Taiwan Journal of Linguistics
Vol. 1.1, 29-68, 2003

INVESTIGATING THE PHONETICS OF MANDARIN TONE SANDHI*

James Myers and Jane Tsay

ABSTRACT

The goal of this paper is to discuss phonetic evidence relating to the production of tone sandhi in Mandarin. Two views of tone sandhi are contrasted: the standard "categorical" view, whereby tone 3 sandhi is treated similarly to *a/an* allomorphy in English, and an alternative "gradient" view, which treats tone 3 sandhi as similar to flapping in English. Evidence is presented for two main claims. First, based on a review of the phonetic literature, it appears that native speakers of Beijing Mandarin process tone sandhi more in line with the gradient view, while speakers of other varieties (including Taiwan Mandarin) process tone sandhi in accordance with the categorical view. Second, a new phonetic study of both tone 3 sandhi and the idiosyncratic tone sandhi found with the morpheme *yi* ("one") provides further support for the claim of categoricality of tone 3 sandhi in Taiwan Mandarin, yet also illustrates the empirical and theoretical pitfalls of phonetic research on tone sandhi.

1. INTRODUCTION

Phonetics has traditionally been treated as a branch of physics, not linguistics (thus the division made in the title of Crystal (1996).

* A much briefer review of the tone 3 sandhi literature is presented in Myers and Tsay (submitted). The experimental results were presented, with less detailed analyses, at the Sixth International Conference on Chinese Linguistics at Leiden University, Netherlands, June 1997. We would like to thank the audience there, including Eric Zee, for comments, Duanmu San for helpful suggestions about materials to use in the experiment, Pam Beddor for being kind enough to allow use of her equipment for a pilot version of this experiment conducted at the University of Michigan, Joyce Huichuan Liu for helping to run the experiments at National Chung Cheng University, and Xu Yi for providing comments on the tone 3 sandhi literature. Two anonymous reviewers also provided very helpful comments. All gaps, mistakes, and misunderstandings are our own responsibility.

Phonetics is often defined as the study of speech articulation and acoustics, which sounds rather irrelevant to cognitive science, while phonology is defined as the study of linguistic sound patterns, apparently closer to the sort of research necessary to truly understand the mind, and not just the mouth and ears. However, as is well-known to anyone who has explored the issue to any depth, the boundary between the two disciplines can become quite fuzzy (a fuzziness celebrated, for example, in the *Laboratory Phonology* series, now up to its seventh volume in Gussenhoven and Warner 2002). Thus given the complex cognitive processes involved in both speech perception and speech production (as well as the perception and production of sign languages), a more accurate view of phonetics is that it is a branch of experimental psycholinguistics. That is, phonetic studies test hypotheses about how phonological knowledge (competence) is actually used (performance), and as such they can be highly relevant to phonological theory and to our understanding of how the mind works.

In this paper we adopt this perspective with respect to the production of tone sandhi in Mandarin. Phrasal sandhi phenomena are inherently interesting to the study of phonological competence and performance, since they represent an almost ideal natural disproof of the "null hypothesis" of Kenstowicz and Kisseberth (1979), namely the hypothesis that phonological knowledge does not actually exist. The typical sort of data collected by phonologists (such as virtually the entire data set of Chomsky and Halle 1968) is by itself incapable of disproving this null hypothesis. Simply observing that there are pairs of words showing alternations, such as *divine-divinity* and *serene-serenity*, is not sufficient to rule out the possibility that English speakers simply memorize these words as wholes and show patterns in their speech solely by virtue of the regular sound changes that occurred historically in the development of the English lexicon. To demonstrate that knowledge of such phonological patterns are psychologically real, one must look at novel forms, not memorized forms. These include borrowings from other languages, invented words used in experimental studies, and novel juxtapositions of morphemes and phonemes produced through speech errors (see Myers 1993 for a survey of the literature on such novel forms and their implications for English phonology). Novel forms also include syntactic phrases and sentences, which are usually entirely original combinations of words. Phrasal sandhi phenomena could not exist if the null hypothesis were entirely correct, and yet they do; therefore, human cognition must contain some sort of phonological processor.

In the case of tone 3 sandhi in Mandarin, traditionally described as the replacement of tone 3 (a low dipping tone in isolation) by tone 2 (a rising tone) before another tone 3 (Chao 1948), there is no doubt that the processor is capable of actively applying the appropriate tone alternations on-line. This is clear not only from speakers' ability to adjust tone in novel phrases (including phone numbers, unfamiliar names, and transliterated foreign terms). It is also well-known that tone 3 sandhi is sensitive to prosodic factors that define the domain within which tone sandhi occurs (Shih 1986, Hsiao 1991). Sensitivity to such factors, which are dependent on the particular combination of words and syntactic structures chosen on-line by speakers during the production of utterances, could not occur unless speakers were actually applying tone sandhi on-line as well. Evidence for this conclusion also comes from unstressed-initial words code-switched from English, which are treated as beginning with a low tone and thus trigger tone sandhi in the preceding tone 3 syllable (Cheng 1968), and from speech errors, since when a tone is erroneously changed into or from a tone 3, the tone 3 sandhi pattern automatically applies or fails to apply as appropriate (Wan and Jaeger 1998).

What is less clear is precisely how the on-line tone sandhi process operates at the phonetic level. That is, once the process has been triggered by the appropriate context (i.e. the appearance of two syllables with lexical tone 3, standing in the required prosodic relationship), there are at least two distinct ways by which the tone change itself might be realized. In the usual way it is described, it involves a substitution of one lexical tone by another, namely a tone 3 is replaced by a tone 2. According to this view, tone 3 sandhi operates basically the same way that *a/an* allomorphy does in English, where speakers flip categorically between *a* (if the following word is vowel-initial) and *an* (if the following word is consonant-initial), a process whose productive, on-line nature is also supported by speech errors (Fromkin 1973). We will call this the "categorical" view.

An alternative view would treat tone 3 sandhi more like flapping in English, where one would not want to say that speakers substitute /t/ and

/d/ with a flap segment in words like *latter* or *ladder*, but instead that the flap sound arises through processes that occur at a later, more phonetic stage in speech production (see e.g. Charles-Luce 1997 for evidence that American English speakers actively adjust the degree of flapping depending on the pragmatic context). Applying this view to Mandarin, tone 3 sandhi would involve adjusting the articulation of a lexical tone 3 so that it sounds more similar to a lexical tone 2 (just as a flap sounds more like a /d/ than a /t/), rather than replacing one set of articulatory instructions with an entirely distinct set. We will call this the "gradient" view.

The goal of this paper is to investigate the phonetic evidence that may help to distinguish between these two competing views of the production of Mandarin tone sandhi. Although we cannot settle the issue entirely in a brief paper like this, we still hope to make two useful contributions. First, we provide an annotated review of the phonetic literature relating to Mandarin tone sandhi. This literature suggests what we believe is a novel and plausible hypothesis, namely that Mandarin tone sandhi is in fact processed in quite distinct ways by different groups of speakers. Specifically, native speakers of Beijing Mandarin seem to apply tone sandhi in accordance with the gradient view, where tone 3 is phonetically modified to sound more like tone 2, without in fact being replaced by it.[1] By contrast, speakers of varieties of Mandarin influenced by other Sinitic languages, in particular Taiwan Mandarin (influenced by Southern Min), seem to process tone 3 sandhi in a categorical fashion.

Second, we describe a new phonetic experiment with speakers of Taiwan Mandarin to test not only the phonetic nature of tone 3 sandhi, but also what we call *yi* sandhi. This is the restricted tone pattern affecting just the morpheme *yi* (一 "one"). This morpheme undergoes its own variety of context-dependent tone alternations (Chao 1948), described more fully later. Since *yi* sandhi is restricted to just one morpheme, it is expected to be processed in the same way as English *a/an* allomorphy, thus seeming to provide a categorical standard against

[1] By "Beijing Mandarin" we mean precisely what the name states: the variety of Mandarin spoken by people who grew up in Beijing. A person who grew up in Shanghai may be just as much a nonnative speaker of Beijing Mandarin as one who grew up in Taipei.

which the behavior of tone 3 sandhi can be compared. As often happens in empirical studies, however, we will see that this view of *yi* sandhi will prove to be somewhat naive, and our results, combined with those for tone 3 sandhi, provide a more complicated (and hence interesting) picture both of tone sandhi and of how it can be investigated phonetically.

The organization of the rest of this paper is as follows. We begin in section 2 with a review of the history of phonetic research into Mandarin tone sandhi. In section 3 we describe our phonetic experiment on Mandarin tone sandhi, with special focus on *yi* sandhi. Finally, in section 4 we discuss the implications of our findings for the processing of tone sandhi, and of the role that phonetic research can play in phonology theory.

2. A HISTORY OF MANDARIN TONE SANDHI PHONETICS

The study of tone 3 sandhi is much longer than one might expect. Although the first careful experimental phonetic studies were not conducted until the second half of the twentieth century, impressionistic descriptions of the phenomenon in Mandarin were given by structural linguists earlier in the century, and by Asian linguists centuries before that. In this section, we review the phonetic work that has been done on Mandarin tone 3 sandhi, organizing it into impressionistic studies, perceptual phonetic studies, and acoustic phonetic studies. The general conclusions will be that one of the common arguments in favor of a categorical view of Mandarin tone 3 sandhi is not valid, and that there seems to be a dialect difference in the phonetic nature of this process: ironically only speakers of "nonstandard" Mandarin seem to behave in accordance with the standardly given categorical rule.

