THE VOCAL EXPRESSION OF EMOTION: AN ACOUSTIC ANALYSIS OF ANXIETY

Mark Jones, Flora Anagnostou & Jo Verhoeven

Department of Language and Communication Science, City University London, UK
Mark.Jones.1@city.ac.uk; florraan@hotmail.com; jo.verhoeven@city.ac.uk

ABSTRACT

During conversations, speakers are assumed to provide information about their emotional state. This study is a systematic investigation of the voice characteristics of the emotion of anxiety. For this purpose speech samples of 63 subjects were collected in a situation which induced mild anxiety.

On the basis of the acoustic measurements it appears that anxiety affects vocal parameters and showed a significant increase in mean F0 and in noise-to-harmonic ratio. In addition, a significant decrease in jitter and F0 range are observed, while there was no change in shimmer. These results applied equally to male and female participants.

Keywords: voice analysis, indexical information, emotion

1. INTRODUCTION

The interest in the vocal expression of emotion is a long standing one and can be traced back as far as the ancient Greeks who used several examples of how the voice should be used to express different emotions.

Anxiety is an emotion which results from a personally threatening situation, either physical or psychological, which triggers a physiological response. Feelings such as tension and nervousness together with physiological arousal (cardiovascular, digestive, endocrine systems) are associated with anxiety [4, 5, 8]). Previous studies of emotional characteristics in speech have tended to neglect anxiety in favour of more 'primary' emotions such as happiness, sadness, anger, and fear, e.g. [3]. The speech characteristics of anxiety are worthy of study because this emotional response may arise in a number of situations where voice professionals such as speech and language therapists or forensic phoneticians operate. In these situations, anxiety may act to alter particular voice characteristics which may themselves be the object of investigation for some other purpose. Knowing what characteristics anxiety may therefore allow a more realistic appraisal of the underlying voice characteristics of an individual, whether for the purposes of clinical intervention or speaker comparison.

The study of emotional voice characteristics faces a particular problem in how to collect controlled data for analysis. Actors may be utilised to simulate certain emotions while providing speech data under controlled conditions, but while these portrayals presumably owe something to the 'true' emotion, they may be affected by cultural traditions which obscure or misrepresent the true emotions. For example, an emotional state may be portrayed using exaggerated cues to ensure that the audience is in no doubt as to the actor's intentions. The method adopted here is one of mood induction, which utilises a specific task considered to place the subject in a situation in which he or she may experience a low level of stress and therefore of anxiety.

2. MATERIALS AND METHODS

Speech samples of 63 native speakers of Greek were collected. Voices were assessed subjectively and analysed objectively by means of acoustic signal processing techniques.

2.1. Participants

A total number of 63 participants were recruited as a convenience sample: they were asked individually whether they wanted to take part in the study. 35 participants were female, 28 were male. The participants ranged between 20 and 64 years of age (x = 42). Ethnic and socio-economic background were not strictly controlled, but 58 subjects were Greek, while 5 had another ethnic background. All participants were nevertheless native speakers of Greek. The voices of the participants did not show any abnormalities as assessed by means of the GRBAS (Grade, Roughness, Breathiness, Asthenia, Strain) scale [1]. The participants were randomly assigned to one of two groups: the experimental group consisted of 32 participants (15 male and 17 female) who had

undergone mood induction to invoke mild levels of anxiety previous to the recording of the speech data. The control group consisted of 31 participants (13 male and 18 female) who had not undergone such mood induction previous to the recordings.

2.2. Materials

The text that was used for speech sampling was 'The Greek Islands'. This passage was originally created to be used in laryngographic voice analysis [2] and has been used in a wide range of phonetic applications. The text contains all the Greek phonemes and phoneme combinations in a wide range of phonetic contexts. It is considerably rich lexically and grammatically and it can be easily read by the participants. Reading of the text takes approximately 2.5 minutes.

2.3. Mood induction

Several techniques have been developed to induce affective states in a controlled manner. These range from the reading of positive or negative self-statements, the use of music, the presentation of films or slides and interactive tasks and computer games (e.g. [11]). In this study the participants in the experimental group were told that they were going to have to present a brief speech while being video recorded for later analysis by a professional screening panel. It is assumed that the emotion experienced by the subjects can be classified as 'anxiety'. This type of mood induction has been used successfully previously [6, 7].

