for bandwidth 2 discriminability. Unless one can hear the difference, bandwidth change cannot carry any information, even about emotional state.

REFERENCES

Couper-Kuhlen, Elizabeth (1986) An introduction to English prosody. Edward Arnold, London.

Fant, C. Gunner M. (1967) On the predictability of formant levels and spectrum envelopes from formant frequencies. In Lehiste, I. (ed) Readings in acoustic phonetics. The M.I.T. Press, Cambridge, Mass.

Flanagan, J.L. (1972) Speech Analysis Synthesis and Perception 2nd ed. Springer-Verlag, Berlin.

Kent, Ray D.& Read, Charles, (1992) The acoustic analysis of speech. Singular Publishing Group, San Diego.

Penny, L. & Barrett, C. (1993) Some acoustic correlates of heightened emotional arousal: reporting bushfire experiences. The 2nd Australian Voice Symposium, Melbourne.

IS FOREIGN ACCENT V!SIBLE?

Duncan Markham

Department of Linguistics and Phonetics
Lund University, Sweden

ABSTRACT - The paper presents a pilot experiment testing speakers' ability to identify language and native/non-native status of other speakers from visual stimuli.

INTRODUCTION

The present experiment is motivated by Honikman's (Honikman, 1964) thesis that there exist language-specific 'articulatory settings' which must be adopted in order to attain native-speaker-like phonetic competence. The pilot experiment in this paper tests a method for assessing the role of peripheral articulation — primarily lip and jaw movements/aperture — in the perception of foreign accent and the identification of languages. This research method and the resultant data may prove useful in the development of visual speaker/language recognition systems and synthesis which deal with foreign/non-standard accents.

Most speakers utilise visual cues as well as the acoustic signal in the perception of speech. Research has indicated that speech perception is degraded where listeners do not have access to visual information for the speaker (eg Hashimoto and Seki, 1994) and it is a common experience amongst polyglots that a familiar language can be recognised at a distance just from lip movements, without word identification playing a role. I seek here to gain some measure of the degree to which an unconscious knowledge of 'typical' visible behaviour is accessible. Of particular interest is the transfer of the peripheral articulatory setting from L1 to L2, whether this is visible, and whether near-native pronunciation precludes such transfer. There exists a modest body of literature relating to the transfer of phonetic characteristics (as distinct from phonological transfer) from L1 to L2, whereof only a very small part is directly concerned with phonetic and imitative processes (the emphases in the existing research lie clearly within sociolinguistics and second language acquisition, especially with regard to the question of a critical period of acquisition).

PROCEDURE

Six speakers (4 females, 2 males) with some experience in a number of languages were selected for the experiment. They were chosen so as to represent a mixture of native, non-native, and near-native language abilities for a variety of languages. The speakers were between 20 and 35 years of age. Four of the six speakers were native speakers of Swedish, one of German and one of Thai. The two non-native speakers of Swedish had lived in Sweden for more than 8 years. All speakers had learnt English and German in school at least to a level which allowed them to read a simple text aloud. Two of the informants also had sufficient knowledge of French to read a French text. All informants had normal hearing and no speech disorders according to self assessment.

—Speakers M(ale)1 and M2 speak English and German with marked accent, especially M2. M1 had spent a few weeks in English and German speaking countries, but was only confident in English.
—F(emale)1 has L1 Thai, and moved to Sweden at the age of nine.She speaks at least near-native Swedish, and is a confident reader of English, French, and German. She began learning English at approximately six years of age and speaks easily. She has been to language courses in France.
—F2 speaks English confidently, having spent time in the USA, but with marked accent. She reads German hesitantly, and has trouble speaking it.
—F3 has L1 German. She speaks Swedish confidently, having resided in Sweden for more than eight years. She speaks English with marked accent. Her German phonology shows some effects of Swedish L2 influence.
—F4 speaks German, French and English very confidently, especially the latter. She shows mild foreign accent for all L2s.