THE EFFECT OF PROSODIC PROMINENCE ON DIPHTHONG REDUCTION IN TAIWAN MANDARIN SPONTANEOUS SPEECH –USING /aɪ/ AS AN EXAMPLE

Chieh-Ching Chen[1], Janice Fon[1,2,3]

[1]Graduate Institute of Linguistics, National Taiwan University (NTU), [2]Graduate Institute of Brain and Mind Sciences, NTU, [3]Neurobiology and Cognitive Science Center, NTU
r10142003@ntu.edu.tw, jfon@ntu.edu.tw

ABSTRACT

Vowel reduction is frequently observed in spontaneous speech. Previous studies have shown that acoustic reduction of monophthongs was manifested through formant undershoot and shorter duration. For diphthongs, which differ from monophthongs by having an additional vowel target, little is said about whether and how reduction should take place. This study thus examines how the reduction of the diphthong /aɪ/ is realised in Taiwan Mandarin and how stress influences the reduction pattern. Results showed that /aɪ/ reduction is mostly in the form of monophthongization with two vowel targets merged. Besides, diphthongs without stress were shorter in duration with formant undershoot on both vowel targets. Finally, the reduction of /aɪ/ likely lacks a negative connotation since no gender difference was found.

Keywords: vowel reduction, diphthong, prosodic prominence, spontaneous speech, Taiwan Mandarin.

1. INTRODUCTION

Vowel reduction can be classified as lexical vowel reduction and acoustic vowel reduction [1, 2]. While lexical reduction is a phonological process that can only be observed on particular words (tel**e**graphy [ɛ] vs. tel**e**graphic [ə]), acoustic reduction is a phonetic process in that vowels are realised with shorter duration and formant undershoot [3]. Previous studies showed that reduction appears more frequently in spontaneous speech [4], and prosodic prominence greatly influences the reduction of vowels. For instance, vowels without lexical stress were more likely to be reduced than those with lexical stress in English and Swedish [1, 3]. In languages without a lexical stress distinction, such as French [5], vowels in non-final positions of a word were more reduced [6]. This study thus intends to examine whether a similar relationship between stress and vowel realisation could be found in Taiwan Mandarin, which is prosodically disparate, and if so, what the exact pattern is.

As a tone language, Mandarin assigns one of the four lexical tones to most of its syllables [7]. The tones are realised fully in prosodically strong positions but are reduced in prosodically weak positions [8]. As prosodic strength is manifested through tonal realisation, not vowel reduction itself, as is more commonly found in stress-timed languages like English, it would be interesting to see whether and how vowel reduction is affected by stress.

Moreover, previous studies on vowel reduction mainly focused on monophthongs [1, 2, 3, 6]. How diphthongs are reduced is less clear. Diphthongs differ from monophthongs by having two vowel targets [9], between which is an additional vowel transition [10], whose duration is vowel-dependent and language-specific [11]. Using 18-min spontaneous speech, Su [12] found that 12% of the four diphthongs in Mandarin were monophthongized. However, the study did not examine whether the reduction was caused by prosodic weakening, nor did it look into how reduced diphthongs were realised. Therefore, in this study, we would like to examine the actual realisation of diphthongs when reduced with larger spontaneous speech data.

/aɪ/ was chosen in this study for two reasons. First, /aɪ/ is the most productive diphthong in Taiwan Mandarin [13]. Therefore, its reduction could show us how diphthongs are generally reduced in the language. Secondly, the distance between the two vowel targets is large for formant measurements. We can visualise different states of diphthong reduction by examining /aɪ/. For example, there might be formant undershoot on either of the two vowel targets, which results in /aɪ/ reducing into other diphthongs. On the other hand, /aɪ/ could be monophthongized with the two targets merged as [e], one target dropped as [a], or centralised as [ə].

The aim of this study could thus be specified as whether and how diphthong /aɪ/ is reduced and affected by stress in Taiwan Mandarin. It was expected that /aɪ/ at a lower stress level would have a higher reduction rate, shorter duration, and more reduced spectral results. Also, males and females would be analysed separately since gender difference was observed in acoustic reduction [6]. In our prediction, males would have a higher reduction rate since in stable sociolinguistic stratification, men use a higher frequency of nonstandard forms than women [14].

