SIMULATED EMOTIONS: AN ACOUSTIC STUDY OF VOICE AND PERTURBATION MEASURES

S. P. Whiteside

Department of Human Communication Sciences, University of Sheffield, UK.

ABSTRACT

This brief study presents a set of acoustic correlates for a number of vocal emotions simulated by two actors. Five short sentences were used in the simulations. The emotions simulated were neutral, cold anger, hot anger, happiness, sadness, interest and elation. The seventy sentences were digitised and a number of acoustic analyses were carried out, which included a number of perturbation measures. The acoustic parameters investigated were: i) mean overall fundamental frequency (Hz); ii) overall mean energy (dB); iii) overall mean standard deviation of energy (dB); iv) mean overall jitter (%); and v) mean overall shimmer (dB). These acoustic parameters were used to profile the vocal emotions.

Results showed that the actors displayed similarities in their acoustic profiles for some emotions like anger and sadness, for example. The results are presented and discussed in brief.

1. INTRODUCTION

Speech and voice characteristics are shaped by a wide variety of complex and inter-related factors. These complex and inter-related factors include stress [1, 2] and human vocal emotion [3, 4]. It is widely documented that different emotions can be signalled and communicated through a speaker's vocal characteristics. See [4, 5 and 6] for reviews of the literature on vocal emotions.

Research into how naturalistic situations shape vocal emotions and the characteristics associated with them [7] would be preferred to the use of simulations by actors [3, 7, 8, 9, 10]. However, truly naturalistic data is difficult to obtain, because the collection of such data is often fraught with serious ethical and moral considerations.

The ecological validity of studying vocal emotion using actors has been questioned by some researchers [11]. However, some recent studies into the vocal expression of emotion [3, 9, 10, 12], have demonstrated that actors are able to simulate vocal emotions, which are on the whole, successfully decoded by listeners at better than chance levels. Banse and Scherer [3] for example, found that from a set of 14 emotions portrayed by actors (hot anger, cold anger, panic fear, anxiety, despair, sadness, elation, happiness, interest, boredom, shame, pride, disgust and contempt), that the only two emotions which were not accurately identified, were shame and disgust. They suggest that these two emotions rely more on visual rather than vocal cues [3]. In the case of shame for example, people are likely to avoid speaking while in this emotional state and therefore there are few perceptually salient cues which listeners would associate with this emotion. With regard to disgust, the signal of this emotion often takes the form of brief affect bursts (e.g. "yuck!") rather than a speaker adopting a particular vocal setting and/or voice quality [3].

Anger, fear, sadness, joy and disgust are among the emotions that have been frequently studied and their acoustic characteristics have been documented in a number of sources [5, 6, 9, 10, 13, 14]. The main acoustic cues that have been documented as characterising these emotions can be summarised as follows.

- **Anger**: There are two types of anger that are described in the literature [e.g. 3]. These are hot anger and cold anger and although similar in quality, differ in intensity. The former, which is more intense in quality, is characterised by an increase in mean fundamental frequency (F0) and mean intensity; an increase in F0 variability and range; an increase in high-frequency energy, falling F0 contours and increased articulation rate. Acoustic characteristics which typify cold anger include: an increase in mean F0, an increase in mean intensity, an increase in high frequency energy and falling F0 contours.

- **Fear**: Fear is characterised by an increase both in mean F0 and F0 range and by an increase in both high frequency energy and articulation rate;

- **Sadness**: This emotion is characterised by a decrease both in mean F0 and F0 range; a decrease in mean intensity, falling F0 contours and decreases in high-frequency energy and articulation rate and lower levels of articulatory precision; and

- **Joy** or **Elation**: These emotions are characterised by an increase in the following: mean F0 and F0 range, F0 variability, mean intensity, and in some cases, an increase in high-frequency energy and articulation rate.

