The Role of Pitch Variability in Holistic Language Processing and its Connection to Usage-based Grammatical Competence

Alvin Cheng-Hsien Chen

Department of English, National Taiwan Normal University (Taiwan)

While human speech is varied and complex, it is restricted by the biomechanical system's limitations. Research suggests a maximum speed at which humans can voluntarily alter their pitch [1-3]. Combined with Lindblom's [4] concept of effort economy, it is likely that pitch variation is utilized most effectively when its pragmatic function is most needed in communication. We propose that pitch variation within a single linguistic unit is minimized compared to the pitch variation within a complex structure involving multiple fundamental units. Specifically, we investigate the relationship between pitch variability of a lexical unit in spontaneous speech and its distributional properties in language use, with a particular focus on Taiwan Mandarin.

To estimate lexical distributional properties, we utilized the Corpus of Contemporary Taiwan Mandarin (COCT), and for spontaneous speech data, we used the Sinica Phone-Aligned Chinese Conversational Speech Database (SPCCSD). The COCT is an extensive collection of 185 million words covering a broad spectrum of topics and genres. The SPCCSD database comprises approximately 3.5 hours of spontaneous, in-person conversations from 16 speakers aged 16 to 46. We defined disyllabic words above a frequency cut-off in the COCT and identified a subset used in the SPCCSD to analyze the relationship between their pitch variability in the speech corpus and their distributional properties in the native corpus.

This study expands the duration-based Pairwise Variability Index (PVI) [5, 6] to include pitch, proposing two versions of pitch-related PVI to measure pitch variability within a speech segment. We used Praat's autocorrelation-based pitch tracking algorithm to extract and transform raw f0 values into semitones for each syllable of the disyllabic words found in the SPCCSD. We then calculated f0 pairwise variability index (f0PVI), using a representative f0 value at the maximum ($f0dbmax_k$) energy of each syllable in the word. This energy-based extraction of f0 values has been proven to produce more reliable f0 values and reduce tracking errors [6-9]. Our two proposals differed mainly in their normalization schemes. The first method, mean-based f0PVI, measured the variability of f0 differences between each sequential pair of syllables, normalizing the pairwise f0 difference by each syllable's f0 means ($f0mean_k$) as formulated in (1). The second method used each syllable's f0 standard deviations ($f0sd_k$) for normalization (i.e., [2]).

We assessed our pitch variability metrics' efficacy through three comprehensive analyses. Firstly, we compared the metric's values in disyllabic words with non-words (i.e., bigrams across word boundaries) to evaluate its ability to distinguish lexical units. Secondly, we examined the correlation between the f0PVI values of disyllabic words and their lexical frequency in the reference corpus. Lastly, we investigated the relationship between the word's f0PVI values and its lexical associations based on its preceding and subsequent linguistic context, using the normalized conditional probability values of delta P to measure the contingency of the disyllabic word given its preceding and subsequent words.

Our analyses revealed that disyllabic words had significantly lower f0PVI values than non-words and words with higher frequency demonstrated less pitch variability. Both the mean-based and SD-based f0PVI values yielded consistent results. The third analysis showed that lexical associations significantly affected a disyllabic word's pitch variability, with backward-directed lexical associations having a more significant effect on pitch variability than forward-directed ones. These findings suggest that the proposed pitch variability metric, f0PVI, can reliably measure structural integrity of a linguistic unit as well as its degree of entrenchment in use. This study underscores the importance of speakers' usage-based competence in modulating pitch variability in speech production and offers significant insights for psycholinguistic research.

Examples

(1) Mean-based f0PVI = $100 \times \left[\frac{\sum_{k=2}^{n}\left|\frac{f0dbmax_k - f0dbmax_{k-1}}{(f0mean_k + f0mean_{k-1})/2}\right|}{n-1}\right]$

(2) SD-based f0PVI = $100 \times \left[\frac{\sum_{k=2}^{n}\left|\frac{f0dbmax_k - f0dbmax_{k-1}}{(f0SD_k + f0SD_{k-1})/2}\right|}{n-1}\right]$

References

[1] Xu, Y. and X. Sun, *Maximum speed of pitch change and how it may relate to speech.* The Journal of the Acoustical Society of America, 2002. **111**: p. 1399-1413.

[2] Sundberg, J., *Maximum speed of pitch changes in singers and untrained subjects.* Journal of Phonetics, 1979. **7**(2): p. 71-79.

[3] Ohala, J.J. and W.G. Ewan, *Speed of pitch change.* The Journal of the Acoustical Society of America, 1973. **53**(1): p. 345-345.

[4] Lindblom, B., *Economy of speech gestures*, in *The production of speech*, P.F. MacNeilage, Editor. 1983, Springer: New York, NY. p. 217-245.

[5] Ling, L.E., E. Grabe, and F. Nolan, *Quantitative characterisation of speech rhythm: Syllable timing in Singapore English.* Language and Speech, 2000. **43**(4): p. 377-401.

[6] Grabe, E., G. Kochanski, and J. Coleman, *Connecting intonation labels to mathematical descriptions of fundamental frequency.* Language and Speech, 2007. **50**: p. 281-310.

[7] Ladd, D.R. and C. Johnson, *Metrical factors in the scaling of sentence-initial accent peaks.* Phonetica, 1987. **44**: p. 238-245.

[8] Chen, A.C.-H. and S.-C. Tseng, *Prosodic encoding in Mandarin spontaneous speech: Evidence for clause-based advanced planning in language production.* Journal of Phonetics, 2019. **76**: p. 1-22.

[9] Chen, A.C.-H., *F0-based pairwise variability index:A prosodic metric for holistic language processing*, in *20th International Congress of Phonetic Sciences*. 2023: Prague, Czech Republic. p. 1381-1385.