First the participants in both groups recorded a reading of the 'Greek islands' passage. After this first recording of the passage, the participants in the experimental group underwent mood-induction and were informed of the video recording task they were to perform. The experimental subjects were then asked to read the passage again just before they were supposed to give their speech. It was only after the recording that they were told that they did not have to present the speech. Although none of the participants explicitly expressed any distress in response to the task, they were given the contact details of the appropriate counselling services.

The control group participated in a second recording of the 'Greek Islands' passage without any form of mood induction.

2.4. Recordings

Recordings were made in the Voice Lab of the Athens Medical Centre-Private Hospital. The laboratory was quiet but not soundproof. All recordings were made directly onto a computer using the Kay Elemetrics CSL model 4150 and a high-quality microphone (AKG C-420PP) which was positioned a few inches away from the speaker's mouth.

2.5. Acoustic analysis

The acoustic analyses were made by means of Kay Elemetrics MDVP Advanced. The following measurements were made: F0, the frequency range, pitch perturbation (jitter), amplitude perturbation (shimmer) and the noise-to-harmonic ratio.

3. RESULTS

The different measures associated with voice quality were analyzed by means of a series of repeated measures ANOVAs with 'gender' (2 levels: male vs. female) and 'condition' (2 levels: control vs. mood-induced) between-subjects variables and 'time of recording' (2 levels: time1 vs. time 2) as the within subjects variable. All the analyses were carried out in SPSS (version 17.0).

In the first analysis, the mean F0 was the dependent variable. Given the space constraints of this paper, we will only focus on the strongest effects as expressed by Partial Squared Eta. The strongest significant effect in this analysis (Partial Squared Eta: 0.742) was the interaction between the variable 'condition' and 'time of recording' ($F(1, 59) = 169.584$, $p < 0.0001$).

This interaction indicates that F0 is significantly higher after mood induction in the experimental group. Before mood induction the mean F0 in this group was 163 Hz while it was 173 Hz after mood induction. In the control group there were no significant differences in F0 (time 1: 162 Hz vs. time 2: 159 Hz).

In the second analysis, the independent variable was the fundamental frequency range of the speakers. In this analysis, the strongest effect was the interaction between 'condition' and 'time of recording' (Partial Squared Eta = 0.79; $F(1, 59) = 5.070$, $p=0.028$). This interaction indicates that the F0 range in the experimental group is significantly different after mood induction: anxiety induces a narrowing of the F0 range (265 Hz vs. 232 Hz). In the control group the mean F0 range remains

approximately the same in the two recordings (281 vs. 280 Hz).

In the third analysis, the independent variable was pitch perturbation or jitter. This analysis revealed no significant effects nor interactions between any of the variables. This is to say that jitter is the same before and after mood induction in the experimental group (pre mood induction: 12.38% vs. post mood induction: 12.38%). The same applies to the control group (time1: 12.54% vs. time 2: 12.38%).

The fourth analysis is one of amplitude perturbation or shimmer in the two groups. This analysis showed a significant interaction between 'condition' and 'time of recording' as the most important effect (Partial Squared Eta = 0.091; $F(1.59) = 5.934$, $p = 0.018$). This interaction is such that shimmer is significantly bigger in the experimental group after mood induction: pre-induction shimmer = 1.54%, post-induction shimmer = 1.38%. In the control group shimmer stands at 1.53% for the first recording and at 1.51% for the second speech sample.

Finally, an analysis was made of noise-to-harmonics ratio (NHR). The strongest effect was the interaction between 'condition' and 'time of recording' (Partial Squared Eta = 0.077; $F(1, 59) = 4.924$, $p = 0.030$). This interaction shows that the noise to harmonics ratio is significantly higher in the experimental group after mood induction: pre-induction NHR=0.22 %, while post-induction NHR was up to 0.24 %. In the control group, NHR was 0.22 % in both recordings.

4. DISCUSSION

The results of this study offer a further demonstration that the mood induction technique does alter acoustic characteristics of speech for the purposes of controlled speech analysis. The emotion that we have called here 'anxiety' has uniform speech properties across the male and female subjects tested.

The characteristics of anxiety were a raised F0, a reduction in F0 range, a reduction in shimmer, and an increase in NHR. No significant differences in jitter were found.

