Proceedings of the 22nd North American Conference on Chinese Linguistics (NACCL-22) & the 18th International Conference on Chinese Linguistics (IACL-18). 2010. Vol 1. Clemens, L.E. & C.-M. L. Liu, eds. Harvard University, Cambridge, MA. 283-300.

Testing the Allomorph Selection Hypothesis in Taiwanese Tone Sandhi

Szu-wei Chen, James Myers and Jane Tsay
National Chung Cheng University

This study examines Taiwanese tone sandhi, which is hypothesized here as a process of choosing between two allomorphs that are both listed in the lexicon. A production experiment was designed to test this hypothesis. Effects of age group, allomorph frequency, and positions of the morpheme were examined in the experiment. The results showed that (1) older speakers had significantly higher accuracy rate than younger ones, (2) Taiwanese TTS production was significantly influenced by allomorph frequency even with morpheme frequency factored out, (3) for young speakers, the allomorph frequency effect was modulated by the position of the target item, (4) a majority of the tone production errors were allomorph selection errors, (5) there was no consistent pattern in productivity across the five tone categories.

1.Introduction

Taiwanese Tone Sandhi (TTS) has been a challenging issue in both rule-based analyses (Wang 1967, Yip 1980, Tsay 1994, Chen 2000) and constraint-ranking analyses (Moreton, 2004). Such grammar-based analyses have been strongly criticized by several researchers as lacking psychological evidence (Hsieh 1970, 1975, 1976, Wang 1995). In this paper, a short review on Taiwanese Tone Sandhi is first introduced, followed by some crucial previous research on this issue. Then, a production experiment is described to test an alternative lexicon-based hypothesis for Taiwanese Tone Sandhi.

1.1 Taiwanese Tone Sandhi

Taiwanese is a branch of Southern Min Chinese spoken in Taiwan, also known as Minnan, closely related to the Xiamen (Amoy) dialect. Over seventy percent of people in Taiwan speak Taiwanese as their first or second language (Huang 1993). It has seven lexical tones, including five long tones and two short ones, as shown in **Table 1.** Tone 4 and Tone 8 are short tones ending with unreleased voiceless stops and thus called checked tones. It should be noted that almost all morphemes in Taiwanese are monosyllabic and carrying a lexical tone. TTS is a tone alternation between the syntactically-defined phrase-final (juncture) and non-final (context) forms of a morpheme, where juncture tones occur at the right edge of an XP (Chen 1987, Lin

1994). For example, when two Tone 2 morphemes *cui*^{53} “water” and *ko*^{53} “fruit” are combined toform a word *cui*^{55} *ko*^{53} “fruit”, the morpheme *cui*^{53} “water” will have a context tone *cui*^{55}. Similarly, when *cui*^{55} *ko*^{53} “fruit” is followed by another morpheme *tiam*^{21} “store” to form *cui*^{55} *ko*^{55} *tiam*^{21} “fruit store”, *ko*^{53} will have a context tone *ko*^{55}.

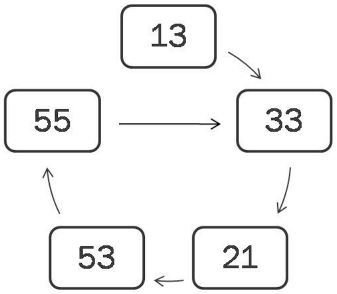
Table 1. Taiwanese Tone Inventories[1]
Underlined tones (T4 and T8) are checked tones.
The tonal values on juncture and context positions are based on a 5-point scale.

Tone	Morpheme	Juncture	Context
1.Ying Ping	/si/‘poem’ 詩	55 讀詩	33 詩文
2.Ying Shang	/si/ ‘death’ 死	53 驚死	55 死人
3.Yin Qu	/si/ ‘four’ 四	21 第四	53 四本
4.Yin Ru	/sik/ ‘color’色	3 白色	53 色彩
5.Yang Ping	/si/ ‘time’ 時	13 當時	33時間
6.Yang Shang	Neutralized with Tone 7		
7.Yang Qu	/si/ ‘temple’寺		21 寺鐘

1 熟識

The five non-checked tones participate in a set of five alternations. TTS does not create new tones and the alternations form a tone circle (Bodman 1955, Wang 1967), as shown in **Figure 1**. In rule-based analyses, juncture tones become context tones. In this study, only the five non-checked tones were included, because we attempted to control other nuisance variables as much as possible. First, checked tones have much lower morpheme frequency than the long tones, and there are some dialectal variations (Hsieh, 1970). Second, since checked tones are not in the tone circle, their tone sandhi patterns are different. Nonetheless, this is not intended to imply that the hypothesis we will test would not also apply to checked tones.

[1] The tonal system is based on the Taiwan Southern Min Pinyin System announced by Ministry of Education (2008), Taiwan.



Tone sandhi rules

a. 55 → 33
b. 13 → 33
c. 53 → 55
d. 21 → 53
e. 33 → 21

Figure 1. Taiwanese tone circle

1.2 Previous research on TTS

Hsieh (1970) first questioned the psychological reality of the sandhi rules. He conducted an experiment to test Taiwanese speakers' ability of producing tone sandhi in disyllabic accidental gaps. His results showed that only actually occurring morphemes received 100% correct tone productions whereas other types of nonce forms showed only 10% to 30% accuracy rates. He claimed that this difference was due to the fact that there were no sandhi forms of the accidental gaps stored in speakers' lexicon. He therefore proposed that the two morphophonemic alternants of each single morpheme were both listed in the lexicon and hence were part of the speakers' knowledge.

