Articulatory Dynamics of Lexical Stress in L2 English: A Case Study of Taiwanese Mandarin Speakers

Paul McGuire, Feng-fan Hsieh, Yueh-chin Chang

National Tsing Hua University, Taiwan
graemepaulmcguire@gmail.com, ffhsieh@mx.nthu.edu.tw, ycchang@mx.nthu.edu.tw

Abstract

In this study involving ten participants, we simultaneously recorded acoustic and articulatory data using electromagnetic articulography (EMA) to investigate lexical stress production in L2 English among Taiwanese Mandarin (TWM) speakers. Analysis of dynamic time series data uncovered hyper-articulation of lingual articulators and the jaw in stressed syllables, alongside distinct tonal contours related to stress and syllable position. Furthermore, a gestural analysis revealed longer consonant gesture plateaus and longer CV lag in stressed syllables but no difference in gesture velocity.

Keywords: L2 English, Lexical stress, Taiwanese Mandarin, EMA

1. Introduction

Taiwanese Mandarin (TWM) is an East Asian tone language that lacks discernible word-level prominence. This study investigates the articulatory and acoustic correlates of stressed and unstressed syllables in L2 English as produced by native TWM speakers. While there have been similar studies, such as Kim's (2021) study involving speakers of Beijing Mandarin (BJM) and Shanghai Mandarin (SHM), our research departs substantially in its methodology and focus. We focus on TWM, a dialect which exhibits very limited use of the neutral tone—often cited as evidence of a trochaic foot in BJM. Crucially, our methodology differs by analysing dynamic trajectories of articulators over time through generalised additive mixed modelling (GAMM), moving beyond the static midpoint measurements used in earlier studies. In addition, we present an analysis of gestural duration and CV lag.

In L1 English, stressed syllables are described as 'hyper-articulated' (de Jong, 1995), while in Greek, they are said to involve 'longer, larger, and faster gestures than their unstressed counterparts' (Katsika and Tsai 2021). Acoustically, the key components used to distinguish stress in L1 English include intensity and duration (Fry 1955), vowel reduction (Delattre 1969), and fundamental frequency (F0) (Lieberman 1960) (although Pierrehumbert 1980 argues that F0 is related to pitch accent rather than being a direct correlate of stress). This exploratory study aims to determine which of the above articulatory and acoustic correlates, implicated in native stress production, are used by L1 TWM speakers in their L2 English production.

2. Methods

2.1. Participants, stimuli and recording procedures

Ten native speakers of Taiwanese Mandarin were recruited for this study. All participants were in their twenties and spoke only Mandarin in their daily life in Taiwan. This study focused on three disyllabic minimal pairs which differ only in stress location (CONflict - conFLICT, PROject - proJECT, DIgest - diGEST). The target words were embedded in the carrier phrase "Please say ____ again" and read in randomised order from a screen in a soundproof room . Eight participants completed ten repetitions of each word, whereas the remaining two could only complete seven repetitions of each due to time constraints. Articulatory data were recorded using EMA (Carstens AG501) at a sampling rate of 2,000 Hz, later down-sampled to 250 Hz. Sensors were attached to the lips, tongue, and lower incisor (for tracking jaw movement), as well as to the right and left mastoid processes and upper incisor (to correct for head movement). The sensors relevant to this study are TT (tongue tip), TB (tongue body), TD (tongue dorsum), LA (lip aperture - the euclidean distance between the upper and lower lip senors) and JAW (lower incisor). Acoustic data were recorded simultaneously at 24 kHz.

