UCLA
Working Papers in Phonetics

Title
WPP, No. 108: Production and Perception of Taiwan Mandarin Syllable Contraction

Permalink
https://escholarship.org/uc/item/6kj264nh

Journal
Department of Linguistics, UCLA, 108

Author
Kuo, Grace

Publication Date
2010-09-20

Peer reviewed

UCLA Working Papers in Phonetics, No. 108, pp. 1-34

Production and Perception of Taiwan Mandarin Syllable Contraction

Grace Kuo

(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. In this present study, it is shown that syllable contraction is optional with the trough depth distribution as the evidence. Trough depth is also employed as the measure for gradience. In the perception experiment, listeners were asked to do a forced-choice identification task and the accuracy was generally high. The results from the production experiment not only verifies that syllable contraction is optional and gradient, but also show that durations and F0 range are the acoustic cues that contribute to the distinction between the fully-contracted tokens and the lexical tokens.

1. Introduction

Contraction is a process that combines two or more words into one and occurs in rapid or connected speech. Often it also combines two syllables into one. Discussions of contraction typically characterize it as a lenition process with articulatory reduction or elision. In English, *don't* and *wanna* are contractions used very frequently in informal speech and each of them involves segment reduction and resyllabification. Likewise, 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. For instance, the disyllabic word *ɕy*^{55}*-jaw*^{51} 'need' in the phrase *pu*^{51} *ɕy*^{55}*-jaw*^{51} 'no need' can be contracted as *ɕiaw*^{55}.

Syllable contraction can have two properties: (a) it is optional (b) it is gradient. The occurrence of syllable contraction is optional, that is to say, the meanings of *wanna* and *want to* are not distinctive so that they can substitute for each other interchangeably. In addition, the sound variations of syllable contraction can be gradient. With an increase in speech rate, a disyllabic phrase is contracted into a single word with gradual sound changes. Take *want to* for example. The sound variations start from [wʌnt tu] and change into [wʌ̃ntə], [wʌ̃nə], [wʌ̃ɾ̃ə] gradually. Taiwan Mandarin syllable contraction also contains these two properties. In order to obtain a quantitative report of the optionality and gradience of syllable contraction, a gradient measure of syllable count is employed and will be introduced later.

Other than the occurrence of syllable contraction, the other main issue in this study is neutralization. For instance, the disyllabic phrase in *na*^{51} *jaŋ*^{51} 'such' could be fully contracted into a monosyllabic utterance *niaŋ*^{51}, which happens to also be an eligible and existing Mandarin word. Accordingly, the interpretation of a monosyllable *niaŋ*^{51} can be either a lexical word with the meaning 'to brew' or a fully-contracted form derived from the disyllabic phrase *na*^{51} *jaŋ*^{51} 'such' by means of segment deletion and resyllabification. Because Taiwan Mandarin syllable contraction involves segment deletion and because the fully-contracted outcome shares the same syllable structure with an actual lexical word, this process has been thought to be a process of neutralization.

The goal in this paper is to better understand syllable contraction in Taiwan Mandarin from both the aspects of perception and production. We will first present the results of the perception experiment which show how well the speakers could differentiate the fully-contractied disyllabic items from their lexical counterparts, and then we will present the results of the production experiment which show how different the two types of items are in essence.

1.1 Previous studies on Taiwan Mandarin syllable contraction

The phenomenon of syllable contraction in Chinese languages has attracted much attention among Chinese linguists due to its productivity and systematicity. Some studies have been descriptive (Taiwan Mandarin: Tseng, 1999; Tseng, 2005, 2006; Chung, 2006) whereas some were analytic (Cheng, 1985 on Taiwan Southern Min; Chung, 1997 and Hsiao, 2002 on both Taiwan Mandarin and Taiwan Southern Min; Hsu, 2003 on Taiwan Southern Min).

The descriptive studies found that the syllable contraction is affected by some lexical, phonological, contextual, and morphological factors. Lexically, high-frequency words are more likely to be contracted, and low-frequency words such as interrogative sentence-final particle *ma*^{55}, are less prominent in phonetic realization and are less likely to be contracted. Likewise, information not in focus is more likely to be contracted, including function words and adverbial modifications (e.g. *tʰu ʐan* → *tʰɹan* 'suddenly'). In addition, words in utterance-final position are more vulnerable to contraction. Phonologically, open syllables and falling tone are favored when syllables are contracted (Tseng, 2005). Chung (2006) indicated that although Mandarin is a syllable-timed language, the contraction is a mixed system varying with speech style. For instance, news reading is closer to syllable-timing, whereas casual conversation is closer to stress-timing. Accordingly, contracted utterances spoken in fast speech would be stress-timed. Finally, words with a closer internal morphosyntactic relationship, and a disyllabic phrase within a longer phrase might have relatively more occurrences of contraction and a greater extent of contraction (Wong, 2006).

The analytic studies have been of two types: phonological and phonetic. In the phonological analyses, the optionality of the occurrence of syllable contraction was hardly mentioned. The focus was mainly on the comparison between the citation input and the fully-contracted output, attempting to characterize regularities in this mapping (Cheng, 1985 on Taiwan Southern Min; Chung, 1997 and Hsiao, 2002 on both Taiwan Mandarin and Taiwan Southern Min; Hsu, 2003 on Taiwan Southern Min). The phonological studies of syllable contraction in Mandarin indicated that cross-dialectally, the fully-contracted form of a disyllabic word combines segmental and prosodic characteristics from the two syllables: segmentally, the edge elements (first syllable onset and the second syllable coda) tend to be preserved whereas the intervocalic elements (the first syllable coda and the second syllable onset) tend to be deleted. In addition, when the second word has a glide as its onset, then the glide tends to be preserved and become an onglide. Moreover, the vowel in the first syllable is vulnerable to elision, whereas the vowel in the second syllable tends to surface in the fully-contracted form. Prosodically, according to Hsiao (2002), the tonal contours tend to preserve all the tonal elements of the source syllables (i.e. starting from the first syllable and ending at the second syllable in the pitch contours). Take the disyllabic phrase *ʈʂʅ*^{55} *taw*^{51} 'to know' for example. The fully-contracted forms of this phrase is *ʈʂaw*^{51}. The form demonstrates the edge segment preservation, vowel preference, and ambiguously in this example, the edge tonal preservation.

In the phonetic analyses, the dependent variables were not considered categorical, but gradient instead (Wong, 2006 on Cantonese; Myers & Li, 2009 on Taiwan Southern Min; Cheng & Xu, 2009 on Taiwan Mandarin). The phonetic study of Wong (2006), for example, coded tokens as showing two levels of syllable contraction ('syllable fusion', in her term). One contained tokens with two remaining vowels (disyllabic fusion) and the other included tokens with only one vowel (vowel coalescence). She concluded that Cantonese syllable contraction is a prosodically-driven process, and 'foot' is the intermediate level of constituent in the prosodic hierarchy that captures Cantonese syllable contraction. In addition, the occurrence and the degree of contraction were highly dependent on speech rate which was defined as the average number of syllables per second. The other determining factors were word frequency, word length, morphosyntactic relationship, and prosodic position. Our study complements Wong's by using a truly gradient measure of contraction, following Myers & Li (2009), in our case in Taiwan Mandarin.

Myers & Li (2009) examined Taiwan Southern Min syllable contraction using a measure called 'trough depth' which varies from some positive number (for disyllables or partially-contracted forms) to zero (for fully-contracted forms) as a gradient criterion of syllable contraction. They found that both segment reduction and tonal merger are positively correlated with lexical frequency; furthermore, the relationship between frequency and the degree of contraction is gradient, rather than categorical. Our study applies this gradient method to Taiwan Mandarin.

The only previous study of Taiwan Mandarin syllable contraction is the very recent paper by Cheng & Xu (2009). They coded syllable contraction into three groups, based on a combination of acoustic criteria concerning the medial consonants: non-contracted, semi-contracted and contracted. They found that the degree of contraction varied with speech rate but not with the order or length of the carrier phrases. They then measured the duration of the V+(C)+(C)+V interval and the F1+F2 excursion size. Both of these measures varied gradiently across tokens, and appear to be positively correlated, suggesting that greater formant undershoot occurs with shorter duration. Thus, Taiwan Mandarin syllable contraction appears to be a gradient undershoot of the articulatory target because of time pressure, rather than a categorical shift. In this paper, a wide variety of acoustic measures are analyzed, though in contrast to Cheng & Xu, the 'trough depth' measure is used to characterize gradient contraction.

Syllable contraction in Taiwan Mandarin is a typical contraction because its occurrence is optional and its variation is gradient. In Tseng's (2005) conversational dialogue corpus study, she reported that only 32% of the eligible syllables in her data of normal speech were contracted, showing that syllable contraction is indeed an optional process. She indicated that the optionality may be due to some external factors such as individual idiosyncrasy, speech rate or speech style. Furthermore, among the contracted data, the most frequent occurrences of syllable contraction were disyllabic contractions, which make up approximately 74% of her contraction data. Some trisyllabic phrases can be contracted but normally only to a disyllabic phrase; very rarely do they form a monosyllabic word. If they do form a monosyllabic word, that was predicted to result from the further contraction of the contracted disyllabic phrase. The preference for disyllabic contractions implies that the occurrence of syllable contraction may be prosodically-driven. Our study complements Tseng's corpus study with a production experiment in which speech rate is varied, and its effect on contraction is observed.

The property of gradience has not drawn attention until recently (Wong, 2006; Myers & Li, 2009; Cheng & Xu, 2009). A disyllabic phrase could have various surface forms by means of different degrees of contraction. The variation being gradient, the forms of Mandarin syllable contraction would not pattern like all-or-nothing. Instead, the forms of the output vary with intermediate degrees of contraction – the forms might not be strictly subject to legal syllable sequence constraints (i.e. $C_1V(C_2)$; C_2 can only be an alveolar nasal or a velar nasal). Our study measures the degree of contraction in Taiwan Mandarin by the quantitative acoustic measure 'trough depth', which was employed in Myers & Li's Taiwan Southern Min study.

1.2. Trough Depth (TrDep)

When speakers read the disyllabic contractible items, the two syllables in each item might remain separate or they might be contracted to different extents, i.e. different degree of elision of medial consonants. Therefore, it is necessary to have a criterion to measure the degree of contraction.

The degree of contraction was measured by the algorithm named ‘trough depth’ (Mermelstein, 1975; Myers & Li, 2009). Each syllable has a peak of inherent sonority between two minimums 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.

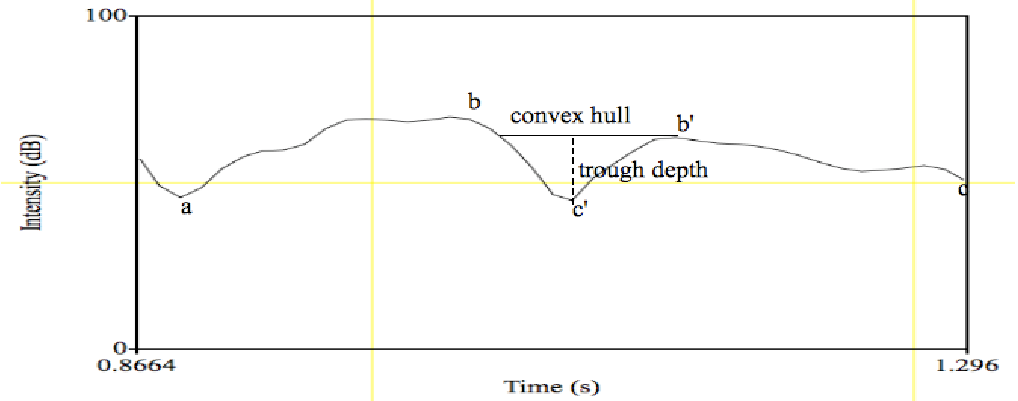


Figure 1 An example of intensity contour displaying the convex hull and trough depth for a noncontracted disyllabic word *tʂə*^{51} *tʂoŋ*^{21} 'this kind'.

In Figure 1, the utterance over the interval (a-b-c) possesses an intensity function with (b) as the maximum point. The convex hull is determined by drawing a horizontal line from the next peak (b') to the hill of the maximum peak. Over the utterance (a-b-c), the maximum hull-loudness difference is the vertical dotted line at (c'), and this is called a 'trough depth'.

In this study, the trough depths were measured from the intensity display in PRAAT with its default settings, as it is shown in Figure 1. We first found the intensity value at the second highest peak (b'), which is either in the first vowel or in the second vowel, and the valley (c'), which is the lowest point between two peaks, and then subtracted (c') from (b'). The result then is the value of trough depth. Trough depth was measured by hand with reference to the spectrogram because sometimes the peaks appear on noise, fricatives or release bursts of stops at the edges. The peaks were always measured from the syllable nucleus. If there was no (c') seen in a token, then there was no (b'). If there was no (b') seen in a token, trough depth was by definition zero.

