Extreme reductions: Contraction of disyllables into monosyllables in Taiwan Mandarin

Chierh Cheng, Yi Xu

Department of Speech, Hearing and Phonetic Sciences, University College London, UK

chierh@phonetics.ucl.ac.uk, yi.xu@ucl.ac.uk

10.21437/Interspeech.2009-162

Abstract

This study investigates a severe form of segmental reduction known as contraction. In Taiwan Mandarin, a disyllabic word or phrase is often contracted into a monosyllabic unit in conversational speech, just as "do not" is often contracted into "don't" in English. A systematic experiment was conducted to explore the underlying mechanism of such contraction. Preliminary results show evidence that contraction is not a categorical shift but a gradient undershoot of the articulatory target as a result of time pressure. Moreover, contraction seems to occur only beyond a certain duration threshold. These findings may further our understanding of the relation between duration and segmental reduction.

Index Terms: contraction, reduction, undershoot, duration, speech rate, Taiwan Mandarin

1. Introduction

Severe segmental reduction is one of the least understood aspects of speech. A number of linguistic studies on allophonic variations in English and German have provided a starting point concerning the relationship between phonetic context, target attainment and articulatory effort [9, 10]. Lessons from pronunciation models for automatic speech recognition systems also indicate that model construction with segmental context and duration features can improve estimations on how syllables are pronounced in context [3, 4]. The common perception of such varied phonetic phenomena is that they stem from interactions between a number of segmental and durational variables suggested by linguistic and engineering literature, but the exact underlying mechanism is still unclear. Taiwan Mandarin has been reported to have widespread canonical forms in shorthand [1, 2, 5, 6, 7, 11]. For example, *wo zhi dao* [wo˨˩˦ tʂɨ˥ tau˥˩],'I know', is contracted into *wo zhao* [wo˨˩˦ tʂau]. In more extreme cases, trisyllables can also be contracted into monosyllabic units, such as *wo bu zhi dao* [wo˨˩˦ pu˥˩ tʂɨ˥ tau˥˩], 'I don't understand', being contracted into *wo bao* [wo˨˩˦ pwau].[1] This study attempts to build on existing research and offer new insights into the underlying mechanisms of segmental reduction by examining the effects of duration on syllable contraction in Taiwan Mandarin.

If, as hardly anyone would deny, articulatory movements in speech take time, for each movement there must be a minimum duration below which its execution becomes impossible. This physical fact actually underlies the notion of intrinsic duration of segments which, according to Klatt [8], is "an absolute minimum duration D_{min} that is required to execute a satisfactory articulatory gesture." From this notion we have derived two hypotheses about syllable contraction in Taiwan Mandarin. The first is that, contrary to the common perception that it is only a feature of casual speech, contraction can be elicited from nonsense sequences in the laboratory by directly manipulating speech rate. The second hypothesis is that it is possible to identify a particular duration below which contraction is regularly observed. This study is to test both of these hypotheses in Taiwan Mandarin.

Our first strategy was to simply ask subjects to speed up their production in the laboratory and see if contraction occurs in their utterances. This may help reveal whether contraction is directly related to time pressure. Secondly, in order to investigate the minimum requirement for articulatory implementation and thus its allocated time when severe reduction such as contraction occurs, syllable sequences with different syllabic structures and consonant types were used. This may help reveal whether segmental reduction occurs as a particular minimal duration.

2. Methodology

2.1. Stimuli

A set of 32 nonsense disyllabic sequences was constructed as testing material, as shown in Table 1. They form four groups according to the level of obstruction by the intervocalic consonant: 1) zero obstruction — CV+V; CV+VN; CV+VV, 2) nasal consonant — CVN+V; CV+NV, 3) non-nasal consonant — CV+CV, where C is a plosive (*pl.*), fricative (*fr.*) or affricate (*af.*), and 4) nasal + non-nasal consonant — CVN+CV. All intervocalic consonants are chosen to have a similar place of articulation but different manners of articulation in order to control for obstruction levels. The vowels in these sequences are chosen to be the corner vowels /i/, /ɑ/ and /u/ in order to maximize the amplitude of formant movement. All the disyllabic units are given the high-level tone in order to minimize any potential tonal effects [12].