The above summary indicates that there are primary and salient acoustic characteristics for vocal emotions. These characteristics include: i) fundamental frequency characteristics which include, the level, range and contour patterns; ii) vocal energy or intensity of voice; iii) the distribution of energy in the spectrum; iv) the location of formant frequencies; and v) temporal characteristics such as speech rate [3]. Whether there are robust emotion-specific vocal characteristics is investigated by Banse and Scherer [3], who report on a number of acoustic profiles that indicate the degree of intensity typical for different emotions together with their valence and quality. In addition, these vocal profiles are powerful enough to indicate those emotions that are distinctly different (e.g. hot anger versus

boredom) or those that have a family resemblance (e.g. elation and happiness, despair and sadness).

This preliminary study aims to examine a set of acoustic parameters which include a number of voice and perturbation measures for 70 sentences representing seven vocal emotions (neutral, hot anger, cold anger, happiness, sadness, interest and elation) simulated by an actor and an actress. The following acoustic parameters were investigated: overall mean energy; standard deviations of energy; mean fundamental frequency values; mean jitter (%) values; and shimmer (dB) The results of this brief study are presented and discussed within a framework that attempts to characterise the simulations. In addition, the similarities and differences between the portrayals of the actor and actress are profiled and discussed in brief.

2. METHOD

2.1. Subjects

RT is a 27 year old female with a standard southern British English accent. RP is a 32 year old male, with a standard British English accent with some northern colouring. Both speakers are non-smokers and have normal hearing, speech and language. At the time of recording, RT and RP had a total of three and twelve years respectively, of amateur and professional acting experience.

2.2. Speech Material

The speech material used to elicit the emotional data consisted of 5 brief sentences (*Weigh your yellow ruler; Wheel your wallaroo away; Rule your row warily; Reel your wheel away; You will reel your wool*). The 7 emotions that were simulated were *neutral, cold anger, hot anger, happiness, sadness, interest* and *elation*. These emotions with the exception of *neutral,* were chosen from the set of emotions studied by Scherer and his colleagues [3, 9, 10] to represent a balance of different emotional strength, valence and activity.

2.3. Speech Data Collection

Several weeks prior to the recording session, both RT and RP were given the speech material to rehearse the portrayal of the 7 emotions listed above. All recordings were carried out in a sound proof room using a Sony DAT recorder. All 70 sentences (7 emotions x 5 sentences x 2 speakers) were digitised (sampling rate of 10kHz) onto a Kay Computerized Speech Lab (CSL) model 4300, which was used to carry out all the acoustic analyses and derive the acoustic parameters which are described below.

2.4. Acoustic Analysis

A total of 5 parameters were derived for each of the 70 sentences (2 speakers x 5 sentences x 7 emotions). The details of these are given below.

2.4.1. Mean energy and mean energy standard deviation (dB) - 2 parameters

The mean energy (in decibels of sound pressure- dB SPL) of the speech samples was calculated using a 20ms frame length. The dB SPL are computed from the speech pressure waveform as 20 times the log of the square root of the energy which has been divided by the number of sampled data points in the frame. The standard deviation values calculated by this algorithm were subsequently used to derive the mean of the standard deviations to provide an indicator of the variability of energy.

2.4.2. Mean Fundamental Frequency (Hz)- 1 parameter

Mean fundamental frequencies were obtained for each phrase using an auto-correlation method and a frame size of 20 ms with a 20 ms frame advance.

2.4.3. Mean Jitter (%)- 1 parameter

Mean jitter values (%) were obtained for each phrase using Koike et al.'s [15] perturbation formula.

2.4.4. Mean Shimmer (dB)- 1 parameter

Mean shimmer values (dB) were obtained for each phrase, as cycle-to-cycle variations in amplitude.