According to Mermelstein (1975), if the loudness difference (dB) of the trough exceeds the threshold 2 dB, then the utterance (a-b-c) should be split up into segment (a-b-c') and segment (c'-b'-c). That is, there are two separate syllables in this utterance. In this study, this assumption is elaborated: if the value of trough depth is larger than 2 dB, it is counted as noncontracted; if the value of trough depth equals 0 dB, there is a full contraction; if the value of trough depth is between 0 dB and 2 dB, there is a near contraction. Figure 2 is the intensity contour of the same utterance *ʈʂə*^{51} *ʈʂoŋ*^{21} with near contraction. In Figure 2, the intensity value at *(b')* is 61.1 dB and the value at *(c')* is 60.9 dB, which means the trough depth at *(c')* is 0.14 dB, a measurement that is far below Mermelstein's threshold 2 dB. Therefore, this speech utterance is considered an example of near contraction, and it is named as 'a partially-contracted item' in this paper.

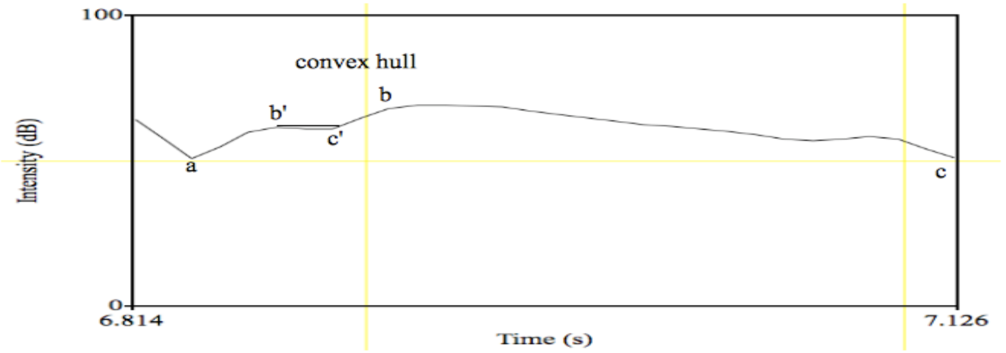


Figure 2 An example of intensity contour displaying the convex hull and trough depth for the partially contracted *ʈʂə*^{51} *ʈʂoŋ*^{21} 'this kind'

Nonetheless, it is noticeable that some items with a trough depth more than 2 dB were not necessarily noncontracted (i.e. two separate syllables). They could be partially contracted instead. For instance, the utterance in Figure 3 is an example that would be considered a noncontracted item since it has a 4.37 dB trough depth. But this utterance is actually a partially-contracted item in that the onset of the second syllable is elided. In this study, we use the term 'noncontracted' to refer to the tokens whose trough depths were over 2 dB, which means that there are clearly 2 separate syllables though they might be partially contracted.

As a measure of the degree of contraction, tough 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 partically contracted or nonontracted. In sum, the trough depth of the tokens which are not fully contracted necessarily varies according to the intervocalic consonants.

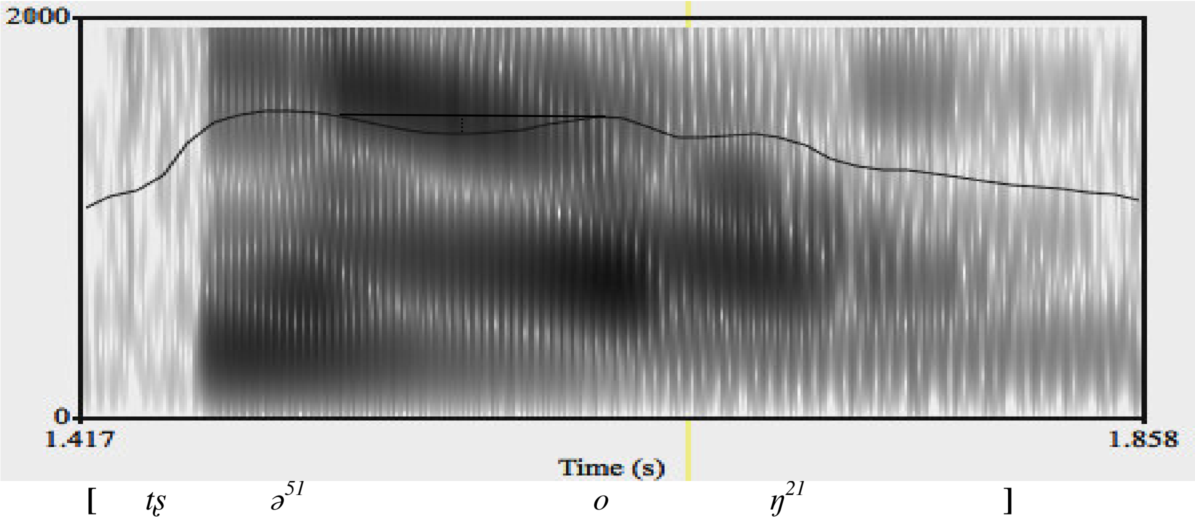


Figure 3 Spectrogram and trough depth of the utterance *tʂə*^{51} *tʂoŋ*^{21} 'this kind'. The solid contour line is the intensity curve. This utterance has the trough depth equivalent to 4.37 dB and is considered non-contracted. Yet the onset of the second syllable /tʂ/ has been elided so that it has undergone contraction to some degree.

2. Perception experiment

Since the a fully-contracted item shares the same syllable structure with it lexical counterpart, it would be interesting then to see how this kind of difference is perceived by listeners. Are they able to distinguish the two items? In order to answer this question, a forced-choice task was designed.

2.1 Stimuli

The stimuli were 80 pre-recorded sentences and all began with the same carrier phrase *t*^{h}*a*^{55} *ʂuo*^{55} 'he says'. The target tokens always appears right afer the carrier phrase. Among the 80 sentences, 18 contained fully-contracted tokens (TrDep = 0 dB), 12 contained non-contracted tokens (TrDep ≥ 2 dB), 10 contained partially-contracted tokens (0 dB < TrDep < 2dB) and 40 contained lexical tokens. Half of the stimuli were female recordings and the half of were male recordings. In addition, half of the recordings were extracted from fast speech and the other half from slow speech. (1a) shows a sentence with a contractible target item and (1b) shows a sentence with a lexical target item.

(1) a. *t*^{h}*a*^{55} *ʂuo*^{55} ***na***$^{\mathbf{51}}$***–jaŋ***$^{\mathbf{51}}$ *tɕiou*^{21} *pu*^{51} *ʐoŋ*^{35}*–ji*^{51}
he say **such** long not easy
'He says it is not easy to last for such a long time.'

b. *t*^{h}*a*^{55} *ʂou*^{55} ***niaŋ***$^{\mathbf{51}}$ *tɕiou*^{21} *pu*^{51} *ʐoŋ*^{35}*–ji*^{51}
he say **brew** wine not easy
'He says it is not easy to brew wine.'

2.2 Participants

19 female and 16 male UCLA students between the ages of 25 and 31 participated in this study. They

were all fluent native speakers of Taiwan Mandarin.

2.3 Procedures

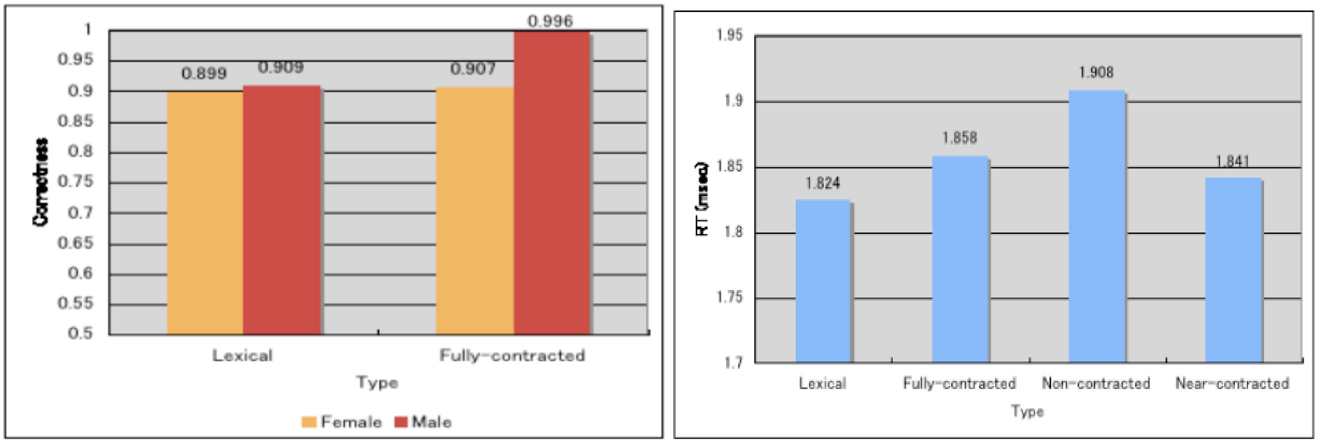
It was a forced-choice task with 80 trials. In each trial, the participants were presented with two bars labeled two similar sentences as (1a) and (1b) on the computer monitor. They were instructed to identify which utterance was being said by clicking one of the two bars after exposed to an audio file of an utterance. Prior to the experiment, three practice trails were conducted to familiarize the participants with the task.

2.4 Results & Discussions

The results of both correctness and reaction time (RT) were entered into a linear mixed-effects model (LME) in R in which Voice Sex (Male vs. Female), Speed (Fast vs. Slow), and Type (Contractible vs. Lexical) were fixed variables and Item and Speaker were random factors. The results are shown in Figure 4.

In correctness, both female and male listeners could correctly identify the fully-contracted tokens. 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 listener shows no difference ($p = 0.944$). Thus, listeners can identify both lexical and fully contracted tokens well, meaning that syllable contraction in these tokens cannot be a complete neutralization.

The measure of reaction time (RT) started from the beginning of the target item until the decision was made. No significant difference were found for the measure of RT.



(a) Correctness (b) Reaction Time

Figure 4 Perception Results (a) Correctness (b) Reaction Time

Interestingly, Taiwan Mandarin syllable contraction is thus apparently like word-final devoicing from the perception point of view. Word-final devoicing is widely attested among languages of the world and has often been assumed to be completely neutralized. Take German word-final devoicing for example (e.g. Charles-Luce, 1985; Fourakis & Iverson, 1984; Port & O'Dell, 1985). In German, the underlying /d/ surfaces as [t] in word-final position and the underlying /t/ surfaces as [t] everywhere. Therefore, the minimal pair *Rad* 'wheel' and *Rat* 'advice' have the same surface structure [rat] and may be indistinguishable. However, this pair is often recognized as intended in connected speech despite the impoverished acoustic information. Port & O'Dell (1985) found that listeners were able to differentiate devoiced final obstruents from the voiceless final obstruents 60% of the time, which is better than chance. They also found that although there was considerable overlap between the final voiced-voiceless obstruent pairs, the acoustic parameters they examined, particularly duration, showed significant, though weak differences. Therefore, German word-final devoicing is not a process of complete neutralization since the voiced-voiceless obstruents can be different from each other acoustically and perceptually.

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. Does duration show significant difference as well? Therefore, a production experiment was designed to investigate acoustic measures, including durational, spectral and prosodic cues.

3. Production experiment

From the production experiment, we would like to seek evidence for Mandarin syllable contraction being optional and gradient and as well to know which acoustic cues are responsible for distinguishing the fully-contracted items from their lexical counterparts. Therefore, we measured 'trough depth' to determine the incidence and degree of contraction in speakers' productions. Some segmental and prosodic acoustic properties were also measured for different token types.

3.1 Materials

Materials consisted of 84 sentences. 40 test sentences and 44 filler sentences. The test sentences were divided into two groups: the first group contained sentences with 20 disyllabic contractible items and the other group contained sentences with 20 correspondent monosyllabic lexical items. The contractible items were phrases frequently heard in daily conversations. Gahl (2008) found that homophone pairs differed in duration as a function of word frequency. In this study, the frequency of the target items (both lexical and disyllabic contractible) was not controlled but the frequency of the whole paradigm is nearly evenly mixed.

The fully-contracted forms of the disyllabic contractible items 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), repeated below.

(1) a. *t*^{h}*a*^{55} *ʂuo*^{55} ***na***$^{\mathbf{51}}$***–jaŋ***$^{\mathbf{51}}$ *tɕiou*^{21} *pu*^{51} *ʐoŋ*^{35}*–ji*^{51}
he say **such** long not easy
'He says it is not easy to last for such a long time.'

b. *t*^{h}*a*^{55} *ʂou*^{55} ***niaŋ***$^{\mathbf{51}}$ *tɕiou*^{21} *pu*^{51} *ʐoŋ*^{35}*–ji*^{51}
he say **brew** wine not easy
'He says it is not easy to brew wine.'

