Syllable Timing and Pausing: Evidence from Cantonese
Conrad Perry, Richard Kwok-Shing Wong and Stephen Matthews

Language and Speech 2009 52: 29
DOI: 10.1177/0023830908099882

The online version of this article can be found at:
http://las.sagepub.com/content/52/1/29

Published by:
SAGE
http://www.sagepublications.com

Additional services and information for Language and Speech can be found at:

Email Alerts: http://las.sagepub.com/cgi/alerts

Subscriptions: http://las.sagepub.com/subscriptions

Reprints: http://www.sagepub.com/journalsReprints.nav

Permissions: http://www.sagepub.com/journalsPermissions.nav

Citations: http://las.sagepub.com/content/52/1/29.refs.html
Syllable Timing and Pausing: Evidence from Cantonese

Conrad Perry¹, Richard Kwok-Shing Wong², Stephen Matthews³

¹ Swinburne University of Technology, Australia
² Hong Kong Institute of Education
³ The University of Hong Kong

Abstract
We examined the relationship between the acoustic duration of syllables and the silent pauses that follow them in Cantonese. The results showed that at major syntactic junctures, acoustic plus silent pause durations were quite similar for a number of different syllable types whose acoustic durations differed substantially. In addition, it appeared that CV: syllables, which had the longest acoustic duration of all syllable types that were examined, were also the least likely to have silent pauses after them. These results suggest that cross-language differences between the probability that silent pauses are used at major syntactic junctures might potentially be explained by the accuracy at which timing slots can be assigned for syllables, rather than more complex explanations that have been proposed.

1 Introduction
Languages differ extensively with respect to the segmental phonology of their syllables and the timing of those syllables in speech (e.g., Ramus, Nespor, & Mehler, 1999), with both the acoustic duration of syllables and silent pause duration being important (e.g., Black, Tosi, Singh, & Takefuta, 1966; Duanmu, 1996; Ferreira, 1993; Meyer, 1994; Price, Ostendorf, Shattuck-Hufnagel, & Fong, 1991). Interestingly, a compensatory relationship has been noted between syllable and silent pause duration in some languages (e.g., Ferreira, 1993; Meyer, 1994), where words with short acoustic durations tend to have longer silent pauses after them than words with long acoustic durations. Ferreira (1993) interpreted this compensatory relationship as evidence that

Acknowledgments: We would like to thank Marija Tabain for very helpful advice and discussion. We are also very grateful to two anonymous reviewers for very helpful suggestions. A full list of the stimuli with translations is available from conradperry@gmail.com.

Address for correspondence. Conrad Perry, Swinburne University of Technology, School of Life and Social Sciences (Psychology), Internal Mail H31, John Street, Hawthorn, 3122, Victoria, Australia; <conradperry@gmail.com>
the system that generates the timing for syllables is not sensitive to their segmental phonology (see also Levelt, 1989). In particular, she suggested that a prosodifier creates timing slots, and only after that is segmental phonology added to the slots. Because of this, when the acoustic duration of a syllable is intrinsically longer than normal due to properties such as having a long versus short vowel, it is likely to fill more of the timing slot than when a syllable has an intrinsically short duration, and hence a compensatory relationship is found.

These results suggesting that timing slots are calculated independently of segmental phonology did not hold for long. In a number of experiments run in Dutch, Meyer (1994) found that although a compensatory relationship appeared to hold, with silent pause durations being shorter after syllables with long acoustic durations and vice versa, it was not complete. The acoustic plus silent pause duration was slightly longer for words that had intrinsically long acoustic syllable durations than for words that did not. Meyer interpreted this to mean that if timing slots are generated, they must be at least partially based on segmental information.

Meyer (1994) also offered an alternative explanation of the results, appealing to the possibility of cyclic vowel production (e.g., Cummins & Port, 1998; Fowler, 1983; MacNeilage, 1998; Vousden, Brown, & Harley, 2000). The idea is that people’s articulatory motor systems can be understood as a system of oscillators operating in a cyclical rhythmic manner, which causes people to try to calculate the timing for stressed vowels in relatively evenly spaced rhythmic sequences. Each point in time calculated is known as a P-center (Fowler, 1979, 1983; Morton, Martin, & Frankish, 1976), and the segmental phonology of syllables is overlaid on these timing points, with the stressed vowel of each word tending to be aligned with a P-center. Because of this tendency to try and produce evenly timed vowels with words that differ in acoustic duration, a relationship between the duration of silent pauses and the acoustic durations of syllables is found, without any need to create timing slots. An additional prediction that such a system makes is that quantal effects, where the duration between vowels tends to be similar to a multiple of the hypothesized cycle length (see, e.g., Fant & Kruckenberg, 1996), may be found due to people waiting more than one cycle at major syntactic junctures.

2 Metrical timing and Mandarin Chinese

That there is a relationship between silent pause and acoustic duration is certainly an important finding, since it provides data that helps constrain various models of speech production (e.g., Leivelt, 1989). It therefore seems reasonable to consider whether the results of Ferreira (1993) and Meyer (1994) are generalizable across languages with highly different timing patterns, such as Mandarin Chinese. Mandarin is interesting because differences compared to other languages in the timing of syllables and the silent pauses between words at major syntactic junctures have been reported. Duanmu (1996), for instance, noted that when results from English and Mandarin studies where speech was examined in laboratory conditions (Shen, 1992; Streeter, 1978) are compared, it appears that cross-linguistically similar syllables in similar clause-final positions do not tend to be lengthened as much in Mandarin (around 20% for Mandarin Chinese versus 70% for English). Thus, Mandarin represents the
opposite end of the lengthening spectrum from English, at least when using data from the languages reported in Vaissière (1983) as a guideline. In addition, he noted that at certain comparable syntactic junctures, English speakers follow straight on (resulting in a no-break utterance at a major syntactic juncture) much more frequently than Mandarin speakers.

Duanmu (1996) accounted for the observed cross-language differences based on the notion of a prosodic foot and how it interacts with syntax. The basic idea he suggested is that in certain conditions, both Mandarin and English use a prosodic foot that takes two syllables. However, in Mandarin, the slots in a prosodic foot need to be filled with individual syllables or silent pauses, whereas in English, two slots may be filled by one syllable. He suggested that this is because English has a large and complex set of syllables whose acoustic durations vary more than Mandarin, and syllables can be “stretched” across two slots in some circumstances. Alternatively, the lesser variability in the timing durations of syllables in Mandarin is due to a phonological restriction on syllable lengthening, which stops syllables from crossing two slots in a prosodic foot without violating a syllabic constraint.

