

Neutralization in Taiwanese Tone Sandhi

James Myers and Jane Tsay
Graduate Institute of Linguistics
National Chung Cheng University
Min-Hsiung, Chia-Yi 621
Taiwan

Summer 2002 version

Contact information:

James Myers
Graduate Institute of Linguistics
National Chung Cheng University
Min-Hsiung, Chia-Yi 621
Taiwan
Lngmyers@ccu.edu.tw [please also CC to Lngtsay@ccu.edu.tw]
Phone: 886-5-242-8251
Fax: 886-5-272-1654

Running title: Neutralization in Taiwanese Tone Sandhi

Abstract

Taiwanese tone sandhi, although a phrase-level pattern, has been argued to be essentially lexicalized. We thus expected it to be phonetically categorical and insensitive to pragmatic context. To test these predictions, we had native speakers read aloud pairs of sentences identical except for morphemes expected to be neutralized by tone sandhi, both across and within prosodic positions. Speakers were recorded in two pragmatic conditions: alone and in the presence of an attentive listener. The across-position comparisons showed no difference in f_0 , but the within-position comparisons did find a small but significant difference. However, this latter result is not consistent with true incomplete neutralization, since the slopes did not preserve the putatively underlying tone contours, and there was no effect of pragmatic condition. Moreover, in a second experiment, the same participants were unable to discriminate or identify the tones in the within-position comparisons. Taiwanese tone sandhi thus seems to involve the manipulation of categorical phonological units, though gradient adjustments can appear in reading pronunciations.

1. Introduction

It seems to be generally accepted that, as an empirical phenomenon, speakers will sometimes give instrumentally distinct pronunciations to words otherwise thought to be phonologically identical. It also seems to be established that at least some of these subphonemic distinctions are perceivable by native listeners. 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, in press; Ernestus & Baayen, 2002), flapping in American English (Charles-Luce, 1997; Charles-Luce, Dressler, & Ragonese, 1999; Charles-Luce & Dressler, 1999), and coda aspiration in Eastern Andalusian Spanish (Gerfen & Hall, submitted); more complete lists of studies can be found in these sources and in the reviews by Manaster Ramer (1996) and Port (1996). 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, Karen, & 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.*, in press); 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, 2002).

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). 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 & Goldston, 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 this or that 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 Taiwanese tone sandhi

The example of lexicalized phrasal phonology that we examine in this paper is tone sandhi in Taiwanese (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 I 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 they 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 I AND FIGURE 1 ABOUT HERE]

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 Taiwanese 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). Simplifying somewhat, their general consensus is that Taiwanese tone sandhi is better viewed as a set of essentially arbitrary alternations between stored allomorphs. If so, Taiwanese 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 Taiwanese 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 Taiwanese 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 Taiwanese. Like Taiwanese 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. A thorough examination of phonetic studies of this pattern is beyond the scope of this paper, but it seems fair to say 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 Taiwanese) 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 f_0 of sandhi tone 3 was significantly lower than that of lexical tone 2, but by a much smaller amount than what has been reported for Beijing Mandarin (2.3 Hz, as compared with the 17.5 Hz found by Zee, 1980).

Taiwanese tone sandhi itself has attracted less attention in the phonetic literature. 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 Taiwanese tone sandhi, basing her analysis on the production and perception of nonsense disyllables.

However, she mixed speakers from different dialects (one of her six speakers had a different tone sandhi system from the rest), did not control the segmental environment, and only presents her results graphically (juncture tones for /do/ are given in her 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 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 prediction of Tsay & Myers (1996) that Taiwanese 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 msec 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, Charles-Luce, & Guo (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 ($ps < 0.01$). This experiment, perhaps the best controlled in the literature, seems to imply that at least one aspect of Taiwanese 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 Taiwanese 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), Taiwanese 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 fluency of the college-aged speakers; Taiwanese is less commonly used among this generation than among older speakers (Huang, 1993), and this possibly may have lead to more artificial reading pronunciations. This is especially likely since Taiwanese is not generally considered by its speakers to be a written language; for historical reasons Chinese orthography as used in Taiwan is optimized for the Mandarin lexicon, though it can be used for Taiwanese as well.

To examine the question of neutralization in Taiwanese tone sandhi more carefully, we first carried out a replication of the second experiment of Tsay & Myers (2001) using improved materials and older (and we hoped more fluent) speakers. We then carried out a perceptual tone identification task using the same participants and items from this production experiment.

2. Experiment 1

This experiment had two primary goals. The first was to test the categoricity hypothesis (i.e. that Taiwanese 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 Tables II and III). Three pairs represented the comparison across prosodic position: in one sentence the target morpheme had juncture yangqu tone (1a, 2a, 3a in Table II), while in the other the target morpheme had context yinping tone (1b, 2b, 3b in Table II). Four pairs represented the within-position comparison, contrasting context yinping tone (1a, 2a, 3a, 4a in Table III) and context yangping tone (1b, 2b, 3b, 4b in Table III).

[INSERT TABLE II AND TABLE III ABOUT HERE]

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 (phrase-final lowering was found in Taiwanese by H.-B. Lin, 1988; Peng, 1997; Tsay, Charles-Luce, & Guo, 1999; Tsay & Myers, 2001). 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 two to four weeks after their participation in Experiment 2) 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 III) 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). All were native speakers of Taiwanese who regularly used Taiwanese 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 (s.d. 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 Taiwanese-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, s.d. 5 years). 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 III) 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. For example, we were surprised to find that fluency was negatively rather than positively correlated with age ($r(28) = -0.32, p = 0.09$); hence in further analyses, we relied solely on fluency scores, not ages.

The recordings were entered into the digital waveform analyzer of Computerized Speech Lab 4300B (Kay Elemetrics) at a sampling rate of 10 kHz per second. 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 msec 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.

The syllable durations and f_0 values of each speaker were averaged across items within each tone category (juncture yangqu vs. context yinping, context yinping vs. context yangping) for each condition (listener absent vs. listener present), yielding 32 data points for each speaker (2 conditions \times 2 comparison types \times 2 tone categories per comparison \times 4 measurements: duration, beginning f_0 , middle f_0 , end f_0) and a total of 768 data points (32 \times 24 speakers).

2.2 Results

No significant differences were found between the two presentation orders of the items in the listener-absent condition for any of our measures (one-way factorial ANOVAs, all $ps > 0.05$), so in further analyses we put all speakers into one group. Separate repeated measures ANOVAs (with speakers as the random variable) of the syllable durations and f_0 contours were performed for the across-position comparisons (juncture yangqu vs. context yinping) and the within-position comparisons (context yinping vs. context yangping). Because the categoricity hypothesis predicts nonsignificant differences, in the following we also report nearly significant

effects ($0.05 < p < 0.1$) and F and p values for nonsignificant effects ($p > 0.05$) where relevant.

Figure 2 shows the mean syllable durations for the across-position comparison (juncture yangqu vs. context yinping) with listener absent and listener present. A two-way ANOVA found that juncture tones were longer than context tones ($F(1,23) = 53.39, p < 0.0001$), by an average of 63 msec. There was also a nearly significant effect of listener ($F(1,23) = 3.35, p = 0.08$), with speakers apparently talking a bit faster with a listener present, but there was no interaction between the two factors ($F(1,23) = 0.17, p = 0.69$), indicating the lack of an effect of discourse context on duration.

