MODELLING EXTREME TONAL REDUCTION IN TAIWAN MANDARIN BASED ON TARGET APPROXIMATION

Chierh Cheng, Yi Xu & Santitham Prom-on

Department of Speech, Hearing and Phonetic Sciences, University College London, UK
chierh.cheng@googlemail.com; yi.xu@ucl.ac.uk; santitham.prom-on@ucl.ac.uk

ABSTRACT

This study explores the underlying mechanisms of tonal reduction in Taiwan Mandarin through the use of computational modelling. Using the quantitative target approximation model (qTA) we tested the hypothesis that *extreme tonal reduction stems from severe time pressure, despite speakers' attempting to achieve the same underlying targets.* A series of modelling tests were performed on experimental corpora containing a large amount of severe tonal reduction cases produced by six Taiwan Mandarin speakers. The results showed support for the hypothesis. This provides further evidence for the time pressure account of phonetic reduction in general.

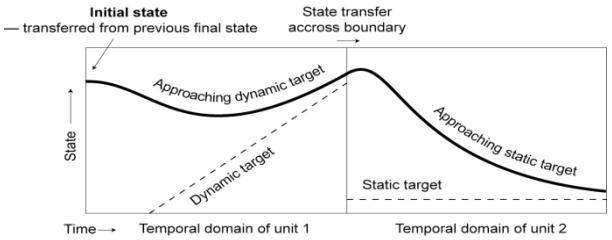
Keywords: tone reduction, modelling, Taiwan Mandarin, quantitative target approximation (qTA)

1. INTRODUCTION

There has been a plethora of research regarding the sources of phonetic reduction [1, 6, 8, 10]. One of the most important factors influencing phonetic reduction is speech rate, which can more specifically be referred to as the local rate of attempted target segments/words [5, 9, 11, 15]. It has also been suggested that speakers constantly reach their physiological limit of articulation when speech rate is high, often resulting in severe reduction [7]. This paper aims to explore the underlying mechanisms of tonal reduction in Taiwan Mandarin, focusing on the relation between duration and F_0 realisation.

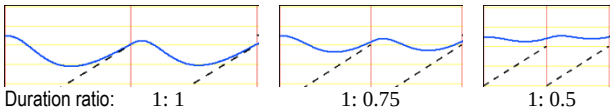
Tonal languages, such as Taiwan Mandarin, have been reported to have two types of reduction involved when speech rate is high, i.e. segmental reduction [3, 12, 14] and tonal reduction [2, 4, 12, 14]. Previous research has shown evidence that extreme segmental reduction is the direct result of speaker's attempt to achieve the underlying segments under severe time pressure [3]. It has also been demonstrated that when two syllables are merged into one, speakers still seem to make an effort to produce the original tones within the contracted syllable [4]. The present study is to further test the underlying mechanism of extreme tonal reduction in Taiwan Mandarin, by using an articulatory-based model to simulate F_0 contours. The model used is the quantitative target approximation model (qTA), which is an implementation of the theoretical target approximation (TA) model, shown in Fig. 1.

Figure 1: Target approximation model, adapted from [19].



According to TA, lexical tones are produced by a process of articulatorily approaching successive underlying pitch targets, each synchronize with its host syllable [19]. The degree to which the tonal target is achieved depends on a) the distance between the initial F_0 and the target, b) the rate of target approximation (determined by articulatory strength) and c) the duration of the syllable. Thus when a) and b) remain constant, shortening syllable duration alone can lead to undershoot of the tonal target [18]. As shown in Fig. 2, qTA can simulate increased flattening of F_0 contours in two consecutive Rising tones by simply shortening the duration of the syllables.

Figure 2: Effect of syllable shortening on two consecutive Rising tones (preceded by a High tone, not shown here), simulated by an interactive demo of qTA at http://www.phon.ucl.ac.uk/home/yi/qTA/



To investigate possible underlying mechanisms of this phenomenon, the present study was designed to test, by using qTA modelling, *whether or not extreme tonal reduction stems from severe*

time pressure despite speakers' attempting to achieve the same underlying targets.

2. METHODOLOGY

The basic modelling strategy of the study was first to train the qTA model using the non-reduced tones to obtain raw tonal target parameters (Sec. 2.2). These parameters were then averaged across the repetitions subject by subject in order to obtain respective canonical parameters for each subject. Three subsequent simulations were then conducted with the canonical target parameters of each subject. The first was to simulate F_0 contours of each of the individual non-reduced bi-tonal sequences, using the canonical parameters along with their own duration and initial F_0 (Sec. 3.1). The second was to simulate the reduced tones, again using the canonical parameters and the duration and initial F_0 of the reduced items (Sec. 3.2). In the third simulation, the duration of the reduced tones was again used, but the target values applied were randomly selected from the canonical target sets. This was to check against the possibility that the reduced F_0 contours were simply flattened (Sec. 3.3). The performances of the simulations were evaluated in terms of goodness of fit to the original F_0 contours, measured in Root Mean Squared Error (RMSE) and Pearson's correlation coefficient (correlation).

