Judging Homophony in Chinese: The Influence of Tones

Marcus Taft
University of New South Wales

Hsuan-Chih Chen
The Chinese University of Hong Kong

When making judgements about whether one character is homophonic with another, both Mandarin and Cantonese speakers encounter difficulties in saying "no" when the characters have the same phonemes but different tones, regardless of whether the task is performed silently or aloud. These difficulties were demonstrated using two different experimental paradigms. Mandarin speakers were poorer than Cantonese speakers at deciding whether two presented characters were homophonic, and compared to previous data collected with English materials, were extremely poor at deciding that a presented character was not homophonic with any other character. Theoretical consideration was therefore given to the difficulty observed when tones differed, and to the difference in performance between Mandarin and Cantonese speakers, as well as to the general difficulty observed in the homophone-generation task.

The impetus for the research to be described in this chapter came from an experience that the first author had when lecturing in China. I (the first author) presented a seminar on the topic of lexical processing in Chinese during which I described an experiment that I had carried out many
years ago using Chinese (Mandarin) speakers and materials. Subjects were required to discriminate real characters from non-characters by pressing a "yes" or "no" button to individually presented stimuli. The real characters were of two types: Those which were pronounced identically to another character (i.e., homophones, like '金') and those which had a unique pronunciation (i.e., non-homophones, like '晃'). Even though I am not at all proficient in the Chinese language, I was able to construct what I thought was a well-designed set of materials using a Chinese dictionary and the assistance of a Chinese-speaking friend. It came as a bit of a shock, therefore, that when I presented examples of my non-homophones during the seminar, many members of the audience immediately called out that these were actually homophones. For example, they said that '晃', which I had classified as being a non-homophone, was actually pronounced in the same way as '黄'.

While I stood there disconcerted, however, the audience began to waver on their claim. They qualified what they had said by acknowledging that such characters were only homophonic if one were to ignore the tone. Now, this seemed to me to be a very odd thing to say, given what I understood about Chinese. Why would one ignore the tone when deciding how a character is pronounced? One of my constant problems when attempting to speak the occasional Chinese phrase has been that I fail to produce the correct tone and therefore the listener does not understand what I am saying. From this I had assumed that the tone was as crucial to the identity of the syllable as were the individual segments (i.e., phonemes). But now I was finding that the tonal structure appeared to be relatively unimportant in identifying the pronunciation of the character. Why should this be?

This experience in China was not the first occasion where I had observed difficulty in judging that two utterances were homophonic. Some years ago, I reported an experiment using English stimuli whereby subjects were presented with a series of words and asked to say whether there existed another word which was pronounced identically to each (Taft, 1984). For half of the items for which a "yes" response was correct, the homophonic word had a different morphological structure to the presented word, and for half it did not. For example, the word GOLD is a single morpheme while the homophonic word GOALED comprises two morphemes (GOAL and ED); whereas both the word HEAL and its homophone HEEL are single morphemes. What I found was that, when the task was performed silently, there were a considerable number of errors in responding to the
morphologically different homophones (i.e., subjects said "no") as compared to the morphologically congruent homophones. On the other hand, when the subjects could read the words aloud before making their decision, the morphologically different homophones were handled just as well as the morphologically congruent homophones.

The way I interpreted these results was to suggest that the silent task was tapping into a more abstract representation of the word's pronunciation than was the vocalized task. The abstract representation of a polymorphemic word is structured differently to that of a monomorphemic word (e.g., GOALED is represented as something like [gol * past tense], while GOLD is represented as something like [gold]), and therefore, at this abstract level, the two morphologically different homophones do not correspond while the two morphemically congruent homophones do (e.g., both HEAL and HEEL would be represented as something like [hil]).