In addition, a carrier sentence consisting of three phrases was devised. The same target sequence was embedded in each phrase and hence produced three times within this carrier (also in Table 1). The first phrase consists of eight syllables, the second 13 syllables and the third 15 syllables. The aim was to impose different amounts of time pressure on the target disyllabic sequences.

2.2. Subjects and procedure

Four bilingual Taiwanese/Mandarin male speakers were recorded. The speakers were aged between 21 and 28 and had no self-reported speech disorders and little professional vocal training. The recordings were conducted in the anechoic chamber of University College London. Speech was recorded with a Shure SM10A microphone placed approximately 30 centimeters from the subjects' mouth. All stimuli were presented to the subjects in traditional Chinese characters and each time only one carrier sentence with the embedded stimuli was shown on the screen in front of the seated subject.

Copyright © 2009 ISCA

To control the level of time pressure, subjects were instructed to articulate the material at three speaking rates, **slow/clear** as if reciting in class, **natural** as if having a conversation with a friend, and as **fast** as possible. The exact speed of articulation, however, was left to the subjects' own discretion. (The average speech rates of slow, natural and fast speech across the four subjects were 4.5, 6.1 and 9.3 syllables per second, respectively.) In this fashion, we aim to obtain both canonical and elliptic forms of each target sequence depending on its position within the carrier sentence and rate of speech. In order to observe whether repetition time contributes to contraction rate, we used three randomized blocks of the above 32 sentence sequences. In total, the number of target sequences for analysis was 32 (stimuli) × 4 (subjects) × 3 (positions in the carrier) × 3 (speech rates) × 3 (blocks/repetitions) = 3456 tokens.

Table 1. *Stimuli and carrier sentence.*

Intervocalic obstruction levels from low to high				
Disyllabic structure		Phonetic presentation and characters shown to the subjects		
1. Zero obstruction				
CV+V		/ti/+/i/ 滴依	/ta/+/a/ 搭阿	/tu/+/u/ 督巫
CV+VN		/ti/+/in/ 滴因	/ta/+/an/ 搭安	/tu/+/un/ 督溫
CV+VV		/ti/+/ai/ 滴哀 /tu/+/au/ 督凹	/ti/+/au/ 滴凹	/tu/+/ai/ 督哀
2. Nasal consonant				
CV<u>N</u>+V		/tan/+/i/ 單依	/tan/+/u/ 單巫	
CV+<u>N</u>V		/ta/+/ni/ 搭妮	/ta/+/nu/ 搭奴	
3. Non-nasal consonant				
CV+<u>C</u>V where C is a	*pl*	/ta/+/ti/ 搭滴	/ta/+/tu/ 搭督	/ta/+/ta/ 搭搭
	pl^{h}	/ta/+/t^{h}i/ 搭踢	/ta/+/t^{h}u/ 搭禿	/ta/+/t^{h}a/ 搭他
	fr	/ta/+/ɕi/ 搭悉	/ta/+/su/ 搭蘇	/ta/+/sa/ 搭撒
	af	/ta/+/tɕi/ 搭激	/ta/+/tsu/ 搭租	/ta/+/tsa/ 搭紮
	af^{h}	/ta/+/tɕ^{h}i/ 搭戚	/ta/+/ts^{h}u/ 搭粗	/ta/+/ts^{h}a/ 搭擦
4. Nasal+ non-nasal consonant				
CV<u>N</u>+<u>C</u>V		/ɕin/+/ti/ 新滴	/sun/+/ti/ 孫滴	/san/+/ti/ 三滴

Carrier Sentence

Character:「你說的是□□是吧！我當然不吃□□沙拉那種東西，因為我最不喜歡他家出的□□沙拉！」

Pinyin: "ni shuo de shi □□ shi ba! wo dang ran bu chi □□ sha la na zhong dong xi, yin wen wo zui bu xi huan ta jia chu de □□ sha la."

