F0 STATISTICS FOR 100 YOUNG MALE SPEAKERS OF STANDARD SOUTHERN BRITISH ENGLISH

Toby Hudson[1], Gea de Jong[1], Kirsty McDougall[1], Philip Harrison[2] and Francis Nolan[1]

[1]Department of Linguistics, University of Cambridge
[2]JP French Associates / Department of Language & Linguistic Science, University of York
[1]toh22|gd288|kem37|fjn1@cam.ac.uk, [2]pth@jpfrench.com

ABSTRACT

This paper presents statistical data for the fundamental frequency of 100 young male speakers of Standard Southern British English producing spontaneous speech under cognitive stress. The material comes from the new DyViS database, for which subjects underwent a simulated police interview. The distribution of F0 in a large homogeneous group of speakers is of forensic significance since it provides a framework for understanding the significance of F0 measurements in casework. Long-term F0 for the 100 speakers yielded a mode of 102 Hz, a mean of 106 Hz and a median of 105 Hz, and had a near-normal distribution. We demonstrate the limitations of F0 as a discriminatory feature for the majority (60%) of our speaker group, whose long-term F0 occurred within a narrow window of 20 Hz. Conversely, we see the forensic implications for recordings where a speaker's F0 is outside this window.

Keywords: fundamental frequency, F0, variation, SSBE, forensic phonetics.

1. INTRODUCTION

Long-term fundamental frequency (F0) is a feature commonly used in forensic cases of speaker identification. It has received attention from a number of phoneticians over the years [e.g. 9, 5, 3, 2]. Although it is highly variable within a single speaker – at the mercy of emotion, state of health, time of day and loudness amongst other factors – and may even be altogether masked in imitation or disguise, mean and standard deviation of F0 have played a key role in many cases [9: 124]. Its robustness is due to the facts that it is to some extent anatomically determined, that it is relatively undisturbed by background noise, and that its measurement is unaffected by telephone transmission [2]. However, interpretation of F0 measurements is dependent on population data for inter- and intra-speaker variation. The newly created DyViS database provides a sizeable representation of a single English-speaking speech community, of 100 young male speakers of Standard Southern British English (SSBE), undertaking the same vocal tasks. The task analysed in the present study is a simulated police interview. The material is therefore spontaneous speech elicited under cognitive stress not dissimilar to the forensic scene. This paper reports descriptive statistics for an investigation into F0 using this database.

2. METHOD

100 subjects were recorded at a sampling rate of 44.1 Hz on a Marantz PMD670 portable solid state recorder in a sound-treated studio. These were speakers of SSBE aged 18-25 years. Each subject was seated about 20 cm in front of a Sennheiser ME64-K6 cardioid condenser microphone. The audio format employed was Microsoft Wave.

The recordings are simulated police interviews of approximately 15-25 minutes. Further details about the content of the database and elicitation techniques are given in [10]. By means of a *Praat* script [1] and further manual editing, 3-5 minutes (according to quantity available) of continuous speech were extracted for each subject (the interviewer's voice was eliminated along with laughter and other intrusive sounds, whispered speech and much of the silence). The speech was extracted from the end of each interview since it is possible that the speaker's F0 (amongst other characteristics) settles down as he moves from responding to quick-fire answers to more lengthy interrogation. In all cases this is more time than is necessary: around one minute has been shown to be the required minimum time frame [9: 123]. The audio files were analysed using a long-term pitch analysis *Praat* script which gave as its output F0 mean, standard deviation and median for each speaker. It also generated the distribution of F0 for each recording using 2.5 Hz bins within a 50-

300 Hz range (avoiding undesirable low frequencies but giving a very generous upper limit), from which a mode for each speaker was obtained.

Two later adjustments were made. First, a histogram showing the mode revealed that three speakers had apparent modes of 51.25, 58.75 and 63.75 Hz, and correspondingly very different means: 91, 110 and 120 Hz respectively. Histograms of the F0 distribution for these three individuals showed each distribution to be bimodal, which was understood to be due to a notable amount of creaky voice. Therefore the first peak for these subjects was rejected, and the second was adopted as the relevant mode. The second adjustment to the output was to double the bin size universally to 5 Hz in order to smooth out perturbations.

3. RESULTS & DISCUSSION

The mode F0 for each individual is shown in Fig. 1, the mean in Fig. 2.