[INSERT FIGURE 2 ABOUT HERE]

The f_0 values at three measurement points for across-position comparison (juncture yangqu vs. context yinping) are shown in Figure 3. A three-way ANOVA (tone category \times listener [absence/presence] \times measurement point [beginning/middle/end]) uncovered only one pattern that reached significance at the 0.05 level, namely a main effect of measurement point ($F(2,46) = 77.69, p < 0.0001$), which merely indicated that the contours were not flat. The effect of listener was nearly significant ($F(1,23) = 3.85, p = 0.06$); speakers spoke with a somewhat lower overall f_0 when a listener was present. Crucially, however, there was no effect of tone category ($F(1,23) = 0.034, p = 0.86$). In fact, the relative mean f_0 of juncture and context were reversed across the two listener conditions, as is clear from Figure 3; however, the listener \times tone category interaction did not reach significance ($F(1,23) = 2.06, p = 0.17$). There was no interaction between tone category and measurement point ($F(2,46) = 0.82, p = 0.45$), showing that the slopes of the two tones were not different. The only interaction with the listener factor that even approached significance was listener \times tone category \times measurement point (i.e. listener \times relative slope) ($F(2,46) = 2.49, p = 0.09$), but the pattern as shown in Figure 3 does not admit of any simple interpretation.

[INSERT FIGURE 3 ABOUT HERE]

For the mean syllable durations in the within-position comparison (context yinping vs. context yangping), shown in Figure 4, a two-way ANOVA found no main effects for tone category ($F(1,23) = 2.03, p = 0.17$) or listener ($F(1,23) = 0.28, p = 0.60$); as with the across-position duration analysis, the interaction was also nonsignificant ($F(1,23) = 0.85, p = 0.37$).

[INSERT FIGURE 4 ABOUT HERE]

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 three-way ANOVA (tone category \times listener [absence/presence] \times measurement point [beginning/middle/end]) found a highly significant effect of measurement point ($F(2,46) = 53.35, p < 0.0001$) indicating that the slopes were not flat, and no main effect of listener ($F(1,23) = 0.25, p = 0.62$) indicating the lack of an overall effect of discourse factors, but there was also a surprisingly significant main effect of tone category ($F(1,23) = 14.70, p = 0.0009$): context yinping was an average of 2.5 Hz higher than context yangping. However, none of the interactions even approached significance (in particular, the listener \times tone interaction, $F(1,23) = 0.11, p = 0.74$), except for listener \times tone category \times measurement point (i.e. listener \times relative

slope) ($F(2,46) = 3.00, p = 0.06$). As can be seen in Figure 5, it appears that speakers kept the slopes for yinping and yangping parallel in the listener-absent condition, but when a listener was present, they distinguished the two tone categories more towards the end than the beginning of the contour.

[INSERT FIGURE 5 ABOUT HERE]

2.3 Discussion

The categoricity hypothesis predicts that f_0 should be fully neutralized in Taiwanese 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, at least in f_0 . All of these predictions were confirmed, except one: contradicting (2), context yinping and context yangping were not completely neutralized in f_0 . However, as we will explain, we have reasons for thinking that this finding does not represent an outright contradiction of the categoricity hypothesis.

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.

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 63 msec longer than context [33]. Yet the lack of a difference in f_0 slope normalized for duration (i.e. across measurement points defined relative to syllable edges) implies that the instructions for the articulation of the tonal contour are the same for juncture and context tones; in juncture position, the implementation of these instructions is merely slowed down.

Moreover, conforming to expectation (4), discourse context had no clear effects on the contrast between the two tone categories; 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. Interestingly, there was not only no effect of discourse context on overall f_0 , but also no effect on duration, which might be thought to be under greater speaker control; possibly the prosodic conditioning of duration is too automatized to allow for much pragmatic variability. The only pragmatic effect we found was a nearly significant and difficult to interpret pattern with f_0 slope ($p = 0.09$), but if anything, it appears that the relative slopes of the two tones were more neutralized with a listener present, suggesting that this effect, if real, was more haphazard than functional.

However, as noted above, the within-position comparisons did seem to suggest incomplete neutralization between context yinping and context yangping, with a very small (2.5 Hz) but

highly significant difference in mean f_0 . Moreover, 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 precisely what we found. 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). Instead, we found no difference in f_0 slope between these two tones.

Second, 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). At best there was a nearly significant change in relative slope ($p = 0.06$), but it went in 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. We are tempted to conclude that these variations, like the equally arbitrary variations in f_0 slope in the across-position comparisons, are signs that the speakers were struggling with an essentially impossible task: trying to make a contrast between two surface phonetic categories that they actually consider to be identical.

Third, 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.* (in press) 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).

Fourth, there is a real concern that our results may have reflected reading processes. The influence of orthography in incomplete neutralization studies is well-known; Warner *et al.* (in press) in particular 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 Taiwanese than of Mandarin, since as noted earlier Taiwanese does not have a long written tradition; reading Taiwanese 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 speaker/reader of Taiwanese 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 ($r(24) = -0.26, p = 0.22$). We can't ascribe the overall f_0 difference to disfluency alone, since reading character by character should also have resulted in f_0 slope differences that we did not find; we will address this issue further in the General Discussion.

One final reason for suspecting that our results are not due to preservation of an underlying juncture tone contrast is that the difference in overall f_0 between context yinping and context yangping was extremely small, a mere 2.5 Hz. It is not immediately clear if such a small difference (especially embedded in the variability noted above) can be perceivable or useful to listeners. The 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). Peng (2000) found a 2.3 Hz difference between tone 2 and sandhi tone 3 in the productions of Taiwan Mandarin speakers, and listeners could not discriminate between them. 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. Taiwanese 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 Taiwanese 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 Taiwanese speakers. If so, this difference must be an experimental artifact, probably due to reading pronunciations.

Our hypothesis that the small difference in overall f_0 found in Experiment 1 for the within-position comparisons was due to reading pronunciations, with no general implications for natural speech, would be strengthened if we could show that fluent Taiwanese speakers show no ability to perceive it. This was the purpose of Experiment 2.

3. Experiment 2

In this experiment, listeners were presented with within-position contrasts (context yinping vs. context yangping). We expected that they would neither be able to consistently discriminate between the two tone categories, nor be able to correctly identify them. In order to provide a baseline for comparison, we also presented listeners with the across-position contrast (juncture yangqu vs. context yinping), which we expected to be easy to distinguish. Although f_0 is neutralized in this pair, duration is not; juncture tones are much longer than context tones. The validity of our test as a measure of tone discrimination can thus be established if listeners show the ability to discriminate between these two tone categories.

3.1 Methods

3.1.1 Materials

The materials for Experiment 2 were whole sentences selected from the recordings produced in Experiment 1. Items were chosen to maximize ease of discrimination and identification. Thus they were chosen only from the listener-present condition, since these items were produced with the intent purpose of being discriminable and identifiable. The experimental items were taken from only three of the original speakers (whom we will identify as speakers 1, 2, and 3). These speakers were the only ones to meet our requirements of fluency scores in the top 50% percentile (speakers 1 and 2 in the top five) and clean production of the sentences chosen for the perception

experiment, with no extraneous noise and no reading errors (hesitations, repetitions, mispronunciations or misidentification of characters, whether target characters or not). A fourth speaker also had a high fluency score and made no reading errors, but there were extraneous noises on his recording, so his productions were used as practice items. Each speaker provided one token each of all three context yinping items and their three matching context yangping items (see Table III). He also provided one token each of two juncture yangqu items and their two matching context yinping items, namely, the [sun³³] and [ti³³] pairs (sentences 2a-b and 3a-b in Table II); the [kim³³] pairs were not used, since speaker 2 failed to produce these items correctly.

The productions of the three speakers chosen were for the most part typical of the set of speakers recorded for Experiment 1 (see Table IV). Speaker 2, in fact, was the "most typical" of all speakers, in the sense that his values for the four key measures (i.e. differences in duration and f_0 for the across-position and within-position comparisons) were on average closer to the sample means than any other speaker. All three speakers showed large duration differences in the same direction as the average for all speakers (juncture longer than context) and f_0 values slightly higher in juncture position than context position, as was paradoxically also the case for the sample means in the listener-present condition (see results section for Experiment 1). Similarly, for the within-position comparisons, f_0 values for all three speakers were higher in context yinping than in context yangping, and in fact these differences were greater and more consistent than for the speakers as a whole. However, there were differences among speakers for the syllable duration in the within-position comparison; the more fluent speakers 1 and 2 produced longer context yinping syllables than context yangping syllables, in the same direction as all speakers on average, while the less fluent speaker 3 showed a duration difference in the opposite direction. Finally, in the within-position comparison, all three speakers showed a greater f_0 contrast at the end of the tone contour than at the beginning, consistent with the nearly significant trend found in the listener-present condition (see results section for Experiment 1).