2. METHODS

Four hours of monologue recordings including 8 speakers (4M, 4F) of the Taipei dialect were chosen from the Taiwan Mandarin-Min Spontaneous Speech Bilingual Corpus [15]. All speakers were young (20-35) fluent Mandarin-Min bilinguals. Each speaker talked for about 30 minutes.

2.1. Vowel labelling

All realisations of the vowel /aɪ/ were manually labelled and transcribed by two transcribers and inter-labeller consistency was later checked by the first author. Visually identifiable F1 and F2 were used as major cues to label the vowels regardless of voice quality.

2.2. Stress labelling

The stress level of each syllable was labelled primarily based on principles outlined in the Pan-Mandarin Tone and Break Indices (M-ToBI) [8]. There are three levels of stress, S1 to S3. S1 is the lowest level of stress and is used when a tone has completely lost its tonal shape. S2 is the next higher level. It is used when the tone still retains its distinctive contour, even though some of its tonal specifications have been lost. S3 is the highest level of stress. It is used to label tones that are realised with a full-fledged contour. However, necessary modifications were made to accommodate the peculiarities of spontaneous speech [16], as the original version of M-ToBI was solely based on read speech. In addition to tonal realisation, amplitude was considered an additional important cue in stress labelling.

In spontaneous speech, S2 is the most common and could be considered the default stress level. Detailed labelling criteria can be seen in Table 1.

Stress	Tone	Amplitude
S1	loss of original tonal shape	soft
S2	default	default
S3	tone expanded/raised	loud

Table 1: Labelling criteria of stress suggested for spontaneous speech in Chuang [16].

2.3. Measurements

Ten F1 and F2 values were extracted at equal time intervals from each token with Praat [17]. Formant settings were adjusted to accommodate gender differences.

3.RESULTS

In total, 2,835 tokens of /aɪ/ were labelled, in which males had 1,370 tokens and females had 1,465 tokens. Table 2 summarises the distribution of /aɪ/ realisations. Six tokens were excluded from the following analyses due to laughter or noise resulting in unrecognisable formant values. A total of 47 tokens longer than 300 milliseconds were regarded as hesitation [6] and were thus excluded. Five truncated tokens were also excluded since they were out of the domain of this study. In the end, there were 2,777 tokens for later analyses. Detailed distribution is shown in Table 3.

Retained	Monophthongs			Diphthongs		Else
[aɪ]	[e]	[ə]	[a]	[eɪ]	[aə]	-
1,274	968	286	135	87	24	3

Table 2: Distribution of different realisations of /aɪ/.

Gender	Stress	[aɪ]	[eɪ]	[aə]	[e]	[ə]	[a]
M	S1	76	6	5	92	55	24
	S2	466	24	14	396	94	54
	S3	27	1	-	1	-	-
F	S1	62	7	1	86	56	22
	S2	562	47	4	388	81	35
	S3	81	2	-	5	-	-

Table 3: Number of phonetic realisations of /aɪ/ across gender and stress.

3.1. Overall diphthong reduction rate

In the four-hour recording, only 45.89% tokens of /aɪ/ were retained. 50.04% were monophthongized and 4.07% were reduced to other diphthongs. Monophthongization is mainly in the form of merging rather than deletion, as [e] occupied 69.69% of the reduced monophthongs, and 20.59% were centralised as [ə]. Only 9.72% were reduced to [a], in which the second vowel target was dropped.

3.2. The ratio of reduced /aɪ/ across groups

The ratio of /aɪ/ reduction was calculated from the number of neutralised /aɪ/, including monophthongs and diphthongs, divided by the sum of all tokens spoken in the group. Logistic regression was performed to analyse the effect of Stress (3) and Gender (2) on the reduction rate. Table 4 summarises the statistical results. Two main effects Stress1 and Stress3 were significant. Specifically, S3 had more /aɪ/ retained than S2 ($p < .001$), while in contrast, S1 had more /aɪ/ reduced than S2 ($p < .001$). Nevertheless, gender difference was not found (*p*

= .213). The result of the ratio of retained and neutralised /aɪ/ is presented in Fig. 1.

