Durational variability of spontaneous and read speech: Comparison between English and Japanese

Yoichi Mukai, Daniel Brenner, Benjamin V. Tucker

VANCOUVER ISLAND UNIVERSITY

NAU NORTHERN ARIZONA UNIVERSITY
Communication Sciences and Disorders

NASCL
Northern Arizona Speech Communication Laboratory

1. Introduction

Background:

- Phonetic research often relies on careful laboratory speech despite spontaneous conversational speech being the most commonly used in everyday interactions [1].
- Compared to read speech, spontaneous speech has been reported to have a higher articulation rate, lower f0 variation, and greater f0 declination, as well as more frequent hesitations, approximated articulation, shorter segment durations, and shorter prosodic units [1].
- Japanese and English have distinct phonological and temporal structures]: Stress vs. Pitch accent [2, 3].

Research objective: An exploratory analysis using durational variability measures to investigate differences between the production of spontaneous and read speech in Japanese and English.

2. Method

Data:

- **Read Speech:** Used Grenon & White's data [4]
  - **60 sentences:** 6 English speakers read 5 sentences; 6 Japanese speakers read 5 sentences.
- **Spontaneous speech:** Used Warner et al.'s data [5]: Phone conversation with someone familiar
  - **180 utterances:** 6 Japanese speakers produced 15 utterances each; 6 English speakers produced 15 utterances each.

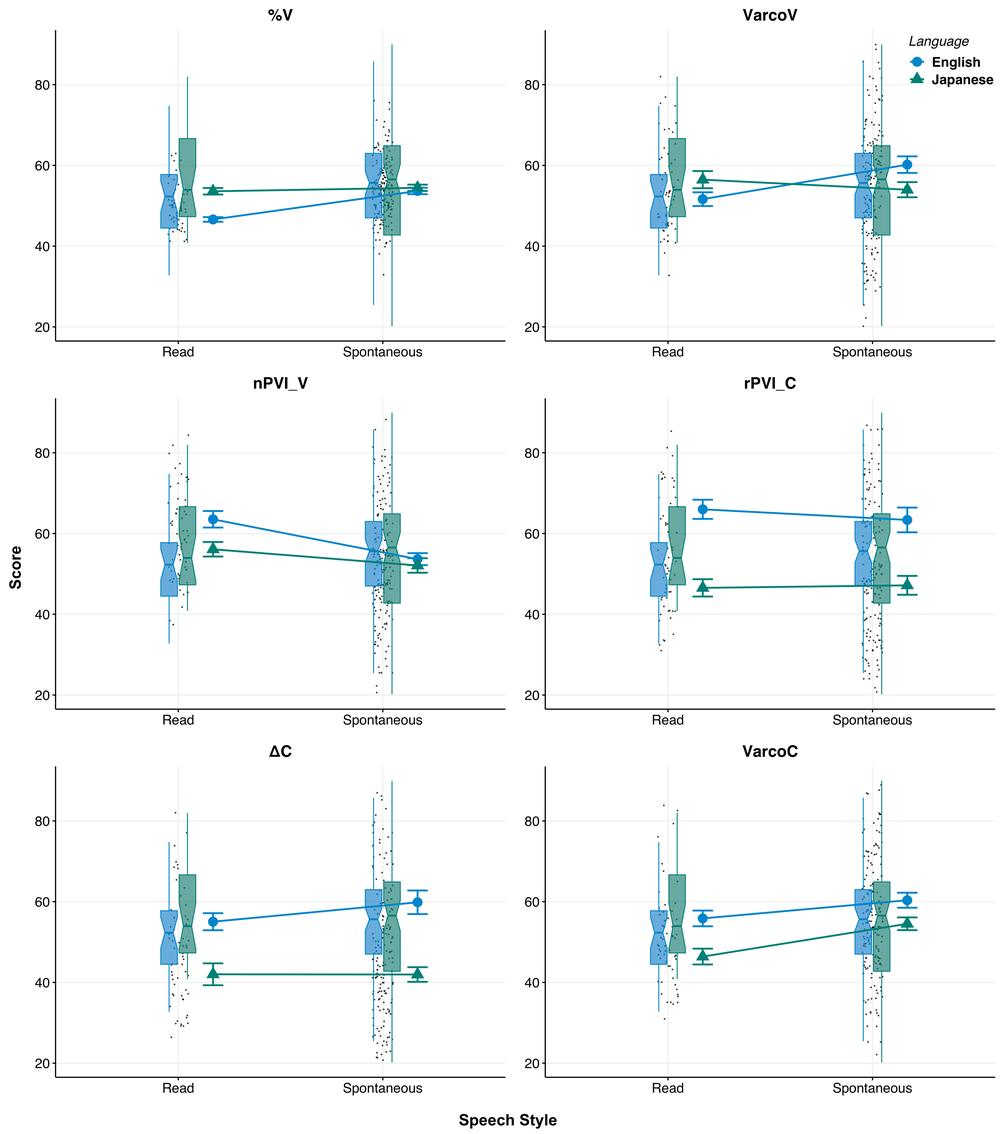
Segmentation:

- Segmented into "vocalic" and "consonantal" intervals, each of which may include one or more segments of the same type.
  - Recording of "fixed creation" read as [fɪkst.kɹi.eɪʃən] was segmented the consonantal intervals, [f], [kstkɹ], [ʃ], and [n]; and the vocalic intervals, [I], [ieɪ], and [ə].
- Pauses and disfluencies were excluded from the intervals.
- The read speech data was segmented in a similar manner (see Grenon and White [4]).

Durational Variability Measures:

- ΔC (standard deviation of consonantal interval duration)
- %V (percentage of total utterance duration composed of vocalic intervals)
- VarcoV (coefficient of variation of vocalic intervals)
- VarcoC (coefficient of variation of consonantal intervals)
- nPVI-V (mean of the differences of successive vocalic intervals divided by their sum and multiplied by 100)
- rPVI-C (mean of the differences of successive consonantal intervals)

2aSC8: Contact: yoichi.mukai@viu.ca



3. Results

Interaction Effects:

- The difference between spontaneous and read speech was significant only in English (higher **%V** for spontaneous speech); English and Japanese differed only in read speech for **%V**.
- Spontaneous speech showed higher **VarcoV** than read speech only in English; The languages differed only in spontaneous speech (higher **VarcoV** in English).

Main Effects of Speech Style:

- Spontaneous speech showed higher variability (**VarcoC)** than read speech.
- Read speech showed higher vowel variability (**nPVI-V**) than spontaneous speech.
- No main effect of speech style was found for **ΔC** or **rPVI-C**.

Main Effects of Language:

- Main effects of language were found for **VarcoC**, **ΔC**, and **rPVI-C**, but not for **nPVI-V**.
- **VarcoC** and **ΔC**: English showed higher variability compared to Japanese.
- **rPVI-C**: English showed higher variability than Japanese.
- **VarcoC**: variability was higher in spontaneous speech than read speech, and higher in English than in Japanese; it was the only measure that differentiated between spontaneous and read speech in Japanese.

4. Discussion

- Consonant variability (VarcoC) increased in spontaneous speech for both languages might suggest a potentially language-independent pattern of temporal variation [6]
- The different behaviours observed for vocalic variability patters highlight language-specific ways in which phonetic reduction manifests.
- **Limitations:** the use of data from different studies, potential dialectal variation within the Japanese speakers, and the limited amount of read speech data.
- **Future Research:** Use additional acoustic measures and investigating a wider range of languages is needed to better understand these patterns and the factors driving phonetic reduction.

References:
[1] Tucker, B. V., & Mukai, Y. (2023). Spontaneous Speech. Cambridge University Press.
[2] Hammond, M. (1999). The Phonology of English: A Prosodic Optimality-Theoretic Approach. Oxford University Press, Oxford, New York.
[3] Vance, T. J. (1987). An introduction to Japanese phonology. New York: University of New York Press.
[4] Grenon, I. and White, L. (2008). Acquiring rhythm: A comparison of L1 and L2 speakers of Canadian English and Japanese. In Proceedings of the 32nd Boston University conference on language development, pages 155–166, Boston. Boston University.
[5] Warner, N., Simonet,M., Tucker, B. V., Sonja Bird, Daniel Brenner, and Maureen Hoffmann (2015). Speech reduction across languages and dialects.
[6] Barry, W. and Andreeva, B. (2001). Cross-language similarities and differences in spontaneous speech patterns. Journal of the International Phonetic Association, 31(1):51–66.

The joint 188th Meeting of the Acoustical Society of America and 25th International Congress on Acoustics in New Orleans, Louisiana, USA on May 20, 2025