To draw a parallel to the situation with Chinese tones, one could suggest that in the abstract representation of the Chinese syllable, the tone is treated separately from the segmental information (i.e., the phonemic information), and is only integrated with the segmental information when the syllable is converted into a form that is overtly vocalized. If this were so, the tone would play an important role when attempting to identify a spoken character, but might be ignored when making a judgement about the pronunciation of an unvocalized character. Such a view accords with linguistic theories of the abstract representation of Chinese syllables, whereby segmental information is kept quite distinct from tonal information (e.g., Cheng, 1966; Wang, 1967). In Cheng's account of Mandarin, for example, the underlying representation of a syllable is separable into segments and tone with independent rules being applied to each in order to generate the surface form. In fact, the tone that is actually produced is seen as being generated from an underlying tone. In particular, Cheng claims that the 2nd (rising) tone is the same as the 1st (level) tone in the underlying representation, but the addition of a "dynamic" feature causes it to emerge as a different tone in the surface pronunciation. Similarly, the 3rd (dipping) tone is supposedly generated from the 4th (falling) tone by the addition of this dynamic feature.

The aim of the experiments to be reported in this chapter was to empirically demonstrate what I had anecdotally observed, namely, that when asked to make judgements about homophony in Chinese, subjects would be quite insensitive to tonal differences. Further, it was expected that if this insensitivity arose at the abstract phonological level, it would be manifested
more strongly when the syllables which were to be judged were spoken silently than when spoken aloud.

**Experiment 1**

In the first experiment, the homophone judgement was made on two characters presented sequentially. There were 90 such pairs. Half of the time the two characters were pronounced identically in Mandarin (e.g., '保' and '饱', both pronounced bāo) and half of the time they were not. The non-homophonic pairs were the items of interest and fell into three conditions. Either the two members of the pair had the same segments, but different tone (e.g., '曲' and '去', i.e., qū and qù respectively), the same tone, but a different vowel (e.g., '气' and '去', i.e., qì and qù), or different segments and tone (e.g., '年' and '去', i.e., nián and qù). There were three lists of items constructed such that no list repeated the same character, while each list contained 15 non-homophonic pairs in each condition. So, for example, List 1 contained 曲去, List 2 气去, and List 3 年去. Characters were presented in their simplified form as used in the People’s Republic of China, and the characters in the three different conditions were approximately matched on their complexity in terms of number of strokes and their frequency of occurrence as determined from a compilation of Chinese word frequency norms (National Institute for Compilation and Translation, 1967). Ten practice item pairs were given to subjects at the beginning of the experiment.

Subjects were divided into 6 groups. Three groups performed the task silently while the other three were instructed to read the characters aloud before making their response. There were three groups for each task condition in order that each of the three lists could be presented to the same number of subjects. There were 6 subjects in each of the three "silent" groups, and 5 subjects in each of the three "aloud" groups. All subjects were students from the People’s Republic of China aged in their 20’s and studying at the University of Arizona.

The characters were presented via computer on a T.V. monitor. The first member of a pair appeared for 1 sec on its own, and remained on the screen when the second member of the pair appeared next to it. After a further 1 sec the two characters were removed. By pressing a foot pedal, subjects were able to bring up the next pair of characters. Subjects were asked to decide whether or not the pairs of characters were pronounced
identically in Mandarin, by pressing a "yes" or "no" button as quickly but as accurately as possible. Response latencies were measured from the onset of the second member of the pair. The reaction times for correct responses were analyzed, as well as the error rates.

It was predicted that when the task was performed silently, the subjects' ability to say "no" to the non-homophones would be poorer when the characters differed in tone than when they differed in segmental structure, but that this difference would be reduced when the characters were spoken aloud.

### Table 1
Reaction times for correct responses (RT in msec) and percentage error rates (%E) for the pairs of characters in Experiment 1 (Mandarin).

