Clustering lexical tones with intonation variation
Katrina Kechun Li¹, Francis Nolan¹ & Brechtje Post¹
¹University of Cambridge
kl502@cam.ac.uk, fjn1@cam.ac.uk, bmbp2@cam.ac.uk

Intonation and tones are interwoven in tone languages, both conveyed through the fundamental frequency (f0). A prevailing assumption in the previous literature is that the preservation of tonal categories is prioritised over any intonation manipulation that conveys pragmatic functions. As a result, these studies tend to adopt a ‘top-down’ approach, examining how intonation modifies the f0 contour for each tonal category (see [1] and references therein). In this paper, we present a ‘bottom-up’ analysis, using a contour clustering technique, to investigate how tonal contours are grouped into categories based solely on f0, without prior knowledge of tonal categories.

Our study utilises data from 14 Cantonese speakers who read sentences under different prosodic focus conditions. The sentences adhere to a Subject-Verb-Object structure, with the subject comprising a name prefix /a/ followed by a syllable representing personal names. These name syllables, our analysis targets, encompass the full inventory of Cantonese tones (including checked tones T7-T9), and are either focused or unfocused. We extracted f0 data from 10 equidistant points of the rhyme part of the syllable using Praat, and identified tracking errors using the method proposed in [2]. F0 was converted to semitones with each participant’s average f0 as their base. We employed a hierarchical agglomerative clustering technique implemented in the R application [3] with the complete linkage criterion. The optimal number of clusters was determined by minimising the within-cluster variance while maximising the between-cluster variance based on Euclidean distances.

After removing outlier contours, we obtained 863 unlabelled f0 contours (93.5% of the data), which were subsequently grouped into four clusters (Figure 1). By mapping predefined tonal categories to these clusters (Figure 2), we observed that the clustering primarily reflected tonal register. The majority of two high tones (high-level T1 and high-checked T7) were grouped together as Cluster 3, while most of the two rising tones (high-rising T2 and low-rising T5) were placed in Cluster 4. The low register tones were collapsed into Cluster 2, including low-falling T4, low-level T6 and low-checked T9. Cluster 1 appeared to represent the mid-register tones, encompassing a substantial portion of mid-level T3 and mid-checked T8, although many T3 and T8 contours also fell into the low register cluster.

Our findings concur with a ‘top-down’ analysis using GAMM modelling, which shows that focus marginally influences f0 contour in Cantonese [4]. However, focus does seem to improve the consistency of the clustering. For instance, when focused, T2 was classified into Cluster 4 in 77% of the cases, in contrast to 58% when unfocused. This supports the view that focus induces hyper-articulation of the tonal target and thereby enhancing tonal contrasts [5]. Interestingly, increasing the number of clusters does not further contribute to the distinction of the intonation or tonal categories within each cluster. This suggests that specific tone pairs can be challenging to discern due to their greater surface similarities in continuous speech, which is also evidenced by the ongoing tone mergers of these pairs (T2/T5, T4/T6, T3/T6) [6].

In summary, this study supports the view that lexical tonal contrast is largely maintained across different intonation conditions. Nonetheless, the emergent clusters are unlikely to fully distinguish all lexical tones when relying solely on f0, due to the influence of both tonal similarities and intonation variations. Our next step involves applying the clustering technique to two other languages in our dataset, Chengdu and Changsha. In these languages, intonation might lead to more pronounced variations in tonal targets. Our study will therefore reveal how a bottom-up approach using contour clustering can provide further insight into the interaction of tone and intonation across tone languages.
Figure 1: Cluster means (dark lines) and the contributing f0 contour (light lines) of each cluster.

Figure 2: The proportion of each tone that contributes to each cluster (rows indicates lexical tones represented using Chao number, 1-lowest, 5-highest). Coloration indicates the frequency of tones in each cluster, with numeric label.

References


