The Role of Initial F0 Rise in Speech Segmentation: A Cross-linguistic Study

Shu-chen Ou[1], Zhe-chen Guo[2]

[1]sherryou@mail.nsysu.edu.tw, National Sun Yat-sen University, Taiwan; [2]zcadamguo@utexas.edu, University of Texas at Austin, USA

This paper investigates the cross-linguistic use of initial F0 rise in speech segmentation. Previous studies using real word detection experiments have provided suggestive evidence that F0 rise serves as a universal cue to word beginnings. We examined if there would be more direct support for this claim by conducting an experiment with listeners of four typologically different languages: English, French, Japanese, and Taiwanese Southern Min. Participants learned an artificial language by listening to speech streams in which the language's words were concatenated without pauses in between and then identified the words in a test. Analysis of identification accuracy indicated that the words with an F0 rise on the initial syllables were identified more accurately than those without any F0 cue, suggesting that initial F0 rise facilitated segmentation. This was found for all listener groups, adding support to the view that F0 rise is a universally accessible cue to word beginnings.

1. Introduction

- Speech segmentation: division of continuous speech into discrete units.
- Segmentation strategies are shaped by language-specific experience, e.g., experience with phonotactics (McQueen, 1998), distribution of prosodic prominence (Tyler & Cutler, 2009), etc.
- However, a rise in F0 is exploited to locate word beginnings by listeners of several different languages, including French (Welby, 2007), Korean (Kim, 2003), and Japanese (Warner et al., 2010).
- F0 rise may be a universally salient cue to word onsets (Warner et al., 2010)
- Direct evidence is still lacking: (1) No direct cross-linguistic comparisons; (2) it is difficult to control for lexical confounds.
- Artificial language (AL) learning experiment: learning an AL by listening to tokens of its words concatenated together without pauses.
- Goal of this study: to examine whether F0 rise is a cross-linguistic cue to word beginnings with an AL learning experiment.
- Four typologically different languages: (1) English (lexical stress); (2) Japanese (lexical pitch accent); (3) Taiwanese Southern Min (TSM; lexical tone); (4) French (no lexical prosody).

2. Experiment

➢ ***Materials***

- AL lexicon (6 CVCVCV nonsense sequences):
  [pakime], [tulepi], [kemina], [mapeku], [nitamu], and [linute]
- The syllables were recorded individually; their F0 contours were flattened at 132 Hz and their durations were normalized to 341 ms.
- Half of the AL words had an initial F0 rise (3.5 semitones) and the other half did not have any prosodic cue.