3. RESULTS

The mean and standard deviation values of the acoustic parameters are given in Tables 1 and 2 for speakers, RT (F) and RP (M), respectively. Further to this, the mean z-scores for jitter (%) and shimmer (dB) were calculated for each speaker. These are profiled by emotion, in Figures 1 and 2, for speakers RT and RP, respectively. The jitter and shimmer data for speakers RT and RP were subsequently classified by machine using a discriminant analysis routine. The mean number of cases correctly classified for the 7 vocal emotions, were 45.7% and 37.1%, for speakers RT and RP, respectively. The details of this discriminant analysis are given in Table 3. Subsequent to this, all five parameters were used in the discriminant analysis routine. Here, correct classification scores were raised to 88.6% and 85.7%, for speakers RT and RP, respectively.

Emotion / Parameter	Neutral	Cold Anger	Hot Anger	Happiness	Sadness	Interest	Elation
Energy (dB)	53.3 (1.6)	55.5 (1.6)	65.0 (.9)	59.9 (1.3)	47.7 (1.4)	59.3 (1.2)	59.5 (1.8)
Energy sd (dB)	5.8 (2.1)	6.3 (1.2)	10.2 (.4)	8.1 (.8)	4.3 (.5)	7.4 (1.0)	7.6 (.7)
Mean F0 (Hz)	188.8 (4.2)	188.0 (5.0)	171.7 (9.2)	172.9 (12.9)	183.3 (6.7)	167.9 (13.2)	156.5 (14.2)
Jitter (%)	1.9 (1.1)	2.8 (.8)	9.5 (1.6)	4.5 (1.4)	1.8 (.8)	6.1 (1.2)	8.4 (1.2)
Shimmer (dB)	.9 (.1)	1.2 (1.0)	1.7 (.1)	1.2 (.2)	.8 (.1)	1.3 (.2)	1.6 (.2)

Table 1: Mean and standard deviation values for acoustic parameters - speaker RT (F).

Emotion Parameter	Neutral	Cold Anger	Hot Anger	Happiness	Sadness	Interest	Elation
Energy (dB)	56.2 (1.8)	57.7 (1.6)	60.8 (.7)	59.8 (1.4)	46.9 (2.3)	58.1 (1.2)	63.0 (1.2)
Energy sd (dB)	6.2 (.9)	7.1 (.8)	7.6 (.8)	7.1 (.4)	3.5 (.9)	6.5 (.9)	7.5 (.7)
Mean F0 (Hz)	133.6 (1.8)	141.0 (10.7)	153.7 (6.8)	158.0 (10.2)	112.5 (11.4)	151.0 (15.6)	151.1 (6.6)
Jitter (%)	9.9 (1.9)	10.2 (2.2)	14.6 (3.1)	8.8 (2.6)	10.6 (1.9)	14.1 (2.5)	15.4 (3.0)
Shimmer (dB)	1.9 (.2)	1.8 (.2)	2.2 (.1)	1.8 (.2)	1.7 (.2)	2.2 (.3)	2.0 (.4)

Table 2: Mean and standard deviation values for acoustic parameters - speaker RP (M).

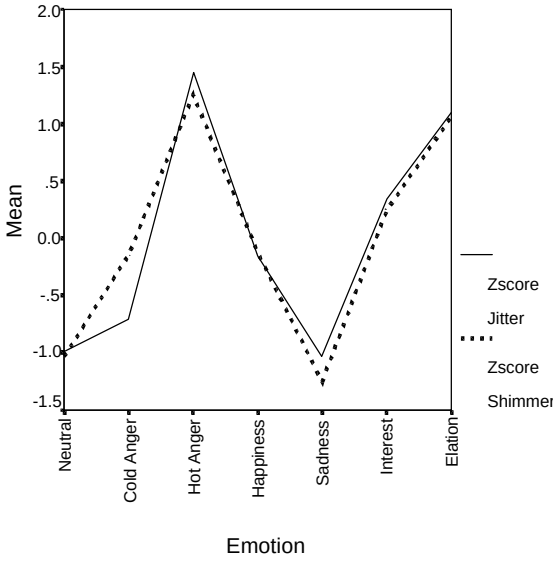

Figure 1: Mean z-scores for jitter (%) and shimmer (dB) by emotion: speaker RT (F).