Hsieh (1975) further confirmed the role of lexically stored tonal allomorphs in TTS production by both children and adults. His first experiment examined three children's tone sandhi production on the stimuli of trisyllabic real and artificial compounds (#XY+Z#). For example, the child was given the name of the fruit #kin^{33}cio^{55}# "banana", and then he or she was asked to offer the name of the store, that is, #kin^{33}cio^{33}+tiam^{21}# "banana store", or given the name of the store, the child had to offer the name of the fruit. The result showed that the accuracy rates were very different across speakers whether compared across all four tests (real word/nonce word × forward/backward operation), in a particular test, or even within a tone category. His second experiment on five adults Taiwanese speakers also presented similar results. Different items governed by the same rule may receive different treatments from the same speaker. Even though two items may receive the same types of responses, the rate of rule application for them may differ greatly. Moreover, for the same test item or the same category, various speakers may react with different degrees of accuracy. However, regardless of this great variability across speakers and different items, one important finding in his study was that for almost every speaker, their rates of applying rules a (55→33) and b (13→33) in **Figure 1** were higher than others rules such as c (53→55), d (21→55) and e (33→21). Based on a computerized Chinese dialectological database, Hsieh (1975) found that the high level tone (55) and the rising tone (35) are the most frequent ones in the distribution of Xiamen syllables among the five long tones. Therefore, he argued that a speaker who had different degrees of lexical familiarity with those tone categories might have different performance. The power of association or analogical power provided a better explanation than variable rules. Since speakers relied on actual words in their lexicon for the association, the higher the frequency of a tone category, the easier an artificial member of this category can be associated with the actual items. In Experiment 3, a native speaker of Taiwanese who moved to the U.S. when she was eleven was tested with the same stimuli in the Experiment 1. The only difference was she went through three trials, which took place at different times. Her scores in all tests improved progressively through trial 1 to trial 3. Hsieh (1975) thus concluded that in the process of learning, both children and adults treated each allomorph of a new morpheme in natural acquisition as an independent item rather than a token of a category.

Wang (1995) also agreed that tone sandhi rules had a certain degree of productivity, but might apply only to familiar lexical items. Therefore, he included lexical familiarity as one of the independent variables. A longitudinal experiment was conducted to investigate tone sandhi behaviors of Taiwanese native speakers. He used mostly nonwords composed of accidental phonotactic gaps, not tonotactic gaps, which when substituted with another tone, could become a real word. Those words and their pretended meanings were taught to the participants in random order in the first meeting. In the following five meetings, the targets were reviewed and carrier sentences were used to elicit speakers' production in both context form and juncture form. His results showed that the speakers' rule-application rates were higher at the end of the experiment than they were at the beginning. The overall productivity was over 50%, though the range of correct responses was rather large. He questioned if those tone sandhi rules really defined competence as the generative linguists assume, since the native speakers' performance demonstrated such a great variability and individual differences were also great across tones. He argued that a simple lexicon was built at the expense of complex psychological operation, by which he meant people's ability of analogy. Speakers were not easily able to allocate a new item in the phonological system, but after a while it was analogized to one of the patterns in the system and found its relations with other items. Thus, TTS may be an analogical chain, rather than a system of rule applications.

Building on Hsieh (1970, 1975) and Wang (1995), Tsay & Myers (1996) argued that Taiwanese Tone Sandhi is a case of Lexical Phrasal Phonology, based on its sensitivity to syntactic structure and its lexical properties. First, following Chen (1987) and Lin (1994), the tone group is syntactically defined, rather than prosodically defined. Secondly, TTS has three lexical properties: lexical idiosyncrasies, semi-productivity, and categoricality, the third point being supported by phonetic studies (Tsay, Charles-Luce & Guo 1999, Tsay & Myers 2001, Myers & Tsay 2008). Moreover, since the explanatory power of previously proposed TTS rules is very limited, they proposed that the only lexical process in TTS is the choosing between two allomorphs that are both listed in the lexicon.

Zhang, Lai & Turnbull-Sailor (2006) proposed a somewhat different view of the TTS productions, arguing that the productivity of TTS rules was affected by phonetics. Following Hsieh (1970)'s experimental design, they constructed five types of disyllabic words. First, they predicted tone sandhis within the tone circle were unproductive whereas phonotactically driven 13 →33, the one outside the circle, was productive. Secondly, they predicted that productivity within the tone circle was subject to a phonetic effect, so that productivity should be the highest when the sandhi changes a longer tone into a shorter tone, since syllables in context position are phonetically shorter than in juncture position. Thus, the shorter tones 21 and 51 should be the preferred sandhi tones. Their results supported their predictions, except that the phonotactically driven rule (13→33) was not as productive as expected. Zhang et al. (2010) replicated and extended these basic findings, and proposed a formal model that had not only a phonetic component for what they see as the universally motivated aspects of TTS, but also allomorph-specific constraints to capture the well-established lexical idiosyncrasy and lack of full productivity of TTS.