English: "You meant □□, didn't you! Of course I won't eat □□ salad that kinda stuff, because I dislike □□ salad made by his family the most!"

2.3. Measurement

One difficulty with research in this area is that, as of yet, there are no standard quantitative methods for measuring contraction rate. We have therefore adopted a set of working definitions for labelling the occurrence of contraction. Since definitive segmentation is impossible, tokens were labelled as 'contracted', 'semi-contracted' and 'non-contracted' according to their degree of segmental weakening or loss. ***Non-contracted*** units were those with clear interruption of formants by the intervocalic consonant, presence of nasal murmur or a clearly lowered F1. ***Contracted*** units are those with continuous F1 with neither interruption by intervocalic consonants nor presence of nasal murmur. ***Semi-contracted*** units are those for which the above segmentation criteria are difficult to apply and no straightforward delimitation of the spectrogram can be made. To be able to calculate contraction rate and to perform statistical analysis, non-contracted, semi-contracted and contracted units were given contraction values of 0, 0.5 and 1, respectively.

It is important to note that in the zero obstruction group, most tokens were classified as contracted because the disyllabic sequence consists of an open CV syllable followed by a syllable with a vowel onset. The vowel-to-vowel junction is usually not interrupted by a consonant in Taiwan Mandarin. Hence, unless there was a clear sign of glottal stop or glottalization between the two vowels, they were all technically classified as contracted. Importantly, we need to emphasize that these contraction labels serve only as indicators of the degree of departure from their canonical non-contracted forms. They are based on technical definitions with no direct theoretical claims.

3. Results and Discussion

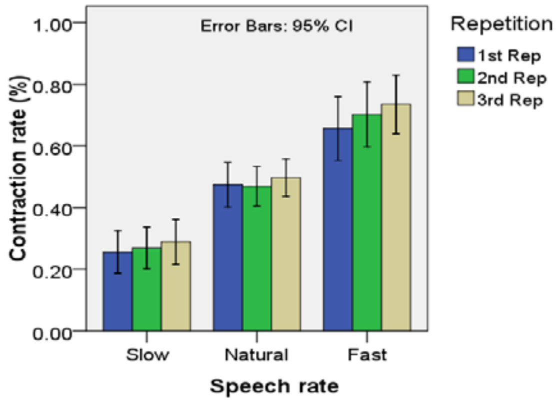

Figure 1: *Contraction rate at different speech rate and repetition time.*

Figure 1 shows the relationship between contraction rate and speech rate. The x-axis shows three different speech rates and the y-axis shows contraction rates in terms of percentage of occurrence. Within each speech rate, contraction rates of the three repetitions are shown separately.

As shown in Figure 1, as speech rate increased, speakers contracted disyllabic units more frequently. However, within each speech rate, there is no evident difference in contraction rate across the three repetitions. Only at the fast rate can we see a small increase in contraction rate across the repetitions. A three-way repeated measures ANOVA on contraction rate was performed with repetition, speech rate and position in carrier as independent variables. The results are shown in Table 2.

Table 2. *Results of three-way repeated measures ANOVA: repletion time, speech rate and position in carrier.*

Source of Variation	*df*	*F*	*Sig*
Repetition	2, 6	2.011	.215
Speech rate	2, 6	28.060	***.001***
Position	2, 6	3.329	.107
Repetition * Speed	4, 12	1.315	.320
Repetition * Position	4, 12	.418	.793
Speed * Position	4, 12	1.410	.289
Repetition* Speed * Position	8, 24	1.275	.302

Table 2 shows that speech rate has a highly significant effect on contraction rate. Neither repetition nor position in carrier has a significant effect. And there are no interactions. Since such a highly consistent relationship between contraction rate and speech rate was present in our data, the results are consistent with the hypothesis that contraction is not a feature unique to casual speech and can be elicited from nonsense disyllabic units in the laboratory. In other words, having a fast speech rate is a sufficient condition for contraction to occur, be it in spontaneous speech or recitation performed in the laboratory.