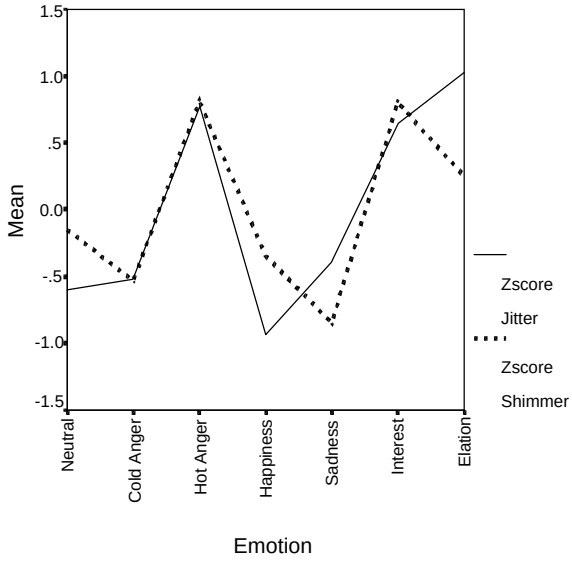

Figure 2: Mean z-scores for jitter (%) and shimmer (dB) by emotion: speaker RP (M).

Classification Emotion	Neutral	Cold Anger	Hot Anger	Happiness	Sadness	Interest	Elation
Neutral	0* 40**	20* 0**	0* 0**	20* 20**	60* 40**	0* 0**	0* 0**
Cold Anger	20* 20**	80* 20**	0* 0**	0* 0**	0* 40**	0* 20**	0* 0**
Hot Anger	0* 20**	0* 0**	60* 0**	0* 0**	0* 0**	0* 40**	40* 40**
Happiness	0* 0**	20* 0**	0* 0**	40* 60**	0* 20**	40* 20**	0* 0**
Sadness	40* 0**	0* 20**	0* 0**	0* 20**	60* 40**	0* 0**	0* 20**
Interest	0* 0**	0* 0**	0* 20**	40* 0**	0* 20**	40* 40**	20* 20**
Elation	0* 0**	0* 0**	40* 0**	0* 0**	0* 20**	20* 20**	40* 60**

Table 3: Results of discriminant analysis (%) for speakers RT (F - indicated by ***)** and RP (M - indicated by ****)**.

4. DISCUSSION

There were both similarities and differences in the acoustic profiles of the vocal emotions of the actors RT and RP (Table 1 and 2). The similarities and differences could be due both, to idiosyncrasies and, differences in the interpretation of the various emotions by the actors. The similarities included:

- Energy: an increase for hot anger compared to neutral, cold anger and sadness; an increase in for hot anger compared to neutral; a decrease for sadness compared to neutral, cold anger, hot anger and happiness;

- Energy sd: an increase for interest and elation compared to sadness; a decrease for sadness compared to hot anger and happiness.

- Fundamental frequency (F0): an increase for elation compared to sadness; a decrease for elation compared to neutral and cold anger.

- Jitter: an increase for hot anger compared to neutral, cold anger, happiness and sadness; a decrease for sadness compared to hot anger and elation.

- Shimmer: an increase for hot anger compared to happiness and sadness; a decrease for cold anger compared to hot anger.

The discriminant analysis routine using all 5 acoustic parameters showed high levels of machine recognition for both speakers. However, when the statistical procedure was carried out by only using the 2 perturbation measures, jitter and shimmer, lower levels of classification were achieved for both speakers. However, these lower levels were still above chance (14.3%), which suggests that the perturbation measures were successful in classifying some of the emotions. There were however, differences in the classification profiles of the two actors. These disparities are reflected in the z-score profiles of the 7 emotions of both speakers (Figures 1 and 2). For example, the similarities in the profiles (Figure 1) for speaker RT for Hot Anger-Elation, Neutral-Sadness and Happiness-Interest explain the patterns of misclassification (Table 3). Misclassifications in the data of speaker RP (Table 3), included confusions among the emotions Neutral-Sadness, Cold Anger-Sadness, Hot Anger-Interest-Elation, and can also be explained by the extent of the similarities and differences in the z-score profiles of these emotion groups (Figure 2).