Figure 1
Two different hypotheses concerning why there are cross-language differences in the number of no-break utterances. The top picture illustrates the prediction from Duanmu (1996). The bottom picture illustrates a prediction based on the theory of Ferreira (1993).
Based on these observations, Duanmu (1996) suggested that when a word occurs at a major syntactic juncture, it typically occurs as the first syllable of a prosodic foot in both English and Mandarin due to its underlying word stress (this is disputed by Dell, 2004, with a reply from Duanmu, 2004). This means that the second slot of the prosodic foot is free. In Mandarin, this slot is normally filled by a silent pause, whereas in English, the syllable that occurs in the first position may lengthen to fill the second slot, thus eliminating any potential silent pause. Hence more no-break utterances occur in English than Chinese. This explanation appears in the top half of Figure 1.

It is possible to suggest an alternative explanation of the cross-language differences based on an examination of the data in Duanmu (1996) and a consideration of the theory proposed by Ferreira (1993). In particular, it appears that the variance in the acoustic duration of syllables in Mandarin is much less than in English (around half in Duanmu’s data), and the English syllables also appear somewhat longer. Thus, the variability in the acoustic durations of syllables is much greater in English than in Mandarin. This difference in variance across languages means that there may be a statistical tendency for English speakers to follow on across syntactic junctures more often than Mandarin speakers due to the amount of time assigned to each syllable slot and the probability that the acoustic duration of certain syllables may run close to the end of the slots assigned for them. Thus, this hypothesis differs from Duanmu’s in suggesting that differences should be found simply due to the amount of variance in the acoustic durations of syllables in different languages, rather than because of different constraints on how syllables integrate into higher-level prosodic constituents. This possibility is depicted in the bottom half of Figure 1.

It is possible to make a direct prediction as to the pattern of data that should be found based on the ideas of Duanmu (1996) to do with prosodic feet and restrictions on syllable lengthening. In particular, speakers of other languages that have a similar prosodic foot structure as Mandarin should also not commonly cross major intonational boundaries if the syllables in their languages also have a syllabic restriction such that syllables cannot be lengthened across two slots of a prosodic foot. An indirect measure of this restriction, as pointed out by Duanmu for Mandarin, is to show that the variance in acoustic durations is quite small compared to languages such as English, when comparing syllable durations at major syntactic junctures versus syllable durations not at major syntactic junctures. Thus, if syllables at major syntactic junctures in a given language are lengthened by a similar amount as in Mandarin (i.e., around 20%) versus English (i.e., around 70%), it would provide evidence that syllables in that language are generally not lengthened across more than one slot in a prosodic foot.

3 Metrical timing and Cantonese

Given that there appear to be meaningful cross-language differences in the way that clause final lengthening occurs, we chose to investigate this in Cantonese. There are a number of similarities and differences between Cantonese, Mandarin, and English. In particular, single syllables in Cantonese, like Mandarin, almost always map into single morphemes, and all syllables are stressed at an output level (there is also no neutral tone, unlike Mandarin). Thus, in Cantonese, unlike English, it is not typically obvious

Language and Speech
which syllables have word level stress. Unlike Mandarin, in Cantonese, there is quite a large amount of variability in the duration of syllables, with Bauer and Benedict (1997) reporting that syllables in neutral stress conditions may vary from 100–400 ms, which is quite similar to some reports about English (e.g., Greenberg, Carvey, Hitchcock, & Chang, 2003). However, as far as we are aware,\(^1\) at least for typical speech styles, for all word classes except particles (described below), this additional variance in neutral stress conditions compared to Mandarin is mainly due to variability across different syllable types, unlike English, where more variability occurs than in Mandarin not only across syllable types (e.g., compare *tip* versus *triumphs*), but also *within* the same type of syllable.

If we assume that the variation in the duration of Cantonese syllables in neutral stress conditions is mainly due to variation across (rather than within) syllable types, then it allows a good test of the hypothesis of Duanmu (1996) about why speakers of Mandarin tend to use silent pauses at major syntactic junctures more often than their English counterparts. His hypothesis predicts that Cantonese speakers should also generally use silent pauses at major syntactic junctures, as long as the prosodic foot structure is similar to Mandarin. This is because if Cantonese syllables are also not typically lengthened to the extent that English ones are, then single Cantonese syllables should also not be able to cross two syllable-slots created by a single prosodic foot.

There is good reason to believe that the prosodic foot, and properties thereof, is similar in Cantonese and Mandarin, at least according to the analysis offered for Mandarin by Duanmu (1999, 2002, 2004). In particular, Duanmu (1999) proposes a number of syntactic tests that examine how metrical patterns and syntax interact in Mandarin. According to Duanmu, the results of these tests support the contention that Mandarin has a disyllabic prosodic foot and that the underlying stress patterns of words can be determined. We note that, at least from a preliminary analysis, some of the tests that Duanmu (1999) proposes for Mandarin show a very similar pattern of results for Cantonese: these include restrictions on word length in verb–object phrases, restrictions on word length in modifier–noun compounds, and preferred synonym choice with the usage of “de” (“ge” in Cantonese).\(^2\) Hence, if one assumes that the interpretation Duanmu (1999) gives to his tests are reasonable for Mandarin, it also seems reasonable to assume that very similar phenomena, in terms of underlying word stress and prosodic foot usage, occur in Cantonese.

---

1 We have confirmed this via experiments similar to those reported below.

2 Cantonese examples we have examined that produce similar results to Mandarin include verb–object: [zung3zik6][liaaldeo2] [to plant flowers]; modifier–noun: [lou5fu2][maa5ai5][tiger][ant]; and “ge” usage restrictions: [waai6jan4] ge3 [hei1pin3] [bad-person]'s [cheating]. Note that we use the Linguistic Society of Hong Kong’s Romanization scheme for spoken Cantonese, which is known as Jyutping. The alphabetical letters represent the segmental phonology of the syllable and the number after them (1–6) represents the tone. The phonology and tone numbers should be taken as an indication only, as there is a good deal of variation amongst speakers in Hong Kong (Bauer & Benedict, 1997).
4. Cantonese particles

Most of the previous work about Cantonese was done on word classes not known for their timing variability. However, Cantonese has one commonly used word class, known as sentence(-final) particles, that typically occur at the end of major syntactic/semantic units, and whose timing can vary considerably. In particular, the same monosyllables can vary from less than 100 ms with a neutral intonation to almost 1 second when stressed (Chan, 1998; Matthews & Yip, 1994), and which syllables are lengthened and the extent to which they are lengthened is relatively idiomatic (e.g., Kwok, 1984). Thus some cannot be lengthened as much as others. This feature of Cantonese provides an interesting contrast to other types of syllables that Cantonese has and also an interesting contrast compared to other languages that do not have syllables with such properties.

