Chapter 3

Experiment One

3.1 Introduction

This chapter introduced a lexical decision experiment that was used to investigate lexical processing in conditions where tonal information was partially or completely manipulated.

In Experiment 1, trials of different tonal conditions are mixed and randomly presented to the subjects, so the presentations of real words randomly involve three tonal conditions. There are two hypotheses for tonal retrieval. First, if lexical tones are dispensable during lexical processing, it is predicted that the segments are able to activate the lexical level even when tonal information is insufficient. On the contrary, if tonal information is very important, lexical processing without tones is expected to be impeded. Accuracy is predicted to decrease and reaction times to increase as tonal information reduces. Besides, high frequency words are expected to have better performances than low frequency words.

3.2 Method.

3.2.1 Subjects

Twenty university students (9 females, 11 males) were paid to participate in the experiment. Their age was between 20 and 25. All of them were native speakers of Mandarin and no one reported any history of speech, hearing or hand disorders.

3.2.2 Material

The present study aimed to examine processing of bisyllabic words. Therefore, all the target words used in our experiment were bisyllabic compounds. Because Mandarin included four tones, there were 16 tonal patterns for the bisyllabic words. However, this study did not aim to examine processing of different tonal patterns. In order to counterbalance the tonal effects, we selected three or more pairs of bisyllabic words for each tonal pattern. To ensure that only one lexical representation was selected during lexical processing, two criteria were set for the selection of stimuli based on phonological and semantic reasons: (1) only one tonal pattern was allowed for the selected word; words that shared the same segmental sequences with them were excluded; (2) only one semantic meaning was permitted; therefore, words that had homophones were avoided; (3) each stimulus had a counterpart which shared similar acoustic features, including F0 contours, but different word frequency.

Two pretests were conducted to examine whether the selected stimuli met our criteria.

3.2.2.1 Tonal pattern pretest

Seven undergraduate students were recruited from National Taiwan Normal University and they were paid for participating in this experiment. They were all native speakers of Mandarin, and no one reported any history of speech or hearing disorders. A total of 125 pairs of high and low frequency words which shared acoustic similarities were examined (as seen in Appendix A). A female native speaker was asked to record these words through a Shure SM10A head-mounted microphone connected to a SONY DAT PCM-M1 digital audio recorder. The recording data were digitalized at a sampling rate of 44100 Hz via CoolEdit Pro 2.0. The stimuli were auditorily presented by the JVC LL9700 multimedia system through YAMAHA speakers in a quiet room. The subjects were instructed to listen to the stimuli and write down the Chinese characters on the answer sheet. The pretest lasted approximately for 25 minutes. Later on, the subjects were asked to work in groups to examine whether there were any other possibilities of tonal combinations or homophones for each word. Extra money would be paid for the subject if he/she found other possibilities of homophones or tonal combinations. Syllables which had homophones or more than one tonal pattern were excluded from our stimuli and 29 words were removed. To create equal number of phonetically similar word pairs, another 56 words (in Appendix B) were further examined by another 4 subjects and 42 words (in Appendix C) were examined by 3 subjects. The procedures were identical to the previous one, removing 13 words. In order to pair acoustically similar high and low frequency words, eventually only 73 counterparts of high and low frequency words were in our consideration.

3.2.2.2 Word frequency pretest

The bisyllabic words which passed the tonal pattern pretest were then tested in the word frequency pretest. Among the selected words, each word had a counterpart in acoustic features and they shared identical tonal patterns. However, they were distinctive in word frequency. Unfortunately, there is no word frequency dictionary or database done with these words. Therefore, this pretest was necessary to clarify the word frequency of the selected candidates of our stimuli.