2.2. Articulatory measurements

Articulatory measurements were made in Matlab using Mview (Tiede 2005). For the vowel analysis, vocalic portions of the articulatory trajectories were identified using the acoustic data as a guide. For the gestural duration and CV lag analyses, articulatory gestures were identified using the *findgest* algorithm in Mview, which identifies gestural landmarks based on a peak velocity threshold of 20%. The following sensors were used to measure the syllable-initial consonant gestures: TDz (where 'z' indicates vertical movement) for [k]on, TBz for [dʒ]est and [dʒ]ect, TTz for [d]i and LA for [f]lict and [p]ro. For the vowel in the CV lag analysis, TDz was used for d[aɪ]. The gestural plateau, specifically the hold phase, was defined as the NOFFS (nucleus offset) timestamp minus the NONS (nucleus onset) timestamp, as shown in Figure 1. CV lag was defined as the interval between the NONS of the consonant and the NONS of the vowel. To control for speech rate, gesture durations were time normalised using an anchor point in the following word. In addition to the duration of gestural plateaus, peak velocity (towards the closure, i.e. sensor velocity at the PVEL (peak velocity) timestamp in Figure 1) and amplitude normalised peak velocity (stiffness; see Roon et al. 2021) were also measured.

 10.21437/issp.2024-61

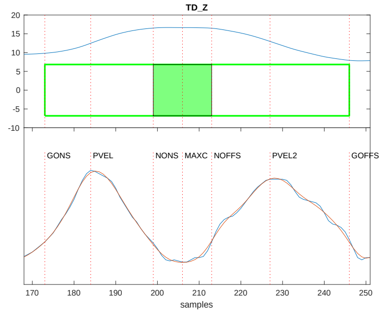

Figure 1: *Example of a [k] gesture—Upper panel: Vertical movement of tongue dorsum sensor; Lower panel: Velocity of sensor*

2.3. Acoustic measurements

The acoustic study investigated the realisation of stressed versus unstressed syllables in terms of three suprasegmental dimensions: intensity (dB), duration (ms), and F0 (Hz) and two segmental dimensions: F1 & F2 (Hz). Tokens were segmented in Praat (Boersma 2007) using text grids aligned to the start and end of the vocalic section within each syllable. Values were extracted with the help of ProsodyPro (Xu 2013) and FormantPro (Xu and Gao 2018).

2.4. Statistical analysis

Tokens were labelled according to the presence or absence of stress (`stress` = 1 or 0), and according to a combination of their stress value and their position in the disyllabic word (`syllpos` = l0, l1, r0 or r1), where 'l1' refers to a stressed syllable in the left position (i.e. con1) and 'r0' to an unstressed syllable on the right (i.e. flict0). Dynamic articulatory and acoustic data (F0, F1 & F2) were time-normalised, within-speaker z-scored and compared using generalised additive mixed modelling (GAMM) in R (based on recommendations from Wieling 2018). Statistical analyses of articulatory gestures and averaged acoustic data were within-speaker z-scored and carried out using linear mixed-effects modelling with the lme4 package (Douglas Bates, Bolker, and Walker 2015) and post hoc pairwise comparisons were calculated using the EMMEANS package (Lenth et al. 2018).

3. Results

3.1. Consonant gestures

Density plots for gestural plateau duration, stiffness and peak velocity are shown in Figure 2. Separate linear mixed-effects models were fitted for each measurement, with stress as the predictor. Random slopes and intercepts were included for syllable type. Among these three variables, only gesture duration demonstrated a statistically significant association (p = 0.005) - indicating that gesture duration is longer in stressed syllables. Stiffness (p = 0.299) and peak velocity (p = 0.415) did not show a significant relationship with stress. A linear mixed-effects model was fitted with `syllpos` as the independent variable (see Figure 3), and post hoc pairwise comparisons were conducted using estimated marginal means (EMMs) with Tukey adjustment for multiple comparisons. Out of the six combinations, only r0 - r1 was found to be significant (est. -0.564, p = 0.0063).

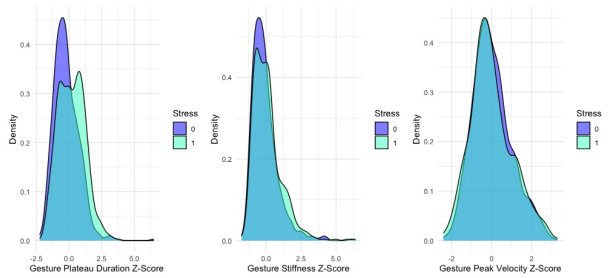

Figure 2: *Density plots of syllable initial consonant measurements; 0 = unstressed, 1 = stressed*