<table>
<thead>
<tr>
<th>Condition</th>
<th>Example</th>
<th>Silent RT</th>
<th>%E</th>
<th>Aloud RT</th>
<th>%E</th>
</tr>
</thead>
<tbody>
<tr>
<td>Different Tone</td>
<td>曲去 qū qū</td>
<td>1115</td>
<td>27.8</td>
<td>1057</td>
<td>43.1</td>
</tr>
<tr>
<td>Different Vowel</td>
<td>氣去 qì qǜ</td>
<td>947</td>
<td>5.2</td>
<td>967</td>
<td>9.3</td>
</tr>
<tr>
<td>Control</td>
<td>年去 nián qǜ</td>
<td>896</td>
<td>4.0</td>
<td>922</td>
<td>7.1</td>
</tr>
<tr>
<td>Homophones</td>
<td>保飽 bǎo bǎo</td>
<td>917</td>
<td>16.9</td>
<td>899</td>
<td>15.8</td>
</tr>
</tbody>
</table>

The mean reaction times and error rates are presented in Table 1. It can be seen that the tone different items did cause problems, but this difficulty was not systematically greater in the silent task compared to the aloud task. While there was a tendency toward such a reduction in difficulty on reaction times, the reverse was true for errors.

The difference between the Different Tone condition and the Different Vowel condition was significant on both reaction times and errors for both the silent and the aloud task, with all F's significant at $p < .02$, as was the difference between the Different Tone and Control conditions. The difference between the Different Vowel condition and the Control was not quite so reliable: While the reaction time difference was highly significant
in the silent task, \( p < .001 \) for both the subject and item analyses, it was
significant by subjects but not by items in the aloud task, \( F(1,12)=7.23, \ p < .02; F(1,42) = 3.34, \ p > .05 \). In addition, the error difference was far
from significant in both the silent and the aloud tasks, all \( F \)'s at \( p > .1 \).
There were no significant interactions between conditions and task, \( p > .1 \)
in every case.

It seems then that tones do tend to be ignored when judgements are
made about the pronunciation of characters. The fact that this is the case
even when the character is spoken aloud, might suggest that the explanation
has nothing to do with abstract phonological representations. Even if it does
(see later discussion for further consideration of this question), there is no
support for the particular account of abstract phonological representations
given by Cheng (1966). According to Cheng, the underlying forms of the
1st and 2nd tones are identical apart from one feature, as are the 3rd and
4th tones. Yet there was no sign in the present experiment of greater
confusion between these tones and any others. There was a reaction time
difference of 144 msec and an error difference of 22.1% between the
Different Tone and Different Vowel conditions in the silent task for the 9
cases where the tone difference was between the 1st and 2nd tone or the 3rd
and 4th tone. These differences were of the same (or if anything, smaller)
magnitude as the remaining 36 cases, where the differences were 174 msec
and 22.7%.

In attempting to explain why tonal information is frequently ignored
in making homophone judgements, certain artifactual accounts must be
addressed. First, it is likely that two syllables which differ only in tone are
acoustically (or phonetically) more similar than two syllables which differ
by a single segment. If so, it could be argued that there is no special
difficulty with tones per se, but rather, that the more acoustically similar
two sounds are, the more readily they will be confused. The argument
against this, however, is that the members of the Different Vowel pairs
were much more acoustically similar to each other than those of the Control
pairs, yet the confusion between them was not dramatically greater, and not
even significantly so when looking at errors. Thus, it would be hard to
sustain the argument that the much greater difficulty observed in the
Different Tone condition than the Different Vowel condition, particularly
in terms of errors, was purely a result of greater acoustic similarity.

A second argument that could be mounted against the present
experiment is that the subjects came from a variety of regions in China. As
a result, a number of the subjects did not speak Mandarin as their primary
Judging Homophony
dialect, and may have spoken it with a non-standard accent. Perhaps this fact could explain the difficulties observed with the Different Tone items, since some of the pairs used in this condition may actually be pronounced the same in some dialects. It is unlikely, however, that this could provide a complete explanation for the results. Differences in pronunciation would be logically more likely to affect responses to homophones than to non-homophones. It seems more likely that two syllables that are normally pronounced the same might differ in another dialect, rather than the other way around. If two syllables which are pronounced differently in standard Mandarin turn out to be pronounced the same in a particular dialect, this would be pure coincidence and certainly infrequent, unless that dialect systematically collapsed together two different sounds (e.g., 1st and 2nd tone). However, the difficulty with tone differences observed in the present experiment generalized statistically across both subjects and items, and this suggests that the majority of subjects and items showed the same pattern (despite the fact that a variety of tone contrasts were used).