5. Hypotheses

The differences in timing associated with different word classes in Cantonese allow us to examine the relationship between acoustic durations and silent pausing in situations where the amount of variance in syllable durations differs extensively (particles versus other word classes). This, and other differences between Cantonese, English, Dutch, and Mandarin, can be used to inform us about three hypotheses.

First, it is possible to see how Cantonese compares to English with respect to clause end timing. In particular, we want to see whether silent pauses between words still compensate for differences between the acoustic durations of syllables in a language with very different timing than English or Dutch (e.g., Cantonese has no perceivable word level stress and each syllable must use a lexical tone; see Bauer & Benedict, 1997; Flynn, 2003, for a description of some of the phonetic and phonological properties of Cantonese; and Gordon, 2004, for ways in which Cantonese syllables may be exceptional even compared to other tonal languages). If similar results are found, it would allow the generality of Ferreira’s (1993) results to be extended. The idea here is that it is not obvious that languages that are not prototypically stress timed and have no perceptually audible word stress should show a compensatory pattern between syllable and silent pause duration. This is especially so for certain theories of speech timing that suggest that timing may be based on the distance between stressed syllables (e.g., Fowler, 1978) or distances that include additional information categories (e.g., Fant & Kruckenber, 1996), as it is not clear how these principles would apply to languages like Cantonese.

Second, the large variation in the acoustic duration of particles allows us to test a prediction that can be made from slot-based timing mechanisms such as that of Ferreira (1993). Such a theory predicts that the duration of the slot created for an item of a given word class in a given prosodic condition will typically be at least as long as the most commonly occurring longest duration item in that class: otherwise, segmental phonology would often overlap into the next timing slot. This means that in Cantonese, the silent pause plus acoustic duration of clause end verbs, adjectives, and nouns will not be as long as the silent pause plus acoustic duration of particles.
Third, it should be possible to investigate the extent that no-break utterances occur at major syntactic junctures in Cantonese. This is of interest because it is possible to compare the effect of syllables that may potentially be very long and have large amounts of variance in their durations (i.e., particles) with the effect of syllables where there is not so much variance in their duration. By doing this, it should be possible to test two different hypotheses concerning cross-language differences in the number of times people use no-break utterances at major syntactic junctures. In particular, if (1) the number of no-break utterances at syntactic junctures is simply related to variance in syllable durations, the number of no-break utterances found after Cantonese particles should be much more than after other word classes; and if (2) the smaller number of no-break utterances at syntactic junctures in languages such as Mandarin compared to English is related to a restriction in the way syllables may behave with respect to prosodic feet, where they cannot fill two slots, then non-particle syllables in Cantonese that occur at major syntactic junctures should generally be followed by a silent pause.

6 Experiment 1

This experiment sought to examine to what extent Cantonese is similar to Mandarin, English, and Dutch, in terms of timing at major syntactic junctures, on words other than particles. For this purpose, we examined words (non-particles) at major syntactic junctures with contrastive and neutral stress in Cantonese. This was done in a similar but not identical way to the contrastive stress manipulation of Ferreira’s (1993) Experiment 4, where critical words were examined at major syntactic junctures with neutral and contrastive stress. We did this using short vowel CVC, long vowel CV:C, and long vowel CV: syllable groups. The coda consonant in the CVC and CV:C words was always an unreleased stop. The expected average acoustic durations for syllables of each type, based on Bauer and Benedict (1997), were for short vowel CVC syllables to be the shortest, then CV:C syllables and finally CV: syllables. We did not select CVC or CV:C with nasal final consonants, since, at least in Mandarin, it has been argued that CVC syllables with nasals might in fact only be CV: syllables (Duanmu, 2002) and we are also not able to predict a priori what the duration of such syllables is likely to be compared to other types of syllables as accurately (although differences do exist, Bauer & Benedict, 1997). All of the sentences followed the same basic pattern:

[Start of sentence][Clause end critical syllable in neutral stress], [rest of sentence]

[Start of sentence][Clause end critical syllable in contrastive stress], [rest of sentence]

An example sentence is as follows (SFP = sentence final particle; CV: zaa3 (炸; bomb); CV:C: zaat3 (繫; tie); zat6 (窒; to mock )):

我唔想俾你炸/繫/窒，你走開呀！
ngo5 m4 soeng2 bei2 nei5 zaa3/zaat3/zat6, nei5 zau2 hoi1 aa3!

I not want by you bomb/tie/mock, you run away SFP!

I don't want to be bombed/tied/mocked by you, go away!

If compensation occurs between silent pauses and the acoustic durations of syllables at major syntactic junctures, as reported in English and Dutch, we should find that the duration of the silent pause after the syllables is negatively correlated with the duration of the syllables. Thus CV: syllables should have shorter silent pauses after them than CV:C syllables, which in turn should have shorter silent pauses after them than CVC syllables. That should be true of both the neutral and contrastive stress conditions, even though the overall syllable plus silent pause duration in the contrastive stress group should be significantly longer.

Even with sentences matched the way we did, semantic differences between the sentences might still cause differences in the placement of intonational breaks (see e.g., Frazier, Clifton, & Carlson, 2004, for an example of this in speech perception). Therefore, to ensure that participants always knew where a syntactic break was supposed to occur, a comma was used to signal the break between the two major clauses in each sentence. The sentences were also constructed in such a way that the two major clauses separated by the comma corresponded to the two most easily divisible semantic parts of the sentence. This was done via a manipulation of the information type presented in the first and second clause of the sentences. In the first main clause, information indicating what the focus of the sentences was or simply a description of the speaker’s perception was presented. In the second main clause, a more diverse array of functions were used. These included being a comment on the first topic in the sentence and being an indication of the requested action, wishes, or emotional states of the speaker.

The stimuli differed in this study compared to Ferreira (1993) and Meyer (1994) in two important ways. First, we marked the major division where we wanted to examine the silent pause with a comma. This may affect the prosody people use, as the way people interpret prosody from written language is affected by punctuation (e.g., Steinhauer & Friederici, 2001). The comma should therefore make it less likely for people to use no-break utterances. Second, whereas Ferreira and Meyer asked participants to memorize a sentence and then speak the sentence out aloud, we asked participants to read the sentence to themselves first but did not remove the sentence. Leaving the sentence visible allows participants to read parts of the sentence that they have forgotten, and hence should reduce the error rate. It also reduces the memory load on subjects, and thus should encourage participants to use appropriate mid-utterance intonational breaks, even when this may have otherwise caused memory difficulties due to the extra time needed in production. The potential trade-off is that the utterances participants produce may sound more like reading than would otherwise have been the case.

