Articulatory limit and extreme segmental reduction in Taiwan Mandarin[a)]

Chierh Cheng[b)] and Yi Xu
Department of Speech, Hearing and Phonetic Sciences, University College London, 2 Wakefield Street, London WC1N 1PF, United Kingdom

(Received 11 February 2013; revised 25 September 2013; accepted 30 September 2013)

The present study investigated whether extreme phonetic reduction could result from acute time pressure, i.e., when a segment is given less articulation time than its minimum duration, as defined by Klatt [(1973). J. Acoust. Soc. Am. **54**, 1102–1104]. Taiwan Mandarin was examined for its known high frequency of extreme reduction. Native speakers produced sentences containing nonsense disyllabic words with varying phonetic structures at different speech rates. High frequency words from spontaneous speech corpora were also examined for severe reduction. Results show that extreme reduction occurs frequently in nonsense words whenever local speech rate is roughly doubled from normal speech rate. The mean duration at which extreme reduction begins occurring is consistent with previously reported minimum segmental duration, maximum repetition rate and the rate of fast speech at which intelligibility is significantly reduced. Further examination of formant peak velocities as a function of formant displacement from both laboratory and corpus data shows that articulatory strength is not decreased during reduction. It is concluded that extreme reduction is not a feature unique only to high frequency words or casual speech, but a severe form of undershoot that occurs whenever time pressure is too great to allow the minimum execution of the required articulatory movement.
 [http://dx.doi.org/10.1121/1.4824930]

PACS number(s): 43.70.Fq, 43.70.Bk, 43.70.Jt, 43.70.Mn [SAF] Pages: 4481–4495

I. INTRODUCTION

To articulate a consonant or a vowel, a speaker has to move all the required articulators from their current position to the targeted position (Birkholz *et al.*, 2011; Lindblom, 1963; Perrier *et al.*, 1996; Saltzman and Munhall, 1989), and this process takes time. As postulated by Klatt (1973, p. 1103), each segment has a minimum duration D_{min}, which is the "minimum time of execution of the required articulatory program." Minimum duration is related to, yet different from the more widely known notion of intrinsic or inherent duration, which usually refers to the typical or average duration of a segment (Klatt, 1975; Peterson and Lehiste, 1960; Siguard, 1973; van Santen, 1994; van Santen and Shih, 2000). According to the estimate of Klatt (1973), D_{min} is roughly 0.45 times the inherent duration. Additionally, through measurement of maximum repetition rate (MRR) of different meaningless syllables, Tiffany (1980) observed a highly rigid "barrier" in articulatory rate: about 13.5 phones/s. Similar observations have been summarized in the comprehensive review of Kent *et al.* (1987). Assuming minimum segmental durations do exist, an intriguing question can be asked: What if, for whatever reason, the speaker has to go beyond the duration barrier and gives a segment an articulation time which is shorter than its minimum duration? One prediction would be that the articulatory integrity of the segment is compromised, resulting in phonetic undershoot (Lindblom, 1963), which would affect the intelligibility of the speech. Adank and Janse (2009) found that in fast speech, where average syllable duration was reduced to 46.0% of normal speech, recognition accuracy was reduced by 15%. Krause and Braida (2002) found that when professional speakers were trained to speak clearly, their intelligibility improved in slow and normal speech, but not in fast speech, suggesting again that there is some kind of articulatory limit that cannot be easily overcome even when trying to speak clearly. What is not clear from the intelligibility studies, however, is the severity of the reductions in fast speech, and how directly they are related to speech rate. As pointed out by Tiffany (1980), speech rate measurements based on sentences or words could be partially attributable to cases of severe reductions in which entire segments are omitted. This may seem to be the case given that the mean speech rate at the fast speech condition of Adank and Janse (2009) is 10.2 syllables per second; three syllables/s faster than the highest MRR summarized in Kent *et al.* (1987) (7.0 syllables/s for /ta/). Thus further understanding of fast speech may require closer examinations of cases of segmental deletion, known as massive or extreme reductions.

Extreme reductions have been well documented for many languages, for example, Dutch—Schuppler *et al.* (2011); German—Kohler (1990); French—Adda-Decker *et al.* (2005); and English—Dalby (1986), Johnson (2004). In some cases the reduction is so severe that entire syllables are lost or merged into other syllables. Reduction of a sequence of two or more syllables into one is also pervasive in the Sinitic language family (Chung, 2006; Myers and Li, 2009; Tseng, 2005), and such reductions are often referred to as "syllable contraction." In Taiwan Mandarin,[1] for example, *wo zhi dao* [wo tʂɨ tau], "I know" can be reduced to *wo zhao*

[a)]Part of this research was presented at INTERSPEECH 2009, September 6–10, 2009, Brighton, UK, pp. 456–459.
[b)]Author to whom correspondence should be addressed. Electronic mail: chierh.cheng@gmail.com

[wo tʂau], and *wo bu zhi dao* [wo pu tʂɨ tau], "I do not know" can be reduced to *wo bao* [wo pwau]. What is still unclear, however, is the exact cause of extreme reductions. In spontaneous speech, where extreme reductions are typically found, it is difficult to separate the effect of time pressure due to fast speech rate from that due to functional load and word frequency (Aylett and Turk, 2006; Bybee, 2002; Myers and Li, 2009; Pluymaekers *et al.*, 2005). When a word is highly frequent (hence familiar to both speakers and listeners) and also used in highly predictable contexts (hence carrying low functional load), there is an incentive to reduce both its duration and articulatory effort (Lindblom, 1990). This could suggest that whenever reduction occurs, effort and duration are both reduced. However, there is some evidence of disassociation of time pressure and articulatory effort. Krause and Braida (2002) found that, for example, when trained to speak clearly, i.e., under an incentive to apply full articulatory effort, professional speakers made no genuine improvement in intelligibility by speaking clearly at a fast rate over speaking conversationally at a fast rate. In this case, it is likely that time pressure rather than weak articulatory effort is responsible for phonetic reductions resulting in the reduced intelligibility. However, as no phonetic details are examined in the study, it is not clear how severe the reductions actually were.

The present study is an investigation into the relation between phonetic reduction, time pressure, and articulatory effort. This is done by examining in Taiwan Mandarin (in which extreme reduction is known to occur frequently) whether nonsense words, i.e., words with the lowest familiarity and the highest information load, are also subject to extreme reduction at high speech rates. We will explore, in particular, two specific questions: (1) *Is there evidence of a duration barrier beyond which extreme reduction frequently occurs regardless of word frequency and information load?* (2) *If evidence of such a barrier exists, is there also evidence of a decrease in articulatory effort when the barrier is approached?*

The study consists of two parts. Part 1 involves two controlled experiments examining Taiwan Mandarin speakers' production of nonsense disyllabic words with varying phonetic structures at different speech rates. Experiment 1 tests whether reduction can be directly elicited by increasing speech rate, and experiment 2 explores whether articulatory effort is strengthened or weakened when extreme reductions occur. In part 2, three spontaneous speech corpora are examined in terms of extreme reduction in two sets of high frequency words. This is to test whether the evidence found in part 1 can be generalized to spontaneous speech.

II. EXPERIMENT 1—LABORATORY DATA

There are two objectives to this experiment. The first is to test the prediction that extreme reduction can occur to nonsense disyllabic words if speech rate is sufficiently high. The main strategy is quite straightforward: To ask subjects to vary their speech rate and see if and when reduction occurs. The second objective is to assess the minimum duration beyond which extreme reductions is regularly observed.

A. Stimuli

Thirty-two nonsense disyllabic sequences in Taiwan Mandarin,[2] shown in Table I, were designed as testing

TABLE I. Stimuli and carrier sentence used in experiment 1.

Intervocalic obstruction levels from low to high				
Disyllabic structure		Phonetic presentation and characters		
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	/ti/ + /au/ 滴凹	/tu/ + /ai/ 督哀
		/tu/ + /au/ 督凹		
Nasal consonant				
CV̲N + V		/tan/ + /i/ 單依	/tan/ + /u/ 單巫	
CV + N̲V		/ta/ + /ni/ 搭妮	/ta/ + /nu/ 搭奴	
Non-nasal consonant				
CV + C̲V	fricative	/ta/ + /ɕi/ 搭悉	/ta/ + /su/ 搭蘇	/ta/ + /sa/ 搭撒
where C is a	plosive	/ta/ + /ti/ 搭滴	/ta/ + /tu/ 搭督	/ta/ + /ta/ 搭搭
	plosive^{h}	/ta/ + /t^{h}i/ 搭踢	/ta/ + /t^{h}u/ 搭禿	/ta/ + /t^{h}a/ 搭他
	affricate	/ta/ + /tɕi/ 搭激	/ta/ + /tsu/ 搭租	/ta/ + /tsa/ 搭紮
	affricate^{h}	/ta/ + /tɕ^{h}i/ 搭戚	/ta/ + / ts^{h}u/ 搭粗	/ta/ + /ts^{h}a/ 搭擦
Nasal + non-nasal consonant				
CV̲N + C̲V		/ɕin/ + /ti/ 新滴	/sun/ + /ti/ 孫滴	/san/ + /ti/ 三滴
Carrier sentence				
Pinyin	"ni shuo de shi _____ shi ba! wo dangran bu chi _____ shala nazhong dongxi, yinwei wo zui bu xihuan ta jia chu de _____ shala."			
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!」			
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!"			

materials. Target sequences were divided into four groups with regard to level of obstruction by intervocalic consonant: (1) zero obstruction—CV + V; CV + VN; CV + VV, (2) nasal consonant—CVN + V; CV + NV, (3) non-nasal consonant—CV + CV, and (4) nasal + non-nasal consonant—CVN + CV, where C is a plosive (pl.), fricative (fr.) or affricate (af.). Other combinations with non-nasal consonants as codas (i.e., CVC + NV or CVC + CV) were not considered, due to their impossibility in Mandarin phonotactics. All intervocalic consonants had a similar place of articulation (alveolar) but different manners of articulation in order to focus on the effect of obstruction level. The vowels in these sequences were /i/, /a/, and /u/ in order to maximize variability in the amplitude of formant movement. Note that not all possible sequences of the selected vowels in all obstruction conditions were tested. Only in obstruction level 3, non-nasal consonant—CV + CV, a balanced vowel sequence, was used to examine articulatory demand and relative formant displacement. All disyllabic units had the high-level tone (Tone 1) in order to minimize potential tonal effects (Xu, 2001).

Time pressure was controlled in two ways. The first was the manipulation of durational variation related to position in sentence and phrase (Klatt, 1975; Xu and Wang, 2009). This was achieved by devising a carrier sentence consisting of three phrases, each having a slot for the same target sequence (see Table I). The first phrase consists of 8 underlying syllables, the second 13, and the third 15 underlying syllables, all with the disyllabic target words embedded. The second way of controlling time pressure was to elicit different speaking rates with direct instructions to the subjects (as detailed below).

