

Neutralization in Taiwan Southern Min Tone Sandhi*

James Myers and Jane Tsay
National Chung Cheng University

Phonologists have long assumed that tone sandhi in Southern Min is neutralizing, both across prosodic positions (e.g., juncture *yangqu* vs. context *yinping*) and within prosodic positions (context *yinping* vs. context *yangping*). Yet phonetic studies of other phonological processes often find that neutralization is only incomplete. Thus we had native speakers of Taiwan Southern Min read aloud pairs of sentences identical except for morphemes expected to be neutralized by tone sandhi. Speakers were recorded in two pragmatic conditions to test if the degree of neutralization is under speaker control. Neither across-position comparisons nor within-position comparisons showed statistically significant differences in f_0 height or slope, though there were effects on syllable duration. This null result is consistent with complete neutralization for a number of reasons: the statistical method used is the most powerful and reliable currently available, f_0 slopes did not preserve the putatively underlying tone contours, there was no effect of pragmatic context on f_0 , and the raw differences in f_0 were much smaller than those found in other tone languages where neutralization has been claimed to be incomplete. Tone sandhi in Taiwan Southern Min thus seems to involve categorical phonological units, just as has long been assumed.

Key words: Southern Min, tone sandhi, incomplete neutralization

1. Introduction

Theoretical phonologists generally assume that phonological representations and processes involve categorical representations composed of discrete units like features and phonemes, an assumption that Port & Leary (2005) consider core to what they call “formal phonology.” Yet it has long been known that speakers will sometimes give

* Basic results for Experiment 1 were presented at the Phonetics Workshop at Chiao Tung University, Taiwan, in March 2002. The experiment was conducted with the help of Joyce H.-C. Liu. We would like to thank Yueh-chin Chang, Mirjam Ernestus, Elliott Moreton, Shu-hui Peng, Natasha Warner, and anonymous reviewers of a previous incarnation of this paper for insights and useful references, and Jan Charles-Luce for providing the paradigm that we followed. Research was funded by National Science Council (Taiwan) grant NSC 90-2411-H-194-023.

instrumentally distinct pronunciations to words otherwise thought to be phonologically identical, a phenomenon known as incomplete neutralization. Classic and recent examples in the growing literature on incomplete neutralization include final devoicing in German (Charles-Luce 1985, Port & O'Dell 1985, Port & Crawford 1989), Catalan (Charles-Luce 1993) and Dutch (Warner, Jongman, Sereno & Kemps 2004, Ernestus & Baayen 2006), and flapping in American English (Charles-Luce 1997, Charles-Luce, Dressler & Ragonese 1999, Charles-Luce & Dressler 1999); more complete lists of studies can be found in these sources and in the reviews by Manaster Ramer (1996), Port (1996), and Port & Leary (2005). If we include sociolinguistic near-mergers (which would not be accepted by some, such as Manaster Ramer 1996), we can add cases such as vowel tenseness/laxness in Southwest American English (Di Paolo & Faber 1990, Faber & Di Paolo 1995) and others summarized by Labov, Karan & Miller (1991) and Labov (1994).

Incomplete neutralization is also known to be under the partial control of speakers, since it systematically varies depending on discourse-specific factors. Thus when explicitly asked to produce minimal pairs, as for an attentive listener, speakers tend to neutralize less (Port & Crawford 1989, Charles-Luce 1997); when producing target words in semantically biasing contexts they tend to neutralize more (Charles-Luce 1993, 1997); when reading items aloud, they may produce subphonemic distinctions that derive from purely orthographic distinctions not reflected in the underlying phonological representations (Warner et al. 2004, Warner, Good, Jongman & Sereno 2006); and speakers actively adjust the degree of neutralization depending on the particular items that have been encountered during the course of an experiment (Ernestus & Baayen 2006).

What remains highly controversial is what such phenomena imply for phonological theory. At one extreme are those who have argued that the existence of incomplete neutralization (along with other phenomena) reveals that knowledge of phonological patterns must be phonetically detailed, rather than involving the categorical symbols of traditional phonological theory (Bybee 1994, Port 1996, Port & Leary 2005, Yu 2007). At the other extreme are those who have claimed that any effects that have been found are uninteresting artifacts of reading pronunciations (Fourakis & Iverson 1984, Manaster Ramer 1996).

1.1 Incomplete neutralization and postlexical phonology

However, a compromise of sorts exists, which is that incomplete neutralization, while genuine, is restricted to a late stage of phonological production processing, in a module variously termed phonetic implementation or postlexical phonology (in the sense of Kiparsky 1982). This modular view of the phonology-phonetics interface is quite traditional, but perhaps the most thorough investigation of its implications for the

(incomplete) neutralization issue can be found in Zsiga (1993, 1995, 1997). For example, Zsiga (1995) compared the phonetic behavior of lexical and postlexical palatalization in English; the former is exemplified by *confess-confession* (related by semi-productive derivational morphology), the latter by *confess-confess your* (a word vs. a compositional phrase with no obvious claim to lexical storage). She found, from both acoustic and electropalatographic data, that the palatals in derived words like *confession* showed no difference from underlying palatals like that in *fresh*, while the palatals in phrases like *confess your* did, since they began like /s/ and only became palatal-like at the end of the segment.

This modular approach to incomplete neutralization is immune to standard attacks from both of the extreme camps. On the one hand, the discovery that German final devoicing, say, shows incomplete neutralization does not really “pose a threat to phonological theory” (in the words of Port & Crawford 1989:257) if one simply moves the final devoicing process from the lexical to the postlexical module. Unlike lexical patterns, which consist of the distributions and alternations of categorical units (e.g. the autosegmental features of Zsiga’s model), and hence should not be capable of showing incomplete neutralization, postlexical processes need not produce lexically distinctive outputs (Kiparsky 1982), nor indeed phonetically categorical ones (Kiparsky 1985). On the other hand, the observation that discourse-specific factors affect the degree of neutralization, used by skeptics to argue that speakers are simply “making an artificial effort to distinguish homophones” (Manaster Ramer 1996:487), can be turned into an argument in favor of the postlexical hypothesis, since many prototypical postlexical processes are quite variable and discourse-dependent. Word-final deletion of /t/ and /d/ in American English, for example, displays postlexical diagnostics such as phonetic gradience (Browman & Goldstein 1990) and sensitivity to word-external information (Guy 1980); it also varies systematically in its rate of application (Guy 1980).

A fundamental problem with the modular approach is the well-known fact that the border between the lexical and postlexical modules has proven to be quite fuzzy. The diagnostics given to distinguish one from the other (e.g. Kaisse & Hargus 1993:16-17) do co-occur much more often than one would expect to happen by mere accident, but they often disagree. For example, moving German final devoicing out of the lexical phonology to account for its incomplete neutralization is not entirely unproblematic, since this process has been argued to be sensitive to morphological structure, a lexical diagnostic (Rubach 1990). Similar points can be made about American English flapping (Steriade 2000) and American English t/d-deletion (Guy 1991). Borowsky (1993), Bybee (2000) and Steriade (2000) collect many further examples of nonphonemic or even phonetically gradient patterns that are sensitive to morphological structure or lexical frequency.

In this paper we address another classic problem for the lexical/postlexical distinction, namely what has been variously called P1 rules (Kaisse 1985), precompiled lexical phonology (Hayes 1990), or somewhat more neutrally, lexicalized phrasal phonology (Tsay & Myers 1996). In contrast to the problematic cases mentioned above, which display essentially postlexical behavior (notably gradience), lexicalized phrasal phonology apparently exhibits only one postlexical property, namely being phrasal rather than word-internal. For example, in Kimatuumbi (Odden 1987, 1990, Hayes 1990), a word-final long vowel shortens before a phrasal complement, a process that is also sensitive to the word's morphological structure and lexical phonological history (i.e. whether the long vowel is underlying or derived by lexical rule). As with prototypical lexical phonology, the pattern in Kimatuumbi is described as manipulating lexically distinctive contrasts (here, vowel duration). Lexicalized phrasal phonology thus has an important phonetic implication that to our knowledge has never been tested: being inherently lexical rather than a phonetic implementation process, it should not show incomplete neutralization. We will refer to this as the categoricity hypothesis.

1.2 Taiwan Southern Min tone sandhi

The example of lexicalized phrasal phonology that we examine in this paper is tone sandhi in Taiwan Southern Min (a cover term for the varieties of Southern Min Chinese spoken in Taiwan). The variety we describe is that spoken in Chiayi County in southern Taiwan; historically it is derived from the Southern Min spoken in Zhangzhou and Xiamen (also called Amoy or Hagu) in Fujian province, across the Taiwan Strait (Yuan 1960, Cheng 1968, Ting 1970).

Southern Min tone sandhi is a phrase-level rather than word-internal phenomenon, as has been made well-known by work on its interaction with syntactic structure (Chen 1987, J.-W. Lin 1994). The pattern involves alternations between tones as they appear in juncture position (i.e. the right edge of a prosodic constituent called a tone group) and in context position (elsewhere). Tone groups are primarily defined syntactically, aligning at the right edge with the right edge of syntactic phrases, unless the phrases are in particular syntactic relations with other elements (namely if the phrase is lexically governed, according to J.-W. Lin (1994)). Tone groups can also be defined purely prosodically, particularly in poetry (Hsiao 1991).

The tone alternations themselves are quite complex, affecting each of the seven contrastive tones differently. Particularly notable (and rather rare even among Sinitic tone sandhi systems, as the survey in Chen (2000) reveals) is the fact that the alternations do not create nonlexical tones (i.e. they are structure-preserving, in the terminology of Kiparsky (1982)). Table 1 lists the seven lexical tone categories with their phonetic tone

values in juncture and context positions using the traditional 5-point numeric scale (Chao 1930), where 1 = lowest and 5 = highest. We use the traditional philological names for the tone categories rather than numerical tone category labels since such labels differ considerably across authors. Note that no surface phonetic tone appears in context position unless it also appears in juncture position. The first five tone categories listed are often called long tones (appearing on open syllables and syllables ending in sonorants) and the last two are short tones (appearing on syllables ending in /p/, /t/, /k/, or glottal stop). It has been argued that the short tones are merely long tones that happen to appear on “short” syllables (Cheng 1968); we avoid this issue by focusing here on the long tones only. Figure 1 illustrates how the tone sandhi alternations of these five long tones form the notorious Southern Min tone circle (Bodman 1955).