Although these studies had somewhat different aims, all explicitly tested and confirmed the role of lexical familiarity in TTS production, and they also agreed that great variability existed across speakers and across different tonal categories. In general, Taiwanese tone sandhi rules seem largely unproductive when tested on nonce words. This makes the lexical nature of TTS a better explanation than rule application. Another piece of evidence is that the speakers in TTS production studies only seem to have two options when producing nonce forms, juncture or context tones; non-sandhi errors are very rare. According to Hsieh (1970, 1975) and Wang (1995), most of the incorrect responses were due to the non-application of rules; that is, speakers produced the juncture tone instead of its context counterpart. Similarly, Zhang et al. (2006) reported only 11.5% of nonce words in the tone circle had the correct sandhi; 82.9% had non-applications. This showed that only 5.6% were other tonal error types. The observation that speakers only have these two choices when producing tones is consistent with the allomorph selection hypothesis.

This hypothesis also predicts that TTS production should be affected by allomorph frequency. The lexical frequency reflects the amount of prior experience that a native speaker has had with a lexical element, and it is well established that this factor affects the retrieval of lexical items from memory (Phillips 1984, Losiewicz 1992, Jescheniak & Levelt 1994, Myers & Guy 1997, Bybee 2001). Although this concept is usually discussed with relation to words or morphemes, the unit relevant to the allomorph selection hypothesis is the allomorph. Given that there are two tonal allomorphs for each morpheme in Taiwanese, morpheme frequency is the sum of juncture allomorph frequency and context allomorph frequency. We thus predict that accuracy in TTS production will be higher for allomorphs of higher frequency. For example, if morpheme X has higher context allomorph frequency than morpheme Y, we predict that TTS production will be more accurate in the nonce compound XZ than in YZ. In our experiment, allomorph and morpheme frequency estimates were based on the Taiwanese Spoken Corpus (Myers & Tsay 2010) of National Chung Cheng University, which is the largest available corpus of spontaneous conversations in Taiwanese, transcribed from radio talk shows in Southern Taiwan. As of May 2010, it contained about 607,000 word tokens.

Finally, based on Taiwanese tone sandhi's lexical nature and the effects of lexical experience on TTS observed by Hsieh (1975) and Wang (1995), we expect that older speakers, who have processed TTS more often, will have higher accuracy rate in tone sandhi production compared with younger speakers, though many other factors may also affect cross-age differences besides lexical experience.

Summarizing the above discussion, our predictions for Taiwanese native speakers' tone production are as follows:

1. Subjects will tend to produce more errors for lower allomorph frequency items and vice versa.
2. Older speakers should have significantly better performance than younger speakers in TTS production.
3. Most of the tone production errors should be allomorph selection errors if all the other phonetic environments are controlled. Other errors may be lexical retrieval errors or errors at the phonetic implementation level.

2. Method

2.1 Participants

Two groups of 12 Taiwanese speakers each were recruited and paid for their participation. One group of speakers (1 male, 11 females) were younger, with an average age of 21 (range: 19-28). The other group of speakers (6 males, 6 females) were older, with an average age of 51 (range: 43-58). None of them reported having any speech or hearing disorders. All of them had to pass a proficiency pretest, in which they had to listen to auditory instructions in Taiwanese, read Chinese characters, and produce sentences in Taiwanese. If the experimenter thought the participant could not produce Taiwanese sentences fluently and correctly, then he or she would be removed from the production task. The purpose of this proficiency pretest was to make sure that the speakers had reached a certain level of proficiency.

2.2 Stimuli

Our wordlist was based on Taiwanese Spoken Corpus at National Chung Cheng University (Myers & Tsay 2010). The pseudowords in our wordlists were composed of two actually occurring morphemes in Taiwanese. The first morpheme was always /ti^{55}/ “pig” followed by our target morphemes. To get both context and juncture positions in a sentence, we used the carrier sentence [gua^{55} be^{55} khi^{53} ***XY + Z*** # be^{55} ***XY***] “I am going to the *XY* shop to buy *XY*”. **Table 2** displays two sets of the target morphemes. These two types of stimuli were based on token allomorph frequency in the corpus: (i) juncture-preferring morpheme set, in which each morpheme had higher juncture allomorph frequency than its context counterpart, as shown in **Table 2a**, and (ii) context-preferring morpheme set, in which each morpheme had lower juncture allomorph frequency than its context counterpart, as shown in **Table 2b**. For instance, since the Tone 1 morpheme *chia* “car” occurs more often in juncture position (e.g. *hue*^{55} *chia*^{55} “train (fire-car)”, *sai*^{55} *chia*^{55} “driving a car”) than in context position in the corpus, it belongs to the juncture-preferring morpheme group. The confounding between morpheme and allomorph frequency seen in the table is inherent to the Taiwanese lexicon, but can be teased apart statistically, as explained below.

