UCLA Working Papers in Phonetics, No. 109, pp. 40-59

Prosodic Boundaries and the Taiwanese Tone Sandhi Group

Grace Kuo
gracekuo@humnet.ucla.edu

Abstract

The goal of this study was to examine the Taiwanese Tone Sandhi Group domain. Replicating Carlson et al.'s (2005) perception study, an experiment was conducted aiming to see whether Taiwanese listeners were able to predict the occurrence and strength of upcoming boundaries - Word, Tone Sandhi Group and Utterance. Stimuli were selected from a corpus of spontaneous Swedish speech and Taiwanese read speech, so as to vary in Boundary types (no break vs. weak break vs. strong break) Fragment size (long vs. short) and Filtering was also varied (normal vs. low-pass filtered). These fragments were presented to two Taiwanese listeners who were instructed to guess whether a prosodic break would follow each fragment. Results revealed that Boundary type is a factor that influenced the judgments when stimuli with all boundary types, fragment sizes and filtering conditions were combined. Taiwanese listeners are able to differentiate strong break vs. weak break vs no break in Swedish and strong break vs. weak break in Taiwanese, but not weak break vs. no break in Taiwanese. In addition, Taiwanese listeners, like American English listeners in Carlson et al's study, were able to hear the boundaries in Swedish in normal speech. Acoustic and prosodic correlates from the Taiwanese stimuli were also examined. There was a small but significant correlation found between judgments and the pause duration. Final word duration ratio and Energy seem to be the most reliable cues to differentiate all Boundary types. Moreover, voice quality analysis suggests that breathiness, rather than creakiness, may have been used to distinguish Tone Sandhi Group boundaries from Utterance boundaries.

1. Introduction

Pitch contours/declination, final lengthening, intensity drop and glottalization are found to be common *pre-boundary* markers in the conveyance of prosodic grouping (Klatt 1975, Lehiste 1979, Kreiman 1982, Grosjean 1983, Wightman et al. 1992, Dilley et al. 1996, Swerts 1997, Barron et al. 2002, Ferrer et al. 2002, Mo et al. 2008). *Post*-boundary cues include initial lengthening and strengthening, and f0 reset (Pierrehumbert & Talkin 1991, Jun 1993, Swerts & Geuykens 1994, Fougeron & Keating 1997, Swerts 1997). The cue observed *at the* boundary is sometimes the presence of a silent pause (Shen 1992, Ferreira 1991&1993, Tseng 2008). This paper discusses the perception of Taiwanese boundaries from these prosodic cues. Given that Taiwanese has a language-specific prosodic unit, the Tone Sandhi Group (TSG) (Chen, 1987), these prosodic cues might also be reliable markers of the boundary of a Tone Sandhi Group.

The size of a Tone Sandhi Group is similar to a phonological phrase (Chen, 1987) and its domain is considered to be determined by the syntactic structure. Chen (1987) proposed two relevant rules to describe it.

(1) a. Tone Sandhi Rule: T → T'/__T within a Tone Group, where T=base tone; T'=sandhi tone
b. Tone Group Formation: Mark the right edge of every XP with #, except where XP is an adjunct c-commanding its head[1]. (# denotes Tone Sandhi Group boundary)

Therefore, an XP (such as Noun Phrase and Verb Phrase) can be a Tone Sandhi Group, and different ways of grouping the phrases give different interpretations to the same utterance. (2) provides examples of how the words are grouped syntactically and how the meanings of the utterance are different (adopted from Hsiao, 1991).

(2) a. 'Doing it seven times, and failing it eight times.'
underlying tones: [*chhit31 cho31*]$_{VP}$ # [*beh31 m33-tioh51*]$_{VP}$ #
seven make eight NEG right
surface tones: [*chhit51 cho31*]$_{VP}$ # [*beh51 m31-tioh51*]$_{VP}$ #

b. 'Equating seven to eight is false.'
underlying tones: [*chhit31 cho31 beh31*]$_{S}$ # [*m33-tioh51*]$_{VP}$ #
seven make eight NEG right
surface tones: [*chhit51 cho51 beh31*]$_{S}$ # [*m31-tioh51*]$_{VP}$ #

In (2), the underlined tones are the sandhi tones. As shown in these examples, if a word is not at the right edge of an XP, the tone of that word has to undergo tone sandhi. In (2a), there are two verb phrase, and the literal meaning is that someone does something seven times but fails on that thing eight

1 The exceptions mainly consist of adjuncts. An adjunct (including an adverbial phrase and an adjective phrase) does not qualify to be a Tone Sandhi Group in that it usually bears a sister-relation and c-commands the head. For instance, in (i), the AP does not form a Tone Sandhi Group and so every word in the AP undergoes tone sandhi (the sandhi tones are underlined). This paper only deals with NP and VP, so these adjunct variations will not be covered.
(i) 'white rose'
underlying tones: [*beh51 sek31*]$_{AP}$ # [*mui13 gui55*]$_{NP}$ #
surface tones: [*beh31 sek51*]$_{AP}$ [*mui33 gui55*]$_{NP}$ #

times, meaning this person keeps failing. In (2b), *beh31* 'eight' is part of the first group of words, which happens to be a sentence, and the meaning of this whole utterance is 'it is not right to take the number seven as the number eight.'

(2) shows that different phrasing leads to different interpretations of the same utterance. The phrasing difference is particularly clear when (2a) and (2b) are produced in speech. Apart from the tone differences between *cho31* and *beh31*, other acoustic cues also facilitate the correct interpretation of the sentence. For instance, the places where the pauses occur – after the first VP in (2a) and after the S in (2b). Furthermore, since the sizes of the prosodic phrases preceding the pauses are different - the Tone Sandhi Group domain and the domain of IP for VP and S respectively - some acoustic features such as duration might be giving hints about the phrase sizes. For instance, one would expect to observe more final lengthening at the IP boundary in (2b) than at the Tone Sandhi Group boundary in (2a), as the former is a bigger prosodic domain.

Taiwanese tone sandhi is intriguing not only in that it is positionally conditioned, but also in that the sandhi patterns are characterized by circular opacity: tone changes are applied in a structure-preserving and circular way. To be more specific, the form of the sandhi tone is not a newly-created tone; instead, a sandhi tone is drawn from the existing set of lexical tones. And the tone change relations can be displayed with a circle, as shown in (3). When a word with a high falling tone (51) as the underlying tone appears in a non-final position in a prosodic phrase, it has to undergo tone sandhi, the outcome of which would be a high level (55) tone. And if the base tone is high level (55), the sandhi outcome has to be a mid level tone (33), and so on. Actual examples are shown in (4). In (4a), the word *gua* 'I' bears a high falling tone underlyingly. However, when it appears in a non-final position within a prosodic phrase, the tone is changed to a high level tone, as shown in (4b). One thing worth mentioning is that Taiwanese closed syllables (with unreleased coda, /p, t, k, ʔ/) have only two tone alternatives, high falling and low falling (with the tone values 51 and 31 respectively). They do not participate in the circular tone pattern, but instead comprise each other's tone sandhi form. They are comparatively short in duration so that they are never confused with the falling (51) and low falling (31) tones in the circular pattern for open syllables.

(3) Taiwanese tone sandhi circle

(a) open syllables

13 → 33 ← 55 ← 51

↘ ↗

31

(b) closed syllables

51 ⇆ 31

(4) 'I enjoy studying Taiwanese tone sandhi.'

a. underlying tones: [*gua51 chin55-ai31 gen13-kiu31 tai13-gi51*] [*e13 pen31-tiao33*]

I enjoy study Taiwanese GEN tone sandhi

b. surface tones: [*gua55 chin33-ai51 gen33-kiu51 tai33-gi51*] [*e33 pen51-tiao33*]

Let's take a closer look at (4) again. The whole sentence contains two Tone Sandhi units and within each unit, tones of all the syllables except for the last have undergone tone sandhi. In (4a), all the syllables are labeled with the underlying tones (also called “base tones”, “citation tones” or “juncture tones”); (4b) transcribes the surface tone (also called “sandhi tone”) for each syllable with the sandhi tones underlined. It is obvious that all the words except for the last within a prosodic domain have to carry sandhi tones.