Nevertheless, the fact that subjects spoke in a variety of accents makes one somewhat wary of drawing strong conclusions from this experiment. For this reason, and also to explore the generalizability of the results beyond Mandarin, we decided to repeat the experiment in Cantonese, where all of the subjects spoke with the same accent.

Experiment 2

The experiment run with Cantonese materials and native speakers was basically of the same design as the experiment run with Mandarin materials. There was, however, an additional non-homophone condition. Rather than just having a Different Vowel condition to compare to the Different Tone condition, a Different Consonant condition was included. In this condition, the two characters of a pair were pronounced with the same tone and vowel, but differed on their initial consonant. The only other change to the design of the materials was, of course, that homophony was determined on the basis of Cantonese pronunciation.

Stimuli were 36 homophonic pairs and 36 non-homophonic pairs. There were four types of non-homophonic pairs: Different Tone pairs (e.g., ‘告’ and ‘高’, i.e., /go/ and /go/), Different Vowel pairs (e.g., ‘家’ and ‘高’, i.e., /ga/ and /go/), Different Consonant pairs (e.g., ‘租’ and ‘高’, i.e., /jo/ and /go/), and Control pairs (e.g., ‘魚’ and ‘高’, i.e.,
M. Taft and H. C. Chen

/\textipa{yue}/ and /\textipa{go}/. All stimuli were presented in their traditional form as used in Hong Kong, and the four different types of pairs were matched as closely as possible on their complexity and their frequency of occurrence according to a Hong Kong Chinese word frequency count (Educational Research Establishment, 1986).

Forty-eight native Cantonese speakers (with a mean age of about 20) drawn from the introductory psychology subject pool at the Chinese University of Hong Kong participated in the experiment. Half of them performed the silent task and the other half the aloud task. Each subject was presented with the 36 homophonous pairs and nine of each of the four types of non-homophonic pairs.

The stimuli were presented on the screen of an IBM PC/AT compatible computer. At the beginning of the experiment, instructions and four practice trials were given. Each trial began with the presentation of a star signal for 500 msec in the centre of the computer screen followed immediately in the same position by the display of the first character of a pair. After another 500 msec the second character of the pair appeared next to the first character for 2 sec. Then the two characters were covered by a masking field. The subjects were instructed to press the space bar on the keyboard to start a trial and to respond by pressing a "yes" or "no" key.

The results can be found in Table 2. In terms of reaction time, the conclusions that can be drawn from these data are the same as those arising from Experiment 1. The Different Tone condition took longer to respond to than the other non-homophone conditions in both the silent task, $F(1,22)=14.06, p<.01$, $F(1,35)=24.36, p<.001$, and the aloud task, $F(1,22)=8.82, p<.01$, $F(1,35)=5.08, p<.05$. While there was a tendency for the size of this effect to be reduced in the aloud task compared to the silent task, the interaction between task and conditions failed to reach significance, all $F$'s at $p>.1$. Unlike the Mandarin experiment, the Different Vowel condition did not statistically differ from the Control condition, and neither did the Different Consonant condition, all $F$'s at $p>.1$. This lack of difficulty with the pairs which differed by one segment accentuates the point that the primary difficulty in judging homophony comes from judging tone differences. It seems, therefore, that in Cantonese as well as in Mandarin, information about tones is not of primary importance when determining the pronunciation of a character.

Interestingly, the error data are very different to those of Experiment 1. The difficulty in judging tone differences does not manifest itself in the error rates, though there is a non-significant tendency for the effect to be
Table 2

Reaction times for correct responses (RT in msec) and percentage error rates (%E) for the pairs of characters in Experiment 2 (Cantonese).