2.1 Impressionistic phonetics

Centuries before Chao (1948) published the now standard rule for Mandarin tone 3 sandhi (i.e. tone 3 becomes tone 2 before tone 3), the diachronic ancestor for this pattern had already been described. Mei (1977) collects descriptions of tone sandhi found in the 16th century notes to a Korean textbook for learning Chinese, and in comments on the proper way for reading poetry written by a late Ming dynasty scholar. As with historical phonology in general, it is difficult to be certain about the phonetic characteristics of the tones being discussed in these works, but it is clear that the tone alternations they describe occur in precisely the locations where tone 3 sandhi occurs in present-day Mandarin, namely when characters today pronounced with tone 3 are adjacent. In both of these historical descriptions, the tone sandhi is given not in phonetic terms, but in terms of lexical tone categories (namely "Rising" 「上」, corresponding to modern tone 3, and "Level" 「平」, corresponding to modern tone 2). In Mei's translation of the relevant passage in the Korean textbook, for example, the rule is given as follows (pp. 238-9): "If both syllables are in the Rising tone, ... pronounce the first syllable like the voiced variety of the Level tone" (original on p. 256: 「但連兩字皆上聲，...則呼上聲如平聲濁音之勢」). Similarly, the early Ming poetry-reading advice states (translation on p. 246): "When one Rising syllable is repeated after another, however, the first syllable sounds like a Level tone" (original on p. 257: 「上上疊用，則第一字便似平聲」). Although both descriptions seem essentially to adopt the categorical view, it is interesting that neither rule is literally given as a substitution, but instead the output of the rule is merely described as producing something that is "like" (「如」、「似」) the other lexical tone category.

Apparently the first researcher to challenge the categorical view of Mandarin tone sandhi presented in these historical works and in Chao (1948) was a non-native speaker, Charles Hockett. Hockett (1947) claimed that tone 3 becomes what he called "tone 5" in some contexts (from his examples, including before tone 3). This "tone 5" (not to be confused with the "toneless" tone that is sometimes given the same name) is claimed to be neutralized with tone 2 when unstressed, but when stressed "[t]he rise for /2/ begins and ends somewhat higher than that for /5/" (p. 219 in Joos 1964). Taken literally, Hockett's claim means that he considered Mandarin tone 3 sandhi to be categorical, but that it resulted in the output of a lexical category other than the standard four. However, from a contemporary perspective it is clear that his argument for the existence of a fifth lexical category is entirely dependent on acceptance of his now-outmoded structuralist phonemic theory, which did not allow for the possibility that morphemes undergoing morphophonemic alternations could have a single underlying representation (such as morphemes appearing with tone 3 in one compound, but "tone 5" in another). If his observations are instead taken to imply merely that tone 3 can be phonetically distinct from a lexical tone 2, then we can see that the phonetic description that he gives is more consistent with the gradient analysis supported by the results of instrumental acoustic phonetic experiments on Beijing Mandarin conducted decades later, as we will show below. Nevertheless, Hockett's credibility as an observer is challenged somewhat by his claim that he also hears the same tone 5 (p. 220) in words that he transcribes as /'ta5 pan4/ "make-up" (presumably 打扮) and /'ma5 fan4/ "annoying" (presumably 麻煩), implying a number of strange things: that tone 3 sandhi also applies before tone 4 (in 打扮), that 「煩」 has lexical tone 4, and that lexical tone 2 can also be pronounced as "tone 5" before tone 4 or tone 2 (麻煩).

Ten years later, Martin (1957) voiced support for Hockett's observations concerning tone 3 sandhi, now explicitly claiming (pp. 215-216) that tone 3 normally becomes tone 2 before tone 3, unless it is stressed, in which case it becomes tone 5 (again adopting the structuralist assumptions that led Hockett to treat tone sandhi as technically categorical while actually gradient). He is also more explicit about giving the values of these tone categories, writing tone 2 in Chao (1930) notation as [35], but tone 5 as [24]. That is, like Hockett he claimed that sandhi tone 3 is slightly lower than lexical tone 2, but only when stressed. The relevance of stress to the phonetic study of tone sandhi is an important issue that we will discuss again below.

2.2 Perceptual phonetic studies

Wang and Li (1967), skeptical of the claims of Hockett (1947) and Martin (1957), conducted the first perceptual study of Mandarin tone 3 sandhi. Two native speakers of Beijing Mandarin produced a set of minimal pairs (disyllabic compounds with tone 3 on the second syllable and either tone 2 or tone 3 on the first) that were then played to sixteen listeners, including the two original speakers plus fourteen instructors of Beijing Mandarin. Accuracy rates in identifying the rising tone as tone 2 or tone 3 ranged between 49.2% to 54.2% for the fourteen instructors, very close to the 50% accuracy rate expected by chance (the two original readers scored slightly better distinguishing the tones produced in their own voices, respectively 56.9% and 67.3%). This study has since become the most-often cited study on the phonetics of Mandarin tone sandhi, and is considered by many to be conclusive proof of the validity of the categorical view (see e.g. Duanmu 2000: 237).

Doubts can be raised about this study, however. Most importantly, no information is given on the acoustic phonetic properties of the actual stimuli presented to the listeners, neither in the published paper nor in the original 63-paged research report (Wang, Li, and Brotzman 1963). It is therefore unknown whether the tones were actually distinct in production. Second, merely failing to display the ability to distinguish the tones in this test does not prove that the listeners lacked the ability entirely (that is, absence of evidence is not the same as evidence for absence, in this case the absence of an ability). Research by sociolinguists (e.g. Labov 1994) has shown that speakers can display an acoustically measurable distinction in their productions of nearly, but not entirely, merged phonemes, without the speakers themselves being able to consciously distinguish the sounds when given isolated minimal pairs. Only when listeners are given the minimal pairs in a discourse context, thus allowing them to process the speech in a pragmatically natural way, may they reveal their latent ability to discriminate between the two phonemes. Another pragmatic factor that may have made it hard for Wang and Li's listeners to distinguish sandhi tone 3 from lexical tone 2 may have lain in the set of listeners themselves. Neither the published paper nor the original report provide any information on the linguistic abilities of the listeners, who are only described as teachers of Beijing Mandarin. It is therefore unclear if they were native speakers of this variety of Mandarin, and it is conceivable that as teachers, presumably trained to teach tone sandhi using the categorical version of the rule, they considered the task inherently impossible and guessed at random.

Perhaps because the findings of Wang and Li (1967) have been considered so decisive, no other perceptual study using Beijing Mandarin speakers and listeners seems to have been conducted. The perceptual studies of Chang and Su (1994) and Peng (2000) both used speakers and listeners of Taiwan Mandarin, and neither revealed any ability to discriminate between sandhi tone 3 and lexical tone 2. However, both of these papers also provide ample information about the acoustic properties of the speech stimuli, and in both cases the differences between the two tone categories were so small as to make the listeners' failure to hear them quite understandable. More information on these acoustic studies is given below.

In short, arguments for the categorical view of Mandarin tone 3 sandhi that rely on perceptual studies are currently much weaker than is generally acknowledged. Clearly more careful work needs to be done here, particularly for Beijing Mandarin.

2.3 Acoustic phonetic studies

Given the discussion in the previous section, questions about the nature of Mandarin tone sandhi production can be more fruitfully addressed by examining the productions themselves, rather than listeners' perceptions of them. As we show in this section, several studies have independently (sometimes inadvertently) provided evidence that tone 3 sandhi is in fact gradient, with sandhi tone 3 differing phonetically from lexical tone 2 in much the way that Hockett (1947) and Martin (1957) described, except that in addition to being slightly lower, it often also differs slightly in slope. This suggests that speakers apply tone sandhi by phonetically modifying tone 3 rather than by replacing it with tone 2. However, this pattern seems to be found most robustly in studies conducted on Beijing Mandarin, suggesting that speakers of other varieties of Mandarin (who are either nonnative speakers or who acquired Mandarin in a sociolinguistic setting where many of the language models are nonnative speakers) behave more in accordance with the standard categorical rule.

2.3.1 Beijing Mandarin speakers

An important (though not the first) acoustic phonetic study using Beijing Mandarin speakers was Zee (1980). Two native speakers of Beijing Mandarin produced minimal pairs (disyllabic compounds with tone 3 on the second syllable and either tone 2 or tone 3 on the first) and fundamental frequency (f_0) was measured at the beginning, end, and at the dip (if one was present). Based on general descriptions of his data, Zee concludes that sandhi tone 3 is not identical to lexical tone 2, and our own reanalysis of his raw data (included in the paper) confirm this: f_0

for sandhi tone 3 was an average of 17.5 Hz lower than that of tone 2, and its pitch contour also rose less steeply.[2] We provide further information on the particulars of Zee's procedures in section 3, since we used his study as a model for our study of tone 3 sandhi and *yi* sandhi in Taiwan Mandarin.

A reasonable objection from skeptics is that artificial differences between sandhi tone 3 and lexical tone 2 only arise when speakers are asked to read minimal pairs, but this objection is met by the very interesting and original study of Kratochvil (1986), who describes a phonetic analysis of spontaneous speech. Specifically, he analyzed a spontaneous monologue about a word list produced by a native speaker of Beijing Mandarin, using a statistical method called discriminant analysis, which allows one to determine the distinctions and similarities among items defined by a number of parameters rather than just one; in this case, items included the first syllables of tone 2 + tone 3 and tone 3 + tone 3 sequences with measurements of their f_0 and amplitude at six points. The discriminant analysis grouped most tokens of sandhi tone 3 together with lexical tone 3 rather than with lexical tone 2, suggesting that the speaker merely adjusted the phonetic form of sandhi tone 3 rather than replacing it by tone 2. The f_0 means given on p. 267 suggest that sandhi tone 3 begins higher than tone 2 and ends around the same or a little lower, a pattern roughly consistent with Zee (1980) in that the slopes were different and sandhi tone 3 was flatter. Kratochvil (1984) provides data that lead to the same general conclusion, though no detailed analysis is given in this paper itself.