[INSERT TABLE IV ABOUT HERE]

3.1.2 Participants

The twenty-two listeners in Experiment 2 consisted of all of the original speakers from Experiment 1 that we were able to contact, including the three who provided the experimental stimuli (Experiment 2 was conducted 10 months after Experiment 1). Two of the participants in Experiment 2 were the two from Experiment 1 who were not recorded due to experimenter error, but for the remaining 20 we have their fluency scores. The mean fluency scores for this subset were not significantly different from those for the complete set of speakers recorded in Experiment 1.

3.1.3 Procedure

The experiment took place in a sound-attenuated room. Using the E-Prime experimental control software (Psychological Software Tools), spoken sentences were presented over headphones and pairs of sentences written in Chinese characters were presented side by side on a computer screen. In each visually presented sentence, the target character that distinguished it from the other sentence was highlighted in red (other characters were black against a white background). Each trial required participants to listen to the sentence, silently read the two paired sentences (the one on the left labeled "1" and the one on the right labeled "2"), and then press the appropriate button on a button box to indicate which written sentence matched the

spoken sentence. Participants were given 10 seconds from the offset of the auditory stimulus to respond; written sentences appeared on the screen at the same time as the spoken sentences began, giving participants up to 15 seconds to read (silently) the two sentences. Reaction times were not recorded.

A practice session was first conducted using the productions of one speaker as the auditory stimuli, followed by the experimental session using three speakers' productions. Trials were presented in the same order for each participant. In the experimental session items produced by the same speaker were presented in blocks (first speaker 1, then speaker 2, then speaker 3); this order reflected the speakers' relative fluency scores, from higher to lower. Each speaker block consisted of four sub-blocks containing trials representing the five discriminations of interest, always presented in the same order ([sun³³], [ti³³], [kun³³], [tsin³³], [t^hau³³]); note that within a block the two across-position trials always came before the three within-position trials. Blocking by speaker and comparison type and maintaining the same order of trials across sub-blocks was intended to make the task easier. For each across-position or within-position pair, participants were presented with all four combinations (2 spoken sentences \times 2 locations of the matching written sentence, i.e. left or right on screen) distributed randomly across the four blocks. This made a total of 60 trials for each participant (3 speakers \times 5 sentence pairs \times 4 combinations), 24 of them across-position trials and 36 within-position trials.

3.2 Results

There were two issues posed by the data, namely whether listeners could discriminate between tone categories and how they identified tone categories. The first issue was addressed with signal detection theory, while the second was addressed with ANOVAs using the proportion of particular responses for a given tone category.

Discrimination was measured using the d' statistic. For each listener, we calculated separate d' values for the across-position comparisons and the within-position comparisons, arbitrarily choosing to count "hits" whenever participants correctly identified juncture yangqu in the across-position comparisons, and context yinping in the within-position comparisons. The d' values represented the listeners' ability to discriminate between two matched tone categories, even if their identifications were (consistently) incorrect. The mean d' value for the across-position comparisons (1.79) was higher than that for the within-position comparisons (0.31), a highly significant difference by a paired t-test ($t(21) = 9.17, p < 0.0001$).

We then calculated 95% confidence intervals for these d' values for each listener using the method described in Macmillan & Creelman (1991:271) (originally proposed by Gourevitch & Galanter, 1967). This allowed us to determine whether the d' values indicated discrimination significantly better than chance ($d' = 0$). For the across-position comparisons, all but four of the 22 listeners showed significant d' values. For the within-position comparisons, however, only one listener showed a significant d' value, and this was one of the three whose productions were used as stimuli (speaker 1; more on this below).

Given the large size and consistency of the duration differences between juncture and context tones, we assumed that any listener unable to discriminate in the across-position comparisons was simply not a sufficiently fluent listener/reader of Taiwanese to meet the challenges of this experiment. This assumption was supported by a relatively large and nearly significant positive correlation between d' and participant fluency scores (based on their productions as speakers in Experiment 1) for the across-position comparisons ($r(20) = 0.43, p = 0.06$). By contrast, correlations between fluency and within-position d' came nowhere near

significance, nor did correlations between across-position and within-position d' ($ps > 0.4$).

The remaining analyses involving tone identification were conducted only on the 18 listeners who showed significant d' values for the across-position comparisons. These analyses used as data points the proportion of responses in which listeners chose the juncture yangqu sentence (for the across-position comparisons) or the context yinping sentence (for the within-position comparisons); again the choice of these as the target categories for analysis were arbitrary. Separate three-way speaker \times tone category \times item repeated measure ANOVAs, with participant as random variable, were conducted for the across-position and within-position comparisons.

Mean proportions of juncture yangqu responses in the across-position comparisons are shown in Figure 6, divided by item ([sun³³] and [ti³³] target sentences) and by target tone category (juncture yangqu vs. context yinping). There was a highly significant main effect of tone category ($F(1,17) = 309.48, p < 0.0001$), with juncture sentences far more likely to elicit juncture responses than context sentences; the mean proportion of juncture responses to juncture sentences was 94%, while the mean proportion of juncture responses to context sentences was 32%. There was also a highly significant main effect of item ($F(1,17) = 58.78, p < 0.0001$), with [sun³³] items more likely to elicit juncture responses than [ti³³] items (78% vs. 48%, respectively). The item factor also showed a highly significant interaction with tone category ($F(1,17) = 40.14, p < 0.0001$); the difference in proportions of juncture responses to the contrasting [ti³³] sentences (juncture vs. context) was larger than that for the [sun³³] sentences (92% vs. 34%, respectively). These asymmetries seem to be entirely attributable to the semantic biasing described in the Materials section for Experiment 1; in the [sun³³] pair, the juncture sentence was judged by native speakers to be more natural than the matched context sentence, while the [ti³³] pair, if biased at all, was biased more towards the juncture sentence.

[INSERT FIGURE 6 ABOUT HERE]

Although there was no main effect of speaker for the across-position comparisons ($F(2,34) = 0.87, p = 0.43$), this factor did interact with the other two factors. Significant speaker \times tone category ($F(2,34) = 5.70, p = 0.007$) and nearly significant speaker \times item ($F(2,34) = 3.13, p = 0.06$) interactions both imply a change in how the listeners responded during the course of the experiment: the difference in proportions of juncture responses for juncture sentences and for context sentences increased across the speakers, while the difference in responses to [sun³³] and [ti³³] sentences (apparently due to semantic biasing) decreased across the speakers. Apparently these changes were due to increased experience with the demands of the experiment; over time attention was focused more on correct identification, and response biases were suppressed. There was no significant three-way interaction between Speaker, Tone, and Item ($F(2, 34) = 0.39, p = 0.68$).

Mean proportions of context yinping responses in the within-position comparisons are shown in Figure 7, divided by item ([kun³³], [t^hau³³], [tsin³³]) and by target tone category (context yinping vs. context yangping). There was no main effect of tone category ($F(1,17) = 2.89, p = 0.11$), with the mean proportion of yinping responses to yinping sentences (47%) quite close to the mean proportion of yinping responses to yangping sentences (42%). Unlike the across-position comparisons, there was a marginal main effect of speaker ($F(2,34) = 2.40, p = 0.045$); speaker 2 elicited more yinping responses than the other two speakers. It's not obvious if this means anything; speaker 2's context yinping and context yangping productions were acoustically

less distinct than speaker 1's, both in overall f_0 and in duration. As with the across-position comparisons, there was a highly significant main effect of item ($F(2,34) = 10.34, p = 0.0003$); [kun³³] sentences tended to elicit yinping responses, [t^hau³³] sentences tended to elicit yangping responses, while [tsin³³] sentences showed no bias either way. These bias patterns matched the naturalness judgments described in the materials section for Experiment 1; the [tsin³³] sentences involved two proper names, and thus unlike the other pairs, both were considered equally natural. There was also a highly significant tone category \times item interaction ($F(2,34) = 10.30, p = 0.0003$): for [kun³³] sentences, listeners correctly gave yinping responses more often to yinping sentences than to yangping sentences, [t^hau³³] sentences showed the opposite pattern, and there was virtually no difference for [tsin³³] sentences. No other effects in this ANOVA approached significance ($ps > 0.1$).