	Est.	SE	z value	Pr
(Intercept)	-0.00	0.12	0.00	.998
stress1	-1.03	0.16	-6.43	.000***
stress3	2.40	0.40	6.02	.000***
genderM	-0.21	0.17	-1.25	.213
stress1:genderM	0.43	0.22	1.93	.053
stress3:genderM	0.45	0.84	0.54	.590

Table 4: Logistic regression results of the ratio of reduced /aɪ/.

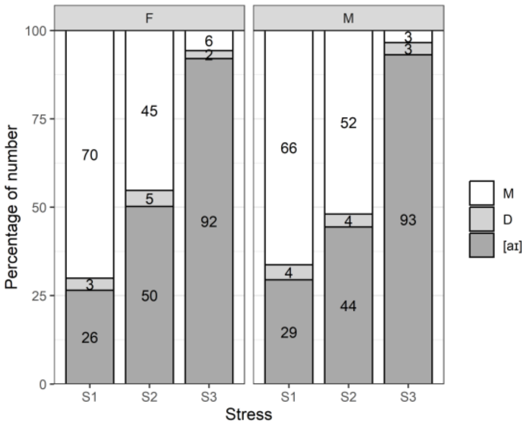

Figure 1: Proportions of neutralised monophthongs (M), neutralised diphthongs (D), and retained /aɪ/ across stress and gender.

3.3. Duration

To investigate whether differences exist across realisations and three levels of stress on duration across groups, a linear mixed effects model (LMM) was performed. Reduction (2) × Stress (3) × Gender (2) were set as independent variables, in which Reduction was defined as whether the realisation was [aɪ] or not. Retained /aɪ/, S2, and female was set as the reference level. Table 5 summarises the result of the main effects. Results showed that reduced /aɪ/ was shorter (p < .001). Besides, S1 was significantly shorter than S2, while S3 was significantly longer than S2. No specific gender effect was found for duration. Fig. 2 presents the mean duration across the three levels of stress and realisations.

	Est.	SE	t value	Pr
Intercept	0.13	0.01	25.44	.000***
Reduction	-0.03	0.00	-15.15	.000***
Stress1	-0.02	0.01	-4.86	.000***
Stress3	0.03	0.00	6.83	.000***
GenderM	0.00	0.01	0.50	.615

Table 5: The results of LMM on duration.

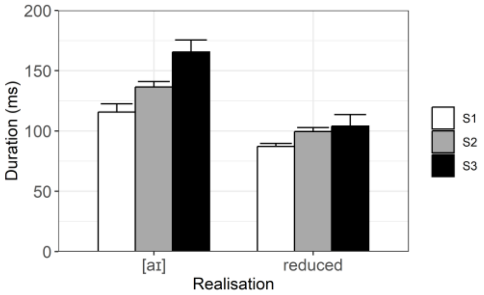

Figure 2: Mean duration across realisations and stress.

3.4. Spectral results

To examine the effect of stress on formant frequencies, F1 and F2 trajectories of retained [aɪ] in S1, S2, and S3 were evaluated. The package lme4 for linear mixed effects models in R was fitted [18]. Stress and Time (t) were the independent variables, and the F1 and F2 values of males and females were the dependent variables. In the analysis, the reference category for Stress was S2.

For both females and males, the interaction of Stress3 × Time and two main effects Stress3 and Time were found in F1. Specifically, S3 was significantly higher than S2 at the beginning; however, as the time went by, S3 went significantly more downwards than S2. Additionally, females had lower and flatter F1 for S1. Tables 6 and 7 present the LMM results of F1 for both genders. On the other hand, results of F2 in Tables 8 and 9 show a main effect Stress1 for females and a Stress3 × Time interaction effect for both males and females. Specifically, S3 went higher than S2 as time went by. Fig. 3. presents the result of formant frequencies across three levels of stress.

	Est.	SE	t value	Pr
Intercept	872.35	30.57	28.54	.000***
Stress1	-50.43	11.22	-4.50	.000***
Stress3	78.78	9.97	7.90	.000***
Time	-38.64	0.66	-58.36	.000***
Stress1:t	8.17	2.10	3.89	.000***
Stress3:t	-15.00	1.87	-8.04	.000***

Table 6: Results of female F1.