The mean of the modes, as well as the mean of the means and the mean of the medians, was calculated to give an impression of the central tendency of F0 for this speech group. The three measures were found to be within 5 Hz of each other: 102.2, 106 and 105 Hz respectively. This compares well to [4] where a mean F0 of 101 Hz for British English conversation is quoted and [8] which presents a mean of 105.2 Hz for young adult male speakers of Australian English. The cumulative percentages for the mode and mean for the DyViS data is presented in Fig. 3.

The curves in Fig. 3 are steep in their central section but have a gentle gradient at both ends. Observing the spread of the modes, we see that the lowest 20% of the speakers had an F0 of under 94 Hz and the highest 20% an F0 above 113 Hz. This leaves a notably narrow intervening band of only 19 Hz in which the majority of the speakers are found (where the trajectory on the cumulative graph is particularly steep). This is comparable to a range of 27 Hz for 60% of the German speakers in [6]. For the means the pattern is much the same: the bottom 20% of speakers have a mean F0 below 99 Hz, the top 20% have a mean above 120 Hz, with 21 Hz intervening. We would therefore state that F0 does not discriminate very well the majority of speakers – namely the 60% with F0 modes in the central area. For those speakers, however, who fall in the outlying areas, F0 is a more salient discriminatory variable.

There is little skewing of the *overall* distribution of the group of speakers; the very slight positive skewing for *individual* speakers is demonstrated by the calculation of mean minus median for each speaker. This is on average only 1 Hz, but 76 of the 100 subjects had a positive skewing, which was within the range of 1-8 Hz. Since it is a characteristic of modal voice that a speaker's F0 is typically found at the lower end of his frequency range, it is not surprising to see a drop at this end and a 'tapering off' in the higher frequencies. This also explains why the mean is slightly higher than the mode throughout as it is 'pulled up' by the higher values. The mode therefore gives a truer indication of central tendency; but one should also be aware that the specific value of the mode will always depend on the bin size.

Our averages are somewhat dissimilar to those of Künzel [6] who reports a mean of 115.8 Hz for 105 German male speakers, and Lindh [7] who reports a mean of 120.8 Hz for 109 Swedish male speakers from the Swedia database. However, in the former case the subjects were engaged in reading, which tends to elicit larger pitch excursions than spontaneous speech. Johns-Lewis [4] reports an average of 128 Hz for British English reading (and 142 Hz for acting). Nevertheless the Swedish 120.8 Hz comes from spontaneous speech. This may be slightly too high a value: Lindh himself refers to octave jumps on the part of the pitch tracker and measurement errors, and suggests that the mean of the medians, 115.8 Hz, may offer a more accurate estimate. The difference between the DyViS statistics and those of [7] is likely to be due to a difference in emotional states, physiological factors, or a code-specific difference. Van Bezooijen [12] has demonstrated on the one hand a similar average F0 for Japanese and Dutch women, but, on the other, a likely correlation between pitch and stereotypical feminine characteristics within each of the two groups. Although it is beyond the scope of this paper to investigate a possible agreement on pitch within a linguistic community, we should note the possibility that SSBE men may be conforming to a perceived social ideal. On more solid ground, the differences in the intonational patterns of English and Swedish, along with the presence of lexical tone in Swedish, could be a factor behind the disparity of F0.

Figure 1: Histogram showing mode F0 distribution for 100 male speakers of SSBE aged 18-25 years, using 3-5 minutes of spontaneous speech per speaker from the DyViS database.