In (1a), the fully-contracted form of the disyllabic item *na*^{51}*jaŋ*^{51} (meaning 'such') is *niaŋ*^{51}, whose pronunciation has a lexical correspondent, meaning 'to brew', as shown in (1b). The disyllabic item in (1a) and the monosyllabic item in (1b) are called 'target items'. For each pair, the syllable that followed the target item was manipulated to be matched phonologically, yet they could be different characters with different meanings. For instance, *tɕiou*^{21} in (1a) was a character with the meaning 'long', whereas the *tɕiou*^{21} in (1b) was a character with the meaning 'wine'. Therefore, the phrase with the fully-contracted disyllabic item and the phrase with the monosyllabic lexical item could sound alike and were potentially ambiguous between two meanings: 'such a long time' and 'to brew wine'.

The sentences in each pair ended with the same words. For instance, the two sentences in (1) both ended with *pu*^{51} *ʐoŋ*^{35}*–ji*^{51} 'not easy'. All the sentences were initiated with the same carrier phrase *t*^{h}*a*^{55} *ʂuo*^{55} 'he says'. The target item directly followed the carrier phrase in each sentence. The average lengths of the test sentences with contractible tokens and with lexical tokens were 8.4 words and 9.4 words, respectively. All the sentences are listed in Appendix A. These materials were presented in a random order across speakers.

3.2 Participants

Five female and five male UCLA students between the ages of 25 and 27 participated in this study. All of them were fluent native speakers of Taiwan Mandarin

3.3 Procedures

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

The experiment consisted of two sessions. The aim of the first session was to elicit slow and careful speech whereas the aim of the second session was to elicit fast and casual speech. At the beginning of each session, speakers heard a recorded demonstration in which a metronome prompt (88 beats/minute in the first session and 144 beats/minute in the second session) was played for 10 seconds and when the prompt stopped, a female voice read three sentences out loud following the tempo just played. In the demonstration of the second session (with faster prompt), the female voice contracted the disyllabic items in a natural way. The reason to use metronome beats was to make the speakers produce sentences at the same range of speech rate and have a similar distribution of syllable duration so that it is straightforward to compare across subjects.

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. This sequence of listening and reading was repeated three times at each session. There were thus a total of 1200 contractible tokens and 1200 lexical tokens.

The slow session always went before the fast session because the participants became familiarized with the list of sentences after each repetition, and it was natural to have them read the list slowly at the first session and faster at the following session. A break was given between two sessions if needed.

3.4 Measurements

Syllable contraction is considered a lenition process with vowel weakening and consonant deletion. For instance, a $C_1V_1C_2.C_3V_2C_4$ disyllabic phrase could potentially be fully contracted as $C_1V_2C_4$, given that the edge consonants (C_1, C_4) tend to be preserved, and that the intervocalic consonants (C_2, C_3) tend to disappear. Therefore, the potentially gradient variations between the vowels were the main concern. In order to measure the gradient variations, a quantitative measure of degree of contraction 'trough depth', (Mermelstein, 1975) is used. In addition, the production of the vowels is affected by the change in speech rate in two ways: temporal and spectral. The former concerns the changes in segment duration and the latter concerns the changes in formant values. Tones are also affected by speech rate and in contraction. Therefore, the durational measures, vowel duration, target duration and sentence duration along with vowel formant measures F1 and F2, and the fundamental frequency were investigated. PRAAT was used for display, editing, playback and measurement procedures.

3.4.1 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.

In studies of fast speech rate, the duration of a single vowel has been the most common dependent variable (Lindblom, 1963, for Swedish monophthongal vowels; Gay, 1978, for American English vowels; Tuller et al.,1982, for American and New Zealand English vowels; Fourakis, 1986, for Greek vowels). However, in this study, vowel duration (VD) was measured as in Cheng & Xu (2009). That is, for the monosyllabic lexical items, VD was measured as the interval between the onset and the offset of the overall vowel periodicity. For the disyllabic items, VD was measured from the onset of the vowel periodicity in the first syllable to the offset of the vowel periodicity in the second syllable (see Figure 5). The measure of VD of some of the disyllabic items might include the consonant parts (i.e. the nasal coda of the first syllable or the onset of the second syllable) depending on the degree of contraction. An incomplete contraction contained vowels and partially deleted consonants. A complete contraction contained only vowel and vowel transitions (e.g. Figure 3). There are two reasons why the measurement of VD includes the optional consonants: First, the occurrence of the first coda and the second onset was optional, yet they were definitely sandwiched by the first vowel and the second vowel. Second, the onset of the first syllable and the coda of the second syllable are intact in both bisyllabic contractible items and their monosyllabic lexical counterparts, which means that the remaining parts of the syllables were the variable element that played a crucial role in the lenition process.

Although fast speech reduction primarily affects vowels, consonants clearly also can reduce. In this study, the duration of the entire monosyllabic or disyllabic target items was measured as target duration (TD). Target duration was measured from the onset of the oral consonantal constriction to the offset of the constriction or the offset of the voicing of the target items, as shown in Figure 5.

The sentence duration (SD) was also measured. It was measured from the onset of the consonantal construction for the word *t^h^a* 'he' to the offset of the constriction or the voicing of the end of the last word.

According to Port (1976), change in speech rate alters not only the absolute durations of vowels, but also their relative durations. In this study, since the results were obtained from two different tempos (slow and fast), the relative durations were considered. With the increase of speech rate, the durations in both the monosyllabic items and the disyllabic items will necessarily be shortened; there is little point comparing the absolute durations of these targets from different speech rates. As a result, the proportion of VD over TD (VD/TD ratio) and the proportion of TD over SD (TD/SD ratio) at different speech rates were calculated. Figure 5 displays the vowel duration and the target duration intervals measured in this study. It is hypothesized that the fully-contracted items are still distinguishable from their lexical counterparts in terms of these duration measures based on the fact that the two tokens are underlyingly different.

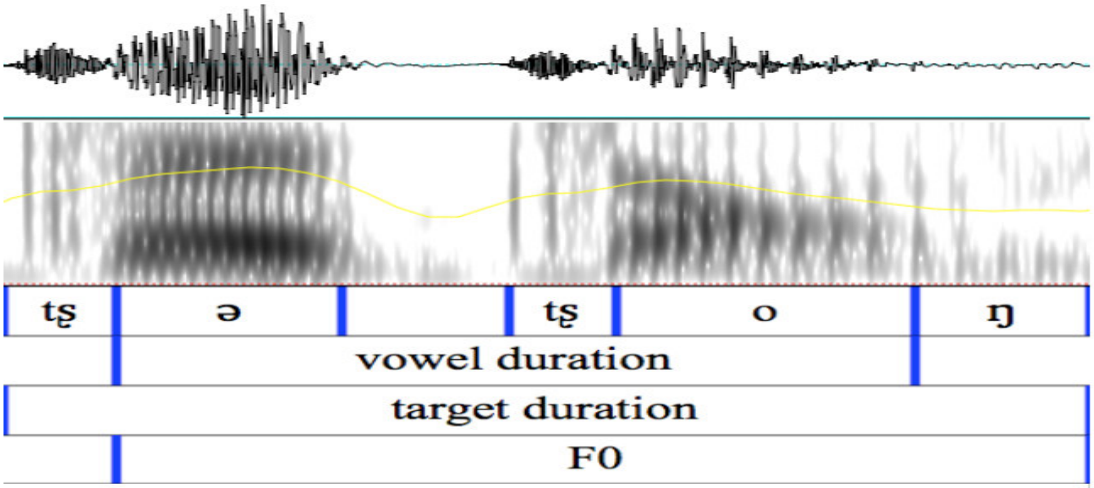


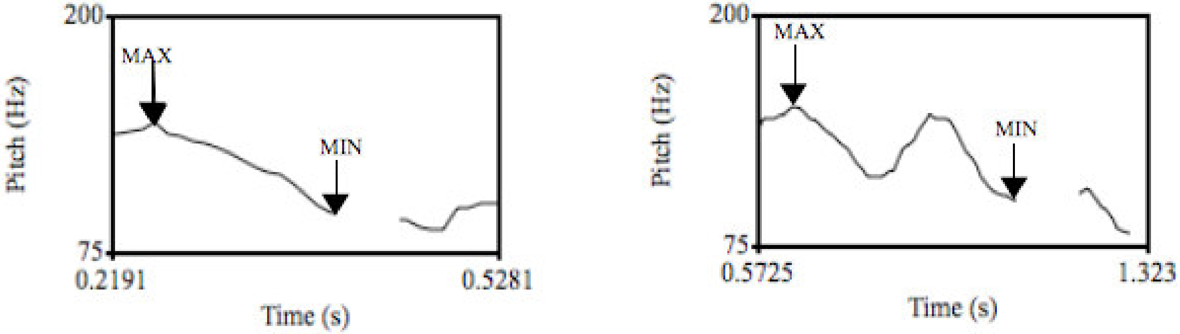
Figure 5 A display of how durations and F0 are measured. The vowel duration (VD) starts from the vowel onset of the first syllable and ends at the vowel offset of the second syllable, including the first syllable coda and the second syllable onset (here, the onset /tʂ/ in the second syllable). The target duration (TD) is the duration of the whole target phrase/word. The minimum F0 and the maximum F0 were extracted from the interval called F0, which excludes the first syllable onset.

3.4.2 F0 measurement

Previous studies mentioned that tone contraction is involved when syllable contraction occurs. Items with contracted tones tend to preserve the edge tonal elements, similar to the process of edge segment preservation in syllable contraction. However, studies from the acoustic perspective are very limited.

It has been found that French speakers use a number of strategies for producing F0 contours while achieving an increase in speech rate (Fougeron & Jun, 1998). It was shown that fast speech was largely characterized by a reduction in overall pitch range and in the amplitude of individual rising and falling pitch movements as well as a simplification of the tonal structure achieved by the non-realization of underlying tones. Therefore, we expect similar effects of speech rate here, and potentially more extreme effects of contraction depending on which tones were involved.

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 6. The F0 range (F0 RANGE) was calculated by subtracting the minmum F0 from the maximum F0. Each speaker's average minimum F0, maximum F0 and F0 range at different speech rates were then calculated across all utterances. Examples of these measurement are shown in Figure 6.



(a) [n ia ŋ] [tɕ iou] (b) [n a j a ŋ] [tɕ iou]

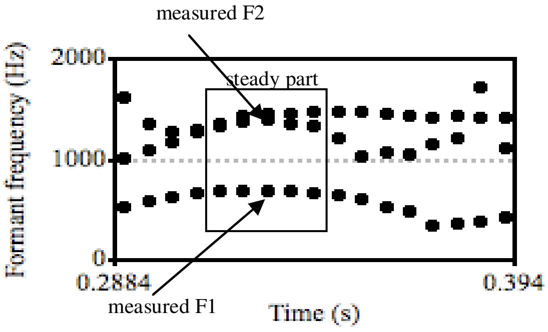
Figure 6 MIN F0 and MAX F0 measurements. (a) is a monosyllabic lexical word [niaŋ^{51}] spoken in fast speech; (b) is a disyllabic contractible phrase [na^{51} jang^{51}] spoken in slow speech by a male speaker. In both (a) and (b), [tɕiou] is a phonological matched word following the target.

3.4.4 Vowel space measurement

Twelve items containing one of the three cardinal vowels /i/, /u/, /a/ were identified from the stimuli for the purpose of making formant frequency measures. The measures of vowel space were F1 and F2, and the total area of the two-dimensional vowel space spanned by these three vowels.

a. F1 & F2: F1 and F2 were measured by hand on the PRAAT display of a formant track made with the window length 0.025s. In order to reduce the influence of vowel nasalization, for the cases that vowels were followed by a nasal coda, the number of formants was adjusted to 7 (default=5). By doing so, F1 and the nasal formants are separated.

The midpoint of F1 and F2 at a steady vowel interval were extracted as the F1, F2 measures. A steady-state is defined as the steady section of the vowel nucleus, not including the CV transition. The vowels /i/ and /u/ tended to be glides /j/ and /w/ in certain items; if that was the case, then the midpoint of the overall glide was measured. Figure 7 shows the measurement of a lexical /a/ produced by a male.



[p a n]

Figure 7 F1 and F2 measurements

b. Vowel area: Lindblom (1963) showed that vowel reduction occurred 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, the articulators had less time to attain the target, so 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)

$$\text{vowel area} = \tfrac{1}{2}\,|x_1y_2 + x_2y_3 + x_3y_1 - y_1x_2 - y_2x_3 - y_3x_1|$$

4. Results

The main interests lie in (a) whether the speakers contracted all the contractible tokens and to what extent; the latter was measured by trough depth; (b) taking all the acoustic measures into account, whether the fully-contracted tokens and their lexical counterparts are different from each other even though they share the same monosyllabic structure.