[INSERT FIGURE 7 ABOUT HERE]

The difference in proportion for the within-position [kun³³] sentences (23%) was far smaller than the mean difference for the across-position sentences (63%) and was not significant by a Tukey HSD performed on the tone category \times item interaction ($p = 0.11$), but nevertheless there were more yinping responses to genuine yinping items. After careful relistening to the stimuli, it appears that the speakers emphasized the rarer word ([kun³³-tui³³], literally "group team") via a slight preceding pause and increased amplitude, but didn't do this for the other sentences, which either already had a preceding pause (the [tsin³³] sentences) or had pragmatically more balanced target items (the [t^hau³³] sentences). Even if some listeners were able to pick up on this tendency, this is quite different from being able to identify the tone categories themselves.

Since three of the participants in the perception experiment were the same three speakers who provided the stimuli, we performed separate two-way (speaker \times tone category) factorial ANOVAs for each participant. For the across-position comparisons, all three showed a significant main effect of tone category ($ps < 0.01$) but no effect of speaker and no interaction, suggesting that distinguishing juncture yangqu and context yinping was equally easy regardless of the speaker. For the within-position comparisons, only speaker 1 showed a significant main effect of tone category ($F(1,30) = 4.74, p = 0.04$); the others all had F s < 1.5 and $ps > 0.2$ for main effects and interactions. However, speaker 1 was also the only one to show a nearly significant speaker \times tone category interaction ($F(2,30) = 2.76, p = 0.08$), since his success was due solely to his ability to correctly identify context yinping vs. context yangping in his own speech. The fact that speaker 1's identification ability disappeared with the speech of speaker 2, whose productions showed virtually identical acoustic cues, suggests that speaker 1 may have relied more on latent memories of his participation in Experiment 1 than on a genuine ability to pick up on the cues. Speaker 3 himself also showed a trend to identify his own tone productions more accurately than productions by the other two speakers, but it was nonsignificant. Speaker 2 showed no ability to identify tones correctly at all, not even the ones that he himself had produced; intriguingly, though, he showed a nonsignificant trend to reverse the identification of tones produced by speaker 3, whose productions of the within-context contrast showed a duration pattern (but not f_0 pattern) opposite to the other two speakers.

3.3 Discussion

The results of the perception experiment confirmed our expectations: listeners demonstrated no ability to discriminate between context yinping and context yangping, let alone identify them

correctly. This was so in spite of the facts that the differences in f_0 were greater in the perception stimuli than the average that emerged from Experiment 1, and that the listeners had read and produced all of the experimental items themselves before, as participants in Experiment 1; even the three listeners who were presented with their own productions did not perform well. Their overall failure with the within-position comparisons cannot be ascribed to difficulty with the experimental procedure alone, since most listeners had no trouble discriminating and correctly identifying the across-position comparisons between juncture yangqu and context yinping. It is true that four of the 22 participants failed to discriminate the across-position contrasts significantly better than chance, but problems in making this discrimination were apparently due to fluency; by contrast, fluency did not show any sign of being responsible for the listeners' problems discriminating the within-position comparisons. Moreover, even the listeners who easily discriminating the across-position comparisons showed no ability to discriminate or identify the tone categories in the within-position comparisons, except for one listener when listening to his own productions.

Of course, failure to demonstrate an ability does not mean that the ability does not exist. Listeners in Experiment 2 were only given 10 seconds after the presentation of the auditory target sentence to make their decision about which of the two visually presented sentences was correct. Even with the longest sentences, however, 10 seconds would give them over 300 msec to read each character, far longer than is necessary for fluent readers to read Chinese. Fixation times and length of saccades are roughly comparable between Chinese and English readers, namely about 250 msec and 1.8 words respectively (see review in Hoosain, 1991:56-7); even with the extra time probably needed to read in Taiwanese rather than Mandarin, these numbers still imply that a considerable amount of time (between 3 and 6 seconds) was left to make a decision after reading both sentences. Recall also that the target characters distinguishing the two sentences were highlighted in red. There can be no doubt that the participants knew the pronunciations of the syllables represented by these highlighted characters; after all, it was the significant difference in f_0 in their own productions from Experiment 1 that prompted Experiment 2. Moreover, while participants were not permitted to listen to an item more than once during a trial, each particular auditory stimulus was presented four times. Familiarity with the written stimuli also must have increased during the course of the experiment, as each sentence appeared on screen twelve times.

Even if our task was artificially difficult, however, this alone cannot explain why listeners performed so much better with the across-position comparisons than the within-position comparisons. We claim that the reason for this difference was that the former involved a reliable cue (i.e. duration) also used by speakers and listeners in natural settings, while the latter did not. The significant f_0 pattern found in Experiment 1 for the within-position contrast, while perhaps having interesting implications for the nature of Chinese reading pronunciations, is not useful to listeners.

4. General discussion

Our overall results are consistent with the claim that Taiwanese tone sandhi is a categorical (i.e. neutralizing) pattern, as expected from its essentially lexicalized nature. Experiment 1 found no difference in f_0 contours between juncture and context tones (i.e. "basic" and "sandhi" tones), replicating the findings of Tsay, Charles-Luce, & Guo (1999) and Tsay & Myers (2001). We did also replicate the small difference between the lexically distinct mid level tones in context position found by Tsay & Myers (2001), but as we argued in the Discussion section for Experiment 1, this seems to be better ascribed to the nature of the task (reading rather than

spontaneous speech). Moreover there was only a small, not quite significant discourse-dependent variation in the degree of neutralization in context position, and this was not in overall f_0 but in a change in slope that went opposite to the direction expected from incomplete neutralization. Experiment 2 showed that listeners who have no difficulty distinguishing juncture and context tones (presumably due to the large differences in syllable durations) cannot discriminate, let alone identify, the two mid level tones in context position (where the only phonetic contrast lies in the small f_0 difference).

The fact that Experiment 1 found no overall difference in slope between the two mid level tones in context position is worth highlighting. 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). 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 Taiwanese). The lack of slope differences in our Taiwanese results are thus incompatible with the proposed mechanisms of incomplete neutralization.

As an alternative explanation of our results, we suggest that a simple model of reading pronunciations may account both for the small overall f_0 difference and the lack of a slope difference. First, readers carry out visual lexical access of a character, which leads to the activation of its citation (juncture) form. Further processing of the character in sentential context leads to access of the context form, and termination of the processing of the juncture form. By this point, however, some of the articulatory program required to initiate tone production would have already begun. Since juncture yinping [55] begins at a higher point than juncture yangping [24], the result would be an overall difference in f_0 , but since the target articulation for both tone categories in context position is now the level tone [33], the slopes would be identical. This model would not predict overall f_0 differences in the juncture vs. context comparison, since in the juncture yangqu sentences, target words were followed orthographically with a comma, making them clearly distinct from the context sentences; this should have reduced the activation of the competing juncture forms during the reading of the context sentences. This model makes a number of further predictions. In particular, we would expect that juncture tones whose contours differ more at the offset than the onset should be fully neutralized even in context position; neutralizing tones like this don't exist in any variety of Southern Min, but they may in other languages with lexicalized phrasal tone patterns. Moreover, even in Taiwanese a similar interference effect may be found for the reading pronunciations of homographic morphemes that differ in tone (a subclass of what are known in Chinese linguistics as *poyinzi*, literally "broken-sound characters"); that is, interference on dynamic parameters like tone may be restricted to the onset setting only.