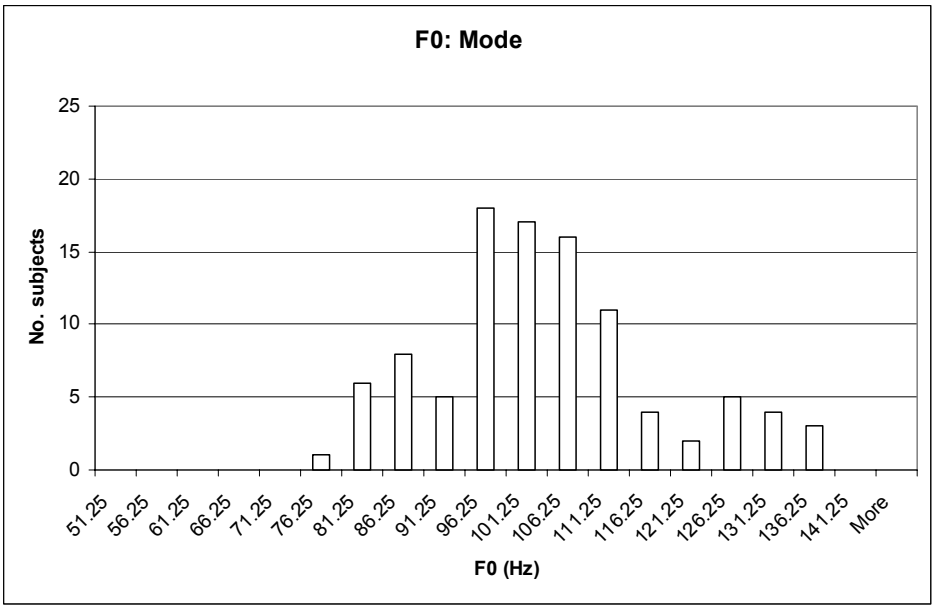
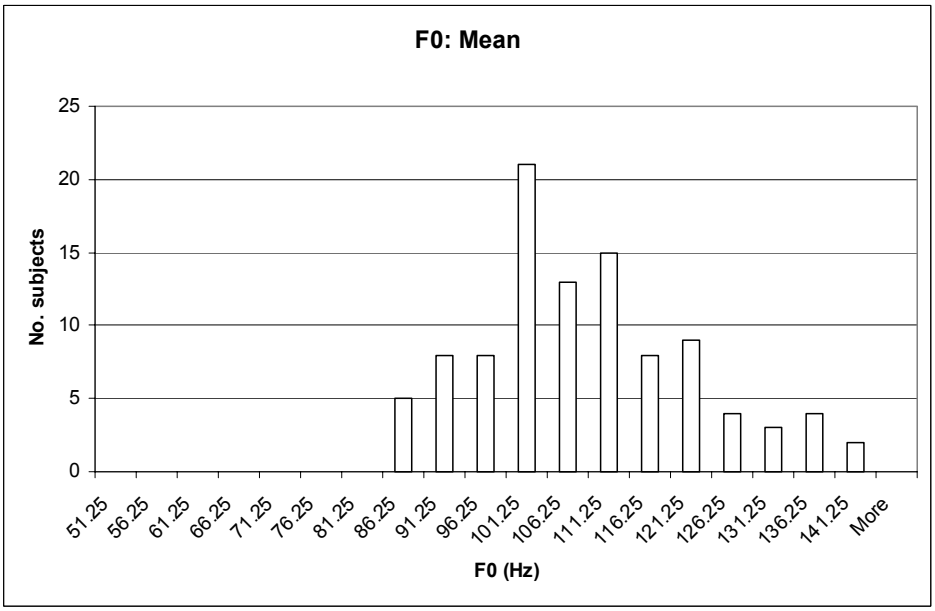


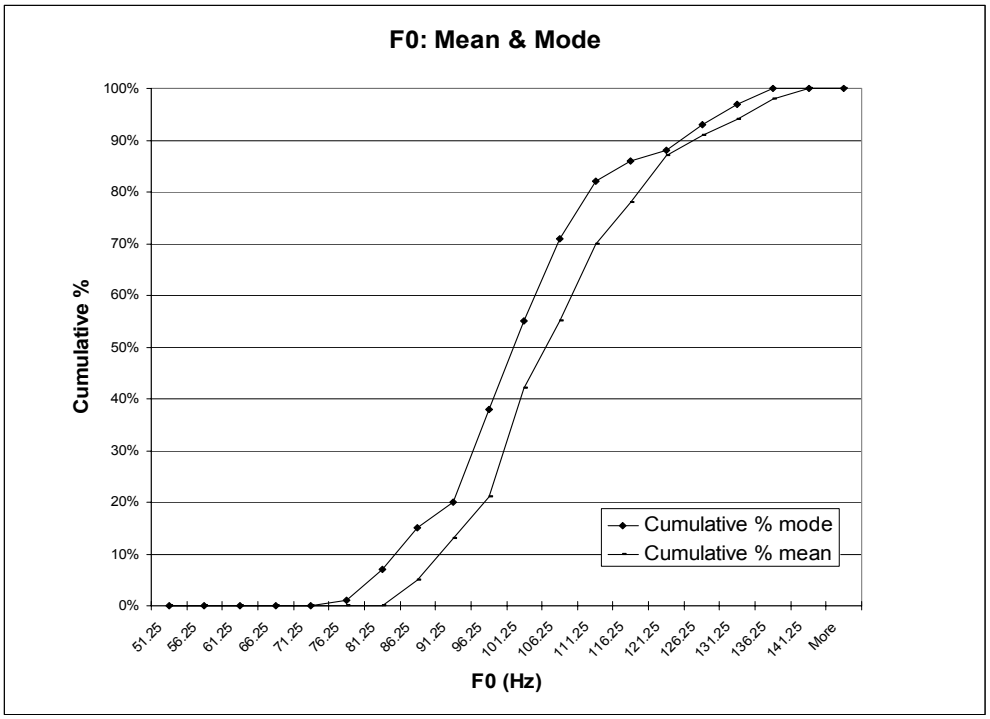
Figure 2: Histogram showing mean F0 distribution for 100 male speakers of SSBE aged 18-25 years, using 3-5 minutes of spontaneous speech per speaker from the DyViS database.