B. Subjects and procedure

Six male Taiwan Mandarin speakers were recorded. They were aged between 21 and 28 and had no self-reported speech disorders or professional vocal training. The speakers were all postgraduate students studying in London whose prior education was in Taiwan. They had been in England for less than 2 years at the time of recording. Only male speakers were used because their formants are easier to track than those of female speakers. The recordings were conducted in an anechoic chamber at University College London. Speech was recorded with a Shure SM10A microphone placed approximately 30 cm from the subjects' mouth. The speech signals were recorded into a computer using the software package Adobe Audition v.1.5 with a sampling rate of 44.1 kHz. All stimuli were presented to the subject 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.

Subjects were instructed to articulate the material at three speaking rates, slow and clear as if reciting in class, in a natural manner as if conversing with a friend, and as fast as possible. No explicit instruction was given as to whether syllables can or should be contracted. During each trial the speaker read out the sentences at the three speeds in the above order. The exact speed of articulation was left to the subjects' discretion. (The mean speech rates of slow, natural and fast across the six subjects were 4.9, 6.8, and 9.3 underlying syllables per second, respectively.) To increase the size of the data sets, three randomized blocks of the above 32 sentence sequences were used. In total, the number of target sequences produced was 32 (sentence sequences) × 3 (positions in the carrier) × 3 (speech rates) × 6 (subjects) × 3 (repetitions) = 5184 tokens. Among these 5184 tokens, 31 (6%) were discarded from further analysis due to inadequate voice quality such as creaky voice or poor recording quality.

C. Labeling reduction levels

The segmental labeling and measurements were conducted in Praat (Boersma and Weenink, 2010). Figures 1–3 display example spectrograms of the tokens produced in experiment 1. It can be seen that as the duration of word/

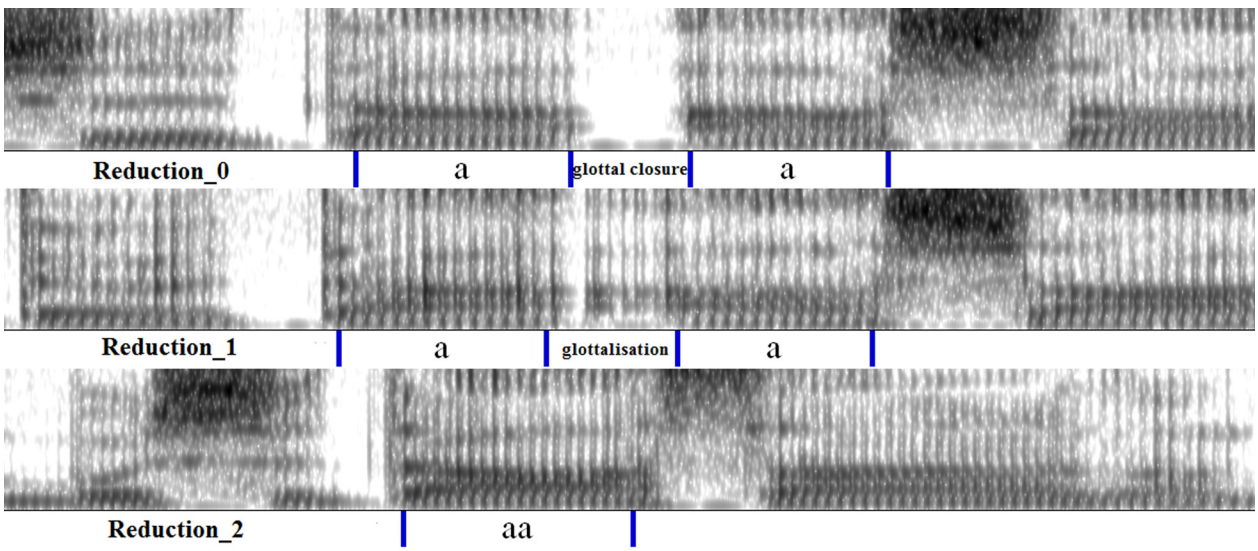

FIG. 1. (Color online) Examples of labeling for /ta/ + /a/ (zero intervocalic obstruction). From top to bottom: Reduction_0 (realized with a full glottal stop), reduction_1 (realized with glottalization), and reduction_2 (with continuous formants). The *x*-axes are over a similar time scale from 0 to 1 s and the *y*-axes have the same frequency scale from 0 to 5000 Hz.

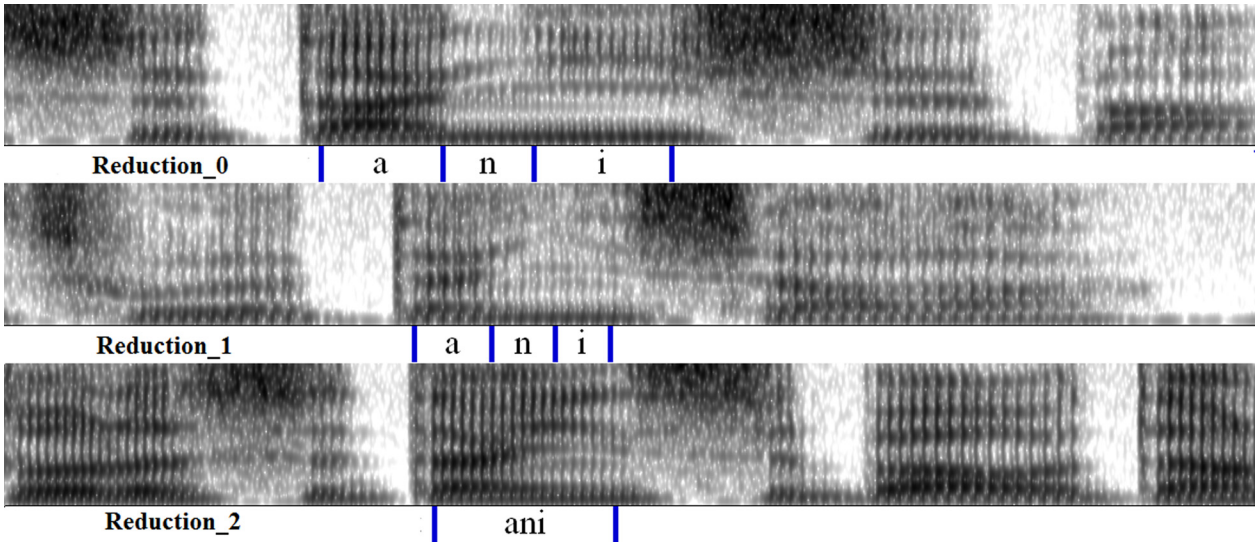

FIG. 2. (Color online) Examples of labeling for /tan/ + /i/ (intervocalic nasal consonant). From top to bottom: Reduction_0, reduction_1, and reduction_2. The *x*-axes are over a similar time scale from 0 to 1 s and the *y*-axes have the same frequency scale from 0 to 5000 Hz.

sequence decreases, the spectrographic patterns become increasingly simplified until little or no trace of the intervocalic consonant is left when duration is the shortest. We categorized the severity of reduction into three levels: "reduction_0," "reduction_1," and "reduction_2," according to their degree of intervocalic segmental weakening or loss (which roughly corresponds to the non-contracted, semi-contracted and contracted levels previously reported in the literature). Reduction_0 units were those with a clear interruption of formants by the intervocalic consonant, presence of nasal murmur or a clearly lowered F1. Reduction_2 units were those with continuous F1, without interruption by either intervocalic consonants or nasal murmur. Reduction_1 units were those for which the above segmentation criteria were difficult to apply and no straightforward delimitation of the spectrogram could be made.

All sound files were segmented and labeled by the first author, a native Taiwan Mandarin speaker. The consistency of labeling the reduction levels was double checked 1 month following the initial labeling. Uncertainty in the labeling occurred only very occasionally, and in all the cases the uncertainty was related to reduction_1. A handful of tokens were relabeled from reduction_0 or reduction_2 to reduction_1 upon rechecking. No tokens of reduction_0 were relabeled as reduction_2 or vice versa. Specific examples of the three degrees of reduction are shown in Figs. 1–3 for different intervocalic obstruction levels.

It is important to note that in the zero obstruction level, most tokens were marked as reduction_2 because the disyllabic sequence consisted of an open CV syllable followed by a syllable with a vowel onset. Hence, unless there was a clear sign of a glottal stop or glottalization between the two

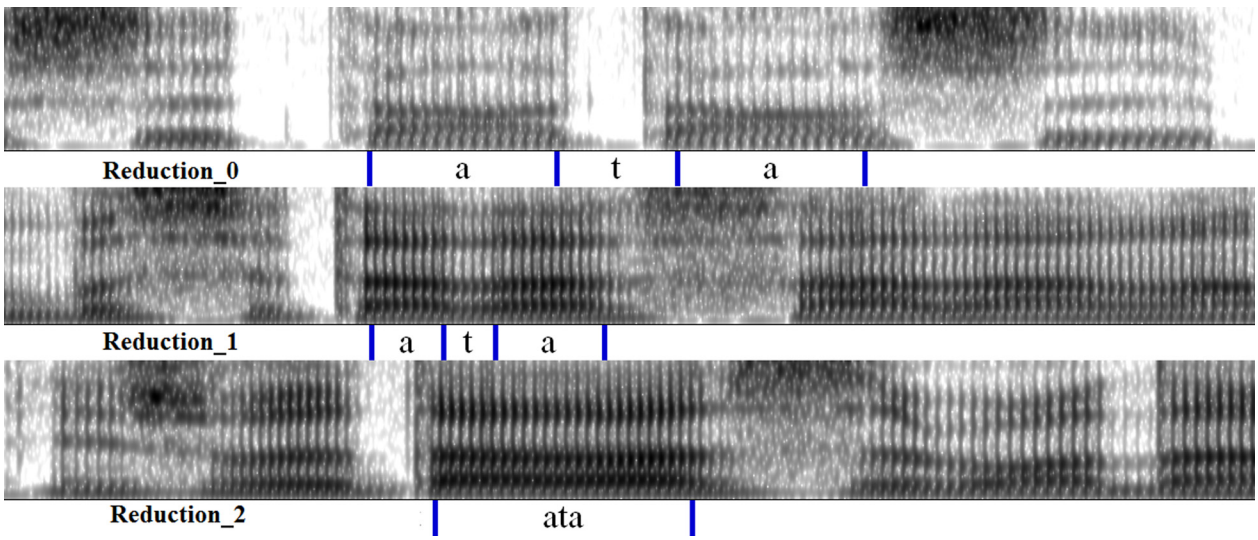

FIG. 3. (Color online) Examples of labeling for /ta/ + /ta/ (non-nasal intervocalic consonant). From top to bottom: Reduction_0, reduction_1 (with dipping F1, despite continuous formants), and reduction_2. The *x*-axes are over a similar time scale from 0 to 1 s and the *y*-axes have the same frequency scale from 0 to 5000 Hz.

vowels, as shown in Fig. 1, they were marked as reduction_2.