<table>
<thead>
<tr>
<th>Condition</th>
<th>Example</th>
<th>Silent</th>
<th></th>
<th></th>
<th></th>
<th>Aloud</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>RT</td>
<td>%E</td>
<td>RT</td>
<td>%E</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Different Tone</td>
<td>/go₁/ /go₁/</td>
<td>910</td>
<td>2.3</td>
<td>1103</td>
<td>4.6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Different Vowel</td>
<td>/ga₁/ /go₁/</td>
<td>816</td>
<td>2.8</td>
<td>1023</td>
<td>2.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Different Consonant</td>
<td>/jo₁/ /go₁/</td>
<td>808</td>
<td>1.4</td>
<td>1050</td>
<td>2.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Control</td>
<td>/yue₁/ /go₁/</td>
<td>794</td>
<td>1.9</td>
<td>1022</td>
<td>2.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Homophones</td>
<td>/sik₃/ /sik₃/</td>
<td>870</td>
<td>6.0</td>
<td>1021</td>
<td>3.8</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

greater in the aloud task than the silent task (a trend also observed in the Mandarin experiment), all F's at $p > .1$. Not only do the Different Tone pairs present little difficulty for the Cantonese readers as far as accuracy goes, but the error rates overall are much lower than for the Mandarin readers. One can only speculate about why this might be the case.

Firstly, it might have something to do with the structure of Cantonese compared to the structure of Mandarin. The fact that there are nine tones in Cantonese as opposed to four in Mandarin means that homophony is less frequent in Cantonese. As a result, homophones may be specifically noted in lexical memory in some way, just as English speakers may have noted the homophony between THEIR and THERE, for example. However, this would not explain why the problem with the Different Tone condition was reflected in reaction times and not in errors. In fact, one might have predicted that Cantonese would be even harder than Mandarin since there are more tones to get confused between.

A second possible explanation lies in the way in which reading is taught in Hong Kong compared to China. In China, an alphabetic system of writing (pinyin) is typically used at the early stages of learning to read, while in Hong Kong the children are only taught character-to-pronunciation mappings, without the mediation of an alphabetic system. Whether this fact could have led to the pattern of differences in accuracy observed between
the two groups is something that may be worth pursuing (perhaps by testing Cantonese speakers from China). Since the pinyin system distinguishes the segments from the tones, it is possible that readers from China are more likely to distinguish segments and tones in their lexical representations than readers from Hong Kong. However, this would not explain why there was a tone effect in the Cantonese experiment on reaction times, even though not on errors, and neither would it explain the overall lower error rate compared to the Mandarin experiment. Moreover, adult readers from China have had little or no need for pinyin after childhood, and most would claim that it plays no role in their thinking about characters. Nevertheless, it would be interesting to ascertain whether there actually is an unconscious influence of these childhood practices on adult lexical representations.

When it comes to comparing homophone decision performance in Chinese with that in English, there is little that can be concluded in relation to Experiments 1 and 2. There are not to our knowledge any studies reporting a homophone decision task on pairs of English words (like SAIL and SALE). Previous experiments using the homophone decision task have typically used single nonwords as the stimuli, like BRANE and RIST (e.g., Taft, 1982; McCann, Besner, & Davelaar, 1988). However, as mentioned earlier, Taft (1984) did report a study using single English words as items. For example, subjects were asked whether there is another word that is pronounced identically to SAIL. On finding a different pattern of results when the words were spoken silently compared to when they were spoken aloud (depending on the morphological structure of the words), Taft concluded that the silent task was tapping a more abstract level of lexical representation than the aloud task. In the Mandarin and Cantonese experiments reported here, however, there was no difference in the pattern of responses observed between the silent and the aloud task. There are several interpretations that can be given to this result.

First, it may be the case that tones are represented in the same way in both the abstract and the surface representation of a syllable in Chinese. Alternatively, it may be the case that there are no such things as abstract representations in Chinese, or that if there are, they are not tapped into when reading silently. Note that these explanations have difficulty explaining why tone differences are problematical, since they suggest that homophone judgements are being made on the basis of a representation that is equivalent to the surface phonetic representation of the syllable, and tones should be an integral part of this phonetic representation. Another explanation that can be offered is that, even in the aloud task, responses are
made at an abstract level where tonal information is imperfectly represented. If so, this would mean that the abstract level of phonological representation has more impact for Chinese readers than English readers since the aloud task seems to be performed by the latter at the surface level.