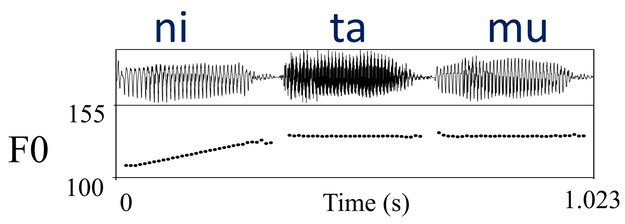

FIG. 1 Waveform and F0 contour of a sample word

➢ ***Procedure***

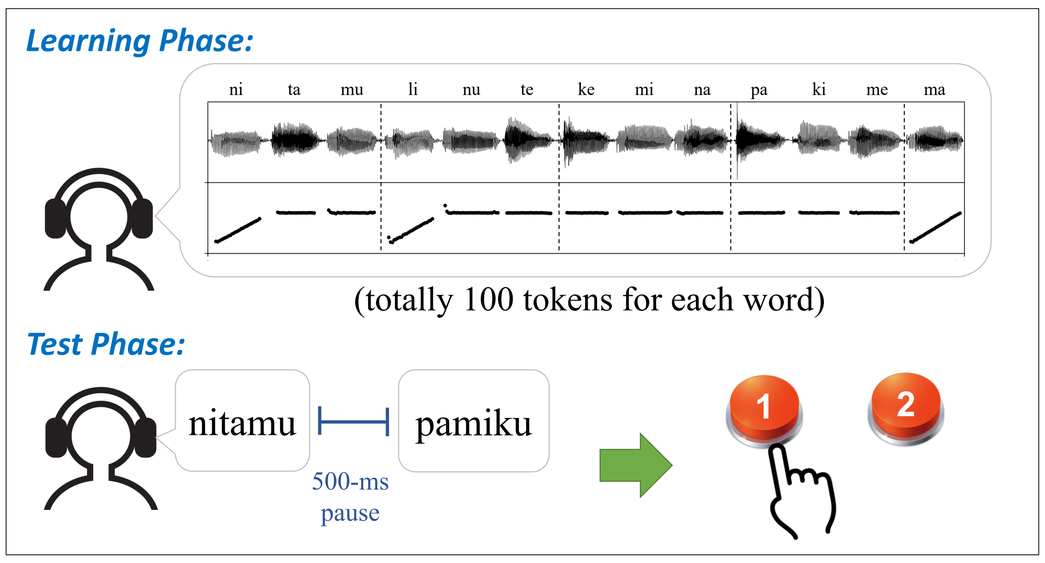

FIG. 2 Procedure of the experiment. First, subjects learned the AL by listening to speech streams in which tokens of the 6 words were concatenated without pauses in between. Next, they heard an AL word and a "nonword" and selected the former.

- 6 AL words × 6 nonwords (sequences not part of the AL) = 36 trials.
- Higher identification accuracy in the test → better segmentation performance during the learning.

➢ ***Participants***

- 15 English, 26 Japanese, 26 TSM, and 21 French listeners.

3. Hypothesis and Prediction

- Hypothesis: F0 rise is a cross-linguistically useful cue for locating word beginnings.
- Prediction: For all listener groups, AL words with initial F0 rise would be identified significantly more accurately than those without any prosodic cue.

4. Results

- A mixed-effects logistic regression analysis (Bates et al., 2015) was conducted on identification responses separately for each group.

  Response ~ Prosody + Trial + (1 + Prosody | Subject) + (1 | Word) + (1 | Nonword)
- Prosody was significant for all groups (English: β = .766, SE(β) = .368, z = 2.081, $p < .05$; Japanese: β = .795, SE(β) = .234, z = 3.390, $p < .001$; TSM: β = .455, SE(β) = .200, z = 2.273, $p < .05$; French: β = .411, SE(β) = .202, z = 2.031, $p < .05$).