As one might have noticed from (3) and the relevant discussion, there are seven tone categories in Taiwanese, and each has two tonal alternations – one base tone and one sandhi tone. A paradigm example is shown in Table 1. It is interesting that sandhi is structure-preserving and therefore might create some lexical ambiguities. For example, given that the sandhi outcome of *si51* 'to die' is *si55*, and that *si55* has a lexical word correspondent 'poetry', then how do listeners know, without hearing the whole utterance, whether the speaker is talking about death or poetry when they only hear *si55?* Traditionally, the sandhi *si55* 'to die' and the base *si55* are considered completely neutralized. If this is really the case, do listeners get confused when they hear an ambiguous sentence in which 'die' and 'poetry' are both plausible interpretations? If it is neutralizing but listeners don't get confused, the listeners must have detected the boundaries, in which case we would like to know how they do so. Are there any acoustic cues that they rely on to detect the boundaries?

Base tone		Sandhi tone	
si55	"poetry"	***si33****-pun51*	"poetry and prose"
si13	"time"	***si33****-kan55*	"time span; time"
si51	"to die"	***si55****-lang13*	"dead people"
si31	"four"	***si51****-tiam51*	"four o'clock"
si33	"temple"	***si31****-seng55*	"temple monk"
sik31	"color"	***sik51****-chhai51*	"color"
sik51	"ripe"	***sik31****-te13*	"baked tea"

Table 1. Taiwanese tone categories with base tone and sandhi tone examples (adopted from Myers et al. 2008)

Therefore, two main questions about Taiwanese tone sandhi arise; one is about tone neutralization and the other is about the phonetic correlates used to detect the prosodic boundaries.

(i) Are sandhi tones indeed neutralized with their lexical correspondents? If yes, is it complete or partial neutralization? If sandhi tone and its lexical correspondent are neutralizing, whether completely or incompletely, how do listeners differentiate them when processing an utterance? Ie., how do they know that this *si55* they hear is a final word with base tone 'poetry', or that *si55* they hear is a non-final word with sandhi tone 'die'?

(ii) Since position (final vs. nonfinal) is strongly related to the appearance of boundary, what are the phonetic correlates that the listeners use to detect a Tone Sandhi Group boundary?

There are three hypotheses: (a) If it is complete neutralization, every tone is potentially ambiguous and listeners would not be able to differentiate the two without being given any lexical or grammatical information. (b) If it is incomplete neutralization, the difference between the two should be gradient and should be potentially distinguishable some of the time, but most of the time, they could still be categorized as the same tone. (c) If there is no neutralization, the two are distinguishable and the traditional sandhi circle is wrong.

1.1 Neutralization

Myers & Tsay conducted a series of studies (2001 and 2008) on the issue of neutralization. Their most recent paper compared the two *kim33* with the well-designed sentence pairs as shown in (5). The sentence pairs are syntactically ambiguous and the difference is revealed when they are spoken – in (5a) the aunt and the sister-in-law are different individuals whereas in (5b), 'Akim' is the name of the sister-in-law. Myers et al. found that in terms of f0, speakers did neutralize base tone and sandhi tone. However, base tone and sandhi tone were distinguished by syllable duration, with a strong effect of phrase-final lengthening. The authors argued that the observed duration effect is consistent with gradient, incomplete neutralization of the two tone categories, so tone sandhi is neutralizing, but not categorically so.

(5) a. "These are aunt, sister-in law and elder brother."

underlying tones:	[*che51*	*si33*	***a55-kim33***],	#	[*so51-a51*]	#	[*kap31* *a55-hiaN55*]	#
	this	is	aunt		sister-in-law		and elder brother	
surface tones:	[*che55*	*si31*	***a33-kim33***],	#	[*so51-a0*[2]]	#	[*kap51* *a33-hiaN55*]	#

b. "These are sister-in-law Akim and elder brother."

underlying tones:	[*che51* *si33*	***a55 kim55-so51 a51***]	#	[*kap31*	*a55 hiaN55]*	#
	this is	Akim(name) sister-in-law		and	elder brother	
surface tones:	[*che55* *si31*	***a33 kim33-so51-a0***]	#	[*kap51*	*a33 hiaN55*]	#

Moreover, they also compared sandhi tone 33 when derived from 55 vs. from 24. As shown in the circular tone sandhi pattern in (3a), 55 and 13 have the same sandhi tone outcome: 33. This is the only case where two underlying tones might be neutralized to the same surface tone in the sandhi context. Their finding was that speakers produced overall higher f0 for sandhi tones derived from 55 than for sandhi tones derived from 13 and that the tone derived from 55 was 8 ms longer on average than the tone derived from 13. Their interpretation of this finding was that there might still be some preservation of the underlying contrast, or that maybe there were other factors involved, such as word frequency, neighborhood size, semantics, etc.

In the previous studies, the two checked tones (tones in closed syllables) were often ignored. Kuo (2010) examined how listeners discriminated the two checked tones in Taiwanese with a gating paradigm. In this study, the stimuli were selected from three "prosodic positions" and two "context conditions": in a carrier sentence or not, as shown in (6).

(6) a. Three prosodic positions: citation, juncture and sandhi; the stimuli are in bold

- citation: words in isolation
  (e.g. **pak51** 'to tie')
- juncture: last syllable in a sandhi domain
  (e.g. sio55 pak51 → sioh33 **pak51** 'to tie together')
- sandhi: first syllable in a disyllabic sandhi domain
  (e.g. pak31 hiong33 → **pak51** hiong33 'northward')

b. Two context conditions: in the following carrier sentence or in isolation

2 Note that the diminutive expressions (a51→ a0), like some function words, do not participate in typical tone sandhi.

'He listens to the tone of ___.'
underlying tones: [*i33*] # [*thiaN55* _____] # [*e13* *siaN55 tiau33*]
he hear GEN tone
surface tones: [*i33*] # [*thiaN33* _____] # [*e33* *siaN33 tiau33*]

The consonant of each stimulus was always the first gate, and the following gates were formed in 5 ms increments. Listeners were presented with stimuli arranged in a duration-blocked format and their task was to decide which tone they heard. The surface tones were considered correct; that is, "high falling" was the correct answer when listeners hear *pak51* from the sandhi position whose underlying form was *pak31* in *pak31 hiong33*. The results revealed that correctness was always very much above chance, and there was no significant main effect of prosodic position. In other words, the two checked tones are neutralized so that when listeners heard a a word with a derived tone, they identified it by its surface tone. There are two possible explanations: first, listeners consistently have a bias toward the surface tones; second, because the stimuli were presented in isolation, listeners did not realize that they had undergone tone sandhi.

Myers et al.'s and Kuo's neutralization studies suggested that every tone token is potentially ambiguous. Therefore, we shall move on to question (ii): “What are the potential phonetic correlates that listeners use to differentiate a sandhi tone from a base tone?" But first, let us look at the language-specific prosodic unit - the Taiwanese Tone Sandhi Group domain.

1.2 Taiwanese Tone Sandhi Group

It is known that the prosodic structure is closely related to, but not isomorphic to, the syntactic structure. For example, a relative clause can be analyzed (here, bracketed) differently according to these two structures, as in (7). In (7b), the prosodic phrases are usually separated by breaks, which are usually realized as pauses.

(7) a. syntactic configuration
[This [is [the cat that caught [the rat that stole [the cheese]$_{NP}$]$_{NP}$]$_{NP}$]$_{VP}$]$_{S}$
b. prosodic phrasing
[This is the cat] [that caught the rat] [that stole the cheese]

The issue of prosodic domain in Taiwanese is even more interesting because of the tone sandhi phenomenon. It seems that the interpretation of a sentence depends heavily on the prosodic domains, but the prosodic domains do not always line up with the syntactic structures. Figure 1 is an example from Chen (1987) demonstrating how a sentence is phrased syntactically and prosodically. In this diagram, # denotes a Tone Sandhi Group boundary, assuming that it is an independent prosodic domain. The fragment before each # forms a Tone Sandhi Group, and there are three Tone Sandhi Groups in this sentence (including the last fragment without a sentence-final #). According to the tone sandhi rule (i.e. only the last syllable in a tone group retains the base tone), all the preceding syllables undergo tone change and bear sandhi tones.