Table 2 Two sets of target morphemes and the carrier sentence

a. Juncture-preferring morpheme targets

Tone category	Juncture preferring	Gloss	Juncture Frequency	Context frequency	Morpheme Frequency
1	chia (車)	“car”	299	53	352
2	chiu (手)	“hand”	187	76	263
3	chiunn (唱)	“sing”	359	144	503
5	cinn (錢)	“money”	402	11	413
7	hun (份)	“one share”	209	18	227

b. Context-preferring morpheme targets

Tone category	Context preferring	Gloss	Juncture frequency	Context frequency	Morpheme Frequency
1	chiu (秋)	“autumn”	18	22	40
2	hue (火)	“fire”	53	86	139
3	siu (秀)	“elegant”	5	14	19
5	hue (回)	“return”	43	49	92
7	chiu (樹)	“tree”	2	42	44

X= ti^{55} “pig”
Y= target morpheme in our wordlists
Z= tiam^{21} “store, shop”

Carrier sentence	[gua^{55} be^{55} khi^{53} ***XY + Z*** # be^{55} ***XY***]
Gloss	I am going to the ***XY shop*** to buy ***XY***.

Our stimuli were presented in written form; thus, a pre-test was run to test the transparency of Chinese characters. Although Taiwanese native speakers may have little experience reading Chinese characters as their orthography, auditory forms of stimuli involve homophones and thus may not trigger the right lexical entry. In total, our stimuli included ten target morphemes with five different tones in each group and ten fillers of the same tones. There were three main nuisance variables we intended to control when selecting the target morphemes: (1) their Chinese character transparency, (2) their phonetic context, and (3) their morpheme frequency. The first two seem reasonably well controlled in our materials, and we show below how we factored out the third in our statistical analysis.

2.3 Elicitation procedure

Participants sat in front of a computer with 17-in. monitor in a sound-treated phonetics lab of the Institute of Linguistics at National Chung Cheng University. After they passed the proficiency test, they moved on to do the production experiment. The whole process (proficiency test + production experiment) took about 25 minutes. There were three sessions in the production experiment: one practice trial, the main production task, and a post-test, each with a short break in between. The practice trial was used to familiarize participants with the procedure that would appear later in the production task; the wordlist used was different from that in the production task. During the break, the experimenter would check if they had problems with the procedure. In the production task, a disyllabic word consisting of two actually occurring morphemes was visually presented and subjects had to put it into the carrier sentence and read out the whole sentence. For example, when a disyllabic word ***tua***^{21} ***cun***^{13} “big ship” was presented, subjects had to read out loud [gua^{55} be^{55} khi^{53} ***tua***^{21}***cun***^{33} ***tiam***^{21} be55 ***tua***^{21}***cun***^{13}]. Finally, a post-test was conducted to compare the F_0 of the actual occurring morpheme when in the real and the pseudo word and to check if the speaker knew how to pronounce the actually occurring morpheme. These twenty pseudo words (10 target words + 10 fillers) were randomized and presented on the screen to elicit the speakers’ production. There were four repetitions divided into two blocks with a short break in between. The stimuli were shown on the computer screen one by one after an affixation cross with about 5000 ms interval.

2.4 Recording and F0 extraction

E-prime (Schneider, Eschman & Zuccolotto 2002) was used to present the stimuli and record productions, which were directly digitized into the computer. A microphone (MUD-326) was placed about 5-10 inches in front of the speaker's lip. 90 sentences were recorded for each speaker, including 40 target pseudo word sentences, 40 filler sentences and 10 sentences with real words in the post-test. In each target sentence, two tokens (one context and one juncture form) were included in our analysis. In total, 1920 target tokens (24 participants × 10 target morphemes × 4 repetitions × 2 positions) were included in our analyses. The extraction of F_0 contours was done using a script (Xu, 2009) in Praat (Boersma & Weenink, 2009), which generated a time-normalized mean F_0 contour, mean F_0, maximum and minimum F_0 and duration of each syllable.

3. Data Analysis and Results

The only dependent variable in this study was tone production accuracy, that is, whether subjects produced the correct tone or not. Productions were transcribed by a native speaker using the time-normalized F_0 contours and mean F_0 of all the target items as references. Whenever the transcriber was not entirely sure about which tone category the target item belonged to, she referred to the mean F_0 or F_0 contours of the same subject's production in real words. **Figure 2** and **Figure 3** display the time-normalized F_0 contours of five juncture and context tones by one of the older speakers. As **Figure 3** shows, both Tone 1 and Tone 5 were produced as a mid-level tone, since there is no rising tone in the context position. In our data, each age group had 960 tokens (12 subjects × 10 target morphemes × 4 repetitions × 2 positions). The older speakers had 4 recording errors, while the younger speakes had 12 recording errors, which were excluded from the valid tokens. The recording errors were mainly caused by the delay of speakers' production, since there were only 5000 ms to record their production.

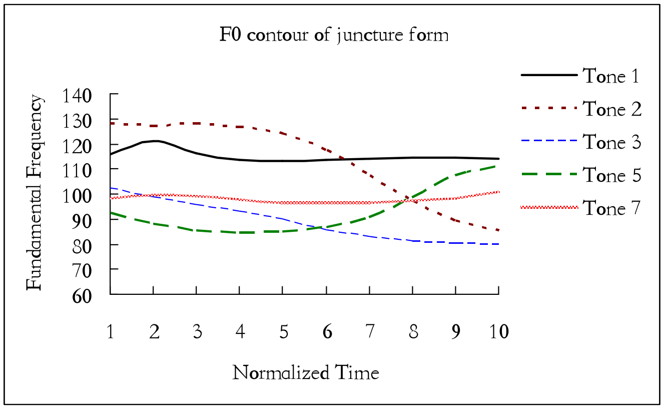


Figure 2. Time-normalized mean F_0 contour of five juncture tones produced by one of the speakers in the old generation group. Each curve is an average of 4 repetitions.