Shen (1990a) represents an example of a study that provides evidence for the gradient nature of Mandarin tone 3 sandhi, even without specifically examining this question at all. This study is focused on the phonetics of tonal coarticulation rather than tone sandhi, but in the graphs reporting the f_0 contours for the two native Mandarin speakers (p. 286), it is clear that sandhi tone 3 is on average approximately 6 Hz

[2] Throughout the paper, we use the terms "f_0", "fundamental frequency", and "pitch" interchangeably. The conclusions given here are based on a four-way factorial ANOVA (speaker × tone class × item × measurement point) that found a highly significant effect of tone class ($F(1, 120) = 141.58$, $p < 0.0001$), implying differences in overall pitch level, and a highly significant interaction between tone class and measurement point ($F(1, 120) = 51.883$, $p < 0.0001$), implying differences in slope.

lower than, and rises less steeply than, tone 2, a pattern quite consistent with that of Zee (1980). While Shen (1990a) does not analyze tone sandhi specifically, Shen (1990b) expresses the clear opinion (p. 61) that "tone sandhi is completely a phonetic phenomenon". Shen (1990b) also approvingly cites studies by Lin, Lin, and Sun (1979), Coster and Kratochvil (1984), and Kratochvil (1984) that imply that tone 3 often appears in a form phonetically distinct from tone 2 in tone sandhi contexts.

Finally, in another study of tonal coarticulation, Xu (1993) had four mainland Chinese Mandarin speakers (three native speakers of Beijing Mandarin, plus Xu himself) produce disyllabic forms composed of morphemes pronounced /ma/; none of the forms ending in tone 3 were real words. Minimum and mean f_0 values were measured for the /m/ and /a/ segments separately. Xu found that sandhi tone 3 was lower than tone 2 in the vocalic portion by a highly significant degree (ps < 0.005), but the actual amount was much smaller than what Zee (1980) had found (merely 3.2 Hz), and there was also no significant difference in slope (as measured by the difference between minimum and maximum f_0). It should be noted that the author himself (Xu Yi, personal communication) does not believe that these observations (made in his dissertation, but not in the paper published as Xu 1997) can be taken as reliable evidence for the gradience of Mandarin tone 3 sandhi. Just as with perceptual studies, then, the definitive acoustic phonetic study on tone sandhi in Beijing Mandarin has yet to be done.

It also should be noted that there are phonetic studies of Beijing Mandarin speakers that apparently failed to find evidence for gradience in tone sandhi. These studies include Shen, Chao, and Peterson (1961) and Rumjancev (1972) (both described in Kratochvil 1986), but unfortunately we have not been able to obtain copies to examine them ourselves (Lyovin 1978 is a long review of Rumjancev 1972, but does not mention tone sandhi at all).

2.3.2 Other Mandarin speakers

In contrast to the above studies on Beijing Mandarin, no acoustic phonetic study using only speakers of other varieties of Mandarin seems to have uncovered strong evidence for phonetic differences between sandhi tone 3 and lexical tone 2. The earliest of these is perhaps

Brotzman (1964), the appendix of which gives raw acoustic data from two Mandarin speakers from Shanghai, whose parents were not native Mandarin speakers, and who themselves went to high school outside of mainland China (respectively, Taiwan and the United States). The f_0 data do not seem to show any clear difference between the two tone categories for these speakers.

Chang and Su (1994), whose perceptual study was mentioned above, used speakers of Taiwan Mandarin; one of the two speakers, in fact, was a native speaker of Taiwanese, not Mandarin. Minimal pairs (disyllabic compounds) produced by the speakers showed no significant differences in f_0 at any of the six measurement points. Although their summary data seem to show a tendency for sandhi tone 3 to be slightly lower in f_0 than tone 2 (though by a miniscule 0.4 Hz), our own statistical analyses run on these means could extract no significant effects.

Fon (1997) also used Taiwan Mandarin speakers, finding that sandhi tone 3 had the same average f_0 value, the same f_0 slope, and the same duration as tone 2 in utterance-initial position (in trisyllabic items), but in utterance-medial position it was actually slightly higher and shorter than tone 2 (the slopes were still the same). Most likely, those utterance-medial differences were not related to tone sandhi at all, but were instead an effect of intonation or variations across speakers or utterances in the application of tone sandhi in trisyllabic items (see e.g. Dow 1972).

The only study to find any reliable difference between sandhi tone 3 and lexical tone 2 in Taiwan Mandarin is Peng (2000). Ten speakers from Taiwan but who spoke no Taiwanese produced minimal pairs (disyllabic compounds); f_0 was measured at 10 different points. Sandhi tone 3 was an average of 2.3 Hz lower than tone 2, an even smaller difference than that found by Xu (1993) (and which was reported by Peng as being merely "marginally significant ... $p < 0.05$", p. 155), and there was no difference in slope.

While on the face of it, Peng's results seem to imply that speakers of a variety of Mandarin other than Beijing Mandarin can also process tone sandhi in a gradient fashion, the small size of the f_0 difference and the lack of any difference in slope cast doubt on this conclusion. First, all of the acoustic phonetic studies on Beijing Mandarin summarized above, with the exception of Xu (1993), found much larger differences in f_0

(from 3 to 7 times larger) and also differences in slope. The size is relevant, since there are limits to the size of f_0 differences that can be readily discriminated in the complex environment presented by speech, and it is not clear that a 2.3 Hz difference is learnable (certainly the listeners in Peng's own perceptual study, as mentioned earlier, failed to demonstrate an ability to discriminate between the two tone classes). The lack of slope differences is also an important clue that the process here is not the same as for Beijing Mandarin, since it is difficult to imagine how a mechanism involving gradient adjustments of a dynamic tone contour could manage to change the contour from some underlying tone 3 form (whether the dipping tone of its citation form or the low dropping tone as which it appears in most contexts) into an almost perfect replica of a genuine tone 2. The slope differences observed for Beijing Mandarin are more what one would expect from gradient phonetic adjustments. Finally, it is striking that the very small difference in f_0, coupled with a lack of slope differences, are precisely what was found by Myers and Tsay (submitted) for Taiwan Southern Min, a language whose tone sandhi process has been argued (Tsay and Myers 1996, Tsay 2002) must be categorical rather than gradient. Myers and Tsay (submitted) in fact present arguments that the small f_0 difference that they found could not be due to gradient processing of Southern Min tone sandhi, since the lack of a slope difference was not what would be predicted if the underlying tones were being preserved, and since unlike genuine gradient processes like English flapping (Charles-Luce 1997), the small f_0 difference was entirely unaffected by pragmatic context. Hence they conclude that the miniscule difference was an artifact of the task, which involved reading aloud minimal pairs (orthography has been shown to give rise to tiny phonetic differences in productions even in the absence of genuine differences in underlying phonology; see e.g. Warner, Jongman, Sereno, and Kemps, in press). These arguments apply equally well to the acoustic results of Peng (2000), though not to the studies conducted on Beijing Mandarin, which showed much larger differences, including slope differences, and which include studies of spontaneous speech.

2.4 Summary

The general point seems to be that for Beijing Mandarin speakers, at least, tone 3 sandhi need not be categorical, but instead may involve phonetic adjustments of the pitch contour that make sandhi tone 3 sound roughly similar to lexical tone 2, though acoustically it is generally lower and may also be flatter in slope. By contrast, speakers who acquire Mandarin as non-natives, or who acquire it from non-natives (directly or indirectly), tend to process tone sandhi in a more categorical fashion.

To our mind the simplest explanation for this pattern is that varieties of Mandarin other than Beijing Mandarin have absorbed the categoricality of the tone sandhi in the contact languages, such as Southern Min. More precisely, when monolingual Southern Min speakers learn Mandarin, the inherently gradient Mandarin tone sandhi pattern is reinterpreted as being as categorical as that of Southern Min (a reinterpretation no doubt encouraged by the use of the categorical rule explicitly taught in Mandarin classes). Children who acquire Taiwan Mandarin from such speakers would then also treat tone sandhi as categorical. Assuming the phonetic studies are in fact reliable, this explanation seems more likely than other alternatives. For example, building on the observations of Hockett (1947) and Martin (1957) that sandhi tone 3 and lexical tone 2 only fail to neutralize in stressed syllables, one might instead suppose that the lack of neutralization is an artifact of overstressing. If this were the case, however, we would expect a greater, not lesser, difference between these two tone categories in Taiwan Mandarin than in Beijing Mandarin, since Taiwan Mandarin pronunciation is notorious for lacking the destressing processes typical of Beijing Mandarin (hence 東西 *dong1xi1* is pronounced the same whether one means "thing" or "east and west").

We admit, however, that the evidence is far from conclusive, and that valid objections can be made about studies that seem to show gradient processing of Mandarin tone sandhi. Not only is there a worry about reading pronunciations, but there is also a certain amount of variability in the application of tone sandhi rules in some contexts (Dow 1972) which, when averaged for statistical analysis, may give the misleading impression that speakers are producing gradiently distinct pronunciations. However, the experiments showing acoustic phonetic differences between sandhi tone 3 and lexical tone 2 in Beijing Mandarin primarily used disyllabic strings, in which tone sandhi applies very regularly. Moreover, if reading pronunciations are solely responsible for the differences between sandhi tone 3 and lexical tone 2, it is rather a mystery why these differences are found primarily with native Beijing Mandarin speakers. If anything, one would expect reading pronunciations to be especially prominent with speakers of "nonstandard" varieties of Mandarin. Nevertheless, our goal here is not to win an argument, but rather to inspire further research on the phonetic processing of Mandarin tone sandhi, an open issue that some Chinese phonologists seem to have prematurely considered settled.