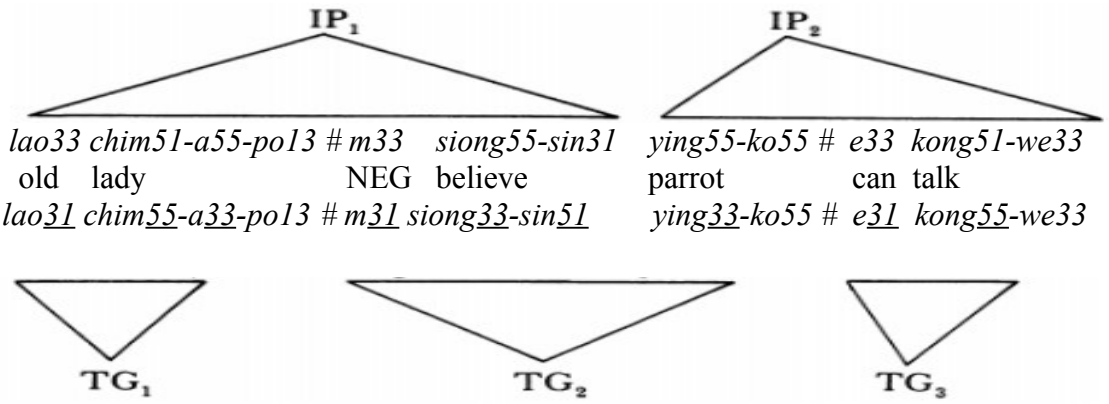

Figure 1. Syntactic configuration and prosodic phrasing of the sentence "The old lady doesn't believe that parrots can talk." IP is determined in syntax, not Intonational Phrase; TG refers to Tone Sandhi Group. (Chen, 1987)

In Figure 1, the syntactic break occurs between "believe" and "parrot", but prosodically, these two words are within the same prosodic domain. The mismatch of the syntactic phrases and the prosodic domains is interesting and somewhat surprising, yet it is not unusual. A similar example showing syntactic-prosodic domain mismatch is French Liaison (Nespor & Vogel, 1986), as illustrated in (8). In syntactic terms, *meilleurs* and *intolérables* are the complements of *amis* and *perrochets, s*o"*meilleurs amis*" and "*perrochets intolérables*" are of the same type of constituent. However, liaison can only apply in the first case and never in the second case. This mismatch again shows us that we cannot identify the domains of the application of phonological rules solely using syntactic constituents.

(8) a. *Les giraffes et les éléphants sont ses meilleurs͡amis*
'Giraffes and elephants are his best friends.'
b. *Claude a des perrochets / intolérables*
'Claude has some intolerable parrots.'

However, whether or not the Tone Sandhi Group is a prosodic domain in the prosodic hierarchy remains controversial. In Chen's (1987) analysis, he claimed that the Tone Sandhi Group appeared to correspond closely to the Phonological Phrase, but he also pointed out that if the Tone Sandhi Group was equivalent to a Phonological Phrase, it would violate the Strict Layer Hypothesis – the Tone Sandhi Group could straddle the boundary dividing two Intonational Phrases, as shown in Figure 1.

Hsu and Jun (1996) tried to provide an answer to this controversy using the VOT and closure duration of word-initial /p^h/ in different prosodic domains (Intonational Phrase and Tone Sandhi Group) at different prosodic positions (initial, medial and final). The hypothesis was that if Tone Sandhi Group was not a prosodic domain, it should not have shown a pattern of domain boundary marking. Interestingly, although the Tone Sandhi Group did not have the domain-initial strengthening effect (which marked the Intonational Phrase boundary), it did have a domain-final weakening effect. However, Shu and Jun were reluctant to incorporate the Tone Sandhi Group into the prosodic hierarchy because it was still a problem for the Strict Layer Hypothesis.

In Keating et al.'s (2003) cross-linguistic study on domain-initial strengthening, they examined the linguopalatal contact and articulatory seal duration of initial /n/ and initial /t/ in five different domains (Utterance, Intonational Phrase, Small Phrase, Word and Syllable). A Small Phrase was a heavy subject Noun Phrase which happened to be a strictly layered Tone Sandhi Group. They predicted that Taiwanese would show larger edge-marking compared to other languages, because the domain head was not marked with tone. However, the result showed that this was not the case. More linguopalatal contact and seal duration were observed in the initial position of higher prosodic domains than lower prosodic domains. To be more specific, they found that one speaker distinguished Intonational Phrase from Small Phrase when the initial consonant was /n/, and did not distinguish Utterance from Intonational Phrases when initial consonant was /t/. The other speaker distinguished all four pairs with both initial /t/ and /n/, failing only to distinguish Small Phrase vs. Word. These results suggested that the Small Phrase (= Tone Sandhi Group) might be successfully incorporated into the prosodic hierarchy.

Pan (2003) investigated the degree of nasalization in final nasal and initial voiced stops across four prosodic domains (Intermediate Phrase, Tone Sandhi Group, Word and Syllable). She found that the durations of final nasal and initial voiced stops of the Tone Sandhi Group and Word domains were not distinctive. However, when considering the amplitude of nasal peak, the Word domain showed a speaker-dependent pattern, whereas the Tone Sandhi Group showed a consistent pattern, with a higher amplitude of the nasal peak than of the Syllable domain. Therefore, she concluded that the Tone Sandhi Group, though maybe not distinct from the Word domain, was an important prosodic boundary in Taiwanese. Furthermore, in her f0 study on the two checked tones (Pan, 2006), she found that f0 decreased more slowly before Intonational Phrase (IP) and Tone Sandhi Group (TG) boundaries than before Word (WRD) and Syllable (SYL) boundaries, and f0 falls fastest after IP, followed by TG, SYL and then WRD. It thus appears to be plausible to treat Tone Sandhi Group as a prosodic domain, since its f0 velocity patterns are different from other prosodic domains.

In fact, when it comes to prosodic annotation of Taiwanese, tone sandhi is always taken into account. In a recent ToBI convention for Taiwanese (Peng & Beckman, 2003), the tone sandhi boundary has to be particularly marked with a break index 'b3'. Unlike Taiwanese, Mand-ToBI (p.c. Janice Fon) and C-ToBI (online resource), referring to Mandarin-ToBI and Chinese-ToBI respectively, neither mark sandhi tones nor use a break index to indicate tone sandhi; they simply transcribe the surface tones.

b4	intonation phrase boundary, either utterance finally (all figures) or medially (e.g., in utterance yws17)
b3	tone sandhi group (TSG) boundary
b3m	percept of TSG boundary without sandhi tone
b2m	base tone without percept of the TSG ending
b2	ordinary "word-internal" syllable boundary
b1	resyllabification (e.g., /het^{21} e/ → [he^{53}.le] in yws19)
b0m	syllable fusion (e.g. /hun^{51}-aŋ^{24}-a-sik^{21}/ → [hun^{55}.aã^{33}.sik^{21}] in nls107)

Table 2. Taiwanese ToBI break index

1.3 Boundary

The boundary of a cumulatively structured prosodic domain tells us about the size of the domain. Research on boundaries, as mentioned above, has shown that the perception of boundaries and degrees of boundary strength are dependent on factors such as pauses, f0 contour, intensity, strengthening effects (e.g. initial strengthening and final weakening) and glottalization. Carlson et al. (2005) reported a perceptual experiment investigating the cues to upcoming prosodic boundaries in spontaneous Swedish speech. In their study, they had Swedish and American English listeners predict the occurrence and the strength of upcoming boundaries in Swedish, using a 5-point scale. Prior to the experiment, researchers (Heldner & Megyesi 2003) labeled boundary presence and strength of in a female politician's radio interview without access to spectrographic analysis. Sixty 2-second long utterance fragments were selected – one third were from positions which they transcribed as being followed by a strong break, one third by a weak break and the rest by having no break. The inter-rater reliability (pair-wise agreement and Kappa value) was fairly high. From the sixty 2-second long fragments, they made another sixty short versions of the stimuli which contained only the final word in the fragments.