50 native listeners came to the experiment in a group of six or fewer people. All of them were paid for their participation. No one reported any history of speech or hearing disorders. 73 pairs of phonetically similar words chosen from the tonal pattern pretest were controlled by the JVC LL9700 multimedia system through YAMAHA speakers in a quiet room. Words of the same acoustic-phonetic pair were broadcasted with a 1.3-second interval, and each different pair had a 4-second interval as the subjects' response times. On the answer sheet (see Appendix D), two columns were presented for each pair. Column A stood for the former word presented while Column B indicated the latter word. The subjects were instructed to choose the high frequency

words among the acoustic-phonetic similar counterparts on the answer sheet. Two versions of word order concerning word frequency were created. For example, half of the subjects would hear *nian*^{2}*xin*^{1} ‘yearly salary’ (low frequency) first followed by *nian*^{2}*qing*^{1} ‘young’ (high frequency) and half subjects would hear the reverse version. Every word only appeared once and the experiment lasted for 8 minutes.

After the two pretests, there were 37 pairs whose voting ratio of the high and low frequency words was 4:1 and 9 pairs in ratio 3:1 were included in the stimuli (see Appendix E). That is, the votes for the high frequency words were four times or three times of the low frequency words. A later concern on the Tone Sandhi effect further removed a pair of high and low frequency word. The Tone Sandhi effect was generally known that the bisyllabic pattern of T3-T3 sounded like T2-T3 pattern in spoken Mandarin. To avoid confusions caused by the Tone Sandhi effect, the pattern of T3-T3 was excluded from our stimuli. Besides, it should be noted that seven pairs of words were later found problematic and excluded in our data analysis. The six words (*jing*^{4}*sheng*^{1}, *lian*^{3}*dan*^{4}, *feng*^{1}*huo*^{3}, *yan*^{4}*hui*^{4}, *yi*^{1}*jia*^{4}, *and yu*^{4}*xi*^{2}) passed the pretests, yet subjects in the production task (which would be discussed later) found other possibilities of homophones or tonal patterns for these pairs. Eventually, 39 pairs of bisyllabic words which met our criteria were used. Unfortunately, the selected candidates of T4-T2 and T4-T3 patterns were not included in our experiment analysis.

Only 13 bitonal combinations were utilized in our experiment (T3-T3 was excluded in consideration of Tone Sandhi effect).

3.2.2.3 Recording and manipulation

In order to manipulate every syllable, each of the 45 pairs of bisyllabic words obtained from the tonal pattern pretest and word frequency pretest was then separated into two monosyllables before the recording. For example, *da³gong¹* 'to have part time job' was separated into *da³* 'to hit' and *gong¹* 'worker'. The monosyllables were then replaced by their T1 counterparts (high level tone) such as *da¹* 'to construct' and *gong¹* 'a worker' and these T1 monosyllables were included in the recording list. For those that did not have T1 counterparts such as *nian²*, the T1 syllable was presented by *Zhuyinfuhau* (a local phonetic system). A female native speaker was asked to record these T1 monosyllables. In order to control the reading speed, the syllables were randomized in the list and each syllable was presented to the speaker for 2 seconds through PowerPoint on the PC screen. Before the presentation of each monosyllable, a fixation in the center appeared for 1 second. The recorded T1 syllables would then be used as the base forms in syllable resynthesis.

Because the first phonemes of the syllables included nasals, stops, fricatives and affricates whose voice onset times (VOT) were obviously different. In order to make the resynthesized syllables sound more natural, we chose three syllables (liao, piao, and shang from the wordlist) that were existent and pronounceable Mandarin syllables in four tones and contained nasal, stop or fricative in the first phoneme. These syllables were presented in four Mandarin tones as well as a convex tone in *Zhuyinfuhau*. The speaker was asked to read these syllables in four Mandarin tones (e.g. *liao*[1] 'to hold up', *liao*[2] 'chat', *liao*[3] 'understand', *liao*[4] 'material') and a convex tone (e.g. *liao*[5]) which was a non-existent rising-falling tone in Mandarin system. The F0 contours extracted from the syllable, *piao,* would be extracted to resynthesize stop initial syllables, *shang* for fricative and affricate initial syllables and *liao* for the resting syllables. The convex tone was used for the non-tone condition in the present study. The five tones were illustrated in Figure 2.