3. TONE 3 SANDHI AND *YI* SANDHI

While the phonetics of tone 3 sandhi has been extensively (though inclusively) studied, this is not the case for what we call *yi* sandhi, namely the modification of the tone of the morpheme meaning "one" 「一」 before the other lexical tones of Mandarin. The basic alternations are illustrated below in (1), with pronunciations given for convenience in Hanyu Pinyin (with digits representing the tones rather than diacritics) since the focus here is on tone categories rather than segmental phonetics.

(1)

In isolation	tone 1	一、二、三 yi1, er2, san1 ("one, two, three")
Before tone 4	tone 2	一件 yi2jian4 ("one CL")[3]
Elsewhere	tone 4	一隻;一門;一碗 yi4zhi1, yi4men2, yi4wan3 ("one CL", "one door", "one bowl")

No other morpheme follows precisely this pattern (Chao 1948). The negation morpheme *bu* (不) shows the same tone changes in context settings, but in isolation it is pronounced with falling tone 4; for many speakers, the numbers *qi* (七 "seven") and *ba* (八 "eight") also change to rising tone 2 before falling tone 4, but do not change to falling tone 4 in other contexts. Certainly the *yi* sandhi pattern is not found with other morphemes that share the same lexical tone and segmental content (e.g. 衣 "clothing"，醫 "doctor").

[3] CL stands for classifier. Semantic differences between the classifiers are not relevant.

The lexical idiosyncrasy of *yi* sandhi makes it important in the comparison of the categorical and gradient views towards Mandarin tone sandhi. This is because it appears that there are tight correlations between the phonetic categoricality or gradience of a phonological process and its status in the grammar. These correlations are particularly relevant to any phonological theory that makes a distinction between lexical phonology (morphophonology) and postlexical phonology ("pure phonology" or "phonetics", i.e. language-specific articulatory processes). The most famous of such theories is of course the theory of Lexical Phonology (e.g. Kiparsky 1982, Hargus and Kaisse 1993), but similar distinctions are made in psycholinguistic models as well, for example Levelt (1992).

One problem that seems to be posed by *yi* sandhi is that, like tone 3 sandhi, it applies across word boundaries (to the extent that word boundaries are clear in Mandarin), and thus it should be a postlexical rule. Yet postlexical rules should not be sensitive to lexical information the way that lexically idiosyncratic *yi* sandhi clearly is. However, it turns out that *yi* sandhi is hardly unique in this regard. There is in fact an entire class of phrase-level rules which display lexical characteristics (Kaisse 1985, Hayes 1990). English *a/an* allomorphy is another example, as is Southern Min tone sandhi, which displays exclusively lexical characteristics except for the fact that it applies at the phrasal level (Tsay and Myers 1996, Tsay 2002). In particular, Southern Min tone sandhi is structure-preserving, a typical characteristic of lexical patterns in which the outputs are lexically contrastive phonemes or tonemes (see e.g. Kiparsky 1982, Borowsky 1993). Thus Southern Min tone sandhi results in surface tones that are acoustically identical (except for the minor differences mentioned above) to lexically contrastive tones (Tsay, Charles-Luce and Guo 1999, Tsay and Myers 2001, Myers and Tsay 2001, Myers and Tsay, submitted).

In a sense it is not that devastating to find that phrasal rules can show lexical characteristics; there are reasonable modifications of Lexical Phonology that can deal with this (e.g. Kaisse 1985, Hayes 1990). Tsay and Myers (1996) and Tsay (2002) argue, for example, that the application of Taiwanese tone sandhi involves the postlexical selection of lexically stored allomorphs, which means that Taiwanese tone sandhi can be both phrasal (the selection is postlexical) but otherwise show exclusively lexical characteristics (since the allomorphs are stored in the lexicon).

Although the property of being phrasal does not reliably correlate with being a "true" postlexical rule, two other properties do seem to cooccur (most of the time): phonetic categoricality and lexical sensitivity. This correlation is in a sense "deeper" than others used to distinguish lexical and postlexical phonology, since lexical rules are thought of as describing distributional patterns of categorical elements (e.g. phonemes, tonemes, or at least combinations of categorical features) within the lexicon rather the change of lexical representations into phonetically gradient patterns. Thus a serious theoretical problem would arise if we found a tone sandhi pattern that is sensitive to lexical information and yet was phonetically gradient. The only way we could handle this would be to suppose either that lexical representations may contain gradient information, or that postlexical, "phoneticky" rules may be sensitive to lexical information.

With this as background, it is now clear why Mandarin *yi* sandhi plays an important theoretical role. This form of tone sandhi shows lexical sensitivity, like Southern Min tone sandhi, and yet it is unknown whether the pattern is also structure preserving, as Lexical Phonology would predict. That is, is it in fact true that the tone 2 that appears before tone 4 is phonetically identical to a lexical tone 2 in the same context? Is it the case that the tone 4 that appears before all other tones is identical to the lexical tone 4? If not, then *yi* sandhi may pose a challenge to the theory of Lexical Phonology, and more generally would then provide further illumination into the nature of on-line phonological processing.

This section of the paper describes an acoustic phonetic study of Mandarin *yi* sandhi that aims to address these questions.

3.1 Methods

The methodology of the study was based as much as possible on previous work on the phonetics of Chinese tone sandhi, in particular that of Zee (1980). Essentially, the design was as follows. Materials were chosen to provide minimal pairs between disyllabic compounds beginning with the morpheme *yi* (「一」 "one") or with another morpheme with the same segmental content. The question was thus whether the sandhi form of *yi* (either with tone 2 or with tone 4, depending on the context) was acoustically identical to the other morpheme with an underlying tone 2 or tone 4.

Thus two aspects of the first syllable of these compounds were examined. The first, of course, was the pitch; thus f_0 was measured. Because of possible differences in slope of the pitch contour, rather than an overall change in pitch, f_0 was measured at three points.

The other important aspect that was measured was the duration of the first syllable. This was done in order to determine the extent of the influence of stress, which is most reliably marked by duration in Mandarin (Shen 1993). Unlike Mandarin tone 3 sandhi, we might expect that the morpheme involved in *yi* sandhi may show a different stress pattern from otherwise homophonous morphemes because it is a closed-class item. As we will see, the possible involvement of stress poses interesting challenges to the phonetic study of *yi* sandhi.

3.1.1 Participants

The participants in the study were ten native speakers of Mandarin, all students at National Chung Cheng University in Chiayi county in Taiwan. All also spoke Southern Min, but they all acquired Mandarin as a first language. Five participants were female and five were male. All speakers were naive to the purposes of the study and had had no linguistic or phonetic training.

3.1.2 Materials

The materials were designed to test two kinds of tone sandhi in Mandarin, both tone 3 sandhi and *yi* sandhi. The minimal pairs chosen were the following.

(2) a. Tone 3 sandhi:

土改 tu3 gai3 ("land reform")	塗改 tu2 gai3 ("alter (text)")

b. *yi* sandhi:

一世 yi1 shi4 ("all one's life")	儀式 yi2 shi4 ("ceremony")
一線 yi1 xian4 ("one string")	胰腺 yi2 xian4 ("pancreas")
一己 yi1 ji3 ("oneself")	異己 yi4 ji3 ("opponent")

一隻狗 yi1 zhi1 gou3 ("one-CL dog")	億隻狗 yi4 zhi1 gou3 ("100 million-CL dogs")

The pair chosen to represent tone 3 sandhi had previously been used by Zee (1980) and Peng (2000) and seemed to contain prosodically quite comparable items (N-N compound vs. V-N compound). The pairs chosen to study *yi* sandhi seem to be the best available in the Mandarin lexicon, though there is the obvious problem of a difference in lexical frequency: both the morphemes and the words listed in the column on the right seem to be less common than those in the column on the left. All attempts were made to ensure that the compounds were of comparable prosodic structure as well, so as to minimize differences in stress, but this seems not to have been successful.

These five pairs were presented in written form along with six other pairs that were not analyzed. These filler pairs are listed below.

(3)

a	買馬 mai3 ma3 ("buy a horse")	埋馬 mai2 ma3 ("bury a horse")
b	倚老賣老 yi3 lao3 mai4 lao3 ("exploit one's seniority")	遺老遺少 yi2 lao3 yi2 shao4/3 ("old and young diehards")[4]
c	一名 yi1 ming2 ("one person")	異名 yi4 ming2 ("different name")
d	一門課 yi1 men2 ke4 ("one-CL class")	億門課 yi4 men2 ke4 ("100 million-CL classes")
e	一碗湯 yi1 wan3 tang1 ("one bowl of soup")	億碗湯 yi4 wan3 tang1 ("100 million-CL bowls of soup")
f	一件事 yi1 jian4 shi4 ("one-CL affair")	億件事 yi4 jian4 shi4 ("100 million-CL affairs")

The pairs in (3a)-(3e) were not included in the analysis because vowel duration was not consistently measurable in the phonetic environment (i.e. before a sonorant). The pair in (3f) was of course not used since the tones on the first syllables are not predicted to be the same.

All of the items were presented in the same sentence frame used by

4 The English translations for these idioms come from DeFrancis (1996).

Zee (1980), namely 我現在讀「_」給你聽 *wo3 xian4zai4 du2 "__" gei3 ni3 ting1* ("I now read __ for you to hear").

