Perception of Thai tones in context

Amber B. Camp & Amy J. Schafer
Department of Linguistics, University of Hawai'i at Mānoa

Research Questions

Q1: Is perception of Thai lexical tone categorical when the words are presented in a sentential and intonational context?

Q2: Does the categorical perception of lexical tone change across intonational contexts, e.g. because the categorical boundary shifts location?

Background: Categorical perception of tones

- Categorical perception (CP) studies attempt to account for the mapping of multiple acoustic cues to discrete phonological categories
- Classic CP research focuses on segments (see Lisker & Abramson 1967, Pisoni 1985, and many others)
- Acoustic landmarks cue perception. Consonants tend to be perceived categorically while vowels are not (Liberman 1957, Fry et al. 1962)
- CP of suprasegmentals has been less clear:
  - Register tones are perceived continuously (Abramson 1979a, Francis et al. 2003)
  - Contour tones are perceived (more) categorically (Francis et al. 2003)
- Contour tones have acoustic landmarks that register tones lack, and sentence frames can provide further context, e.g. for the speaker's F0 range
- Past studies have shown sentence frames provide speaker-specific context for pitch and speech rate, facilitating CP of tones (Moore & Jongman, 1997; Wong & Diehl, 2003; Sjerps et al., 2018)
- Sentence intonation may also cause level tones to become more contoured (see Thai background)
- Francis et al. (2006) tested un-modified mid level tones in Cantonese sentence frames modified to varying degrees (semitones). Those mid level tones were identified differently depending on the frame, supporting the idea that extrinsic context is relevant to lexical tone processing

Background: Thai

- Lexical tone in Thai (Morén & Zsiga, 2006) and other languages (e.g. Cantonese in Ma et al., 2006) can change shape in different intonational contexts
- Limited research to date on the perception of Thai lexical tone & intonation
  - Luksaneeyanawin (1998) described Thai intonation based on production data
    (ex. raised, narrow pitch range for questions & final fall for statements)
  - Abramson (1979b) and Morén & Zsiga (2006): phonetic realization of lexical tone differs across different intonational/sentential contexts

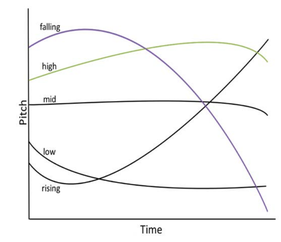

Fig. 1 Five canonical Thai tones, based on Abramson 1962

- Abramson (1979b, and other work), Morén & Zsiga (2006), and Zsiga & Nitisaroj (2007) postulate an underlying L tone on Thai falling tones in final position (i.e. citation form), and that tones in non-final position are compressed or truncated
- Alternatively, since there is agreement that the phonetic realization of tones is different sentence-medially and sentence-finally, and it is generally assumed that there is a final fall in declaratives (attributed to L%), we can also assume that tones are augmented by L%, similar to how tones are influenced by H% in Cantonese (Ma et al., 2006)
  - Therefore, we assume a sentence-final fall (an L% intonational boundary tone) causes high and falling tones, respectively, to take on a lower contour with a final fall, and a greater fall (see Fig. 2). Additionally, these tones can potentially be described as having their canonical contours sentence-medially

Predictions: Thai tones in context

P1: Sentence frames provide acoustic context such as relative pitch, which will facilitate categorical perception of tones

P2: A final L% boundary tone will cause the categorical boundary between high and falling tones to shift; more steps will be identified as falling-tone words as compared to the medial position, indicating a strong influence of bottom-up acoustic information. E.g., listeners will preferentially map falling contours to falling lexical tones, rather than high tone exemplars in an L% context

Materials

- Two naturally-produced words, /ɔː/-high (wheel) and /ɔː/-falling (mule), were recorded in sentence-medial and sentence-final frames

Medial Context Frame
prod leuk phab khong (target) hai chan
LOW FALL FALL RISE FALL RISE
'Select the image of the ____ for me'

Final Context Frame
prod leuk phab khong (target)
LOW FALL FALL RISE
'Select the image of the ____'

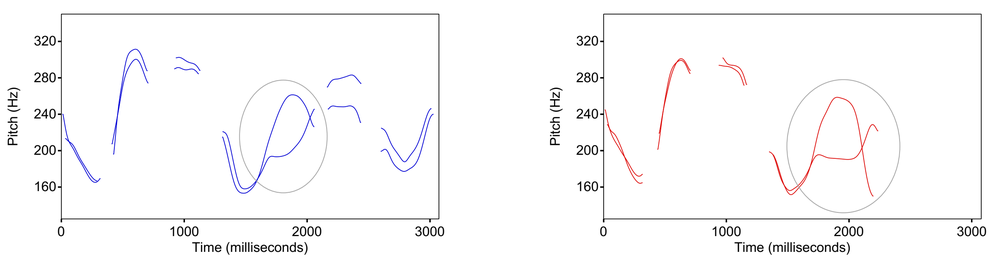

Fig. 2 Naturally produced sentences in the sentence-medial (left panel) and sentence-final (right panel) frames. In both panels, the target words are circled and the falling tone is represented by the line with the higher midpoint.

