# Contour clustering: a tool for exploring prototypical f0 patterns

Constantijn Kaland
*Institute of Linguistics, University of Cologne*

This work presents an automatic data-driven analysis for describing prototypical f0 patterns. This is particularly suitable as an exploratory tool in initial stages of prosodic research and language description. The approach has several advantages over traditional ways to investigate prosody and intonation, which are sometimes based on auditory impressions or limited empirical research to support phonological claims. Contour clustering is applicable to spontaneous and scripted speech of any language. There is no restriction as to which prosodic domain (intonation unit, (intermediate) phrase, word, syllable) can be investigated and there is limited need for annotation prior to analysis. The core of this approach is a cluster analysis on time-series of f0 measurements and consists of two scripts (Praat and R). Graphical user interfaces can be used to perform the analyses and speaker variability can be accounted for. As determining the number of clusters is a key part of the analysis, graphical feedback (plots) is provided for each clustering round (example in Figure 1). After cluster analysis, Praat textgrids can be generated with the cluster number annotated for each individual contour in the data. Although further confirmatory analysis is still required, the outcomes provide useful and unbiased directions for any investigation of prototypical f0 contours based on their acoustic form. These features make the approach particularly useful for language documentation, where the description of prosody is often lacking.
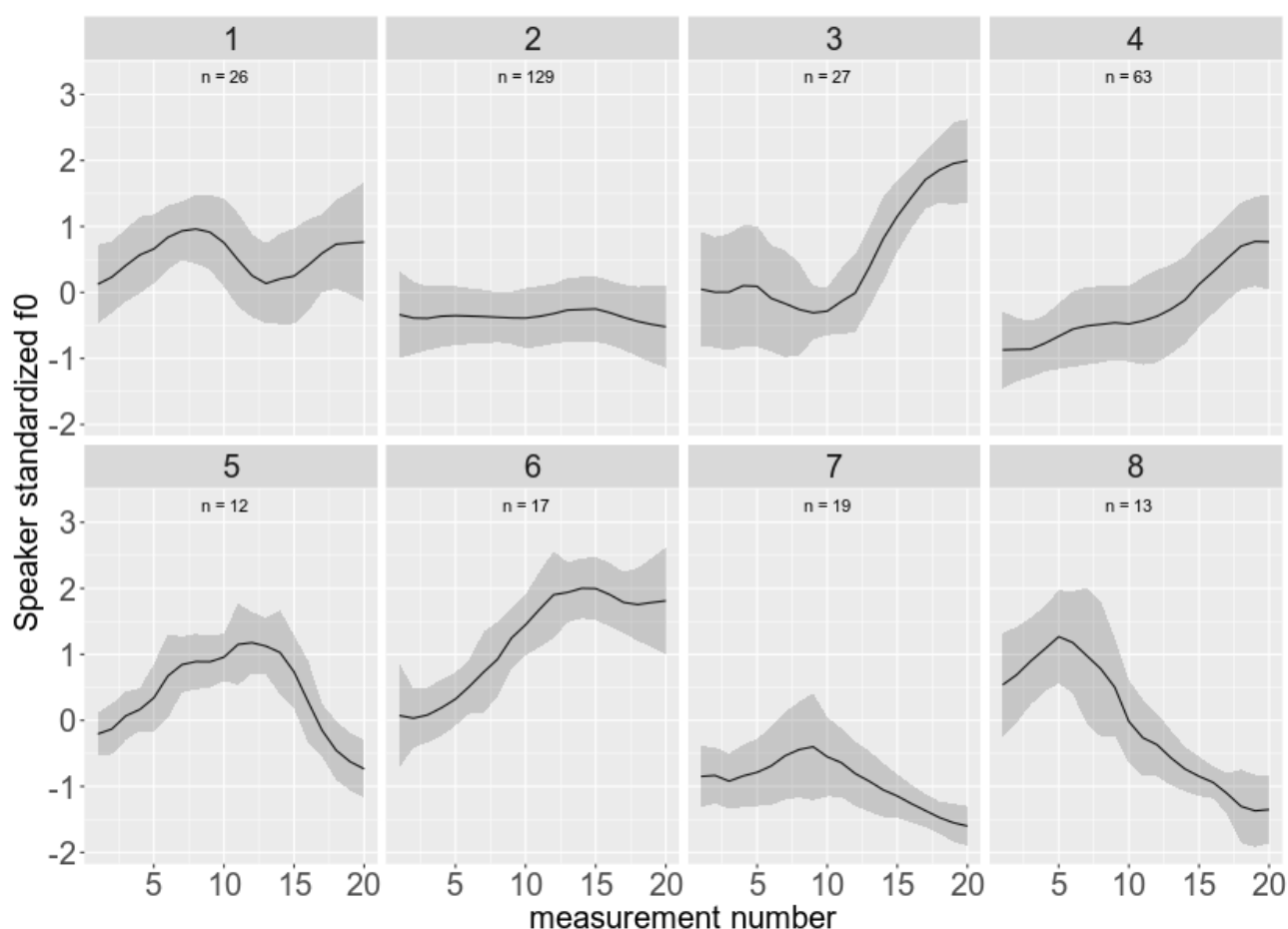


Figure 1. Example of a graphical output based on spontaneous Papuan Malay data assuming eight clusters. Each contour represents the mean speaker standardized f0 (and standard deviation) in a cluster.