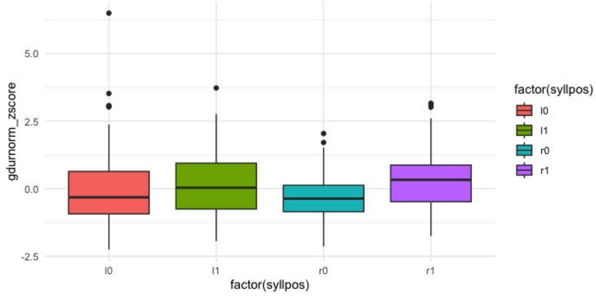

Figure 3: *Gesture plateau duration by syllable position; l0 = left position - unstressed, l1 = left position - stressed, r0 = right position - unstressed, r1 = right position - stressed*

3.2. CV lag

Only the syllable pair 'DI/di' was subjected to CV lag analysis due to the challenges in delineating gestures for other stimuli. These challenges arose from shared articulators between consonants and vowels, along with complex syllable onsets, which made it difficult to clearly identify relevant gestures. Figure 4 presents density plots for stressed and unstressed pair: 'DI/di'. Another linear mixed-effects model was fitted, using the same formula as that used for the consonant gestures, with z-scored, time-normalised CV lag as the dependent variable. The stressed syllable 'DI' had significantly longer CV lag than unstressed 'di' (p = 0.0001).

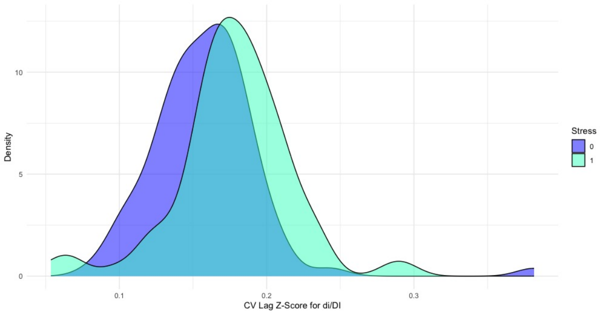

Figure 4: *Density plots of z-scored time normalised CV lag for 'di/DI'; 0 = unstressed, 1 = stressed*

3.3. Acoustic means

Linear mixed-effects models were fitted for each of the suprasegmental measurements of interest (mean values of in-

Table 1: Vowel GAMM articulatory / F1-F2 results (x = front/back; z = high/low)

	TDz	TDx	TBz	TBx	TTz	TTx	JAWz	JAWx	F1	F2
CON	* inferior		* inferior		* inferior		* inferior			
FLICT	* superior	* anterior		* anterior						
DI	* superior		* inferior		* inferior	* posterior	* inferior	* posterior		
GEST					* inferior		* inferior			
PRO	* inferior		* inferior			* posterior	* inferior	* posterior	* high	* low
JECT					* inferior		* inferior		* low	* high

tensity, F0 and duration) using the same formula used in the consonant measurement analysis. All acoustic measurements showed significant differences between stressed and unstressed syllables. Stressed syllables were associated with higher values of mean intensity (p = 1.7e-06), mean F0 (p = 0.0001) and mean duration (p = 0.005). As with gesture durations, a linear mixed-effects model was fitted with the factor `syllpos` as the independent variable and post hoc pairwise comparisons were computed. For intensity, all combinations of stressed versus unstressed syllables differed significantly, only l0 - r0, and l1 - r1 showed no significant difference. F0 differed significantly for all combinations other than l0 - r0. The combination of l1 - r1 differed significantly, indicating higher F0 in stressed syllables in the left position of the disyllabic word than those in the right position (p = 0.0055). Duration differed significantly between l0 - l1, l1 - r0 and, as with F0, between l1 - r1 (with stressed syllables on the left being longer. p = 0.0386). Interestingly, the difference in acoustic duration between r0 and r1 did not reach statistical significance, contrasting with the gestural duration analysis, in which r0 - r1 was the only significantly different combination.