Another possible explanation for the difficulty that Chinese subjects encountered in both the aloud and silent tasks with the Different Tone condition, is that the subjects simply mispronounced some of the characters, and that this happened to be particularly so for the Different Tone items. The possibility that mispronunciations had an influence on the results was examined in the Cantonese study, where the pronunciations given by subjects were monitored by asking them to read out the full list of characters after they had performed the homophone decision task. It turned out that errors of pronunciation were reasonably common (12.2% of responses), but that the least number of errors were made with the Different Tone condition items (5.6%). There were 319% errors of pronunciation amongst the Different Consonant, Control, and Homophone items, and 8.3% errors amongst the Different Vowel items. It seems, therefore, that there was no special difficulty in pronouncing the Different Tone pairs, and in fact, the relative lack of difficulty with these items compared to the other non-homophonic pairs could possibly have counteracted the tone effect on error rates. In point of fact, an explanation in terms of inappropriate pronunciations is really only relevant to the examination of error rates, and not to the differences observed in reaction times. If a character is not pronounced correctly, it could lead to an error in the judgement of homophony (depending upon the mispronunciation), but should have no impact on the time taken to make a correct response.

Perhaps the difference in experimental paradigms accounts for the contrast between the aloud/silent difference in English and the lack of an aloud/silent difference in Chinese. In Experiments 1 and 2, subjects were asked to compare the pronunciations of two presented characters. In the English study (Taft, 1984), subjects were presented only one word and asked if it was possible to generate another word with the same pronunciation. Since it is conceivable that these two paradigms tap into somewhat different processes, we decided to try out the latter paradigm with Mandarin readers to see if problems with tone differences still emerged, and also if the pattern was still the same regardless of whether the syllables were read silently or aloud. Furthermore, since the conclusions drawn from the English study were based on incorrect responses given to homophones rather than to non-homophones, Experiment 3 included a manipulation of homophones in addition to the manipulation of non-homophones.
Experiment 3

In Experiment 3, native Mandarin speakers were presented with single characters and asked to respond by button press whether or not there existed in their vocabulary another character pronounced identically to the presented character. Half of the items were homophones for which a "yes" response was correct, and the other half were non-homophones for which a "no" response was correct. There were two types of non-homophones: A Competing Tone condition, where there existed at least one other character pronounced with the same segments as the presented character, but with a different tone (e.g., '肉' which is the only character pronounced ròu, though '柔' is pronounced róu), and a No Competing Tone condition, where there were no other characters pronounced with the same segments as the presented character (e.g., '日' being the only character pronounced diǎ, while no characters are pronounced diú, diū, or diù). The two conditions were approximately matched overall on number of strokes and frequency of occurrence according to the "Modern Chinese Frequency Dictionary".

Similarly there were two types of homophones: A Competing Tone condition, where there was more than one character pronounced with the same segments as the presented character, both with the same tone and different tones (e.g., '洪' pronounced hóng, while '红' is also pronounced hóng, but '江' is pronounced jiāng), and a No Competing Tone condition where there was more than one character pronounced with the same segments as the presented character, but only with the same tone (e.g., '若' is pronounced ruò, as is '弱', but there are no characters pronounced ruó, ruó or ruò). The two conditions were approximately matched overall on number of strokes and frequency of occurrence according to the "Modern Chinese Frequency Dictionary".

Items were presented on a video display unit under computer control. Each item was displayed for 2 sec, and because the results from a pilot subject suggested that the task was very difficult, an inter-stimulus interval of 8 sec was used. Subjects were instructed to decide whether or not the presented character was pronounced identically in Mandarin to any other character. Response was to be made by button press. The same subjects were used in the silent and the aloud task, with the former always being conducted before the latter. That is, in the first phase the subjects were told not to read the character out aloud, while in the second, they were told that now they should do so and that they should listen to themselves in order to
make their decision. By using the same subjects, it was anticipated that performance in the aloud task would be better than performance in the silent task, but the question of interest was whether this would be equally true across all four conditions.

Fifteen subjects participated in the study. All were graduate students from the People's Republic of China studying at the University of New South Wales. Only 5 of the subjects were not from the north part of China.