3.1.3 Procedure

The eleven pairs (five experimental items and six filler items) were presented to the speakers in written form (the first two participants, one male and one female, read the items off of sheets of paper, while the remainder read the items off a computer screen, set up to display only one item as a time).

In the experiment reported in Zee (1980), items were separated into two lists, so that one list consisted entirely of tone 3 + tone 3 compounds, while the other consisted entirely of tone 2 + tone 3 compounds. In this way no minimal pair was ever directly compared in the reading. However, it was felt that this method made it possible for random factors to give rise to significant differences in tone production for the two types of compounds.

Thus the order of list presentation in our study was varied in a number of ways. First, as in Zee's study, minimal pairs were separated into two lists, labeled A and B. However, the lists did not consistent entirely of items with a particular tone pattern. Rather, an attempt was made to balance the lists for the number of items of each type. The experimental items found in lists A and B are shown below.

(4)	List A	List B
	塗改	土改
	儀式	一世
	一線	胰腺
	異己	一己
	億隻狗	一隻狗

Each list was then randomized four times, creating Block A and Block B, each of 44 items (11 items × 4 repetitions). Participants were then evenly assigned to read either in the order Block A and then Block B, or Block B and then Block A (three males read in order A-B and two in B-A; three females read in order B-A and two in A-B). Participants were presented with five sentences to read for practice and to help set the recording level. Then they would read either Block A or Block B. After a short break, they would then read the other block. The first repetition (i.e. the first eleven items in a block) were not analyzed, as it was assumed that the participant was still becoming accustomed to the items.

The remaining three repetitions for each experimental item were entered into the digital waveform analyzer of Computerized Speech Lab 4300B (Kay Elemetrics) at a sampling rate of 10 kHz per second. Duration was measured from the beginning to the end of periodicity of the vowel portion of the target syllables. Fundamental frequency was calculated for 20 msec frames using the CSL pitch-tracking algorithm. 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. The values for the three repetitions were then averaged, resulting in a single score for each participant for each item for each of four measures (duration, beginning f_0, mid f_0, and end f_0).

3.2 Results

Separate analyses were conducted for the tone 3 sandhi pair (tu2gai3 vs. tu3gai3), for the *yi* sandhi pairs involving the possible neutralization between tone 1 and tone 2 (yi1shi4 and yi1xian4 vs. yi2shi4 and yi2xian4), and for the *yi* sandhi pairs involving the possible neutralization between tone 1 and tone 4 (yi1ji3 and yi1zhi1 vs. yi4ji3 and yi4zhi1). In each of these three comparisons, duration was analyzed separately from f_0 using analysis of variance (ANOVA), a standard statistical method that allows us to examine not only the separate contributions of various factors that may influence duration or f_0, but also the way that these factors might interact. This is particularly useful in the case of f_0, since tones may differ not only in overall pitch but also in slope, which can be analyzed as an interaction between tone class (e.g. tu2 vs. tu3) and measurement point (i.e. beginning, middle, or end of the pitch contour). A significant interaction between these two factors shows that the different tone classes do not show the same relationship among the three measurement points, which means that the slopes of their pitch contours are not the same. All of the analyses were conducted by participant (the number of items was too small to do by-item analyses). Males and females were analyzed together, in spite of the obvious differences in their f_0 ranges, because the nature of the statistical tests (repeated measures) made this difference irrelevant (what are crucial are the differences across items produced by the same speaker). The data that we used to conduct our analyses are given in the appendix.

The results for the tone 3 sandhi pair are given below, with average durations shown in Figure 1 and average f_0 contours in Figure 2.

Figure 1. Durations of the first syllables in 塗改 vs. 土改

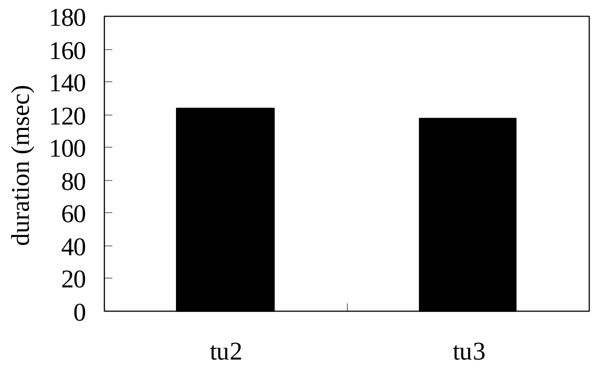
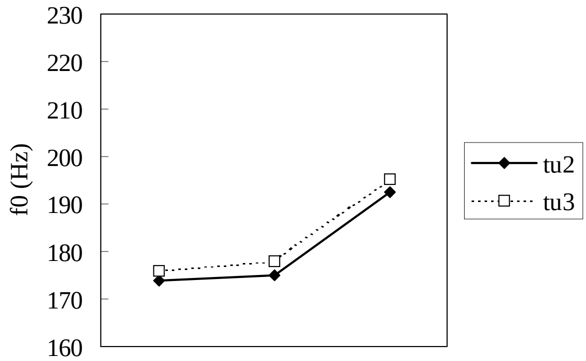


Figure 2. f_0 contours of the first syllables in in 塗改 vs. 土改



A one-way ANOVA comparing the durations (in milliseconds) of the first syllables in the tone 3 sandhi pair (i.e. tu2 compared with tu3) found no significant difference (tu2 mean duration = 123 msec, tu3 mean duration = 118 msec; $F(1, 9) = 1.53$, $MSe = 117.7$, $p > 0.24$). For f_0, a two-way ANOVA (tone class × measurement point) found a significant main effect of measurement point by itself ($F(2, 18) = 9.93$, $MSe = 221.7$, $p = 0.0012$), which merely expresses the unexciting finding that the slopes for the two tones were not flat, but instead rose slightly, as can be seen in the graph. However, there was no main effect of tone class ($F < 1.3$, $p > 0.29$; i.e. the slight difference between the two lines in the graph was not large enough to stand out significantly from the variation across speakers) and no interaction ($F < 0.08$, $p > 0.92$; i.e. the slopes of the two tone contours were not significantly different from chance). In other words, none of the slight differences seen in Figures 1 and 2 mean anything, and we must say that we failed to demonstrate that tone 3 sandhi is not categorical.

We now turn to the duration results for the *yi* sandhi pairs. The average durations for the first syllables in the yi1（一）vs. yi2 comparisons are shown in Figure 3, and those for the first syllables in the yi1（一）vs. yi4 comparisons are shown in Figure 4.

Figure 3. Durations of the first syllables in 一世 & 一線 vs. 儀式 & 胰腺

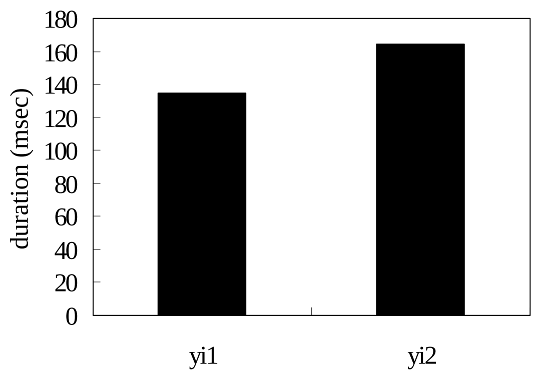
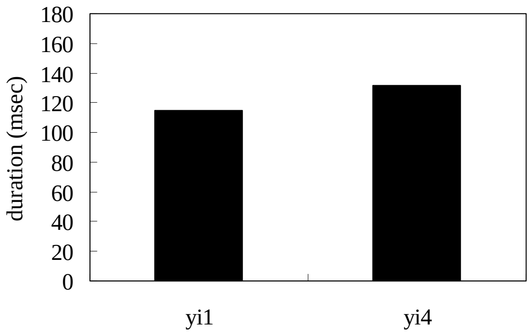


Figure 4. Durations of the first syllables in 一己 & 一隻狗 vs. 異己 & 億隻狗



For each duration comparison, we conducted a two-way ANOVA (tone class × word pair). This allowed us to take advantage of the greater amount of data provided by our use of two word pairs for each comparison, without the loss of information that would occur if we averaged across the word pairs. The ANOVA for yi1 vs. yi2 found a nearly significant effect of word pair ($F(1, 9) = 4.44$, $MSe = 211.9$, $p = 0.064$), showing that syllables in the two word pairs differed somewhat in duration. More importantly, there was a highly significant effect of tone class ($F(1, 9) = 30.42$, $MSe = 298.5$, $p = 0.0004$), with yi1 syllables being reliably shorter (mean duration = 135 msec) than yi2 syllables (mean duration = 165 msec). Moreover, this pattern was highly consistent across the two word pairs, since there was no significant interaction between tone class and word pair ($F < 0.6$, $p > 0.47$). The ANOVA for the yi1 vs. yi4 comparison showed a similar duration pattern: a significant effect of word pair ($F(1, 9) = 8.79$, $MSe = 698.1$, $p = 0.016$), a marginally significant effect of tone class ($F(1, 9) = 5.08$, $MSe = 555.4$, $p = 0.051$), again with yi1 syllables being shorter (mean duration = 115 msec) than yi4 syllables (mean duration 132), but no interaction ($F < 0.7$, $p > 0.44$), which implies that the pattern was consistent across word pairs.

Finally, we present the results for the f_0 contours for the *yi* sandhi pairs. The average f_0 contours for the first syllables in the yi1 vs. yi2 comparisons are shown in Figure 5, and those for the first syllables in the yi1 vs. yi4 comparisons are shown in Figure 6.