4. CONCLUSION

This study provides the forensic community with F0 information for a homogeneous English-speaking speech group, speaking under cognitive stress, of a format useful for extrapolating population data. The overall spread of F0 for our 100 young male speakers of SSBE is near to normally distributed. The average mode, mean and median are 102, 106 and 105 Hz respectively. A slight positive skewing in the *individual* F0 distributions is due to the lack of an upper limit – that is, higher frequencies pulling up the mean. 60% of the speakers' modes came within approximately a 20 Hz range; the lowest modeoverall was 76 Hz, the highest 136 Hz, giving a range of 60 Hz. Our findings are consistent with previous data, either finding confirmation or diverging with reason.

Figure 3: Cumulative percentages of mean and mode F0 for 100 male speakers of SSBE aged 18-25 years, using 3-5 minutes of spontaneous speech per speaker from the DyViS database.



This study is a platform for further investigation. In the first instance, this will be to examine within-speaker F0 distributions with the current data. Since the same speakers are recorded on the database engaging in other speaking styles on the same occasion, namely a familiar conversational style and reading a passage and disconnected sentences, it is our intention to compare F0 of these styles with that of the speech produced under cognitive stress in the present study.

5. ACKNOWLEDGEMENTS

This research is supported by the UK Economic and Social Research Council as part of the project 'Dynamic Variability in Speech [DyViS]: A Forensic Phonetic Study of British English' [RES-000-23-1248].

6. REFERENCES

[1] Boersma, P., Weenink, D. 2005. *Praat: Doing Phonetics by Computer*. <http://www.praat.org/>.

[2] Braun, A. 1995. Fundamental frequency - how speaker-specific is it? In: Braun, A., Köster, J.-P. (eds.), *Studies in Forensic Phonetics: BEIPHOL* 64, 9-23.

[3] Hollien, H. 1990. *The Acoustics of Crime*. New York: Plenum Press.

[4] Johns-Lewis, C. 1986. Prosodic differentiation of discourse modes. In: Johns-Lewis, C. (ed.) *Intonation in Discourse*. London: Croom Helm. 199-219.

[5] Künzel, H.J. 1987. *Sprechererkennung. Grundzüge forensischer Sprachverarbeitung*. Heidelberg: Kriminalistik Verlag.

[6] Künzel, H.J. 1989. How well does average fundamental frequency correlate with speaker height and weight? *Phonetica* 46: 117-125.

[7] Lindh, J. 2006. Preliminary descriptive F0-statistics for young male speakers. *Lund Univ. Working Papers* 52, 89-92.

[8] Loakes, D. 2006. Variation in long-term fundamental frequency: measurements from vocalic segments in twins' speech. *Proc. 11th SST*. Auckland. 205-210. <http://www.assta.org/sst/2006/>.

[9] Nolan, F. 1983. *The Phonetic Bases of Speaker Recognition*. Cambridge: Cambridge University Press.

[10] Nolan, F. McDougall, K., de Jong, G., Hudson, T. 2006. A forensic phonetic study of 'dynamic' sources of variability in speech: the DyViS project. *Proc. 11th SST*. Auckland. 13-18. <http://www.assta.org/sst/2006/>.

[11] Traunmüller, H., Eriksson, A. 1995. The frequency range of the voice fundamental in the speech of male and female adults (unpublished manuscript).<www.ling.su.se/staff/hartmut/aktupub.htm>. Visited 1-Mar-07.

[12] van Bezooijen, R. 1995. Sociocultural aspects of pitch differences between Japanese and Dutch women. *Language and Speech* 38(3): 253-265.