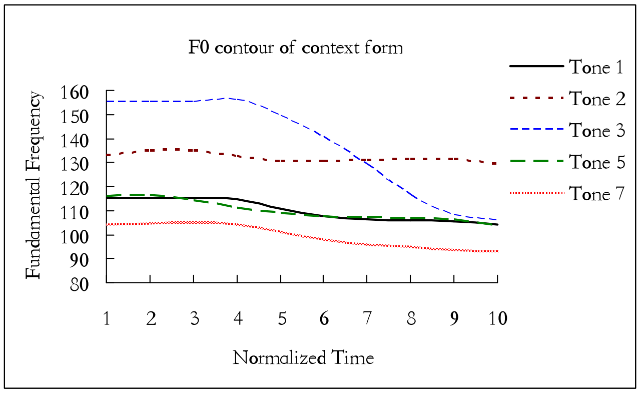


Figure 3. Time-normalized mean F_0 contour of five context tones produced by one of the speakers in the old generation group. Each curve is an average of 4 repetitions.

3.1 Accuracy rate and error type

Figure 4 shows the average accuracy rates of two different generations. The old generation has an overall accuracy rate of 94.56%, ranging from 87.34% to 100% across speakers, while the young generation has only 68.78%, ranging from 48.75% to 87.75%. The denominators are the valid tokens, which are total tokens minus recording errors.

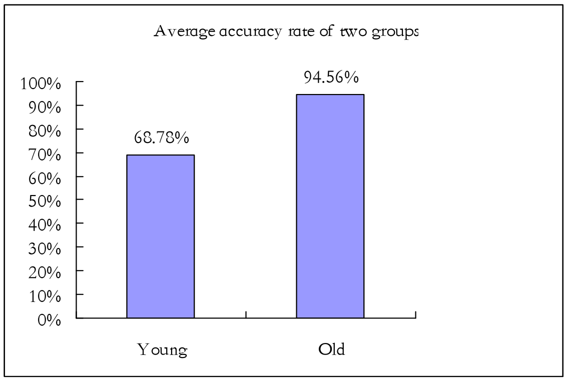


Figure 4. Average accuracy rates of tone sandhi production by old and young generation

Figure 5 displays the number of accurate tokens in both context and juncture positions produced by old and young speakers. Older speakers did well in both positions, with 452 accurate tokens in each position, but younger speakers had worse performance, with 260 accurate tokens in the context position and 392 ones in the juncture position.

Figure 6 illustrates the number of accurate tokens in five non-checked tones by the old and young speakers. The order on the number of accurate tokens from high to low for older speakers is T3>T1>T2=T5>T7, while it is T5>T2>T1>T3>T7 for younger speakers. There is no consistent pattern across two groups except that T7 had the lowest accuracy rate.

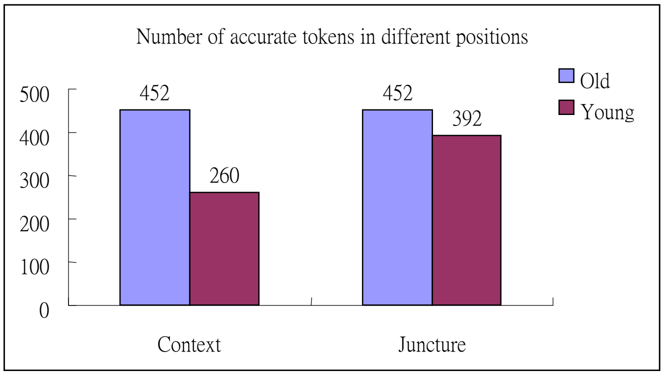


Figure 5. Number of accurate tokens in different positions by older and younger speakers

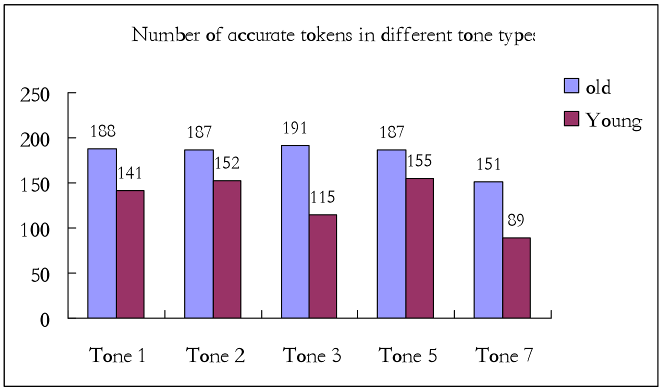


Figure 6. Number of accurate tokens in five non-checked tones by older and younger speakers

Most of the speakers' errors were allomorph errors, that is, substituting context tones with juncture ones or vice versa. As shown in **Figure 7**, younger speakers made 243 allomorph errors and 53 other tonal errors out of 948 valid tokens, whereas older speakers had only 20 allomorph errors and 16 other tonal errors out of 956 valid tokens. 16 error tokens produced by older speakers belonged to the wrong target word type, which means speakers did produce both context and juncture forms, but of words different from the target ones. For example, the target Tone 7 morpheme was pronounced as Tone 1 morpheme, but speakers changed the context and juncture form consistently. This may be an effect of orthography; our target morpheme *hun*^{33} "份" should be pronounced as Tone 7 while another similar character *hun*^{55} "分" should be pronounced as Tone 1. Other tonal errors were those where the target tone was pronounced as a tone other than its allomorph. For example, the context position of Tone 1 should be mid level tone. When it was pronounced as a falling, rising, or low tone, then it was regarded as other tonal error. When it was pronounced as a high level tone, then it was an allomorph error.