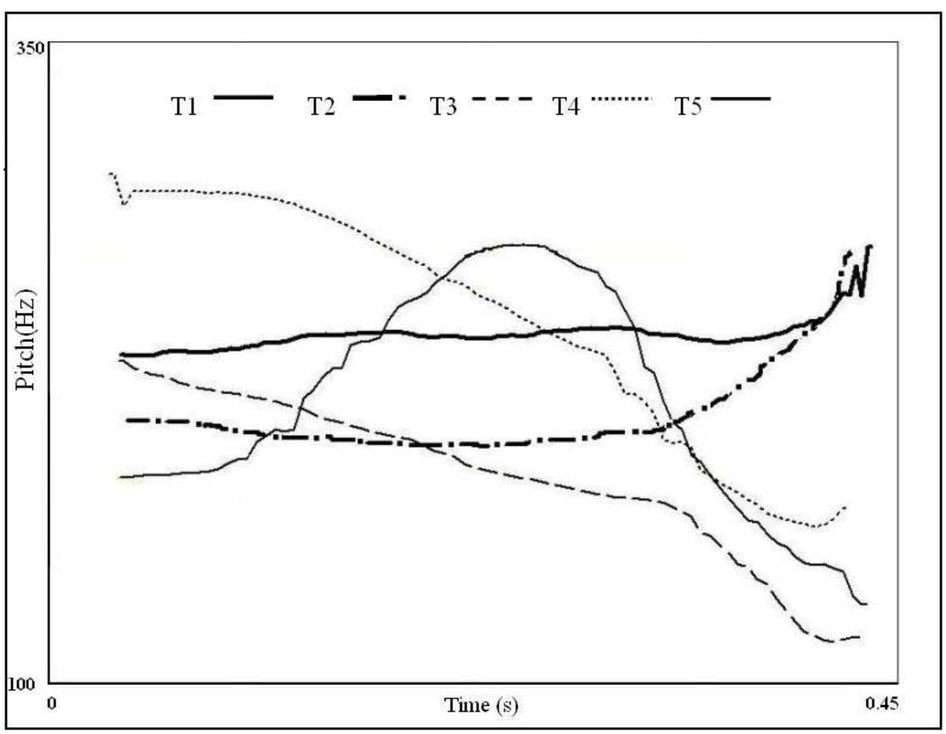

Figure 2 Five extracted F0 contours used in the syllable resynthesis.

We resynthesized every syllable with five tones consistently by Praat 4220 (2006). The resynthetic process was intended to eliminate tonal coarticulation on production which served as cues in language perception (Gottfried & Suiter, 1997; Xu, 1997). A pretest was then used to examine whether the native speakers were able to recognize these resynthesized syllables correctly. In the pretest, the resynthesized monosyllables of five tones were presented auditorily. Five subjects were asked to write down the *Zhuyinfuhao* and the tones of the stimuli. They were instructed to use the symbol (^) to represent the convex tone. The results showed that all the resynthesized syllables were well-identified.

The resynthesized syllables were then recombined to construct the bisyllabic compound words. During the reconstruction, the duration of every first syllable of the compound words was manipulated into 450 ms and the duration of the second syllables was manipulated into 550 ms. Besides, there was a 50-ms-pause inserted between the first and the second syllables, as shown in Figure 3. In this case, those words which had plosives in the second syllables would sound more natural. In any case, the total duration of each bisyllabic word would reach 1.050 seconds.