Table 1 shows Taiwan Southern Min tone categories in juncture (tone group final) and context (elsewhere) positions. Transcriptions for surface phonetic tones follow Peng (1997) using the 5-point scale of Chao (1930). The tone categories are listed somewhat out of their traditional order so that the three which can be realized as a mid level [33] (*yinping*, *yangping*, *yangqu*) are grouped at the top.

Table 1

Tone categories	Juncture form	Context form
<i>yinping</i>	si ⁵⁵ “poetry”	si ³³ -bun ²⁴ “poetry and prose”
<i>yangping</i>	si ²⁴ “time”	si ³³ -kan ⁵⁵ “time span; time”
<i>yinshang</i>	si ⁵¹ “die”	si ⁵⁵ -lan ²⁴ “dead people”
<i>yinqu</i>	si ²¹ “four”	si ⁵¹ -tiam ⁵¹ “four o’clock”
<i>yangqu</i>	si ³³ “temple”	si ²¹ -tsiŋ ⁵⁵ “temple monk”
<i>yinru</i>	sik ²¹ “color”	sik ⁵³ -ts ^h ai ⁵¹ “color”
<i>yangru</i>	sik ⁵³ “ripe”	sik ²¹ -te ²⁴ “baked tea”

Figure 1 represents the Southern Min long tone circle as found in our speakers’ variety of Taiwan Southern Min. Arrows follow the traditional assumption that context tones are derived from juncture tones.

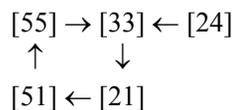


Figure 1

Note that we are careful to label the alternate tone forms “juncture tone” and “context tone”, rather than the more commonly used “basic tone” and “sandhi tone”. This

is because the nature of Taiwan Southern Min tone sandhi makes it difficult to be sure that the juncture tone truly is basic. The claim that it is depends on the assumption that the alternations must involve rewrite rules, but several researchers have challenged this assumption (Hsieh 1970, 1975, 1976, S. H. Wang 1995, Chen 1996, Tsay & Myers 1996, Peng 1998, Moreton 1999, Tsay 2002). Simplifying somewhat, their general consensus is that Taiwan Southern Min tone sandhi is better viewed as a set of essentially arbitrary alternations between stored allomorphs. If so, Taiwan Southern Min tone sandhi represents a particularly strong example of lexicalized phrasal phonology.

Their arguments can be summarized as follows. First, unlike prototypical postlexical phonology (or most lexical phonology for that matter), the Taiwan Southern Min tone sandhi alternations do not follow any overall pattern; the ad hoc exploitation of complex technical devices is required to reduce the five long-tone alternations to one rule by W. S.-Y. Wang (1967), to three rules by Yip (1980), or to three different rules by Tsay (1994). The difficulty is not solved by using a constraint-based analysis, since as Moreton (1999) shows, Optimality Theory is incapable of handling circular chain shifts. Second, the tone sandhi alternations seem to be at best only semi-productive. When given novel word combinations or combinations involving nonce words (phonotactic or tonotactic gaps), native speakers rarely produce the correct tone sandhi alternations with accuracy rates above 80%, dropping below 10% for some tasks and tone categories (Hsieh 1970, 1975, 1976, Tseng 1995, S. H. Wang 1995, see also H.-B. Lin 1988:126). This makes Taiwan Southern Min tone sandhi look quite different from automatic processes like German final devoicing or American English flapping, and much more similar to word-internal morphophonological patterns, which also show low rates of application with novel items (e.g. Ohala 1974, Steinberg & Krohn 1975, Myerson 1978). Third, mixing tone sandhi patterns from different varieties of Southern Min is not uncommon, giving rise to idiolects that show distinct alternations for morphemes of the same tone class (Hsieh 1976); this should not happen if the application of tone sandhi involved general postlexical rules. Finally, in a concept formation task, Peng (1998) found that participants tended to class context *yinping* [33] with juncture *yangqu* [33] rather than with its putative underlying form (juncture *yinping* [55]), revealing a tendency to take the surface tone as basic (although curiously, context *yangping* [33] did tend to be classed with juncture *yangping* [24]).

Instrumental phonetics can provide important evidence bearing on this issue, since descriptions of Southern Min tone sandhi have always assumed that the alternations are categorical in two ways: both across prosodic positions (e.g. juncture *yangqu* [33] vs. context *yinping* [33], among other possible comparisons) and within context position (i.e. with the pair context *yinping* [33] vs. context *yangping* [33]). This assumption is implicit in Bodman's tone circle and in most generative phonological analyses. It is also

assumed in phonetic studies such as that of Peng (1997), who states that the sandhi alternations are “categorical” (p.375), although she gives no phonetic justification for this assumption because her focus lies elsewhere. Naive native speakers seem to share this assumption, reporting no differences when given minimal pairs to consider and repeat to themselves (Tseng 1995).

1.3 Incomplete neutralization and Sinitic tone sandhi

The question of incomplete neutralization in tone sandhi has been addressed more thoroughly in another Sinitic language, Mandarin, than in Southern Min. Like Southern Min tone sandhi, Mandarin tone 3 sandhi has traditionally been assumed to be categorical; Chao (1948) describes it as the replacement of tone 3 (a low dipping tone in citation form) by tone 2 (rising) when before another tone 3. According to the thorough examination of phonetic studies of this pattern given in Myers & Tsay (2003), it seems that contrary to Chao (1948), sandhi tone 3 is systematically lower than lexical tone 2 in the speech of native speakers of Beijing Mandarin, consistent with the hypothesis that it preserves aspects of its citation form (Zee 1980, Kratochvil 1986, Xu 1993; see also the noninstrumental observations in Hockett 1947, and Martin 1957). However, there are unresolved questions about the perceptibility of this difference and whether it is found in other varieties of Mandarin. In an often-cited perceptual study with Beijing Mandarin speakers, W. S-Y Wang & Li (1967) concluded that lexical tone 2 and sandhi tone 3 are homophonous. Studies with speakers of Taiwan Mandarin (the variety of Mandarin spoken in Taiwan and strongly influenced by Taiwan Southern Min) generally do not find differences between sandhi tone 3 and lexical tone 2, whether in production (Chang & Su 1994, Fon 1997) or in perception (Chang & Su 1994, Peng 2000). Peng (2000) did find that the mean fundamental frequency (f_0) of sandhi tone 3 was significantly lower than that of lexical tone 2, but by a much smaller amount than has been reported for Beijing Mandarin (2.3 Hz, as compared with the 17.5 Hz found by Zee (1980)). Myers & Tsay (2003) suggest that this apparent dialect difference, where Mandarin tone sandhi in Taiwan is categorical while Mandarin tone sandhi in Beijing is not, provides indirect support for the claim that tone sandhi in Taiwan Southern Min is truly categorical, assuming that this property is carried over to the Mandarin acquired by Southern Min speakers as a second language.

In a recent study, Yu (2007) discusses a different sort of subphonemic tonal phenomenon in Cantonese, this time involving the near merger of distinct lexical categories (see e.g. Labov 1994) rather than incomplete neutralization due to morphophonemic alternation (e.g. tone sandhi). Impressionistic studies had long noted an apparent difference in tonal contour between phonemically specified [35] tones and the diachronically

derived [35] that appears, among other places, in nominalizations of certain verbs (e.g. [sou³⁵] “a broom”, related to [sou³³] “to sweep”). Yu’s instrumental analysis demonstrated that indeed, the latter have an f_0 , on average, 4.8 Hz higher than the former, though without a difference in contour shape. Moreover, as is typical for near mergers, native speakers listening to isolated syllables cannot reliably distinguish the two categories. Although this case doesn’t shed light directly on tone sandhi, it does seem to demonstrate that it is indeed possible for tonal systems to involve subphonemic lexical contrasts.

Turning now to tone sandhi in Southern Min, here there seem to be neither impressionistic observations nor instrument evidence posing any serious challenge to the assumption of categoricity. Cheng (1968:39-40) and Du (1988) provide phonetic data on tones in both juncture and context positions, but unfortunately not in a form that can be reanalyzed to address the categoricity hypothesis. As already noted, Peng (1997) simply assumed that tone sandhi is categorical, and so in her analysis of prosody and tonal coarticulation treated tones like juncture *yangqu* and context *yinping* as if they were phonologically identical mid level [33] tones.

H.-B. Lin (1988) does address the question of neutralization in the production of Taiwan Southern Min tone sandhi, basing the analysis on the production and perception of nonsense disyllables. However, speakers were mixed from different dialects (one of the six speakers had a different tone sandhi system from the rest) and did not systematically control the segmental environment; the use of nonsense syllables rather than real words may also have given rise to somewhat artificial productions. Interpretation of the results is also complicated by the fact that they are only presented graphically (juncture tones for /do/ are given in Figure 2.5, p.50, and context tones for /si/ in Figure 2.6, p.52). Interestingly, in the graphs, context *yinping* [33] appears to be approximately 8 Hz higher than context *yangping* [33], consistent with incomplete neutralization if context is derived from juncture (juncture *yinping* [55] is higher than juncture *yangping* [24]). Nevertheless, in an identification task, error rates were quite high (over 50%) (Table 4.3, p.119); another identification experiment showed greater accuracy (Table 4.7, p.128), but only because the stimuli contained many nonapplications of the tone sandhi alternations, artificially making lexical tones distinct in context position (Table 4.6, p.126).

Chang (1988) examined the apparent neutralization within context position (*yinping* vs. *yangping*) in disyllabic lexical compounds. Separate comparisons for f_0 at five different points within the target syllable showed no significant differences. Intriguingly, the nonsignificant difference in overall f_0 , though extremely small (0.89 Hz), was again in the expected direction if incomplete neutralization had occurred: context *yinping* was higher than context *yangping*. However, the sets of context *yinping* and context *yangping* items were not matched in size, in segmental environment, or in tonal environment (voiced onsets and adjacent tones can both affect f_0 contours), making this study, like

that of H.-B. Lin (1988), less than conclusive.

