The contribution of dynamics to the perception of tonal alignment

Leonardo Lancia[1], Cristel Portes[2]
[1]*Laboratoire de Phonétique et Phonologie (CNRS – Sorbonne Nouvelle), Paris, France*
[2]*Aix Marseille Univ, CNRS, LPL, Aix-en-Provence, France*

It is still quite unclear how listeners combine the different features of the f0 signal in the perception of the intonational categories. For example, because the durations of phonologically pertinent f0 movements are weakly correlated to their extension [1], in the perception of pitch accents listeners should be able to factor out the effects of segmental durations on the slopes of f0 rises and falls. Still, the perception of an f0 peak seems to be affected by differences in the slope of the rising (or falling) movement both when these are due to higher starting (or ending) f0 values [2] and when these are due to differences in the curvature of the f0 movement [3]. This sensitivity does not affect the attribution of basic phonological traits (e.g. high/low) to f0 movements, but the perception of their anchoring to the segmental material, i.e, it affects the perception of tonal alignment [2, 3]. To explain these intricate behavioural patterns, it has been proposed [2, 3] that the perceptual location of tonal events depends in some way on the average location of high f0 values in the speech signal. However, this explanation is unable to account for asymmetries in the effects of f0 shapes. Indeed the perception of tonal alignment is more sensitive to changes in the rise preceding a high tone than to changes in the following fall [2].

We present a model that overcomes this limitation by building on the role played by perceptual dynamics in the mapping of continuous f0 values onto abstract intonational events. More specifically, we conceived a simple variant of the Drift Diffusion Model (see [4] for a review) and we conducted a simulation experiment replicating the results obtained in published works ([2] and [3]). In the model, the current state of the perceptual system is characterized by the value of the variable y in the following way: if y exceeds a positive threshold, an upward pitch movement is perceived, while if it crosses a negative threshold, a downward pitch movement is detected. An f0 peak is perceived when both an upward f0 movement and a downward f0 movement have been detected. Governed by the following law of change of y: $\dot{y} = c\dot{f}0 - \lambda y + \xi$, the model states that at a given instant the derivative of y (i.e. its amount of change) is equal to the current rate of change of the f0 signal weighted by the free parameter c (determining the sensitivity of the perceptual system to changes in *f0*), minus the current value of y weighted by the parameter λ (determining how fast the perceptual system returns to a neutral state when *f0* stops changing), plus a random term ξ (representing perceptual noise).

By showing that the proposed model replicates the known effects of f0 shapes on the perception of tonal alignment (see Figure 1) we demonstrate the relevance of dynamics in the explanation of the mechanisms underlying intonation perception. Disregarding this dimension may lead to faulty interpretations of listeners' perceptual behaviour. For example, our model explains the relation between the shape of the f0 curve and the perceived tonal alignment without assuming that listeners explicitly store information on the shape of f0 trajectories. In that, our account resonates with previous models based on the average location of high f0 values, whose partial success can easily be explained as resulting from the accumulative nature of the perceptual drift process. The simplicity of dynamical models like the drift-diffusion model proposed here is due to the fact that they are specific to the constraints of a particular perceptual task and, as shown by a prolific research tradition (see [4] for a review), it guarantees that model parameters are easy to interpret. Additional work is required to account for the perception of more complex f0 patterns and for the effects of spectral and amplitude changes on the perception of f0 changes (e.g.: [5]), and to show how the observed low-dimensional dynamics emerge from the complex interactions between the different factors affecting the perception of intonation.