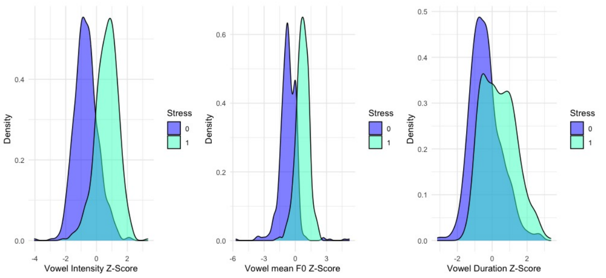

Figure 5: *Density plots of acoustic measurements of vowels; 0 = unstressed, 1 = stressed*

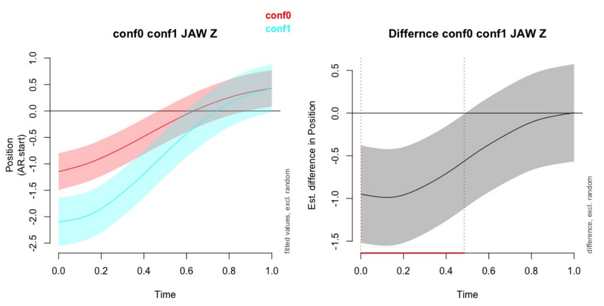

Figure 6: *GAMM smooth and difference plots of JAWz (vertical movement) for 'con'/'CON'*

3.4. Dynamic articulatory & acoustic analysis

3.4.1. Articulator and formant trajectories

The results of the vowel analysis are presented in Table 1. Asterisks denote significantly different articulator trajectories between the stressed and unstressed vowels that are continuous for a portion comprising at least 15% of the vocalic section (see Figure 6). Anatomical directions—superior, inferior, anterior, and posterior—refer to the position of the articulator in the stressed syllable (i.e. 'CON') relative to its position in the syllable's unstressed counterpart (i.e. 'con'), during the portion where significant difference is observed. Similarly, formant values are labelled 'high' or 'low' to denote the stressed syllable's relative value during the window of significant difference.

The results indicate that stressed vowels were most consistently associated with larger jaw displacement, with all stressed vowels other than 'FLICT' showing significantly more inferior jaw positions than their unstressed counterparts. Hyper-articulation of the lingual articulators was also observed in at least one dimension in every stressed syllable. Significantly different formant trajectories were found only for 'PRO' and 'JECT'. The other four syllables showed significant differences in articulatory trajectories despite showing no acoustic difference in terms of the first and second formants.

3.4.2. F0 trajectory

A GAMM analysis of F0 trajectories over time across the four levels of `syllpos` revealed that stressed syllables are not only produced with higher F0 but also appear to be produced with distinct tonal contours as a function of the interaction between stress and the position within the disyllabic word. As shown in Figure 7, stressed syllables in the left position are produced with a steady high tone, whereas stressed syllables on the right drop abruptly from their highest point, roughly approximating Mandarin Chinese's 'first' and 'fourth' tones, respectively. Unstressed syllables on the right descended into creaky voice phonation for most of the speakers, whereas those on the left we produced with a more level tone in the middle of the speakers' pitch range.

4. Discussion and conclusion

The results of the acoustic analysis showed that, suprasegmentally, stressed syllables were found to be positively correlated with F0, duration and intensity. Segmentally, significant differences were observed in the vowel formant trajectories between 'PRO/pro' and JECT/ject', whereas no such differences were found between 'CON/con' and 'GEST/gest', despite these syllables sharing a phonemic vowel and occupying identical positions within the word. That stress is a associated with F1

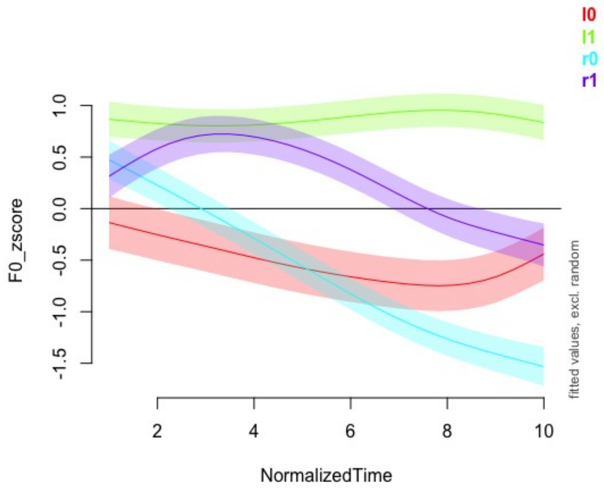

Figure 7: *F0 contours by syllable position and stress; l0 = left position - unstressed, l1 = left position - stressed, r0 = right position - unstressed, r1 = right position - stressed*

