ACOUSTIC EFFECTS OF AUTHENTIC AND ACTED DISTRESS ON FUNDAMENTAL FREQUENCY AND VOWEL QUALITY

Lisa Roberts

Department of Language and Linguistic Science, University of York, UK
lsr501@york.ac.uk

ABSTRACT

This study investigates variation in fundamental frequency and vowel quality from speech produced in distress and non-distress conditions by victims of genuine attacks in emergency service forensic recordings and by trained actors. Recordings from two criminal cases involving violent attack were compared with recordings from four actors re-enacting the forensic scenarios. Non-distress speech was available from the victims by chance, and from the actors by design. Findings show that: a) F0 increases and becomes more variable in distress conditions by both actors and victims; b) all speakers show a target undershoot of /i:/ realisations, though for actors this is characterized by a more open production (increase in F1) whereas in victims a retraction (a decrease in F2) is observed; and c) /əʊ/ productions appear more complex and harder to interpret, nevertheless two tentative trends - one in the form of a peripheralisation of the nuclei and the other in the form of a centralisation of the offglides - may be observed across speakers.

Keywords: forensic phonetics, vowel quality, f0, variation, emotional speech

1. INTRODUCTION

The analysis of speech and sound has an ever-increasing presence within criminal investigations. Events such as violent attacks are frequently recorded by victims, witnesses and occasionally by the perpetrators themselves. Police officers and lawyers often ask forensic speech scientists whether recordings of calls to the emergency services purporting to represent violent events are real or hoaxes, and also question whether and to what extent vocalisations in those recordings reflect real distress. Practitioners are currently discouraged from conducting such analyses by the International Association of Forensic Phonetics and Acoustics code of practice (clause 9, [3]) presumably owing to limited research in this area.

1.1. Emotional speech

Most previous studies of emotional speech have examined vocal cues to laboratory-induced emotion in the modal speech of actors (for overview see [8]). Few attempt to examine extreme emotion and fewer use authentic data [1]. Most existing studies, although insightful, have the following limitations:

1. they employ a simplistic categorisation of emotion;
2. they have limited ecological validity;
3. they may merely reflect stereotypical behaviours that actors are trained to adopt

The present research seeks to redress these limitations by analysing distress speech and vocalisations from victims in real-life emergency situations and comparing them with corresponding reference (non-distress) material from the same speakers. In addition, by contrasting these with productions from actors attempting to reproduce the same forensic material, the study investigates the relationship between authentic distress and potential stylistic emotional behaviours. By combining carefully controlled laboratory distress material with real-life 'messy' forensic data, i.e. recordings that are frequently of brief duration and variable quality - typical of data used by forensic practitioners, the present study represents a first step towards results that might ultimately be used to substantiate forensic expert opinions in this area.

2. DATA AND METHOD

2.1. Authentic data

Authentic forensic material derives from a corpus of recordings collected by J. P. French Associates[2]. Two suitable cases were selected based on the criteria that they pertain to a life-threatening attack, contain both distress and reference non-distress material from the victim, and have already been resolved in the courts.

- **Case A** concerns an audio recording of a group of three men chatting in a car before one of

them pulls out a gun and attempts to shoot the other two. The majority of the material contains non-distress conversation between the participants (the other two were not aware of the gun until the last moment). Towards the end of the recording, distress speech from the victim (Victim A), a man aged 34 years from Yorkshire, UK, is heard as he reacts to the unfolding events. A recording device was being operated in the front of the car.

- **Case B** concerns a call to the emergency services made by a 42-year man from the Midlands, UK (Victim B) from his mobile phone [3] after having been stabbed in the stomach. During the call, the attacker returns to run over the victim with a car. Non-distress material is available for Victim B from a recording of a police custody interview.

All recordings were received from JPFA in an unedited, digitised, .wav file format version of the original file (sampling rate 44.1kHz; 16 bit depth). Recordings were edited and normalised using *SoundForge 9.0*.

