SYLLABLE CONTRACTION IN TAIWAN MANDARIN

Grace Kuo

Department of Linguistics, University of California, Los Angeles, USA
gracekuo@humnet.ucla.edu

ABSTRACT

Taiwan Mandarin syllable contraction is a lenition process which involves the elision of the intervocalic segments and the merger of the tonal elements of two syllables. Trough depth is employed as the measure for gradience of syllable contraction in this study and the distribution of trough depth shows that syllable contraction is optional. In the perception experiment, listeners were asked to do a forced-choice identification task and the accuracy was generally high. The production results verify that syllable contraction is optional and gradient, and the duration and F0 range are the acoustic cues that contribute to the distinction between the fully-contracted tokens and the lexical tokens.

Keywords: syllable contraction, trough depth, gradience, optional

1. INTRODUCTION

Contraction is a process that combines two or more words into one and occurs in rapid or connected speech. Discussions of contraction typically characterize it as a lenition process with articulatory reduction or elision. Syllable contraction is also very prevalent in Taiwan Mandarin. Some contractions from earlier forms of Mandarin are even fossilized and can appear in both formal and informal speech.

The occurrence of Taiwan Mandarin syllable contraction can be optional and the sound variations of syllable contraction can be gradient. For instance, with an increase in speech rate, a disyllabic phrase *na51-jaŋ51* can be contracted into a single word with gradual sound changes, and the meanings of *na51-jaŋ51* and the fully-contracted *niaŋ51* are not distinctive so that they can substitute for each other interchangeably. However, an issue of neutralization may arise. In some cases, such as the contracted *niaŋ51*, whose form happens to also be an eligible and existing Mandarin word, meaning 'to brew'. In others words, the interpretation of a monosyllable *niaŋ51* can be either a lexical word or a fully-contracted form derived from the disyllabic phrase *na51-jaŋ51* by means of segment deletion and re-syllabification.

In the current study, we would like to see if Taiwan Mandarin listeners could differentiate the fully-contracted tokens from the lexical tokens. If they do, how are the two tokens different from each other in terms of the acoustic measurements. A perceptual experiment with a forced-choice task and a production experiment with duration, spectral and prosodic measures are conducted.

2. PERCEPTION EXPERIMENT

A perception experiment with a forced-choice task was conducted to see if Taiwan Mandarin listeners are able to distinguish fully-contracted tokens from the lexical tokens.

2.1. Methods

Thirty-five native Taiwanese listeners participated in the experiment. The experiment was conducted in a sound treated booth at the Phonetics Lab. Listeners were presented with an audio file with 10 fully-contracted tokens (e.g. (1a)), 40 lexical tokens (e.g. (1b)) and 30 contractible but not fully-contracted tokens (e.g. (1a)) half of each token type were spoken by males and the other half were spoken by females.

(1) a. "He says it is not easy to last for so long."
tʰa55 ʂuo55 ***na51-jaŋ51*** *tɕiou21 pu51 ʐoŋ35-ji51*
he say **such** long not easy

b. "He says it is not easy to brew wine."
tʰa55 ʂuo55 ***niaŋ51*** *tɕiou21 pu51 ʐoŋ35-ji51*
he say **brew** wine not easy

In each trial, they would see two sentences presented on the screen and then they were presented one token from the stimuli. Their task was to identify which sentence they had heard.

2.2. Results

The result shows that listeners could correctly identify the fully-contracted tokens 90.29% of the time and correctly identify the lexical tokens 90.86% of the time. This result can be tested in

two ways: First, a binomial test shows that this level of accuracy is above chance. Second, a paired t-test comparing correctness for lexical vs. fully-contracted tokens by listeners shows no difference *(p* =.944*)*. Thus, listeners can identify both lexical and fully contracted tokens well, meaning that syllable contraction in these tokens cannot be a complete neutralization.