- Target words extracted and were used to create two continua: sentence-medial and sentence-final, using PSOLA in Praat (Mitterer et al., 2011, Xu et al. 2006, Moulines & Laroche, 1995)
- Pitch and intensity manipulated linearly across 9 steps; duration averaged and equated

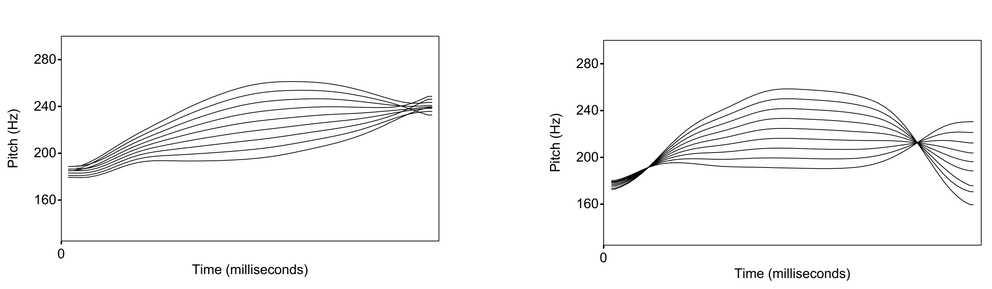

Fig. 3 Tone continua for sentence-medial (left) and sentence-final (right) positions

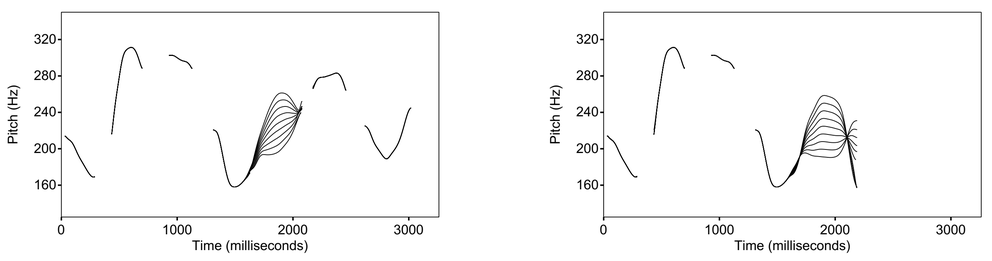

Fig. 4 Continua spliced into sentence-medial (left) and sentence-final (right) frames. Note that the portion of frame before the target word is identical in both cases

- Continua endpoints are minimal tone pairs: /ɔː/-high (wheel) vs. /ɔː/-falling (mule)
- Tokens then concatenated with natural sentence frames
- In each trial, participants heard individual tokens presented in full sentence frames

Participants

- 18 adult native Thai speakers (10 female)
- No musical training; self-reported normal hearing and vision

Method

Task 1: Identification

- 4 practice items
- 54 randomized trials
  (9 steps x 2 Frames x 3 reps)
- Participants hear a token from the continuum embedded in a Frame, and select the picture corresponding to the high- or falling-tone word

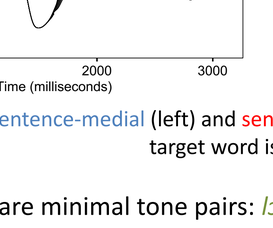

Fig. 5 Sample visual scene

Task 2: Discrimination

- 4 practice items
- ABX discrimination task
  (one interval between A and B)
- 500 ms ISI (interval between offset/onset of sentences)
  - 3114 ms between targets in Medial context
  - 2152 ms between targets in Final context
- Participants hear A, B, X, and then select '1' or '2' to indicate X is identical to A or B
- 128 randomized trials
  (8 step pairs x 4 ABX orders x 2 Frames x 2 reps)