There were three important differences between this study and that of Duanmu (1996). First, instead of having participants repeat a small number of sentences many times, we used a greater number of different sentences. This was intended to minimize...
effects of repeating stimuli, as we had noted from test studies that when our partici-
pants repeated stimuli, they also had a tendency to increase their speech rate on the
repeated item. Second, our sentences were longer, averaging around 13 syllables the
three (sentences Duanmu used were five, five, and nine syllables long). This means
our sentences would be likely to have more and longer intonational breaks within
them. Third, the syntactic junctures we examined were different in that they always
marked the division in a topic-comment style utterance. Again, this property should
make it less likely that our participants would use a no-break utterance.

In sum, differences in the methodology between this study and some of the
others mentioned should tend to lead to the minimization of no-break utterances
in our study. Thus, if this type of utterance is found, it suggests that it is also likely
to occur in less artificial Cantonese speech. Alternatively, if the silent pause at the
syntactic juncture is longer than in other tasks, it may be partially due to task specific
differences.

6.1 Participants
Twelve students from the University of Hong Kong served as participants in the study.
All were native speakers of Cantonese with English as their second language.

6.2 Stimuli
Fifteen carrier sentences were constructed such that there was a critical word that was
an adjective, verb, or noun at a major syntactic juncture that was not the end of the
sentence. The critical words were chosen in triplets for each sentence such that each
triplet had a word with a CVC, CV:C, and CV: structure. The CVC and CV:C words in
a triplet always had the same onset and coda consonant, and the onset consonant was
shared across all syllable types in a triplet (e.g., bat, baat, baa). In a small number of
the triplets, slightly different sentences were used when classifiers for nouns differed.
All sentences were written in colloquial Cantonese. The sentences were generally of
a topic-comment type, which is very typical of Cantonese speech (Matthews & Yip,
1994), and quite a diverse range of structures within such a description was used. A
further six practice stimuli of a similar grammatical type were also constructed.

6.3 Method
All of the sentences were printed on paper using 24-point font to indicate words with
contrastive stress and 12-point font to indicate words without such stress. Sentences
were counterbalanced across three groups, such that two critical syllable from each
triplet (e.g., baa, baat, bat) one with contrastive and the other with neutral stress,
appeared in each group. Thus, in each counterbalanced group, the same carrier
sentence was used twice, but with a different critical syllable. In addition, the two
occurrences of each carrier sentence in each group were arranged such that one
always fell in the first half of the list and the other in the second, and, across the
three groups, all critical syllable were used twice, once with contrastive and once
with neutral stress.
All participants were given the three lists of sentences (and hence saw each sentence six times, but with different words in different stress conditions in the critical position). The order in which the lists were given to participants was counterbalanced. In terms of individual items, participants were told to first read each sentence that appeared in the list silently to themselves, make sure they knew what the sentence was, and then to read the sentence out aloud in a way similar to that which they would hear in normal speech. They were told to read words printed in the larger font using contrastive stress, and were given a specific example of how this might be done verbally by the experimenter, who made sure they knew what contrastive stress meant. They were told not to worry if they saw a sentence more than once, but simply to say it out aloud again. In case they made an error, they were told to repeat the entire sentence.

6.4 Data treatment

The sound editing package Praat (Boersma, 2001) was used to analyze the stimuli. Syllable and silent pause durations were measured by hand (i.e., no automatic boundary finding algorithm was used). The onset of voiced stops was marked after the occlusion. When syllables had no silent pause between them, the duration was determined as the first point in time at which the onset of the second syllable could be identified. All word and silent pause durations above 1500 ms were removed from the analysis, as were incorrect syllable pronunciations. This led to 1.94% of the data being removed. For each participant, mean durations were calculated and all items that had durations 2.5 standard deviations (SDs) above or below them were removed from the analysis.

Table 1

Means (ms) and SDs of syllable length (acoustic duration), silent pause duration, syllable plus silent pause duration, and proportion of times there was a no-break utterance after the critical word in Experiment 1, as a function of syllable and stress type

<table>
<thead>
<tr>
<th>Stimuli type</th>
<th>Syllable length</th>
<th>Silent pause duration</th>
<th>Syllable plus silent pause duration</th>
<th>No-break utterance proportion</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
<td>M</td>
<td>SD</td>
</tr>
</tbody>
</table>

Neutral Stress:

CV: 320.9  57.8  283.7  280.3  604.8  281.8  .26 .44
CV:C 246.4  55.9  381.1  249.8  543.3  234.7  .14 .35
CVC 192.5  50.1  350.8  235.5  627.5  627.5  .17 .38

Contrastive Stress:

CV: 414.2  119.2  476.5  275.7  890.7  284.1  0.06 .23
CV:C 327.7  107.3  523.8  258.8  857.7  291.3  0.03 .18
CVC 254.8  95.9  563.3  256.8  824.1  284  0.03 .17
This was done for both acoustic and silent pause durations. This led to 0.93% of the data being removed. Means and SDs were derived within syllable groups separately (i.e., for CV:, CVC and CV:C groups separately), since we had an a-priori reason to believe that they would differ in timing (i.e., CV: syllables would be the longest and CVC syllables the shortest). In calculating whether a no-break utterance occurred, if the measurable amount of silent pause between the end of the syllable and the onset of the next was less than 100 ms, a no-break utterance was assumed to have occurred with CVC and CV:C syllables. With CV: syllables, a no-break utterance was assumed to occur only if there was no measurable silent pause between them and the onset of the next syllable. The average syllable-to-syllable duration of each sentence was also calculated by taking the total duration of each utterance, subtracting the duration of the silent pause in the middle, and then dividing that number by the number of syllables in each sentence. We will refer to this below as speech rate, although it is important to note that we removed the duration of the silent pause from this calculation. The results appear in Table 1.

6.5 Results

6.5.1 Syllable and silent pause durations

The data were analyzed using a 12 Participants × 3 Syllable type (CV:, CV:C, CVC) × 2 Stress Type (Contrastive/Neutral) between-groups Analysis of Variance (ANOVA) for both the syllable and syllable plus silent pause durations. Speech rate was used as a covariate. Sampling was considered random for participants and fixed for both syllable and stress type. We did not examine interactions between participant and stress type when examining only syllable durations since it was not possible to control for the level of contrastive stress used by our participants (i.e., interactions may occur simply because some participants used a much greater level of contrastive stress than others).