D. Results

From Figs. 1–3 we can see various effects of progressive reduction on the spectrographic integrity of the target sequences. In Fig. 2, for example, /n/ in /ani/ gets weakened in reduction_1 and virtually disappears as a separable segment in reduction_2. At the same time, there is severe undershoot of the vowel /a/, as its F1 and F2 become very similar to those of the surrounding consonants. The same is true for the sequence /ata/ in Fig. 3. Thus, there seem to be two different processes in this continuous reduction process: (a) disintegration of the intervocalic consonants and (b) undershoot of the flanking vowels. Experiment 1 is designed to mainly scrutinize process (a) by examining the conditions under which intervocalic consonants become severely reduced in a variety of VCV sequences, while process (b) will be more closely examined in experiment 2. More specifically, in experiment 1 we will test the prediction that extreme reduction can occur to nonsense disyllabic words if local speech rate is sufficiently high. As part of this test, we will assess the mean duration below which deletion of intervocalic segments is regularly observed.

Figure 4 displays the distribution of the three reduction levels. Reduction_0 and reduction_2 occurred more frequently (43.63% and 47.04%, respectively) than reduction_1 (9.33%), leading to a binomial distribution of reduction types.

A multinomial logistic regression is performed with reduction level as the ordinal dependent variable, and speed, obstruction level, and position in the carrier sentence as the predictors. Results show that speed is positively associated with reduction level (Coef. = 1.59, S.E. = 0.05, $p < 0.001$). For a unit increase in speed, the expected ordered log odds increases by 1.59 as one moves to the next reduction level (from reduction_0, reduction_1 to reduction_2). On the other hand, obstruction level is negatively related to reduction level (Coef. = −1.86, S.E. = 0.05, $p < 0.001$). For a unit increase in obstruction level, the expected ordered log odds decreases by 1.86 as one moves to the next reduction level. Position has no effect on reduction level (Coef. = 0.02, S.E. = 0.04, $p = 0.58$).

Figure 5 displays the relationship between reduction level and speed. The x axis shows three different reduction levels and the y axis shows relative frequency count of each reduction levels and speech rate. Within each reduction level, frequency counts for the three speeds are shown separately. For reduction_0, a decline in frequency count is seen as speed increases. Conversely, for both reduction_1 and reduction_2, as speed increases frequency count also increases. The two largest distributions are: slow speed in reduction_0 (23.79%) and fast speed in reduction_1 (20.98%). This is in accordance with the previous statistics: Significant positive relation between speed (from slow to fast) and reduction level (from reduction_0 to reduction_2). Note that 8.56% of the cases in reduction_2 occurred at slow speed. A major contributor here is the zero-obstruction group, which involves glides and vowels as syllable onset in the second syllable and was therefore often labeled as reduction_2. To test the contribution of the zero-obstruction cases in reduction_2, a follow-up logistic regression was conducted with zero-obstruction cases removed from all reduction levels. The variables used are the same as the above multinomial logistic regression. Results remained similar to those of the previous analysis, showing a positive relation between reduction level and speed (Coef. = 1.62, S.E. = 0.06, $p < 0.001$), a negative relation between reduction level and obstruction level (Coef. = −0.93, S.E. = 0.07, $p < 0.001$), and no influence of position on reduction level (Coef. = −0.02, S.E. = 0.05, $p = 0.71$).

Table II displays rates of extreme reduction in different phonetic structures in terms of the number of reduction_2 cases (reduction_1 items are excluded due to its ambiguous reduction status). The left column indicates the obstruction level of the intervocalic consonants from low to high. The middle column shows reduction rates with respect to each phonetic structure, and the rightmost column shows the overall mean reduction rates for different levels of obstruction. In

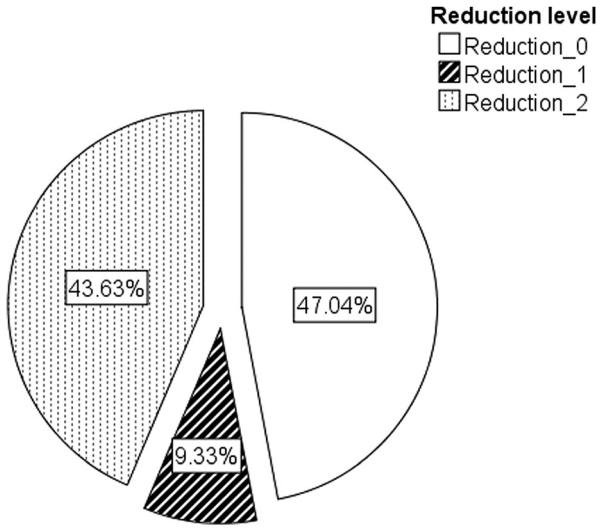

FIG. 4. Distribution of the three reduction levels of experiment 1.

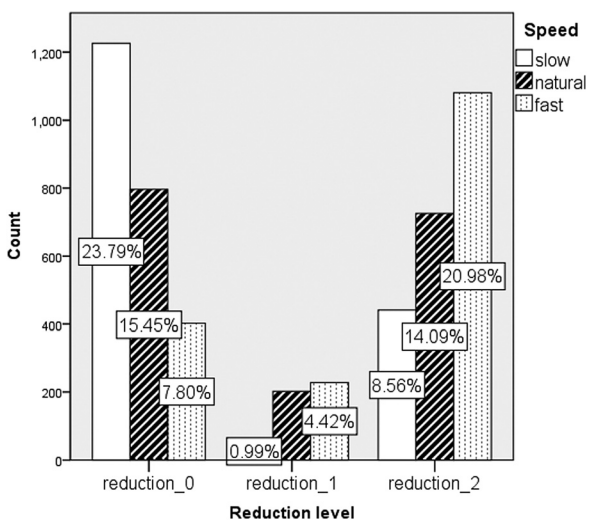

FIG. 5. Reduction rates at different speed.

TABLE II. Rate of extreme reduction (%) at different levels of intervocalic obstruction in experiment 1.

Intervocalic obstruction levels from low to high			
Disyllabic structure		Rate of extreme reduction (%)	Mean overall (%)
Zero obstruction			
CV + V		88.41%	89.77%
CV + VN		93.15%	
CV + VV		88.27%	
Nasal consonant			
CV<u>N</u> + V		62.35%	42.04%
CV + <u>N</u>V		21.67%	
Non-nasal consonant			
CV + <u>C</u>V where <u>C</u> is a	fricative	18.71%	19.80%
	plosive	20.87%	
	plosive^{h}	19.21%	
	affricate	23.97%	
	affricate^{h}	16.22%	
Nasal + non-nasal consonant			
CV<u>N</u> + <u>C</u>V		10.54%	10.54%

the zero-obstruction level, the reduction rate of nearly 90% was a natural consequence of the lack of canonical consonantal obstruction to interrupt the vowel-to-vowel formant movements as mentioned earlier. A mean reduction rate of 42.04% was seen in the nasal consonant level. It appeared that it was easier to lose coda nasals (62.35%) than initial nasals (21.67%) under time pressure. As for the non-nasal consonant level, reduction rate varied with the manner of consonant articulation. Unaspirated obstruents (plosive and affricate) had higher reduction rates (avg. 22.42%) than their aspirated counterparts (avg. 17.72%). In the nasal + non-nasal consonant level, the highest intervocalic obstruction yielded the lowest reduction rate (10.54%) among all of the phonetic structures included in this experiment.

Table II demonstrates that as the level of obstruction increases the rate of extreme reduction decreases. This inverse relation between obstruction level and reduction rate might indicate that reduction rate is determined by the level of articulatory demand, but time pressure is actually a more likely determining factor. As the CV*N* + *C*V group indicates, reduction rate is related to the time allocated to the consonant: When there are two adjacent consonants (of similar articulatory demands, i.e., /n/ and /t/), presumably twice as much time is allocated to the closing gesture. When duration is proportionally shortened under time pressure, these two consecutive obstruents are the last ones whose combined allocated time is reduced to the point when no closure of the vocal tract is possible.

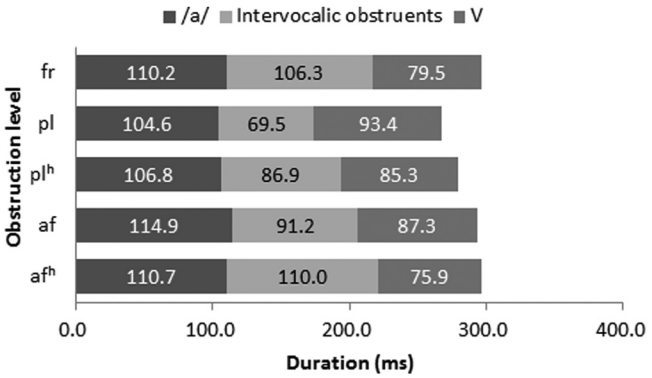

FIG. 6. Varying segmental durations (*x* axis) in /ta/ + CV units (*reduction_0*) with different levels of consonantal obstruction.

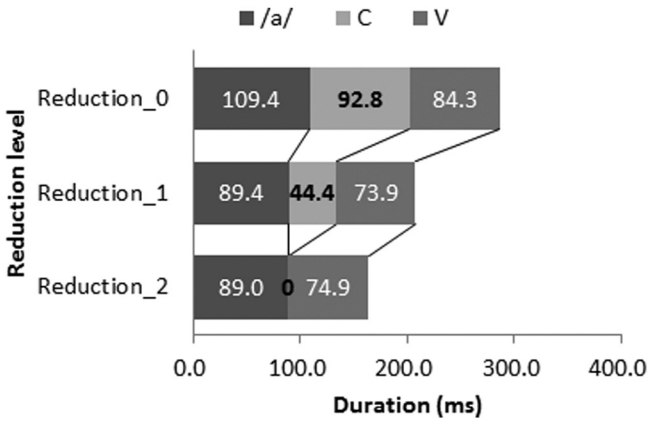

FIG. 7. Segmental durations at different reduction levels in *all* /ta/ + CV units.

This interpretation is further supported by Fig. 6 which shows mean segmental duration of /ta/ + CV units in the reduction_0 group according to the manner of articulation of the intervocalic consonant. The *x* axis displays consecutive durations of the preceding /a/, the intervocalic consonant and the following vowel. The *y* axis shows different levels of consonantal obstruction in the /ta/ + CV sequences. In the reduction_0 units the duration of the intervocalic consonants varies with their level of obstruction. Moreover, the duration of the second vowel varies compensatorily with the onset duration ($r = -0.97$, $p < 0.01$).

To see the effect of time pressure in a more straightforward manner, mean consecutive segmental durations of all stimuli in the /ta/ + CV sequences are plotted in Fig. 7 according to the degree of segmental reduction. Here the duration values at reduction_0 may reflect the amount of time used in canonical articulations. At reduction_1, the durations of all segments are reduced, with the most severe reduction in the intervocalic obstruents (44.4 ms). Finally, at reduction_2, the overall duration of disyllabic words is

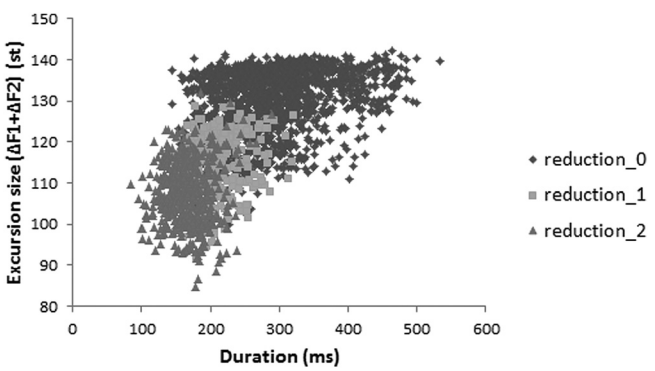

FIG. 8. Scatter plot of formant displacement (ΔF1 + ΔF2, *y* axis in semitone) over formant duration (*x* axis in ms): /ta/ + CV.

compressed to the point where no intervocalic consonantal closure is possible. Thus, a duration of 44.4 ms in the reduction_1 case seems to be the mean minimum duration below which intervocalic consonants are virtually "lost."