One interesting indirect argument for the validity of our conclusions is the observation that speakers of Taiwan Mandarin seem to show a greater degree of neutralization between tone 2 and sandhi tone 3 than Beijing Mandarin speakers, as observed in section 1.3 of the introduction. The greater degree of neutralization in Taiwan Mandarin, if genuine, may indicate that native speakers of Taiwanese (as well as Mandarin speakers growing up in an environment rich in Taiwanese-accented Mandarin) have reanalyzed Mandarin tone sandhi, nonneutralizing in its original form, as categorical under the influence of their own categorical tone sandhi system.

As noted in the introduction, there are several independent arguments supporting the claim

that Taiwanese 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, Taiwanese 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 on-line 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 Taiwanese 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, Taiwanese 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). While we believe that this trend is quite exciting and promising, categorical phonological knowledge does exist and its origin and processing remain important challenges.

Acknowledgments

Basic results for Experiment 1 were presented at the Phonetics Workshop at Chiao Tung University, Taiwan, in March 2002. Experiment 1 was conducted with the help of Joyce H.-C. Liu. Experiment 2 was conducted with the help of Wen-Ling Wang. We would like to thank Yueh-chin Chang, Mirjam Ernestus, Elliott Moreton, Shu-hui Peng, and Natasha Warner for insights and useful references, and Jan Charles-Luce for providing the paradigm that we followed in Experiment 1. Research was funded by National Science Council (Taiwan) grant NSC-90-2411-H-194-023.

References

- Bodman, N. C. (1955) *Spoken Amoy Hokkien*. Revised and published 1987, Ithaca, NY: Spoken Language Services.
- Boersma, P. (1998) *Functional phonology*. The Hague: Holland Academic Graphics.
- Borowsky, T. (1993) On the word level. In *Phonetics and phonology: studies in lexical phonology* (S. Hargus & E. M. Kaisse, editors), pp. 199-234. New York: Academic Press.
- Browman, C. P. & Goldstein, L. M. (1990) Tiers in articulatory phonology, with some implications for casual speech. In *Papers in laboratory phonology 1: between the grammar and physics of speech* (J. Kingston & M. E. Beckman, editors), pp. 341-376. Cambridge: Cambridge University Press.
- Bybee, J. L. (1994) A view of phonology from a cognitive and functional perspective, *Cognitive Linguistics*, **5**, 285-305.
- Bybee, J. L. (2000) The phonology of the lexicon: evidence from lexical diffusion. In *Usage-based models of language* (M. Barlow & S. Kemmer, editors), pp. 65-85. Stanford: CSLI Publications.
- Chang, Y.-C. (1988) Sandhi tonal des syntagmes dissyllabiques du Min-nan parlé à Taiwan [Tone sandhi in disyllabic compounds in Minnan spoken in Taiwan], *Cahiers de Linguistique Asie Orientale*, **17**, 193-234.
- Chang, Y.-C. & Su, Y.-C. (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, Y.-R. (1930) A system of tone letters, *Le Maître Phonétique*, **45**, 24-27.
- Chao, Y.-R. (1948) *Mandarin primer*. Cambridge, MA: Harvard University Press.
- Charles-Luce, J. (1985) Word-final devoicing in German: effects of phonetic and sentential contexts, *Journal of Phonetics*, **13**, 309-324.
- Charles-Luce, J. (1993) The effects of semantic context on voicing neutralization, *Phonetica*, **50**, 28-43.
- Charles-Luce, J. (1997) Cognitive factors involved in preserving a phonemic contrast, *Language and Speech*, **40**, 229-248.
- Charles-Luce, J. & Dressler, K. M. (1999) The effects of semantic predictability in non-pathological older adults' production of a phonemic contrast. *Clinical Linguistics and Phonetics*, **13**, 199-217.
- Charles-Luce, J., Dressler, K. M. & Ragonese, E. (1999) Effects of semantic predictability on children's preservation of a phonemic voice contrast, *Journal of Child Language*, **26**, 505-530.
- Chen, M. Y. (1987) The syntax of Xiamen tone sandhi, *Phonology Yearbook*, **4**, 109-149.
- Chen, M. Y. (1996) Tone geometry -- a Chinese perspective. In *New horizons in Chinese linguistics* (C.-T. J. Huang & Y.-H. Audrey Li, editors), pp. 21-48. Dordrecht: Kluwer Academic Publishers.
- Chen, M. Y. (2000) *Tone sandhi: patterns across Chinese dialects*. Cambridge: Cambridge University Press.
- Cheng, R. L. (1968) Tone sandhi in Taiwanese, *Linguistics*, **41**, 19-42.
- Di Paolo, M. & Faber, A. (1990) Phonation differences and the phonetic content of the tense-lax contrast in Utah English, *Language Variation and Change*, **2**, 155-204.
- Du, T.-C. (1988) *Tone and stress in Taiwanese*. PhD thesis, University of Illinois at Urbana-Champaign.
- Ernestus, M. & Baayen, H. (2002) The functionality of incomplete neutralization in Dutch: the

- case of past-tense formation. Paper presented at the Eighth Conference on Laboratory Phonology, Yale, June 27-30.
- Faber, A. & M. Di Paolo (1995) The discriminability of nearly merged sounds. *Language Variation and Change*, **7**, 35-78.
- Fon, J. (1997) *What are tones really like? An acoustic-based study of Taiwan Mandarin tones*. MA thesis, National Taiwan University.
- Fourakis, M. & Iverson, G. K. (1984) On the 'incomplete neutralization' of German final obstruents, *Phonetica*, **41**, 140-149.
- Gerfen, C. & Hall, K. (2001) Coda aspiration and incomplete neutralisation in Eastern Andalusian Spanish. University of North Carolina at Chapel Hill ms., submitted.
- Gourevitch, V. & Galanter, E. (1967) A significance test for one-parameter isosensitivity functions. *Psychometrika*, **32**, 25-33.
- Guy, G. R. (1980) Variation in the group and the individual: the case of final stop deletion. In *Locating Language in Time and Space* (W. Labov, editor), pp. 1-54. New York: Academic Press.
- Guy, G. (1991) Explanation in variable phonology: an exponential model of morphological constraints, *Language Variation and Change*, **3**, 1-22.
- Hayes, B. (1990) Precompiled phrasal phonology. In *The phonology-syntax connection* (S. Inkelas & D. Zec, editors), pp. 85-108. Chicago: University of Chicago Press.
- Hockett, C. F. (1947) Peiping phonology, *Journal of American Oriental Society*, **67**, 253-267. Reprinted 1964 in *Readings in Linguistics I, Fourth Edition* (M. Joos, editor), pp. 217-228. Chicago: University of Chicago Press.
- Hoosain, R. (1991) *Psycholinguistic implications for linguistic relativity: a case study of Chinese*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Hsiao, Y. E. (1991) *Syntax, rhythm and tone: a triangular relationship*. PhD dissertation, University of California, San Diego.
- Hsieh, H.-I. (1970) The psychological reality of tone sandhi rules in Taiwanese, *Proceedings of Chicago Linguistics Society*, **6**, 489-503.
- Hsieh, H.-I. (1975) How generative is phonology? In *The transformational-generative paradigm and modern linguistic theory* (E. F. K. Koerner, editor), pp. 109-144. Amsterdam: John Benjamins.
- Hsieh, H.-I. (1976) On the unreality of some phonological rules, *Lingua*, **38**, 1-19.
- Huang, S.-F. (1993) *Yuyan, shehui yu zuqun yishi*. [Language, society and ethnic identity]. Taipei: Crane Publishing.
- Kaisse, E. M. (1985) *Connected speech: the Interaction of syntax and phonology*. New York: Academic Press.
- Kaisse, E. M. & Hargus, S. (1993) Introduction. In *Phonetics and phonology 4: studies in Lexical Phonology* (S. Hargus & E. M. Kaisse, editors), pp. 1-19. New York: Academic Press.
- Kiparsky, P. (1982) Lexical phonology and morphology. In *Linguistics in the morning calm* (I. S. Yang, editor), pp. 135-160. Seoul: Hanshin Publishing Company.
- Kiparsky, P. (1985) Some consequences of Lexical Phonology, *Phonology Yearbook*, **2**, 85-138.
- Kratochvil, P. (1986) The case of the third tone. In *Wang Li Memorial Volumes (English Volume)* (Chinese Language Society of Hong Kong, editors), pp. 253-276. Hong Kong: Joint Publishing Co.
- Labov, W. (1994) *Principles of linguistic change: internal factors*. Oxford: Blackwell.
- Labov, W., Karen, M. & Miller, C. (1991) Near-mergers and the suspension of phonemic