They found that Swedish and American English listeners are both able to predict upcoming boundaries in Swedish based on the properties of the preceding word alone or the preceding 2-second phrase. This showed that longer stimuli do not lead to greater accuracy; in other words, the information contained in the final word of the fragment was sufficient to facilitate the prediction of the upcoming breaks, though there may well be other information distributed earlier in the phrase. The result from the American English listeners is especially interesting because it suggests that listeners are using prosodic cues to detect the boundaries, rather than semantic or lexical cues. In addition, they also found that there were significant correlations between listeners' judgments and the presence/absence of final creak and final f0 change.

In an earlier version of their 2005 paper, Carlson et al. reported that a third language group (Mandarin) was tested using the same procedure. There was no statistical result reported, but they indicated that their preliminary results showed a pattern for the 2-second stimuli that was similar to that of the Swedish and American English listeners, but Mandarin listeners were less able to make correct predictions based on one-word stimuli. It seems that for Mandarin listeners, who like American English listeners had, had no lexical or semantic information, it was easier to guess an upcoming boundary when they were given a larger fragment. In other words, the information in the final word was not sufficient for them to make correct judgments. However, the fact that they could do the task when presented with longer pieces suggested that Mandarin listeners seemed to rely on the information in the fragments before the last word more than the information from the last word alone. That is, there must be additional useful information in the interval before the last word, which Mandarin listeners used either exclusively, or in combination with the information in the last word.

The findings in Mandarin become very interesting if the same pattern is found in Taiwanese, because in Taiwanese, all tones except for that on the final syllable have to undergo tone sandhi within a Tone Sandhi Group. If the same finding is also discovered in Taiwanese, would that mean that when Taiwanese listeners detect the phrase boundary, they are paying more attention to the piece that undergoes sound change rather than the piece that retains the original property? In order to get a preliminary results, I conducted a pilot experiment, replicating Carlson et al. (2005), withe Taiwanese listeners. If Taiwanese tone sandhi is a perceivable domain, and if listeners are able to differentiate the Tone Sandhi Group from other prosodic domains such as Utterance and Word, what signal cues do they

seem to rely on?

2. A pilot study

This study replicates Carlson et al.'s (2005) study on the prediction of upcoming prosodic boundaries in Swedish, in the present case with Taiwanese listeners. In addition, I added some Taiwanese stimuli (adapted from Hayashi et al. 1998) to the pilot listening materials in order to observe how Taiwanese listeners detect their own language's boundaries. Moreover, all the stimuli were presented in two versions – normal and low-pass-filtered. The addition of the latter was to observe whether listeners could predict the occurrence of the boundaries when segmental information is removed.

2.1 Participants

Two native speakers of Taiwanese (one male and one female) were recruited for this pilot study. They are also fluent native speakers of Taiwan Mandarin and the length of their stays in the U.S. were around 4.5 years. They were paid 10 dollars each for participation in the one-hour experiment.

2.2 Stimuli

There were 312 listening stimuli in total – 240 Swedish and 72 Taiwanese. The Swedish stimuli were obtained from Carlson et al.'s experiment, and were recorded by one female speaker. The Taiwanese stimuli were selected from the corpus of Hayashi et al.'s study on Taiwanese domain-initial strengthening. In their corpus, there were five different prosodic domains, of which we only selected three – Utterance, Small Phrase and Word, as shown in (9). The three prosodic domains provide a three-way classification (no break, weak break and strong break), by which we could make a parallel comparison to the Swedish data's with three breaks.

(9) a. *Utterance (Utt)*
[*gua55 u31 khuaN51-tioh31 pa33-**pa55***]$_S$[*ta31-ta33*][*khai51 iah51 be31 lai13*]
I PAST see-ASP Dad Tata why yet NEG come
"I saw Dad. Why hasn't Tata got here yet?"

b. *Small Phrase (SP)*
[*hit51 e33 lang13*] [*e33 pa33-**pa55***]$_{NP}$[*ta31-tioh31 chit31-chiah51 ka33-chuah51*]
that CL person GEN Dad step-on one-CL cockroach
"The person's dad stepped on one cockroach."

c. *Word (Wrd)*
[*gua55 kah31 li55 kong51*][*pa33-**pa55***]$_{NP}$[*ta31-tioh31 chit31-chiah51 ka33-chuah51*]
I KAH you tell Dad step-on one-CL cockroach
"Let me tell you. Dad stepped on one cockroach."

Three repetitions from each speaker of the three target utterances starting from the beginning and continuing until the second "pa" were selected as the long version. (In the corpus, there were 10 repetitions for each utterance. The length of the long version was no more than 2 seconds). To be comparable with the Swedish data, I also created the one-word version by selecting the second '*pa*' from each long version token, which was shorter than the Swedish single words.

Another addition to the experimental design was the use of low-pass filtered stimuli. Low-pass filtered speech preserves the rhythmic timing and pitch of the original speech but removes all the segmental information. The purpose for the use of this version was to see if Taiwanese listeners were able to correctly detect Taiwanese boundary breaks without being given any lexical information. Recall that in Carlson et al.'s results, American English and Mandarin speakers could predict the occurrence of the upcoming boundary when exposed to Swedish stimuli, suggesting that listeners do not rely on lexical/semantic information to find the boundaries. Having said that, this should still hold true when listeners hear utterances in their own language with the segmental information removed. All the natural speech was low-pass filtered using Praat (Boersma & Weenink 2008) at a frequency cut-off of 400 Hz with 50 Hz smoothing, and the intensity was adjusted to 70dB. This will be called the 'filtered speech' condition, as opposed to the 'normal speech' condition.

There were 240 fragments from the Swedish stimuli (3 breaks x 1 speaker x 2 fragments x 20 items x 2 filterings). The Taiwanese stimuli contained 72 fragments (3 break x 2 speakers x 2 fragments x 3 repetition x 2 filterings).

2.3 Procedure

The experiments were presented in Praat in a quiet room. Fragments were presented one at a time over headphones at an average intensity of 70 dB. Each participant heard the fragments in the following order: 120 low-pass filtered Swedish fragments, 120 normal Swedish fragments, 36 low-pass filtered Taiwanese fragments and 36 normal Taiwanese fragments. The fragments within each section were presented in a randomized order. Subjects were told to express their judgment on a 5-point scale after hearing a spoken utterance fragment. If they thought there would be a strong break after the last word, they should respond with 5. If they felt that there would be no break after the last word, they should respond with 1. They could use the rest of the scale to mark in-between categories.

2.4 Results

Perceptual judgments

The data combined all tokens from both listeners and the results were analyzed in ANOVA with four within-subject factors: Boundary types (no break vs. weak break vs. strong break), Fragment size (long vs. short), Filtering (normal vs. low-pass filtered) and Language (Swedish stimuli vs. Taiwanese stimuli). Note that in terms of prosody, 'no break' refers to 'Word boundary' and 'strong break' refers to 'Utterance boundary' in both languages, and that 'weak break' in Swedish refers to a phrase boundary, while the 'weak break' in Taiwanese refers to 'Tone Sandhi Group boundary'. The results showed that these two Taiwanese listeners could make reliable judgments about upcoming boundaries from both the short and long fragments and from both low-pass filtered and normal speech. Only Boundary type ($F(2, 309)=75.45$, $p < .01$) had a significant effect on the judgments, while Fragment size, Filtering, and Language did not. A Tukey HSD post-hoc test showed that judgments regarding the three boundary types were all significantly different from one another ($p <.01$).

Interestingly, although Language didn't influence the judgments, the interaction between Boundary type and Language did have a slight but significant influence ($F(2, 612)=23.54$, $p = .04$). This was the only interaction effect observed and it is shown in Figure 2. Post-hoc tests of the interaction showed that the

significant difference was mainly due to the language difference at the strong break ($t(1, 206) = 1.97$, $p < .05$), not at the weak break or no break. Taiwanese listeners assigned bigger boundary strength to Taiwanese stimuli than to Swedish stimuli when they heard fragments with a strong break.