2.2. Acted data

Actors were recorded performing scripts based on the events in Case A and Case B during a drama workshop held at Jerwood Space Studios, London, in July 2010. Two actors worked on each script: Actors 1 and 2 worked on Case A; Actors 3 and 4 on Case B. The actors were male, aged between 23years and 28 years, and speakers of Standard Southern British English. They had all trained on a National Council Drama Training accredited programme. Each actor was recorded in two conditions - unrehearsed and rehearsed. In the latter condition, which is analysed and compared here, the script was performed after a series of vocal, physical and emotional warm ups. In their first recording session, the actors read the beginning of the standardised reading passage 'The Rainbow Passage' as control material. They were recorded individually in a small room with corkboard flooring and minimal furniture away from the main rehearsal space. Recordings were made using a head mounted DPA 4066 microphone and a Marantz PMD 670 solid-state digital machine (44.1kHz; 16 bit). Performances were also filmed using a Panasonic SDR-H90 video camcorder.

2.3. Variables

In addition to F0, the first three formants of vowel phonemes were measured. The brevity of the authentic distress material meant that the only two present in both victims' control and distress recordings were the (British) English phonemes /i:/ and /əʊ/ (the latter being realised as a monophthong in the accents represented in both authentic recordings). Between 1 and 5 tokens of these vowels were available for each of the speakers in each condition - non-distress *vs* distress.

Table 1: The number of analysable vowel tokens available per speaker and per condition. N=77

Speaker	Non-distress		Distress	
	FLEECE	GOAT	FLEECE	GOAT
Victim A	1	2	1	3
Victim B	2	5	3	4
Actor 1	4	3	1	3
Actor 2	4	3	1	3
Actor 3	4	2	4	5
Actor 4	4	2	3	5

2.4. Analysis

Normalised digital recordings were analysed in *Praat v5.1.25*. Since the in-built pitch tracker at times struggled to find the true F0, the contour was checked and corrected using *Praat*'s pitch object facility. Vowel formants were measured using a combination of spectrographic displays and spectral (FFT) slices. /i:/ tokens, a spectral slice was produced from a selection of the vowel during which F1 was stable at its maximum and F2 stable at its minimum. For /əʊ/ tokens, a spectral slice was produced of the vowel nucleus, roughly defined as being the location of stable F1 and F2 as soon after the consonant transition, and the vowel off-glide, taken at F1 and F2 ultimate values before fading. The duration of the selection for each spectral slice was typically one period [1].

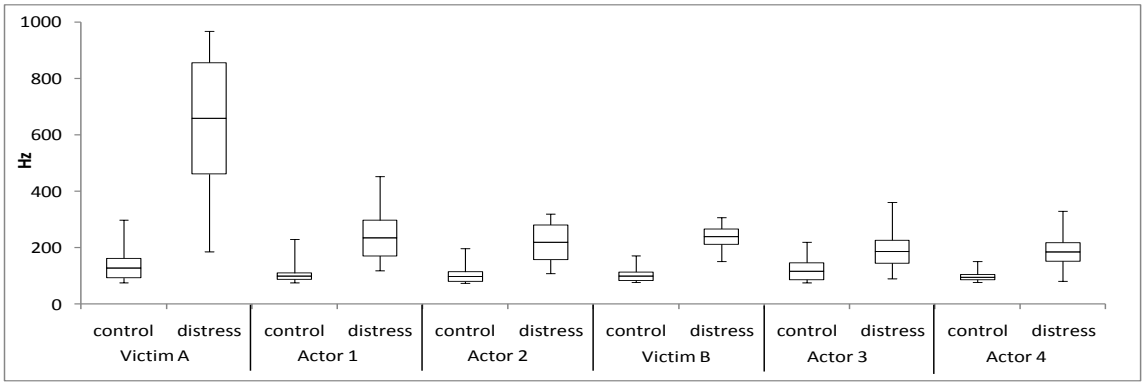
3. RESULTS AND DISCUSSION

3.1. F0

In line with previous studies, e.g. [7, 9], F0 was shown to increase in distress conditions by all speakers (Fig. 1). The extent of the increase varies across individuals. Within-speaker variability of F0, calculated as absolute range and standard deviation, is also seen to increase, as found by [9]. Victim A ranges from 185Hz to 968Hz when vocalising in distress,[4] whereas Victim B shows an F0 range more typical with those from the actors,

roughly 150Hz to 350Hz. A possible explanation is that although potentially fatal attacks may lead to a uniform physiological response [4], the victim's level of injury and pain and his/her psychological evaluation of the event may mediate that response, leading to cross-individual variation [6].