In the absence of any direct investigation of laryngeal behaviour, the mechanisms underlying these changes can only be speculated about. The increase in F0 and reduction in F0 range suggests higher vocal fold tension throughout. A reduction in shimmer indicates more regular vocal fold vibration. These characteristics can perhaps be most easily linked to an increase in tension in the laryngeal musculature. An increase in NHR suggests incomplete vocal fold adduction, and this characteristic sits less well with the idea that the vocal folds are tensed. Changes in subglottal pressure (Psg) may also play a role. A higher Psg would be expected to raise F0 and also induce a longer open quotient [9], and would therefore also contribute to a greater volume velocity of airflow, increasing turbulence throughout the vocal tract. Increased tension in the intercostal muscles may therefore accompany an increase in laryngeal tension.

Impressionistically, anxiety would seem to involve a more 'robotic' voice quality as variability in F0 range and shimmer is reduced; natural sounding synthesis requires some degree of variability in voice quality.

An increase in F0 is not unexpected with some emotions. [10] found no increase in F0 in speech elicited under conditions of increased cognitive stress, although the low end of the F0 range (the F0 'floor') was reported to be higher for subjects judged to be more 'anxious' in the stress condition. The reduction in variability in F0, and the absence of any significant effect on jitter in particular, runs counter to certain stereotypical impressions of anxious or fearful speech.

Assuming that the stereotypes do indeed bear some relation to reality, one possibility to explain this discrepancy is that rather than measuring anxiety directly, as is usually assumed in such studies, these results may reflect subjects' attempts to maintain control over their voice production to eliminate any such overt markers of emotion.

While further research is needed to determine the likelihood of this strategy within the mood induction elicitation task, the results remain useful as an indicator of when a subject feels emotional pressure, even if what surfaces in the speech signal is an attempt to suppress a manifestation of that emotion. Most research on emotional cues in speech takes the view that such cues serve a communicative function which is useful to both the speaker and to the listener. As such, the speaker does not attempt to suppress or inhibit emotional markers in speech. This situation may be far from the truth. Speakers are unlikely, after all, to want to communicate all kinds of emotion in all communicative situations; in public, an outward show of some emotions may be undesirable. Far from being useful to the speaker, a show of anxiety

may allow a non-anxious listener to take advantage of a situation to the speaker's detriment.

In the case of situations in which a speech professional is engaged with the speaker for some reason, such as clinical intervention or forensic speaker comparison, knowing what counts as vocal suppression of emotion may serve a particularly useful purpose.

In the absence of further indications to the contrary, the results of this study can be interpreted as showing that anxiety has acoustic characteristics which differentiate it from other emotions.

5. CONCLUSIONS

Speech elicited from subjects under conditions likely to induce anxiety shows acoustic characteristics which can be taken as indications of greater tension in the laryngeal and respiratory musculature. These characteristics result in a higher but less variable F0, less variability in shimmer, and greater noise.

6. REFERENCES

[1] Hirano, M. 1981. *Clinical Investigation of Voice*. New York: Springer Verlag.

[2] Konstantopoulos, K. 2004. Unpublished PhD thesis.

[3] Laukka, P. 2005. Categorical expression of vocal emotion expressions. *Emotion* 5, 277-295.

[4] Leary, M.R. 1982. Social anxiety. In Wheeler, L. (ed.), *Review of Personality and Social Psychology*. Beverley Hills: Sage, 97-119.

[5] Levitt, E.E. 1980. *The Psychology of Anxiety*. Hillsdale, NJ: Lawrence Erlbaum.

[6] Reidy, J., Richards, A. 2003. *Anxiety and Explicit Memory for Emotionally Toned Words.* Unpublished manuscript.

[7] Richards, A., French, C.C., Young, A.W., Calder, A.J., Webb, B., Fox, R. 2002. Anxiety-related biases in the classification of emotionally ambiguous facial expression. *Emotion* 2, 273-287.

[8] Spielberger, C.D. 1972. Anxiety as an emotional state. In Spielberger, C.D. (ed.), *Anxiety: Current Trends in Theory and Research*. New York: Academic Press, 23-49.

[9] Titze, I.R. 1989. On the relation between subglottal pressure and fundamental frequency in phonation. *J. Acoust. Soc. Am.* 85, 901-906.

[10] Tolkmitt, F.J., Scherer, K.R. 1986. Effect of experimentally induced stress on vocal parameters. *J. Experimental Psychology: Human Perception and Performance* 12, 302-313.

[11] Velten, E. 1968. A laboratory task for induction of mood states. *Behaviour, Research, & Therapy* 6, 473-482.