2.1. Corpus

The speech material consists of disyllabic /ma/+/ma/ nonsense sequences with a total of 16 (4x4) tonal combinations, which were embedded in two carrier sentences (target items were preceded by H/L tones). The material was recorded by six male native Taiwan Mandarin speakers. Various conditions were imposed to elicit different degrees of tonal reduction, including positions in a carrier sentence (initial, mediate and final), repetition time (1^{st}, 2^{nd}, and 3^{rd}) and speech rate (slow, normal and fast). Three reduction types were labelled according to the integrity of the intervocalic /m/, *non-reduced* for a clear presence of an intervocalic nasal, *reduced* for a clear absence of nasal murmur and *semi-reduced* for intermediate cases. More details about the corpus can be found in [4].

2.2. Extracting qTA parameters from non-reduced bi-tonal sequences

The modelling was done with a Praat script that implements the qTA model [13]. The script is a modified version of the publicly released PENTAtrainer [17]. It simulates the F_0 contours of an utterance by applying qTA through automatic analysis-by-synthesis. For each interval to be simulated, the script extracts three target approximation parameters: *m* and *b*, which define the slope and height of a linear target, and λ, which specifies the rate of target approximation [13]. Parameter estimation was done automatically in the script by minimizing the sum of squared errors between the simulated and original F_0 contours.

The Praat script was applied to all the non-reduced bi-tonal sequences to extract target parameters (*m*, *b* and λ) from each syllable in a disyllabic word. Table 1 lists the extracted target parameters and the RMSE and correlation values in comparison with the original F_0 contours (averaged across all the subjects).

Table 1: Canonical qTA target parameters *m*, *b* and λ, extracted from all non-reduced bi-tonal sequences, together with mean duration and local RMSE for each interval, and overall correlation values. Note that each subject has their own canonical parameters and Table 1 presents the summary. The last column is the random tone dyads for simulation 3 detailed in Sec. 3.3.

1^{st} sylb.	m	b	λ	Duration (s)	RMSE	2^{nd} sylb.	m	b	λ	Duration (s)	RMSE	Cor.	Random
H	-4	4	41	0.185	0.21	**H**	2	3	30	0.187	0.30	0.94	RF
H	2	5	41	0.228	0.28	**R**	19	-2	37	0.214	0.45	0.98	FH
H	3	5	42	0.226	0.30	**L**	-3	-5	35	0.201	0.72	0.97	LR
H	12	4	47	0.213	0.27	**F**	-44	0	45	0.200	0.30	0.98	RL
R	26	-1	24	0.223	0.27	**H**	10	4	47	0.199	0.36	0.98	LF
R	31	1	24	0.211	0.28	**R**	28	-1	33	0.202	0.39	0.96	FL
R	56	3	23	0.231	0.31	**L**	-27	-4	42	0.205	0.57	0.98	FR
R	42	-1	25	0.224	0.30	**F**	-40	2	56	0.209	0.39	0.98	HH
L	12	-9	25	0.223	0.42	**H**	15	1	64	0.212	0.47	0.98	LL
L	-10	-8	27	0.224	0.42	**R**	27	-3	50	0.208	0.57	0.96	HL
L	43	3	31	0.210	0.35	**L**	-25	-4	42	0.197	0.55	0.97	FF
L	14	-10	27	0.207	0.50	**F**	-22	-1	55	0.210	0.59	0.96	HR
F	-46	1	41	0.224	0.29	**H**	-3	2	50	0.202	0.35	0.98	RR
F	-65	0	38	0.230	0.31	**R**	6	-1	42	0.214	0.48	0.99	LH
F	-67	1	36	0.223	0.35	**L**	-5	-2	41	0.193	0.67	0.98	RH
F	-35	3	46	0.191	0.29	**F**	-23	0	52	0.193	0.35	0.97	HF

A mean RMSE of 0.32±0.14 semitone was obtained in the first syllable and of 0.47±0.22 semitone in the second syllable. The overall correlation is 0.97±0.25. The average RMSE of the two syllables and correlation values are shown as the leftmost points in Fig. 3. These low RMSE and high correlation values indicate that the qTA model could use the extracted target parameters to accurately *resynthesize* natural F_0 contours.