To more closely examine the relation between time pressure and contraction rate, we computed contraction rates at varying levels of consonantal obstruction. As shown in Table 3, as the level of obstruction increases, contraction rate decreases. This inverse relation might indicate that contraction rate is determined by level of articulatory demand. Further analysis shows that time pressure is in fact a more direct determining factor as more demanding manners of articulation is assigned longer intrinsic time.

Table 3. *Contraction rate (%) at different levels of intervocalic obstruction.*

Intervocalic obstruction levels from low to high			
Disyllabic structure		**Contraction rate (%)**	**Average overall (%)**
1. Zero obstruction			
CV+V		86%	87.1%
CV+VN		90%	
CV+VV		85.3%	
2. Nasal consonant			
CV<u>N</u>+V		64.5%	44.8%
CV+<u>N</u>V		25%	
3. Non-nasal consonant			
CV+<u>C</u>V where C is a	plosive	26%	22.6%
	plosive^h	25%	
	fricative	20%	
	affricate	25.7%	
	affricate^h	16.3%	
4. Nasal+ non-nasal consonant			
CV<u>N</u>+<u>C</u>V		13%	13%

Figure 2 shows mean segmental duration of non-contracted t<u>a+CV</u> units according to the manner of articulation of the intervocalic consonant. The x-axis represents the duration of preceding /a/ and that of the intervocalic consonant and the following vowel. The y-axis shows the different levels of consonantal obstruction in the t<u>a+CV</u> sequences, from low to high, showing plosive, aspirated plosive, affricate, aspirated affricate, respectively.

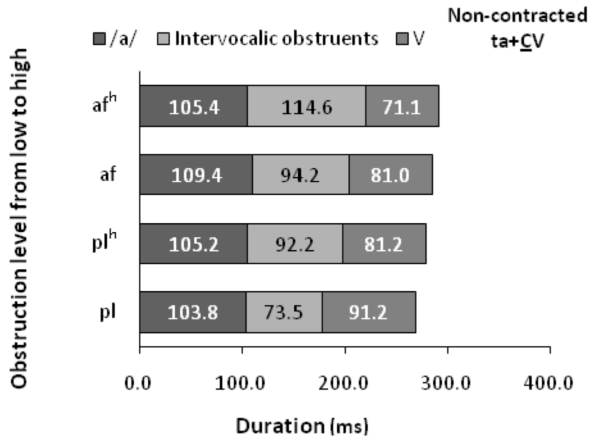

Figure 2: *Varying segmental durations in t<u>a+CV</u> units with different levels of consonantal obstruction.*

It is of interest to see in Figure 2 that in non-contracted units, the duration of the intervocalic consonant varies with their levels of obstruction. This indicates that the time allocated to a consonant is related to its level of obstruction. Consequently, as speech rate increases, those consonants with shorter allocated time would be more likely to be reduced to the point of losing all its spectral cues of consonantal obstruction.

To show the relationship between duration and formant excursion size, a scatter plot of all tokens in Group 3 is shown in Figure 3 (since Group 3 contains the core variant phenomena we are interested in). The x-axis represents the combined duration of /a/ and the second syllable in t<u>a+CV</u> sequences. The y-axis represents the sum of the excursion sizes of F1 and F2 within this interval. The three contraction types, non-contracted, semi-contracted and contracted, are represented by different symbols as indicated in the legend.