& F2 difference in 'PRO' but not 'CON', could be to do with the effect of the articulation of the rhotic in 'PRO'. Regarding 'JECT' and 'GEST', data visualisation revealed a pattern of centralisation in 'GEST' similar to that observed in 'JECT'; however, this difference did not reach statistical significance. GAMM analyses of F0 trajectories further uncovered distinct F0 contours arising from the interaction between stress and syllable position, details that would have been overlooked in a study focusing on measurements from static points.

The results of the articulatory analysis suggest that in Taiwanese Mandarin-accented English, stressed vowels correlate with greater jaw displacement and exhibit significant 'hyper-articulation' of the lingual articulators, even when the vowels are segmentally identical in terms of their first and second formants. Furthermore, consonant gesture plateau duration and CV lag were found to be significantly longer in stressed syllables. Interestingly, Kim's (2021) study suggests that the stressed syllables do not involve substantial supra-glottal hyper-articulation in L2 English by speakers of Standard Chinese. This discrepancy could be attributed to several potential confounding factors: the contrast between spontaneous and laboratory speech, the difference between point-to-point comparison and trajectory analysis of EMA sensors, and variations across Mandarin dialects.

5. Acknowledgements

This study was supported by a research grant (MOST 110-2410-H-007-025), awarded to the second author, to which we are grateful.

6. References

Boersma, Paul (2007). "Praat: doing phonetics by computer". In: *http://www. praat. org/*.

Delattre, Pierre (1969). "An acoustic and articulatory study of vowel reduction in four languages". In.

Douglas Bates, MM, Ben Bolker, and Steve Walker (2015). "Fitting linear mixed-effects models using lme4". In: *Journal of Statistical Software* 67.1, pp. 1–48.

Fry, Dennis B (1955). "Duration and intensity as physical correlates of linguistic stress". In: *The Journal of the Acoustical Society of America* 27.4, pp. 765–768.

Katsika, Argyro and Karen Tsai (2021). "The supralaryngeal articulation of stress and accent in Greek". In: *Journal of phonetics* 88, p. 101085.

Kim, Boram (2021). "Lexical Stress Realization in Mandarin Second Language Learners of English: An Acoustic and Articulatory Study". In.

Lenth, Russell V., Henrik Singmann, Jonathon Love, Paul Buerkner, and Maximilian Herve (2018). *Package "Emmeans"*. R package version 4.0-3. URL: `http://cran.r-project.org/package=emmeans`.

Lieberman, Philip (1960). "Some acoustic correlates of word stress in American English". In: *The Journal of the Acoustical Society of America* 32.4, pp. 451–454.

Pierrehumbert, Janet Breckenridge (1980). "The phonology and phonetics of English intonation". PhD thesis. Massachusetts Institute of Technology.

Roon, Kevin D, Philip Hoole, Chakir Zeroual, Shihao Du, and Adamantios I Gafos (2021). "Stiffness and articulatory overlap in Moroccan Arabic consonant clusters". In: *Laboratory Phonology: Journal of the Association for Laboratory Phonology* 12.1, pp. 1–23.

Tiede, Mark (2005). "MVIEW: software for visualization and analysis of concurrently recorded movement data". In: *New Haven, CT: Haskins Laboratories*.

Wieling, Martijn (2018). "Analyzing dynamic phonetic data using generalized additive mixed modeling: A tutorial focusing on articulatory differences between L1 and L2 speakers of English". In: *Journal of Phonetics* 70, pp. 86–116.

Xu, Yi (2013). "ProsodyPro—A tool for large-scale systematic prosody analysis". In: Laboratoire Parole et Langage, France.

Xu, Yi and Hong Gao (2018). "FormantPro as a tool for speech analysis and segmentation". In: *Revista de Estudos da Linguagem* 26.4, pp. 1435–1454.