Tsay, Charles-Luce & Guo (1999) and Tsay & Myers (2001) together describe four experiments that explicitly attempted to test the prediction of Tsay & Myers (1996) that Taiwan Southern Min tone sandhi, as lexicalized phrasal phonology, should be fully neutralizing both across prosodic positions (juncture vs. context) and within position (context *yinping* vs. context *yangping*). Larger numbers of speakers were used than in previous studies and materials were carefully controlled for both segmental and tonal environment. Moreover, acknowledging the phrasal nature of tone sandhi, matched items were whole sentences rather than isolated compounds. Across-position comparisons consistently showed large and significant differences in syllable duration (juncture tones being over 40 ms longer than context tones), but differences in f_0 were rare, and always attributable to phrase-final lowering rather than incomplete neutralization. For example, in the second experiment of Tsay et al. (1999), juncture *yinshang* [51] was lower at the onset than context *yinqu* [51], even though the juncture form of *yinqu* is [21]; if aspects of an underlying [21] were preserved in context *yinqu*, its onset should not have been higher as it in fact was. Two of the experiments that examined the within-position comparisons (i.e. context *yinping* vs. context *yangping*) found no significant differences between these tones in duration or f_0 at any of the points measured, but in one of these experiments, there was again a nonsignificant tendency for context *yinping* to be higher than context *yangping* (the other seemed to show a trend the opposite way). However, the second experiment in Tsay & Myers (2001), using more speakers than the three previous ones (30 vs. 10 or fewer) found higher f_0 values at all measurement points for context *yinping* (mean 2.7 Hz), and this time the differences were highly significant ($p < 0.01$). This experiment, perhaps the best controlled in the current literature, seems to imply that at least one aspect of Taiwan Southern Min tone sandhi is in fact incompletely neutralizing. Nevertheless, as we note below, there are reasons to want further confirmation of these results.

This second experiment in Tsay & Myers (2001) also examined the effect of discourse factors on Taiwan Southern Min tone sandhi; as lexicalized phrasal phonology, tone sandhi was expected to be insensitive to such factors. Building on the methods of Charles-Luce (1997), neutralization was encouraged by first putting speakers alone in the recording room to read aloud target sentences that were randomly mixed among fillers; afterwards they read the same sentences, but now arranged in minimal pairs with an attentive listener present. In spite of the overall f_0 differences noted above, no interaction was found with factor of absence/presence of a listener. This suggests that unlike American English flapping (which Charles-Luce (1997) did find to be affected by this same pragmatic factor), Taiwan Southern Min tone sandhi is not subject to discourse-dependent variation.

Problems with the studies of Tsay, Charles-Luce & Guo (1999) and Tsay & Myers (2001) leave room for improvement, however. The lack of significant differences for f_0 in most of the experiments could be due to the small number of speakers or to excess variability in the data caused by averaging across categories that should have been kept distinct. For example, the across-position comparisons in the second experiment of Tsay & Myers (2001) compared juncture *yangqu* with both context *yinping* and context *yangping*, since the authors prematurely assumed that the lack of a within-position difference had already been established by earlier experiments. Criticizing this experiment from the opposite perspective, it is possible that the incomplete neutralization observed in context *yinping* and context *yangping* could have been due to the lesser mastery of the phonological system by the college-aged speakers; Taiwan Southern Min is less commonly used among this generation than among older speakers (Huang 1993). There was also a technical statistical problem. Like most research in phonetics and psycholinguistics, the studies by Tsay and colleagues analyzed the results with repeated-measures analysis of variance (ANOVA), but this technique is known to produce artificially low p values if factors have more than two levels (Max & Onghena 1999), as did the factor representing measurement points within the tone contours. More on this statistical problem, and its solution, will be described in §2.1.3.

To examine the question of neutralization in Taiwan Southern Min tone sandhi more carefully, we carried out a replication of the second experiment of Tsay & Myers (2001) using improved materials and older (and we hoped more fluent) speakers.

2. Experiment

This experiment had two primary goals. The first was to test the categoricity hypothesis (i.e. that Taiwan Southern Min tone sandhi is indeed neutralizing) using more carefully controlled materials and more fluent speakers than in previous studies. The categoricity hypothesis predicts no difference in f_0 for juncture *yangqu* vs. context *yinping*, and no difference for context *yinping* vs. context *yangping*; all of these tones should behave as realizations of the same phonological category [33]. The categoricity hypothesis has nothing to say about the durations of the syllables containing these tones, however, since variation in duration is a separate phenomenon, applied in production after tone categories have been selected and their f_0 trajectories prepared for articulation. Based on previous research (e.g. Peng 1997), we expect that durations should be significantly longer in juncture position than in context position, and that syllables with context *yinping* and context *yangping* tones should not differ in duration.

The second goal was to examine a further prediction of the categoricity hypothesis, namely that speakers should be unable to adjust the distinctiveness of the neutralized

tone categories. In particular, it should not matter if speakers produce items in random order alone in a room (a situation that would allow or even encourage neutralization), or if they produce items in minimal pairs with the explicit instruction to distinguish the pairs for a listener (a situation that should reduce neutralization, if it is truly under speaker control).

2.1 Methods

2.1.1 Materials

Seven pairs of sentences were created that differed from each other in a single morpheme (represented by a single orthographic character); these target morphemes all had tones standardly transcribed as a mid level tone [33] (see Table 2 and 3). Three pairs represented the comparison across prosodic position: in one sentence the target morpheme had juncture *yangqu* tone (1a, 2a, 3a in Table 2), while in the other the target morpheme had context *yinping* tone (1b, 2b, 3b in Table 2). Four pairs represented the within-position comparison, contrasting context *yinping* tone (1a, 2a, 3a, 4a in Table 3) and context *yangping* tone (1b, 2b, 3b, 4b in Table 3).

Table 2: Sentence pairs for the across-position comparisons (juncture *yangqu* vs. context *yinping*). The highlighted syllables in the (a) sentences are juncture *yangqu*, and in the (b) sentences they are context *yinping*. Syntactic constituents relevant to the definition of tone groups are marked. Commas appear in the same locations as they appeared in the versions written in Chinese characters used in the experiments. COP represents a copula.

- 1a tse⁵⁵ si²¹ [a³³**kim**³³]_{NP}, sə⁵¹ a²¹ kap⁵³ a³³ hĩã⁵⁵
 this COP Aunt Sister-in-law and Elder Brother
 ‘These are Aunt, Sister-in-law and Elder Brother.’
- 1b tse⁵⁵ si²¹ [a³³**kim**³³-sə⁵¹ a²¹]_{NP} kap⁵³ a³³ hĩã⁵⁵
 this COP (name) Sister-in-law and Elder Brother
 ‘These are Sister-in-law Akim and Elder Brother.’
- 2a tse⁵⁵ si²¹ [lim³³**sun**³³]_{NP}, ts^hun³³ kiau⁵⁵ kap⁵³ sə⁵¹ hua²⁴
 this COP (name) (name) and (name)
 ‘These are Limsun, Tshunkiau and Soohua.’
- 2b tse⁵⁵ si²¹ [lim³³**sun**³³ ts^hun³³ kiau⁵⁵]_{NP} kap⁵³ sə⁵¹ hua²⁴
 this COP (name) and (name)
 ‘These are Limsuntshunkiau and Soohua.’

- 3a tse⁵⁵ si²¹ [tɿ³³]_{NP}, baŋ²¹ kap⁵³ ũa⁵¹
 this COP chopstick meat and bowl
 ‘These are chopsticks, meat and bowl.’
- 3b tse⁵⁵ si²¹ [tɿ³³-baŋ²¹]_{NP} kap⁵³ ũa⁵¹
 this COP pig meat and bowl
 ‘These are pork and bowl.’

Table 3: Sentence pairs for the within-position comparisons (context *yinping* vs. context *yangping*). The highlighted syllables in the (a) sentences are context *yinping*, and in the (b) sentences they are context *yangping*. Syntactic constituents relevant to the definition of tone groups are marked. COP represents a copula, MOD represents a modifier marker, and PASS represents a function word appearing before the agent in a passive construction. The proper names both glossed “Tsinkesian” in 3a-b are actually distinct in their first characters. The sentence pair 1a-b was not included in the analysis due to insufficiently high fluency scores for sentence 1a (see Procedures section for Experiment 1).

- 1a gua⁵⁵ be⁵⁵ [kuan³³ tə⁵⁵ a⁵¹]_{VP} kap⁵³ kuan³³ i⁵⁵ a⁵¹
 I want donate table and donate chair
 ‘I want to donate tables and chairs.’
- 1b gua⁵⁵ be⁵⁵ [kuan³³ tə⁵⁵ a⁵]_{NP} kap⁵³ kuan³³ i⁵⁵ a⁵¹
 I want high table and high chair
 ‘I want high tables and high chairs.’
- 2a i³³ ai⁵¹ ts^ham³³ ka³³ [kun³³-tui³³]_{NP} e³³ ua²¹ taŋ³³
 he like participate military team MOD activity
 ‘He likes to participate in military activities.’
- 2b i³³ ai⁵¹ ts^ham³³ ka³³ [kun³³-tui³³]_{NP} e³³ ua²¹ taŋ³³
 he like participate group team MOD activity
 ‘He likes to participate in group activities.’
- 3a tsit⁵³ e⁵⁵ si²¹ pɔ⁵¹ te²¹ hi²¹ e³³ [tsin³³ ke³³ sian⁵⁵]_{NP}
 this COP puppet show MOD (name)
 ‘This is Tsinkesian in the puppet show.’
- 3b tsit⁵³ e⁵⁵ si²¹ pɔ⁵¹ te²¹ hi²¹ e³³ [tsin³³ ke³³ sian⁵⁵]_{NP}
 this COP puppet show MOD (name)
 ‘This is Tsinkesian in the puppet show.’