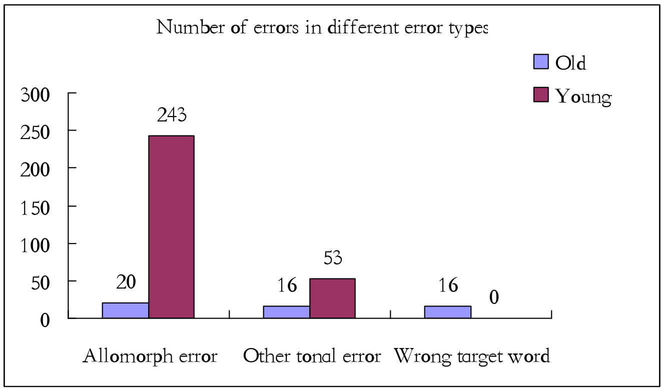


Figure 7. Number of tonal errors produced by 24 speakers. There are 348 error tokens out of 1904 valid tokens in total.

3.2 Statistical analyses

The two key factors in the experimental design were allomorph frequency and age group. Since frequency distributions are skewed, we took the logarithm to make them more normally distributed. Our data were analyzed by mixed effects logistic regression, a statistical technique that can handle both categorical data (accuracy as a binary measure) and repeated measures at the same time (Agresti, Booth, Hobert & Caffo 2000). The free statistical software R (R Development Core Team 2006) and its additional package lme4 (Bates & Sarkar 2007) were used to run the test. Our models analyzed the effect on accuracy of age group, allomorph frequency, morpheme frequency, position, and the interactions among them. As noted earlier, allomorph and morpheme frequency are correlated in the Taiwanese lexicon. To make sure these two values could still be statistically distinguished, we computed the condition number *k* (Baayen 2008), where values of 30 or higher indicate serious confounding. In our data, the condition number of allomorph and morpheme frequency was merely *k*=13.9, which implies that these two factors were not seriously confounded. Since mixed effects regression has the advantage of being able to deal with more than one random variable, separate by-speaker and by-speaker-and-item analyses were run. A likelihood ratio test showed that the by-speaker-and-item was significant better than the simpler model ($\chi^2(1)=109.49$, *p*<.001).

The result of the by-speaker-and-item analysis showed significant main effects of age group (b=-3.02, *z*=-2.00, *p*<.05) and allomorph frequency (b=0.75, *z*=2.00, *p*<.05), but no significant main effects of position (b=-1.56, *z*=-0.58, *p*>.05) or morpheme frequency (b=-0.11, *z*=-0.33, *p*>.05). There was also a significant two-way interaction between position and allomorph frequency (b=-2.19, *z*=-2.22, *p*<.05), and a significant three-way interaction among age group, position, and allomorph frequency (b=2.17, z=2.36, *p*<.05), but no interactions with morpheme frequency (*p* > .05). The logic of the statistical analysis thus suggests that it is indeed allomorph frequency that affected accuracy, not morpheme frequency.

As **Figure 8** shows, in context position, the higher the allomorph frequency, the higher the accuracy rate, whereas accuracy rate was less influenced by allomorph frequency in the juncture position.

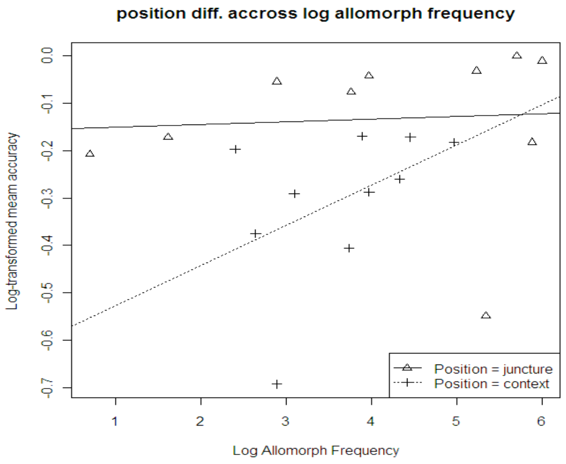


Figure 8. Interaction plot of position vs. allomorph frequency

Figure 9 and **Figure 10** illustrate the three-way interaction among age group, position, and allomorph frequency. Younger speakers had a consistent general pattern of lower accuracy rate in lower allomorph frequency items in both context and juncture positions, but the frequency effect was larger in context position than juncture position. Older speakers showed a similar pattern with younger speakers in the context position. However, in the juncture position, older speakers had essentially identical accuracy rates for higher allomorph frequency items.