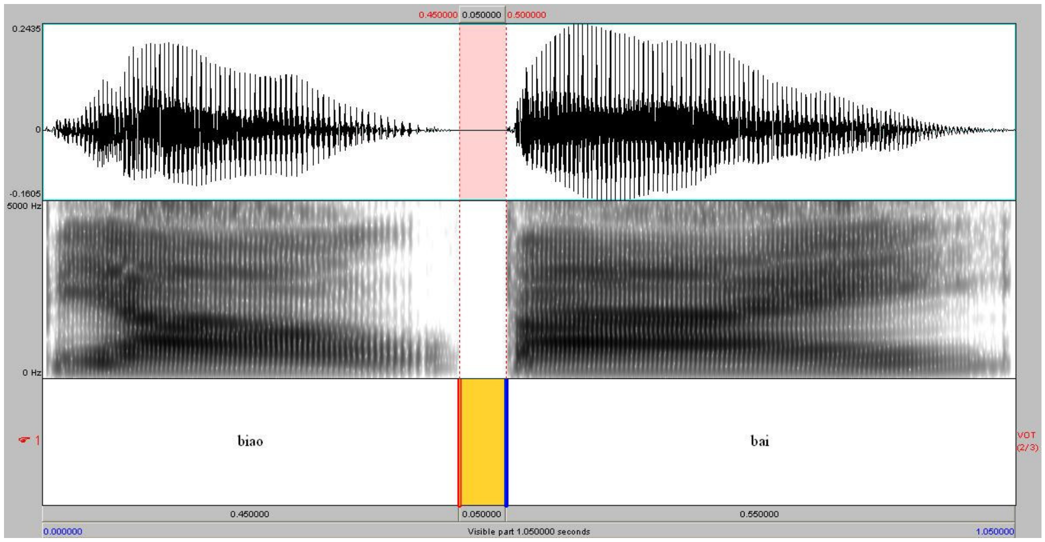

Figure 3 Spectrogram of the bisyllablic target word, *biao*^{3}*bai*^{2} ‘to bare one’s heart’.

3.2.2.4 Production pretest

The production task was designed to make sure that the manipulated auditory stimuli were able to elicit our intended bisyllabic words. Six native speakers recruited from university were paid for participating in the experiment. 90 target words were used. Stimuli of three manipulated tonal conditions were averaged for each subject. The subjects were tested individually in a quiet room. A writing test was used. All the stimuli were randomly presented at a comfortable level through SONY dynamic stereo headphones MDR-7502. The subjects were instructed to listen to the auditory stimuli and write down the target words on the answer sheet. Before the experimental block, there was a practice block containing 12 practice trials. Besides, a 400-ms beep and a 400-ms silence were presented prior to each target word. The response times were 8 seconds for each trial and the experiment lasted for 8 minutes.

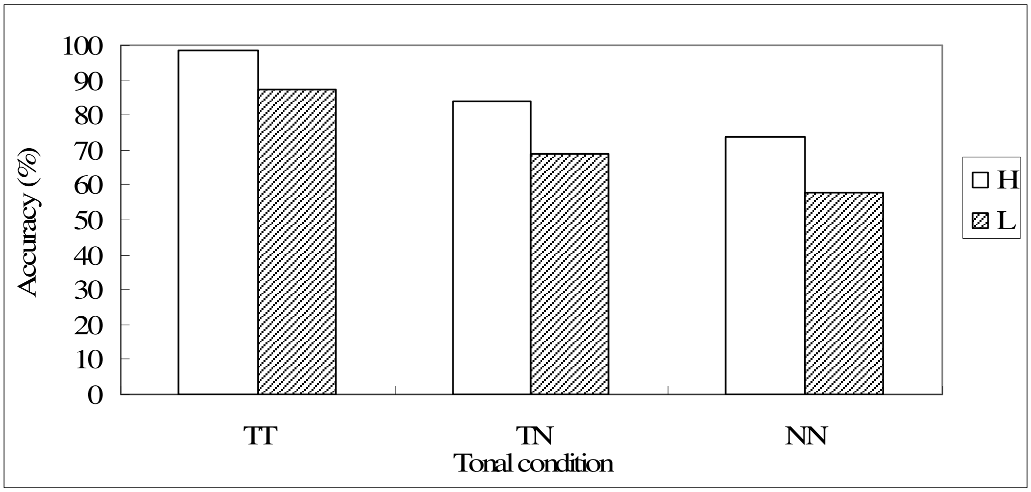

Figure 4 Accuracy of responses to high (H) and low frequency (L) words in different tonal conditions (TT: tone-tone condition, TN: tone-nontone condition, and NN : nontone-nontone condition) in the production task.