- 4a i³³ ki³³ sut⁵³ bə³³ hə⁵¹ [t^hau³³ tsit²¹ kai⁵¹]_{VP} tə²¹ hɔ²¹ laŋ³³ lia⁵¹ tiə²¹
 he skill not good steal one time then PASS people catch
 ‘His skills were not good; he was caught when he once stole.’
- 4b i³³ ki³³ sut⁵³ bə³³ hə⁵¹ [t^hau³³ tsit²¹ kai⁵¹]_{VP} tə²¹ hɔ²¹ laŋ³³ lia⁵¹ tiə²¹
 he skill not good first one time then PASS people catch
 ‘His skills were not good; he was caught the first time he did (something).’

The materials improved on those used in the second experiment of Tsay & Myers (2001) by including only one type of surface [33] tone in the within-position comparison, namely *yinping*. This simplified the interpretation in case we found no difference between f_0 contours for juncture [33] and context [33]. Incomplete neutralization should make context *yinping* (whose juncture form is a high [55]) higher than juncture [33] in an additional effect on top of phrase-final lowering (found in Taiwan Southern Min by Peng (1997), Tsay, Charles-Luce & Guo (1999), Tsay & Myers (2001), Hsieh (2004); cf. H.-B. Lin (1988) who found sentence-final lowering). Thus if we found no difference in f_0 this result could not be dismissed as a mere canceling out of these two effects.

It was quite difficult to create sentences that matched perfectly in surface phonetic forms (aside from prosody in the across-position pairs and possibly incomplete neutralization) without also affecting lexical, semantic, or pragmatic naturalness. The native-speaker judgments of the second author, as well as those of our assistants and a subset of the participants in this experiment (asked a few weeks afterwards) were quite consistent about which sentence in each pair seemed more natural. These biases were distributed evenly across the pairs of both types, and one within-position pair (with [tsin³³]; sentences 3a-b in Table 3) showed no bias at all.

2.1.2 Participants

Twenty-four male staff at the National Chung Cheng University (Chiayi, Taiwan) were paid to participate in the experiment (data from six additional paid participants were dropped before analysis, for reasons explained in the Procedure section). Only male speakers were used in an attempt to minimize cross-speaker variation in f_0 . All were native speakers of Taiwan Southern Min who regularly used Taiwan Southern Min in daily life, although they also spoke Mandarin when necessary. They reported no history of speech or hearing problems. The mean age was 38 years (sd 7 years).

2.1.3 Procedure

All speakers participated in two experimental conditions. In the listener-absent

condition, conducted first, speakers made recordings alone in a room. The fourteen experimental sentences were mixed among eighteen additional filler sentences (like the experimental items, designed as minimal pairs differing in only one character/morpheme, but most also differed in segmental content or surface phonetic tone). Participants were presented with all items in one of two arbitrarily assigned orders, one random and the second the reverse of the first. In the listener-present condition, conducted after the listener-absent condition for all participants, a Taiwan Southern Min-speaking assistant was present in the room with the speaker, attentively listening and pretending to try to determine which of two possible sentences correctly matched the speaker's utterance. The sixteen experimental sentences were presented in minimal pairs, without any filler items. To avoid overpracticing, there were no repetitions in either condition.

Both conditions were carried out in the same sound-attenuated room. Speakers were shown one sentence at a time, displayed in Chinese characters on a computer screen. They were asked to read aloud each written sentence into a microphone connected to a Sharp MiniDisc digital recorder. However, due to experimenter error two participants were not recorded in the listener-present condition and so were not included in the analysis.

As an additional check of speaker fluency, the naturalness of all recorded sentence tokens (both targets and fillers, for both listener-absent and listener-present conditions) was judged by a separate group of six naive native speakers, all female staff at National Chung Cheng University (mean age 35, sd 5 years). Only female listeners were used due to the limited size of the participant pool (all of the males participated as speakers). Tokens were played on a Sharp MiniDisc player in the same order in which they were recorded. Listeners judged them on a 5-point scale, where 1 = "very unnatural" (written in Chinese characters on the answer sheet) and 5 = "very natural". A score of 3.5 was chosen as our threshold. One target sentence (with [kuan³³]; sentence 1a in Table 3) received a mean score of 3.4 and so it and its match (sentence 1b) were not analyzed. Three speakers were also removed before analysis for having mean fluency scores below 3.5. In addition to acting as a filter, the fluency scores also allowed us to perform additional analyses. In particular, reading fluency turned out to be negatively rather than positively correlated with age ($r(28)=-0.32$, $p=0.09$), perhaps because older speakers had slower cognitive processing overall or because they were less accustomed to reading Southern Min aloud.

The recordings were entered into the digital waveform analyzer of Computerized Speech Lab 4300B (Kay Elemetrics) at a sampling rate of 10 kHz. Syllable duration was measured from the beginning to the end of periodicity of the vowel portion of the target syllables. Note that all target syllables began with a voiceless onset and were followed by a syllable with an obstruent onset, making the location of the beginning and

end of the target syllables quite straightforward. Fundamental frequency was calculated for 20 ms frames using the CSL pitch-tracking algorithm. If a measured value differed from the preceding value by more than 30 Hz, it was assumed to be an artifact of the pitch-tracker and was replaced by the mean of the two surrounding values. After adjusting values in this way, the respective means of the first two, middle two, and last two f_0 values were computed to serve as measures of the f_0 contour.

Analyses were only carried out on a given sentence pair for a given speaker if the target syllables in all four of the items necessary for the analysis (2 conditions \times 2 sentences in each pair) were correctly pronounced and pitch-tracked. This criterion meant that one further speaker had to be removed from analysis, due to the high rate of pitch-tracking errors for his recordings in the listener-present condition.

Because the logic of this study involved searching for differences that we predict not to exist, it was important to use the most powerful and reliable statistical method currently available, so that if no significant differences were found, we could interpret this as suggesting that, perhaps, such differences truly do not exist after all. This logic precluded the use of the most familiar statistical method, namely analysis of variance (ANOVA), which despite its popularity has a number of serious limitations. First, the fact that some tokens had to be dropped from the analysis meant that running ANOVA would require us to average across items (ANOVA cannot handle “unbalanced” data sets, e.g. a 2-by-3 design involving only five cells rather than the expected six). Averaging would result in the loss of possibly crucial information. Second, we needed to test, across speakers, f_0 values measured at three points, but this situation risks violating a crucial assumption of repeated-measures ANOVA called sphericity, which stipulates that the variance of the differences between any two levels in a multi-level factor (as we have here with the three measurement points) must be the same. If sphericity violations are not taken into account, the resulting p values may be artificially low (Max & Onghena 1999). Yet if they are taken into account with the standard post-hoc corrections, the p values may be artificially high (Huynh & Feldt 1976). Finally, treating f_0 measurement point as a simple multi-level factor misses the fact that the three points are actually ordered, a fact that cannot be captured in ANOVA.

Repeated-measures ANOVA is no longer the gold standard in the analysis of cross-speaker data (Baayen 2004) since the development of a much more powerful and appropriate method in the past few decades. This method is called linear mixed-effects modeling (LME, Pinheiro & Bates 2000, Baayen, forthcoming). Like repeated-measures regression (Lorch & Myers 1990), LME treats participants as random, but it encodes participants as a single random variable rather than as a set of dummy variables, thereby improving sensitivity and avoiding the need to conform to sphericity assumptions. As a type of regression, LME includes all data points in the analysis rather than averages,

and has no problem with “unbalanced” data sets. Finally, the independent variables can include both nominal independent variables as in ANOVA (juncture = 1, context = -1; *yinping* = 1, *yangping* = -1; listener present = 1, listener absent = -1) but also ordinal ones (beginning f_0 = 1, middle = 2, end = 3). LME is quite simple to run using the free statistical software R (R Development Core Team, 2007); see Baayen ([forthcoming](#)) for tutorials.

LME analyses find regression coefficients, conventionally symbolized B ; sign indicates the direction of effect (e.g. a negative B for measurement point indicates a falling contour). Significance (i.e. that B is not zero) is tested using t values. There continues to be some disagreement over how to compute the degrees of freedom (df), which depends on sample size; higher df values mean potentially lower p values (greater significance). Since we predict that there will be no significant effects of tone category on f_0 , the safest test would be to assume higher df values; we expect that we will still fail to find significant results even under these ideal conditions. The easiest way to do this is to pretend that our sample size is infinite, which we can do by treating the t values as if they were z values for the standard normal distribution. Hence in the statistical results below, we report three values for each result: B , z , and p .

2.2 Results

By-speaker analyses were run on all available syllable durations and f_0 values for each tone category (juncture *yangqu* vs. context *yinping*, context *yinping* vs. context *yangping*) in each discourse condition (listener absent vs. listener present). There were 696 data points in the juncture vs. context f_0 analyses, and 744 data points in the *yinping* vs. *yangping* f_0 analyses. Both f_0 data sets showed significant violations of sphericity by Mauchly’s test (for the measurement point factor and its interactions, $W < 0.76$, $p < .05$), confirming that ANOVA was not the appropriate analysis.

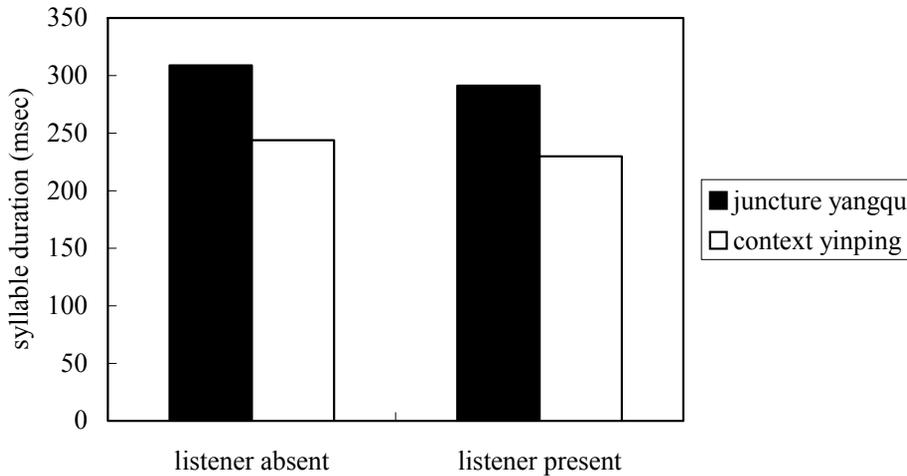


Figure 2: Mean syllable durations for juncture *yangqu* vs. context *yinping* in listener absent vs. present conditions.