Taiwan Mandarin syllable contraction is like word-final devoicing from the perception point of view. Word-final devoicing is widely attested among languages and has often been assumed to be completely neutralized. Take German word-final devoicing for example (e.g. [1, 2, 5]). In German, the underlying /d/ surfaces as [t] in word-final position and the underlying /t/ surfaces as [t] everywhere. However, /d/ and /t/ are often recognized as intended in connected speech despite the impoverished acoustic information. In fact, listeners were able to differentiate devoiced final obstruents from the voiceless final obstruents 60% of the time, which is better than chance (e.g. [5]). Although there was a considerable overlap between the final voiced-voiceless obstruent pairs, the acoustic parameters, particularly duration, showed significant though weak differences. Therefore, German word-final devoicing is not a process of complete neutralization since the voiced-voiceless obstruents are different from each other acoustically and perceptually.

In the current study, given that Taiwan Mandarin listeners could differentiate the fully-contracted and lexical tokens very well, we would like to see how the fully-contracted tokens and lexical tokens in Taiwan Mandarin are different from each other acoustically. Therefore a production experiment was carried out and duration, spectral and prosodic measures were investigated. There were two issues that are concerned: whether Taiwan Mandarin syllable contraction is optional and gradient and whether it is neutralizing, i.e. whether there are some acoustic cues that distinguish the fully-contracted tokens from their lexical counterparts.

3. PRODUCTION EXPERIMENT

3.1. Methods

3.1.1. Speakers

Five female and five male Taiwanese native speakers participated in this study.

3.1.2. Materials

Materials consisted of 84 sentences, among which 40 sentences were test sentences and the rest were filler sentences. Half of the test sentences contained disyllabic contractible items and the other half contained correspondent monosyllabic lexical items. The fully-contracted forms of the disyllabic contractible tokens could potentially be homophonous with their monosyllabic lexical correspondents. Each disyllabic monosyllabic pair was put into the phonologically same frame sentence, as shown in (1a) and (1b). The average lengths of the test sentences with contractible tokens and with lexical tokens were 8.4 words and 9.4 words, respectively. These materials were presented in a random order across speakers.

3.1.3. Procedures

Speakers were recorded in a sound attenuated room wearing a headphone connected to an XaudioBox for PCQuirer. The recorded signal was digitized at 32 bits with a 44.1 kHz sampling rate.

The experiment consisted of two sessions, slow and fast. At the beginning of each session, speakers heard a recorded demonstration in which a metronome prompt (88 beats/minute in the slow session; 144 beats/minute in the fast session) was played for 10 seconds and when the prompt stopped, a female voice read three sentences following the tempo just played. In the demonstration of the fast session, the female voice contracted the disyllabic items in a natural way. After each demonstration, the session began. The speakers first heard a metronome prompt that lasted for 10 seconds. When the prompt was off, they read the whole list of sentences on paper with no metronome in the background.

3.2. Measurements

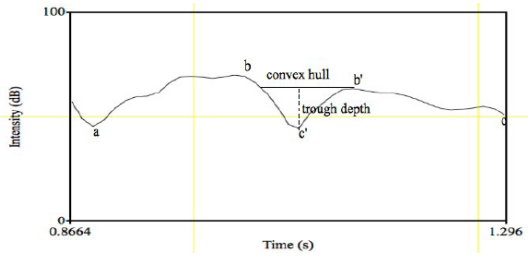
3.2.1. Trough depth

When speakers read the disyllabic contractible items, the two syllables might remain separate or be contracted to different extents, i.e. different degree of elision of medial consonants. Therefore, an algorithm named ‘trough depth’ was used to measure the degree of contraction (e.g. [3, 4]).

Each syllable has a peak of inherent sonority between two minimum of sonority. Therefore, on the premise that sonority is realized in intensity, a disyllabic word with no or near contraction should display two clear peaks in intensity. On the other

hand, if the target item is a fully-contracted disyllabic word or a monosyllabic lexical word, there should be only one peak displayed in intensity.