	Est.	SE	t value	Pr
Intercept	759.62	22.07	34.41	.000***
Stress1	7.00	7.59	0.92	.356
Stress3	30.09	12.09	2.49	.013*
Time	-29.00	0.53	-54.75	.000***
Stress1:t	0.71	1.41	0.50	.617
Stress3:t	-8.40	2.26	-3.71	.000***

Table 7: Results of male F1.

	Est.	SE	t value	Pr
Intercept	1801.10	15.31	117.63	.000***
Stress1	-34.81	15.43	-2.26	.002*
Stress3	-25.83	13.72	-1.88	.006
Time	50.81	0.91	55.76	.000***
Stress1:t	-2.37	2.89	-0.82	.411
Stress3:t	23.77	2.57	9.26	.000***

Table 8: Results of female F2.

	Est.	SE	t value	Pr
Intercept	1482.03	47.20	31.40	.000***
Stress1	-2.74	12.24	-0.22	.823
Stress3	-27.83	19.51	-1.43	.154
Time	47.78	0.85	55.91	.000***
Stress1:t	-3.60	2.28	-1.58	.114
Stress3:t	23.85	3.65	6.53	.000***

Table 9: Results of male F2.

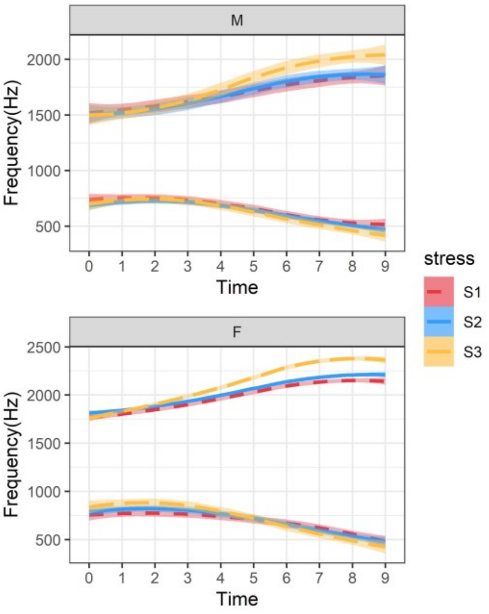

Figure 3: Spectral results of retained [aɪ] across S2 and S3.

4. DISCUSSION

In this study, we found that the diphthong /aɪ/ is prone to be reduced to monophthongs in spontaneous speech, and reduction is mainly caused by target merging, not target drop. Besides, with larger data, we observed a higher monophthongization rate than what was indicated in Su [12]. The difference lies in the fact that we looked at all tokens of /aɪ/ rather than the first 100 occurrences in the speech. Also, our speakers talked for 30 minutes instead of 3 minutes. It is very likely that our result differed from Su because speakers become more relaxed in longer monologues.

Moreover, we found that vowel reduction was correlated with tonal reduction. The effect of stress was manifested in the reduction rate, duration, and formant frequencies. In general, S2 and S1 have higher reduction rates and shorter duration, while in formant frequencies, S2 has more centralised spectral results compared with S3. Our result of unstressed vowels having shorter duration is similar to the result of lexical stress discussed in English, Swedish, and Dutch [1, 2, 3]. Besides, we found that the shorter duration was not only manifested between retained and reduced /aɪ/ but also observed across the three levels of stress within retained /aɪ/ and reduced /aɪ/. Retained /aɪ/ is longer than reduced /aɪ/. S3 is the longest, while S2 is second, and S1 is the shortest.

In spectral results, both F1 and F2 of the retained /aɪ/ were affected by stress. In the result of F1, S2 was lower at the beginning and higher in the later part compared with S3. In other words, in unstressed [aɪ], the tongue position for [a] was higher and [ɪ] was lower. The effect was further highlighted in females' S1, in which the tongue position for [a] was even higher and that for [ɪ] much lower compared with their tokens in S2.

On the other hand, the result of F2 shows that S2 went lower than S3 in the later part of the trajectory, which means that [ɪ] was less fronted when unstressed. Our result thus followed the pattern observed in the reduction of monophthongs [1, 2, 3, 6], in which unstressed vowels were more likely to be centralised.

Interestingly, gender difference was not observed either in terms of reduction rate or duration. Our result differed from what was found in French in that males had more reduced vowel space [6]. However, we only examined /aɪ/ and its reduction in this study. Therefore, it can be inferred that the gender difference in reduction might be manifested in the overall vowel space rather than the mere reduction of /aɪ/. Moreover, since we found that males produce nonstandard forms as frequently as females, it is likely that the reduction of /aɪ/ did not have a negative connotation [14].

