

**Like a square peg in a round hole:
Why contour shape matters for learning new intonation patterns**

Mariapaola D’Imperio^{1,2} & James Sneed German¹

¹Aix-Marseille Université, CNRS, LPL UMR 7039, 13100, Aix-en-Provence, France

²Institut Universitaire de France, Paris, France

mariapaola.dimperio@lpl-aix.fr, james.german@lpl-aix.fr

Imitating intonation across dialects requires speakers to approximate patterns that do not correspond to categories in their native inventory. For this reason, it can reveal the amount and type of phonetic detail that can be accessed from recently experienced speech events. It can also reveal limitations imposed by the role of articulatory practice. Here we explore how Singapore English (SgE) speakers imitate the final contour of American English (AmE) Y/N questions. The SgE contour for this context is similar to the AmE one, in that it involves a rise from a low f0 before the last syllable to a high f0 at the boundary. Phonologically, however, the two patterns are distinct. In AmE, the low preboundary f0 corresponds to a low pitch accent (L*) which is aligned to a stressed syllable and followed by two high tones aligned directly to the edge (H-H%). Phonetically, this translates to a low region that rises from a stressed syllable, and levels out again near the end of the utterance. Phonologically, SgE is an edge-based language [1] and does not have pitch accents. This is reflected phonetically in that the f0 rise does not begin on a stressed syllable, but remains low throughout the final Accentual Phrase, rising only near the onset of the final syllable of the utterance. This difference is illustrated in Figure 1 by the red (AmE) and green (SgE) lines.

The similarities and differences between these two varieties present an interesting test of cross-dialect imitation for several reasons. First, in [2], we showed that for utterance-initial contours in declaratives, SgE speakers were able to imitate detailed aspects of an AmE speaker’s f0 peak alignment. This was in spite of the fact that the inventory of SgE lacks an intonational category that corresponds phonetically to the targeted American pattern (an f0 peak aligned to a stressed syllable). Moreover, they were able to match the specific f0 peak alignment of the model speaker on a token-by-token basis. Thus, we can ask for the present data whether SgE speakers can imitate dynamic aspects of a different type of AmE contour (low rising, L* H-H%) in a different context (utterance-final). Second, in the previous study, the targeted AmE contour was phonetically very distinct from any sentence-initial contour in the SgE inventory, whereas for the present data, the two contours show broad phonetic resemblance. Based on principles of perceptual assimilation [3] we can therefore hypothesize that imitators will perform poorly on the Y/N question contour either because they do not hear the differences between their native contour and the AmE contour, or because they deem their native contour to be “good enough”. Finally, the present data allows us to directly explore the role of contour shape. [4] and [5] showed that contour curvature can be important for marking intonational contrasts in Neapolitan Italian and German. We can therefore ask whether SgE speakers can imitate rise *shape* of the AmE Y/N question contour or if they can only approximate scaling properties of specific regions of that contour, given that their phonology lacks the relevant shape specification.

To explore these issues, 19 male SgE participants read sentences aloud in two different tasks: (i) in their native dialect in the absence of audio prompts (*Baseline*), and (ii) in response to audio prompts which they tried to imitate (*Imitation*). Twelve trisyllabic target words with initial stress appeared as the last word in a Y/N question (e.g., *Does sleep improve your memory?*). Audio prompts were prerecorded by a native speaker of AmE and controlled for utterance tune (L* H- L* H-H%). Mean f0 was extracted at 12 equal intervals through each target word utterance, yielding a time-normalized estimate of the final portion of the f0 contour. Figure 1 shows mean f0 at each time step for the AmE speaker, the Baseline task, and the Imitation task. These plots show that participants modified their pattern towards that of the AmE speaker, since at nearly every time step, the mean f0 shifts away from the baseline value in the expected direction. This effect was confirmed by a linear mixed effects regression analysis (*task*: est.=6.93, t=2.30, p<0.0001; *time*: est._{max}=76.2, t_{max}=30.8, p<0.0001; *interaction*: est._{max}=26.1, t_{max}=7.83, p<0.0001).

