Automatic acquisition of jitter and shimmer measurements across large sets of sustained vowel productions

Fredrik Karlsson & Jan van Doorn
Umeå University, Sweden

Introduction
- Measurements of jitter and shimmer are frequently employed to quantify laryngeal control and stability during voice production in patients.
- Estimates of jitter and shimmer may be obtained from acoustic recordings of sustained vowels produced by patients using computer software implementing algorithms for their extraction (e.g. Praat MDVP).
- The jitter and shimmer algorithms do not, in themselves, exclude non-stable productions and are influenced by the inclusion of silence in the analyzed intervals.
- As a consequence, reliable measurements of jitter and shimmer are made by manually opening each sound file and selecting an interval for the computations.

The method
- The first step of the method outlined here attempts to find the sustained vowel production among other sounds possibly recorded within the same acoustic recording, and to extract start and end times for a discovered vowel.
- Rationale of the first step of the procedure
  - If the vowel is not voiced, jitter measurements should not be conducted. Portions with no voicing should therefore not be considered.
  - A recording is likely to be conducted with the microphone closest to the desired source, increasing the likelihood of that source being prominent in the recording.
  - A sustained vowel is likely to be less variable in pitch compared to other productions.
  - A sustained vowel is likely to have a comparably long duration compared to other productions with low pitch variability.
- In a second step, jitter and shimmer estimates are extracted in a way that affords estimation of the robustness of the obtained measurements and their interpretation.

Evaluation of the vowel extraction method
- The results obtained from the automatic procedure were evaluated through applying it to 1271 sound files:
  - 1253 containing a sustained vowel production
  - 18 files not containing recorded speech
- The results were compared to that of manually corrected version of the vowel identification output

Results
- Of the 1253 sound files with vowels, 982 were left unaltered in the manual control procedure
- 271 files were subject to some adjustment in vowel start and end times (mean adjustments = 0.23s)
  - Mainly due to short but highly stable pitch portions (249 productions)
  - Overlapping speech (leading into the vowel production, 22 productions)

Conclusion
- The proposed extraction method provides a reasonable robust way to isolate sustained vowels from rival sounds in an acoustic recording
- The procedure is sensitive to short voiced portions with very stable pitch
  - Which are easily identified by their short durations
- Simultaneous speech may seriously interfere with the localisation of the produced vowel.
  - Don't talk while your patient is being recorded!