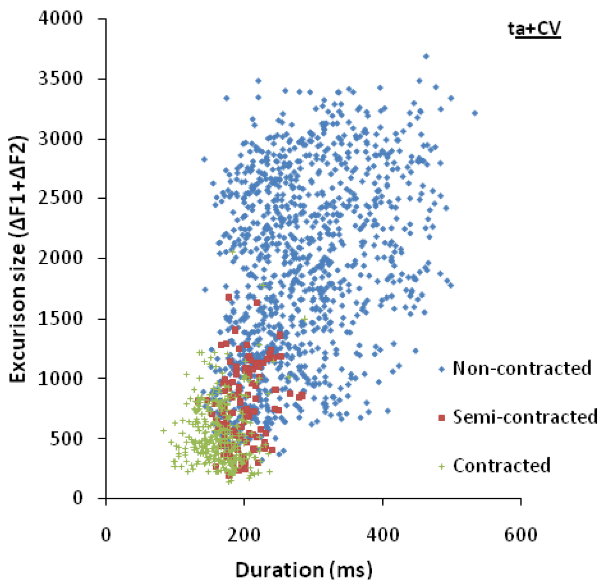

Figure 3: *Scatter plot of formant excursion size (ΔF1+ΔF2) over formant duration: ta+CV.*

The scatter plot demonstrates a positive relation between duration and formant excursion size in t<u>a+CV</u> sequences: the shorter the duration, the smaller the excrusion size. When the duration becomes extremely short—under 200ms, contraction per our technical definition given earlier occurs. It is intersting to note that semi-contraction cases cluster about the duration of 200 ms. More severe reductions can be seen in faster local rates of articulation, i.e. in units of less than 200 ms. Therefore, the 'duration threshold' for integrity of disyllables seems to be approximately 200ms.

To see the time demand of varying articulatory functions in a straightforward manner, mean segmental durations of all stimuli in Group 3 are plotted in Figure 4 according to degree of segmental reduction — **non-contracted**, **semi-contracted**, and **contracted**. The figure exhibits a horizontal compression along the time dimension. **Non-contracted** units may reflect the amount of time used in canonical articulations. In **semi-contracted** units, durations of all segments are reduced, with the most severe reduction in the intervocalic obstruents (47.9 ms). Finally, in **contracted** units, the overall duration of the disyllabic words is compressed to the point at which no intervocalic consonantal closure is possible. Thus the 47.9 ms in the semi-contracted case seems to be the duration at which the intervocalic consonants begin to be severely modified, and below which they are virtually 'lost.'

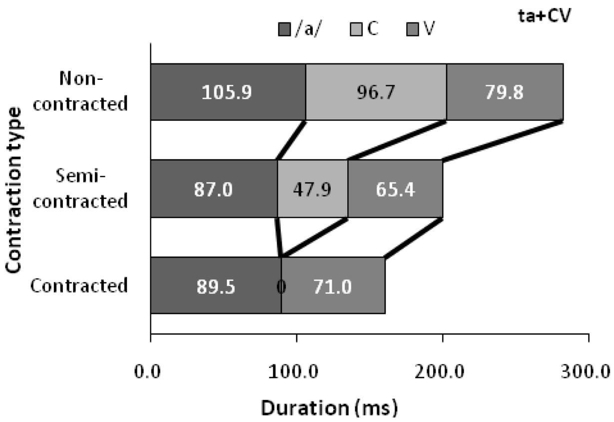

Figure 4: *Segmental durations at different contraction levels.*

The results of the above analyses provide initial evidence that duration is the most direct determining factor for the occurrence of contraction in Taiwan Mandarin. This finding is in agreement with Lindblom's duration-dependent target undershoot model [10], which predicts that greater undershoot occurs with shorter duration. Contraction, as the most severe form of reduction, is just one step further from undershoot. We have seen evidence that as time pressure increases, execution of the articulatory movement toward a consonantal target becomes difficult, resulting in what we call semi-contraction. As time pressure further increases, the articulatory movement simply becomes impossible, thus resulting in contraction. It is apparent from these results that contraction is not a categorical shift, but a gradient undershoot of articulatory targets as a function of time pressure.