Figure 4 displayed the accuracy of the production experiment. The averaged percentages of high (H) and low (L) frequency words in three tonal conditions were 98.6 % (H_TT), 84.1 % (H_TN), 73.9 % (H_NN), 87.3% (L_TT), 68.8 % (L_TN), and 57.9 % (LNN) respectively. The accurate percentage increased as word frequency and tonal information increased. The result showed that the subjects were able to write down the intended target words. However, six pairs of words were found problematic in the production task. The subjects found homophones or counterparts that shared the same syllables with the target words. These pairs were then excluded.

3.2.2.5 Stimuli

39 pairs of resynthesized bisyllabic words were used as the stimuli (see Appendix G). Each pair contained high and low frequency counterparts that shared

acoustic similarities (see Appendix D). The high (H) and low (L) frequency counterparts may have (a) identical initial syllable such as *nian²qing¹* 'young' of high frequency and *nian²xin¹* 'yearly salary' of low frequency, (b) identical final syllable such as *kong¹jian¹* 'space' of high frequency and *gong¹jian¹* 'storm fortifications' of low frequency, or (3) at least the same vowels such as *qing¹song¹* 'relaxed' of high frequency and *xing¹kong¹* 'starry sky' of low frequency. Besides, three tonal conditions were set for each stimulus: (1) tone-tone condition (TT): correct tonal information was given to the bisyllables; (2) tone- nontone condition (TN): the first syllable with the intact tonal contour whereas the second one was replaced by a convex tone; the syllables with the convex tone were defined as nontone syllables in the present study; (3) nontone-nontone condition (NN): the original tones of both syllables were substituted by convex tones. Besides, 78 nonwords were added as fillers. These nonwords were created from two existent Mandarin syllables but the combination did not form any meaningful words. For instance, *jiang¹* 'thick liquid' and *zheng¹* 'compete' were combined to form *jiang¹zheng¹* which was not a meaningful word in Mandarin.

3.2.2.6 Apparatus

Recording were done with a Shure SM10A head-mounted microphone connected to a SONY DAT PCM-M1 digital audio recorder, using Maxell DM120X DAT tapes.

The recording data were then digitalized at a sampling rate of 48000 Hz and was later downsampled to 44100 Hz through CoolEdit Pro 2.0. Manipulation of tone and duration was done by Praat 4220 (2006).

During the lexical decision task, all the stimuli were controlled by E-Prime 1.1 program. SONY dynamic stereo headphones MDR-7502 were used. Reaction times were recorded by a SRB200-03248 Serial Reaction Box connected to the PC.

3.2.2.7 Procedure

Subjects were tested individually in a quiet room. All the auditory stimuli were presented at a comfortable level through the headphones. One practice block containing twelve practice trials was presented to each subject before the experiment (see Appendix E). All the stimuli were randomly presented in a different order for each subject, and no stimuli from the practice block were used in the experimental blocks. The subjects were divided into three groups. Each subject received all the three experimental conditions and each stimulus appeared in each of its three tonal conditions across different groups. For example, $nian^{2}qing^{1}$ 'young' contributed TT pattern to one group, but was presented in TN pattern to another group, and NN pattern to still another group. Before the presence of each target word, a beep lasting for 400 ms was presented, followed by a 400 ms silence. The subjects were asked to decide as rapidly and accurately as possible whether the test item was a real word or a nonword by pressing the buttons. Across subjects, dominant and nondominant hands were balanced across word and nonword responses. Besides, the response window for each trial was 2800 ms. No matter what decisions were made, each trial terminated after the response window. The entire experiment lasted for 28 minutes.