Figure 2 shows the mean syllable durations for the across-position comparison (juncture *yangqu* vs. context *yinping*) with listener absent and listener present. A listener \times tone LME found that juncture tones were significantly longer (by an average of 61 ms) than context tones ($B=30.43$, $z=11.39$, $p<.0001$). There was also a significant effect of listener ($B=-7.73$, $z=-2.89$, $p<.01$), with speakers apparently talking a bit faster with a listener present (mean duration was 15 ms shorter), but there was no interaction between the two factors ($B=-1.55$, $z=-0.58$, $p=.56$), revealing no effect of discourse context on duration. Error bars are not shown since the statistics of LME does not depend on them.

The f_0 values at three measurement points for across-position comparison (juncture *yangqu* vs. context *yinping*) are shown in Figure 3. A tone category \times listener [absence/presence] \times measurement point [1/2/3] LME uncovered only one pattern that reached significance at the 0.05 level, namely a main effect of measurement point ($B=-3.26$, $z=-9.35$, $p<.0001$), which merely indicated that the contours, though transcribed as level [33], were actually acoustically falling, not flat. The effect of listener was not significant ($B=-0.87$, $z=-1.16$, $p=.25$), though on average, speakers spoke with a slightly lower overall f_0 (by 2.1 Hz) when a listener was present. Crucially, however, there was not even a hint of any effect of tone category ($B=0.02$, $z=0.02$, $p=.98$), with the overall f_0 difference being a mere 0.2 Hz. In fact, the relative mean f_0 of juncture and context were reversed across the two listener conditions, as is clear from Figure 3, though the listener \times tone category interaction was not at all significant either ($B=-0.09$, $z=-0.27$, $p=.79$). There

was also no interaction between tone category and measurement point ($B=0.03$, $z=0.10$, $p=.92$), showing that there was no reason to reject the null hypothesis that the slopes of the two tones were identical.

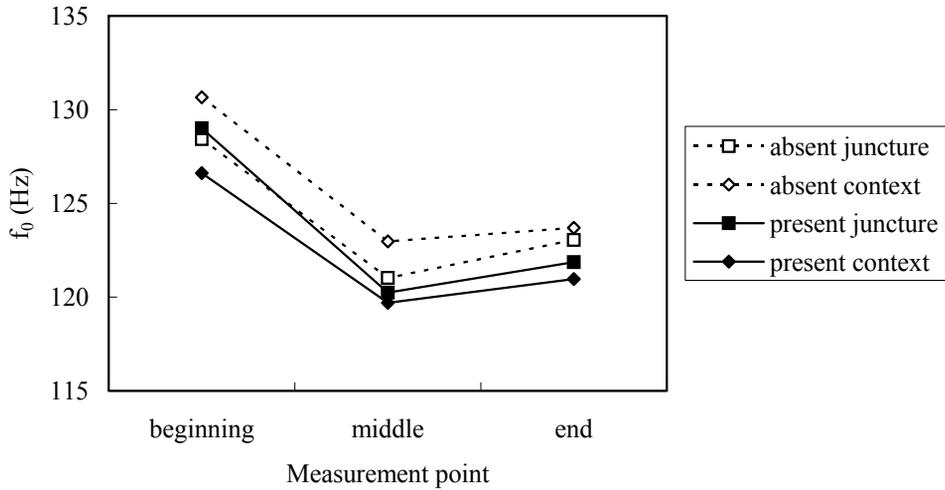


Figure 3: Mean f_0 values at three measurement points for juncture *yangqu* vs. context *yinping* in listener absent vs. present conditions.

For the mean syllable durations in the within-position comparison (context *yinping* vs. context *yangping*), shown in Figure 4, a listener \times tone LME found a significant effect of tone category, with *yangping* syllables 8 ms an average of longer than *yinping* syllables ($B=-3.96$, $z=-3.06$, $p<.01$). There was no main effect of listener ($B=-1.58$, $z=-1.22$, $p=.22$), though on average syllables were 3 ms longer when the listener was absent. There was, however, a significant interaction with the listener condition ($B=-2.71$, $z=-2.10$, $p<.05$), with speakers producing a greater duration difference when a listener was present (13 ms) than when the listener was absent (3 ms).

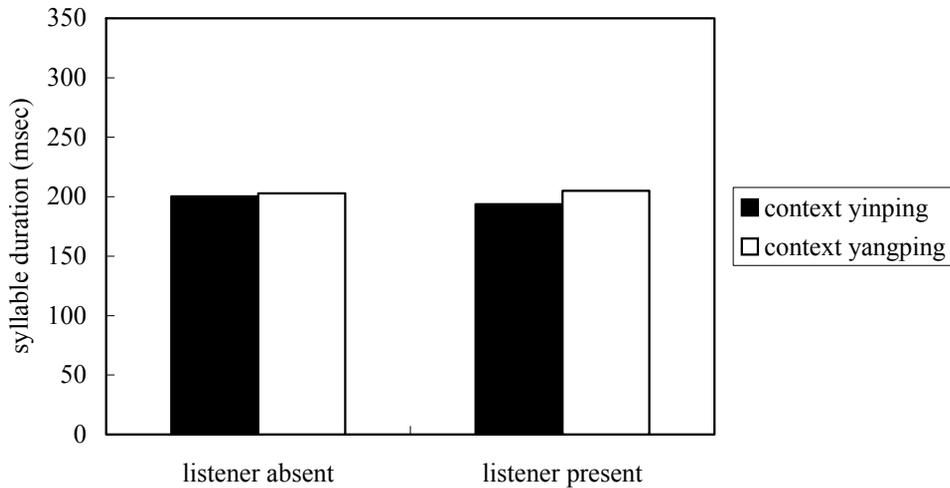


Figure 4: Mean syllable durations for context *yinping* vs. context *yangping*, in listener absent vs. present conditions.

Figure 5 shows the f_0 values for three measurement points in the within-position comparison (context *yinping* vs. context *yangping*). As with the across-position analysis, a tone category \times listener [absence/presence] \times measurement point [1/2/3] LME found a highly significant effect of measurement point ($B=-2.63$, $z=-6.46$, $p<.0001$) indicating that the slopes were falling. Otherwise, as before, there were no significant effects, with no overall effect of listener ($B=0.19$, $z=0.22$, $p=.83$) or tone category ($B=0.90$, $z=1.03$, $p=.31$), with an overall cross-category f_0 difference of a mere 2.3 Hz. There was also no interaction between these two factors ($B=-0.73$, $z=-0.83$, $p=.41$), revealing no discourse effect on contrasting tone by overall pitch, and no interaction between tone category and measurement point ($B=0.11$, $z=0.28$, $p=.78$), showing that slopes need not be assumed to be anything other than identical. Although Figure 5 may seem to imply that with the listener present, the two tone categories differ more greatly in slope than with no listener, the three-way interaction of tone with listener and measurement point failed to reach significance ($B=0.32$, $z=0.80$, $p=.43$).

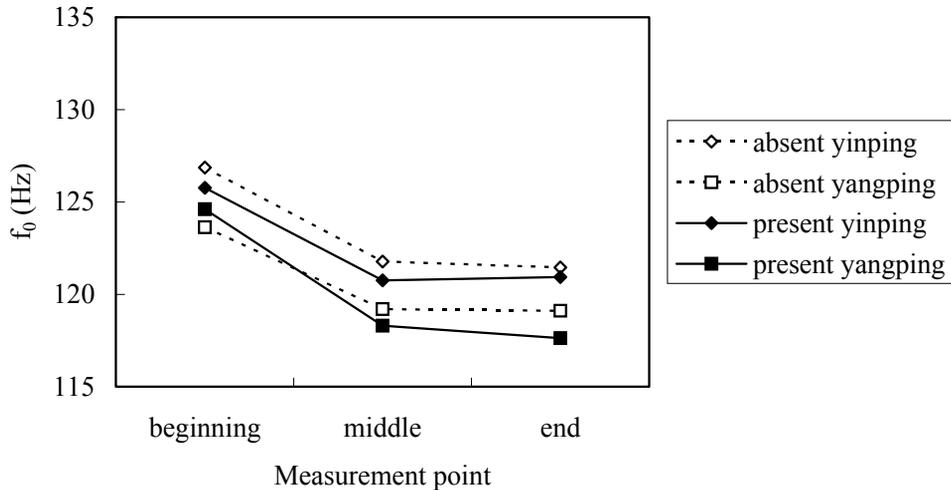


Figure 5: Mean f_0 values at three measurement points for context *yinping* vs. context *yangping* in listener absent vs. present conditions.

3. Discussion

The categoricity hypothesis predicts that f_0 should be fully neutralized in Taiwan Southern Min tone sandhi, and that the alternations should not be affected by pragmatic conditions. Thus we expected that (1) f_0 would be neutralized in across-position comparisons (juncture position vs. context position) and (2) f_0 would also be neutralized in within-position comparisons (context *yinping* vs. context *yangping*). Based on previous research we further expected that (3) syllable duration would be clearly distinct across position but neutralized within position. Finally, we expected that (4) the absence or presence of a listener would have no effect on the degree of neutralization. The first two predictions were confirmed fully, while the second two were apparently only partially confirmed. Nevertheless, as we will explain, we have reasons for thinking that these findings are best understood within the context of the categoricity hypothesis.