A closer look at the data from Taiwanese stimuli showed that the perceived boundary strength of a weak break was not significantly different from that of no break, which means that for the two Taiwanese listeners, there was no boundary strength difference between the Tone Sandhi Group boundary and the Word boundary. However, the boundary strength of the Utterance boundary was significantly different from that of the Tone Sandhi Group boundary ($t(1, 94) = 1.99$, $p < .05$) and from that of the Word boundary ($t(1, 94) = 1.99$, $p < .05$). As for the Swedish stimuli, the two Taiwanese listeners made a clear distinction in boundary strength – the three boundaries were significantly different from one another ($p < .05$).

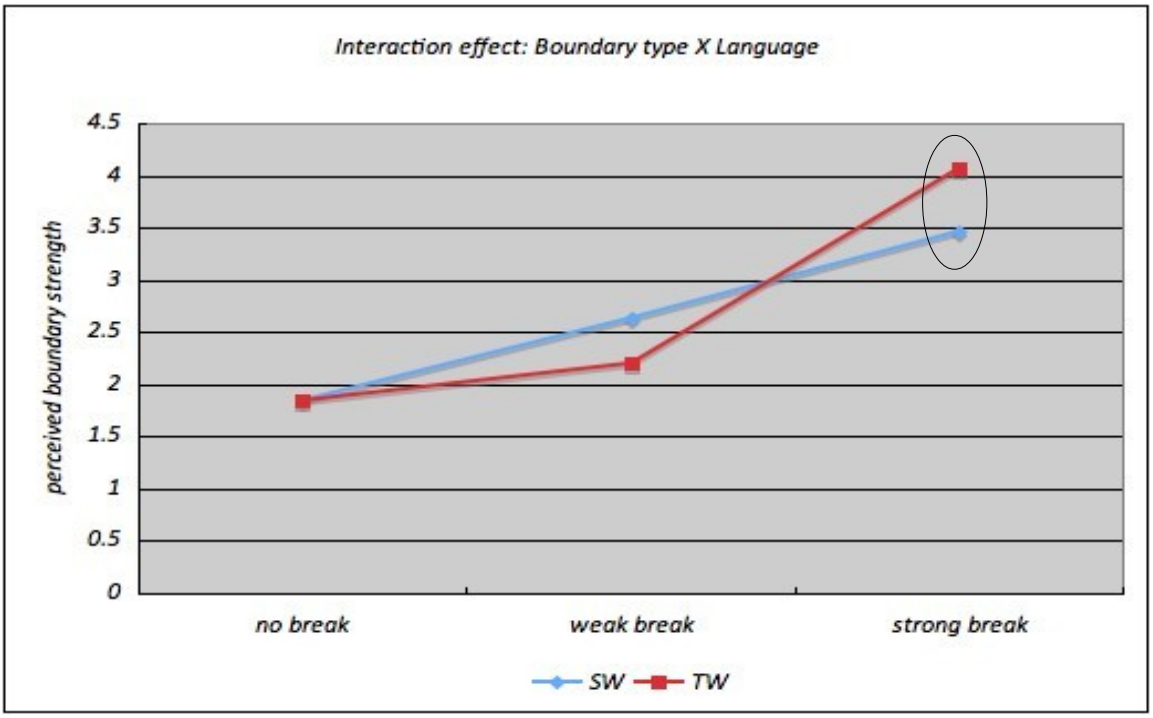

Figure 2. Interaction effect: Boundary type x Language. A post hoc test shows that the language difference is only significant at the strong break.

Figure 3 demonstrates the general judgment results. Data are grouped according to labeled Boundary strength and Fragment size and to Filtering and to Language. Note that there was no three-way interaction observed. The vertical dark dotted lines drawn on each boundary type set indicate the significantly different relations between the three boundary types. For example, Taiwanese listeners did not differentiate the Taiwanese weak break from no break when the fragments were short, regardless of whether they were presented normally or low-pass filtered. However, when they heard short fragments in Swedish, they did not separate weak breaks from strong breaks.

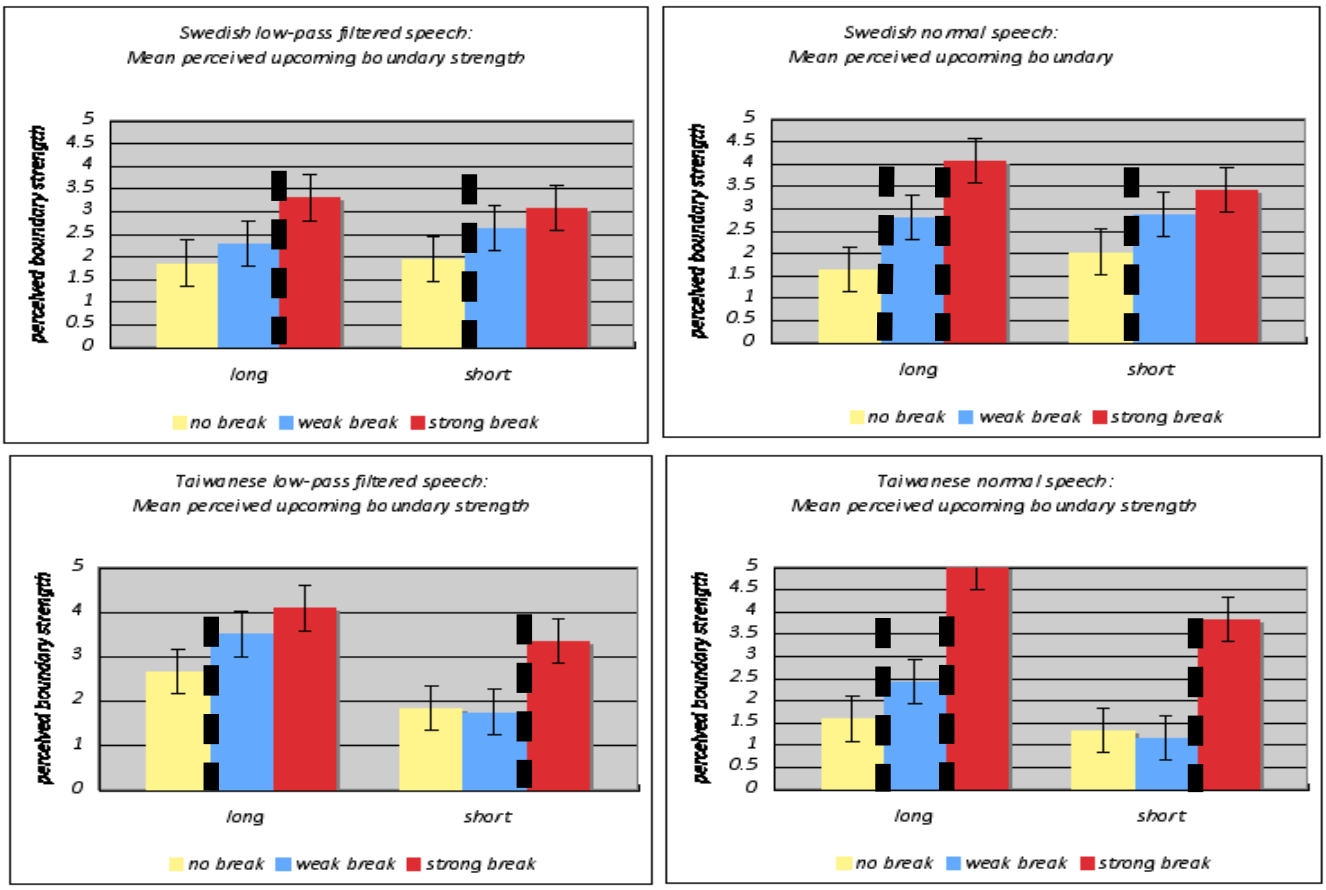

Figure 3. Mean perceived upcoming boundary strength. Vertical lines show significant difference.

Next, we restrict our focus to the data wherein the Taiwanese listeners heard Swedish "normal" speech in order to make a direct comparison with Carlson et al.'s study. Figure 4-1 is the perceived upcoming boundary strength for Swedish and American English speakers (from Carlson et al.); Figure 4-2 is the perceived upcoming boundary strength for Taiwanese listeners.

