Speakers adaptively plan f0 trajectories under rate changes: Evidence from Thai contour tones
Francesco Burroni & James Kirby
Institute of Phonetics and Speech Processing, Ludwig Maximilian University of Munich
{francesco.burroni, jkirby}@phonetik.uni-muenchen.de

An open question in phonetics is how speakers of tone languages adjust the production of f0 contours when faced with changes in speech rate and word duration. This is particularly relevant for Thai contour tones, where previous research has yielded conflicting results. Some studies have suggested that Thai speakers “scale” contour tones to fit the duration of the syllables they are associated with [1]. Other studies have found that speakers truncate contour tones in connected speech, specifically the Falling (F) and Rising (R) tones, producing them without a complete fall or rise [2], [3]. Finally, another possibility is that contours are not truncated, instead they “spill over” onto the consonantal onset of following words, giving the impression of truncation when f0 tracking is limited to rhymes [4], [5]. We provide evidence and analyses supporting this last hypothesis. Using Generalized Additive Mixed Models (GAMMs), we examine entire f0 trajectories and demonstrate that Thai speakers adaptively plan and execute contour tones to accommodate their characteristic f0 shapes within the constraints imposed by speech rate changes, with interesting interactions between certain tonal combinations and speech rates.

Methodology. Two production experiments were conducted with 44 Bangkok Thai speakers. They produced disyllabic nonce combinations of syllables with sonorant onsets bearing Falling/Rising tones (20 speakers) or Mid/Low/High tones (24 speakers) in carrier sentences imitating the speed of a continuous rate cue. The disyllabic combinations were embedded in a carrier phrase consisting of Mid-toned syllables (i.e., M1 [M/L/F/H/R] M2 M3) to minimize coarticulatory effects. Our focus is on the f0 trajectories of the three contour tones F/H/R, which serve as ideal targets to observe the impact of rate changes. We extracted f0 (following the methods of [5]) from the beginning of the preceding Mid-tone (M1) to the end of the consonant onset of the word following the target disyllable (M2). We analyzed the f0 trajectories of each tone separately. The GAMM models incorporated parametric terms for tonal context and duration, as well as smooths for time, duration, tonal context, and their interactions. Subject-specific factor smooths were also included. We chose GAMMs to generate predictions for each contour tone in different tonal contexts based on z-scored duration.

Results. Two key results emerge from the GAMM fits. First, Thai speakers overwhelmingly do not truncate contour tones. Second, reduction of the final portion of contour tones only emerges in specific combinations of durations and tonal context. Figure 1 shows f0 contours for the F/H/R tones at different illustrative durations and in different tonal contexts. Note that the results are confirmed in full model predictions of f0 contours over duration treated as a continuous variable. Starting from the F tone (top row), the final falling portion is never truncated at any duration/ in any context, except for the shorter duration in the F-F context. For the H tone (mid row), no truncation of the final falling portion is observed, except for the H-F context at normal and longer durations. Finally, for the R tone (bottom row), truncation of the final rise is only observed for the R-H and R-R contexts at shorter durations.

Our analysis reveals two key findings. First, Thai speakers do not canonically employ truncation as a strategy to adapt to variations in speech rate. Instead, they demonstrate adaptive planning and execution of f0 contours, ensuring that the rate of f0 change corresponds to the duration of the associated segmental material, preserving characteristic f0 shapes. This entails the need for f0 tracking over larger windows beyond the immediate word. Second, specific combinations of tonal contexts and duration induce changes in contour tone shape that resemble truncation. These contexts share shorter durations and conflicting f0 direction, e.g., F-F sequences, where the final fall of the first tone conflicts with the initial rise of the second tone. Our findings have implications for models of f0 control. The dynamic scaling of f0 contours suggests a relationship between rate of f0 change and the units lexically associated with the tone, a relationship not incorporated in mainstream models of f0 control. The occurrence of truncation-like effects at faster speeds, with conflicting f0 movements, suggests that these effects may arise from competing demands on laryngeal articulation that cannot be met due to insufficient timing resulting in asymmetric blends. We present preliminary modeling of these effects using an f0 control model incorporating aspects of the Fujisaki and Task-Dynamic model ([6]–[8]).
Figure 1: F0 contours from GAMM model fits representing the mean realization of the Falling (top panel), High (mid panels) and Rising tone (bottom panel) followed by different tonal contexts and at different z-scored durations. Orange lines indicate truncation-like effects.

References