Figure 1 also suggests that while imitators achieved a degree of phonetic approximation by lowering f0 early in the contour and raising f0 at the end, the overall shape of the contours does not change. Since the AmE and Baseline contours have a similar amount of overall curvature, a 2nd-order polynomial (i.e., a parabola) is not expected to capture this difference. Instead the shape differences

are best characterized by the degree of inflection (i.e., sigmoid-like curvature), which statistically can be assessed by regression fitting a third order polynomial to each contour and then comparing differences in the coefficient of the third term. This method was applied to each token in the data set, and the distributions of the resulting coefficients were compared. Since the AmE tokens are inflected rightward, those values should be overall more negative than for the baseline tokens. Furthermore, if speakers were able to adjust the shape towards that of the AmE targets, then the distribution of coefficient values should be lower (more negative) for the imitation task than for the baseline task. Figure 2 confirms the first difference in the coefficient values. Indeed, there is strong separation between the SgE Baseline values which fall close to zero (indicating little inflection), while the AmE values are highly negative. The distribution for the Imitation tokens, however, shows a high degree of overlap with the Baseline tokens and very little with the AmE tokens. The difference between Baseline and Imitation was only marginally significant in a linear mixed effects comparison (est.= -4.52, $t=2.35$, $p=0.020$), suggesting that speakers were unable to reliably shift from the simple concave rise of their native system to the more complex, S-shaped curve associated with the target contours.

Overall, our findings show that SgE speakers could match certain phonetic details of the imitated f0 contours, since they could raise or lower the f0 level appropriately across different regions. This implies that they were also able to perceive differences in the targeted contours, though possibly only in terms of scaling effects. The fact that they did not modify the higher order shape of their contours, however, suggests interference at one or more levels. Possibly, they could not perceive the difference in shape due to effects of perceptual assimilation. On this view, the imitated tokens are variants of a native category whose parameters have been adjusted to provide a better phonetic match to the targets. Alternatively, imitating complex contour shapes may require articulatory practice. As the native system involves only fully concave rises, SgE speakers may lack the articulatory control to produce a rise involving an inflection point. In either case, these findings reveal that new contour shapes are not immediately accessible to the perception and/or the production system. Moreover, it suggests that a lack of success in imitating this level of phonetic detail might be at the basis of foreign accent phenomena in L2 acquisition.

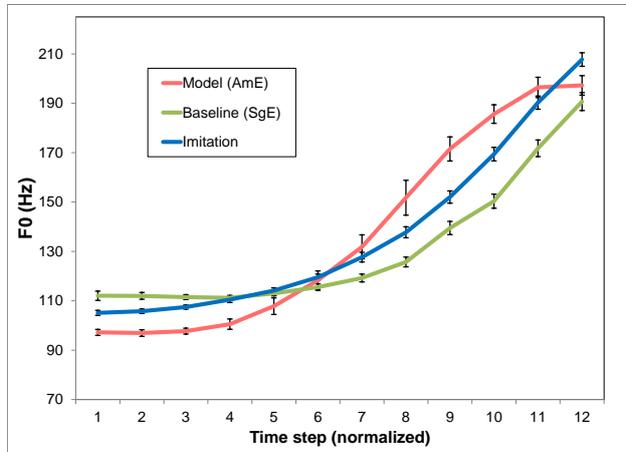


Figure 1. Lineplot of F0 by normalized time step for three tasks.

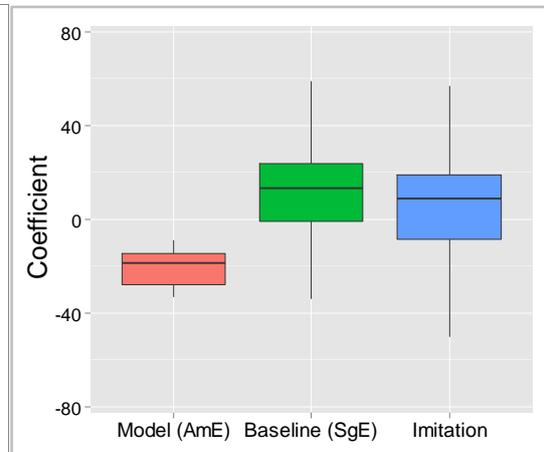


Figure 2. Boxplot of the coefficients of polynomial regression models by task.

References: [1] Jun, S.-A. (2005). *Prosodic Typology: The Phonology of Intonation and Phrasing*. Oxford UP. [2] D’Imperio, M. & German, J. S. (2015). Phonetic detail and the role of exposure in dialect imitation. *ICPhS 2015*. [3] Best, C. T., McRoberts, G. W., & Goodell, E. (2001). Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener’s native phonological system. *J. of the Acoustical Society of America*, 109(2), 775–794. [4] D’Imperio, M. & Cangemi, F. (2009). The interplay between tonal alignment and rise shape in the perception of two Neapolitan rising accents. Presented at *Phonetics and Phonology in Iberia (PaPI) 2009*. / [5] Niebuhr, O., D’Imperio, M., Gili-Fivela, B. and Cangemi, F. (2011). Are there “shapers” and “aligners”? Individual differences in signaling pitch accent category. In *Proceedings of ICPhS XVII*.