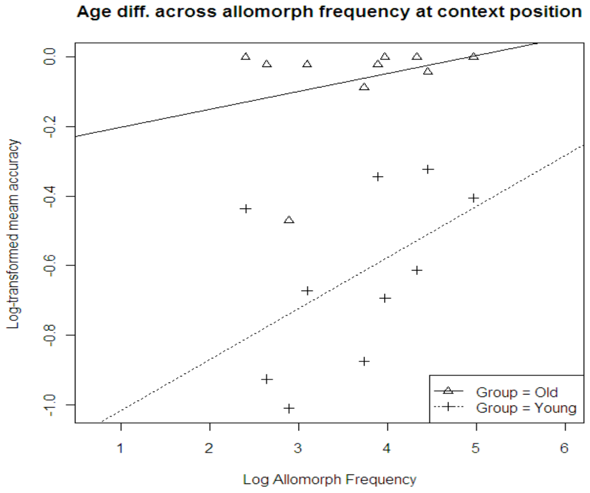


Figure 9. Different age group across allormorph frequency at context position
The solid line represents older speakers and the dotted line represents younger speakers.

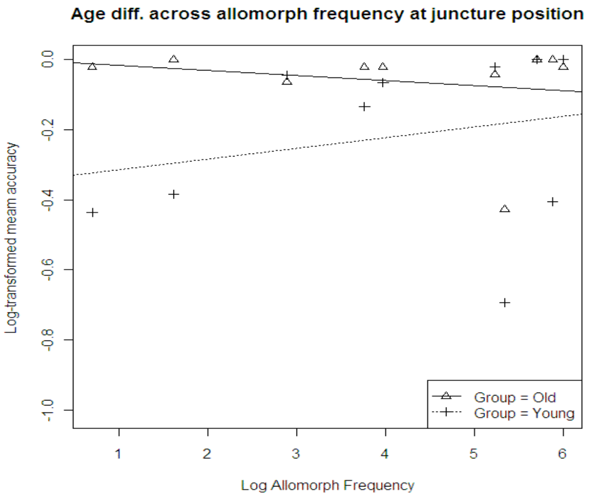


Figure 10. Different age group across allormorph frequency at context position
The solid line represents older speakers and the dotted line represents younger speakers.

4. Discussion

The present study was designed to test the Allomorph Selection Hypothesis, which claims that TTS does not involve rule applications in the tone circle, but rather involves the process of choosing between two allomorphs that are both listed in the lexicon. Going beyond previous studies that only indirectly showed the lexical effect, we directly examined allomorph frequency to investigate how this lexical effect influences native speakers' production, with morpheme frequency factored out statistically.

We found that the older speakers did significantly better (94.56% accuracy rate) than the younger speakers (68.78%) in TTS production. Older speakers also showed greater consistency in TTS production (SD=0.04), while younger speakers showed more fluctuations in their performance (SD=0.15). This is consistent with our claim that lexical familiarity does play a role in TTS production, though of course the two speaker groups probably also differed in their overall Taiwanese competence, not just in their lexicons. Moreover, as **Figure 5** shows, older speakers did equally well in both context and juncture positions, while younger speakers' performance in the context position was significantly worse than that in the juncture position (*b*=3.30, *z*=2.86, *p*<.01). Crucially, this performance discrepancy was modulated by allomorph frequency. As illustrated in **Figure 8**, higher allomorph frequency items had an overall higher accuracy rate. However, in the case of older speakers, since their overall accuracy rate is already very high, there may be a ceiling effect. Although this result confirms that both allomorphs are stored in the mental lexicon, there may be some asymmetry in the storage and/or retrieval of the two forms.

As for the tone categories, similar to the results of previous studies, performances varied across different speakers in both groups. Since different speakers may have different degree of exposure to different lexical items, it is natural to have this kind of individual variances. This poses a problem to Zhang et al. (2006, 2010)'s claim that different tone categories have consistently higher or lower productivity rates. As illustrated in **Figure 6**, though the productivity rates are different in the two speaker groups, there is no consistent pattern across the five tone categories. Young speakers made the most errors in Tone 3 and Tone 7, which is the opposite of Zhang et al. (2006)'s prediction that speakers should favor sandhi rules that shorten tones. Different degrees of lexical familiarity seem to offer a better explanation for the variability across speakers and across tone categories.

Although not all the tonal errors were allomorph errors, 77.01% of the tone production errors were allomorph selection errors. Other tonal errors and wrong target errors account respectively for 19.83% and 5.6% of the total errors. Since allomorph errors were the majority, this pattern seems to provide further evidence to support the Allomorph Selection Hypothesis. However, as mentioned earlier, TTS production may involve asymmetries in lexical retrieval or storage. Further experiments are required to explore such issues, in particular, the time course of lexical retrieval relative to the phonetic implementation of tone production.

5. Conclusion

The results of the TTS production experiment show that: (1) Older speakers had significantly higher accuracy rates than younger ones in tone sandhi production, which might be caused by different degree of lexical familiarity to Taiwanese. (2) Taiwanese TTS production was significantly influenced by allomorph frequency even with morpheme frequency factored out, which supports the claim that both forms are listed in the mental lexicon. (3) The allomorph frequency effect for young speakers was modulated by the position of the target item. (4) A majority of the tone production errors were allomorph selection errors. (5) There was no consistent pattern in productivity across five tone categories.

Thus, our results support the argument that grammar-governed tone alternation is not the nature of TTS. It is therefore better to maintain that TTS involves the process of choosing between two allomorphs that are both listed in the lexicon. Our results of a significant positive main effect of allomorph frequency may provide some new evidence to this claim. Nevertheless, how exactly these two allomorphs are listed in the mental representation begs for more careful research in both TTS perception and production studies.