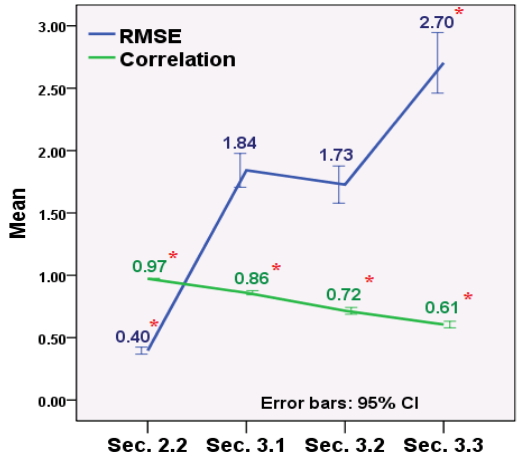
A logistic regression model was performed with tone type as the ordinal dependent variable, and the three target parameters as predictors. Results showed that all three target parameters can easily differentiate tone type (*b*: $\chi^2 = 13.80$, $p < 0.001$; *m*: $\chi^2 = 67.05$, $p < .0001$; *λ*: $\chi^2 = 4.01$, $p < 0.05$). Note that tone sandhi has been considered, i.e. in an LL sequence, the first L was assigned to R category in the statistical test.

3. HYPOTHESIS TESTING

3.1. Using canonical parameters to simulate non-reduced bi-tonal sequences

Each subject's respective canonical parameters were used to simulate F_0 curves of *individual* non-reduced bi-tonal sequences using another Praat script that performs qTA synthesis. Results indicate a high correlation of 0.86±0.12. RMSE values are also fairly low (syll. 1: 1.53±0.79, syll. 2: 2.15±1.00). Mean RMSE of the two syllables and correlation values are shown as the second points from left in Fig. 3.

Figure 3: Results of parameter training and three simulations. Blue line indicates mean RMSE values of the two syllables and green line indicates correlation values. On the x-axis, **Sec. 2.2**: parameter training; **Sec. 3.1**: non-reduced tonal simulation; **Sec. 3.2**: reduced tonal simulation; **Sec. 3.3**: reduced tonal simulation with random targets assignment.

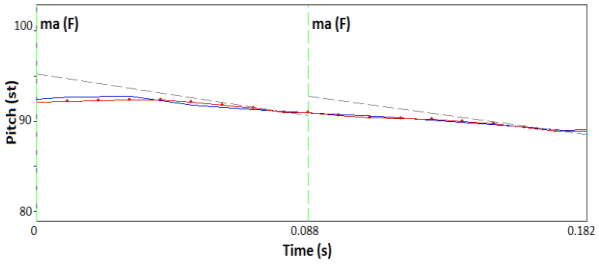


Compared to that of the resynthesis reported in Sec. 2.2, the decreased precision in the goodness of fit is expected, since here all the individual contours of a particular bi-tonal sequence are simulated with the same set of mean target values from respective subjects. The results are still relatively good compared to those of earlier studies [13].

3.2. Simulating F_0 contours of reduced tonal sequences with canonical parameters

In this simulation, we applied the canonical target parameters to the reduced tonal sequences. This took several steps. First, a mean ratio of the relative duration of the two syllables in a bi-tonal sequence was computed from the non-reduced tokens, which were averaged speaker by speaker. Second, each reduced bi-tonal sequence, which consisted of only a single interval due to the loss of the intervocalic consonant, was divided into two intervals, each having the same relative duration as the mean relative duration of the corresponding canonical sequence. In step 3, each subject's respective canonical parameters of non-reduced tone dyads were applied to each of the individual reduced tonal sequence, interval by interval. Fig. 4 shows an example simulation as displayed in the demo window of the Praat script.

Figure 4: An example of simulating a reduced bi-tonal sequence, using the canonical parameters of the tone sequence FF (preceded by a High tone, not shown here). Pitch targets (grey dashed line), synthesized F_0 (red dotted curve) against the original F_0 (blue curve).



The mean RMSE and correlation results are shown as the third points from left in Fig. 3. The evaluation of goodness of fit shows a correlation value of 0.72±0.19, RMSE of 1.23±0.64 (interval 1), and RMSE of 2.23±1.14 (interval 2). These values, as shown in Fig. 3, are close to those of non-reduced sequences in simulation 1 (Sec. 3.1). This seems to suggest that speakers may apply the same underlying tonal targets when time pressure is high.

3.3. Simply flattened? Random target application

The results of simulation 2 (Sec 3.2), despite seemingly agreeing well with our hypothesis, could be due to the fact that tones are simply flattened [4] such that all the tone dyads become similar to each other. To test this possibility, in this simulation the canonical targets are randomly assigned to the reduced tone sequences. The tone types for random pairings are shown in the last column of Table 1. If the reduced tone sequences are no longer related to the canonical sequences, this simulation would show little difference to simulation 2.