Of all the measured dependent variables, sentence duration is the measure that is mostly likely to vary with speed. It is expected that the sentence duration in the slow speech should be longer than that in the fast speech. A paired t-test was performed and there was a significant difference, $t(9) = -15.67$, $p < .05$. The mean sentence duration for fast speech (Mean = 1.15s, sd = 0.2451s) was significantly shorter than the mean sentence duration for slow speech (Mean =2.05s, sd = 0.4071s). This result indicates that the speakers did make a difference when they were asked to speak at different speech rates.

4.1 Trough Depth (TrDep)

'Trough depth' is used to measure the degree of contraction. If trough depth of a contractible phrase is large, the loudness difference between the two syllables is clearer, and the less the degree of contraction. The measure of trough depth of the lexical items is always zero in that there is only one sonority peak in each item. However, when speakers are given contractible disyllabic items, they do not always contract them, and if they do, the measure of trough depth varies. According to Mermelstein, trough depth 2 dB is the threshold for a syllable boundary. When the trough depth of a contractible disyllabic item is larger than 2 dB, then there are by this definition two syllables. If a contractible disyllabic item has trough depth valued zero like lexical items, it is considered fully-contracted. This category of the contractible tokens will be the focus in the following analysis.

Contractible vs. Lexical

To examine the influence of speaker and speech rate on the occurrence of contraction, trough depth was used to classify different types of contraction – full contraction (TrDep = 0 dB), near contraction ($0 \text{ dB} < \text{TrDep} < 2 \text{ dB}$) and disyllables ($\text{TrDep} \geq 2 \text{ dB}$). An error loss of 4.12% (99 items) was removed from the data set prior to the analyses. Errors were defined as missing the target, hesitation at recording and noise in the background. In the data, the trough depth of non-fully-contracted tokens (near contraction and disyllables) varied from 0.05 dB to 45.6 dB.

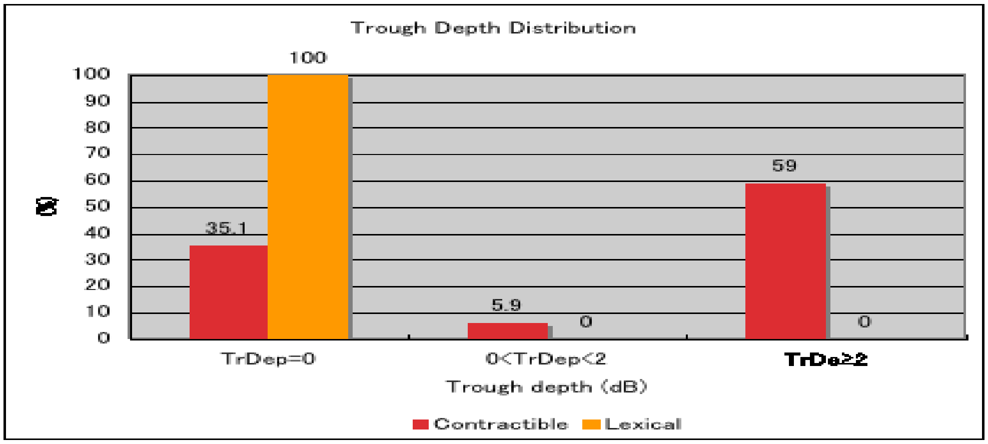


Figure 8 Overall frequency distribution of trough depth in all tokens

Figure 8 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). The result confirms that 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 seems to suggest 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. Figure 9 is a closer look into the trough depth distribution of the contractible tokens in terms of different speech rates. It shows that most of the fully-contracted tokens are from fast speech, while most of the disyllables are from slow speech. Interestingly, some tokens at fast speech are disyllables (17.92%) and some tokens at slow speech are fully contracted (6.83%). The number of partially-contracted tokens is almost the same every 0.5 dB interval, and they do not show a gradient pattern as well. The trough depth distribution in Figure 9 implies that the number of the fast tokens does not drop down gradually nor does the number of the slow tokens climb up increasingly.

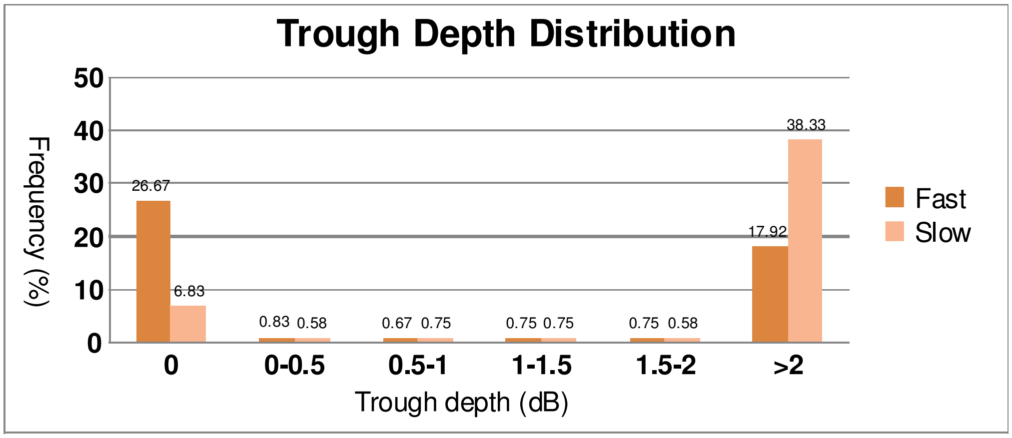


Figure 9 Trough depth distribution of the contractible items

The trough depths for all the items were entered into a linear mixed-effects model (LME) in R in which Sex (Male vs Female) was the fixed between-subject factor, and Speed (Fast vs. Slow) and Type (Contractible vs. Lexical) were the fixed within-subject factors. Furthermore, there were three within-subject random factors, namely Repetitions (3 levels), Items (20 levels) and Speaker (10 levels). The results of the LME analysis of trough depth are shown in Table 1. The current lmer function in R does not report *p* values because there are controversies over the computation of the degree of freedom (*df*). However, since the number of observations in this study is so large (n > 2000), the *df* controversies are not an issue and *t* values could be considered a test of significance. Baayen (2008) states that the absolute *t* value should be greater than 2 at the alpha level of 0.05 when the number of observations is very large.

Table 1 Linear mixed-effects modeling of trough depth (Contractible vs. Lexical)

Parameter	Estimate	Std. Error	t value
(Intercept)	3.1258	0.8808	3.549*
Sex	1.6164	0.7929	2.039*
Type	-3.1328	0.4263	-7.348*
Speed	9.5039	0.4247	22.375*
Type x Sex	-1.5730	0.6074	-2.590*
Type x Speed	-9.4672	0.6010	-15.753*
Sex x Speed	-0.3019	0.6053	-0.499
Sex x Type x Speed	0.2428	0.8539	0.284

* = significant at $p < .05$

Hence for trough depth, significant main effects were observed for Sex, Type, Speed, and the interactions of Type by Sex and Type by Speed. The results showed that males (4.62 dB) had larger trough depth than females (3.91 dB), probably because females tended to contract syllables more. This is born out by the Type by Sex interaction. Moreover, tokens in the slow speech (6.54 dB) had larger trough depth probably because they were less likely to be contracted than those in the fast speech (1.93 dB). Again, it is born out by the interaction of Type and Speed. Since the trough depth of the lexical tokens is always 0 dB, there was more variation in the trough depth of the contractible tokens than that of the lexical tokens. It is not surprising to observe that the contractible tokens (8.55 dB) had larger averaged trough depth than the lexical tokens (0 dB). The interaction of Type by Sex and the interaction of Type by Speed were explored with post-hoc *t*-tests. The Type by Sex interaction results showed that there was a significant difference between males and females in terms of the averaged trough depth of the contractible items [$t(1143) = -2.64$, $p < .05$]. In addition, significant differences were found in the comparison of the contractible tokens and the lexical tokens for both males [$t(563) = 22.05$, $p < .05$] and females [$t(580) = 20.19$, $p < .05$], with the contractible tokens having larger trough depths than the lexical tokens. The results of the Type by Speed interaction showed that the averaged trough depth of the contractible tokens at the slow speech rate is significantly higher than that at the fast speech rate, $t(1143) = -18.69$, $p < .05$. In addition, contractible tokens and lexical tokens are significantly different in both fast speech [$t(1140) = 14.79$, $p <.05$] and slow speech [$t(1146) = 30.85$, $p < .05$]. The interaction plots are illustrated in Figure 10. The stars and the circles indicate the significant pairwise differences. In sum, contraction is not complete for either males or females, and in either fast or slow speech, though both females and fast speech favor contraction.

We have seen that the sentence duration in slow speech tends to be longer and that contractible items in slow speech are less likely to be contracted. So it is possible that sentence duration and trough depth could be positively correlated. Figure 10 below is a scatterplot distribution of trough depth vs. sentence duration with the regression line. The result shows that sentence duration and trough depth are statistically correlated with $p < .05$. However, $r^2 = 0.084$ indicates that only 8.4% of the sentence duration data can be accounted for with trough depth. As a sentence is slower, the trough depth would be larger and less degree of contraction is involved.

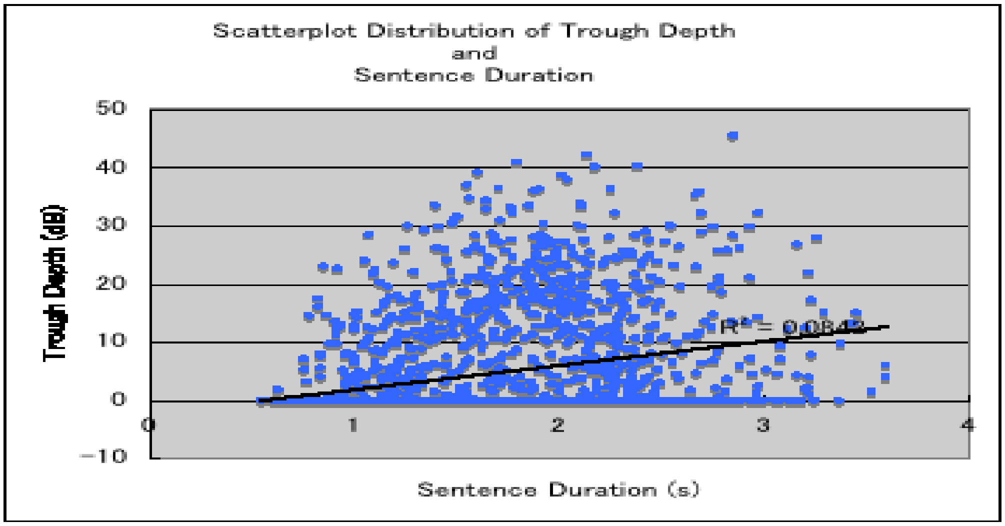


Figure 10 Correlation between trough depth and sentence duration

In the following results, in order to observe the degree of contraction of the disyllabic contractible items, trough depth is treated as a potential predictor of the other dependent variables. For the items with the trough depth value larger than zero (TrDep > 0 dB), the correlation between the trough depth and each dependent variable will be examined.

4.2 Durations

4.2.1 Target duration ratio (TD/SD ratio)

The targets are the disyllabic contractible phrases and the monosyllabic lexical words. Other things being equal, the duration of a lexical word should be shorter than a contractible phrase because of the syllable number difference. Therefore, the proportion of the target duration over the sentence duration of a lexical word is predicted to be smaller than that of a contractible phrase. However, if a contractible phrase is fully contracted, that is, it has a monosyllabic form, then the proportion of the target duration over the sentence duration should pattern like a lexical word. In the following results of TD/SD ratios, we first examined the difference between all contractible phrases and their lexical counterparts, and then we paid particular attention to the difference between fully-contracted phrases and lexical items. The prediction for the former is that there should be an obvious difference between contractible and lexical items because speakers do not always fully contract all the disyllabic targets. In contrast, the prediction for the latter is that the fully-contracted tokens and the lexical tokens should show no significant difference. We would then infer that the fully-contracted tokens are the result of a complete neutralization, with no gradient changes involved, on this duration measure.

Contractible vs. Lexical

The ratios of the target duration (TD) over the sentence duration (SD) were entered into a linear mixed-effects model in which Sex (Male vs. Female), Speed (Fast vs. Slow), and Type (Contractible vs. Lexical) are fixed variables. Repetition, Item and Speaker were the three random factors. An error loss of 4.12% (99 items) was removed from the data set prior to the analyses. The results are displayed in Table 2.

Table 2 Linear mixed-effects modeling of TD/SD ratio (Contractible vs. Lexical)

Parameter	Estimate	Std. Error	t value
(Intercept)	0.197218	0.007077	27.87*
Sex	0.011172	0.005394	2.07*
Type	-0.064391	0.002023	-31.82*
Speed	0.019525	0.002016	9.69*
Type x Sex	-0.010721	0.002883	-3.72*
Type x Speed	-0.026687	0.002852	-9.36*
Sex x Speed	-0.010187	0.002873	-3.55*
Sex x Type x Speed	0.006909	0.004052	1.70

* = significant at $p < .05$

Significant main effects of Sex, Type, Speed and all the two-way interactions were observed. The results indicate that males (mean = 16.99%, sd = 0.056) had larger TD/SD ratios than females (mean = 16.81%, sd = 0.054), slow speech (mean = 16.99%, sd = 0.057) involved higher TD/SD ratios than fast speech (mean = 16.81%, sd = 0.053) and the contractible items (mean = 20.97%, sd = 0.041) have higher TD/SD ratios than the lexical items (mean = 12.87%, sd = 0.033). The means of the TD/SD ratios do not always look very distinct from each other, but there are some small differences which are significant with these large numbers of observations.