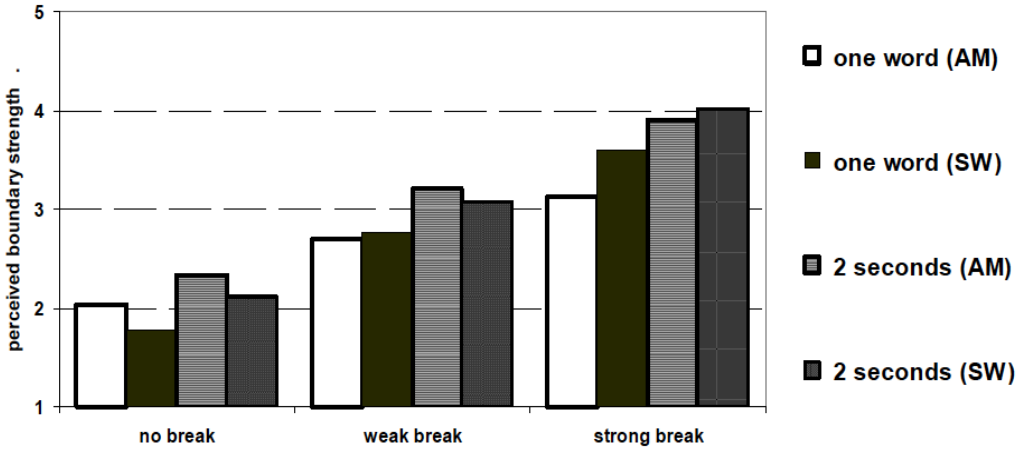

Figure 4-1. Swedish and American English listeners

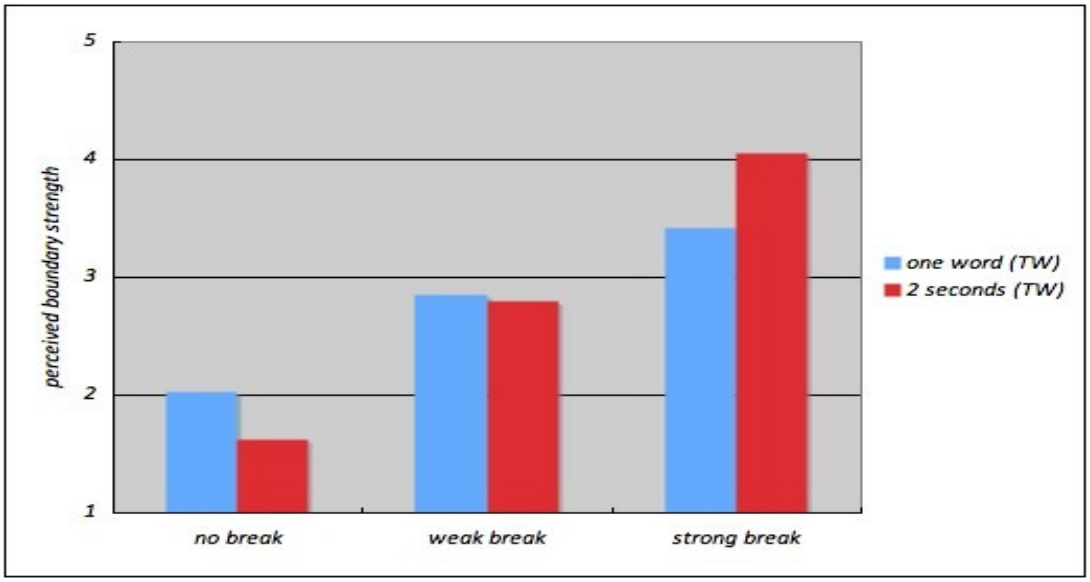

Figure 4-2. Taiwanese listeners. This is the same as the upper right graph in Figure 3.

In Carlson et al.'s study, it was shown that there was no significant difference between Swedish and English listeners. Their repeated-measures ANOVA with within-subjects factors of Boundary type and Fragment size revealed significant effects of Boundary Type and Fragment size on their subjects' perception of boundary strength. All three boundary types were significantly different from one another. In their study with Mandarin listeners (without statistical report), they found that Mandarin listeners differentiate the three boundaries only when presented with longer fragments.

Our results for Taiwanese listeners hearing the same Swedish stimuli showed that Fragment size was not an influential factor, but that there was a significant effect of Boundary type ($F(2, 110) = 43.9$, $p < .05$) on their subjects' perception of boundary strength. In addition, there was an interaction between our Boundary type and Fragment size ($F(2, 234) = 3.26$, $p < .05$), as shown in Figure 5. Like the results shown in the upper right graph in Figure 3, for the long 2-second fragments, all three boundaries are

significantly different from one another, whereas for the short one-word fragments, no break is significantly different from weak break and strong break. In addition, the longer fragments made it easier to detect the stronger break, and so there was a significant effect of Fragment size when the boundary was a strong break (t (1, 78) = 1.99, $p < .05$).

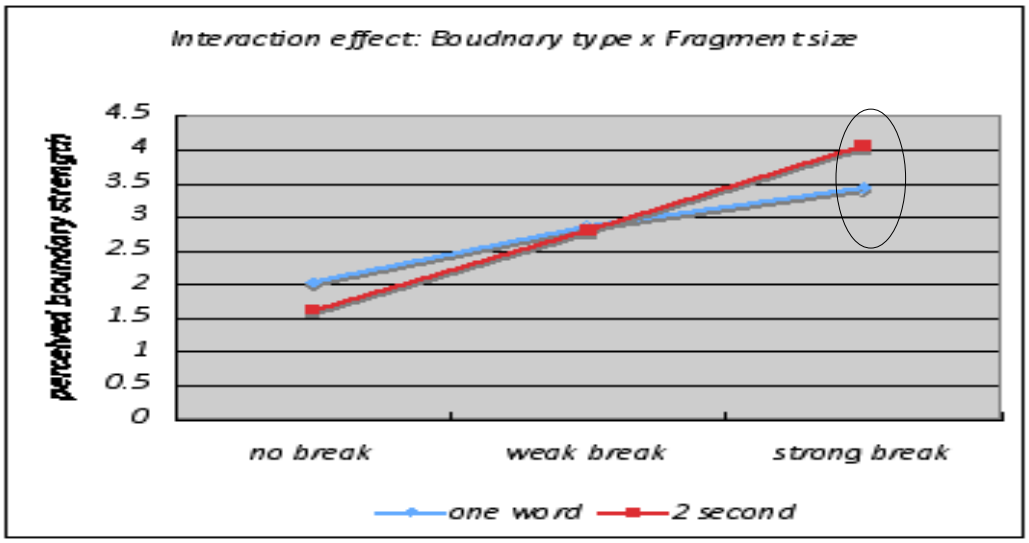

Figure 5. Interaction effect: Boundary type x Fragment size in Swedish stimuli

These results of this pilot study with two Taiwanese listeners showed that (a) Boundary type (with weak break = Tone Sandhi Group) was a factor that influenced the judgments: listeners were able to differentiate three boundary types in Swedish, and could only differentiate the strong break from the other two types of break. (b) Taiwanese listeners, like American English listeners, were able to hear the boundaries in Swedish in normal speech, especially in long fragments. This makes Taiwanese listeners similar to Mandarin listeners in Carlson et al.'s study because the latter can hear the difference only in long fragments. (c) Taiwanese listeners were able to hear some but not all Swedish boundaries even in low-pass filtered stimuli where only the rhythm and pitch information remained. (b) and (c) suggest that listeners can use prosodic information as primary cues with which to detect the boundaries, in addition to lexical and semantic information.

Acoustic and prosodic correlates

Next, several potential phonetic and prosodic cues of the Taiwanese stimuli were examined in an attempt to identify which features of the fragments might influence the listeners' judgments. Those cues included pause duration at the break, final word duration in proportion to the overall duration (final-word lengthening), several voice measures by VoiceSauce (Shue et al. 2011) and f0 slope.