As the degree of vocal tract constriction reduces, the magnitude of formant displacement in the vowels is also reduced. It is therefore possible to further examine the relationship between reduction level and duration by observing the relationship between duration and formant displacement. Figure 8 displays a scatter plot of all tokens in the /ta/ + CV sequences. The *x* axis represents the combined duration of /a/ and the second syllable in /ta/ + CV sequences. The *y* axis represents the sum of F1 and F2 displacements within this interval. The three reduction levels are represented by different symbols. The scatter plot shows a positive relation between duration and formant displacement in /ta/ + CV sequences: The longer the duration, the larger the displacement ($R^2 = 0.41$, $p < 0.001$). The plot also shows an orderly distribution of the reduction levels as a function of duration: The reduction_0 cases all had relatively long durations; when the duration became extremely short—less than about 200 ms, extreme reduction occurred. It is interesting to note that the boundary between reduction_0 and reduction_2 was around the duration of 200 ms, suggesting that the mean duration for the integrity of disyllables was approximately 200 ms (without the duration of the onset consonant /t/ in /ta/ + CV).

It is also interesting that the reduction_0 data is largely horizontally distributed, indicating a ceiling in terms of maximum formant displacement which is not exceeded even when duration becomes very long. In contrast, reduction_2 data is largely vertically distributed, indicating a floor in terms of duration. Similar patterns are seen in previously reported articulatory data (Nelson *et al.*, 1984; Perkell *et al.*, 2002).

III. EXPERIMENT 2—LABORATORY DATA

This experiment is aimed at examining the process of continuous reduction in greater detail than in experiment 1 and testing if there is evidence suggesting whether articulatory effort is strengthened or weakened when reduction occurs. The strategy is to track the formant trajectories and velocity profiles of two quasi-symmetric articulatory movements to determine the relative contributions of duration and articulatory effort. The use of formant dynamics as a measure of articulatory effort, which is rarely done (but see Moon and Lindblom, 1994), needs some justifications.

There is often a concern that it is inappropriate to link formant movements to articulatory movements owing to the lack of a one-to-one relation between articulation and acoustics. However, it is also the case that individual articulators do not represent segmentally relevant articulatory movements as a whole. For example, for the vowel [i], the raising of the tongue blade against the hard palate to form a narrow constriction must be accompanied by the widening of the pharynx. In fact, it has been shown that F2 is more sensitive to pharyngeal width than to constriction at the tongue blade (Fant and Pauli, 1975; Wood, 1986). Thus F2 of [i] is reflective of at least two articulatory maneuvers: Tongue-blade raising and tongue-root fronting. Even the vertical position of the larynx differs across vowels in a manner that would enhance their formant differences (Demolin *et al.*, 2000; Hoole and Kroos, 1998; Wood, 1986). For consonants, the phonetically relevant articulation should also take aerodynamics into consideration. To produce [t], for example, not only the tip of the tongue should be raised against the alveolar ridge but also the sides of the tongue need to be elevated to form an airtight seal. Thus, the movement of any particular articulator is not for its own sake, but to serve as part of a collective maneuver to achieve the overall aerodynamic and acoustic effects that constitute the phonetic category (Mattingly, 1990; Hanson and Stevens, 2002). As a result, specific articulatory kinematics can provide only a partial approximation of the goal-oriented articulatory movement as a whole.

While acoustic measurements such as formant trajectories also provide only a partial approximation of the underlying goal-oriented movements, according to perturbation theory (Fant, 1960; Stevens, 2000), only the lower formants (up to F3) are individually controllable, since direct control of the higher formants would require simultaneous maneuvers of too many parts of the vocal tract (in order to constrict and widen it at all nodes and antinodes whose number correspond to the formant order, i.e., 1 pair for F1, and 2 pairs for F2, etc.). As a result, less critical information is missed if only the first few formants are measured. Interestingly, Hertrich and Ackermann (1997) and Perkell *et al.* (2002), after careful examinations of articulatory dynamics, both suggested that the phonetically most relevant information may be found in the acoustic signal. In fact, much of the argument of the H&H theory (Lindblom, 1990) was based on measured formant movements and their velocity in Moon and Lindblom (1994).

Furthermore, the comparability of articulatory and acoustic measurements can be empirically tested by examining whether acoustic and articulatory movements show similar dynamic patterns. At least for fundamental frequency, highly linear relations between F_0 velocity and F_0 movement amplitude have been found (Xu and Sun, 2002; Xu and Wang, 2009). These resemble the linear relations in articulatory or limb movement (Hertrich and Ackermann, 1997; Kelso *et al.*, 1985; Ostry *et al.*, 1983; Ostry and Munhall, 1985). This is despite the fact that F_0 is the output of a highly complex laryngeal system (Honda, 1995; Zemlin, 1988). It will therefore also be an empirical question as to whether formant kinematics also exhibits similar linear relations to warrant dynamic analyses that have been applied to limb and F_0 movements, which is addressed in this experiment.

A. Stimuli

Two disyllabic sequences, /ta/ + /ja/ and /ta/ + /wa/ with intervocalic glides /j/ and /w/, were devised to allow observation of formant trajectories without interruption by obstruent intervocalic consonants (see Table III). The use of glides can also avoid the issue of articulatory overlap between C and V because glides, owing to their characteristics as

TABLE III. Stimuli used in experiment 2.

Disyllabic structure	Phonetic presentation and characters	
CV + GV	/ta/ + /ja/ 搭壓	/ta/ + /wa/ 搭挖

semivowels, are specified for the entire shape of the vocal tract rather than predominantly at the place of articulation as obstruents. This will help make the interpretation of the relation between articulatory effort and duration in reduction more straightforward. The same carrier sentences as in experiment 1 were used.

B. Subjects and procedure

Four of the six subjects from experiment 1 were re-recruited to participate in this experiment. Two other male subjects with very similar linguistic backgrounds to those of experiment 1 were added. The same procedure as experiment 1 was followed. There were a total of 2 (stimuli) × 3 (positions in the carrier) × 3 (speech rates) × 6 (subjects) × 3 (repetitions) = 324 tokens.

C. Labeling and measurement

The segmental labeling was carried out in a similar manner to experiment 1, but with some modifications that are now detailed. When producing /ta/ + /ja/ and /ta/ + /wa/ within a carrier sentence, the speaker's vocal tract is never fully closed during the semivowels /j/ and /w/. Thus there would be few reduction_0 cases going by the previous criteria, i.e., those with a pause between the first and second vowels (as seen in Fig. 1 for the zero-obstruction level in experiment 1). Therefore, the labeling of reduction level in experiment 2 used F1 dip as a primary indicator and F2 peak or valley as a secondary indicator. In producing intervocalic glides /j/ and /w/ the articulators need to move to the position of the glide from the position of the preceding vowel /a/ and then to the position of the following vowel /a/. Since F1 was very low in both /j/ and /w/, cases with a fall followed by a rise in F1 were marked as reduction_0. In contrast, cases in which both formants (F1, F2) were flat, and no obvious curve could be seen, were marked as reduction_2. Units marked as reduction_1 were cases where the preceding two classifications were not straightforwardly applicable. Such cases commonly showed a slight F1 dip along with a slight F2 rise for /j/ and a slight F2 fall for /w/. Examples of labeling are shown in Figs. 9 and 10.

A set of kinematic measurements, including movement duration (time elapsed between two formant turning points), movement amplitude (difference in semitone between two turning points), and peak velocity (highest absolute value in the continuous velocity profile of the movement), were taken using a Praat script specifically written for the project (a custom version of FormantPro: Xu, 2007–2013). The script uses the Burg algorithm implemented in Praat to extract continuous formants, and applied a trimming algorithm (originally developed for processing F_0 contours, cf. Xu, 1999) to remove excessive and sudden bumps in the formant trajectories. It then computes the velocity (i.e., the first derivative) of the formants using a central differentiation algorithm. Figure 11 shows an example of an F1 movement of /ta/ + /ja/ and the corresponding velocity profile. The script first searched for the F1 minimum (point B in Fig. 11) in the formant track. It then finds peak velocities from within each of the two intervals (interval 1—from A to B; interval 2—from B to C). Similar procedures were applied to F2. In cases where formant trajectories either became effectively flat (as illustrated in Figs. 12 and 13) or took an incorrect direction from the canonical form, the kinematic measurements became erroneous. In those cases (146 out of 324 for F1 and 112 out of 324 for F2), the algorithm could not generate any valid measurements and they were automatically excluded by the script and not processed in further velocity analysis. The remaining tokens were then further analyzed to estimate the articulatory effort involved.

Note that this high failure rate is by design, as without it, we would not have been certain that speakers reached

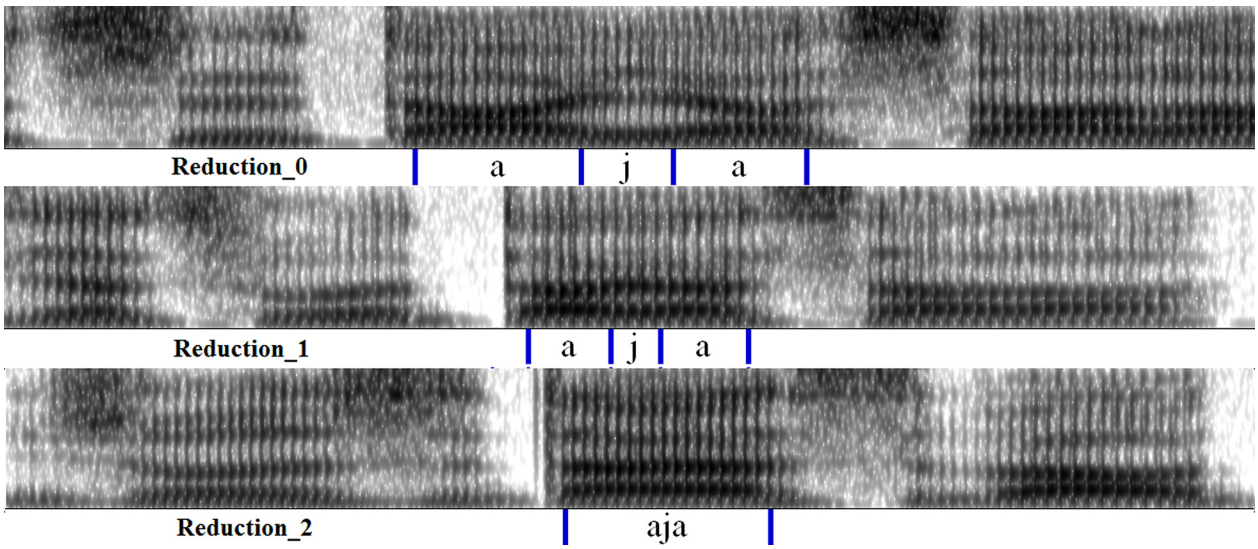

FIG. 9. (Color online) Examples of labeling for /ta/ + /ja/ (intervocalic glide /j/ in-between). From top to bottom: Reduction_0, reduction_1, and reduction_2. The x-axes are over a similar time scale from 0 to 1 s and the *y*-axes have the same frequency scale from 0 to 5000 Hz.