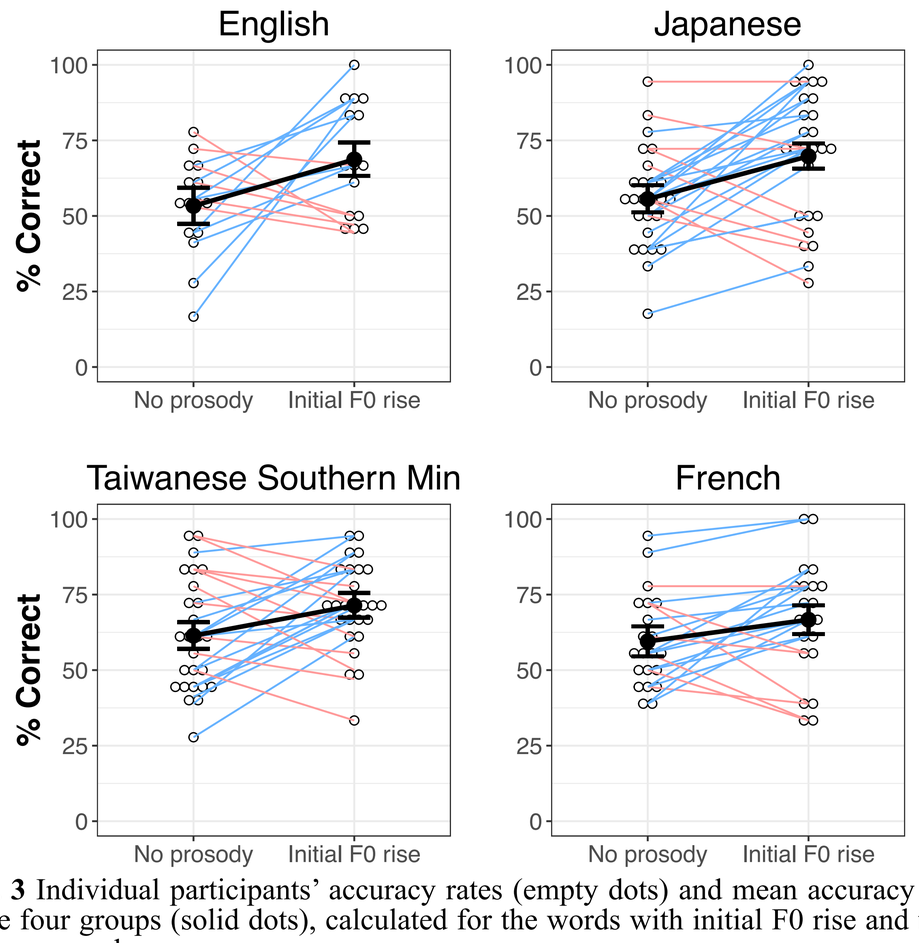

FIG. 3 Individual participants' accuracy rates (empty dots) and mean accuracy rates of the four groups (solid dots), calculated for the words with initial F0 rise and those with no prosody.

5. Discussion and Conclusion

- The results support the hypothesis that a rise in F0 is exploited cross-linguistically to locate word beginnings.
- The findings from French and Japanese listeners are consistent with previous studies with these two listener groups (Warner et al., 2010; Welby, 2007).
- In English, stressed syllables are predominantly word-initial (Cutler & Carter, 1987) and one correlate of lexical stress is F0 rise (Spitzer et al., 2007).
- The English listeners might apply the stress-based segmentation strategy developed from experience with their native language (Tyler & Cutler, 2009).
- However, a language-specific explanation fails to account for TSM listeners' use of initial F0 rise.
- TSM has only one rising tone, which occurs only in the final position of a word, prosodic domain, etc. because of the language's tone sandhi process.
- The cross-linguistic parallel provides some preliminary evidence for F0 rise as a universal cue to word beginnings.
- Further issues and limitations:
  1. If the hypothesis holds, what might be the mechanism(s) underlying the universal use of F0 rise? Is it related to the "trochaic law" (Hay & Diehl, 2007)?
  2. Need more subjects for some language groups.

6. References

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*, 1–48.

Cutler, A., & Carter, D. M. (1987). The predominance of strong initial syllables in the English vocabulary. *Computer Speech and Language*, *2*, 133–142.

Hay, J. S., & Diehl, R. L. (2007). Perception of rhythmic grouping: Testing the iambic/trochaic law. *Perception & Psychophysics*, *69*, 113–122.

Kim, S. (2003). The role of post-lexical tonal contours in word segmentation. *Proceedings of 15th ICPhS* (pp. 495–498), Barcelona.

McQueen, J. M. (1998). Segmentation of continuous speech using phonotactics. *Journal of Memory and Language*, *39*, 21–46.

Spitzer, S. M., Liss, J. M., & Mattys, S. L. (2007). Acoustic cues to lexical segmentation: A study of resynthesized speech. *The Journal of Acoustical Society of America*, *122*, 3678– 3687.

Tyler, M. D., & Cutler, A. (2009). Cross-language differences in cue use for speech segmentation. *The Journal of Acoustical Society of America*, *126*, 367–376.

Warner, N., Otake, T., & Arai, T. (2010). Intonational structure as a word-boundary cue in Tokyo Japanese. *Language and Speech*, *53*, 107–131.

Welby, P. (2007). The role of early fundamental frequency rises and elbows in French word segmentation. *Speech Communication*, *49*, 28–48.