Figure 5. f_0 contours of the first syllables in 一世 & 一線 vs. 儀式 & 胰腺

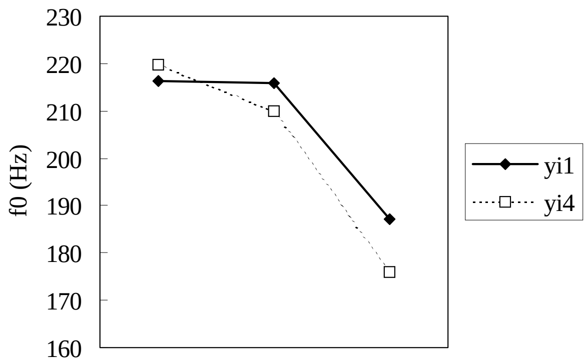
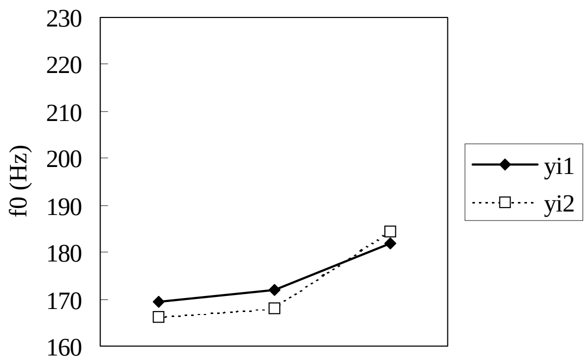


Figure 6. f_0 contours of the first syllables in 一己 & 一隻狗 vs. 異己 & 億隻狗



We conducted separate analyzes for the two tone comparisons, this time using three-way ANOVAs (tone class × word pair × measurement point). For both analyses, the most significant effect was a main effect of measurement point (yi1 vs. yi2: $F(2, 18) = 45.92$, $MSe = 60.0$, $p < 0.0001$; yi1 vs. yi4: $F(2, 18) = 44.68$, $MSe = 348.2$, $p < 0.0001$), but again this merely describes the uninteresting observation that the slopes of the tone contours were not flat. In another not particularly important result, no significant main effect of word pair was found for either analysis (Fs < 2.1, ps > 0.18). Moreover, only in the yi1 vs. yi4 comparison was there a significant interaction between word pair and measurement point (yi1 vs. yi2: $F < 0.38$, $p > 0.69$; yi1 vs. yi4: $F(2, 18) = 4.69$, $MSe = 66.4$, $p = 0.023$), which suggests that the word pairs chosen for the yi1 vs. yi4 comparison differed somewhat in the slopes of their tone contours, while those for the yi1 vs. yi2 did not.

Far more important is that neither comparison found significant main effects of tone class (Fs < 2.7, ps > 0.13), implying that the overall differences in pitch height seen in Figures 5 and 6 may be treated as random. Yet both comparisons found highly significant interactions between tone class and measurement point (yi1 vs. yi2: $F(2, 18) = 7.10$, $MSe = 16.0$, $p = 0.0053$; yi1 vs. yi4: $F(2, 18) = 6.35$, $MSe = 86.4$, $p = 0.0082$), showing that the slopes were more different than what would be expected to happen purely by chance. As can be seen in Figures 5 and 6, these slope differences essentially involved *yi* (i.e. 「一」) being somewhat flatter than the tone it was compared with. More precisely, the tone contour for *yi* doesn't rise as steeply at the end as does yi2 in the comparison shown in Figure 5, while the tone contour for *yi* doesn't fall as steeply at the beginning as does yi4 in the comparison shown in Figure 6. Since lexical tone 1 is of course more flat than either tone 2 or tone 4, these significant differences in slope seem to imply that *yi* sandhi does not in fact involve complete phonetic neutralization, contrary to the theoretical expectations. This surprising result will be discussed more fully below.

3.3 Discussion

The most important results were, first, that no difference was found in the productions of sandhi tone 3 and lexical tone 2, and second, that significant differences were found in the productions of *yi* sandhi forms from their non-sandhi lexical matches, both in duration and pitch contour, though not in overall pitch height. The first result is consistent with previous phonetic studies of tone 3 sandhi in Taiwan Mandarin, but the second result is rather surprising, seeming to conflict not only with our hypothesized explanation for the categorical processing of tone 3 sandhi in Taiwan Mandarin but also with theories like Lexical Phonology that posit a strict correlation between phonetic categoricality and lexical idiosyncrasy.

Addressing our tone 3 sandhi results first, we must consider the possibility that we simply missed a difference that is in fact present in Taiwan Mandarin. There are at least two differences between the method of presentation in this study and in Zee (1980). First, as noted earlier, the order of presentation was split across speakers; some speakers received "tu2 gai3" in the first list while others received "tu3 gai3" first. By contrast, in Zee's study, speakers apparently always received "tu3 gai3" first. However, it is not at all clear why this order difference would have an effect on the production of tone sandhi.

A perhaps more important difference between our study and Zee's is the context in which the tone 3 sandhi pairs appeared. In his study, speakers were given five minimal pairs to test the effect of tone 3 sandhi plus additional items "to avoid monotony" (Zee 1980:100). Thus minimal pairs involving tone 3 sandhi formed the bulk of the items. By comparison, in our study only three items involved tone 3 sandhi; the remaining eight items all involved syllables with the segmental content of *yi*. Similarities between the tone 3 minimal pair items were thus perhaps less obvious to our speakers, and so they were less likely to emphasize their differences. However, given that no other study has found strong evidence that tone 3 sandhi in Taiwan Mandarin is gradient, even when minimal pairs formed the bulk of the items, we think it is safe to say that our study helps confirm that tone sandhi is processed categorically in this Mandarin variety.

Quite the opposite conclusion seems to be presented by the *yi* sandhi results. Not only did the pitch contours for *yi* sandhi forms have different slopes from those with lexical tone 2 or lexical tone 4, but in both cases the *yi* sandhi slope was flatter, somewhat more similar to its canonical tone 1 form. Yet for a number of reasons it does not seem to make sense that speakers would be phonetically modifying the contour of lexical tone 1 in the application of *yi* sandhi. The clearest reason comes from the theoretical considerations that we have been emphasizing: as a lexically idiosyncratic process, *yi* sandhi is expected to involve categorical, not gradient processing. But there are other reasons to be suspicious as well, since even if we allow that theories like Lexical Phonology are wrong, and speakers can in fact apply lexically idiosyncratic processes in a gradient fashion, we would have to explain why precisely the same speakers in precisely the same experiment nevertheless treated tone 3 sandhi in a strictly categorical fashion.

The crucial clue to what may really be going on comes from the differences in duration. In both *yi* sandhi comparisons, the morpheme meaning "one" 「一」 was produced with a shorter duration than the morpheme that we compared with it. This could not have been due to inherent duration differences across tone categories; in particular, lexical tone 4 is generally shorter than lexical tone 1 (see e.g. Tseng 1990), but we instead found that a tone-4-like sandhi yi1 was shorter than a genuine tone 4. It seems, then, that 「一」 was pronounced with less stress than the morphemes it was compared with (see Shen 1993 for the importance of duration to the phonetics of stress in Mandarin). As noted earlier, a difference in stress makes sense here, since all of these comparison morphemes are of higher frequency than 「一」, and at least some of them are clearly content rather than function morphemes. Function morphemes are well known to have prosodic behavior different from content morphemes (e.g. compare the English noun *can* with the auxiliary *can*, where only the latter can reduce); recent discussions can be found in Selkirk (1996) and Hung & Peters (1997), among many other places.

It is necessary to invoke stress, and not duration alone, since shortening the duration of a syllable doesn't necessarily result in a changing of pitch contour. For example, in the acoustic phonetic experiments on Southern Min tone sandhi reported in Myers and Tsay (submitted), differences in duration due to prosody were also observed (in the contrasts involving context vs. juncture position), yet there were never any effects on f_0 slope. However, it seems that the slope differences we observe here in yi sandhi are very similar to those expected from stress differences. Specifically, the usual phonetic effect of stress on Mandarin tone is to widen the tone range, so that when stressed, "rising tones rise higher and falling tones drift lower (Shen 1990b: 60). With reduced stress, then, contour tones should tend to be flatter, just as we observed with 「一」. Hence it seems safer to conclude that *yi* sandhi in Taiwan Mandarin is indeed categorical, as predicted, with postlexical adjustments in tone contour due to stress.

Yet evidence from other languages hints that even if stress-based reduction is responsible for the phonetic differences we observed, the processing of phonology may be somewhat more complex than theories like that of Lexical Phonology seem to imply. For example, in English, vowel reduction in unstressed syllables has been shown to be sensitive to lexical frequency (i.e. how often a word is used). Thus the vowel in the first unstressed syllable of the higher-frequency word *astronomy* is reduced more than the matching vowel in the lower-frequency word *gastronomy* (Fidelholz 1975). In spite of the sensitivity of vowel reduction to a lexical property (frequency), it is nevertheless truly phonetically gradient. Hence the degree of reduction of the marked vowels in *ev**e**ry*, *mem**o**ry*, and *mamm**a**ry* is positively correlated with these words' relative frequencies (Hooper 1976). This suggests that the role of tone reduction in the phonetic study of *yi* sandhi should not be considered an annoying interference, but a possibly important element in a complete understanding of how tone sandhi actually operates as a cognitive process.

4. CONCLUSIONS

In this paper we have discussed the sort of phonetic evidence that must be collected if we are to understand what is happening when Mandarin speakers apply tone sandhi. We argued that the most crucial sort of evidence must come from the careful instrumental study of production, not impressionistic descriptions or even experimental perception studies. The evidence presented in this paper (summaries of previous work on tone 3 sandhi and new data concerning *yi* sandhi) suggest that Mandarin tone sandhi may be processed in a phonetically gradient fashion, somewhat like flapping in English, although there are differences across varieties of Mandarin. Even the lexically idiosyncratic *yi* sandhi may possibly have gradient elements in its processing, though most likely this is primarily under the influence of prosody, and does not necessarily pose serious challenges to standard views of lexical representation.