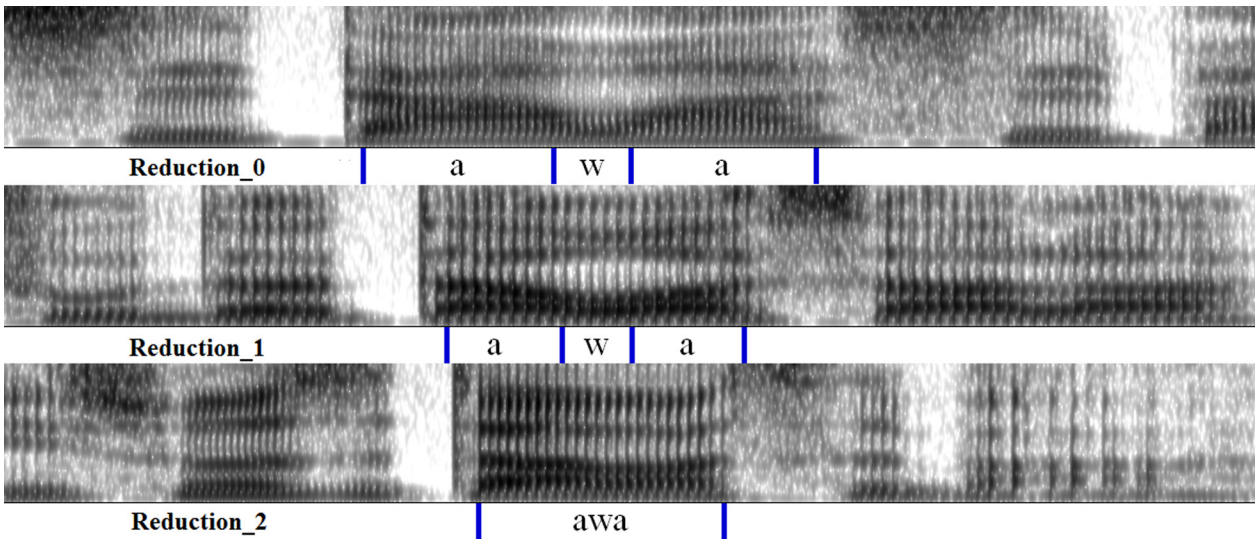

FIG. 10. (Color online) Examples of labeling for /ta/ + /wa/ (intervocalic glide /w/ in-between). From top to bottom: Reduction_0, reduction_1, and reduction_2. The *x*-axes are over a similar time scale from 0 to 1 s and the *y*-axes have the same frequency scale from 0 to 5000 Hz.

their articulatory limit. Extreme reduction, by its very nature, necessarily involves the destruction of the integrity of the underlying articulation. Thus there is an unavoidable trade-off between being able to simulate and systematically analyze extreme reduction in an experimental setting and not having to throw out a substantial amount of uninformative data (i.e., examining only cases well short of extreme reduction). Here we have given priority to the former in the interest of pushing the boundaries of our understanding of speech on this inherently difficult issue.

Figure 14 displays scatter plots of F1 peak velocity as a function of F1 displacement (movement amplitude) computed with data from all subjects at all three speech rates. The relationship between F1 peak velocity and F1 displacement was highly linear (R^2 ranging from 0.80 to 0.98). Such a linear relationship is consistent with previous findings regarding articulatory movements, and it has been considered to directly reflect the stiffness of the assumed mass-spring system during movement execution (Kelso *et al.*, 1985; Ostry *et al.*, 1983; Ostry and Munhall, 1985; Perkell *et al.*, 2002). On this basis, we also used the ratio of peak velocity and amplitude as a measure to access articulatory effort.

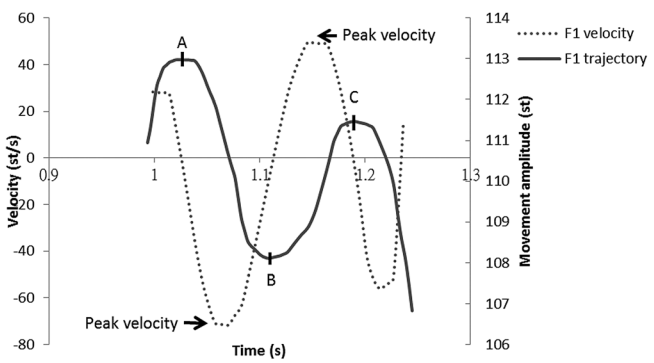

FIG. 11. F1 trajectory (solid line) and its velocity profile (dotted line): A, B and C mark the turning points of the F1 trajectory and delineate two intervals (A-B, and B-C). Two peak velocities were obtained, one within each interval as indicated in the figure. The *y* axis on the right hand side is in units of semitones for the F1 trajectory while the *y* axis on the left is in the units of semitone per second for the F1 velocity. The *x* axis represents the time domain in seconds.

D. Results

This experiment was intended to mainly address the second of the two questions raised in Sec. I, namely, if experiment 1 shows the existence of a minimum duration barrier, is articulatory strength decreased when this barrier is approached? In addition, the relation between formant displacement and duration is also further examined.

Figure 15 displays the distribution of the three reduction levels. Reduction levels 1 and 2 occurred with similar frequencies (27.27% and 19.48%, respectively) and reduction_0 occurred most frequently (53.25%).

A multinomial logistic regression was performed with reduction level as the ordinal dependent variable, and speed as well as position in the carrier sentence as the predictor variables. The results show that speed is positively associated to reduction level (Coef. = 2.51, S.E. = 0.23, $p < 0.001$). For a unit increase in speed, the expected ordered log odds increases by 2.51 in moving to the adjacent higher category of reduction (from reduction_0 to reduction_1, and

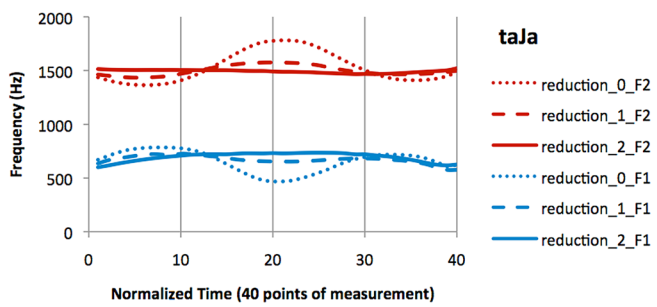

FIG. 12. (Color online) Formant trajectories (F1 and F2) of /ta/ + /ja/ sequences averaged across all six subjects: Dotted lines represent the F1 and F2 trajectories of reduction_0, dashed lines reduction_1 and solid lines reduction_2.

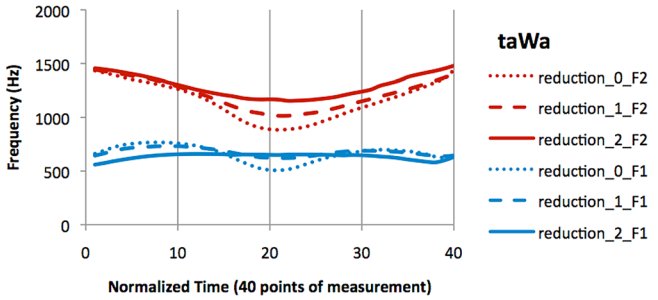

FIG. 13. (Color online) Formant trajectories (F1 and F2) of /ta/ + /wa/ sequences averaged across all six subjects: Dotted lines represent the F1 and F2 trajectories of reduction_0, dashed lines reduction_1 and solid lines reduction_2.

then to reduction_2). Position has no effect on reduction level (Coef. = −0.002, S.E. = 0.16, $p = 0.99$). Thus the OLR results of both experiment 1 and 2 show similar patterns despite the slightly different definitions used for the reduction levels.

Figure 16 shows a scatter plot displaying the relationship between duration and formant displacement for all tokens of /ta/ + /ja/ and /ta/ + /wa/. As in Fig. 8, Fig. 16 again shows a positive relation between duration and displacement ($R^2 = 0.67$, $p < 0.001$), indicating that the two measurements are strongly linked. A mean duration for the integrity of disyllalbes was also observed: reduction_1 units of /ta/ + /ja/ and /ta/ + /wa/ sequences cluster around 200 ms (again, without the duration of the onset consonant /t/ in /ta/ + GV).

The design of experiment 2 allowed us to further examine the relationship between duration, displacement and articulatory effort of the three reduction levels. Table IV

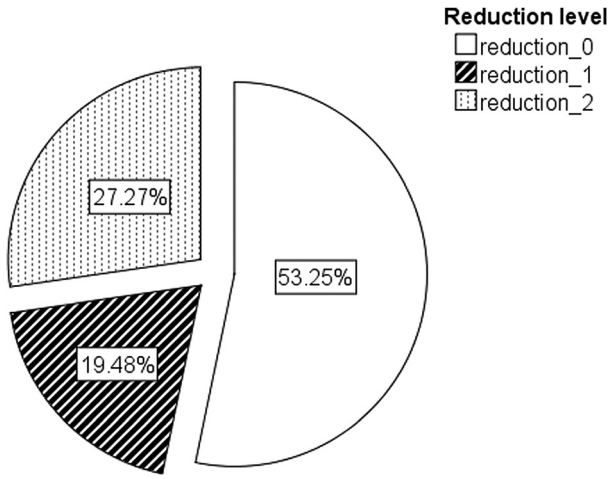

FIG. 15. Distribution of the three reduction levels of experiment 2.

shows a set of one-way analyses of variance (ANOVAs) performed with reduction level as the independent variable and duration, displacement and slope of regression line (peak velocity/displacement) as dependent variables. (In order to see the relative contribution of each formant trajectory, measurements of F1 and F2 are listed separately.)