- contrast. *Language Variation and Change*, **3**, 33-74.
- Lin, H.-B. (1988) *Contextual stability of Taiwanese tones*. PhD thesis, University of Connecticut.
- Lin, J.-W. (1994) Lexical government and tone group formation in Xiamen Chinese, *Phonology*, **11**, 237-275.
- Macmillan, N. A. & Creelman, C. D. (1991) *Detection theory: a user's guide*. Cambridge, UK: Cambridge University Press.
- Manaster Ramer, A. (1996) A letter from an incompletely neutral phonologist, *Journal of Phonetics*, **24**, 477-489.
- Martin, S. E. (1957) Problems of hierarchy and indeterminacy in Mandarin phonology, *Bulletin of the Institute of History and Philology*, **29**, 209-230.
- Moreton, E. (1999) Non-computable functions in Optimality Theory. University of Massachusetts at Amherst ms. Available at roa.rutgers.edu.
- Myerson, R. F. (1978) Children's knowledge of selected aspects of Sound Pattern of English. In *Recent advances in the psychology of language: formal and experimental approaches* (R. N. Campbell & P. T. Smith, editors), pp. 377-402. New York: Plenum Press.
- Odden, D. (1987) Kimatuumbi phrasal phonology, *Phonology Yearbook*, **4**, 13-26.
- Odden, D. (1990) Syntax, lexical rules and postlexical rules in Kimatuumbi. In *The phonology-syntax connection* (S. Inkelas & D. Zec, editors), pp. 259-278. Chicago: University of Chicago Press.
- Ohala, J. J. (1974) Experimental historical phonology. In *Historical linguistics 2: theory and description in phonology* (J. M. Anderson & C. Jones, editors), pp. 353-387. Amsterdam: North-Holland Publishing Company.
- Peng, S.-H. (1997) Production and perception of Taiwanese tones in different tonal and prosodic context, *Journal of Phonetics*, **25**, 371-400.
- Peng, S.-H. (1998) Evidence for a UR-free account of tone sandhi in Taiwanese. Poster presented at the Sixth Conference of Laboratory Phonology, University of York, July 2-4.
- Peng, S.-H. (2000) Lexical versus 'phonological' representations of Mandarin sandhi tones. In *Papers in laboratory phonology 5: acquisition and the lexicon* (M. B. Broe & J. B. Pierrehumbert, editors), pp. 152-167. Cambridge: Cambridge University Press.
- Pierrehumbert, J. B. (2001) Exemplar dynamics: word frequency, lenition and contrast. In *Frequency and emergence of linguistic structure* (J. Bybee & P. Hopper, editors), pp. 137-157. Amsterdam: John Benjamins.
- Port, R. F. (1996) The discreteness of phonetic elements and formal linguistics: response to A. Manaster Ramer, *Journal of Phonetics*, **24**, 491-511.
- Port, R. & O'Dell, M. (1985) Neutralization of syllable-final voicing in German, *Journal of Phonetics*, **13**, 455-471.
- Port, R. & Crawford, P. (1989) Incomplete neutralization and pragmatics in German, *Journal of Phonetics*, **17**, 257-282.
- Rubach, J. (1990) Final devoicing and cyclic syllabification in German, *Linguistic Inquiry*, **21**, 79-94.
- Shen, X. S. (1990) Tonal coarticulation in Mandarin, *Journal of Phonetics*, **18**, 281-295.
- Steinberg, D. D. & Krohn, R. K. (1975) The psychological validity of Chomsky and Halle's vowel shift rule. In *The Transformational-Generative Paradigm and Modern Linguistic Theory* (E. Koerner, editor), pp. 233-259. Amsterdam: John Benjamins.
- Steriade, D. (2000) Paradigm uniformity and the phonetics-phonology boundary. In *Papers in laboratory phonology 5: acquisition and the lexicon* (M. B. Broe & J. B. Pierrehumbert,

- editors), pp. 313-334. Cambridge: Cambridge University Press.
- Ting, P-H. (1970) *Taiwan yuyan yuanliu*. [The origins of the languages in Taiwan] Taichung, Taiwan: Press Secretary's Office, Taiwan Provincial Government.
- Tsay, J. (1994) *Phonological pitch*. PhD dissertation, University of Arizona, Tucson.
- Tsay, J., Charles-Luce, J. & Guo, Y.-S. (1999) The syntax-phonology interface in Taiwanese: acoustic evidence, *Proceedings of the XIVth International Congress of Phonetic Sciences* (J. J. Ohala, Y. Hasegawa, M. Ohala, D. Granville & A. C. Bailey, editors), pp. 2407-2410. San Francisco, August 1-7.
- Tsay, J. & Myers, J. (1996) Taiwanese tone sandhi as allomorph selection. *Proceedings of the Berkeley Linguistic Society*, **22**, 394-405.
- Tsay, J. & Myers, J. (2001) Processes in the production of Taiwanese tone sandhi: an acoustic phonetic study. *The Proceeding of 5th National Conference on Modern Phonetics* (Cai Lianhong, Zhou Tongchun, Tao Jianhua, editors). Beijing: Tsinghua University.
- Tseng, C.-C. (1995) *Taiwanese prosody: an integrated analysis of acoustic and perceptual data*. PhD thesis, University of Hawaii.
- Wang, S. H. (1995) *Experimental studies in Taiwanese phonology*. Taipei: Crane Publishing.
- Wang, W. S-Y. (1967) Phonological features of tone, *International Journal of American Linguistics*, **33**, 93-105.
- Wang, W. S-Y. & Li, K.-P. (1967) Tone 3 in Pekinese, *Journal of Speech and Hearing Research*, **10**, 629-236.
- Warner, N., Jongman, A., Sereno, J. & Kemps, R. (in press) Sub-phonemic durational differences in production and perception, *Journal of Phonetics*.
- Xu, Y. (1993) *Contextual tonal variation in Mandarin Chinese*. PhD thesis, University of Connecticut.
- Xu, Y. (1997) Contextual tonal variations in Mandarin, *Journal of Phonetics*, **25**, 61-83.
- Yip, M. (1980) *The tonal phonology of Chinese*. PhD thesis, MIT. Reprinted by Garland in 1990 as *The tonal phonology in Chinese*.
- Yuan, J.-H. (ed.) (1960) *Hanyu fangyang gaiyao*. [Outline of Chinese Dialects] Beijing: Wenzhi Gaige Chubanshe.
- Zee, E. (1980) A spectrographic investigation of Mandarin tone sandhi. *UCLA Working Papers*, **49**, 98-116.
- Zsiga, E. C. (1993) *Features, gestures, and the temporal aspects of phonological organization*. New Haven, CT: Yale University dissertation.
- Zsiga, E. C. (1995) An acoustic and electropalatographic study of lexical and post-lexical palatalization in American English. In *Papers in laboratory phonology 4: Phonology and phonetic evidence* (B. Connell & A. Arvaniti, editors), pp. 282-302. Cambridge: Cambridge University Press.
- Zsiga, E. C. (1997) Features, gestures, and Igbo vowels: an approach to the phonology-phonetics interface, *Language*, **73**, 227-274.