3.3 Results

3.3.1 Accuracy

Figure 5 displayed the accuracy conducted on the tonal conditions and word frequency. The average percentages of high (H) and low (L) frequency words across different tonal conditions were 93% (H_TT), 79% (H_TN), 69% (H_NN), 81% (L_TT), 71% (L_TN) and 52% (L_NN) respectively. High frequency words had higher accuracy percentage than low frequency words. The lexical mapping of tonal words onto the lexicon was more correct while syllables containing non-tone words were more difficult to process, especially in low frequency words.

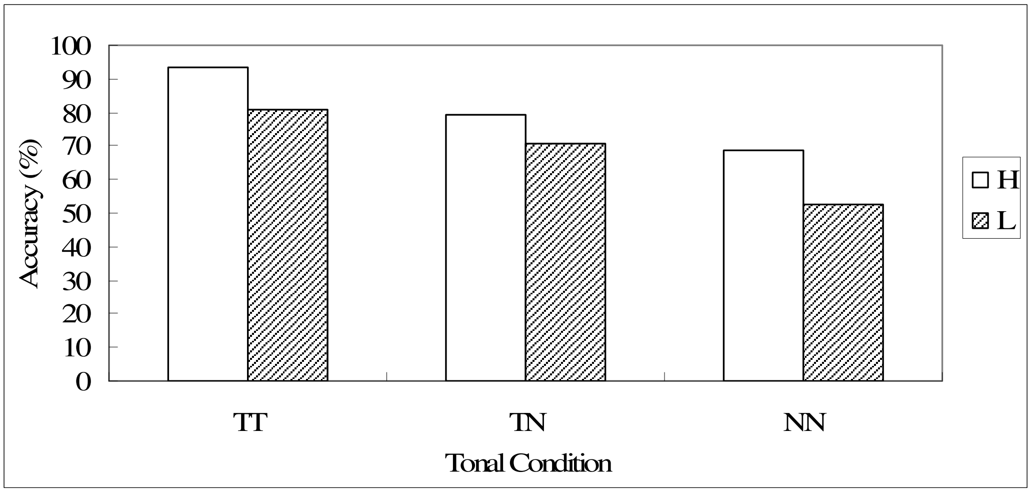

Figure 5 Accuracy of responses to high (H) and low frequency (L) words in different tonal conditions (TT: tone-tone condition, TN: tone-nontone condition, and NN : nontone-nontone condition).

A Tonal condition (3) x Word frequency (2) between-subjects two-way ANOVA was conducted to examine the tonal effect and word frequency effect on accuracy. The percentage data were arcsine transformed before being statistically analyzed using ANOVA. The formula was 2*ASIN (SQRT(percentage)) (Rietveld & Hout, 1993). The result showed that both word frequency and tonal effects were significant [Tonal condition, F (2, 114) = 33.098, $p < .001$; Word frequency, F (1, 114) = 26.464, $p < .001$]. No interaction was found. Multiple comparisons indicated that all three tonal conditions showed significant differences ($p < .01$).

3.3.2 RT

The reaction times were calculated from the target offset to the response onset. Only correct responses were included in the analysis. Reaction times were calculated from the target onset to the response onset. Responses that were (1) less than 200 ms, (2) greater than 2800 ms or (3) beyond three standard deviations were excluded. The percentage of rejected responses was 25.9 %, including 13% incorrect responses from the NN condition, 8.3% from the TN condition, and 4.4% from the TT condition. The relatively high percentage of the excluded items was resulted from the NN condition.

Figure 6 presented the mean reaction times of high (H) and low (L) frequency words across tonal conditions. The averaged RTs were 274 ms (H_TT), 339 ms (H_TN), 365 ms (H_NN), 438 ms (L_TT), 498 ms (L_TN), and 483 ms (L_NN) respectively. The high frequency words took much shorter reaction times compared with the low frequency words. Across conditions, the tone-tone (TT) condition had the best performances on reaction times. The tone-nontone words of the same frequency had approximate reaction times to the nontone-nontone (NN) condition.