The ANOVAs shows that reduction level has a significant effect on duration, F1 displacement and regression slope. A *post hoc* analysis of duration shows that all three reduction levels are significantly different from each other. (Duration: [reduction_0 > reduction_1], $Sig. = 0.0001$; [reduction_0 > reduction _2], $Sig. = 0.0001$; [reduction_1 > reduction_2], $Sig. = 0.008$). The *post hoc* analysis of formant displacement shows that F1

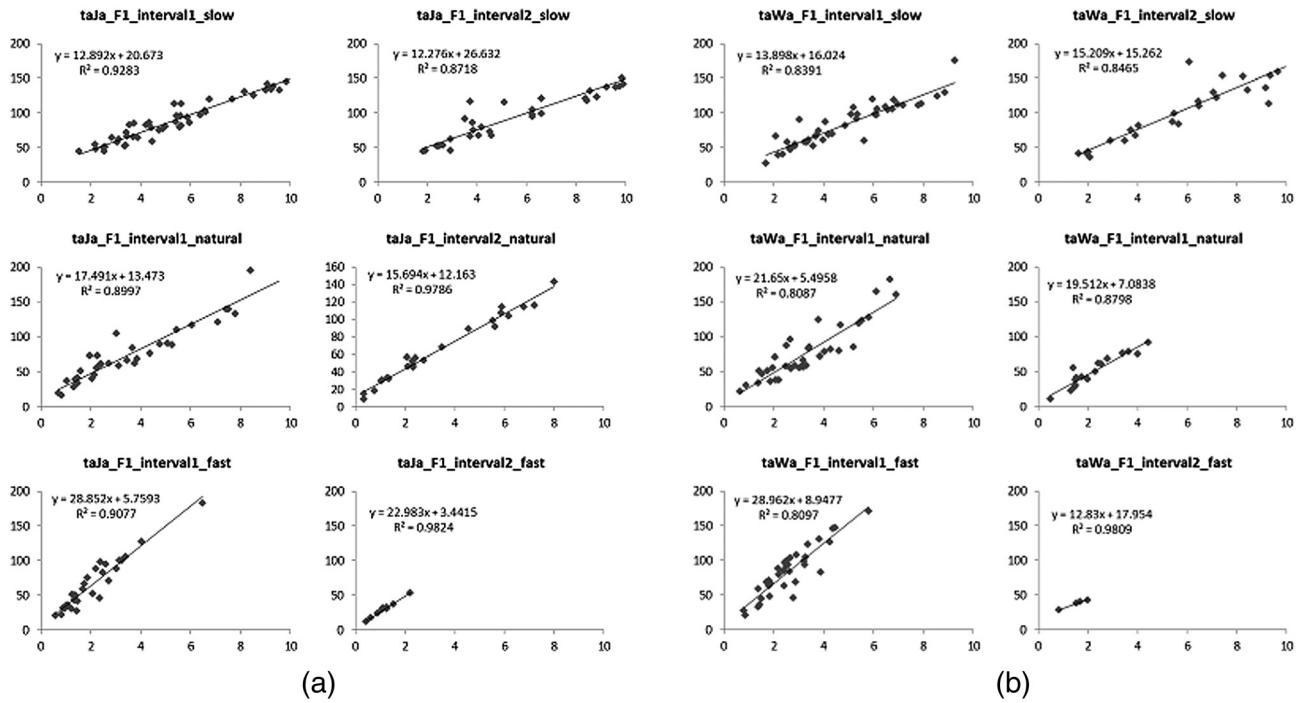

FIG. 14. (a) Linear relation of peak velocity (y axis in semitone/s) to displacement (x axis in semitone) across slow, natural, and fast speech rates: /ta/ + /ja/ sequences. (b) Linear relation of peak velocity (y axis in semitone/s) to displacement (x axis in semitone) across slow, natural and fast speech rates: /ta/ + /wa/ sequences.

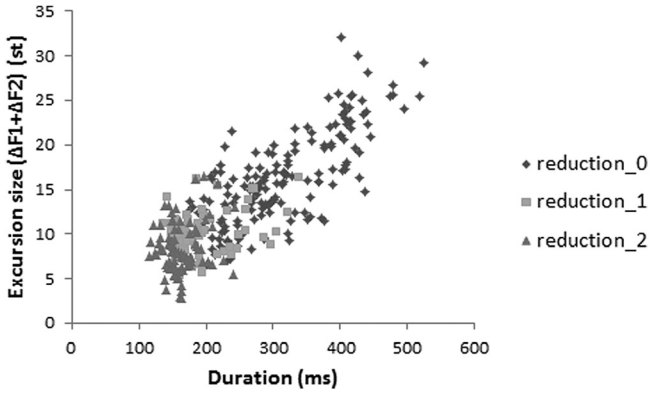

FIG. 16. Scatter plot of formant displacement (ΔF1 + ΔF2, *y* axis in semitone) over formant duration (*x* axis in ms): /ta/ + /ja/ and /ta/ + /wa/.

displacement of reduction_0 units is significantly larger than that of reduction_1 and reduction_2 but the difference is insignificant between reduction_1 and reduction_2 (F1 displacement: [reduction_0 > reduction_1], $Sig. = 0.008$; [reduction_0 > reduction_2], $Sig. = 0.006$). The regression slopes vary greatly across the different reduction levels for F1 ([reduction_0 < reduction_1], $Sig. = 0.003$; [reduction_0 < reduction_2], $Sig. = 0.0001$; [reduction_1 < reduction_2], $Sig. = 0.003$) but the same form of variation is not seen for F2.

The above results show that for F1 the displacements of both reduction_1 and reduction_2 are smaller than that of reduction_0. However, the regression slope is actually greater in reduction_2 than in reduction_0. For F2, although the differences in displacement and slope are in the same direction as F1, no difference is statistically significant across the reduction levels. To further confirm that these continuous variables can be used in predicting the reduction level, a follow-up logistic regression was conducted with reduction level as the ordinal dependent variable (from reduction_0, reduction_1 to reduction_2) and duration and regression slope as the continuous predictors (F1 and F2 displacements were not considered as they were used as reference in labeling reduction levels as mentioned in Sec. III C). Results remained similar to those of the ANOVA analyses, showing a negative relation between reduction level and duration (Coef. = −33.2, S.E. = 2.5, $p < 0.001$), and a positive relation between reduction level and regression slope (Coef. = 0.05, S.E. = 0.01, $p < 0.001$).

IV. EXPERIMENT 3—CORPUS DATA

This experiment is to explore if similar reduction patterns as those found in laboratory data (experiments 1 and 2) can be also observed in spontaneous speech. The spontaneous speech materials were taken from three corpora. A brief summary of the corpora is given in Table V, adapted from Tseng (2008, p. 3, Table I).

From the three corpora, two sets of words were selected for analysis (Table VI), each containing one stem form and one compound form. The stem syllables are [tʃʰɤ jaŋ] and [na jaŋ], in which all the syllables have the falling tone. The compound form is an added suffix [tsɨ] which has the neutral tone. These tokens were selected because, first, they are of similar phonetic structure to the stimuli used in experiment 2 (VGV), and second, both the citation and reduced forms of the four words are frequent in the corpora, giving enough tokens to allow reliable comparisons. They are also high in lexical frequency (zheyang, zheyangzi and nayang rank 48th, 66th, and 547th out of the 11 728th places in the Sinica corpus, respectively. No ranking information for nayangzi is available, but it is intuitively highly frequent to native speakers).[3] A total of 262 tokens from 17 speakers (eight males and nine females) were extracted. These speakers were in their twenties and had similar language backgrounds to those used in experiments 1 and 2.

All tokens extracted from the corpora were again labeled according to their reduction degrees using the same criteria as in experiment 2. For each token, only the segments exhibiting relevant formant trajectories (for calculating articulatory effort, marked as underscored in Table VI) were subjected to further velocity analysis. Three kinematic measurements (movement duration, movement amplitude, and peak velocity), similar to experiment 2, were obtained. As in experiment 2, erroneous measurements due to flattened formant trajectories were excluded from the velocity analysis (155 out of 262 for F1, and 136 out 262 for F2). Note that these corpus tokens are not as symmetrical as those used in experiment 2 in terms of their articulatory movements (i.e., they do not have the same vowels in both syllables), leading to a reduced success rate in generating valid measurements (from 55% for F1 and 65% for F2 in experiment 2 to 41% for F1 and 48% for F2 in the corpus data).

A. Data analysis

As with Table IV for experiment 2, a set of one-way ANOVAs were carried out with reduction level as the independent variable and duration, displacement and slope of regression line (peak velocity/displacement) as dependent variables. The results are shown in Table VII together with mean duration, displacement, and regression slopes of F1 and F2 for the three reduction types in the corpus data.

TABLE IV. Mean duration, displacement (F1, F2) and regression slope of peak formant velocity over formant displacement (F1, F2) for the three reduction levels of experiment 2.

Reduction level	Duration (ms)	F1 displacement (st)	F2 displacement (st)	F1 slope	F2 slope
Reduction_0	313.8	20.31	15.21	21.28	19.00
Reduction_1	199.0	8.86	8.82	26.95	23.95
Reduction_2	160.7	8.14	6.40	32.82	27.25
F value	$F_{(2,3)} = 360.68$	$F_{(2,3)} = 29.71$	$F_{(2,3)} = 1.93$	$F_{(2,3)} = 167.68$	$F_{(2,3)} = 0.31$
p value	$p < 0.001$	$p < 0.05$	$p = 0.289$	$p < 0.001$	$p = 0.756$

TABLE V. Summary of spontaneous Taiwan Mandarin speech corpora.[a]

Corpus	Mandarin Conversational Dialogue Corpus (MCDC)	Mandarin Topic-Oriented Conversation Corpus (MTCC)	Mandarin Map Task Corpus (MMTC)
Scenario	Free conversation between strangers	Subjects knew each other well	Subjects knew each other well
Purpose	Disfluency	Dialogue acts	Phonetic variations
Period	2001.03–2001.07	2002.01–2002.03	2002.01–2002.03
Transcription	Orthographically transcribed and annotated		

[a]Tseng (2008)

Only the difference in duration was significant, showing a progressive decrease from reduction_2 to reduction_0 units. A *post hoc* analysis of duration shows that reduction_0 is significantly different from both reduction_1 and reduction_2 (Duration: [reduction_0 > reduction_1], $Sig. = 0.006$; [reduction_0 > reduction_2], $Sig. = 0.002$). Comparisons of slope and displacement indicate that the F1 slope of reduction_2 is significantly steeper than that of reduction_0 (F1 slope: [reduction_0 < reduction_2], $Sig. = 0.043$). F2 slope and displacement show no significant differences.[4]

Comparing to the laboratory data, similar patterns can be seen (cf. Tables IV and VII and their respective *post hoc* results). Statistical tests on both laboratory and corpus data indicate that duration is a consistent and reliable predictor of extreme reductions. Furthermore, despite the differences in phonetic structure between the stimuli in experiment 2 and the corpus data, i.e., [ta ja] and [ta wa] versus [tʃʰɤ jaŋ] and [na jaŋ], the slope of the regression line for peak velocity over displacement is steeper in reduction_2 than in reduction_0 tokens.

V. DISCUSSION

In this study we explored two main questions related to extreme reductions: (1) *Is there evidence of a duration barrier beyond which extreme reduction frequently occurs regardless of word frequency and information load?* (2) *If evidence of such a barrier exists, is there also evidence of a decrease in articulatory effort when the barrier is approached?* Experiments 1 and 2 both show that intervocalic consonants of disyllable sequences are frequently lost when the measured VCV duration approaches 200 ms or shorter, thus indicating a positive answer to question 1. Additionally, a highly consistent relationship between reduction rate and speech rate is found, as shown in Fig. 5. Further analyses show that reduction rate is closely related to time pressure: The shorter the duration, the more likely extreme reduction occurs. In other words, a very short duration constitutes a sufficient condition for extreme reduction to occur. The positive relation between reduction and time pressure is seen even more clearly in Figs. 8 and 16.