Since we are interested in whether TD/SD ratio differentiates the tokens by Type, only the interactions that involve Type are discussed. The post-hoc *t*-test showed that both females and males had significantly higher TD/SD ratio for contractible tokens than for lexical tokens [Female: $t(1100) = 39.93$, $p < .05$; Male: $t(1088) = 34.72$, $p < .05$]. In addition, the male speakers' contractible tokens had higher TD/SD ratio than the female speakers did, $t(1138) = -2.52$, $p < .05$. The post-hoc *t*-tests for the interaction of Type by Speed showed that the TD/SD ratio for contractible and lexical tokens are significantly different from each other at both speech rates [Fast rate: $t(1076) = 29.28$, $p < .05$; Slow rate: $t(1117) = 4.86$, $p < .05$], and the TD/SD ratio at different speech rates are different from each other for both Types [Contractible: $t(1088) = -6.03$, $p < .05$; Lexical: $t(1121) = 48.03$, $p < .05$].

Fully Contracted vs. Lexical

Next, we focus on the comparison between just the fully-contracted disyllabic tokens and the monosyllabic lexical tokens. The fully-contracted items share the same syllable structure with their lexical counterparts. Based on this assumption, a prediction is that fully-contracted tokens are not significantly different from their lexical counterparts in terms of TD/SD ratios.

The data for this subset of items were entered into an LME model in which the fixed factors were Sex

(Male vs. Female), Speed (Fast vs. Slow) and Type (Fully contracted vs. Lexical), while the random factors were Repetition, Item, and Speaker. Although the data of the observations here is a subset of that in the section above, the number of observations is still large enough to allow the absolute *t* value larger than 2.0 for significance.

The results in Table 3 revealed that all the fixed factors and two-way interactions were significant, much as for the analysis of all contractible tokens. A significant main effect of Type was observed, with fully-contracted tokens (mean = 18.72%, sd = 0.039) involving higher ratios than lexical tokens (mean = 12.86%, sd = 0.032). A significant mean effect of Sex was also obtained. Overall, females (mean = 14.42%, sd = 0.042) had higher TD/SD ratio than males (mean = 14.31%, sd = 0.043). The main effect of Speed was also significant: the fast speech (mean = 15.21%, sd = 0.044) involved higher ratio than the slow speech (mean = 13.26%, sd = 0.038).

The Type effect was explored through the interaction of Type and Sex as well as the interaction of Type and Speed. The post-hoc *t*-tests for the interaction of Type by Sex showed that both males and females differentiate fully-contracted tokens from lexical tokens [Male: $t(268)=20.68$, $p < .05$; Female: $t(330) = 17.32$, $p < .05$]. In addition, females' fully-contracted tokens and males' fully-contracted tokens were distinguishable with males having higher TD/SD ratio, $t(398) = -2.5$, $p < .05$. But females' and males' lexical tokens were not distinguishable.

The post-hoc *t*-tests for the interaction of Type by Speed showed that TD/SD ratios of the fully-contracted tokens were significantly different in fast speech and slow speech, $t(398) = -2.06$, $p < .05$. Likewise, the TD/SD ratios of the lexical tokens were as well significantly different in fast and slow speech – lexical tokens in fast speech had higher TD/SD ratios than lexical tokens in slow speech, $t(1117) = 4.86$, $p < .05$. In addition, in both fast and slow speech, the fully-contracted tokens were significantly different from the lexical tokens [Fast: $t(581)=19.2$, $p < .05$; Slow: $t(94) = 16.86$, $p < .05$]. The interaction results suggests that the main effect of Type is general and thus the TD/SD ratio is an indicator differentiating fully-contracted tokens from lexical tokens for both sexes and both speeds. Thus, it is not complete neutralization.

Table 3 Linear mixed-effects modeling of TD/SD ratio (Fully-contracted vs. Lexical)

Parameter	Estimate	Std. Error	t value
(Intercept)	0.180020	0.006565	27.420*
Sex	0.010128	0.004661	2.173*
Type	-0.047154	0.002147	-21.965*
Speed	0.010672	0.003864	2.762*
Type x Sex	-0.009870	0.003148	-3.135*
Type x Speed	-0.017919	0.004282	-4.185*
Sex x Speed	-0.012794	0.005597	-2.286*
Sex x Type x Speed	0.009718	0.006175	1.574

* = significant at $p < .05$

TD/SD Ratio and Trough Depth Correlation

It is also of interest to explore the relationship between the two variables, TD/SD ratio and trough depth. Trough depth was used to measure the degree of contraction, and whether or not trough depth is a good predictor of the measure of TD/SD ratio is worth examination. The data entered excluded the fully-contracted tokens (TrDep = 0 dB).The data of the TD/SD ratio and of the trough depth were entered into the R function to calculate the correlation between the two variables. The result shows that TD/SD ratio and trough depth are correlated with $p < .05$. However, the $r^2 = 0.055$ indicates that only 5.5 percent of the variation in TD/SD ratio can be explained by trough depth.

4.2.2 Vowel duration ratio (VD/TD ratio)

The duration of the vowel in a disyllabic contractible item started from the onset of the vowel in the first syllable and ended at the offset of the vowel in the second syllable. If the item was fully contracted or it was lexical, then the duration was of a single vowel. The previous section showed that Sex, Speed and Type were predictors for the TD/SD ratios comparing the contractible tokens and the lexical tokens as well as comparing the fully-contracted tokens and the lexical tokens. In this section, the VD/TD ratio for these same tokens is examined. Since the interest is mainly the comparison between fully-contracted tokens and their lexical counterparts, the comparison results of all contractible tokens and lexical tokens are not provided. The prediction is that VD/TD ratio patterns like TD/SD ratio because the only difference between the measurement of VD and TD is whether the onset of the first syllable and the coda of the second syllable, namely the edge elements, were included or not. Whether the intervocalic segments were deleted and whether the vowels of the first and the second syllables were merged into a single vowel were the two main processes relevant here. The edge elements were preserved throughout the process of syllable contraction; therefore, if there is any change on syllable contraction, that must be in the VD. The use of TD as the denominator provides a very local control for speech rate.

Fully Contracted vs. Lexical
The ratios of the vowel duration (VD) over the target duration (TD) were examined with a linear mixed-effects model with the same fixed factors: Sex, Speed, and Type and the same random factors: Repetition, Item and Speaker. The results are displayed in Table 4. The significant main effect of Type was found, and the result shows that the fully-contracted items (mean = 66.63%, sd = 0.146) had larger VD/TD ratio than the lexical items (mean = 59.04%, sd = 0.193). The results regarding VD/TD ratios suggest that VD/TD ratio can differentiate fully-contracted tokens from their lexical counterparts.

The difference between VD and TD is whether or not the periphery consonants in the target phrase are included. If the periphery consonants are not crucial to syllable contraction, the VD/TD ratio should remain constant. However, the observed significant main effect of Type suggests that the periphery consonants are different in that when they are in the fully-contracted tokens, they are more shrunk so that the VD/TD ratio for the fully contracted tokens is larger than the VD/TD ratio for the lexical tokens.

In addition, the lme results for TD/SD ratio and VD/TD ratio are different in that there were more factors with significant effect for TD/SD ratio than for VD/TD ratio. This might be because the former involves a bigger domain (i.e. sentence) which could be affected by various factors, whereas the latter involves a more local domain (i.e. phrase and vowel within the phrase) which is not expected to have as many individual effects as a sentence does. In future work, we will also analyze the raw data of VD, TD and SD as dependent variables.

Table 4 Linear mixed-effects modeling of VD/TD ratio (Fully Contracted vs. Lexical)

Parameter	Estimate	Std. Error	t value
(Intercept)	0.684602	0.036910	18.548*
Sex	-0.003963	0.019734	-0.201
Type	-0.071506	0.010462	-6.835*
Speed	-0.016551	0.018832	-0.879
Type x Sex	-0.025189	0.015340	-1.642
Type x Speed	0.010945	0.020868	0.524
Sex x Speed	0.044928	0.027276	1.647
Sex x Type x Speed	-0.063464	0.030095	-2.109*

* = significant at $p < .05$

VD/TD Ratio and Trough Depth Correlation

In order to observe the relationship between the VD/TD ratio and the trough depth, trough depth and VD/TD ratio of the partially-contracted tokens and the disyllabic tokens (TrDep > 0 dB) were entered to calculate the correlation between the two variables. The VD/TD ratio and trough depth are correlated because $p < .05$ and the result $r^2 = 0.105$ means that around ten percent of the variation in VD/TD ratio can be explained by trough depth.

4.2.3 Interim summary

Inconsistent with our predictions, 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. Moreover, the two ratios and trough depth are correlated, though the correlations are weak.

4.3 F0

The values of F0 were entered into linear mixed-effects models in which Sex, Speed, and Type were the fixed factors while Repetition, Items and Speakers were the random factors. The dependent variables were minimum F0 (MIN F0), maximum F0 (MAX F0) and the F0 RANGE. The F0 RANGE was calculated from the subtraction of MIN F0 from MAX F0. As above, the result reported below concentrates on the F0 comparison of fully-contracted tokens and lexical tokens. If the pitch range of the fully-contracted tokens is like that of the lexical tokens, then the fully-contracted tokens are neutralized on this dimension. But if the pitch range of the fully-contracted tokens is different from that of the lexical tokens, then it seems that speakers are using pitch range to distinguish between the fully-contracted and the lexical tokens.

4.3.1 Minimum F0 (MIN F0)

Fully Contracted vs. Lexical

The minimum F0 data for the fully contracted and the lexical items were entered into a linear mixed-effects model with the fixed factors Sex, Type and Speed and with the random factors Repetition, Item and Speaker. The results are in Table 5. Significant effects were found in Sex and Speed, but not Type. Although the effect of Type was not significant, the interaction of Type by Speed was. The interaction of Sex by Speed was also significant, and so was the three-way interaction with Type.

The results show that slow speech (156 Hz) involved lower MIN F0 than fast speech (173 Hz), and that males (122 Hz) had lower minimum F0 than females (207 Hz), both as expected. However, the MIN F0 of the fully contracted tokens was not significantly different from that of the lexical tokens. The lack of a Type effect indicates that MIN F0 could not be a cue that speakers rely on to distinguish fully contracted tokens from the lexical tokens overall.

Nonetheless, although the main effect of Type was not significant, a significant interaction of Type by Speed was obtained. The post-hoc *t*-tests show that fully-contracted tokens and lexical tokens were significantly different from each other in slow speech [$t(668) = -2.55$, $p < .05$] but not in fast speech. That is, there was a greater Type difference in slow speech than in fast speech. In addition, the fully-contracted tokens in fast speech were significantly higher than those in slow speech [Fully contracted: $t(398) = 4.95$, $p < .05$]; the lexical tokens in fast speech were significantly higher than those in slow speech [Lexical: $t(1154) = 4.95$, $p < .05$].

Table 5 Linear mixed-effects modeling of Minimum F0 (Fully Contracted vs. Lexical)

Parameter	Estimate	Std. Error	t value
(Intercept)	212.640	6.668	31.89*
Sex	-86.026	8.514	-10.10*
Type	2.358	1.752	1.35
Speed	-30.322	3.153	-9.62*
Type x Sex	1.254	2.571	0.49
Type x Speed	17.301	3.494	4.95*
Sex x Speed	12.409	4.568	2.72*
Sex x Type x Speed	-14.502	5.039	-2.88*

* = significant at $p < .05$

Minimum F0 and Trough Depth Correlation

In order to observe the relationship between the MIN F0 and the trough depth, the data of two variables for partially-contracted tokens and disyllabic tokens were entered to calculate the correlation between the two variables. The result shows that MIN F0 and trough depth were correlated with $p < .05$. However, the $r^2 = 0.011$ indicates that only 1.1 percent of the MIN F0 data can be accounted for with trough depth.

4.3.2 Maximum F0 (MAX F0)

Fully Contracted vs. Lexical

The maximum F0 data were entered into a linear mixed-effects model with the same factors – Sex, Speed, Type were the fixed factors; Item, Repetition and Speaker were the random factors. The results of the analysis of the maximum F0 over the comparison between the fully-contracted items and the lexical items are shown in Table 6.

