# Phonetic Variation of Intonation Contours

## A quantitative analysis using the PRInt model and c-means clusters

In the Autosegmental Metrical theory, intonation contours consist in a sequence of tonal targets structured in terms of relative height and position over the segmental tier. One of the core ideas of the theory is that phonological representations of intonation contours can be derived from the observation of $F_0$ variations (Pierrehumbert 1980). This idea underlies the widely used technique of analysis-by-annotation developed in the ToBI system and adapted to many languages. However, the model has been criticized because it relies on transcribers' insight rather than on a purely quantitative and objective analysis of data. It is indeed difficult for transcribers not to be biased by their acquired intonational knowledge when they annotate novel utterances (see Hirst 2005; Martin 2012; Wightman 2002; Xu 2007 2011).

The PRInt model (Bacuez 2012), based on the AM theory, is a computational tool that circumvents the potential effects of subjectivity by the systematic analysis of phonetic variations in large corpora. Segmental and intonational units are defined by a set of features and these features can take any value within the range circumscribed by the data. Any feature can be implemented in the model (e.g. $F_0$, intensity, and duration for intonation). The first module of the PRInt model iteratively discretizes each utterance of a corpus until it finds its tonal contour, a sequence of low (L) and high (H) tones whose structural organization is solely expressed in terms of relative $F_0$ height and syllabic alignment (Figure 1).

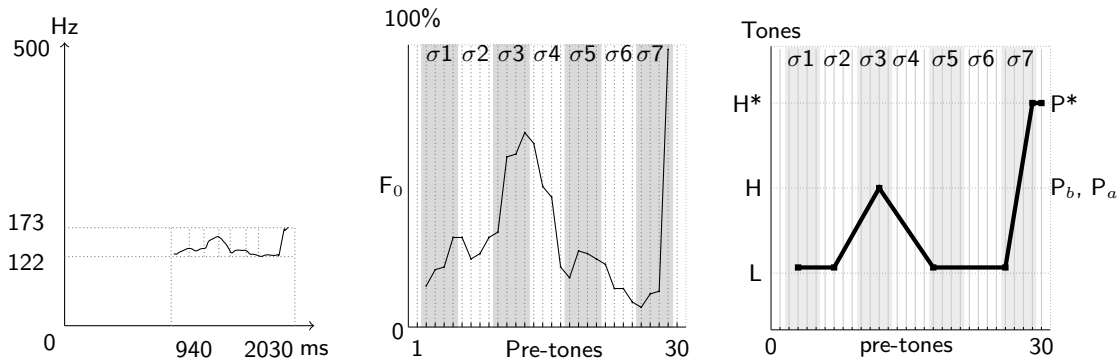
For each feature, the second module of the PRInt model establishes a scale of typicality for the values of this feature in a corpus. Ranking of values is done by c-mean clustering, a process which organizes the data into groups of values around centers of varying population and density, with minimal intra-group distance (similarity) and maximal inter-group distance (difference). A fuzzy function maps the degree of typicality of all the values of a feature onto their cluster organization. This technique has the advantage to establish the scale of typicality not in a linear way but based on the tendency(ies) found in the corpora. From these rankings, the model can estimate the phonological representation of a feature or group of features from the mean of their phonetic variations weighted by their degree of typicality.

In the present paper, the PRInt model was applied to three sub-categories of French closed questions: neutral, with disbelief, and with surprise. These well-documented contours ensured that the model output could be controlled. Data were experimentally elicited and collected among 22 native speakers of metropolitan French for a total of about 6600 utterances. The reciprocal influence between the segmental and intonational tiers is accounted for by degree of typicality. For example, the model treats intonational variations caused by certain segments, such as fricatives (/ʃ/ or /ʒ/), as marginal or a-typical occurrences of contours' variations: the model assigns the lowest degree of typicality to such variations inside their intonational category, and their role in the categorical or phonological structure is minimized. Conversely, variations of segmental duration caused by intonational units are given a very high degree of typicality

in their category and their role in the categorical or phonological structure is maximized. For example, in the case of closed questions, which are signaled by a L+H* tonal structure on the last syllable, it was found that this syllable is consistently lengthened. While the range of variation is continuous, a lengthening of the syllable of more than twice the length of any other syllable in the sentence is the most typical case, and the degree of typicality diminishes as lengthening increases or decreases in relation to the most typical value.

From the ranking of features' phonetic variations by degree of typicality among all samples in the data, the PRInt model calculated the phonological tonal structure of the three contours of French closed questions. It found that disbelief has a contour (L% L+H% L-L%) distinct from the contour of neutral closed questions (L% L+H*H%). The model calculated that surprise is signaled by an allophonic variation of the neutral contour, and is phonetically characterized by an enlargement of the contour in all dimensions. Moreover, the model determined the threshold under which the contour of surprise is not distinguishable from its unmarked counterpart.

The PRInt model is a powerful tool to annotate, organize, and analyze the multifarious parameters of intonational data in large corpora. Once the model has acquired the categorical organization of intonation contours by degree of typicality, it can use this knowledge to automatically annotate novel sentences and make an objective decision about their (grade of categorical) membership among these contours.




(a) Actual (ms, Hz) (b) Scaling (%) and pre-tones (c) Tonal structure

Figure 1 Automated tonal discretization (or annotation) of an utterance by the PRInt model. The actual sample (a) is scaled toward its maximal dimensions and, within each half-syllable, the two most contrastive $F_0$ points, or pre-tones, are located (b). At the utterance level, the subset of the most contrastive pre-tones forms the tonal sequence of the sentence (c). /vu.vu.le.kõ.vjɛn.vu.vwaʁ/, *vous voulez qu'on vienne vous voir ?*, 'Do you want us to come visit you?'

**References**


Bacuez, N. 2012. *Automated Pattern Recognition for Intonation (PRInt). An essay on intonational phonology and categorization*: University of Texas at Austin dissertation.

Hirst, D. 2005. Form and function in the representation of speech prosody. *Speech Communication* 46. 334–347.

Martin, P. 2012. The Autosegmental-Metrical prosodic structure: not fit for French? *Speech Prosody 2012* .

Pierrehumbert, J. 1980. *The phonology and phonetics of English intonation*: MIT dissertation.

Wightman, C. 2002. ToBI or not ToBI? In *First International Conference on Speech Prosody*, Aix en Provence.

Xu, Y. 2007. Speech as articulatory encoding of communicative functions. *In Proceedings of The 16th International Congress of Phonetic Sciences, Saarbrucken* 25–30.

Xu, Y. 2011. Speech prosody: a methodological review. *Journal of Speech Sciences* 85–115.