Figure 1: An example of intensity contour displaying the convex hull and trough depth for a non-contracted disyllabic word *tʂə51* tʂoŋ21 "this kind".



As a measure of the degree of contraction, trough depth is mainly about the presence of the medial consonants because consonants have lower energy than vowels. If the intervocalic consonants of an item are entirely elided, then trough depth is equivalent to 0 dB, and the target item is fully contracted. On the other hand, if the intervocalic consonants are not entirely elided or not elided at all, then trough depth is larger than 0 dB, then the target item is partially contracted or non-contracted. In sum, the trough depth of the tokens which are not fully contracted necessarily varies according to the intervocalic consonants.

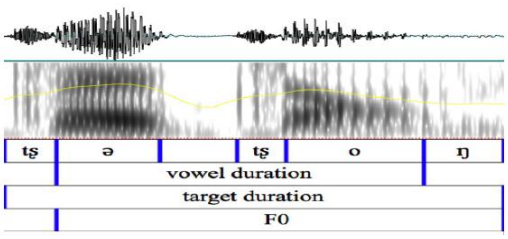
3.2.2. Durational measurements

Vowel duration (VD), target duration (TD) of the target items (i.e. durations of the disyllabic items as in (1a) and their monosyllabic lexical correspondents as in (1b)), and sentence duration (SD) were measured by first segmenting spectrograms in PRAAT.

3.2.3. F0 measurements

The F0 measurements were made on the pitch analysis from PRAAT, whose standard values of the pitch ceiling and the pitch floor are 600 Hz and 75 Hz. For each utterance, the minimum F0 (MIN F0) and the maximum F0 (MAX F0) were extracted from the interval shown as 'F0' in Figure 2.

Figure 2: A display of how durations and F0 are measured.



3.2.4. Vowel space measurements

F1 and F2 were measured by hand on the PRAAT display of a formant track made with the window length 0.025s. The midpoint of F1 and F2 at a steady vowel interval were extracted.

Vowel reduction occurs in speech produced at a fast speech rate. The vowel length decreased as the speech rate increased, and the change of the vowel length was accompanied by change of the formant frequencies. Because there was less time for the vowel to be produced, there would be a failure to reach target formant frequencies, resulting in 'formant undershoot'. The undershoot of F1 and F2 suggests that the vowels should appear centralized and so the vowel area composed of the three cardinal vowels /i, a, u/ is predicted to be relatively smaller in fast speech. The area of the vowel triangle was calculated with the formula in (2):

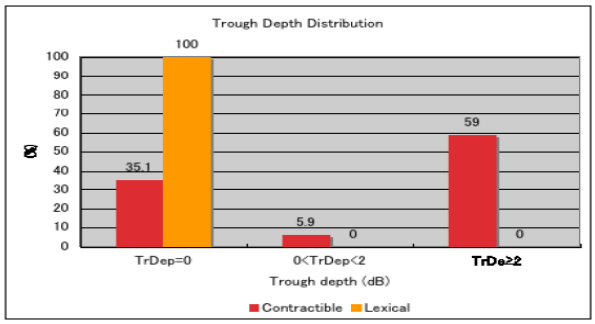
(2) Given that the vowel triangle is surrounded with three points: (x_1, y_1), (x_2, y_2) and (x_3, y_3), the vowel area = $\frac{1}{2} |x_1y_2+x_2y_3+x_3y_1-y_1x_2-y_2x_3-y_3x_1|$

3.3. Results

3.3.1. Trough Depth (TrDep)

Figure 3 is the trough depth distribution on the scale of full contraction (TrDep= 0 dB), near contraction (0 dB < TrDep < 2 dB) and disyllables (TrDep ≥ 2 dB).

Figure 3: Overall frequency distribution of trough depth in all tokens.



Syllable contraction is optional since only less than half of the tokens had full or near contraction (41%). In addition, among these contracted items, 86% were fully contracted. The great proportion of fully-contracted forms suggests that once speakers use contraction, they tend to produce the fully-contracted forms. Therefore the effect of gradience is not salient from this perspective and the distribution seems to be almost categorical.