Figure 1: Boxplots showing absolute F0 mean, s.d., min and max for victims and actors in control and distress conditions.



Physiologically, the increase in F0 may arise from a tensing of the laryngeal musculature and vocal folds. Impressionistic voice quality analysis of distress productions with increased F0 support this view. Interestingly, preliminary results from the same data show that increased vocal effort, airflow and intensity may not invariably accompany the increase in laryngeal tension.

3.2. Vowel variables

3.2.1. /i:/

Victims' productions of /iː/ in distress could be characterised as retracted with a decrease in F2 but F1 remaining stable (Fig. 2). For both victims, /iː/ realisations changed from [i] to [i̟]. For actors, however, distress productions of /iː/ tended to be more open with an increase in F1, whereas F2 typically remained much the same (Fig. 3). One exception (Actor 3) had variable distress productions in no one direction. On the whole, /iː/ productions were [i] to [i̟/e].

Figure 2: Scatterplot of Victim A /iː/ vowels showing a target undershoot of the front-back dimension as demonstrated by both victims.

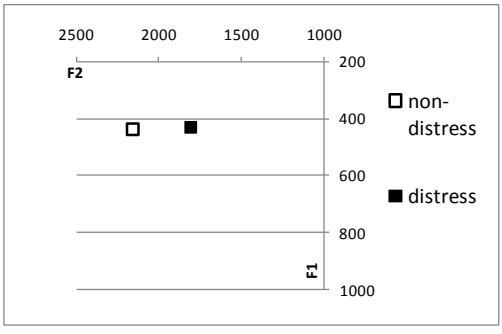
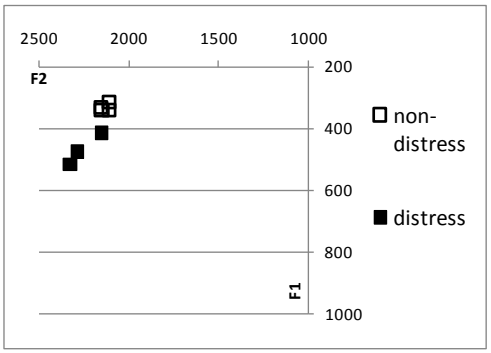


Figure 3: Scatterplot of Actor 4 /iː/ vowels showing a target undershoot of the height dimension as found in most actors.



3.2.2. /əʊ/

With regard to /əʊ/, the tendencies are less clear. No tendency appears to distinguish actors from victims, though two patterns emerge (Figs. 4, 5). The first demonstrates a peripheralisation of the nucleus of a distress token while moving towards the same target off-glide as non-distress productions (as found in Victim A and Actors 3 and 4). A lengthening of the glide can also be observed in this figure though this was not repeated by other speakers. The second shows an alternative pattern, as produced by the remaining speakers, with a centralisation of both nucleus and off-glide targets, as well as a shortening of the glide.

At a very general level, the picture to emerge here is that of speakers undershooting phonetic targets, i.e., contracting the vowel space. Although not directly comparable with the present data, previous studies involving cognitive stress and speaking in noise have also found an

undershooting of phonetic targets [2] and a centralisation of the vowel space [5]. The first is present in both actors and victims, albeit in different directions. The increase in F1 in actors' /i:/ vowel productions may be linked to the overall increase in F0, as well as with overall jaw opening and lowering. The decrease in F2 for victims' /i:/ vowels, although another form of target undershoot, remains harder to explain. Whereas actors are trained to maximise intelligibility while performing, e.g., in terms of increased amplitude and exaggerated jaw movements, thus accounting for some of the non stress-induced causes of F1 increase, victims, of course, receive no such training. The second is partially present in /əʊ/ distress productions with some speakers producing centralised nuclei and off-glide targets, and may be considered a two-pronged form of phonetic undershoot.