In terms of acoustic durations, the results confirmed the findings of Bauer and Benedict (1997). Thus CV: words had the longest duration, CV:C words had a slightly shorter duration and CVC words the shortest (CV:: 368 ms; CV:C: 287 ms; CVC: 224 ms), $F(2, 22.4) = 108.84$, $MSE = 13394$, $p < .001$. Participants appeared to mark contrastive stress by lengthening the duration of syllables (Neutral stress: 253 ms; Contrastive stress: 332 ms), $F(1, 11.04) = 12.25$, $MSE = 108036$, $p < .01$. The interaction between contrastive stress and syllable type was also significant, $F(2, 22.31) = 6.53$, $MSE = 3028$, $p < .01$, with the amount of lengthening weakly correlated with the initial length of the syllable type (i.e., CV:: 93 ms; CV:C: 81 ms; CVC: 62 ms). Thus syllables with longer acoustic durations appeared to be lengthened more than syllables with shorter durations, although since this pattern was rather weak, we did not examine it further. The covariate was not significant, $F < 1$.

Whilst the overall pattern was clear, individual differences amongst participants were observed. In particular, whilst the overall effect of participant was not significant, $F(11, 13.08) = 1.45$, $MSE = 120196$, $p = .26$, the interaction between participant and syllable type was, $F(22, 22.02) = 34.54$, $MSE = 2035$, $p < .001$, suggesting that
differences between the acoustic durations of the three types of syllable may not be similar across all participants.

The syllable plus silent pause durations were quite different from the syllable durations alone, and the pattern was similar to that of Ferreira (1993) and Meyer (1994), where the difference between groups was generally reduced compared to the difference between acoustic durations alone. The results showed a main effect of acoustic plus silent pause duration (CV:: 749 ms; CV:C: 742 ms; CVC: 683 ms): \(F(2, 23.69) = 5.21, \text{MSE} = 53135, p < .05\). Thus there was not perfect compensation between acoustic plus silent pause duration, since, if there was, it should have led to a null-effect. However, the result was much weaker than the acoustic duration alone comparison, suggesting that there was some, if not perfect, compensation. There was also a main effect of contrastive stress (Neutral: 591 ms; Contrastive: 858 ms): \(F(1, 11.16) = 39.88, \text{MSE} = 408182, p < .001\). Thus, as with English and Dutch, contrastive stress increases the total acoustic plus silent pause duration. This difference (267 ms) greatly exceeded the difference in acoustic durations found in the last set of comparisons (79 ms). Thus contrastive stress increases not only acoustic duration, but also silent pause duration. The interaction between syllable type and stress type was marginal, \(F(2, 22.67) = 3.43, \text{MSE} = 23132, p = .05\) (Difference between Neutral and Contrastive stress: CV:: 286 ms; CV:C: 314 ms; CVC: 197 ms).

Overall, the results suggest that Cantonese behaves similarly to English and Dutch, where silent pause durations after syllables are negatively correlated with the acoustic duration of syllables. As with Dutch, this compensation did not appear complete, with CVC words, which had the shortest mean acoustic duration, also having the shortest acoustic plus silent pause duration.

6.5.2 No-break utterances

No-break utterances across syntactic junctures occurred only 11.54% of the time. However, they were not evenly distributed across the groups, with the greatest number of no-break utterances coming in the neutral stress condition with CV: words (Neutral stress, CV:: 25.71%; CV:C: 14.29%; CVC: 17.32%; Contrastive stress, CV:: 5.62%; CV:C: 2.91%; CVC: 3.37%). To examine this pattern, we used a similar ANOVA to that above. The results showed main effects of syllable type, \(F(2, 23.12) = 3.80, \text{MSE} = .42, p < .05\); and stress (Neutral stress: 19.09%; Contrastive stress: 3.09%); \(F(1, 11.22) = 13.54, \text{MSE} = 5.87, p < .005\); and the interaction between them was marginal, \(F(2, 22.43) = 3.06, \text{MSE} = .45, p = .067\). In terms of individual differences, the main effect of participant approached significance, \(F(11, 13.86) = 2.33, \text{MSE} = 1.17, p = .07\), and there was an interaction with syllable type, \(F(22, 22.03) = 2.23, \text{MSE} = .11, p < .05\), suggesting variability in the number of no-break utterances participants produced.

The results suggest that the longer average silent pause durations after syllables in Cantonese compared to English did not stop participants from using no-break utterances. In particular, the syllable type with the longest acoustic duration in the neutral stress conditions (CV:) caused people to use no-break utterances more often. There was no such effect in the contrastive stress condition. This pattern suggests that in at least some prosodic conditions, syllables of longer acoustic durations cause
no-break utterances to occur more commonly. This suggests that no-break utterances may not be due to people simply ignoring major syntactic boundaries and may also not be due to variance in syllable durations alone. Rather, it appears that the results may be due to variance in syllable duration, with no-break utterances occurring most often when syllables that are longer than the typical timing slots assigned for their word class are encountered, but only in prosodic conditions that do not favor long silent pause durations.

Despite the differing pattern of results in terms of the number of no-break utterances in contrastive and neutral stress conditions, the results suggest that Cantonese speakers often use no-break utterances. If we are correct in assuming that non-particle syllables in Cantonese cannot usually span two places in a prosodic foot, as suggested by Duanmu (1996) for Mandarin, it suggests that syllables read aloud by Cantonese speakers should not have crossed these major syntactic junctures so frequently. Given that this did in fact happen, a simpler explanation might be that it is due in part to durational variance across syllable types and the relative independence of segmental phonology from the process that calculates where in time syllables should fall. That is, when syllables happen to cross the edge of the slot they are assigned at major syntactic junctures, which are harder to calculate accurately in Cantonese compared to Mandarin due to the extra syllable types, they are not typically realigned so that there is a silent pause.

6.6 Individual differences

Despite the limited number of participants, there were reasonably large individual differences in the results with respect to the average acoustic and silent pause duration. However, in terms of compensation between these two measures (the focus of this study), the results were quite stable. As can be seen from the bottom two panels of Figure 2, the mean durations of the three types of syllable examined at least followed the same categorical pattern for all participants (i.e., CV: > CV:C > CVC) in both the contrastive and neutral stress conditions. Alternatively, with the acoustic plus silent pause durations, there was no single participant where the range calculated based on the middle 50% of the observations (i.e., the 25th and 75th percentile) for each syllable type was not overlapped at least to some extent by the range calculated the same way for the other syllable types. Thus, while compensation between acoustic and silent pause duration may not have been perfect, it was clearly evident across our participants. The other notable pattern is that the amount of variance caused by differences in the acoustic durations of syllables was much less than that caused by differences in silent pause durations. Thus, the duration of silent pauses is much more variable than the acoustic duration of syllables, across all participants.