Furthermore, the data found here are consistent with previous findings regarding minimum duration of segments and maximum repetition rate and intelligibility of fast speech. As shown in Fig. 7, the mean duration of intervocalic C in reduction_0 is 92.8 ms, while that in reduction_1 is 44.4 ms. Thus the ratio of canonical C duration to the duration beyond which C is disintegrated is $44.4/92.8 = 0.48$. This is similar to Klatt's (1973) estimate that D_{min} is roughly 0.45 times the inherent duration. It is also consistent with the finding of Adank and Janse (2009) that when the average syllable duration is reduced to 46.0% of normal speech, recognition accuracy is significantly reduced. All these studies seem to point to a *double-speed threshold*: When speech rate is doubled, whether locally or globally, some segments can no longer be articulatorily sustained, and are thus virtually lost. Moreover, the 200 ms duration threshold for the VCV sequence found in experiments 1 and 2 is consistent with the maximum repetition rate of 13.5 segments/s found by Tiffany (1980), which means a minimum duration of $1000/13.5 = 74.0$ ms for a single segment, and $3 \times 74 = 222$ ms for three segments. Note that in Tiffany's study, all the segments were required to be auditorily present, while the Reduction_1 level tokens in the present study as shown in Figs. 2 and 3 is already questionable by that standard, which may explain the small difference between Tiffany's 222 ms and the 200 ms seen here.

In regard to the second question: *Is articulatory effort decreased when the duration barrier is approached*, the results of experiments 2 and 3 indicate a negative answer. The data rather suggest that reduction is more likely accompanied by an *increase* in articulatory effort, because a steeper slope of the regression line between peak velocity and displacement was found for tokens labeled as reduction_2 compared to those labeled as reduction_1 and reduction_0. Such steeper slopes, according to Nelson (1983) and Perkell *et al.* (2002), are an indication of increased stiffness,

TABLE VI. Selected conventional reduction units in the corpus data.

Phonetic structure		Characters	Pinyin	Meaning	Count
1.	tʃʰɤ jaŋ	這樣	zhe yang	"this"	148
	tʃʰɤ jaŋ tsɨ	這樣子	zhe yang zi	"this way"	103
2.	na jaŋ	那樣	na yang	"that"	6
	na jaŋ tsɨ	那樣子	na yang zi	"that way"	5

TABLE VII. Mean duration, displacement (F1, F2) and regression slope of peak formant velocity over formant displacement (F1, F2) for the three reduction levels of the corpus data.

Reduction level	Duration (ms)	F1 displacement (st)	F2 displacement (st)	F1 regression slope	F2 regression slope
Reduction_0	359.1	31.15	15.32	20.56	23.04
Reduction_1	230.8	23.00	8.99	25.99	23.37
Reduction_2	221.2	23.22	11.27	33.44	25.58
F value	$F_{(2,8)} = 11.31$	$F_{(2,4)} = 1.45$	$F_{(2,1)} = 0.68$	$F_{(2,7)} = 3.03$	$F_{(2,7)} = 0.21$
p value	$p < 0.01$	$p = 0.336$	$p = 0.651$	$p = 0.113$	$p = 0.819$

and hence greater articulatory strength due to heightened muscle contraction. That articulatory strength seems to be increased when approaching minimum duration when producing nonsense words is consistent with the prediction of H&H theory that greater velocity may be applied to offset the effects of time pressure (Lindblom, 1990). But the finding that a clear duration dependency was observed in both clear and citation speech (Moon and Lindblom, 1994) suggests that the effect of the compensation is both small and orthogonal to the effect of time pressure. This means that Lindblom's (1963) earlier and simpler duration-dependent undershoot model is still applicable. What is shown more clearly by the present data than in previous studies is that when duration is severely shortened, e.g., by more than a half, there is simply no way for speakers to maintain the integrity of a segment or a syllable. Target undershoot under time pressure is thus *inevitable* when the time allocated to a segment is less than its minimum duration.

The present finding of the clear role time pressure plays in reduction does not, however, rule out the possible contribution of other, more widely known factors, which include lexical frequency (Bybee, 2002; Myers and Li, 2009), information load (Karlgren, 1961), listener considerations (Lindblom, 1990) and speech style (Dankovičová and Nolan, 1999; Moon and Lindblom, 1994). Time pressure is likely to be related to these factors in two different ways. First, all of these factors may directly affect duration, which in turn determines the level of time pressure on articulation. In this way time pressure is the *direct* cause of reduction. Support for this relation can be seen by comparing the high correlation between duration and formant displacement shown in Figs. 8 and 16 and the positive yet weak correlation between lexical frequency and spectral reduction reported by Myers and Li (2009, Figs. 3 and 5) for Taiwanese. Second, other factors may also be *orthogonally* related to time pressure, i.e., by affecting articulatory strength, as hypothesized by the H&H theory (Lindblom, 1990). In this way, different levels of reduction may be seen at a given duration, and the same level of reduction may be associated with different durations. This kind of orthogonal relation could probably explain why the mean duration of reduction_1 is 230.8 ms for the high-frequency words in the corpus data (Table VII) as opposed to the 199.0 ms for the nonsense words in experiment 2 (Table IV). Further evidence for the independent control of articulatory strength has been shown in tone studies. As proposed by Chen and Xu (2006) and successfully modeled in Prom-on *et al.* (2012), the severe undershoot as well as high context-sensitivity of the Mandarin neutral tone contributes to a lexically specified weak articulatory strength.

Thus, it is critical to identify, for each factor, its exact effect on reduction, i.e., whether it is through duration or strength, or both. In the case of duration, it is important to recognize that it is affected by many linguistic functions, including, in particular, phrase boundary, stress, within-syllable location, within-word location, within-phrase location, lexical tone, focus, and syllable structure (Berkovits, 1994; Dankovičová, 1997; Gahl and Garnsey, 2004; Klatt, 1975, 1976; van Santen, 1994; van Santen and Shih, 2000; Xu, 2009, to cite only a few). A case in point is the reduction of the Mandarin word *jiao ta che* [tɕiao ta̲ tʃʰɤ] "bicycle" as related to within-word location. Its second syllable is often reduced and the word becomes [tɕiao a̲ tʃʰɤ], despite the fact that it is a noun and is not of particularly high frequency (Chung, 2006). Chung (2006) points out that in general the second syllable is easily elided in tri- or tetra syllabic items. As found by Chen (2006), the middle syllables of a four-syllable word in Mandarin are drastically shortened. And Xu and Wang (2009) further demonstrated that tonal reduction in these syllables is directly attributable to shortened duration. These findings show that linguistically triggered local shortening may also lead to extreme reductions independent of speaking style and information load. In the case of articulatory strength, it is important to identify its effect independent of time pressure. With the exception of Moon and Lindblom (1994) most studies made no attempts to separate the two. The slope of regression between peak velocity and displacement used in the present study is only one of the methods that can be used to isolate the effect of articulatory strength. Other methods may be developed in the future.

Finally, neither time pressure nor articulatory strength could fully account for cases of fossilized lexical contraction such as "don't," "aren't," and "isn't" in English. In Mandarin *beng* [pɤŋ] as a contracted form of *bu yong* [pu jioŋ] is even written as a single character ("甭"), indicating that it is supposed to be spoken as a monosyllabic rather than disyllabic word ("不用"). These fossilized forms can remain monosyllabic even when spoken slowly and/or clearly, thus contrasting with the nonsense words of experiments 1 and 2 as well as the high-frequency words of experiment 3 (all of which show clear variability with duration). In future research there is a need to test each suspected case of fossilized reduction by directly controlling duration.

VI. CONCLUSIONS

To the best of our knowledge, no prior research has systematically examined variations in spectral patterns related to possible articulatory mechanisms underlying extreme reductions in any language. In the present study, cases of extreme reduction were successfully elicited from nonsense disyllabic words at high speech rates, and the rate of reduction as a function of speech rate was found to be similar to that of high-frequency words in a number of spontaneous speech corpora. This indicates that extreme reduction is not a characteristic unique to only casual speech or high word frequency. For both experimental and spontaneous data, spectral analyses show that severe reduction regularly occurs whenever segmental duration is shortened beyond a threshold consistent with previous findings about minimum duration, maximum repetition rate, and intelligibility of fast speech. Taking the average durations of the current data and those reported in Klatt (1973) and Adank and Janse (2009), the threshold for the disintegration of a consonant is around 0.46 of its canonical duration. Regression analysis for peak velocity of formant movement as a function of formant displacement suggests that articulatory effort is not weakened when reduction occurs. Overall, it seems that in speech, time is one of the most important commodities: Other things being equal, the more information is carried by a word, the more time is allocated to its articulation; those words that carry less information are allocated less time, often to the extent that some of its segments are allocated less than the minimum duration, resulting in their virtual elimination from the speech stream.

ACKNOWLEDGMENTS

We would like to thank Dr. Shu-Chuan Tseng and the institute of Linguistics, Academia Sinica Taipei, for providing the corpus data. Archives and linguistic representations of spoken Taiwan Mandarin can be found at http://mmc.sinica.edu.tw/sstm_e.htm (date last viewed September 18, 2013). The first author was supported in part by UCL Graduate School Research Projects Fund.

[1]Taiwan Mandarin here refers to the standard Mandarin natively spoken by people in Taiwan. Due to the constant influence of Southern Min, Taiwan Mandarin has developed its own stable linguistic system, which is distinct from the Mandarin spoken in Beijing.

[2]For readers less familiar with the Chinese language, how characters convey sounds and nonsense words is briefed here. Each Chinese character represents a syllable of sound (i.e., a combination of segments and tones) with a particular meaning or set of meanings. The correspondence between sound and character (i.e., symbol) is highly dependent upon semantic factors. For example, the pronunciation /an/ with a falling tone can be written as 暗, 岸, and 按, generically meaning "dark," "a shore," and "to press." Another Chinese character, 案, is of the same pronunciation and carries sets of meanings such as "a project," "a long table," or "a legal case," subject to different semantic factors. With such flexible connection to the sound of the syllable they present, disyllabic nonsense words used in our study were made in a fashion to meet as semantically unexpected as possible.

[3]Other references regarding Chinese word statistics such as frequency rank and cumulative percentage can be found at http://elearning.ling.sinica.edu.tw/eng_teaching.html (date last viewed September 18, 2013).

[4]A follow-up logistic regression was also conducted with reduction level as the ordinal dependent variable (from reudciotn_0, reduction_1 to reduction_2) and duration and regression slope as the continuous predictors. Results remained similar to those of the ANOVA analyses, showing a negative relation between reduction level and duration (Coef. $= -0.01$, S.E. $= 0.001$, $p < 0.001$) and that regression slope has no significant effect on reduction level (Coef. $= -0.01$, S.E. $= 0.02$, $p = 0.49$).