Clearly further work needs to be done. Given the controversies that have swirled around the phonetic nature of Mandarin tone 3 sandhi (in the modern era alone, for some sixty years), it seems that it is high time for phoneticians to conduct the conclusive experiments that will settle the issue once and for all. One problem is that most previous studies have been conducted in relative ignorance of the others or merely as a small part of a larger research project. With a broader historical view and attention focused solely on the phonetics of tone 3 sandhi itself, it becomes clear that there are certain factors that future researchers should take into careful consideration. Above all, while minimal pairs are crucial for conducting clean statistical analyses, they likely have some effect on speakers' self-awareness. At a minimum, therefore, the target minimal pairs should be hidden among a large number of fillers, particularly fillers that contain both phonologically identical pairs and near-minimal pairs (e.g. differing only in tone, but not in ways that are directly relevant to tone sandhi). Moreover, to test our hypothesis that there is truly a dialect difference in the processing of tone 3 sandhi, a single study, using consistent procedures, materials, and analytical methods, should be conducted jointly across the Taiwan Strait, with speakers of both Beijing Mandarin and Taiwan Mandarin. We have as yet been unsuccessful in organizing a joint project like this ourselves; anyone interested in attempting one may contact us or carry the banner on without us, as they prefer. It doesn't matter to us who does it, as long as it's done.

Second, because of the concern over the influence of reading pronunciations, attempts should be made to develop more "natural" phonetic methods. We feel that precisely the wrong response to this problem would be to return to an emphasis on traditional impressionistic phonetics (as has been argued by some, e.g. Manaster Ramer 1996). Rather, the objective power of instrumental analyses can continue to be taken advantage of if they are employed on spontaneous speech, or if further experiments are performed on precisely how speech production is affected by reading pronunciations or other artificial laboratory conditions. Coster and Kratochvil (1984), Kratochvil (1984), and Kratochvil (1986) represent examples of the first approach, and it is likely that sophisticated statistical analyses of large speech corpora (or data gathered through the elicitation methods developed by quantitative sociolinguists; e.g. Labov 1994) can go even further. Moreover, if efforts are made to understand the processes involved in reading pronunciations and other laboratory artifacts, this would go far towards strengthening the validity of all laboratory phonetics studies, not just ones focusing on incomplete neutralization. For example, if one could develop a quantitative model that predicted exactly how much a reading context affects articulation, one might then be able to adjust for this factor in phonetic studies that rely on reading. A crude example of this sort of approach is given in Myers and Tsay (submitted), where analyzes are conducted to examine how speakers' varying degrees of neutralization of tone categories correlate with independent listeners' judgments of the fluency of these speakers (see also Warner et al., in press).

Third, researchers interested in Mandarin tone 3 sandhi might consider picking up where we left off in our study of *yi* sandhi. While our results were unfortunately rather ambiguous (implying either the exciting finding that lexical processes can be gradient, or the uninteresting finding that unstressed tones tend to flatten out), it seems important for those interested in the phonetics of tone 3 sandhi to have some sort of baseline process to compare it with. After all, testing the standard claim that tone 3 sandhi is categorical faces a logical difficulty: sandhi tone 3 and lexical tone 2 are predicted *not* to be significantly different from chance. The only way to be sure that a null result supports this claim, rather than merely showing the failure of the task to measure anything, is to conduct other studies with the same speakers, the same language, the same procedure, but a different process for which a different pattern is expected.

Finally, the most general lesson we wish to leave is for theoretical phonologists who normally don't pay much attention to phonetic research: phonology cannot be studied in a vacuum. Phonology is inherently an interface system, half mental, half physical. As with all of psychology, we cannot (yet) peer directly into the brain to see how phonology works, so the best evidence we have for what's going on in there comes from physical behavior. Of course this does not eliminate the necessity for theoretical modeling; after all, phoneticians are (or should be) cognitive scientists, not behaviorists. Nevertheless, we feel that phonological theory has matured to the point where future progress can best be made by giving a more prominent role to the psycholinguistics of speech, which requires not only improvements in modeling but also in the collection of empirical data.

REFERENCES

Borowsky, T. 1993. On the word level. In Hargus and Kaisse (eds.), pp. 199-234.
Brotzman, R. L. 1964. *Progress Report on Mandarin Tone Study. Project on Linguistic Analysis, Ohio State University Research Foundation* 8: 1-35.
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 Maitre Phonetique* 45: 24-27.
Chao, Y. R. 1948. *Mandarin Primer*. Cambridge: Harvard University Press.
Charles-Luce, J. 1997. Cognitive factors involved in preserving a phonemic contrast. *Language and Speech* 40: 229-248.
Cheng, C. C. 1968. English stresses and Chinese tones in Chinese sentences. *Phonetica* 18: 77-88.
Chomsky, N. and Halle, M. 1968. *The Sound Pattern of English*. New York: Harper and Row. Reprinted 1990 by Cambridge, MA: MIT Press.
Coster, D. C., and Kratochvil, P. 1984. Tone and stress discrimination in normal Beijing dialect speech. *New Papers on Chinese Language Use*, 119-132.
Crystal, D. 1996. *Dictionary of Linguistics and Phonetics* (4th edition). Oxford: Blackwell Publishers.
DeFrancis, J. (ed.) 1996. *ABC Chinese-English Dictionary*. Honolulu: University of Hawai'i Press.
Dow, Francis D. M. 1972. A discussion on tone sandhi problems in Chinese. *Journal of the International Phonetic Association* 2 (1):13-19.
Duanmu, S. 2000. *The Phonology of Standard Chinese*. Oxford: Oxford University Press.
Fidelholz, J. L. 1975. Word frequency and vowel reduction in English. *Proceedings of CLS 11*, 200-213. University of Chicago.
Fon, J. 1997. *What are tones really like? An acoustic-based study of Taiwan Mandarin tones.* MA thesis, National Taiwan University.
Fromkin, V. A. 1973. *Speech Errors as Linguistic Evidence*. The Hague: Mouton.
Gussenhoven, C., and Warner, N. (eds.) 2002. *Laboratory Phonology 7*. Berlin: Mouton de Gruyter.
Hargus, S. and Kaisse, E. M. 1993. *Phonetics and Phonology Vol. 4: Studies in Lexical Phonology.* Toronto: Academic Press.
Hayes, B. 1990. Precompiled phrasal phonology. In Inkelas, S. and Zec, D. (eds.). *The Phonology-Syntax Connection*, 85-108. Chicago: University of Chicago Press.
Hockett, C. F. 1947. Peiping phonology. *Journal of American Oriental Society* 67:253-267. Reprinted 1964 in M. Joos (ed.) *Readings in Linguistics I*, fourth edition, pp. 217-228. University of Chicago Press.
Hooper, J. B. 1976. Word frequency in lexical diffusion and the source of morphophonological change. In *Current progress in historical linguistics: Proceedings of the Second International Conference on Historical Linguistics*, ed. William M. Christie, Jr., 95-105. Amsterdam: North-Holland Publishing Company.
Hsiao, Y. 1991. *Syntax, Rhythm and Tone: A Triangular Relationship*. Taipei: Crane Publishing Co.
Hung, F. S., & Peters, A. M. 1997. The role of prosody in the acquisition of grammatical morphemes: evidence from two Chinese languages. *Journal of Child Language* 24: 627-650.

Kaisse, E. M. 1985. *Connected Speech: The interaction of Syntax and Phonology*. Toronto: Academic Press.
Kenstowicz, M. and Kisseberth, C. 1979. *Generative Phonology: Description and Theory*. New York: Academic Press.
Kiparsky, P. 1982. Lexical phonology and morphology, in I. S. Yang (ed.) *Linguistics in the Morning Calm*, 135-160. Seoul: Linguistic Society of Korea.
Kratochvil, P. 1984. Phonetic tone sandhi in Beijing dialect stage speech. *Cahiers de Linguistique - Asie Orientale* 13 (2):135-174.
Kratochvil, P. 1986. The case of the third tone. In the Chinese Language Society of Hong Kong (eds.) *Wang Li Memorial Volumes (English Volume)*, pp. 253-276. Hong Kong: Joint Publishing Co.
Labov, W. 1994. *Principles of linguistic change: internal factors.* Oxford: Blackwell.
Levelt, W. J. M. 1992. Accessing words in speech production: stages, processes and representations. *Cognition* 42:1-22.
Lin M. C., Lin J. Z., and Sun G. H. 1979. A study of tone sandhi in standard Chinese with computer. *Proceedings of the 9th International Congress of Phonetic Sciences*, 388.
Lyovin, A. V. 1978. Review of Rumjancev (1972). *Journal of Chinese Linguistics* 6(1):120-168.
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. Taipei.
Mei, T. L. 1977. Tones and tone sandhi in 16th century Mandarin. *Journal of Chinese Linguistics* 5:237-260.
Myers, J. 1993. *A Processing Model of Phonological Rule Application.* PhD dissertation, University of Arizona.
Myers, J., & Tsay, J. 2001. Testing a production model of Taiwanese tone sandhi. *Proceedings of the Symposium on Selected National Science Council Projects in General Linguistics from 1998-2000*, 257-279. National Taiwan University, Taipei.
Myers, J., & Tsay, J. Submitted. Neutralization in Taiwanese tone sandhi. Submitted to *Journal of Phonetics*, August 2002.
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.
Rumjancev, M. K. 1972. *Ton i intonacija v sovremennom kitajskom jazyke* [Tone and intonation in modern Chinese.] Moscow: Izdatel'stvo Moskovskogo (Gosudarstvennogo) Universitets.
Selkirk, E. 1996. The prosodic structure of function words. In *Signal to syntax: Bootstrapping from speech to grammar in early acquisition* (J. L. Morgan & K. Demuth, editors), pp. 187-214. Lawrence Erlbaum Associates.
Shen, X. S. 1990a. Tonal coarticulation in Mandarin, *Journal of Phonetics*, **18**, 281-295.
Shen, X. S. 1990b. *The Prosody of Mandarin Chinese.* University of California Press.
Shen, X. S. 1993. Relative duration as a perceptual cue to stress in Mandarin. *Language and Speech* 36: 415-433.
Shen, Y., Chao, C. Y., & Peterson, G. 1961. Some spectrographic light on Mandarin Tone-2 and Tone-3. *Study of Sounds* 9: 265-314.
Shih, C. 1986. *The Prosodic Domain of Tone Sandhi in Chinese*. PhD dissertation,

University of California, San Diego.