To further explore individual variation in the number of no-break utterances, we examined the average acoustic duration, speech rate, and silent pause duration of each participant and how these factors related to the number of no-break utterances that they gave. The average silent pause duration was calculated after no-break utterances were removed. The results showed a weak correlation between the number of no-break utterances and the variables examined (Syllable duration: \( r = .24 \); Speech Rate: \( r = -.20 \); Pause duration: \( r = -.08 \)). Despite the small number of participants,
the weak correlations suggest that there are large individual differences between the extent that participants use no-break utterances and that this is not simply a function of speech rate or other such measures that were examined (see also Fant, Kruckenberg, & Nord, 1991, who also found wide individual variability in silent pausing/speech rates). The results appear in Figure 3.
Experiment 2

The previous experiment examined the effect of syllable timing at major syntactic junctures using words syntactically similar to those of Ferreira (1993), but in a language with very different prosodic characteristics. Similar results were observed, where the silent pauses tended to compensate for the acoustic durations of the syllables.

Experiment 2 focused on the aforementioned word class known as sentence (-final) particles, which have very different timing properties than other word classes in Cantonese. They exhibit a greater amount of timing variance and may have a very long acoustic duration. As discussed in the introduction, this type of word allows for a rigorous test of how closely the duration of timing slots is related to the segmental phonology of the words that fill them, since these durational differences mean that if slot size is related to typical syllable durations, then very long silent pauses would need to occur after particles that only have a short acoustic duration. It therefore seems meaningful to examine timing in a similar way to Experiment 1, but with particles immediately preceding the prosodic/syntactic boundary. The idea here is that the compensatory relationship between silent pause duration and the acoustic duration of syllables may be weaker with particles than non-particles.

Since it is not possible to manipulate long and short particles via the use of particles with a CVC and CV:C structure without excessive particle duplication (the vast majority are a CV: structure), we only used an explicit stress manipulation. However, for the analysis below, we ordered the particles based on their mean acoustic duration, calculated across the 12 participants, and then split them into two groups, creating an intrinsically long duration and an intrinsically short duration group. These two groups were then used as if a long and short intrinsic duration manipulation had been done. This sorting procedure was adopted because it is difficult to predict a priori which particles are likely to be short and which long. As a third comparison group, we used sentences of a similar structure, except without a particle. A contrastive stress manipulation was not used with this group, since it is possible that particles read aloud with contrastive stress may have idiomatically specified metrical information, and, in addition, stress may be a different phenomenon when used with particles (where...
it might be better termed syllable lengthening) versus other syntactic types. It is thus
difficult to determine the meaningfulness of a comparison between stressed particles
and stressed non-particles. The stimuli were all of the following form:

[Start of sentence 1][clause final word][particle with neutral stress], [end of sentence]

[Start of sentence 1][clause final word][particle with stress], [end of sentence]

[Start of sentence 2][clause final word], [end of sentence]

An example sentence with a particle is ('zaa3 is the particle of interest, it means
“only” in this situation; CL = classifier; ASP = aspect marker; MASS = classifier for
mass nouns; PRT = resultative particle; SFP = sentence final particle):

我淨係食D糖 setInput mou5 sik6-dou3 di1 dan3gou1 aa3
I only eat-ASP MASS sweet SFP, I not-have eat-PRT MASS cake SFP.
I have only eaten the sweets, I haven't touched the cake at all.

An example of a control sentence is (gung1 is the critical syllable):

我聽講佢辭工 keoi5 ci4-zo2 daai6hok6 fan6 gung1, jiu3 heoi3 LV zou6 je5,
I hear say he resign-ASP university CL job, need go LV do something.
I heard he quit his university job, he wanted to join LV instead.

Note that the sorting procedure used does not directly sort the acoustic plus silent
pause durations, only the acoustic durations. Thus, if silent pause duration and acoustic
duration were not negatively correlated, we would expect to find that the short acoustic
duration group would also show a shorter acoustic plus silent pause duration, due
to the inclusion of the acoustic duration component in the acoustic plus silent pause
duration comparison. Furthermore, it should be harder to find a null-effect where silent
pause plus acoustic durations are not statistically different in short and long acoustic
duration groups than when groups are picked a priori based on another factor such
as syllable type, other things being equal. This is because when groups are initially
picked based on another factor such as syllable type, it means that due to variance,
some items that are intrinsically short for reasons not to do with syllable type may be
classified in the group that is expected to be long and vice versa. This adds additional
variance into the acoustic duration part of a long–short acoustic plus silent pause
duration comparison compared to when the groups are post hoc sorted, where the
acoustic duration difference between the two groups is maximized. It should therefore
be easier to find a null-effect of acoustic plus silent pause duration when the groups are chosen based on syllable durations when they are sorted compared to when they are a priori chosen based on expected duration. We therefore consider it reasonable to examine whether compensation still occurs using our sorting procedure, at least when interpreting null-effects of acoustic plus silent pause duration. Alternatively, if there are differences between the acoustic plus silent pause durations of the sorted groups, it is harder to interpret. Such a difference might be due to the fact that the acoustic duration difference is maximized, hence making the acoustic duration component a potentially greater factor in the acoustic plus silent pause duration comparison than the silent pause duration component.

7.1 Participants
Twelve Cantonese-speaking students participated in the study. All except one were from the University of Hong Kong. A number of them had participated in the previous experiment, although there was a break of over a month between this experiment and the previous one.

7.2 Materials
Twenty-eight critical sentences with a large variety of particles were used. All had a structure where a particle was used at a major syntactic juncture, but where the sentence continued after the particle. Each sentence used a different particle, although sentences were repeated in the stressed and neutral stress conditions. A further 32 sentences that were similar in construction to the sentences with the particles, in terms of being broken up into two major syntactic constituents, except without a particle, were used. In addition, the ratio at which different syllable structures were used was similar to the sentences with particles.

7.3 Procedure
The procedure was the same as for the previous experiment, apart from only two counterbalanced groups being used, with each group being given all stimuli, but in a different order. The 32 non-particle sentences were repeated in each counterbalanced group, whereas the sentences with the particles were only shown in one of their different forms (stressed, neutral) in each half of the entire stimuli list.