Table I.

Taiwanese 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.

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

Table II.

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} , sə⁵¹a²¹ kap⁵³ a³³hĩã⁵⁵
 this COP Aunt Sister-in-law and Elder Brother
 "These are Aunt, Sister-in-law and Elder Brother."
- 1b tse⁵⁵ si²¹ [a³³**kim**³³-sə⁵¹a²¹]_{NP} kap⁵³ a³³hĩã⁵⁵
 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^hun³³kiau⁵⁵ kap⁵³ sə⁵¹hua²⁴
 this COP (name) (name) and (name)
 "These are Limsun, Tshunkiau and Soohua."
- 2b tse⁵⁵ si²¹ [lim³³**sun**³³-ts^hun³³kiau⁵⁵]_{NP} kap⁵³ sə⁵¹hua²⁴
 this COP (name) and (name)
 "These are Limsuntshunkiau and Soohua."
- 3a tse⁵⁵ si²¹ [**tʰi**³³]_{NP} , ba^{ʔ21} kap⁵³ ũa⁵¹
 this COP chopstick meat and bowl
 "These are chopsticks, meat and bowl."
- 3b tse⁵⁵ si²¹ [**tʰi**³³-ba^{ʔ21}]_{NP} kap⁵³ ũa⁵¹
 this COP pig meat and bowl
 "These are pork and bowl."

Table III.

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 [**kuan**³³ tə⁵⁵a⁵¹]_{VP} kap⁵³ kuan³³ i⁵⁵a⁵¹
 I want donate table and donate chair
 "I want to donate tables and chairs."
- 1b gua⁵⁵ be⁵⁵ [**kuan**³³ tə⁵⁵a⁵¹]_{NP} kap⁵³ kuan³³ i⁵⁵a⁵¹
 I want high table and high chair
 "I want high tables and high chairs."
- 2a i³³ ai⁵¹ ts^ham³³ka³³ [**kun**³³-tui³³]_{NP} e³³ ua²¹taŋ³³
 he like participate military team MOD activity
 "He likes to participate in military activities."
- 2b i³³ ai⁵¹ ts^ham³³ka³³ [**kun**³³-tui³³]_{NP} e³³ ua²¹taŋ³³
 he like participate group team MOD activity
 "He likes to participate in group activities."
- 3a tsit⁵³e⁵⁵ si²¹ pɔ⁵¹te²¹hi²¹ e³³ [**tsin**³³ke³³sian⁵⁵]_{NP}
 this COP puppet show MOD (name)
 "This is Tsinkesian in the puppet show."
- 3b tsit⁵³e⁵⁵ si²¹ pɔ⁵¹te²¹hi²¹ e³³ [**tsin**³³ke³³sian⁵⁵]_{NP}
 this COP puppet show MOD (name)
 "This is Tsinkesian in the puppet show."
- 4a i³³ ki³³sut⁵³ bə³³hə⁵¹ [**t^hau**³³tsit²¹kai⁵¹]_{VP} tə²¹ hɔ²¹ laŋ³³ lia⁵¹tiə²¹
 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 ki³³sut⁵³ bə³³hə⁵¹ [**t^hau**³³tsit²¹kai⁵¹]_{VP} tə²¹ hɔ²¹ laŋ³³ lia⁵¹tiə²¹
 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)."

Table IV.

Means (and standard deviations) for f_0 differences in Hz and syllable duration differences in msec for the three speakers whose productions were used in Experiment 2.

	Across-position differences (junction yangqu - context yinping)				Within-position (context yinping - context yangping)			
	initial f_0	mid f_0	end f_0	syllable duration	initial f_0	mid f_0	end f_0	syllable duration
speaker 1	18.3 (19.4)	2.8 (12.4)	2.3 (5.3)	128 (41)	4.7 (9.8)	5.5 (7.3)	10.0 (12.3)	-65 (6)
speaker 2	4.8 (9.5)	6.3 (6.0)	2.0 (7.1)	85 (20)	0.3 (7.5)	5.8 (5.3)	2.8 (2.0)	-68 (3)
speaker 3	1.8 (12.4)	-2.5 (0.7)	11.8 (11.7)	106 (7.8)	7.3 (2.4)	9.5 (1.8)	12.2 (2.1)	46 (2)

Figure 1.

The Southern Min long tone circle as found in our speakers' variety of Taiwanese. Arrows follow the traditional assumption that context tones are derived from juncture tones.

Figure 2.

Mean syllable durations for juncture yangqu vs. context yinping in listener absent vs. present conditions.

Figure 3.

Mean f_0 values at three measurement points for juncture yangqu vs. contour yinping in listener absent vs. present conditions.

Figure 4.

Mean syllable durations for context yinping vs. context yangping, in listener absent vs. present conditions.

Figure 5.

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

Figure 6.

Mean proportions of juncture yangqu responses in across-position comparisons (juncture yangqu vs. context yinping), divided by item and target tone category.

Figure 7.

Mean proportions of context yinping responses in within-position comparisons (context yinping vs. context yangping), divided by item and target tone category.

Figure 1.

$$\begin{array}{ccccc} [55] & \rightarrow & [33] & \leftarrow & [24] \\ \uparrow & & \downarrow & & \\ [51] & \leftarrow & [21] & & \end{array}$$

Figure 2.

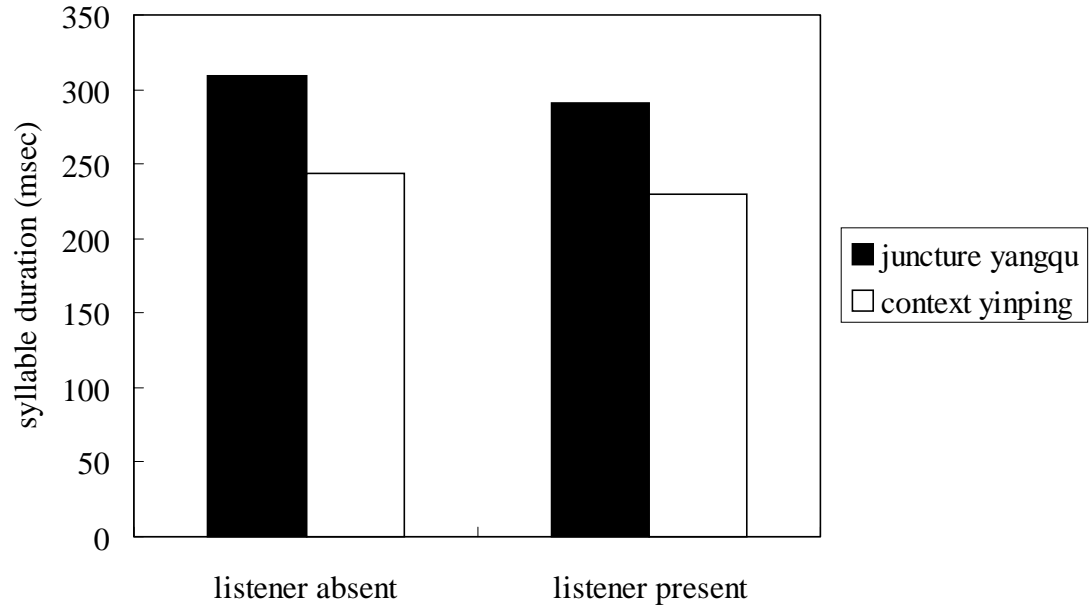


Figure 3.