Trough depths for all the items were entered into a linear mixed effects model (LME) in R in

which Sex was the fixed between-subject factor, and Speed (Fast vs. Slow) and Type (Contractible vs. Lexical) were the fixed within-subject factors. The results showed that males (4.62 dB) had larger trough depth than females (3.91 dB); tokens in slow speech (6.54 dB) had larger trough depth than in fast speech (1.93 dB). Contractible tokens (8.55 dB) had larger averaged trough depth than the lexical tokens (0 dB).

3.3.2. Durations

The ratios of the target duration (TD) over the sentence duration (SD) were entered into a LME model. Comparing fully-contracted tokens and lexical tokens, female had larger TD/SD ratios than males, fast speech involved higher TD/SD ratios than slow speech and the fully-contracted tokens had higher TD/SD ratios than lexical tokens.

The ratios of the vowel duration (VD) over the target duration (TD) of the fully-contracted and lexical tokens were examined in the LMR model. The results showed that only the significant main effect of Type was found – the fully-contracted tokens had larger VD/TD ratio than the lexical tokens.

The duration results presented show that both TD/SD ratio and VD/TD ratio differentiate fully-contracted tokens from lexical tokens. That is, neutralization is not complete.

3.3.3. F0

The values of MIN F0, MAX F0 and F0 Range were entered into LME model. The results showed that significant effects were found in Sex (male < female) and Speed (slow < fast) but not Type for MIN F0. As for MAX F0, significant effects were also found in Sex (female > male) and Speed (fast > slow), but not Type. However, the F0 Range data showed that females (30 Hz) had a wider pitch range than males (20 Hz), fully contracted tokens (40 Hz) had a wider pitch range than lexical tokens (30 Hz) and slow speech (37 Hz) had a wider pitch range than fast speech (30 Hz).

The F0 results showed that F0 Range is an acoustic cue that differentiates fully-contracted tokens from lexical tokens.

3.3.4. Vowel space

The formant values of the three cardinal vowels /i, a, u/ were measured from 12 items – 6 fully-contracted disyllabic items and their lexical counterparts. The results showed female vowel area of the fully-contracted tokens is not different from that of the lexical tokens, but males tended to use a bigger space for the fully-contracted tokens. This may be because the fully-contracted tokens still preserve original vowel quality from both syllables. In addition, tongue moved more upward for /i/ and /u/ and more front for /i/ for fully-contracted tokens.

3.3.5. Correlations

In order to observe the gradience effect, the correlation between trough depth and all these acoustic measures were examined. Results showed that VD/TD ratios, TD/SD ratios and MIN F0 were weakly correlated with trough depth whereas MAX F0 and F0 Range were not correlated with trough depth.

4. CONCLUSION

This study reveals that Taiwan Mandarin syllable contraction is an optional lenition process and the distribution of trough depth is the evidence. Trough depth is also a measure for gradience. In addition, both segmental (durations) and prosodic cue (F0 Range) were responsible for the distinction of fully-contracted tokens and their lexical counterparts.

5. REFERENCES

[1] Charles-Luce, J. 1985. Word-final devoicing in German: effects of phonetic and sentential contexts. *Journal of Phonetics* 13, 309-324.

[2] Fourakis, M., Iverson, G.K. 1984. On the incomplete neutralization of German final obstruents. *Phonetica* 41, 140-149.

[3] Mermelstein, P. 1975. Automatic segmentation of speech into syllabic units. *Journal of the Acoustical Society of America* 58, 880-883.

[4] Myers, J., Li, Y. 2009. Lexical frequency effects in Taiwan Southern Min syllable contraction. *Journal of Phonetics* 37, 212-230.

[5] Port, R., O'Dell, M. 1985. Neutralization of syllable-final voicing in German. *Journal of Phonetics* 13, 455-471.