Tsay, J. 2002. 閩南語連讀變調與詞素變體選擇假設. [Taiwanese tone sandhi and the allomorph selection hypothesis]. *Dangdai Yuyanxue* [Contemporary Linguistics] 4 (3): 176-200.

Tsay, J., Charles-Luce, J. and Guo Y. S. 1999. The syntax-phonology interface in Taiwanese: acoustic evidence. *Proceedings of the International Congress of Phonetic Sciences* 14: 2407-2410.

Tsay, J. and Myers, J. 1996. Taiwanese tone sandhi as allomorph selection. *Proceedings of Berkeley Linguistics Society* 22: 394-405.

Tsay, J., and Myers, J. 2001. Processes in the production of Taiwanese tone sandhi: an acoustic phonetic study. *The Proceeding of the 5th National Conference on Modern Phonetics*, 233-237. Tsinghua University, Beijing.

Tseng, Chiu-yu. 1990. *An Acoustic Phonetic Study on Tones in Mandarin Chinese*. Special Publications No. 94. Institute of History and Philology, Academia Sinica, Taipei.

Wan, I. P. and Jaeger, J. 1998. Speech errors and the representation of tone in Mandarin Chinese. *Phonology* 15: 417-461.

Wang, W. S. Y, Li, K. P, & Brotzman, R. L. 1963. *Research on Mandarin phonology.* Project on Linguistic Analysis, Ohio State University Research Foundation. 6. 1-63.

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*. Doctoral dissertation, MIT. Reprinted in 1990 as *The Tonal Phonology in Chinese*.

Zee, E. 1980. A spectrographic investigation of Mandarin tone sandhi. *UCLA Working Papers* 49:98-116.

James Myers
Institute of Linguistics
National Chung Cheng University
Min-Hsiung, Chia-Yi, Taiwan 621, ROC
Lngmyers@ccu.edu.tw

Jane Tsay
Institute of Linguistics
National Chung Cheng University
Min-Hsiung, Chia-Yi, Taiwan 621, ROC
Lngtsay@ccu.edu.tw

Appendix: Mean values for three repetitions.

Table 1. Speakers.

Number	Gender
1	female
2	male
3	female
4	female
5	female
6	female
7	male
8	male
9	male
10	male

Table 2. tugai3 durations (msec).

Speaker	tu2	tu3
1	143.7	158.0
2	108.0	96.7
3	85.3	111.0
4	107.7	97.7
5	164.0	163.0
6	160.3	151.3
7	130.7	110.0
8	91.3	83.0
9	119.0	96.0
10	124.7	108.0

Table 3. tugai3 f_0 values (Hz).

	tu2			tu3		
Speaker	begin	mid	end	begin	mid	end
1	200.3	209.3	227.7	191.7	193.0	221.3
2	127.7	126.0	137.7	131.7	131.0	135.7
3	188.0	191.3	200.3	197.7	206.0	216.7
4	212.0	210.3	218.3	213.3	211.3	213.0
5	216.3	204.3	270.7	216.0	205.7	270.7
6	213.3	223.0	239.3	215.3	215.7	241.7
7	173.0	182.0	190.0	178.3	184.0	186.7
8	129.0	111.3	115.0	125.3	123.7	129.7
9	120.7	127.7	141.0	121.3	127.0	137.3
10	158.0	164.0	184.0	169.3	182.0	199.3

Table 4. yi1 (一) vs. yi2 duration comparisons (msec).

	yishi4		yixian4	
Speaker	yi1	yi2	yi1	yi2
1	150.7	182.3	148.3	185.3
2	143.0	154.0	147.0	150.3
3	152.7	115.3	112.0	169.7
4	152.3	159.7	131.0	159.0
5	171.3	199.0	122.7	197.7
6	180.3	203.0	181.7	207.0
7	102.7	176.0	103.3	155.0
8	107.7	156.7	90.0	107.7
9	120.7	159.0	125.3	157.0
10	134.3	168.0	113.3	131.3

Table 5. yi1 (一) vs. yi4 duration comparisons (msec).

	yiji3		yizhi1	
Speaker	yi1	yi4	yi1	yi4
1	179.7	172.0	100.7	140.3
2	127.7	122.3	108.3	115.0
3	121.3	131.3	76.3	67.3
4	123.3	121.7	122.3	104.0
5	150.7	152.3	112.7	115.7
6	205.7	197.7	132.7	141.0
7	87.0	122.7	75.7	122.0
8	62.0	122.7	64.3	100.0
9	94.0	133.3	91.0	167.0
10	135.7	149.3	125.7	135.0

Table 6. yi1 (一) vs. yi2 f_0 comparisons (Hz).

a. yi1shi4 vs. yi2shi4

	yi1			yi2		
Speaker	begin	mid	end	begin	mid	end
1	196.7	194.3	209.3	195.0	196.7	208.7
2	131.7	131.0	134.0	120.7	120.0	130.3
3	204.0	207.0	221.0	178.7	181.0	183.3
4	209.3	204.3	219.0	207.0	203.7	217.5
5	211.7	216.0	236.7	217.3	213.0	240.3
6	211.0	211.7	241.5	204.0	201.3	220.7
7	159.0	174.7	167.7	159.7	167.0	185.0
8	111.7	118.3	121.0	101.7	105.0	120.0
9	120.0	119.7	125.0	120.0	118.3	135.0
10	159.3	164.3	170.3	149.7	148.3	172.7

b. yi1xian4 vs. yi2xian4

	yi1			yi2		
Speaker	begin	mid	end	begin	mid	end
1	217.3	221.3	222.7	201.0	204.3	220.7
2	121.5	121.5	132.0	117.3	123.3	146.3
3	176.0	179.0	181.5	191.0	200.0	232.5
4	191.3	196.0	205.7	209.7	200.7	215.7
5	222.3	219.7	249.3	208.3	214.3	244.7
6	209.0	214.0	233.7	204.3	200.3	222.0
7	161.3	166.7	174.7	165.0	168.7	179.3
8	105.0	106.5	110.5	105.7	111.7	118.3
9	126.0	123.3	136.7	116.0	114.0	125.7
10	144.7	148.0	147.7	152.0	171.0	168.0

Table 7. yi1 (一) vs. yi4 f_0 comparisons (Hz).

a. yi1ji3 vs. yi4ji3

	yi1			yi4		
Speaker	begin	mid	end	begin	mid	end
1	275.7	261.0	219.0	280.7	218.3	187.0
2	147.7	149.3	139.0	145.0	138.3	130.0
3	255.0	240.3	206.0	222.0	212.7	186.3
4	250.0	241.7	205.7	254.7	252.3	208.0
5	313.7	311.3	241.3	302.7	298.0	237.3
6	231.3	248.0	246.0	270.3	260.0	200.0
7	205.7	203.0	171.7	214.3	209.7	169.0
8	146.3	147.3	138.0	152.3	137.0	116.0
9	152.7	155.7	135.3	165.7	152.7	124.7
10	204.3	205.7	145.7	199.0	185.7	142.0

b. yi1zhi1 vs. yi4zhi1

	yi1			yi4		
Speaker	begin	mid	end	begin	mid	end
1	262.3	260.7	247.3	260.7	254.7	219.3
2	146.7	143.3	129.0	141.0	136.0	134.0
3	250.7	243.0	223.7	230.3	234.0	215.0
4	240.3	244.0	209.0	249.0	246.0	207.3
5	282.0	294.5	238.5	318.7	300.0	240.3
6	248.0	260.7	230.7	258.3	272.0	235.0
7	208.3	206.7	175.0	214.7	204.0	166.3
8	143.7	140.3	133.7	152.7	137.7	119.0
9	156.7	160.0	136.0	168.7	152.7	129.0
10	206.7	200.3	174.0	194.7	195.3	154.0

華語變調之聲學探究

麥傑、蔡素娟
國立中正大學

本文的目的在討論華語變調發聲的聲學證據。我們比較變調的兩個觀點：標準“範疇性的”觀點認為華語三聲變調類似英語冠詞 a/an 的分詞關係；“階層性的”觀點則認為華語三聲變調類似英語齒齦塞音變拍音的轉換。我們提出證據來支持兩個論點：首先，根據文獻上有關華語三聲變調的聲學研究，北京華語的三聲變調與“階層性的”觀點比較一致，而其他地方（包括台灣）的華語三聲變調與“範疇性的”觀點比較一致；其次，我們的三聲變調與「一」的特殊變調的聲學研究也支持台灣華語的三聲變調是“範疇性的”，雖然這些結果同時也顯示變調的聲學研究在實證與理論上容易犯的錯誤。