7.4 Data processing
All word and silent pause durations above 1500 ms were removed from the analysis, as were words that were incorrectly pronounced. This led to 0.21% of words and 2.5% of silent pauses being removed. For each participant, a mean RT score was calculated and all items with durations 2.5 SDs above or below it were removed from the analysis. This was done for both word and silent pause durations, causing 1.60% of words and 1.46% of silent pauses to be removed. Mean item durations were then sorted into a list. This list was split into two in each of the stressed and neutral stress
Table 2

Means (ms) and (SDs) of syllable length (acoustic duration), silent pause duration, syllable plus silent pause duration, and proportion of times there was a no-break utterance after the critical word in Experiment 2, as a function of stimuli type

<table>
<thead>
<tr>
<th>Stimuli type</th>
<th>Syllable length</th>
<th>Silent pause duration</th>
<th>Syllable plus silent pause duration</th>
<th>No-break utterance proportion</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M</td>
<td>SD</td>
<td>M</td>
<td>SD</td>
</tr>
<tr>
<td>Neutral stress particle</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Short duration</td>
<td>238.6</td>
<td>74.4</td>
<td>417.2</td>
<td>273.2</td>
</tr>
<tr>
<td>Long duration</td>
<td>312.8</td>
<td>91.7</td>
<td>328.2</td>
<td>288.9</td>
</tr>
<tr>
<td>Difference</td>
<td>-74.2</td>
<td>89.0</td>
<td>14.8</td>
<td></td>
</tr>
<tr>
<td>Stressed particle</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Short duration</td>
<td>311.2</td>
<td>114.0</td>
<td>535.3</td>
<td>309.9</td>
</tr>
<tr>
<td>Long duration</td>
<td>425.7</td>
<td>156.3</td>
<td>419.4</td>
<td>305.9</td>
</tr>
<tr>
<td>Difference</td>
<td>-114.5</td>
<td>115.9</td>
<td>1.5</td>
<td></td>
</tr>
<tr>
<td>Final syllable (non-particle)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Short duration</td>
<td>243.5</td>
<td>60.9</td>
<td>196.7</td>
<td>205.1</td>
</tr>
<tr>
<td>Long duration</td>
<td>296.6</td>
<td>60.4</td>
<td>171.0</td>
<td>187.8</td>
</tr>
<tr>
<td>Difference</td>
<td>-53.1</td>
<td>-27.4</td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>

conditions, thus creating two groups of fast-to-articulate particles and two groups of slow-to-articulate particles. A similar procedure was performed on the other 32 sentences without particles, with both presentations of each sentence being considered an independent token, and therefore used 64 individual tokens in the analysis. The results appear in Table 2.

7.5 Results

7.5.1 Syllable and syllable plus silent pause durations

A 12 (Participant) × 3 (Stimuli type) × 2 (Short/Long duration) between-groups ANOVA was used to examine the acoustic duration of the syllables. Speech rate was used as a covariate. Participants were considered to have been randomly sampled
whereas stimuli type and duration were considered fixed factors. There was a main effect of acoustic duration (258 ms vs. 330 ms), which was expected since that is how the stimuli were ordered, $F(1, 11) = 91.32$, $MSE = 460369$, $p < .001$; a main effect of stimulus type (Stressed particle: 368 ms; Neutral particle: 275 ms; Non-particle: 270 ms), $F(2, 22) = 11.61$, $MSE = 888295$, $p < .001$, and a significant interaction, $F(2, 22.16) = 9.70$, $MSE = 82376$, $p < .005$ (short vs. long duration groups: Stressed particle, 328 ms vs. 417 ms; Neutral particle, 239 ms vs. 313 ms; Non-particle, 244 ms vs. 297 ms). The interaction appeared to have been caused by the stressed particle group having a greater acoustic duration and also a greater acoustic duration difference between the short and long groups than the other two groups (Difference between short and long duration groups: Particle with stress: 113 ms; Particle with neutral stress: 68 ms; Non-particle: 53 ms). There was no main effect of participants, $F(11, 24.35) = 1.16$, $MSE = 92014$, $p = .36$.

A similar ANOVA performed on the silent pause plus acoustic durations only showed a main effect of stimulus type (Stressed particle: 846 ms; Neutral particle: 648 ms; Non-particle: 454 ms), $F(2, 22.04) = 31.89$, $MSE = 14418092$, $p < .001$. This appeared to be caused by the stressed particle group having a greater total duration than the other two groups and the non-particle group having the shortest syllable plus silent pause duration. Neither duration group (i.e., short versus long) nor the interaction even approached significance (Group: $F < 1$; Group × Stimulus type, $F(2, 22.16) = 1.71$, $MSE = 126911$, $p = .20$). These results suggest that silent pause duration compensates for acoustic duration very well in the groups examined. As can be seen from Table 2, that was especially true for the particles, where compensation was almost perfect, despite the huge difference in acoustic durations (Stressed particle: 7.26 ms; Neutral stress particle: 3.14 ms). In addition, in terms of individual differences, whilst there were differences in overall means, $F(11, 21.82) = 4.69$, $MSE = 452131$, $p < .005$, the interaction between participants and the long and short groups was not significant, $F(11, 23.58) = 1.25$, $MSE = 90661$, $p = .31$.

In terms of how word durations and silent pauses interact, the results were very similar to those observed in Experiment 1 and the results of Ferreira (1993): when words with long durations were used, only a short silent pause tended to separate them from the next word and vice versa. This was despite the fact that the average actual acoustic duration of the words and silent pauses summed together reached almost a second in one group (stressed particles), almost twice as long as in the results reported by Ferreira. This total length is in fact very similar to what could be predicted a priori by a slot-based theory based on the longest duration particles that are commonly found in normal Cantonese speech, which, according to Matthews and Yip (1994) and Chan (1998), is just under 1 second.

At first glance, these results support the possibility that syllable timing might be relatively independent of a word’s segmental phonology, in the same way as other similar results have done. In particular, the results suggest that particles are processed in much the same way as other types of words, despite their idiosyncratic properties, and to compensate for their greater variability in duration, the prosodifier simply allows a very long timing slot for them to be inserted into.
The results are not completely compatible with the idea that segmental phonology is independent of syntax, however, since there were differences between particles and non-particles. In particular, in neutral stress conditions, the silent pause duration after particles appeared much longer than the silent pause duration after non-particles, despite both groups having similar acoustic syllable durations. Thus, in one case, the silent pause duration was around 1.35 times the length of the acoustic duration, and in the other it was only .68 times the length. This suggests that acoustic duration does not necessarily increase with increasing silent pause duration. Alternatively, in a contrast between neutral and stressed particles, quite similar acoustic/silent pause duration ratios were found (1.30 vs. 1.35), but the size of the syllable and silent pause durations added together was much greater in the stressed group. This suggests that syllable duration cannot be simply a function of the intrinsic duration of a word plus a proportion of the entire timing duration assigned for the segmental phonology and silent pause (i.e., slot size). This is because in one comparison (neutral stress particle vs. non-particle), a longer total slot size did not induce a longer acoustic duration, but in the other comparison (stressed particle vs. neutral stress particle) it did.

Although the pattern of results found with the particles and non-particles may have differed, it is not necessary to give up on the idea that syllable timing is typically independent of segmental phonology. In particular, the extra semantic emphasis that particles may attract compared to the other word classes may perhaps be accommodated via the assignment of more than one timing slot (in terms of slot-based models), or people waiting a greater number of cycles when particles are encountered before beginning to articulate the next word (in terms of P-center models). In both these cases, it could be assumed that the acoustic duration is only related to the first timing slot/P-center cycle, and hence different silent pause durations may be found between particles and non-particles, even though acoustic syllable durations are similar. Such a possibility remains to be explored.