References

Agresti, Alan, James G. Booth, James P. Hobert & Brian Caffo. 2000. Random-effects modeling of categorical response data. Sociological Methodology, 30, 27-80.

Baayen, R. Harald. 2008. Analyzing Linguistic Data: A Practical Introduction to Statistics Using R, Cambridge & New York: Cambridge University Press.

Bates, Douglas & Deepayan Sarkar. 2007. lme4: Linear mixed-effects models using S4 classes. R package version 0.9975-12.

Bodman, N. C. 1955. Spoken Amoy Hokkien. Kuala Lumpur.

Boersma, Paul & David Weenink. 2009. Praat: A system for doing phonetics by computer [Computer program]. http://www.praat.org/

Bybee, J. L. 2001. Phonology and Language use. Cambridge: Cambridge University Press.

Chen, M. Y. 1987. They syntax of Xiamen tone sandhi. Phonology Yearbook 4. 109-149.

Chen, M. Y. 2000. Tone sandhi: Patterns across Chinese dialects. Cambridge University Press, Cambridge, UK.

Huang, Shuanfan. 1993. Language, Society, and Ethnic Identity. Taipei: Crane.

Hsieh, Hsin-I. 1970. The psychological reality of tone sandhi rules in Taiwanese. In Papers from the 6th Meeting of the Chicago Linguistics Society. Chicago. 489-503.

Hsieh, Hsin-I. 1975. How generative is phonology. In E. F. Koerner (ed.) The transformational generative paradigm and modern linguistic theory. John Benjamins, Amsterdam. The Netherlands. 190-144.

Hsieh, Hsin-I. 1976. On the unreality of some phonological rules. Lingua. 38:1-19.

Huang, Shuanfan. 1993. Language, Society, and Ethnic Identity. Taipei: Crane.

Jescheniak, J. D. & Levelt, W. J. M. 1994. Word frequency effects in speech production: Retrieval of syntactic information and of phonological form. Journal of Experimental Psychology: Learning, Memory, and Cognition 20(4). 824-843.

Lin, J.-W. 1994. Lexical government and tone group formation in Xiamen Chinese. Phonology 11. 237-275.

Losiewicz, B. L. 1992. The effect of frequency on linguistic morphology. University of Texas Dissertation.

Ministry of Education. 2008. Taiwan Minnanyu Luomazi Pinyin Fangan Shiyong Shouce. [Guidebook of the Romanization of Taiwan Southern Min.]

http://www.edu.tw/files/bulletin/M0001/tshiutsheh.pdf

Moreton, Elliot. 2004. Non-computable functions in Optimality Theory. In John J. McCarthy (ed.). Optimality Theory in Phonology. Blackwell Publishing, Malden, MA. 141-164.

Myers, J. & Guy, G. R. 1997. Frequency effects in variable lexical phonology. University of Pennsylvania Working Papers in Linguistics, 4 (1). 215- 228.

Myers, J. & Tsay, J. 2008. Neutralization in Taiwan Southern Min Tone Sandhi. In Y. E. Hsiao, H.-C. Hsu, L.-H. Wee, and D.-A. Ho (Eds.) Interfaces in Chinese phonology: Festschrift in honor of Matthew Y. Chen on his 70th birthday (pp. 47-78). Language and Linguistics Monograph Series Number W-8. Taipei, Taiwan: Academia Sinica.

Myers, J. & Tsay, J. 2010. Parallel Experimental and Corpus-based Analyses of Southern Min Phonology (2/3). NSC Project Progress Report. NSC 97-2410-H-194-067-MY3.

Phillips, B. 1984. Word frequency and the actuation of sound change. Language 45. 9-25.

R development Core Team. 2006. R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria. ISBN: 3-900051-07-0, URL http://www.R-project.org.

Schneider, W., Eschman, A. & Zuccolotto, A. 2002. E-Prime reference guide. Pittsburgh: Psychology Software Tools Inc..

Tsay, Suhchuan Jane. 1994. Phonological Pitch. PhD dissertation. University of Arizona.

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. 2407-2410. Berkeley: University of California.

Tsay, J. & Myers, J. 1996. Taiwanese tone sandhi as allomorph selection. 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. 233-237. Peking: Tsinghua University.

Wang, Samuel H. 1995. The tone sandhi phenomenon. Experimental Studies in Taiwanese Phonology. Taipei: Crane.

Wang, William S.-Y. 1967. Phonological features of tone. International Journal of American Linguistics 33 (2). 93-105.

Xu, Yi. 2009. "_TomeNormalizedF0.praat. On-line: http://www.phon.ucl.ac.uk/home/yi/tools.html

Yip, Moira. 1980. The Tonal Phonology of Chinese. Cambridge, MA: MIT Dissertation.

Zhang, Jie, Yuwen Lai & Craig Turnbull-Sailor. 2006. Wug-testing the "tone circle" in Taiwanese. In Donald Baumer, David Montero, and Michael Scanlon (eds.), Proceedings of the 25th West Coast Conference on Formal Linguistics. Cascadilla Proceedings Project, Somerville, MA. 453-461.

Zhang, Jie, Yuwen Lai & Craig Turnbull-Sailor. 2010. Modeling Taiwanese speakers' knowledge of tone sandhi in reduplication. Lingua 121 (2). 181-206.