Results

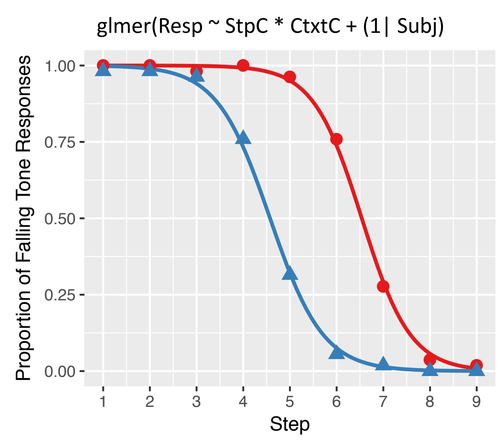

Fig. 6 Identification task results

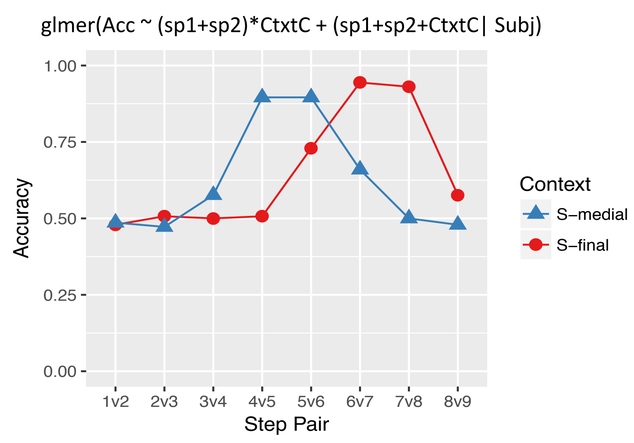

Fig. 7 Discrimination task results

- Identification results show a categorical boundary at about step 4.5 in the sentence-medial context and at 6.5 in the sentence-final context (ME logistic regression; backwards model comparison)
  - Significant effects of Step ($p < .001$, $z = -13$) and Context ($p < .001$, $z = -8.9$); no interaction between Step and Context
- Discrimination results show a significantly higher level of accuracy between steps 4-5-6 and steps 6-7-8, which correlates to the boundaries above (ME logistic regression with orthogonal quadratic term; backwards model comparison)
  - Significant effect of each Step Pair term (sp1: $z = 8.1$, sp2: $z = -2.8$) but not Context. Significant interaction between each Step Pair term and Context (sp1*Ctxt: $z = -5.5$, sp2*Ctxt: $z = -4.4$)

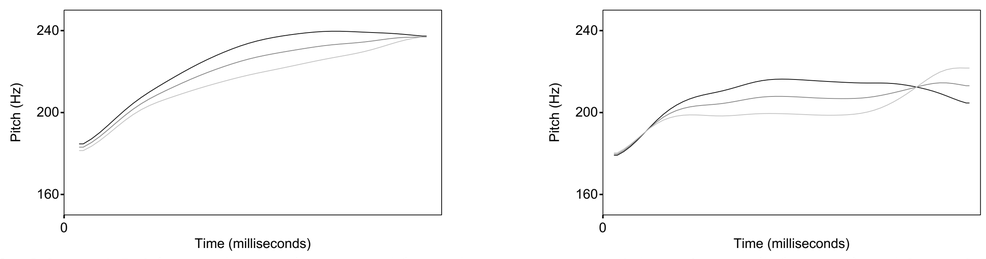

Fig. 8 Steps 4-5-6 (top to bottom) extracted from the sentence-medial continuum (left panel) and Steps 6-7-8 from the sentence-final continuum (right panel). These steps represent step pairs that had high discrimination accuracy

Discussion and conclusions

- Results indicate:
  - (1) CP in both intonational contexts; and
  - (2) a categorical boundary shift dependent on context
- Acoustically lower/falling tones in an L% context were perceived as the falling lexical tone
- Listeners do not fully compensate for L%-influenced phonetic realizations of tones. Pitch contours were preferentially mapped to the lexical tone
- These results suggest that, in regards to F0, lexical meaning takes precedence over sentence intonation
- Step pairs (Fig. 8) show various differences between tones that can explain the high discrimination accuracy, including:
  - Peak alignment (Zsiga & Nitisaroj, 2007)
  - Overall shape (Abramson, 1979b)
- Additional research is necessary to probe the roles of abstract tone shapes and/or peak alignment, across grammatical/affective variation in intonation, and how these might be employed during online processing decisions, e.g. in a hybrid model that incorporates multiple exemplars as well as abstract categories for lexical and intonational distinctions
- Production research in progress from 27 native Thai speakers indicates tone/intonation realizations that pattern similarly to the single speaker in this experiment
- Perception/word recognition research in progress uses eye-tracking to investigate the time course of processing the interaction between tone and intonation