5. CONCLUSION

In this study, we examined the reduction of /aɪ/ in terms of possible realisations, reduction rate, duration, and formant frequencies. The diphthong /aɪ/ was more reduced as two vowel targets merged and reduction does not differ across genders. Also, we found that the prosodic prominence in Mandarin has an effect on the /aɪ/ reduction. The effect was not only observed in reduction rate, but also in the durational and spectral results of retained /aɪ/. In sum, diphthongs realised with more reduced tonal shapes tended to have shorter duration and more centralised spectral results.

6. REFERENCES

[1] M. Fourakis, "Tempo, stress, and vowel reduction in American English," *The Journal of the Acoustical society of America,* vol. 90, no. 4, pp. 1816-1827, 1991.

[2] D. R. Van Bergem, "Acoustic vowel reduction as a function of sentence accent, word stress, and word class," *Speech communication,* vol. 12, no. 1, pp. 1-23, 1993.

[3] B. Lindblom, "Spectrographic study of vowel reduction," *The journal of the Acoustical society of America,* vol. 35, no. 11, pp. 1773-1781, 1963.

[4] M. Nakamura, K. Iwano, and S. Furui, "Differences between acoustic characteristics of spontaneous and read speech and their effects on speech recognition performance," *Computer Speech & Language,* vol. 22, no. 2, pp. 171-184, 2008.

[5] D. J. Hirst, A. Di Cristo, and Y. Nishinuma, "Prosodic parameters of French: A cross-language approach," *Contrastive studies of Japanese and other languages series,* pp. 7-20, 2001.

[6] C. Meunier and R. Espesser, "Vowel reduction in conversational speech in French: The role of lexical factors," *Journal of Phonetics,* vol. 39, no. 3, pp. 271-278, 2011.

[7] C. N. Li and S. A. Thompson, "Chinese," in *The world's major languages*: Routledge, 2003, pp. 811-833.

[8] S.-h. Peng, M. K. Chan, C.-y. Tseng, T. Huang, O. J. Lee, and M. E. Beckman, "Towards a Pan-Mandarin system for prosodic transcription," *Prosodic typology: The phonology of intonation and phrasing,* pp. 230-270, 2005.

[9] C. Gussenhoven and H. Jacobs, *Understanding phonology*. Routledge, 2017.

[10] I. Lehiste and G. E. Peterson, "Transitions, glides, and diphthongs," *The journal of The acoustical society of America,* vol. 33, no. 3, pp. 268-277, 1961.

[11] M. Lindau, K. Norlin, and J.-O. Svantesson, "Some cross-linguistic differences in diphthongs," *Journal of the International Phonetic Association,* vol. 20, no. 1, pp. 10-14, 1990.

[12] T.-t. Su, "Using the same methodology to compare reduction and assimilation phenomena in spontaneous French and Taiwanese Mandarin," in *Proceedings of the 15-th International Congress of Phonetic Sciences, Barcelona, Spain*, 2003.

[13] S.-C. Tseng, "Spontaneous Mandarin production: results of a corpus-based study," in *2004 International Symposium on Chinese Spoken Language Processing*, 2004: IEEE, pp. 29-32.

[14] W. Labov, "The intersection of sex and social class in the course of linguistic change," *Language variation and change,* vol. 2, no. 2, pp. 205-254, 1990.

[15] J. Fon, "A preliminary construction of Taiwan Southern Min spontaneous speech corpus," National Science Council technical report [NSC-92-2411-H-003-050-], 2004.

[16] Y.-Y. Chuang and J. Fon, "The effect of prosodic prominence on the realizations of voiceless dental and retroflex sibilants in Taiwan Mandarin spontaneous speech," in *Speech Prosody 2010-Fifth International Conference*, 2010.

[17] P. Boersma and D. Weenink, "Praat: doing phonetics by computer (Version 6.1. 51)[Computer software]," ed, 2021.

[18] D. Bates, M. Mächler, B. Bolker, and S. Walker, "Fitting linear mixed-effects models using lme4," *arXiv preprint arXiv:1406.5823,* 2014.