Real and acted responses of distress: an auditory & acoustic analysis of extreme stress & emotion

Lisa Roberts
Department of Language and Linguistic Science, University of York, U.K.

Abstract
The study explores vocal responses of distress in authentic and acted forensic situations. Data is drawn from an archive of forensic recordings and comparable scenarios involving actors. The material is subject to acoustic and auditory-phonetic analyses. Results are presented on parameters including fundamental frequency (F₀), tempo, intensity, and vowel quality. Findings show that victims and actors in distress generally demonstrate an increase in F₀ and intensity overall, an increase in variability of F₀ and intensity, and a decrease in tempo. Actors' distress responses differ from those from real victims in vowel quality and their exaggerated increase in F₀.

Key words: forensic phonetics, distress speech, acted responses.

Introduction
Previous studies investigating emotional speech have concerned speech technology applications (Erickson, 2005) and/or observed vocal cues of emotion in everyday speech using actors (Scherer, 2003). Few attempt to characterise extreme emotion and fewer use authentic data. The present study represents a first step towards results that might ultimately be used to substantiate forensic expert opinions in this area.

Data and Method
Data relates to victims of violent attack where an attack is imminent, completed or in progress. Authentic material was drawn from eight forensic cases (victims: 4 male, 4 female; aged 15-47). All but one were calls to the emergency services. 9 actors (6 male, 3 female; aged 27-55) were chosen from filmed material, e.g. Reservoir Dogs, A Clockwork Orange etc. Injuries common to both victims and actors included stabbings, shootings, burnings and being chased. All material was digitised (44.1kHz, 16 bit depth). Auditory and acoustic analyses were conducted using SoundForge and Praat.

Findings
Due to the high amount of inter- and intra-individual variation, a four-way taxonomy was employed to compare like productions across all data.

Proceedings of the 3rd ISCA Workshop ExLing 2010, 25-27 August, Athens, Greece
Table 1. Summary of data classification system.

<table>
<thead>
<tr>
<th>Category</th>
<th>Control data</th>
<th>Distress speech</th>
<th>Distress vocalisation</th>
<th>Scream</th>
</tr>
</thead>
<tbody>
<tr>
<td>Features</td>
<td>non-emotionally</td>
<td>intelligible, emotionally</td>
<td>emotionally aroused production,</td>
<td>emotionally aroused production,</td>
</tr>
<tr>
<td></td>
<td>aroused speech,</td>
<td>aroused speech</td>
<td>questionable intelligibility &amp;</td>
<td>devoid of linguistic content</td>
</tr>
<tr>
<td></td>
<td>e.g. regular</td>
<td></td>
<td>linguistic content</td>
<td></td>
</tr>
<tr>
<td></td>
<td>conversation</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**F₀**

For both acted and authentic data, distress productions typically had a higher \( F₀ \) than the individual’s control data/expected range, with productions increasing in \( F₀ \) as individuals progress through the taxonomy. As expected, females tend to have a higher \( F₀ \) than males, though productions classified as vocalisations and screams tend to be beyond expected ranges for both sexes. Distress vocalisations and screams often had a rise-fall \( F₀ \) contour. These tokens contained voicing irregularities and breaks in phonation. Pitch jumps also occurred, e.g. from 510Hz to 1270Hz (vocalisation from a female victim) and from 559Hz to 810Hz (distress speech from a male actor).

![Figure 4: Graph showing average F₀ (Hz) data for victims and actors.](image)

**Intensity**

On the whole, levels of intensity mimicked F₀ contours, presumably as a result of increased \( F₀ \). A plateau effect was observed in some recordings though this is more likely attributable to the limitations of the recording equipment (e.g. automatic gain control) rather than the individuals’ physiology. However, results were variable across the data, for two possible reasons: first, in the authentic data, the distance and orientation of the
An auditory and acoustic analysis of extreme stress and emotion

individual from the microphone is unknown; secondly, in the acted data, it is uncertain if sound editing took place post-production.

Vowel quality
Measurements of F1 and F2 were taken where vowels in the distress speech condition could be compared with control data, i.e. in all acted data but in only two cases of the authentic data. In the former, a contraction of the vowel space was seen; in the latter, distress vowels tended to show a shifting of the vowel space towards the front dimension. Both patterns can be associated with Lombard speech.

Tempo
Articulation rates decreased as individuals progressed through the categories of the taxonomy. Male actors had a slower AR than male victims whereas female actors tended to have a faster AR than female victims.

Discussion
Acoustic changes in control and distress conditions may in part be caused by unconscious physiological responses to stress, namely changes in respiration rates and muscle tension (Hansen & Patil, 2007), giving rise to a faster tempo, an increase in intensity, and a tensing of the vocal folds with a consequent increase in F0.

Sources of variation between the data sets may be both external, e.g. due to different injuries and circumstances surrounding the victim/actor, or internal, since individuals may appraise sources of stress differently (Andreassi, 2007).

A further consideration is to what extent are distress productions in victims voluntary? It is not always clear if linguistic content is present in the productions. Involuntary responses are presumably purely physiological with
little variation. Voluntary responses, on the other hand, will most likely be greater in range. Whereas distress responses are similar in both data sets, actors may be reflecting an awareness of both a physiological response and the presence of variation amongst victims. In some cases, they seek to exaggerate these features for the benefit of the audience. Where responses differ across the data sets, actors may themselves be unsure of what they are trying to emulate.

Summary and Conclusions
Tendencies common to both real and acted distress responses are an increase in overall $F_0$ and intensity, an increase in variability of $F_0$ and intensity, with a decrease in speech tempo. Acted distress responses tend to use a contracted version of the vowel space, whereas authentic responses see an increase in $F1$ and $F2$ leading towards a shift in the front dimension. Acted distress responses differentiate themselves not only in regard to vowel quality, but also in the extent in which the $F_0$ increases – acted responses may be viewed as exaggerated versions of authentic distress responses.

Acknowledgements
This research was supported by the UK Economic and Social Research Council (PhD +3 studentship). I am very grateful to Paul Foulkes, Peter French, and Dominic Watt for their contributions, and especially J. P. French Associates for authentic forensic audio access and advice.

References