References

Abramson, A. S. (1979a). The noncategorical perception of tone categories in Thai. In B. Lindblom & S. Öhman (Eds.), *Frontiers of Speech Communication Research* (pp. 127-134). London, England: Academic Press. **Abramson, A. S. (1979b).** The coarticulation of tones: An acoustic study of Thai. In T. L. Thongkum, V. Panupong, P. Kullavanijaya, & M. R. K. Tingsabadh (Eds.), *Studies in Tai and Mon-Khmer phonetics and phonology in honor of Eugénie J. A. Henderson* (pp.1 – 9). Bangkok, Thailand: Chulalongkorn University Press. **Francis, A. L., Ciocca, V., & Ng, B. K. C. (2003).** On the (non) categorical perception of lexical tones. *Perception & Psychophysics*, *65*(7), 1029-1044. **Francis, A. L., Ciocca, V., Wong, N. K. Y., Leung, W. H. Y., & Chu, P. C. Y. (2006).** Extrinsic context affects perceptual normalization of lexical tone. *Journal of the Acoustical Society of America*, *119*(3), 1712-1726. **Fry, D. B., Abramson, A. S., Eimas, P. D., & Liberman, A. M. (1962).** The identification and discrimination of synthetic vowels. *Language and Speech*, *5*(4), 171-189. **Liberman, A. M., Harris, K. S., Hoffman, H. S., & Griffith, B. C. (1957).** The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology*, *54*(5), 358. **Lisker, L., & Abramson, A. S. (1967).** Some effects of context on voice onset time in English stops. *Language and Speech*, *10*(1), 1-28. **Luksaneeyanawin, S. (1998).** Intonation in Thai. In D. Hirst & A. Di Cristo (Eds.), *Intonation Systems: A Survey of Twenty Languages* (pp. 379-397). Cambridge, England: Cambridge University Press. **Ma, J. K., Ciocca, V., & Whitehill, T. L. (2006).** Effect of intonation on Cantonese lexical tones. *Journal of the Acoustical Society of America*, *120*(6), 3978-3987. **Mitterer, H., Chen, Y., & Zhou, X. (2011).** Phonological abstraction in processing lexical-tone variation: Evidence from a learning paradigm. *Cognitive Science*, *35*(1), 184-197. **Moore, C. B., & Jongman, A. (1997).** Speaker normalization in the perception of Mandarin Chinese tones. *Journal of the Acoustical Society of America*, *102*, 1864–1877. **Morén, B., & Zsiga, E. (2006).** The lexical and post-lexical phonology of Thai tones. *Natural Language & Linguistic Theory*, *24*(1), 113-178. **Moulines, E., & Laroche, J. (1995).** Non-parametric techniques for pitch-scale and time-scale modification of speech. *Speech Communication*, *16*(2), 175-205. **Pisoni, D. B. (1985).** Speech perception: some new directions in research and theory. *Journal of the Acoustical Society of America*, *78*(1), 381–388. **Sjerps, M. J., Zhang, C., & Peng, G. (2018).** Lexical tone is perceived relative to locally surrounding context, vowel quality to preceding context. *Journal of Experimental Psychology: Human Perception and Performance*, *44*(6), 914. **Wong, P. C., & Diehl, R. L. (2003).** Perceptual normalization for inter- and intratalker variation in Cantonese level tones. *Journal of Speech, Language, and Hearing Research*, *46*(2), 413-421. **Xu, Y., Gandour, J. T., & Francis, A. L. (2006).** Effects of language experience and stimulus complexity on the categorical perception of pitch direction. *Journal of the Acoustical Society of America*, *120*(2), 1063-1074. **Zsiga, E., & Nitisaroj, R. (2007).** Tone features, tone perception, and peak alignment in Thai. *Language and Speech*, *50*(3), 343-383.

Acknowledgements
Data collection was supported in part by the Department of Linguistics Endowment Fund at UH Mānoa. Special thanks to Victoria B. Anderson, Wenyi Ling, Andrea L. Berez-Kroeker, and Gary Holton for providing extensive feedback and assistance. Additional thanks for feedback from attendees of the Language Acquisition Reading Group and of Advanced Laboratory Research.

Presented at LabPhon 16 in Lisbon, Portugal, June 22, 2018, contact: acamp@hawaii.edu

Supplemental Materials