The pause and the proportion of the final word duration in relation to the overall duration are shown in Figure 6. Paired t-tests showed that fragments with a strong break had a significantly longer pause at the boundary than fragments with a weak break or no break ($p < .05$). However, there was no pause duration difference between fragments with a weak break or with no break, which is as expected. On the other hand, the proportion of the final word to the overall fragment duration showed a significant main effect. The fragments with a strong break had a significantly longer final word duration ratio than the fragments with a weak break, which in turn had a significantly longer final word duration ratio than

fragments with no break ($p < .05$ in all comparisons).

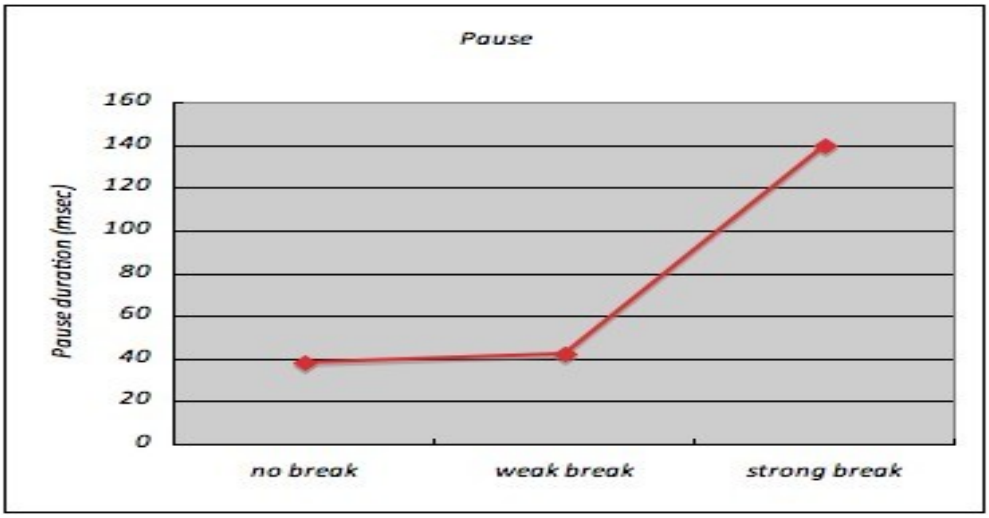

Figure 6-1. Pause duration

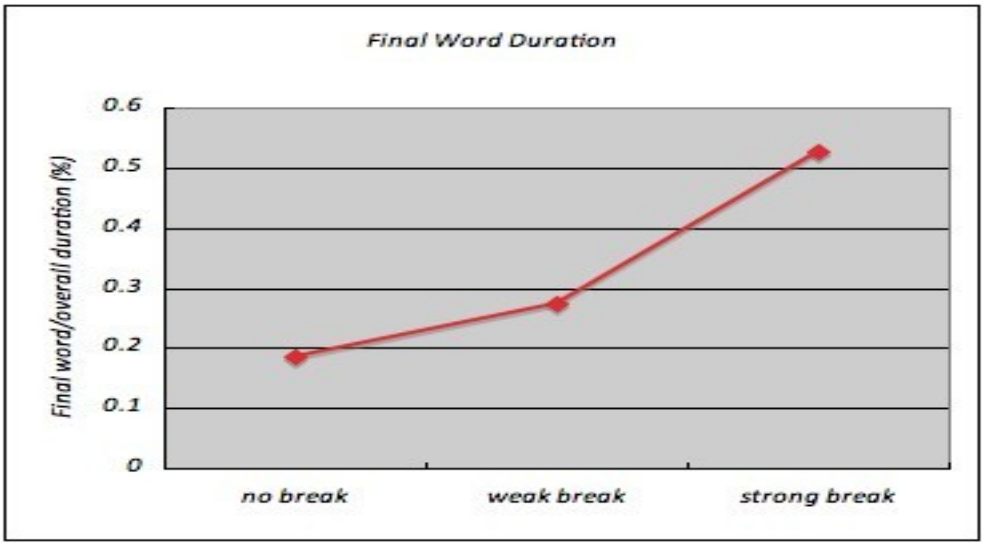

Figure 6-2. The proportion of final-word duration to overall duration

However, a moderate correlation was found between the pause duration and perceived boundary strength ($r = 0.63$; $p < .05$), whereas there was no significant correlation between the final word duration ratio and the perceived boundary strength. These results suggest that pause duration, but not pre-boundary final-word lengthening, is a reliable cue for distinguishing strong break from both weak and no break. Another cue must be used to make a distinction between weak and no break.

Several previous studies have shown that glottalization over the preceding syllable(s) may influence boundary perception. For instance, Carlson et al.'s study counted the number of stimuli which had a creaky sound and found that fragments with more creakiness were more likely to be judged to precede a strong boundary. In my pilot study, creakiness is examined at the vowel of the last syllable in each fragment using the formant-corrected measures H1-H2, H2-H4, H1-A1, H1-A2, H1-A3, CPP, Energy and HNR in VoiceSauce. Each fragment is normalized into nine equivalent periods of time (labeled as

01~09), and VoiceSauce averages the output of the above measures over each interval. Among all the measures, only the following were found to have a significant influence on listeners' ability to distinguish the three boundaries (the number in parentheses indicates at which interval(s) the measures are influential)[3]: H2-H4 (01), H1-A1 (03~09), H1-A3 (02~07), CPP (07~08), and Energy (03~08). In addition, there were significant correlations between the perceived boundary strength and the significant voice measures – over the first interval, H2-H4 (r = 0.53; p < .05) and over the seventh interval, H1-A3 (r = 0.52; p < .05), CPP (r = -0.64; p < .05) and Energy (r = -0.67; p < .05).

In addition, a series of Tukey HSD post hoc tests showed that H2-H4 (01) and H1-A3 (07) facilitate a distinction between fragments with no break and strong break. H1-A1 (07) and CPP (07) facilitate the distinction between strong break and no break, and between strong break and weak break. Energy appeared to be the most useful in that it facilitated a distinction between all three breaks. These results are displayed in Table 3. It appears that the last word preceding a strong break tends to have significantly higher H2-H4, higher H1-A1, higher H1-A3, lower CPP, and lower Energy. This kind of spectral difference indicates that breathier voice appears when the break is that of a bigger prosodic domain.

	strong - weak	weak - no	no - strong
H2-H4 (01)			✔
H1-A1 (07)	✔		✔
H1-A3 (07)			✔
CPP (07)	✔		✔
Energy (07)	✔	✔	✔

Table 3. Significant relations found at different prosodic domain boundaries
For all these significant differences, p < .05.

A close observation of the significance relation of the fragments with a weak break (which refers to the domain of our Tone Sandhi Group) reveals that using H1-A1(07) and CPP (07), we can separate the tone sandhi domain (weak break) from a larger domain, Utterance (strong break). And using Energy, we can separate the Utterance, Tone Sandhi Group and Word domains (no break). Observing the pattern of the results, we found that breathiness, rather than creakiness, was potentially used to distinguish a Tone Sandhi Group boundary from an Utterance boundary, but no cues appear to help differentiate a Tone Sandhi Group boundary from a Word boundary except Energy.

Previous studies (Lin 1988, Peng 1997, Myers 2008 and among others) found that duration and/or f0 are influential for distinguishing Taiwanese sandhi tone and base tone in perception. Therefore, I also examined the f0 slopes on the final word from each boundary condition. However, there was no significant effect of f0 slope, which indicates that f0 is less important than pause duration for Utterance boundary and voice quality for Taiwanese boundary detection, at least in these stimuli.

2.5 Discussion & Implication

The judgment results of this pilot study with two Taiwanese listeners are interesting. The results show

3 Note that these results might not be precise. Alpha correction should be done for multiple comparisons.

that when normal and filtered speech and long and short fragments were combined, listeners could predict the upcoming boundaries in Swedish very well, which suggests that the acoustic and prosodic cues, rather than lexical or semantic information, are the primary cues for the judgment. However, listeners were not able to make a distinction between no break and a weak break when they listened to the Taiwanese stimuli, which suggested that the Tone Sandhi Group domain and the Word domain are not easily distinguished.