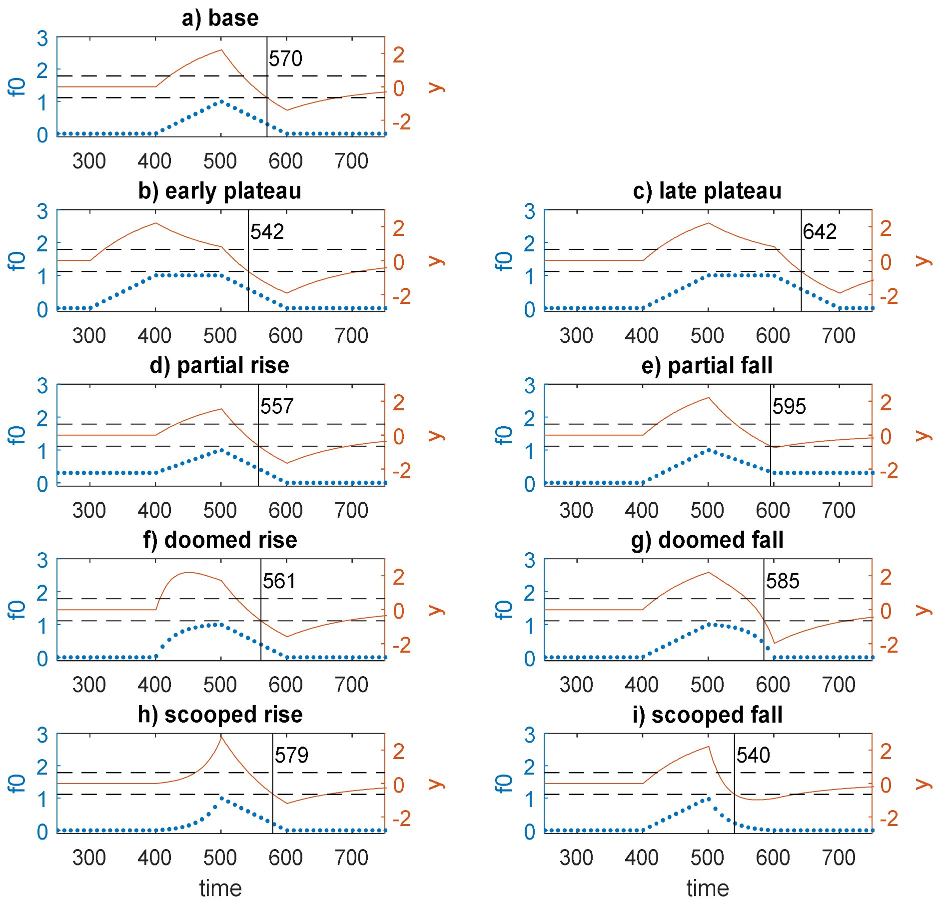

Figure 1. *Behaviour of the model during the presentation of several versions of the basic f0 contour depicted in panel a. The dotted blue curves represent the f0 contour (values in arbitrary units on the left axis) and the continuous red curves represent the state of the perceptual system (values in arbitrary units on the right axis). The horizontal dashed lines represent the thresholds that trigger the detection of pitch movements. The vertical lines indicate the time-point when the fall is detected and the f0 peak perceived. The numbers at their right side indicate the corresponding point in time. Panel a: simulated perception of a symmetric f0 peak. The following modifications to this basic shape anticipate the perception of the f0 peak: early plateau (panel b), partial rise (panel d), doomed rise (panel f), scooped fall (panel i). The perception of the f0 peak is delayed by the following modifications: late plateau (panel c), partial fall (panel e), doomed fall (panel g), scooped rise (panel h).*

[1] Ladd, D. R., Faulkner, D., Faulkner, H., & Schepman, A. (1999). Constant “segmental anchoring” of f0 movements under changes in speech rate. *The Journal of the Acoustical Society of America*, *106*(3), 1543-1554.

[2] D’Imperio, M. (2000). The role of perception in defining tonal targets and their alignment (Doctoral dissertation, The Ohio State University).

[3] Barnes, J., Veilleux, N., Brugos, A., & Shattuck-Hufnagel, S. (2012). Tonal Center of Gravity: A global approach to tonal implementation in a level-based intonational phonology. *Laboratory Phonology, 3*(2), 337-383.

[4] Bogacz, R., Brown, E., Moehlis, J., Holmes, P., & Cohen, J. D. (2006). The physics of optimal decision making: a formal analysis of models of performance in two-alternative forced-choice tasks. *Psychological review, 113*(4), 700.

[5] House, D. (1996). Differential perception of tonal contours through the syllable. In *Proceeding of Fourth International Conference on Spoken Language Processing. ICSLP'96* 4, 2048-2051.