This preliminary study found profiles of acoustic measures of F0, Energy and standard deviation of Energy, to be similar to those of earlier studies [3, 12]. Acoustic profiles of these parameters reflected the high levels of activity associated with emotions like Hot Anger and Elation, and the lower levels of activity in Sadness, for example. In addition, the perturbation measures of jitter and shimmer also showed profiles that reflected the higher levels of activity associated with hot anger compared to the more controlled and restricted nature of Cold Anger, and the low levels of activity associated with Sadness. These preliminary data suggest that perturbation measures deserve some serious consideration in the acoustic profiling of vocal emotion, and their contribution to voice quality.

5. ACKNOWLEDGMENTS

I wish to thank RT and RP without whom this study would not have been possible. The data was devised and collected by the author within the EC TIDE project TP1174.

6. REFERENCES

1. Murray, I. R., Baber, C., and South, A. "Towards a definition and working model of stress and its effects on speech", *Speech Communication*, 20, 1996, 3-12.

2. Ruiz, R., Absil, E., Harmegnies, B., Legros, C., and Poch, D. "Time- and spectrum-related variabilities in stressed speech under laboratory and real conditions", *Speech Communication*, 20, 1996, 111-129.

3. Banse, R., and Scherer, K. R., "Acoustic profiles in vocal emotion expression", *Journal of Personality and Social Psychology*, 70(3), 1996, 614-636.

4. Murray, I. R., and Arnott, J. L. "Toward the simulation of emotion in synthetic speech: A review of the literature on human vocal emotion", *Journal of the Acoustical Society of America*, 93 (2), 1993, 1097-1108.

5. Scherer, K. R. "Vocal affect expression: a review and a model for future research", *Psychological Bulletin*, 99, 1986, 143-165.

6. Scherer, K. R. "Vocal correlates of emotional arousal and affective disturbance", In H. L. Wagner and A. S. R. Manstead (Eds.) *Handbook of Social Psychophysiology*, John Wiley, Chichester, 1989.

7. Williams, C. E., and Stevens, K. N. "Emotions and speech: some acoustic correlates", *Journal of the Acoustical Society of America*, 52, 1972, 1238-1250.

8. Laukkanen, A. M., Vilkamn, E., Alku, P., Oksanen H. "Physical variations related to stress and emotional state: a preliminary study", *Journal of Phonetics*, 24, 1996, 313-335.

9. Scherer, K. R. "How emotion is expressed in speech and singing", *Proceedings of the XIIIth International Congress of Phonetic Sciences*, 3, 1995, 90-96.

10. Scherer, K. R. "Expression of emotion in voice and music", *Journal of Voice*, 9(3), 1995, 235-248.

11. Greasley, P., Setter, J., Waterman, M., Sherrard, C., Roach, P., Arnfield, S., and Horton, D. "Representation of prosodic and emotional features in a spoken language database", *Proceedings of the XIIIth International Congress of Phonetic Sciences*, 1, 1995, 242-245.

12. Johnstone, T., Banse, R., and Scherer, K. R. "Acoustic profiles of prototypical vocal expressions of emotion", *Proceedings of the International XIIIth Congress of Phonetic Sciences*, 4, 1995, 2-5.

13. Pittam, J., and Scherer, K. R. "Vocal expression and communication of emotion", In M. Lewis and J. M. Haviland (Eds.), *Handbook of Emotions*, New York: Guilford Press, New York, 1993.

14. Scherer, K. R., Banse, R., Wallbott, H. G., and Goldbeck, T. "Vocal cues in emotion encoding and decoding", *Motivation and Emotion*, 15, 1991, 123-148.

15. Koike, Y., Takahashi, H., and Calcaterra, T. C. "Acoustic measures for detecting laryngeal pathology", *Acta Otolaryngologica*, 84, 1977, 105-117.