Figure 4: Scatterplot of Actor 3's averaged /əʊ/ nuclei and off-glide positions for distress and non-distress productions representing a nucleus peripheralisation pattern.

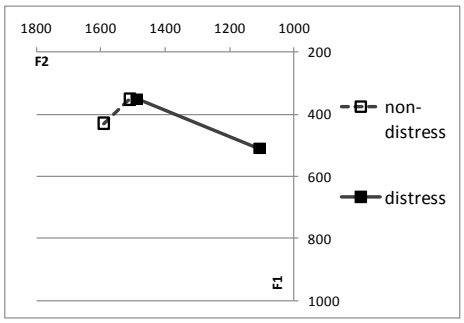
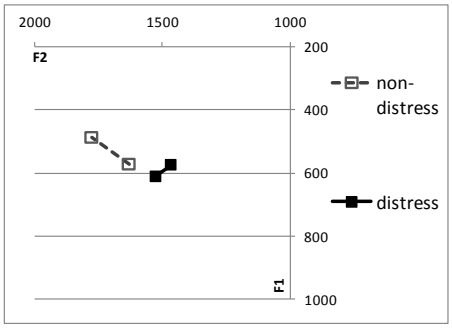


Figure 5: Scatterplot of Actor 1's averaged /əʊ/ nuclei and off-glide positions for distress and non-distress productions representing a centralised pattern.



4. CONCLUSION

Although limited by the brevity and quality of authentic forensic data, observations concerning acoustic differences of distress and non-distress speech show: F0 increases and becomes more variable in distress, and vowel targets may be prone to undershooting. Distinctions between acted and authentic responses of distress merit further investigation but preliminary findings show a discriminating factor in the nature of /i:/ phonetic target undershooting.

5. REFERENCES

[1] Duckworth, M., McDougall, K., de Jong, G., Shockey, L. 2007. The reliability of formant measurements in high quality audio data - the effect of agreeing measurement procedures. *International Association of Forensic Phonetics and Acoustic 16th Annual Conference*, Plymouth.

[2] Hecker, M.H.L., Stevens, K.N., von Bismarck, G., Williams, C.E. 1968. Manifestations of task-induced stress in the acoustic speech signal. *Journal of the Acoustical Society of America* 44, 993- 1001.

[3] International Association of Forensic Phonetics and Acoustics - Code of Practice. *http://www.iafpa.net/code.html*

[4] Jessen, M. 2006. *Einfluss von Stress auf Sprache und Stimme. Unter Besonderer Beruecksichtigung Polizeidienstlicher Anforderungen.* Idstein: Schulz-Kirchner Verlag GmbH.

[5] Karlsson, I., Banziger, T., Dankovicová, J., Johnstone, T., Lindberg, J., Melin, H., Nolan, F., Scherer, K. 2000. Speaker verification with elicited speaking styles in the VeriVox project. *Speech Communication* 31, 121-129.

[6] Kirchhübel, C., Howard, D.M., Stedmon, A. In press. Acoustic correlates of speech when under stress: Research, methods and future directions. *International Journal of Speech, Language and Law* 18(1).

[7] Kuroda, I., Fujiwara, O., Okamura, N., Utsuki, N. 1976. Method for determining pilot stress through analysis of voice communication. *Aviation, Space and Environmental Medicine* 47, 528-533.

[8] Scherer, K.R. 2003. Vocal communication of emotion: A review of research paradigms. *Speech Communication* 40, 227-256.

[9] Williams, C.E., Stevens, K.N. 1969. On determining the emotional state of pilots during flight: An exploratory study. *Aerospace Medicine* 40, 1369-1372.

[1] Exceptions are studies reporting on stress in the speech of aviation personnel, e.g. [7, 9].

[2] A forensic speech/audio laboratory based in York, UK.

[3] The frequency bandwidth and other limitations of mobile phone transmitted speech have been taken into account in the analysis.

[4] The possibility of Victim A's 968 Hz reading being laryngeal whistle has been considered and rejected as the F0 curves show the contour falling uninterruptedly from this value to 519Hz without any perceptible change in phonatory quality. The latter value is well within the pitch range found elsewhere in the distress material from this victim.