3.1 Neutralization of f_0 across prosodic positions

The results for the across-position comparison strongly suggest that the speakers did neutralize juncture *yangqu* and context *yinping*, in accordance with expectation (1). Our use of context *yinping* (rather than *yangping*) to make the comparison with the juncture tone means that the lack of a difference in f_0 cannot be dismissed as a canceling

out of the effects of incomplete neutralization by phrase-final lowering, since with this tone category the effects should be additive as explained above in the Materials section. In fact we found no evidence for either incomplete neutralization or phrase-final lowering. The lack of a phrase-final effect on f_0 is not unprecedented; Peng (1997) reported less lowering in phrase-final position than in utterance-final position.

3.2 Neutralization of f_0 within prosodic position

Consistent with prediction (2), there was also no significant difference in overall f_0 or f_0 contours in context *yinping* vs. context *yangping* tones. True, if the two context tones preserved some aspects of their putative origins as juncture tones, the f_0 for context *yinping* (with juncture form [55]) should be overall somewhat higher than that for context *yangping* (with juncture form [24]), and this is consistent with what we found, with *yinping* f_0 an average of 2.3 Hz higher than *yangping* f_0 . Nevertheless, there are several other results that are inconsistent with the hypothesis that these context forms are truly derived from juncture forms during speech production.

First, if context *yinping* [33] is derived from [55] and context *yangping* [33] from [24], preservation of this contrast should not only produce overall f_0 differences, but also slope differences. In particular, there should be a greater difference in f_0 at the beginning than at the end (since [55] differs more from [24] at the beginning than at the end). Yet we found no significant difference in f_0 slope between these two tones.

Second, variation across speakers seems to be higher for the context *yinping* vs. context *yangping* contrast than for more solid cases of incomplete neutralization described in the literature. A large proportion of our speakers (10 of 24 speakers, or 42%) showed the opposite pattern from the average in at least one of the two pragmatic conditions (i.e. they produced overall higher f_0 for context *yangping* than for context *yinping*). By contrast, in the study of vowel duration effects of coda /t/ and /d/ in Dutch, which is also a relatively small effect, Warner et al. (2004) found that only five of their 15 speakers (33%) failed to show the difference reflected in the average, with only three of these (20%) showing a pattern opposite to the average. We found a large amount of within-speaker variation as well. Even among the speakers with the top three highest fluency scores there was a reversed pattern (i.e. higher f_0 for context *yangping*) in 48% of the comparisons (13 out of 27 = 3 speakers \times 3 sentence pairs \times 3 measurement points).

Third, there is a real concern that our results may have reflected reading processes. The influence of orthography in incomplete neutralization studies is well-known; in particular, Warner et al. (2004, 2006) showed that sub-phonemic duration differences could be induced in speech production by purely orthographic differences in the absence of any underlying phonological distinction. Chinese orthography does not directly

represent tone, but it does use distinct symbols (characters) for distinct morphemes. When naming these characters in isolation, one naturally pronounces the juncture form (i.e. the citation form). This tendency might be even greater for a reader of Southern Min than of Mandarin, since as noted earlier Southern Min does not have a long written tradition; reading Southern Min aloud probably involves less automatized lexical access and speech preparation processes than is the case for reading Mandarin or Dutch aloud. Hence if a less fluent reader of Taiwan Southern Min has a tendency to read character by character, this may result in a tendency to pronounce characters aloud in a form closer to juncture form. If such disfluency was at least partly responsible for the incompleteness of f_0 neutralization that we found, we should find a negative correlation between speakers' fluency scores and their mean differences between context *yinping* and context *yangping* f_0 values. This correlation was indeed negative, although it was small and nonsignificant ($r(24)=-0.26, p=0.22$).

Fourth, the difference in overall f_0 between context *yinping* and context *yangping* was not only nonsignificant, but also extremely small, a mere 2.3 Hz. This quite a bit smaller than the 17.5 Hz difference found by Zee (1980) between lexical tone 2 and sandhi tone 3 in Beijing Mandarin, or even than the 4.8 Hz difference Yu (2007) reports for near merger in Cantonese tones. Instead, it is identical to the 2.3 Hz difference found by Peng (2000) between lexical tone 2 and sandhi tone 3 in Taiwan Mandarin. Moreover, it is not clear if such a small difference (especially embedded in the variability noted above) can be perceivable or useful to listeners; Peng's Taiwan Mandarin listeners could not detect their 2.3 Hz difference. Note that this difference is far smaller than the effect of anticipatory coarticulation found by Peng (1997) on the offset of context *yinping* (surface [33]), which could be up to 19 Hz (see her Table VII, p.386). There are ways that a near merger can be maintained in a speech community even if the phonetic difference isn't easily perceived (Labov 1994), but none seem relevant to this case. Taiwan Southern Min speakers do very little reading and writing in this language, so the putative contrast is unlikely to be maintained by orthography (outside of artificial situations like our experiment), and there are no varieties of Taiwan Southern Min where tone sandhi is absent, which would allow juncture forms to be clearly heard in context position. Thus if the difference in f_0 in the within-position comparison is not perceptible, it cannot be learned and hence cannot be part of the production phonology of Taiwan Southern Min speakers. If so, this difference must be an experimental artifact, probably due to reading pronunciations. In any case, it is not statistically significant.

Finally, the speakers did not demonstrate the ability to adjust the distinctiveness of these two tone categories in response to the difference in pragmatic condition (listener absent vs. present). Note that the slight difference in slope implied in Figure 5 goes precisely the wrong way if speakers were preserving the juncture forms: speakers

dropped *yangping* towards the end relative to *yinping* when a listener was present, rather than making the two contours begin distinctly and having them come together at the end.

3.3 Duration effects

Consistent with expectation (3), juncture *yangqu* and context *yinping* were clearly distinguished by syllable duration, with a strong effect of phrase-final lengthening; juncture [33] was an average of 61 ms longer than context [33]. Yet surprisingly, there was also a significant difference in duration within context position, with context *yangping* 8 ms longer, on average, than context *yinping*. If context *yinping* [33] is derived from [55] and context *yangping* [33] from [24], it is conceivable that some remnant of the base tone's inherent duration could be preserved. In juncture position, level *yinping* [55] is inherently shorter than rising *yangping* [24], with H.-B. Lin (1988) reporting a duration difference around 2% of the mean duration of the two tones and Tsay (2001) reporting a duration difference of around 6%. The observed duration difference here falls right in the middle of these two estimates, namely around 4% of the mean durations (8 ms divided by 201 ms). Thus the observed difference in duration is consistent with gradient incomplete neutralization of the two tone categories in tone sandhi, contrary to the categoricity hypothesis.

Yet we should remember that we found no overall difference in slope between these two tones in context position. It has been argued that the processes giving rise to incomplete neutralization involve the temporal adjustment of gestures and their relations (Zsiga 1993, Port 1996, Port & Lear 2005). If so, one would expect that with respect to incomplete neutralization in tone, such processes would be incapable of raising or lowering overall f_0 without affecting slope in any way, since the production of f_0 across a syllable is by nature a dynamic process. The incomplete neutralization of sandhi tone 3 and lexical tone 2 in Beijing Mandarin does indeed involve slope differences (Zee 1980), and slope differences are also found with the gestural timing adjustments of tonal coarticulation (see Shen 1990, and Xu 1993, 1997, for Mandarin; Peng 1997, for Taiwan Southern Min). The lack of slope differences in our Taiwan Southern Min results is thus incompatible with the proposed mechanisms of incomplete neutralization.

Moreover, duration differences in tone are not unpredictable aspects of their lexical specification, but instead arise as a side effect of tonal articulation (Ohala 1978). If this is truly a case of incomplete neutralization, why then is it only the accidental duration effects that are “preserved”, not the tonal contours that give rise to them? One way to resolve this puzzle would be to adopt a version of the analysis developed by Tsay (2001) in a study of tone productions in children acquiring Taiwan Southern Min. She noted

that children often fail to produce the segmental hallmark of the “short” (rusheng) tones, namely the final voiceless coda (/p, t, k/ or glottal stop). Nevertheless, like adults, the durations of rusheng syllables in child speech were significantly shorter than “long” syllables, whether or not an audible coda was present. This pattern implies that children are capable of picking up on what, in the adult phonology, may be an accidental property of rusheng syllables, namely their short duration, despite lacking control over the voiceless codas that give rise to it. It seems plausible to suppose that adult speakers maintain this kind of durational knowledge about tone categories. Since listeners can use syllable duration as a cue to tone identity in Mandarin (e.g., Blicher, Diehl & Cohen 1990, Tseng, Huang & Jeng 1996), why couldn’t speakers employ such durational cues in production as well, alongside the categorical tonal information?

In other words, speakers could simply be imposing the property of duration, accidentally associated with the tone categories, on top of tonal contours that are lexically specified as identical. This is consistent with the reading-pronunciation scenario sketched earlier, whereby speakers highlight contrasts between otherwise identical sentences by pronouncing written characters closer to their isolation form, that is, their juncture form, which in this case are somewhat longer in duration. Like the duration variation in juncture vs. context tones, duration variation within context position would be applied at a late stage of production, where lexically specified representations are implemented in fluent speech. After all, something like this must already be assumed in order to explain how speakers can freely lengthen or shorten words that they wish to emphasize or deemphasize.

Turning now to expectation (4), discourse context had no apparent effects on the contrast between the two tone categories in f_0 ; unlike what has been found with incompletely neutralizing postlexical phonology (e.g. Charles-Luce 1997), speakers did not show a lesser degree of neutralization when a listener was present. However, given the above discussion of late-stage control over duration, it should not be surprising to find that discourse context did have an effect on duration in the within-position analysis, with the duration difference between context *yinping* and context *yangping* increasing from 3 ms without a listener to 13 ms with a listener. This is just what would be expected if speakers have control over the implementation of the prosodic frames in which lexically specified tonal contours are realized. It is important to recognize that this is possible, but this interesting finding doesn’t undermine the categoricity hypothesis concerning tone sandhi itself.

However, we need not even concede this much, since the results involving within-position duration effects are the only ones in this experiment that even hint at the preservation of an “underlying” contrast under pragmatic control. This makes it hard to be sure that these particular patterns, though significant across speakers, do not result

from some accidental property of the particular items used in the task, perhaps something about the orthographic stimuli (e.g. character frequency, familiarity, neighborhood size, or semantics). Further work will be needed to determine whether the longer duration of context *yangping* relative to context *yinping* is actually caused in some way by the relative durations of their juncture forms, rather than being a coincidence.