Adank, P., and Janse, E. (**2009**). "Perceptual learning of time-compressed and natural fast speech," J. Acoust. Soc. Am. **126**, 2649–2659.

Adda-Decker, M., Boula de Mareüil, P., Adda, G., and Lamel, L. (**2005**). "Investigating syllabic structures and their variation in spontaneous French," Speech Commun. **46**, 119–139.

Aylett, M., and Turk, A. (**2006**). "Language redundancy predicts syllabic duration and the spectral characteristics of vocalic syllable nuclei," J. Acoust. Soc. Am. **119**, 3048–3058.

Berkovits, R. (**1994**). "Durational effects in final lengthening, gapping, and contrastive stress," Lang. Speech **37**, 237–250.

Birkholz, P., Kroger, B. J., and Neuschaefer-Rube, C. (**2011**). "Model-based reproduction of articulatory trajectories for consonant-vowel sequences," IEEE Trans. Audio Speech Lang. Proc. **19**, 1422–1433.

Boersma, P., and Weenink, D. (**2010**). "Praat: Doing phonetics by computer," http://www.praat.org (Last viewed October 4, 2010).

Bybee, J. (**2002**). "Word frequency and context of use in the lexical diffusion of phonetically conditioned sound change," Lang. Var. Change **14**, 261–290.

Chen, Y. (**2006**). "Durational adjustment under corrective focus in Standard Chinese," J. Phonetics **34**, 176–201.

Chen, Y., and Xu, Y. (**2006**). "Production of weak elements in speech—Evidence from F_0 patterns of neutral tone in Standard Chinese," Phonetica **63**, 47–75.

Chung, K. S. (**2006**). "Contraction and backgrounding in Taiwan Mandarin," Concentric: Studies Linguist. **32**, 69–88.

Dalby, J. M. (**1986**). *Phonetic Structure of Fast Speech in American English* (Indiana University Linguistics Club, Bloomington, IN), 85 pp.

Dankovičcová, J. (**1997**). "The domain of articulation rate variation in Czech," J. Phonetics **25**, 287–312.

Dankovičcová, J., and Nolan, F. (**1999**). "Some acoustic effects of speaking style on utterances for automatic speaker verification," J. Int. Phonetics Assoc. **29**, 115–128.

Demolin, D., Metens, T., and Soquet, A. (**2000**). "Real time MRI and articulatory coordinations in vowels," in *5th Speech Production Sem.* (Munich, Germany), pp. 86–93.

Fant, G. (**1960**). *Acoustic Theory of Speech Production* (Mouton de Gruyter, The Hague, The Netherlands), 328 pp.

Fant, G., and Pauli, S. (**1975**). "Spatial characteristics of vocal tract resonance modes," in *Proc. Speech Comm. Sem.* (Stockholm, Sweden), pp. 121–132.

Gahl, S., and Garnsey, S. M. (**2004**). "Knowledge of grammar, knowledge of usage: Syntactic probabilities affect pronunciation variation," Language **80**, 748–775.

Hanson, H. M., and Stevens, K. N. (**2002**). "A quasiarticulatory approach to controlling acoustic source parameters in a Klatt-type formant synthesizer using HLsyn," J. Acoust. Soc. Am. **112**, 1158–1182.

Hertrich, I., and Ackermann, H. (**1997**). "Articulatory control of phonological vowel length contrasts: Kinematic analysis of labial gestures," J. Acoust. Soc. Am. **102**, 523–536.

Honda, K. (**1995**). "Laryngeal and extra-laryngeal mechanisms of F_0 control," in *Producing Speech: Contemporary Issues: For Katherine Safford Harris*, edited by F. Bell-Berti, and L. J. Raphael (AIP, New York), pp. 215–232.

Hoole, P., and Kroos, C. (**1998**). "Control of larynx height in vowel production," in *5th ICSLP* (Sydney, Australia), pp. 531–534.

Johnson, K. (**2004**). "Massive reduction in conversational American English," in *Spontaneous speech: Data and Analysis. Proceedings of the 1st Session of the 10th International Symposium*, edited by K. Yoneyama and K. Maekawa (National International Institute for Japanese Language, Tokyo, Japan), pp. 29–54.

Karlgren, H. (**1961**). "Speech rate and information theory," in *Proc. 4th ICPhS*, pp. 671–677.

Kelso, J. A. S., Vatikiotis-Bateson, E., Saltzman, E. L., and Kay, B. (**1985**). "A qualitative dynamic analysis of reiterant speech production: Phase portraits, kinematics, and dynamic modeling," J. Acoust. Soc. Am. **77**, 266–280.

Kent, R. D., Kent, J. F., and Rosenbeck, J. C. (**1987**). "Maximum performance tests of speech production," J. Speech Hear Disord. **52**, 367–387.

Klatt, D. H. (**1973**). "Interaction between two factors that influence vowel duration," J. Acoust. Soc. Am. **54**, 1102–1104.
Klatt, D. H. (**1975**). "Vowel lengthening is syntactically determined in a connected discourse," J. Phonetics **3**, 129–140.
Klatt, D. H. (**1976**). "Linguistic uses of segmental duration in English: Acoustic and perceptual evidence," J. Acoust. Soc. Am. **59**, 1208–1221.
Kohler, K. J. (**1990**). "Segmental reduction in connected speech in German: phonological facts and phonetic explanations," in *Speech Production and Speech Modelling*, edited by W. J. Hardcastle and A. Marchal (Kluwer Academic Publishers, Dordrecht, The Netherlands), pp. 69–92.
Krause, J. C., and Braida, L. D. (**2002**). "Investigating alternative forms of clear speech: The effects of speaking rate and speaking mode on intelligibility," J. Acoust. Soc. Am. **112**, 2165–2172.
Lindblom, B. (**1963**). "Spectrographic study of vowel reduction," J. Acoust. Soc. Am. **35**, 1773–1781.
Lindblom, B. (**1990**). "Explaining phonetic variation: A sketch of the H&H theory," in *Speech Production and Speech Modeling*, edited by W. J. Hardcastle and A. Marchal (Kluwer Academic Publishers, Dordrecht, The Netherlands), pp. 413–415.
Mattingly, I. G. (**1990**). "The global character of phonetic gestures," J. Phonetics **18**, 445–452.
Moon, S.-J., and Lindblom, B. (**1994**). "Interaction between duration, context, and speaking style in English stressed vowels," J. Acoust. Soc. Am. **96**, 40–55.
Myers, J., and Li, Y. (**2009**). "Lexical frequency effects in Taiwan Southern Min syllable contraction," J. Phonetics **37**, 212–230.
Nelson, W. L. (**1983**). "Physical principles for economies of skilled movements," Biol. Cybern. **46**, 135–147.
Nelson, W. L., Perkell, J. S., and Westbury, J. R. (**1984**). "Mandible movements during increasingly rapid articulations of single syllables: Preliminary observations," J. Acoust. Soc. Am. **75**, 945–951.
Ostry, D., Keller, E., and Parush, A. (**1983**). "Similarities in the control of speech articulators and the limbs: Kinematics of tongue dorsum movement in speech," J. Exp. Psychol. **9**, 622–636.
Ostry, D. J., and Munhall, K. G. (**1985**). "Control of rate and duration of speech movements," J. Acoust. Soc. Am. **77**, 640–648.
Perkell, J. S., Zandipour, M., Matthies, M. L., and Lane, H. (**2002**). "Economy of effort in different speaking conditions. I. A preliminary study of intersubject differences and modeling issues," J. Acoust. Soc. Am. **112**, 1627–1641.
Perrier, P., Ostry, D. J., and Laboissière, R. (**1996**). "The equilibrium point hypothesis and its application to speech motor control," J. Speech Hear. Res. **39**, 365–378.
Peterson, G. E., and Lehiste, I. (**1960**). "Duration of syllable nuclei in English," J. Acoust. Soc. Am. **32**, 693–703.
Pluymaekers, M., Ernestus, M., and Baayen, R. H. (**2005**). "Lexical frequency and acoustic reduction in spoken Dutch," J. Acoust. Soc. Am. **118**, 2561–2569.
Prom-on, S., Liu, F., and Xu, Y. (**2012**). "Post-low bouncing in Mandarin Chinese: Acoustic analysis and computational modelling," J. Acoust. Soc. Am. **132**, 421–432.
Saltzman, E. L., and Munhall, K. G. (**1989**). "A dynamical approach to gestural patterning in speech production," Ecol. Psychol. **1**, 333–382.
Schuppler, B., Ernestus, M., Scharenborg, O., and Boves, L. (**2011**). "Acoustic reduction in conversational Dutch: A quantitative analysis based on automatically generated segmental transcriptions," J. Phonetics **39**, 96–109.
Siguard, B. (**1973**). "Maximum rate and minimum duration of repeated syllables," Lang. Speech **16**, 373–395.
Stevens, K. N. (**2000**). *Acoustic Phonetics* (The Massachusetts Institute of Technology Press, Cambridge, Massachusetts), 618 pp.
Tiffany, W. R. (**1980**). "The effects of syllable structure on diadochokinetic and reading rates," J. Speech Hear. Res. **23**, 894–908.
Tseng, S. C. (**2005**). "Contracted syllables in Mandarin: Evidence from spontaneous conversations," Lang. Linguist. **6**, 153–180.
Tseng, S. C. (**2008**). "Spoken corpora and analysis of natural speech," Taiwan J. Linguist. **6**, 1–25.
van Santen, J. P. H. (**1994**). "Assignment of segmental duration in text-to-speech synthesis," Comput. Speech Lang. **8**, 95–128.
van Santen, J. P. H., and Shih, C. (**2000**). "Suprasegmental and segmental timing models in Mandarin Chinese and American English," J. Acoust. Soc. Am. **107**, 1012–1026.
Wood, S. (**1986**). "The acoustical significance of tongue, lip, and larynx maneuvers in rounded palatal vowels," J. Acoust. Soc. Am. **80**, 391–401.
Xu, Y. (**1999**). "Effects of tone and focus on the formation and alignment of f_0 contours," J. Phonetics **27**, 55–105.
Xu, Y. (**2001**). "Sources of tonal variations in connected speech," J. Chinese Linguist. Mono. **17**, 1–31.
Xu, Y. (**2007–2013**). "FormantPro.praat. Available from: <http://www.phon.ucl.ac.uk/home/yi/FormantPro/>."
Xu, Y. (**2009**). "Timing and coordination in tone and intonation—An articulatory-functional perspective," Lingua **119**, 906–927.
Xu, Y., and Sun, X. (**2002**). "Maximum speed of pitch change and how it may relate to speech," J. Acoust. Soc. Am. **111**, 1399–1413.
Xu, Y., and Wang, M. (**2009**). "Organizing syllables into groups—Evidence from F_0 and duration patterns in Mandarin," J. Phonetics **37**, 502–520.
Zemlin, W. R. (**1988**). *Speech and Hearing Science: Anatomy and Physiology* (Prentice Hall, Englewood Cliffs, NJ), 610 pp.