A significant main effect for Sex was obtained, with greater MAX F0 for females (247 Hz) than for males (148 Hz). The main effect of Speed was also significant with greater MAX F0 for fast speech (203 Hz) than for slow speech (194 Hz). There is no significant main effect of Type and no significant interactions. MAX F0 seems not to be a cue for differentiating fully-contracted tokens from lexical tokens; that is, the two token Types are neutralized on this dimension.

Table 6 Linear mixed-effects modeling of MAX F0 (Fully Contracted vs. Lexical)

Parameter	Estimate	Std. Error	t value
(Intercept)	250.0326	9.1677	27.273*
Sex	-98.8062	12.0768	-8.181*
Type	-1.6411	1.7489	-0.938
Speed	-7.4231	3.1474	-2.358*
Type x Sex	2.0481	2.5655	0.798
Type x Speed	0.7201	3.4879	-0.206
Sex x Speed	-3.6141	4.5589	-0.793
Sex x Type x Speed	3.0826	5.0294	0.613

* = significant at $p < .05$

Maximum F0 and Trough Depth Correlation

The MAX F0 and the trough depth data of the partially-contracted and disyllabic tokens (TrDep > 0 dB) were entered to calculate the correlation. The result showed that MAX F0 and trough depth were not correlated at all (p > .05). In addition, $r^2 = 0$.

4.3.3 F0 Range (F0 RANGE)

Fully Contracted vs. Lexical

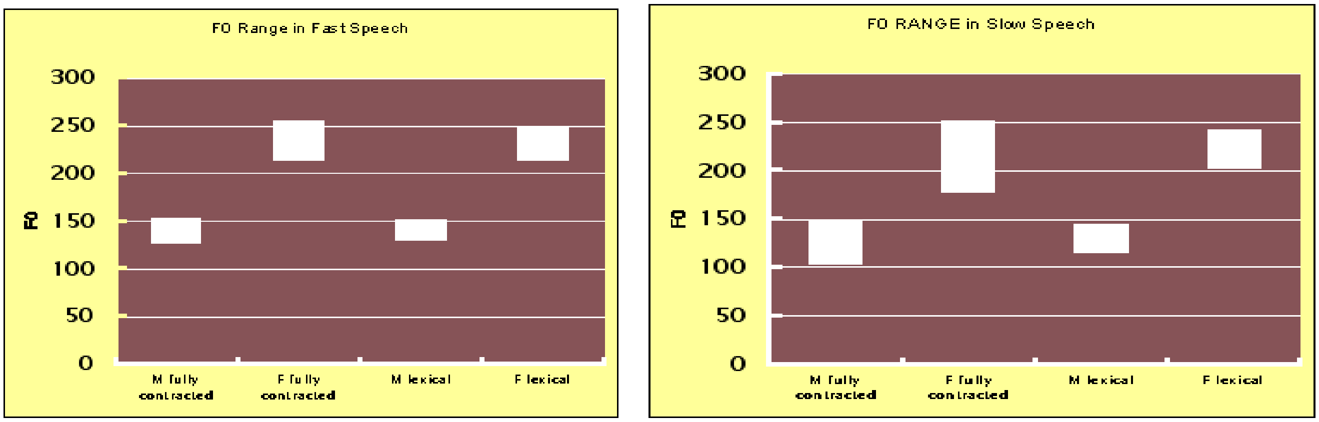
The data for RANGE was calculated from MIN F0 and MAX F0 and was analyzed with a linear mixed-effects model. The fixed factors were Sex, Type and Speed and the random factors were Repetition, Item, and Speaker. The results of the analysis of the pitch range are shown in Table 7. It shows that for the pitch range, all the factors mattered except for the interaction of Type by Sex. 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 fast speech rate (30 Hz) had a narrower pitch range than slow speech rate (37 Hz). The latter result echoed the previous French study (Fougeron & Jun, 1998), showing that fast speech is characterized by a reduction in the overall pitch range.

The post-hoc *t*-tests for the interaction of Type and Speed show that both fast speech and slow speech differentiate fully-contracted tokens from lexical tokens [Fast: $t(518) = 4.15$, $p < .05$; Slow: $t(1154) = -5.59$, $p < .05$]. In addition, the slow speech had significantly wider pitch range than fast speech for both fully-contracted tokens and lexical tokens [Fully contracted: $t(100) = -5.08$, $p < .05$; Lexical: $t(644) = -11.0$, $p < .05$]. This result suggests that fully-contracted tokens and lexical tokens are consistently distinct in the dimension of F0 RANGE, therefore, the fully-contracted tokens are not neutralized.

Table 7 Linear mixed-effects modeling of RANGE (Fully Contracted vs. Lexical); * = significant at $p < .05$

Parameter	Estimate	Std. Error	t value
(Intercept)	37.4684	5.2052	7.198*
Sex	-12.8243	5.6397	-2.274*
Type	-4.1005	1.6856	-2.433*
Speed	22.7114	3.0337	7.486*
Type x Sex	0.8628	2.4725	0.349
Type x Speed	-16.3634	3.3618	-4.867*
Sex x Speed	-16.0232	4.3941	-3.647*
Sex x Type x Speed	17.5523	4.8477	3.621*

Figure 11 indicates that overall, slow speech involved a wider range than fast speech, female speakers had a wider range than males, and most importantly, fully-contracted tokens had a wider range than their lexical counterparts. Despite the significant three-way interaction effect, the post-hoc *t*-tests showed that Type difference can be observed across speech rates and genders. Note that for MIN F0, there was no main effect of Type, but here, when the total F0 RANGE is considered, there is. Thus, total F0 RANGE is a more reliable correlate than MIN F0.



(a) F0 RANGE in Fast Speech (b) F0 RANGE in Slow Speech

Figure 11 F0 RANGE (MAX-MIN) for the Fully Contracted and the Lexical tokens by sex, at two speech rates

F0 Range and Trough Depth Correlation

In order to observe the relationship between the F0 RANGE and the trough depth, the data of the partially-contracted and disyllabic tokens (TrDep > 0 dB) were entered to calculate the correlation between the two variables. The result showed that F0 RANGE and trough depth were not correlated ($p > .05$).

The results from the mixed-effects models for MIN F0, MAX F0 and F0 RANGE show that fully-contracted tokens and lexical tokens are distinctive in the dimension of F0 RANGE, and in the dimension of MIN F0 in slow speech. Otherwise, the fully-contracted tokens seem to be neutralized and not distinguishable from lexical tokens. Two possible reasons pertinent to tone preservation are proposed to account for this observation. It seems that tone is more likely to remain intact, compared to segments, in the process of syllable contraction. Therefore, there are examples in which the syllable structure of the fully-contracted token is monosyllabic, but the tone is realized as a complex contour tone, which combines the two citation tones of the contractible disyllabic words. In addition, the phenomenon of tone preservation may level down or level up the overall pitch realization of a token.

4.3.4 Interim summary

The F0 results presented show that F0 RANGE is an acoustic cue that differentiates fully-contracted tokens form lexical tokens. The time course comparisons of the different token Types in three different Tone Types (HL, LH, HH) show that fully-contracted and lexical tokens are mostly not distinguishable except for the time points close to the end of the LH Tone Type. Moreover, MIN F0 is weakly correlated with trough depth whereas MAX F0 and F0 RANGE are not correlated with trough depth.

4.4 Vowel Space

There were 26 fully-contracted tokens and 57 lexical tokens for /i/; there were 59 fully-contracted tokens and 52 lexical tokens for /u/; there were 14 fully-contracted tokens and 55 lexical tokens for /a/. The formant values (means and standard deviations) of the three cardinal vowels /i/, /a/, /u/ shown in Table 8 were measured from 12 items - 6 disyllabic items that were fully contracted and their 6 lexical counterparts. Since the fully-contracted items had exactly the same syllable structures as the lexical items did, the underlined vowels in the fifth column are the vowels measured in both item types.

Table 8 Fully-contracted and lexical items whose vowels were measured.

Vowel	Item	Fully-contracted	Gloss	Lexical	Gloss
a	#3	/pa^{55} tien^{55}/	'8 o'clock'	/p**a**n^{55}/	'class'
a	#14	/ta^{51} tɕia^{55}/	'everyone'	/t**a**^{51}/	'most'
i	#6	/tɕin^{55} t^{h}ien^{55}/	'today'	/tɕ**i**en^{55}/	'to fry'
i	#16	/ɕin^{55} tien^{51}/	'(place name)'	/ɕ**i**en^{51}/	'county'
u	#11	/hu^{55} ʑan^{24}/	'suddenly'	/h**u**an^{51}/	'to exchange'
u	#20	/ɕia^{51} wu^{21}/	'afternoon'	/ɕ^{j}a**u**^{51}/	'to laugh'

There are three possibilities for the F1 results of /i/ and /u/: (a) if there is no complete neutralization and if the measured high vowels of the aforementioned fully-contracted items are more glide-like, the F1 of /i/ and /u/ of the fully-contracted items should be significantly lower than that of the lexical items; (b) if there is no complete neutralization and if the measured high vowels of the fully-contracted items are more like the lax vowels (i.e. [ɪ] and [ʊ]), then the average F1 of /i/ and /u/ of the fully-contracted items is expected to be higher than the lexical items; (c) if the measured high vowels are completely neutralized, then the average F1 of /i/ and /u/ of the fully-contracted items and their lexical counterparts should show no difference.

Likewise, there are three possibilities for the F2 results of /i/ and /u/: (a) if there is no complete neutralization and if the measured high vowels of the fully-contracted items are more like glides, the F2 of /i/ of the fully-contracted items should be significantly higher than that of the lexical items whereas the F2 of /u/ of the fully-contracted items should be significantly lower than that of the lexical items; (b) if there is no complete neutralization and if the measured high vowels are more like the lax vowels, then the average F2 of /i/ of the fully-contracted items is expected to be lower than the lexical items whereas the average F2 of /u/ of the fully-contracted items is expected to be higher than the lexical items; (c) if the measured high vowels are completely neutralized, then the average F2 of /i/ and /u/ of the lexical items and their fully-contracted counterparts should show no difference. In Table 9, the average formant values of the /i/, /u/, /a/ in fast speech and slow speech are displayed. The values in parentheses beside the means are the standard deviations. In Table (9b), some data in slow speech is missing because there were no fully-contracted data for /i/ and /a/ in slow speech.

Table 9 (a) Formant values in fast speech

		i		a		u	
		F1	F2	F1	F2	F1	F2
Male	Fully contracted	334 (28)	2436 (216)	673 (78)	1143 (144)	436 (68)	856 (175)
	Lexical	400 (53)	2474 (209)	668 (49)	1055 (185)	533 (66)	971 (155)
Female	Fully contracted	303 (30)	2466 (60)	797 (111)	1116 (181)	472 (77)	887 (174)
	Lexical	341 (53)	2324 (75)	721 (79)	1028 (138)	432 (98)	891 (164)

Table 9 (b) Formant values in slow speech

		i		a		u	
		F1	F2	F1	F2	F1	F2
Male	Fully contracted	--	--	651	828	383	742
	Lexical	372	2492	729	1086	561	959
Female	Fully contracted	--	--	--	--	416	748
	Lexical	332	2392	742	1018	488	925

In Table 9, we find that males seemed to have higher formant values than females in the lexical tokens. Males' having this tendency could be a genuine phonetic difference since they had comparatively larger standard deviations. That means, male speakers tended to have more variations.

The following two sections examine whether the F1and F2 of the three vowels in fully-contracted tokens are distinctive from their lexical counterparts in fast speech. Since very few syllables were fully-contracted in the slow speech, only tokens in the fast speech were considered.

4.4.1 F1

Fully Contracted vs. Lexical

The F1 values for each vowel in the fast speech were entered independently into a linear mixed-effects model in which the fixed factors were Sex (Male vs. Female) and Type (Fully contracted vs. Lexical) and the random factors were Item (6 levels), Repetition (3 levels) and Speaker (10 levels).

The results for F1 in vowel /i/ are shown in Table 10. A significant main effect was only observed for Type. The main effects of this fixed factor showed that lexical tokens (370 Hz, sd = 60) had larger F1 than fully- contracted tokens (316 Hz, sd = 32). The Type effect suggests that the fully-contracted tokens were not completely neutralized, moreover, the significantly lower F1 of /i/ in the fully-contracted tokens indicates that a fully-contracted item was produced with raised tongue. The vowel of this item was closer to the glide [j] in terms of height.

Table 10 Linear mixed-effects modeling of F1 of /i/ in fast speech (Fully Contracted vs. Lexical)

Parameter	Estimate	Std. Error	t value
(Intercept)	303.36	21.40	14.175*
Sex	30.99	20.00	1.549
Type	38.80	14.09	2.754*
Type by Sex	26.75	21.33	1.254

* = significant at $p < .05$

The results for F1 in vowel /u/ are shown in Table 11. A significant main effect was only found in the interaction of Type and Sex. The Type by Sex interaction plot is displayed in Figure 18. The post-hoc *t*-tests for the interaction of Type by Sex showed that only males significantly differentiate F1s of /u/ in fully-contracted tokens from in lexical tokens, $t(54) = -5.37$, $p < .05$, with the lexical tokens having higher F1 than the fully-contracted tokens. It suggests that for male speakers, /u/ in fully-contracted tokens was more glide-like. For females, the two tokens were neutralized in terms of F1 of /u/. In addition, only for the lexical tokens did females differ from males, $t(42) = -4.28$, $p < .05$.