4. Conclusions

As noted in the introduction, there are several independent arguments supporting the claim that Taiwan Southern Min tone sandhi is lexicalized, including its phonetic arbitrariness, its only partial productivity, and the fact that lexical items in the same tone category may undergo different sandhi alternations. If we are right in interpreting our phonetic results as consistent with this general claim, Taiwan Southern Min tone sandhi (and lexicalized phrasal phonology in general) may represent a phenomenon with an important role to play in the debate over the nature of phonological representations and processes. Lexicalized phrasal phonology does indeed seem to require the online manipulation of categorical phonological units. We don't believe that this manipulation is done by general rules of the sort that have been rejected on phonetic grounds for patterns like German final devoicing, and in fact Taiwan Southern Min tone sandhi behaves more like phonologically conditioned allomorphy than "pure phonology" (although in many cases the distinction isn't clear, e.g. voicing assimilation in English inflection). Nevertheless, Taiwan Southern Min speakers know how to apply tone sandhi to novel sentences and word combinations (at least semi-productively), and this knowledge does not seem to involve gradient temporal and gestural adjustments.

A growing literature views all phonological representations and processes as phonetically detailed (e.g. Bybee 1994, 2000, Port 1996, Boersma 1998, Steriade 2000, Pierrehumbert 2001, Ernestus & Baayen 2006, Yu 2007). We agree that this research line is quite promising and important. Nevertheless, even advocates of this approach (Boersma 1998, Beckman & Pierrehumbert 2003) have acknowledged that categorical phonological representations do exist, and their origin and processing remain important challenges to phonological theory.

References

- Baayen, R. Harald. 2004. Statistics in psycholinguistics: a critique of some current gold standards. *Mental Lexicon Working Papers* 1:1-45.
- Baayen, R. Harald. 2008. *Analyzing Linguistic Data: A Practical Introduction to Statistics Using R*. Cambridge & New York: Cambridge University Press.
- Beckman, Mary. E., and Janet Pierrehumbert. 2003. Interpreting “phonetic interpretation” over the lexicon. *Phonetic Interpretation*, ed. by John Local, Richard Ogden & Rosalind Temple, 13-28. *Papers in Laboratory Phonology* 6. Cambridge & New York: Cambridge University Press.
- Blicher, Deborah L., Randy L. Diehl, and Leslie B. Cohen. 1990. Effects of syllable duration on the perception of the Mandarin Tone 2/Tone 3 distinction: evidence of auditory enhancement. *Journal of Phonetics* 18.1:37-49.
- Bodman, Nicholas C. 1955[1987]. *Spoken Amoy Hokkien*. Ithaca: Spoken Language Services.
- Boersma, Paul. 1998. *Functional Phonology*. The Hague: Holland Academic Graphics.
- Borowsky, Toni. 1993. On the word level. *Studies in Lexical Phonology*, ed. by Sharon Hargus & Ellen M. Kaisse, 199-234. *Phonetics and Phonology* 4. San Diego: Academic Press.
- Browman, Catherine P., and Louis M. Goldstein. 1990. Tiers in articulatory phonology, with some implications for casual speech. *Between the Grammar and Physics of Speech*, ed. by John Kingston & Mary E. Beckman, 341-376. *Papers in Laboratory Phonology* 1. Cambridge & New York: Cambridge University Press.
- Bybee, Joan L. 1994. A view of phonology from a cognitive and functional perspective. *Cognitive Linguistics* 5.4:285-305.
- Bybee, Joan. L. 2000. The phonology of the lexicon: evidence from lexical diffusion. *Usage-based Models of Language*, ed. by Michael Barlow & Suzanne Kemmer, 65-85. Stanford: CSLI.
- Chang, Yueh-Chin. 1988. Sandhi tonal des syntagmes dissyllabiques du Min-nan parlé a Taiwan [Tone sandhi in disyllabic compounds in Minnan spoken in Taiwan]. *Cahiers de Linguistique Asie Orientale* 17:193-234.
- Chang, Yueh-Chin, and Yi-Ching Su. 1994. La modification tonale du 3ème ton du mandarin parlé à Taiwan [Tone modification of the third tone in Mandarin spoken in Taiwan]. *Cahiers de Linguistique Asie Orientale* 23:39-59.
- Chao, Yuen Ren. 1930. A system of tone letters. *Le Maître Phonétique* 45:24-27.
- Chao, Yuen Ren. 1948. *Mandarin Primer*. Cambridge: Harvard University Press.
- Charles-Luce, Jan. 1985. Word-final devoicing in German: effects of phonetic and sentential contexts. *Journal of Phonetics* 13.3:309-324.

- Charles-Luce, Jan. 1993. The effects of semantic context on voicing neutralization. *Phonetica* 50.1:28-43.
- Charles-Luce, Jan. 1997. Cognitive factors involved in preserving a phonemic contrast. *Language and Speech* 40.3:229-248.
- Charles-Luce, Jan, and Kelly M. Dressler. 1999. The effects of semantic predictability in non-pathological older adults' production of a phonemic contrast. *Clinical Linguistics and Phonetics* 13.3:199-217.
- Charles-Luce, Jan, Kelly M. Dressler, and Elivra Ragonese. 1999. Effects of semantic predictability on children's preservation of a phonemic voice contrast. *Journal of Child Language* 26.3:505-530.
- Chen, Matthew Y. 1987. The syntax of Xiamen tone sandhi. *Phonology Yearbook* 4: 109-149.
- Chen, Matthew Y. 1996. Tonal geometry: a Chinese perspective. *New Horizons in Chinese Linguistics*, ed. by C.-T. James Huang & Y.-H. Audrey Li, 21-48. Dordrecht & Boston: Kluwer.
- Chen, Matthew Y. 2000. *Tone Sandhi: Patterns across Chinese Dialects*. Cambridge & New York: Cambridge University Press.
- Cheng, Robert L. 1968. Tone sandhi in Taiwanese. *Linguistics* 41:19-42.
- Di Paolo, Marianna, and Alice Faber. 1990. Phonation differences and the phonetic content of the tense-lax contrast in Utah English. *Language Variation and Change* 2:155-204.
- Du, Tsai-Chwun. 1988. *Tone and Stress in Taiwanese*. Urbana: University of Illinois at Urbana-Champaign dissertation.
- Ernestus, Mirjam, and R. Harald Baayen. 2006. The functionality of incomplete neutralization in Dutch: the case of past-tense formation. *Laboratory Phonology* 8, ed. by Louis M. Goldstein, D. H. Whalen & Catherine T. Best, 27-49. Berlin: Mouton de Gruyter.
- Faber, Alice, and Marianna Di Paolo. 1995. The discriminability of nearly merged sounds. *Language Variation and Change* 7:35-78.
- Fon, Janice. 1997. *What are Tones Really Like? An Acoustic-based Study of Taiwan Mandarin Tones*. Taipei: National Taiwan University MA thesis.
- Fourakis, Marios M., and Gregory K. Iverson. 1984. On the 'incomplete neutralization' of German final obstruents. *Phonetica* 41:140-149.
- Guy, Gregory R. 1980. Variation in the group and the individual: the case of final stop deletion. *Locating Language in Time and Space*, ed. by William Labov, 1-54. New York: Academic Press.
- Guy, Gregory R. 1991. Explanation in variable phonology: an exponential model of morphological constraints. *Language Variation and Change* 3:1-22.

- Hayes, Burce. 1990. Precompiled phrasal phonology. *The Phonology-syntax Connection*, ed. by Sharon Inkelas & Draga Zec, 85-108. Chicago: University of Chicago Press.
- Hockett, Charles F. 1947. Peiping phonology. *Journal of the American Oriental Society* 67.4:253-267. Reprinted in *Readings in Linguistics I* (4th edition), ed. by Martin Joos, 217-228, 1964. Chicago: University of Chicago Press.
- Hsiao, Yuchau E. 1991. *Syntax, Rhythm and Tone: A Triangular Relationship*. San Diego: University of California dissertation.
- Hsieh, Hsin-I. 1970. The psychological reality of tone sandhi rules in Taiwanese. *CLS* 6:489-503. Chicago: Chicago Linguistics Society.
- Hsieh, Hsin-I. 1975. How generative is phonology? *The Transformational-generative Paradigm and Modern Linguistic Theory*, ed. by E. F. K. Koerner, 109-144. Amsterdam: John Benjamins.
- Hsieh, Hsin-I. 1976. On the unreality of some phonological rules. *Lingua* 38.1:1-19.
- Hsieh, Pei-Yu. 2004. *Variation in Taiwanese Tones*. Chiayi: National Chung Cheng University MA thesis.
- Huang, Shuanfan. 1993. *Yuyan, Shehui Yu Zuqun Yishi [Language, Society and Ethnic Identity]*. Taipei: Crane.
- Huynh, Huynh, and Leonard S. Feldt. 1976. Estimation of the Box correction for degrees of freedom from sample data in randomized block and split-plot designs. *Journal of Educational Statistics* 1.1:69-82.
- Kaisse, Ellen M. 1985. *Connected Speech: The Interaction of Syntax and Phonology*. Orlando: Academic Press.
- Kaisse, Ellen M., and Sharon Hargus. 1993. Introduction. *Studies in Lexical Phonology*, ed. by Sharon Hargus & Ellen M. Kaisse, 1-19. Phonetics and Phonology 4. San Diego: Academic Press.
- Kiparsky, Paul. 1982. Lexical phonology and morphology. *Linguistics in the Morning Calm*, ed. by In-Seok Yang, 135-160. Seoul: Hanshin.
- Kiparsky, Paul. 1985. Some consequences of Lexical Phonology. *Phonology Yearbook* 2:85-138.
- Kratochvil, Paul. 1986. The case of the third tone. *Wang Li Memorial Volumes (English Volume)*, ed. by Chinese Language Society of Hong Kong, 253-276. Hong Kong: Joint Publishing Co.
- Labov, William. 1994. *Principles of Linguistic Change: Internal Factors*. Oxford & Cambridge: Blackwell.
- Labov, William, Mark Karan, and Corey Miller. 1991. Near-mergers and the suspension of phonemic contrast. *Language Variation and Change* 3:33-74.