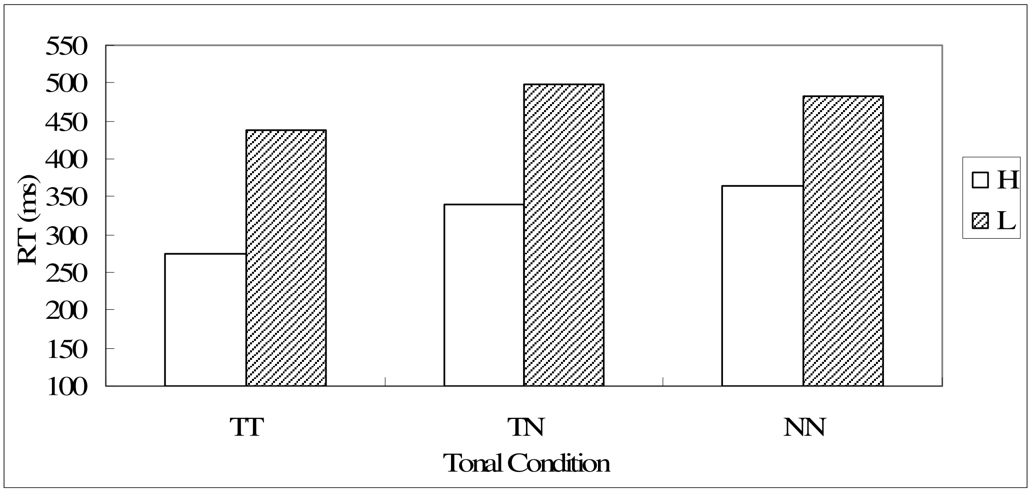

Figure 6 Mean reaction times as a function of word frequency and tonal effects of high (H) and low (L) frequency words across tonal conditions.

A Tonal condition (3) x Word frequency (2) two-way ANOVA was run to test the main effects of tonal effects and word frequency effects. The statistic data showed that both the tonal conditions and word frequency had significant influences on the reaction times [Tonal condition, F (2, 1145) = 7.851, $p < .01$; Word frequency, F (1, 1145) = 82.658, $p < .01$], but there was no interaction found between the two factors. Pair comparisons conducted on the reaction times was displayed to examine the significant differences across conditions. Comparisons of the TT-TN conditions and the TT-NN conditions reported significant tonal effects, but the significant differences were not found in the TN-NN pair on reaction times.

3.4 Discussion

The results showed that word frequency had significant effects across conditions. High frequency words had better performances than low frequency words. The reaction times of high frequency words in NN condition (365 ms) were even shorter than low frequency words in TT condition(438 ms), indicating that high frequency words were retrieved faster than low frequency words. However, the accuracy decreased as tonal information decreased despite of word frequency. The gradually decreased accuracy across different tonal conditions indicated that tonal information was important for lexical processing. If tonal information was manipulated, lexical processing was impeded, especially for low frequency words. For high frequency words, although the accuracy dropped slightly, the accuracy was still relatively high.

Regarding the reaction times, listeners needed more time to deal with the TN words and the NN words compared with the TT condition. Specifically, a main effect was found for tonal conditions, but multiple comparisons conducted on the reaction times showed that the significant effect was mainly performed between the TT condition and the NN condition. In our assumption, first intact syllables of the TN words were expected to facilitate lexical processing and tonal retrieval of the second syllables. As our expectation, the reaction times of the high frequency words in the TN condition were shorter than that in the NN condition. Specifically, paired comparisons conducted on the reaction times showed that there was no significant difference between the TN and the NN conditions.

Besides, it should be noted that the accuracy of the high frequency words in tone-tone (TT) condition did not reach 100 % correctness, suggesting the possible unnaturalness due to resynthesized sounds. In previous research, it was found that subjects could use the secondary cues such as duration, amplitude, or coarticulation during the lexical perception. There is one possibility that native listeners utilize the acoustic information all together in normal speech. Therefore, when we removed parts of the acoustic features in our resynthesized words, the listeners detected the unnaturalness and treat some real words as nonwords, causing a gap of 7 % to reach 100 %.