

## **Preliminary F0 statistics and forensic phonetics**

*Jonas Lindh*

*Department of Linguistics, Göteborg University, Sweden*

*jonas.lindh@ling.gu.se*

Fundamental frequency has been used for a long time in speaker identification (Braun, 1995; Rose, 2003). The within-speaker variation in F0 is affected by several factors. In Braun (1995), they are categorized as technical, physiological and psychological factors. Tape speed, which surprisingly still is an issue for forensic samples, and sample size are examples of technical factors. Smoking and age are examples of physiological factors, while emotional state and background noise are examples of psychological factors. However, fundamental frequency has been shown to be a successful forensic phonetic parameter (Nolan, 1983). To be able to study differences it is suggested to use long-term distribution measures such as arithmetical mean and standard deviation (Rose, 2002). The duration of the samples should be more than 60 seconds according to Nolan (1983), but Rose (1991) reports that F0 measurements for seven Chinese speakers stabilized much earlier, implying that the values may be language specific (Rose, 2002). Positive skewing of the F0 distribution for a speaker is typical (Jassem et al., 1973).

When extracting fundamental frequency information for 109 young male Swedish speakers, it was discovered that octave jumps and other analysis errors were frequent. Manual examination of the results revealed that using median values instead of means was more reliable as it is less sensitive to outliers. In forensic cases the questioned recording and reference are often very different both with respect to speech style and audio quality. Comparing F0 mean in these cases can therefore be problematic.

Since the F0 mean is heavily influenced by paralinguistic variation such as emotions etc. it has been suggested that the average base value ( $F_b$ ) for F0 represents individual speaker characteristics in a better way (Traunmüller, 1994). Instead of using the mean to calculate the baseline, as suggested by Traunmüller, it seems safer in most cases to use the median for reasons mentioned above. The baseline is in 'normal' speech situated 1.43 standard deviations below the mean. This measure is assumed to be more closely related to the carrier frequency and thereby possibly to a given speaker's voice characteristics.

There are no summary statistics for F0 available for Swedish speakers after Kitzing (1979), who reports a mean of 110.3 Hz and a standard deviation of 3 semitones (reported in Traunmüller & Eriksson, 1995:A) for 51 male speakers ranging between 21 and 70 years of age.

The base value is here described together with mean, median and standard deviation for the whole group.

### **Method**

Young males aged 20–30 were chosen because this age group is represented in the Swedia database and because 62% of the convicted criminals in Sweden last year belong to that age group. (<<http://www.bra.se>>).

The software Praat (Boersma & Weenink, 2005) was used to collect F0 data from 109 young male speakers in the Swedia database (<<http://www.swedia.nu>>) and the durations of the recordings range from 17.4 to 116.8 seconds per speaker with a mean duration of 52.3 and standard deviation of 15.2. The parameters extracted from the recordings were F0 mean, median, average baseline value ( $F_b$ ), standard deviation, maximum and minimum in Hz. The range for the F0 tracker was set to 75–350 Hz to be able to cover all possible frequency excursions while at the same time avoiding octave jumps.

## **Conclusions and future work**

The recordings analyzed in this study provide preliminary statistics for the distribution of mean fundamental frequency, base value and standard deviation for young Swedish males. Approximately 65% of the speakers have a mean fundamental frequency between 100 and 130 Hz. The mean of the means is 120.8 Hz. There is a positive skewing (0.6) with five extreme outliers between 150 and 170 Hz. The automatic analysis has a tendency of making positive octave jumps. The median distribution shows the same positive skewing as that for the means (0.6), but the mean (of the medians) is lower (115.8 Hz). Approximately 68% of the speakers have a median F0 of 100–130 Hz. The results suggest the use of the more robust median instead of the mean, since octave jumps influence the arithmetical mean. The mean baseline is 86.3 Hz and 53 % of the speakers have a base value between 80 and 100 Hz. 56% of the speakers' standard deviations fall in the 15–25 Hz range. If expressed in semitones, 59% of the standard deviations fall in the 2.5–3.5 semitone range. The figures are strongly influenced by measurement errors. This resulted in a few extreme outliers.

These preliminary results suggest the baseline as a more representative measure for speaker identification because the distribution is less skewed than the mean. About 30 % of the base lines calculated from data used in the present study fell in the 90–100 Hz interval, which agrees well with the calculated 93.4 Hz per balanced speaker of European Languages reported by Traunmüller and Eriksson (1995:B).

To be able to study between-speaker differences better, intra-speaker F0 distribution should be compared and studied using different measures.

## **References**

- Boersma, P. & Weenink, D. (2005). Praat: doing phonetics by computer (Version 4.3.27) [Computer program] Retrieved October 7, 2005, from <http://www.praat.org/>
- Braun, A. (1995). Fundamental frequency – how speaker-specific is it? In A. Braun & J.-P. Köster (Eds.), *Studies in Forensic Phonetics*, 9–23. Trier: Wissenschaftlicher Verlag Trier.
- Brottsförebyggande Rådet: [www] Retrieved November 26, 2005, from <<http://www.bra.se/>>
- Jassem, W., Steffen-Batog, S., and Czajka, M. (1973). Statistical characteristics short-term average F0 distributions as personal voice features, in W. Jassem (Ed.) *Speech Analysis and Synthesis*, Vol. 3, 209–225. Warsaw: Polish Academy of Science.
- Kitzing, P. (1979). *Glottografisk frekvensindikering: En undersökningsmetod för mätning av röstläge och röstomfång samt framställning av röstfrekvensdistributionen*. Malmö: Lund University
- Nolan, F. (1983). *The Phonetic Bases of Speaker Recognition*, Cambridge: Cambridge University Press.
- Traunmüller, H. (1994) Conventional, biological, and environmental factors in speech communication: A modulation theory, *Phonetica*, **51**, 170–183.
- Traunmüller, H. & Eriksson, A. (1995:A). The frequency range of the voice fundamental in the speech of male and female adults. Unpublished Manuscript (can be retrieved from <http://www.ling.su.se/staff/hartmut/aktupub.htm>)
- Traunmüller, H. & Eriksson, A. (1995:B). The perceptual evaluation of F0-excursions in speech as evidenced in liveliness estimations. *J. Acoust. Soc. Am.*, **97**, 1905–1915.
- Rose, P. (1991). How effective are long term mean and standard deviation as normalisation parameters for tonal fundamental frequency?, *Speech Communication*, **10**, 229–247
- Rose, P. (2002). *Forensic Speaker Identification*. New York, Taylor & Francis.