Table 11 Linear mixed-effects modeling of F1 of /u/ in fast speech (Fully Contracted vs. Lexical)

Parameter	Estimate	Std. Error	t value
(Intercept)	472.37	29.80	15.851*
Sex	-35.70	27.51	-1.298
Type	-34.18	18.59	-1.839
Type by Sex	131.49	26.06	5.046*

* = significant at $p < .05$

The results for F1 in vowel /a/ are shown in Table 12. No significant effect was found in any of the factors nor factor interactions. In sum, the F1 data of the three vowels showed that the fully-contracted tokens were distinctive from the lexical tokens in the height of the high vowels /i/ and /u/.

Table 12 Linear mixed-effects modeling of F1 of /a/ in fast speech (Fully Contracted vs. Lexical)

Parameter	Estimate	Std. Error	t value
(Intercept)	768.97	44.91	17.123*
Sex	-69.09	42.76	-1.616
Type	-45.47	23.98	-1.896
Type by Sex	15.05	38.62	0.390

* = significant at $p < .05$

F1 and Trough Depth Correlation

In order to see if the variation in F1 results from the gradience characteristic in syllable contraction, for each vowel, the data of trough depth and F1 of the partially-contracted and disyllabic tokens (TrDep > 0 dB) were entered into the correlation function in R. Neither F1 of /i/ nor F1 of /u/, are correlated with trough depth. For F1 of /a/, $r^2 = 0.064$ and $p < .05$, showing that although only 6.4% of the data can be explained by trough depth, F1 of /a/ and trough depth are significantly correlated. In other words, as the tongue gets lower, trough depth gets bigger, and the syllable is less contracted.

4.4.2 F2

Fully Contracted vs. Lexical

The F2 values for each vowel were entered independently into a linear mixed-effects model in which the fixed factors were Sex (Male vs. Female) and Type (Fully contracted vs. Lexical) and the random factors were Item (6 levels) , Repetition (3 levels) and Speaker (10 levels).

The results for F2 in vowel /i/ are shown in Table 13. Significant main effects were observed for Type and the Type by Sex interaction. The Type effects showed that lexical tokens (2398 Hz, sd = 172) had smaller F2 than fully-contracted tokens (2454 Hz, sd = 144). The Type effect suggests that the fully-contracted tokens were not completely neutralized, moreover, the significantly higher F2 of /i/ in the fully-contracted tokens indicates that the tongue was more front when the fully-contracted tokens were produced, and that the /i/ is more like a glide [j] in terms of vowel frontness. The post-hoc *t*-tests for the interaction of Type by Sex indicated that females discriminate F2 of /i/ of fully-contracted tokens from that of the lexical tokens, $t(42) = 6.32$, $p < .05$ but males neutralized them. Besides, males and females had significantly different F2 of /i/ for the lexical tokens, but not for the fully-contracted tokens, $t(34) = -3.58$, $p < .05$.

Table 13 Linear mixed-effects modeling of F2 of /i/ in fast speech (Fully Contracted vs. Lexical)

Parameter	Estimate	Std. Error	t value
(Intercept)	2471.30	48.25	51.21*
Sex	-45.74	71.87	-0.64
Type	-146.24	47.84	-3.06*
Type by Sex	194.14	72.82	2.67*

* = significant at $p < .05$

The results for F2 in vowel /u/ are shown in Table 14. No significant effect was observed in any of the factors nor interactions. Therefore, the two token Types seems to be neutralized in terms of F2 of /u/.

Table 14 Linear mixed-effects modeling of F2 of /u/ in fast speech (Fully Contracted vs. Lexical)

Parameter	Estimate	Std. Error	t value
(Intercept)	887.97	58.78	15.106*
Sex	-33.04	60.50	-0.546
Type	18.55	40.71	0.456
Type by Sex	96.88	57.08	1.697

* = significant at $p < .05$

The results for F2 in vowel /a/ are shown in Table 15. None of the factors caused the significant effect. In sum, the data of F2 of the three vowels shows that in addition to /i/, the fully-contracted tokens are completely neutralized and are not different from their lexical counterparts in terms of tongue frontness.

Table 15 Linear mixed-effects modeling of F2 of /a/ in fast speech (Fully Contracted vs. Lexical)

Parameter	Estimate	Std. Error	t value
(Intercept)	1067.045	76.816	13.891*
Sex	32.861	110.252	0.298
Type	-37.593	56.318	-0.668
Type by Sex	-7.445	90.549	-0.082

* = significant at $p < .05$

F2 and Trough Depth Correlation

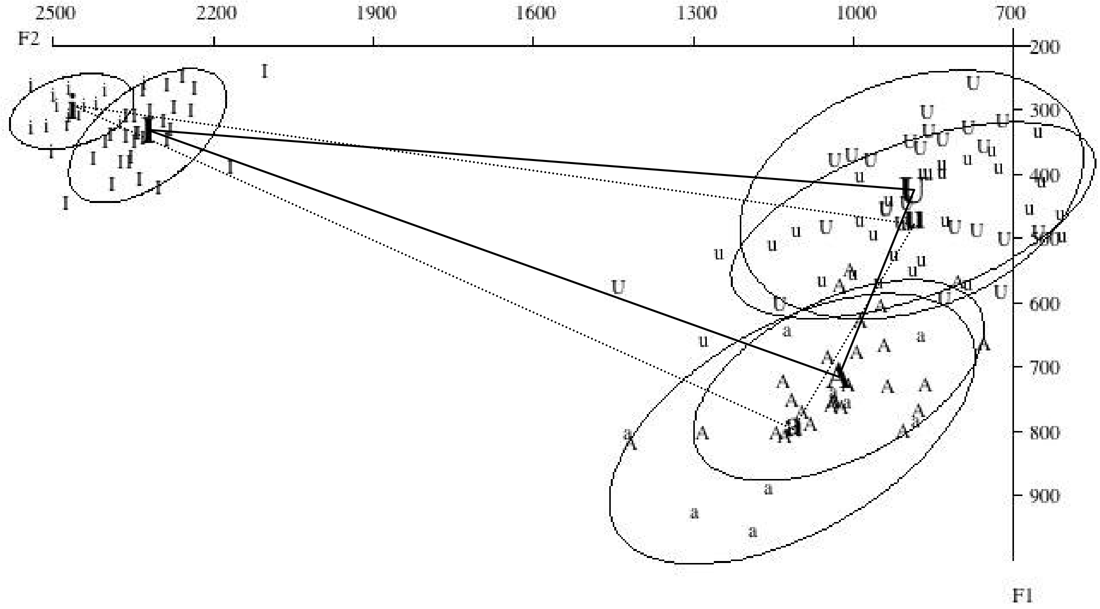
In order to see if the variation in F2 results from the gradience characteristic in syllable contraction, for each vowel, the data of trough depth and F2 of the partially-contracted and disyllabic tokens (TrDep > 0 dB) were entered into the correlation function in R. The results showed that none of the F2 measures was correlated with trough depth in that $p > .05$.

4.4.3 Vowel area

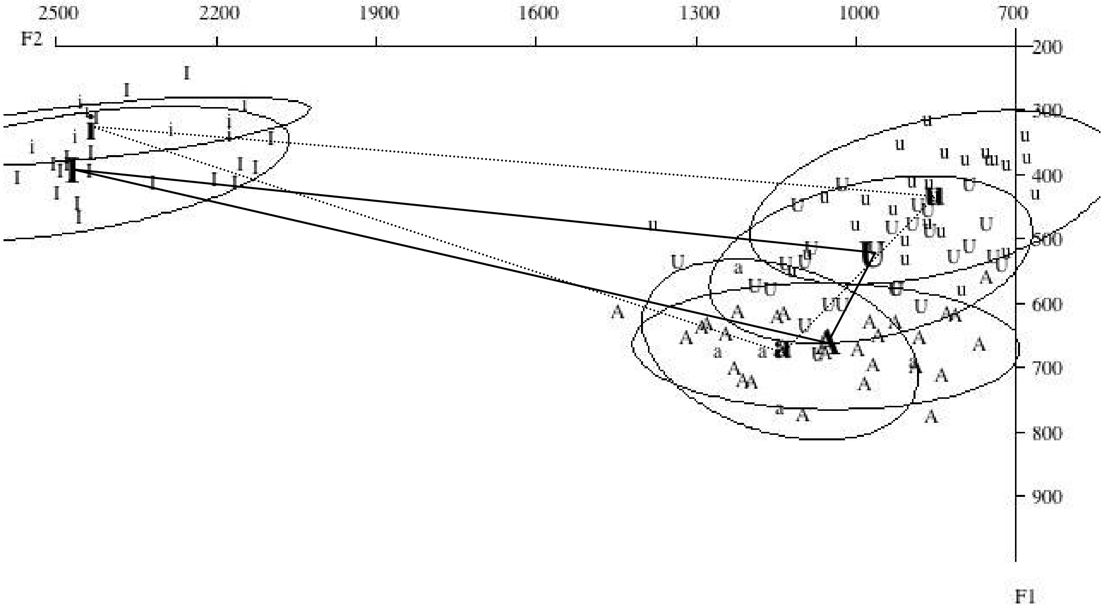
Another look at the difference between fully-contracted tokens and lexical tokens came from the measurement of vowel areas. Vowel areas were calculated from the average F1 and F2 values across Items and Repetitions. By this measurement, differences in F1 and F2 of individual vowels could combine to produce significant differences in overall area.

Recall the Type effect results of the F1 and F2 for each vowel in fast speech. For the two high vowels /i/ and /u/, F1 of the fully- contracted tokens was higher than that of the lexical tokens, suggesting that the tongue raised upward for /i/ and /u/ in the fully-contracted tokens. However, for /a/, there was no F1 difference between the two token Types. As for the F2 data, only /i/ significantly differentiates the F2 in the fully-contracted tokens from the F2 in the lexical tokens. That is, in terms of F2, the /u/ and /a/ in the fully-contracted and the lexical tokens were completely neutralized.

Unlike in the previous analyses, the data of the vowel area were entered into *t*-test with values averaged across Items and Repetitions. Type was thus the only within-subject factor. The result for the female data indicated that the vowel areas of both Types were not different from each other. The females' vowel space is displayed in Figure 12(a), along with all the individual vowel tokens. On the other hand, the areas of the two Types were significantly different from each other in the male data, $t(4) = -3.37$, $p < .05$. The vowel space of males with all the individual vowel tokens is shown in Figure 12(b). In both Figures, the capitalized vowels are data from the lexical tokens, and the small case vowels are data from the fully-contracted tokens. The triangle enclosed with the solid lines is the area of the lexical tokens, and the triangle enclosed with the dotted lines is the area of the fully-contracted tokens.



(a) Females' F1-F2 vowel space in fast speech.



(b) Males' F1-F2 vowel space in fast speech.

Figure 12 F1-F2 vowel space in fast speech for females **(a)** and males **(b)**. The bold capital letters are means for the vowels in the lexical tokens; the bold letters with small cases are means for the vowels in the fully-contracted tokens.

The vowel area result shows that female speakers do not distinguish the fully-contracted and the lexical tokens by means of vowel space size, whereas male speakers tended to use a bigger space in the oral cavity for the fully-contracted tokens.

4.4.4 Interim summary

In this section, F1 and F2 of each vowel and the vowel area were examined. The results showed the vowel area of the fully-contracted tokens is not different from that of the lexical tokens for females, but a Type difference in vowel area was found in the male data. Moreover, the Type effect was also found in the high vowels, which is, the tongue moved upward for fully-contracted tokens. It is worth noticing that for both males and females, the low vowel /a/ of different token Types were neutralized.

5. General discussion

The results of the perception experiment shows that listeners can differentiate the fully-contracted tokens from their lexical counterparts with no problems. On the production of syllable contraction, two major questions were raised at the outset of this study. First, whether syllable contraction is an optional lenition process; second, whether fully-contracted tokens and lexical tokens are distinguishable from the perspective of production and if they are, what are the acoustic cues that differentiate these two. The following discussion will address these questions based on the results of the data analyses described in the previous sections.