- Lin, Hwei-Bing. 1988. *Contextual Stability of Taiwanese Tones*. Storrs: University of Connecticut dissertation.
- Lin, Jo-wang. 1994. Lexical government and tone group formation in Xiamen Chinese. *Phonology* 11:237-275.
- Lorch, Robert F., Jr., and Jerome L. Myers. 1990. Regression analyses of repeated measures data in cognitive research. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 16.1:149-157.
- Manaster Ramer, Alexis. 1996. A letter from an incompletely neutral phonologist. *Journal of Phonetics* 24.4:477-489.
- Martin, Samuel E. 1957. Problems of hierarchy and indeterminacy in Mandarin phonology. *Bulletin of the Institute of History and Philology Academia Sinica* 29: 209-230.
- Max, Ludo, and Patrick Onghena. 1999. Some issues in the statistical analysis of completely randomized and repeated measures designs for speech, language, and hearing research. *Journal of Speech, Language, and Hearing Research* 42.2: 261-270.
- Moreton, Elliott. 1999. Non-computable functions in Optimality Theory. Amherst: University of Massachusetts Manuscript. (Available at roa.rutgers.edu.)
- Myers, James, and Jane Tsay. 2003. Investigating the phonetics of Mandarin tone sandhi. *Taiwan Journal of Linguistics* 1.1:29-68.
- Myerson, Rosemarie Farkas. 1978. Children's knowledge of selected aspects of sound pattern of English. *Recent Advances in the Psychology of Language: Formal and Experimental Approaches*, ed. by Robin N. Campbell & Philip T. Smith, 377-402. New York: Plenum Press.
- Odden, David. 1987. Kimatuumbi phrasal phonology. *Phonology Yearbook* 4:13-26.
- Odden, David. 1990. Syntax, lexical rules and postlexical rules in Kimatuumbi. *The Phonology-syntax Connection*, ed. by Sharon Inkelas & Draga Zec, 259-278. Chicago: University of Chicago Press.
- Ohala, John J. 1974. Experimental historical phonology. *Historical Linguistics*, Vol. 2: *Theory and Description in Phonology*, ed. by John M. Anderson & Charles Jones, 353-387. Amsterdam: North-Holland.
- Ohala, John J. 1978. Production of tone. *Tone: A Linguistic Survey*, ed. by Victoria A. Fromkin, 5-39. New York: Academic Press.
- Peng, Shu-hui. 1997. Production and perception of Taiwanese tones in different tonal and prosodic contexts. *Journal of Phonetics* 25.3:371-400.
- Peng, Shu-hui. 1998. Evidence for a UR-free account of tone sandhi in Taiwanese. Poster presented at the Sixth Conference of Laboratory Phonology. Toronto: University of York.

- Peng, Shu-hui. 2000. Lexical versus 'phonological' representations of Mandarin sandhi tones. *Acquisition and the Lexicon*, ed. by Michael B. Broe & Janet B. Pierrehumbert, 152-167. Papers in Laboratory Phonology 5. Cambridge & New York: Cambridge University Press.
- Pierrehumbert, Janet B. 2001. Exemplar dynamics: word frequency, lenition and contrast. *Frequency and Emergence of Linguistic Structure*, ed. by Joan Bybee & Paul Hopper, 137-157. Amsterdam & Philadelphia: John Benjamins.
- Pinheiro, José C., and Douglas M. Bates. 2000. *Mixed-effects Models in S and S-Plus*. New York: Springer.
- Port, Robert F. 1996. The discreteness of phonetic elements and formal linguistics: response to Alexis Manaster Ramer. *Journal of Phonetics* 24.4:491-511.
- Port, Robert F., and Penny Crawford. 1989. Incomplete neutralization and pragmatics in German. *Journal of Phonetics* 17.4:257-282.
- Port, Robert F., and Adam P. Leary. 2005. Against formal phonology. *Language* 81.4: 927-964.
- Port, Robert F., and Michael L. O'Dell. 1985. Neutralization of syllable-final voicing in German. *Journal of Phonetics* 13.4:455-471.
- R Development Core Team. 2007. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org>.
- Rubach, Jerzy. 1990. Final devoicing and cyclic syllabification in German. *Linguistic Inquiry* 21.1:79-94.
- Shen, Xiaonan Susan. 1990. Tonal coarticulation in Mandarin. *Journal of Phonetics* 18.2:281-295.
- Steinberg, Danny D., and Robert K. Krohn. 1975. The psychological validity of Chomsky and Halle's vowel shift rule. *The Transformational-Generative Paradigm and Modern Linguistic Theory*, ed. by E. F. K. Koerner, 233-259. Amsterdam: John Benjamins.
- Steriade, Donca. 2000. Paradigm uniformity and the phonetics-phonology boundary. *Acquisition and the Lexicon*, ed. by Michael B. Broe & Janet B. Pierrehumbert, 313-334. Papers in Laboratory Phonology 5. Cambridge & New York: Cambridge University Press.
- Ting, Pang-Hsin. 1970. *Taiwan Yuyan Yuanliu [The Origins of the Languages in Taiwan]*. Taichung: Press Secretary's Office, Taiwan Provincial Government.
- Tsay, Jane. 1994. *Phonological Pitch*. Tucson: University of Arizona dissertation.
- Tsay, Jane. 2001. Phonetic parameters of tone acquisition in Taiwanese. *Issues in East Asian Language Acquisition*, ed. by Mineharu Nakayama, 205-225. Tokyo: Kuroshio.

- Tsay, Jane. 2002. Minnanyu liandu biandiao yu cisu bianti xuanze jiashe [Tone sandhi in Minnan and the allomorph selection hypothesis]. *Contemporary Linguistics* 4.3: 176-200.
- Tsay, Jane, Jan Charles-Luce, and Yung-Sung Guo. 1999. The syntax-phonology interface in Taiwanese: acoustic evidence. *Proceedings of the 14th International Congress of Phonetic Sciences (ICPhS'99)*, ed. by John J. Ohala, Yoko Hasegawa, Manjari Ohala, Daniel Granville & Ashlee C. Bailey, 2407-2410. San Francisco, USA.
- Tsay, Jane, and James Myers. 1996. Taiwanese tone sandhi as allomorph selection. *BLS* 22:394-405. Berkeley: Berkeley Linguistic Society.
- Tsay, Jane, and James Myers. 2001. Processes in the production of Taiwanese tone sandhi: an acoustic phonetic study. *Proceeding of the 5th National Conference on Modern Phonetics*, ed. by Lianhong Cai, Tongchun Zhou & Jianhua Tao, 233-237. Beijing: Tsinghua University.
- Tseng, Chin-Chin. 1995. *Taiwanese Prosody: An Integrated Analysis of Acoustic and Perceptual Data*. Honolulu: University of Hawaii dissertation.
- Tseng, Chin-Hsing, Kou-You Huang, and Jing-Yi Jeng. 1996. The role of the syllable in perceiving spoken Chinese. *Proceedings of the National Science Council, Part C*, 6.1:71-86.
- Wang, Samuel H. 1995. *Experimental Studies in Taiwanese Phonology*. Taipei: Crane.
- Wang, William S-Y. 1967. Phonological features of tone. *International Journal of American Linguistics* 33.2:93-105.
- Wang, William S-Y., and Kung-pu Li. 1967. Tone 3 in Pekinese. *Journal of Speech and Hearing Research* 10.3:629-636.
- Warner, Natasha, Allard Jongman, Joan Sereno, and Rachèl Kemps. 2004. Incomplete neutralization and other sub-phonemic durational differences in production and perception: evidence from Dutch. *Journal of Phonetics* 32.2:251-276.
- Warner, Natasha, Erin Good, Allard Jongman, and Joan Sereno. 2006. Orthographic vs. morphological incomplete neutralization effects. *Journal of Phonetics* 34.2:285-293.
- Xu, Yi. 1993. *Contextual Tonal Variation in Mandarin Chinese*. Storrs: University of Connecticut dissertation.
- Xu, Yi. 1997. Contextual tonal variations in Mandarin. *Journal of Phonetics* 25.1:61-83.
- Yip, Moira. 1980. *The Tonal Phonology of Chinese*. Cambridge: MIT dissertation. Reprinted by Garland in 1990 as *The Tonal Phonology in Chinese*.
- Yu, Alan C. L. 2007. Understanding near mergers: the case of morphological tone in Cantonese. *Phonology* 24.1:187-214.
- Yuan, Jia-Hua. (ed.) 1960. *Hanyu Fangyan Gaiyao [Outline of Chinese Dialects]*. Beijing: Wenzhi Gaige Press.

James Myers and Jane Tsay

- Zee, Eric. 1980. A spectrographic investigation of Mandarin tone sandhi. *UCLA Working Papers* 49:98-116.
- Zsiga, Elizabeth. C. 1993. *Features, Gestures, and the Temporal Aspects of Phonological Organization*. New Haven: Yale University dissertation.
- Zsiga, Elizabeth. C. 1995. An acoustic and electropalatographic study of lexical and post-lexical palatalization in American English. *Phonology and Phonetic Evidence*, ed. by Bruce Connell & Amalia Arvaniti, 282-302. Papers in Laboratory Phonology 4. Cambridge & New York: Cambridge University Press.
- Zsiga, Elizabeth. C. 1997. Features, gestures, and Igbo vowels: an approach to the phonology-phonetics interface. *Language* 73.2:227-274.

James Myers
Graduate Institute of Linguistics
National Chung Cheng University
Min-Hsiung, Chiayi 62102, Taiwan
Lngmyers@ccu.edu.tw

Jane Tsay
Graduate Institute of Linguistics
National Chung Cheng University
Min-Hsiung, Chiayi 62102, Taiwan
Lngtsay@ccu.edu.tw