Compared to simulation 2, the correlation value decreased to 0.61±0.18, as shown in Fig. 3. RMSE increased to 2.17±1.37 for interval 1 and to 3.24±1.84 for interval 2. A multivariate analysis of variance (MANOVA) on correlation and RMSE values showed significant differences across the four sections (Correlation: $F_{(3,764)} = 233.2$, $p < .000$; RMSE: $F_{(3,764)} = 140.9$, $p < .000$). A post hoc analysis (Tukey HSD) on correlation further indicated that results from all sections differ significantly from each other. A post hoc analysis (Tukey HSD) on RMSE showed no significant difference between Sec. 3.1 and 3.2 ($p = .745$). Thus the result of simulation 3 (i.e. random target application with reduced duration) is shown to be statistically much worse than that of simulation 2 (i.e. matching target parameters with reduced duration). This confirms that the performance of simulation 2 is not due to simple F_0 flattening.

4. CONCLUSION

This paper presents a series of experimental simulations to demonstrate the proposed hypothesis that extreme tonal reduction stems from severe time pressure when the same underlying targets are attempted by the speaker. The quantitative target approximation (qTA) model was used to explore the underlying mechanism of tonal reduction in Taiwan Mandarin. Results show evidence for a direct link between duration and F_0 realisation. It further indicates that an articulatory-based model is capable of simulating tonal reduction under high time pressure [16].

5. REFERENCES

[1] Beckman, M.E., Pierrehumbert, J. 2003. Interpreting 'phonetic interpretation' over the lexicon. In Local, J., Ogden, R., Temple, R. (eds.), *Phonetic Interpretation*. Cambridge: Cambridge University Press, 13-37.

[2] Cheng, C.E. 2004. *An Acoustic Phonetic Analysis of Tone Contraction in Taiwan Mandarin.* MA, Taipei: National Cheng Chi University.

[3] Cheng, C., Xu, Y. 2009. Extreme reductions: Contraction of disyllables into monosyllables in Taiwan Mandarin. *Proc. Interspeech 2009* Brighton, 456-459.

[4] Cheng, C., Xu, Y., Gubian, M. 2010. Exploring the mechanism of tonal contraction in Taiwan Mandarin. *Proc. Interspeech 2010* Makuhari, 2010-2013.

[5] Flege, J.E. 1988. Effects of speaking rate on tongue position and velocity of movement in vowel production. *Journal of the Acoustical Society of America* 84, 901-916.

[6] Karlgren, H. 1962. Speech rate and information theory. In Sovijärvi, A., Aalto, P. (eds.), *Proc. of the Fourth International Congress of Phonetic Sciences*. The Hague: Mouton & Co, 671-677.

[7] Kohler, K.J. 1990. Segmental reduction in connected speech in German: phonological facts and phonetic explanations; In Hardcastle, W.J., Marchal, A. (eds.), *Speech Production and Speech Modelling.* Dordrecht: Kluwer Academic Publisher, 69-92.

[8] Lindblom, B. 1963. Spectrographic study of vowel reduction. *Journal of the Acoustical Society of America* 35, 1773-1781.

[9] Lindblom, B. 1964. A note on segment duration in Swedish polysyllables. *Speech Transmission Laboratory Quarterly Progress Status Report* 2, 1-5.

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

[11] Moon, S.J., Lindlbom, B. 1994. Interaction between duration, context and speaking style in English stressed vowels. *Journal of the Acoustical Society of America* 96, 40-55.

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

[13] Prom-on, S., Xu, Y., Thipakorn, B. 2009. Modelling tone and intonation in Mandarin and English as a process of target approximation. *Journal of the Acoustical Society of America* 125, 405-424.

[14] Tseng, S.C. 2005. Contracted syllables in Mandarin: Evidence from spontaneous conversations. *Language and Linguistics* 6, 153-180.

[15] Vatikiotis-Bateson, E., Kelso, J.A.S. 1993. Rhythm type and articulatory dynamics in English, French and Japanese, *Journal of Phonetics* 21, 231-265.

[16] Xu, Y., Prom-on, S. 2010. Articulatory-functional modelling of speech prosody: A review. *Proc. Interspeech 2010* Makuhari, 46-49.

[17] Xu, Y., Prom-on, S. 2010-2011. PENTAtrainer.praat. *http://www.phon.ucl.ac.uk/home/yi/PENTAtrainer/*

[18] Xu, Y., Wang, M. 2009. Organizing syllables into groups—Evidence from F_0 and duration patterns in Mandarin. *Journal of Phonetics* 37, 502-520.

[19] Xu, Y., Wang, Q.E. 2001. Pitch targets and their realization: Evidence from Mandarin Chinese. *Speech Communication* 33, 319-337.