The findings of this study are still quite preliminary. More data have been collected and we are planning to examine in even greater detail how exactly the trajectories of formant movements change from canonical forms to semi-contracted and contracted forms. And we plan to assess the contribution of articulatory effort, if any, to contraction in Taiwan Mandarin.

4. Conclusions

To the best of our knowledge, no previous research has been conducted from an articulatory/acoustic perspective to test the basic idea that contraction in Taiwan Mandarin is a consequence of certain articulatory facets controlled by speakers. In this study, contraction was successfully elicited in the laboratory from nonsense disyllabic sequences and evidence was found that contraction is directly related to time pressure, and it occurs whenever segmental duration is shortened beyond a certain threshold. These findings offer supporting evidence for the two hypotheses outlined in the introduction. What we have found in the present study, especially in terms of the mechanisms underlying contraction may not be limited to only Taiwan Mandarin. For example, the reduction form such as "I'm; don't", although apparently fossilized in contemporary English, may have gone through a similar process of gradient contraction under time pressure in earlier English. It is also possible that severe forms of reduction similar to contraction in Taiwan Mandarin are more widespread than has been recognized. Further improvement of our understanding of the underlying mechanisms of phonetic reduction is therefore highly desirable.

5. Acknowledgements

This research was supported in part by NIH grant No. 1R01DC006243 to the second author.

6. Endnotes

[1] It should be noted that in these examples two types of contractions are involved. One is segmental contraction and the other is tonal contraction, as Mandarin is a tonal language. The current study investigates only segmental contraction. Tonal contraction will be studied in the future.

7. References

[1] Chung, K. S., "Contraction and backgrounding in Taiwan Mandarin", Concentric: Studies in Linguistics, 32: 69-88, 2006.

[2] Chung, R., "Syllable contraction in Chinese", Proceedings of the Third Conference on Chinese Languages and Linguistics, 199-235, 1997.

[3] Fosler-Lussier, E., Morgan, N. "Effects of speaking rate and word frequency on pronunciations in conventional speech", Speech Communication, 29: 137-158, 1999.

[4] Fosler-Lussier, E., Williams, G. "Not just what, but also when: Guided automatic pronunciation modeling for broadcast news", in: ADRPA Broadcast News Workshop, Herndon, VA, 1999.

[5] Greenberg, S. "Speaking in shorthand—A syllable-centric perspective for understanding pronunciation variation", Speech Communication, 29: 159-176, 1999.

[6] Hsiao, C. 1986, "A preliminary investigation into phonetic contraction in Mandarin", The World of Chinese Language, 40: 46-50, 1986.

[7] Hsiao, Y. "Tone contraction", The Eighth International Symposium on Chinese Languages and Linguistics, Academia Sinica, 2000.

[8] Klatt, D. H., "Linguistic uses of segmental duration in English: Acoustic and perceptual evidence", J. Acoust. Soc. Am, 59: 1208-1221, 1976.

[9] Kohler, K. J., "Segmental reduction in connected speech in German: phonological facts and phonetic explanation", Speech Production and Speech Modelling, W. J. Hardcastle & A. Marchal, eds., Dordrecht: Kluwer., 69-92, 1990.

[10] Lindblom, B., "Explaining phonetic variation: A sketch of the H&H theory", Speech Production and Speech Modelling, W. J. Hardcastle & A. Marchal, eds., Dordrecht: Kluwer., 403-439, 1990.

[11] Tseng, S. C. "Syllable Contractions in a Mandarin Conversational Dialogue Corpus", International Journal of Corpus Linguistics, 10 (1): 63-83, 2005.

[12] Xu, Y. "Sources of tonal variations in connected speech", Journal of Chinese Linguistics, monograph series, 17: 1-31, 2001.