5.1. Optionality & Gradience

Syllable contraction seems to be optional and the data showed that it is not fully predictable from speech rate. The token distribution of trough depth in Figure 9 demonstrates that: (a) contractible tokens could be non-contracted in fast speech (17.92%); contractible tokens could be fully contracted in slow speech (6.83%). (b) The bimodal distribution of the contractible tokens in fast speech confirms that syllable contraction is optional. In addition, although the number of tokens between the two end frequency peaks for both speech rates is very low, we tried to predict the variations of the other dependent measures among the partially-contracted and disyllabic tokens (TrDep > 0 dB) using the linguistically motivated variable 'trough depth'. None of the dependent measures can explain more than 15% of the variance by using trough depth. Given that trough depth is a measure for gradience, if an acoustic measure and trough depth have a strong correlation, we could say that the acoustic measure also reflects the gradient variations in syllable contraction. In our results, TD/SD ratio, VD/TD ratio, MIN F0 and F1 of /a/ are correlated with trough depth and so they are also cues for gradience. The two durational measures are both positively correlated with trough depth, and so is F1 of /a/. MIN F0 and trough depth, on the other hand, had a weak negative correlation. The finding is rather interesting because it supports the claim that the process of syllable contraction is not entirely all-or-nothing, it appears to be a systematic gradient pattern. Furthermore, the fact that both durational cues are included implies that the variations in duration are able to capture the variations of the syllables at different degrees of contraction.

5.2 Type Effect (Fully-contracted vs. Lexical)

According to the result of a pilot perceptual study, Taiwan Mandarin listeners are able to differentiate the fully-contracted tokens (defined as TrDep = 0 dB) from their lexical counterparts with the correctness of 96%. This present study examines the potential acoustic cues that speakers could use to differentiate tokens. We have investigated eight possible acoustic correlates of Taiwan Mandarin syllable contraction: (a) TD/SD ratio (b) VD/TD ratio (c) MIN F0 (d) MAX F0 (e) F0 RANGE (f) F1 of all the cardinal vowels (g) F2 of all the cardinal vowels (h) vowel area. The first two are relative durational measures; (c-e) are three fundamental frequency cues; (f-h) are three segmental measures related to vowel space with regards to lingual (co-)articulation. These acoustic measures were chosen because they are related to articulatory processes and are from the temporal and spectral domains.

The results demonstrate that both durational measures can be used to differentiate the fully-contracted tokens from the lexical tokens, and the prosodic measure that can do so is F0 RANGE. MIN F0 can be distinguishable only in slow speech. Moreover, in fast speech, the spectral measures F1 of the high front vowel /i/ is able to differentiate two types of tokens as well.

Considering the fact that the fully-contracted tokens had greater TD/SD ratio and VD/TD ratio than the lexical tokens and the fact that fully-contracted tokens had wider pitch range, we find that the fully-contracted tokens tend to preserve the length and the pitches of the underlying forms. These minute preservations in production might then keep the listeners from misinterpreting the syllable they hear. The result of the vowel space measures showed that the Type effect was found particularly in the high front vowel. In addition, as far as the vowel area is concerned, only males differentiate vowel area of the two different tokens, with the fully-contracted tokens having larger vowel space.

In sum, the discussion in 5.1 indicates that syllable contraction in Taiwan Mandarin is optional as a process but gradient in distribution. The discussion in 5.2 indicate that the fully-contracted tokens were not complete neutralization because they are essentially different from their lexical counterparts in duration, pitch range and vowel height of the high vowels. Warner et al (2004) proposed an 'incomplete neutralization' effect for German word-final devoicing, claiming that some other detailed phonetic factors, or some morphological or orthographic distinctions may cause the listeners and speakers to differentiate the sound contrasts successfully. Syllable contraction in Taiwan Mandarin, like German word-final devoicing (e.g. Charles-Luce, 1985; Fourakis & Iverson, 1984; Port & O'Dell, 1985), must have an incomplete neutralization effect.

6. Conclusion

This study has provided answers to two research questions. First, syllable contraction is an optional lenition process and the distribution of trough depth is the evidence. Trough depth seems to be an ideal measure for gradience because it is correlated with VD/TD ratio, TD/SD ratio, MIN F0 and F1 of /a/. Regarding the other research question, we found that both segmental and prosodic cues were responsible for the distinction of the two token Types. The fully-contracted tokens had larger TD/SD ratio, larger VD/TD ratio and wider F0 RANGE. In addition, speakers differentiated the fully-contracted tokens from their lexical counterparts in terms of vowel height and vowel space. The fully-contracted tokens had higher high front vowel /i/ and male speakers tended to enlarge the oral cavity for the fully-contracted tokens. This study has verified that syllable contraction in Taiwan Mandarin is optional in occurrence and gradient in variation. Moreover, the acoustic cues that contribute to differentiating the fully-contracted tokens and the lexical tokens were found. These findings offer supporting evidence for the argument outlined in the introduction that syllable contraction is not a complete neutralization.

Reference

Baayen, R. H. (2008) *Analyzing Linguistic Data: A Practical Introduction to Statistics Using R.* Cambridge University Press.

Charles-Luce (1985) Word-final devoicing in German: effects of phonetic and sentential contexts, *Journal of Phonetics*, 13, pp. 309-324.

Cheng, R. L. (1985) Sub-syllabic morphemes in Taiwanese. *Journal of Chinese Linguistics*, 13, 1, pp. 12-43.

Cheng C. & Xu Y. (2009) Extreme reductions: Contraction of disyllables into monosyllables in Taiwan Mandarin. Interspeech 2009. Brighton.

Chinese Knowledge Information processing Group (CKIPG). 1993. *Corpus-based Frequency Count of Characters in Journal Chinese: Corpus-based Research Series no. 1.* Taipei: Academia Sinica.

Chinese Knowledge Information processing Group (CKIPG). 1993. *Corpus-based Frequency Count of Words in Journal Chinese: Corpus-based Research Series no. 2.* Taipei: Academia Sinica.

Chung, R (1997). Syllable contraction in Chinese. Proceedings *of the Third Conference on Chinese Languages and Linguistics.* pp. 199-235.

Chung, K. S. (2006) Contraction and backgrounding in Taiwan Mandarin. *Concentric: Studies in Linguistics,* 32, 1, pp. 69-88.

Fourakis, M. & Iverson, G. K. (1984) On the incomplete neutralization of German final obstruents. *Phonetica,* 41, pp. 140-149.

Fourakis (1986) An acoustic study of the effects of tempo and stress on segmental intervals in Modern Greek. *Phonetica,* 43, pp. 172-188.

Fougeron & Jun (1998) Rate effects on French Intonation: Phonetic realization and prosodic organization. *Journal of Phonetics,* 26, 1, pp. 45-70.

Gahl, S. (2008). "Time" and "thyme" are not homophones: Word durations in spontaneous speech. *Language,* 84, 3, pp. 474-496.

Gay (1978) Physiological and acoustic correlates of perceived stress. *Language and Speech* 21, pp. 347-353.

Hsiao Y. (2002) Tone Contraction, *Chinese Languages and Linguistics VIII,* Academia Sinica, Taipei, pp. 1-16.

Hsu, H.-C. (2003) A sonority model of syllable contraction in Taiwanese Southern Min. *Journal of East Asian Linguistics,* 12, 4, pp. 349-377.

Lindblom, B. (1963) Spectrographic study of vowel reduction. *Journal of the Acoustical Society of America* 35: 1773-1781.

Mermelstein, P. (1975) Automatic segmentation of speech into syllabic units. *Journal of the Acoustical Society of America,* 58(4): 880-883.

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

Port, R. (1976) The influence of speaking tempo on the duration of stressed vowel and medial stop in English trochee words. Ph.D. Dissertation. University of Connecticut.

Port, R. & O'Dell, M. (1985) Neutralization of syllable-final voicing in German, *Journal of Phonetics*, 13, pp. 455-471.

Tseng C.-C. (1999) Contraction in Taiwanese: Synchronic analysis and is connection with diachronic change. *Chinese languages and linguistics V: Interactions in languages.* pp. 205-232. Taipei: Academia Sinica.

Tseng S. C. (2005) Contracted Syllables in Mandarin: Evidence from Spontaneous Conversations. *Language and Linguistics,* 6, 1, 155-182.

Tseng S. C. (2006) Syllable Contractions in a Mandarin Conversational Dialogue Corpus. *International Journal of Corpus Linguistics*, 10, 1, 63-83.

Warner N., Good, E., Jongman, A., & Sereno, J. (2004) Orthographic vs. morphological incomplete neutralization effects. *Journal of Phonetics* 34, 2, 285-293.

Wong, W. Y. P. (2006) Syllable fusion in Hong Kong Cantonese connected speech. Ph.D. Dissertation. The Ohio State University.

Appendix A
Reading Materials

The whole sentences are written in Pinyin with target phrases/words bold. The target phrases/words are in IPA along with the tones.

1.	a. *ta shuo* ***zhi-dao*** *xiang-ji hen gui*	/tʂʅ^{55} taw^{51}/
	b. *ta shuo* ***zhao***-*xiang-ji hen gui*	/tʂaw^{51}/
2.	a. *ta shuo* ***zhe-zhong*** *tsai bu hao*	/tʂə^{51} tʂoŋ^{21}/
	b. *ta shuo* ***zhong*** *tsai bu hao*	/tʂoŋ^{51}/
3.	a. *ta shuo* ***ba-dian***-*dang shou-huan-ying*	/pa^{55} tien^{21}/
	b. *ta shuo* ***ban***-*dang shou-huan-ying*	/pan^{55}/
4.	a. *ta shuo* ***na-yang*** *jiu bu rong-yi*	/na^{51} jaŋ^{51}/
	b. *ta shuo* ***niang***-*jiu bu rong-yi*	/niaŋ^{51}/
5.	a. *ta shuo* ***wan-shang*** *lu hen wei-xian*	/wan^{21} ʂaŋ^{51}/
	b. *ta shuo* ***wang*** *lu hen wei-xian*	/waŋ^{21}/
6.	a. *ta shuo* ***jin-tian*** *dan bu hao-chi*	/tɕin^{55} tʰien^{55}/
	b. *ta shuo* ***jian*** *dan bu hao-chi*	/tɕien^{55}/
7.	a. *ta shuo* ***zhong-shan***-*tang li-shi you-jiu*	/tʂoŋ^{55} ʂan^{55}/
	b. *ta shuo* ***zhuan***-*tang li-shi you-jiu*	/tʂuan^{55}/
8.	a. *ta shuo* ***hu-ran***-*jian bie-shu kai le*	/hu^{55} ʐan^{35}/
	b. *ta shuo* ***huan*** *jian bie-shu kai le*	/huan^{51}/
9.	a. *ta shuo* ***ge-wei*** *tong-xue xin-ku le*	/kə^{51} wej^{51}/
	b. *ta shuo* ***guei*** *tong-xue sin-ku le*	/kʷej^{51}/
10.	a. *ta shuo* ***bao-gao*** *ban-zhang bu hao*	/paw^{51} kaw^{51}/
	b. *ta shuo* ***bao*** *ban-zhang bu hao*	/paw^{51}/
11.	a. *ta shuo* ***shi-shi***-*shang huai-ren hen-duo*	/ʂʅ^{51} ʂʅ^{35}/
	b. *ta shuo* ***shi***-*shang huai-ren hen-duo*	/ʂʅ^{51}/
12.	a. *ta shuo* ***yu-yen***-*shi zi-liao hen-duo*	/ɥy^{21} jen^{35}/
	b. *ta shuo* ***yuan***-*shi zi-liao hen-duo*	/ɥen^{35}/
13.	a. *ta shuo* ***bu-yao*** *qu tai-bei hen zheng-chang*	/pu^{35} jaw^{51}/
	b. *ta shuo* ***biao*** *qu tai-bei hen zheng-chang*	/piaw^{55}/
14.	a. *ta shuo* ***da-jia*** *dou ai chang-ge*	/ta^{51} tɕia^{55}/
	b. *ta shuo* ***da*** *dou ai chang-ge*	/ta^{51}/
15.	a. *ta shuo* ***gang-gang*** *tsai lai guo*	/kaŋ^{55} kaŋ^{55}/
	b. *ta shuo* ***gang*** *tsai lai guo*	/kaŋ^{55}/
16.	a. *ta shuo* ***xin-dian*** *shi-gong-suo hen-da*	/ɕin^{55} tien^{51}/
	b. *ta shuo* ***xian***-*shi-gong-suo hen-da*	/ɕien^{51}/
17.	a. *ta shuo* ***muo-tuo***-*che hen-gui*	/muo^{35} tʰuo^{55}/
	b. *ta shuo* ***muo***-*che hen-gui*	/muo^{35}/

18. a. *ta shuo* ***da-gai*** *bu-zou le* /ta^{51} kaj^{51}/
b. *ta shuo* ***dai-****bu-zou le* /taj^{51}/
19. a. *ta shuo* ***ming-tian*** *xu-tan hen-you-yi-si* /miŋ^{35} tʰien^{55}/
b. *ta shuo* ***mian****-xu-tan hen-you-yi-si* /mien^{35}/
20. a. *ta shuo* ***xia-wu*** *de ren bi-jiao duo* /ɕia^{51} wu^{21}/
b. *ta shuo* ***xiao*** *de ren bi-jiao duo* /ɕiaw^{51}/