Next, we looked at Taiwanese listeners' judgments on Swedish and Taiwanese stimuli only in normal speech and found that Boundary type still had a significant effect on the judgments. In addition, there is an interaction effect between Boundary type and Fragment size. Listeners are able to differentiate the three boundary types when presented with long fragments, but not able to tell a weak break from a strong break when presented with short fragments.

The results of the prosodic correlates in the Taiwanese stimuli show that pause duration, final-word lengthening and voice quality are all influential. First, the bigger prosodic domain has a longer pause duration. Second, the final-word duration ratio in fragments with a strong break is significantly longer than in fragments with a weak break, and the duration ratio in fragments with a weak break is significantly longer than in fragments with no break. Furthermore, H1-A1, CPP and Energy are important in differentiating the weak break (our Tone Sandhi Group domain) from the strong break. Among the three measures, Energy is the most sensitive correlate in that it could differentiate all three prosodic domains from one another.

3. Conclusion

It is known that Taiwanese tone sandhi domains end with a base tone, and that all tones before it have to undergo tone sandhi. In addition, previous studies suggest that base tone and sandhi tone with the same tone values are neutralized. This raises the question, how do listeners know that the tone they hear is a base tone, and thus the end of a tone sandhi domain? The results of the pilot study, both in terms of perception judgments and acoustic correlates, provide a possible answer to this question. It is found that Taiwanese listeners are able to correctly predict the occurrence of Swedish boundaries, but are not able to differentiate a Word boundary from a Tone Sandhi Domain boundary in Taiwanese. The fact that they can predict Swedish boundaries suggests that prosodic cues might be the primary cues for the detection of boundaries. However, it is still unclear why the Word domain and Tone Sandhi Group domain are not distinguishable. As for the acoustic analyses of the Taiwanese stimuli, we find that listeners use duration and spectral tilt information to disambiguate the domains, which means they are able to disambiguate the tones with these acoustic/prosodic cues.

Acknowledgments
I would like to thank Professor Rolf Carlson, Professor Julia Hirschberg and Professor Marc Swerts who generously shared their Swedish stimuli, provided their working papers and had insightful discussions with me. I am also thankful for the valuable comments I received from Professor Pat Keating, Professor Sun-Ah Jun, Professor Megha Sundara and Professor Jody Kreiman.

References

D. Barron, E. Shriberg & A. Stolcke. (2002). Automatic Punctuation and Disfluency Detection in Multi-Party Meetings Using Prosodic and Lexical Cues, ICSLP2002, Denver, USA.

P. Boersma & D. Weenink. (2008). Praat: doing phonetics by computer (Version 5.0.36) [Computer program]. Retrieved from http://www.praat.org/.

R. Carlson, J. Hirschberg & M. Swerts. (2005). Cues to upcoming Swedish prosodic boundaries: Subjective judgment studies and acoustic correlates. *Speech Communication 46, 3-4*, 326-333.

M. Chen. (1987). The syntax of Xiamen tone sandhi. *Phonology Yearbook 4*, 109-149.

L. Dilley & S. Shattuck-Hufnagel. (1996). Glottalization of word-initial vowels as a function of prosodic structure. *Journal of Phonetics 24,* 423-444.

L. Ferrer, E. Shriberg & A. Stolcke. (2002). Is the Speaker Done Yet? Faster and More Accurate End-of-Utterance Detection Using Prosody. *ICSLP-2002*, 2061-2064.

F. Ferreira. (1991) Effects of Length and Syntactic Complexity on Initiation Times for Prepared Utterances. *Journal of Memory and Language 30*, 210-233.

F. Ferreira (2003). Creation of Prosody During Sentence Production. *Psychological Review 100*, 233-253.

C. Fougeron & P. Keating. (1997). Articulatory strengthening at edges of prosodic domains. *Journal of Acoustical Society in America 101*, 3728-3740.

F. Grosjean. (1983). How long is the sentence? Prediction and prosody in the on-line processing of language. *Linguistics 21*, 501-529.

W. Hayashi, C. Hsu & P, Keating. (1998). Domain-initial strengthening in Taiwanese: a follow-up study. *UCLA Working Papers in Phonetics 96.*

M. Heldner & B. Megyesi. (2003). Exploring the prosody-syntax interface in conversation. *Proceedings of ICPhS 03.*

Y. Hsiao. (1991). Syntax, rhythm and tone: a triangular relationship. Taipei: Crane Publishing Co.

C. Hsu & S. Jun. (1996). Prosodic strengthening in Taiwanese: syntagmatic or paradigmatic? *UCLA Working Papers in Phonetics 96*, 69-89.

S. Jun. (1993). The Phonetics and Phonology of Korean Prosody. Unpublished Ph.D. Dissertation. The Ohio State University, Columbus, Ohio.

P. Keating, T. Cho, C. Fougeron & C. Hsu. (2003). Domain-initial strengthening in four languages. *Papers in LabPhon VI*, Cambridge University Press.

D. Klatt (1975).Voice onset time, frication, and aspiration in word-initial consonant clusters. *Journal of Speech and Hearing Research 18, 4,* 686-706.

J. Kreiman. (1982). Perception of sentences and paragraph boundaries in natural conversation. *Journal of Phonetics 10*, 163-175.

G. Kuo. (2010). Tone recognition of the two checked tones in Taiwanese. Poster, Acoustical Society of America*,* Cancún, México.

I. Lehiste (1979). Sentence boundaries and paragraph boundaries - perceptual evidence. In *The Elements: A Parasession on Linguistic Units and Levels* (Clyne et al. eds), Chicago: The Chicago

Linguistic Society, 99-109.
Y. Mo, J. Cole & E. Lee. (2008). Naïve listeners' prominence and boundary perception. *Speech Prosody* 2008, Campinas, Brazil.
J. Myers & J. Tsay (2001). Testing a production model of Taiwanese tone sandhi. *Proceedings of the Symposium on Selected National Science Council Projects in General Linguistics from 1998-2000.* 257-279. National Taiwan University, Taipei.
J. Myers & J. Tsay (2008). Neutralization in Taiwan Southern Min Tone sandhi. in Y. Hsiao et al. (eds.) *Interfaces in Chinese phonology: Festschrift in honor of Matthew Y. Chen on his 70th birthday* (pp. 47-78). Language and Linguistics Monograph Series Number W-8. Taipei, Taiwan: Academia Sinica.
M. Nespor & I. Vogel. (1986). Prosodic Phonology. Dordrecht: Foris.
H. Pan. (2003). Prosodic hierarchy and nasalization in Taiwanese. *Proceeding of the 15th ICPhS,* 575-578.
H. Pan. (2006). Boundaries and Tonal Articulation in Taiwanese Min. *Proceedings of Speech Prosody 2006*, 51.
S. Peng & M. Beckman. (2003). Annotation conventions and corpus design in the investigation of spontaneous speech prosody in Taiwanese. *Proceedings of SSPR 2003*, 17-22.
Pierrehumbert & Talkin (1991) Lenition of /h/ and glottal stop. Papers in Laboratory Phonology II, Cambridge University Press. Cambridge UK. 90-117.
S. Shen (1992). A pilot study on the relation between the temporal and syntactic structures in Mandarin. In *Journal of International Phonetic Association* 22, 35-43.
Y.-L. Shue, P. Keating, C. Vicenik. & K. Yu. (2011). VoiceSauce: A program for voice analysis. *Proceedings of ICPHS 2011*.
M. Swerts. (1997). Prosodic features at discourse boundaries of different strength. *Journal of Acoustical Society of America 101, 1*, 514-521.
M. Swerts & R. Geuykens. (1994). Prosody as a marker of information flow in spoken discourse. *Language Speech 37*, 21-43.
C. Tseng. & C. Chang. (2008). Pause or No Pause? - Prosodic Phrase Boundaries Revisited. *Tsinghua Science and Technology 13 (4)*, 500-509.
C. Wightman, S. Shattuck-Hufnagel, M. Ostendorf & M. Price. (1992). Segmental durations in the vicinity of prosodic phrase boundaries. *Journal of the Acoustical Society of America 91*, 1707-1717.