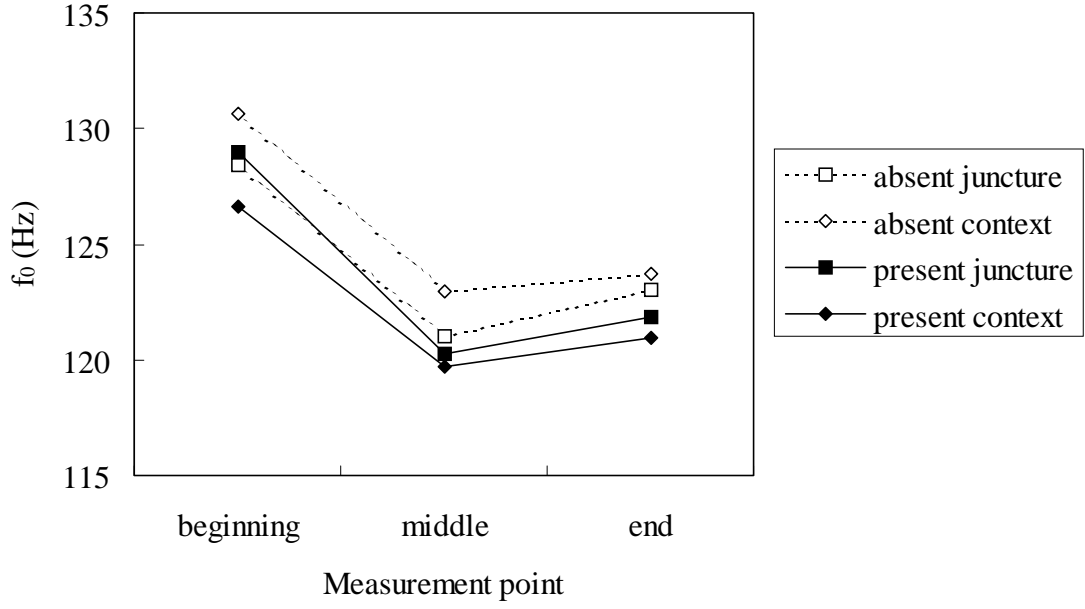


Figure 4.

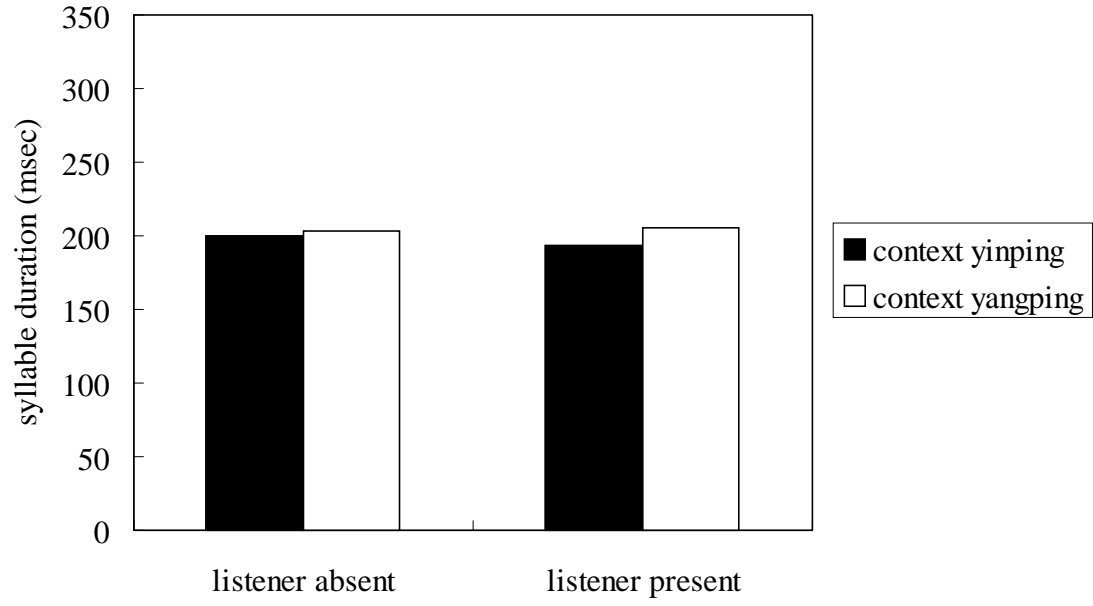


Figure 5.

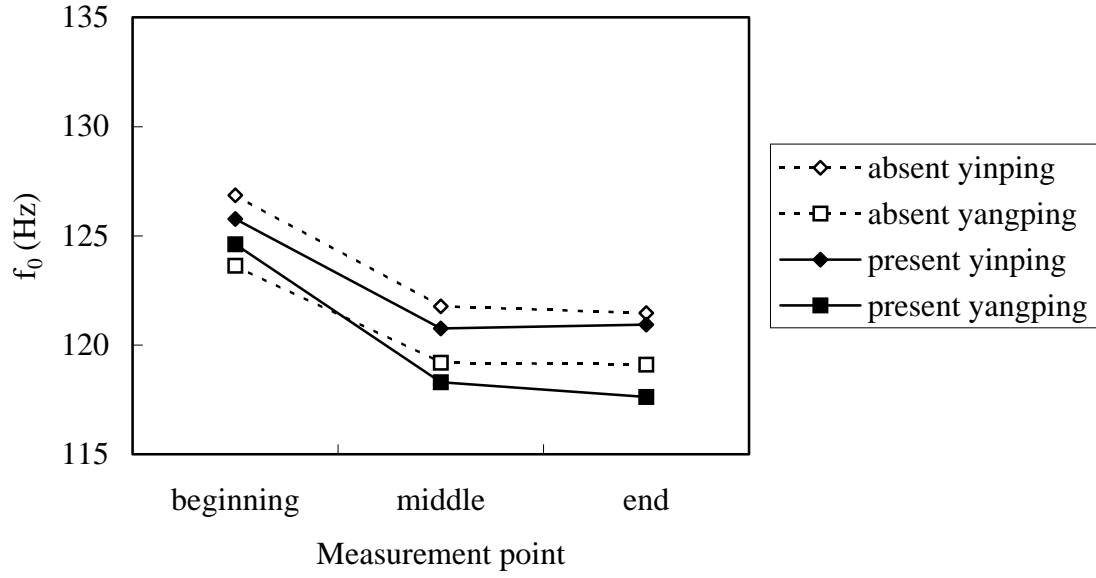


Figure 6.

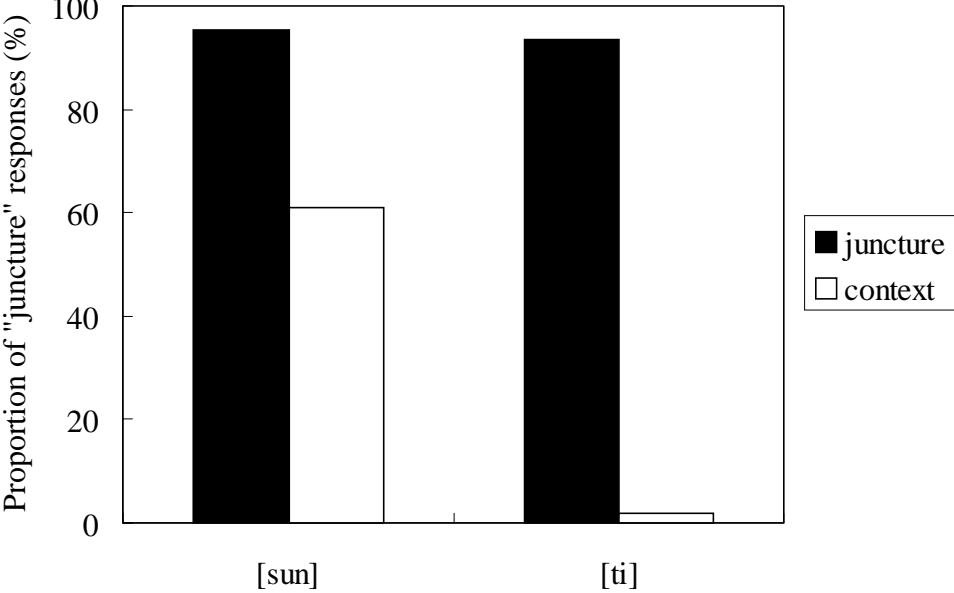


Figure 7.