7.5.2 No-break utterances

As in the analysis of acoustic and silent pause duration, word class also appeared to have an effect on the number of times people did not use silent pauses at syntactic junctures, with speakers using no-break utterances more commonly with non-particles than with particles (Stressed particle: 8.63%; Neutral particle: 12.20%; Non-particle: 33.72%), $F(2, 22.23) = 26.20$, $MSE = 7.61$, $p < .001$. The number of times participants used a no-break utterance after short duration particles and words also appeared less than long duration particles and words (21.67% vs. 24.03%), although the result was not significant, $F < 1$. Like the previous data, the number of no-break utterances differed significantly across participants, and this was reflected in both a main effect of participant, $F(11, 14.73) = 3.21$, $MSE = 7.61$, $p < .05$, and an interaction between participants and stress type, $F(11, 24.26) = 2.65$, $MSE = .42$, $p < .05$.

The results of the no-break utterance analysis support the possibility of a word class distinction between the timing of particles and the timing of non-particle clause-end syllables, for the same reason as the acoustic plus silent pause duration analysis.
This is because the acoustic duration of the neutral stress particles, as found in the previous analysis, was similar to that of the non-particle syllables. However, the number of no-break utterances was far fewer after particles than non-particles. The most reasonable explanation for this is that particles have proportionately more silent pause after them than non-particles, and hence they are less likely to be associated with a no-break utterance. Thus, whilst variability in timing durations has some effect on the number of no-break utterances (being higher for syllables with long acoustic durations compared to syllables with short acoustic durations), word class may also play a role too.

In terms of individual variation, the results were very similar to the previous experiment. As can be seen from Figure 4, very reliable differences between acoustic durations in the short and long groups were basically eliminated once silent pause duration was added, and this phenomenon appeared quite stable across participants. A slight difference emerged in the pattern of correlations with the number of no-break utterances. In particular, whilst the correlation with syllable duration ($r = .14$) and speech rate ($r = .032$) was very weak, the correlation with silent pause duration was moderate ($r = -.60$). The latter finding suggests that participants who tend to use shorter silent pauses also tend to use no-break utterances more. The results appear in Figure 5.

**Figure 4**

Mean individual acoustic and acoustic plus silent pause durations found in Experiment 2. The boxes are defined based on data points from the first, second, and third quartiles. The whiskers represent values two standard deviations above and below the mean. Note that the Y-axis is scaled differently in the acoustic plus silent pause versus acoustic duration only groups. S = Short; L = Long
In this study we examined the relationship between the acoustic duration of syllables and silent pause durations in Cantonese, a language prosodically quite different from others that have been examined. Results from the two experiments revealed a strong compensatory relationship between the acoustic duration of syllables and the duration of the silent pauses after them. That was true both of word classes that have been examined in other languages, and of particles, which can have exceptionally long and variable articulation durations in Cantonese.

The results are of general interest in terms of the rhythmic effects found in speech. First, they show that compensation between acoustic and silent pause duration occurs in conditions where there is comparatively high variability in syllable durations and in conditions where there is comparatively low variability. Second, they show that even in languages where each syllable is fully articulated, compensation still occurs. These results, together with those from other languages, suggest that this compensation effect is relatively independent of the unique features that syllables in different languages may have, such as tone (which Cantonese has), perceptually audible word-level stress (which Cantonese lacks—all syllables are stressed at an output level), and the durational variability of syllables. This suggests that compensation between acoustic and silent pause duration at syntactic junctures is a very general property of speech timing.

The data add some support to models of speech timing that use slot-based mechanisms. This is because the size of the slot created for syllables appeared to be strongly related to word class and prosodic position, rather than segmental phonology. When we used particles, an exceptionally long slot was left compared to when word classes not known for such timing variability were examined. Thus, long slots were made for words in prosodic positions where syllables with potentially very long acoustic durations can occur, even when the syllables themselves were not long.
Whilst the mean results appear relatively clean, in the sense that compensation between acoustic and silent pause duration was found in a number of different conditions, a large amount of variability exists amongst individuals. This was especially so for the duration of silent pauses and the number of no-break utterances used. Interestingly, however, the strength of the correlation between speech rate and both silent pause duration and number of no-break utterances was in fact surprisingly weak. Thus, speech rate was not strongly associated with the length of the silent pauses used (see Fant, Kruckenberg, & Ferreira, 2003; Fant et al., 1991, for a similar observation). One interpretation of these results, based on the idea of cyclical vowel production (e.g., Fowler, 1979), is that at major syntactic junctures, there is considerable individual variability in the number of cycles different speakers leave before beginning to speak again. Thus, whilst faster speakers may produce syllables more quickly, the actual number of cycles they may wait for at syntactic junctures is not necessarily strongly related to this rate. This would add a large amount of variance into a correlation between silent pause duration and speech rate, and thus reduce the strength of any correlation between these variables.

In terms of more specific implications, it is possible to return to our original cross-language question as to why silent pauses occur at major syntactic junctures more commonly in Mandarin than English. The results suggest that the idea that Mandarin may cause more empty slots in prosodic feet to be left than English may be overly complex. In particular, if an identical prosodic foot assignment is assumed, then empty slots should have been able to have been filled in some circumstances by the stressed particles in this study, which have enormous variability in timing (and hence should be able to be “stretched” across a slot). However, that was not case: there was still almost always a break for that type of word. This also appeared true of clause-end words with contrastive stress used in the first experiment, where, again, a silent pause was almost always found. By contrast, it was not true of the syllable type with the longest acoustic duration that we used in neutral stress conditions in Experiment 1 (i.e., CV: syllables): participants often used a no-break utterance with those syllables, despite the syntactic juncture. These results, and the difference between the contrastive stress and word class results, suggests that a number of factors play a role in the extent to which people use no-break utterances at clause-end boundaries, including just simple variation in syllable durations.

Finally, the results reported here should remain as tentative. One problem with the study is that the data were generated in relatively artificial conditions, in that participants read aloud relatively short utterances that were not related in any way. These sorts of conditions may have caused the participants to use less no-break utterances than under more naturalistic conditions, as they may have habituated to the typical intonational division in the stimuli. For the same reason, rhythmic effects may have been exaggerated. Further study on more naturalistic speech is therefore needed to generalize the results, particularly with respect to rhythmic effects in different speech conditions (see e.g., Fant et al., 1991).

References


C. Perry, R. Kwok-Shing Wong, S. Matthews 51

Downloaded from las.sagepub.com at Swinburne Univ of Technology on August 7, 2010
SYLLABES AND SILENT PAUSES