The results are presented in Table 3. Turning first to the non-homophones, there appears to be a delay in response times when there exists a character with the same segments as the presented character, but with different tone. However, the main effect of tone competition was only significant on the subject analysis, \( F(1,14) = 6.34, \, p < .05, \, F(1,48) = 0.17, \, p > .1 \). On the other hand, the analysis of errors did prove to be significant, \( F(1,14) = 19.42, \, p < .001, \, F(1,48) = 6.78, \, p < .02 \). There was no significant interaction between tone competition and task, for either response times or error scores.

It can be seen then that the task of thinking of other characters with the same pronunciation as the presented character produces similar results to the task of comparing the pronunciation of two presented characters, particularly in terms of the errors made. Tonal information appears to be less important than segmental information in making judgements about pronunciation, and it does not make any difference whether the judgement is made on a silently read character or on a spoken character. That is, in making a decision about the homophony of a character, tonal differences are sometimes ignored.

The results for the homophonic items may at first seem paradoxical. It appears that the presence of characters with competing tones actually facilitates both response latencies and accuracy: While the difference between the Competing Tone and No Competing Tone conditions was nonsignificant on response times, neither for subjects nor for items, \( F(1,14) = 1.32, \, p > .1, \, F(1,48) = 0.57, \, p > .1 \), the error data produced a significant difference, \( F(1,14) = 21.66, \, p < .001, \, F(1,48) = 4.59, \, p < .05 \). Again, there was no interaction between tone competition and task on either measure.

The explanation for this result would actually appear to be quite simple. If tonal information is ignored, then the pool of characters which are considered to be homophonic with the presented word will be increased when there exist characters pronounced with the same segments and different tone. In other words, on a number of occasions with the Competing
Table 3
Reaction times for correct responses (RT in msec) and percentage error rates (%E) for the pairs of characters in Experiment 3.

<table>
<thead>
<tr>
<th>Condition</th>
<th>Example</th>
<th>Silent</th>
<th></th>
<th>Aloud</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>RT</td>
<td>%E</td>
<td>RT</td>
<td>%E</td>
</tr>
<tr>
<td>Non-homophones</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Competing Tone</td>
<td>肉 ròu</td>
<td>3214</td>
<td>63.6</td>
<td>2967</td>
<td>59.5</td>
</tr>
<tr>
<td>No Competing Tone</td>
<td>丢 diū</td>
<td>3134</td>
<td>48.2</td>
<td>2730</td>
<td>44.6</td>
</tr>
<tr>
<td>Homophones</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Competing Tone</td>
<td>洪 hóng</td>
<td>2104</td>
<td>12.8</td>
<td>1957</td>
<td>13.3</td>
</tr>
<tr>
<td>No Competing Tone</td>
<td>若 ruò</td>
<td>2209</td>
<td>26.7</td>
<td>1996</td>
<td>27.2</td>
</tr>
</tbody>
</table>

Tone items, subjects will make their "yes" response on the basis of the wrong character. Since this will unavoidably register as a correct response, performance in this condition will appear to be better than in the No Competing Tone condition where such errors do not occur. So, it seems that the homophonic items are reflecting the same phenomenon as the non-homophonic items, namely, that tones are not an important part of the phonological representation of a Chinese syllable.

In addition to the tone effect, however, what is particularly striking about the results of this experiment is the fact that the task was extraordinarily difficult for subjects to carry out. Reaction times of around 3 secs would seem to be extremely long, though one cannot draw a comparison with studies in English since there are no equivalent English data. However, error rates on this task using English materials are available from Taft (1984). On comparing the data from the English and Chinese studies, it can be seen that performance in the homophone decision task is very different in the two languages. The error rate for correct "no" responses in English (e.g., for words like ACHE) was very low (1.9% in the silent task and 0.3% in the aloud task), while in Mandarin, performance was not much better than chance for the equivalent items (i.e., the No Competing Tone condition). The error rate for correct "yes" responses was not so dramati-
cally different, though accuracy was still greater for English than for Mandarin. For the silent task in English there were 17% errors for homophones with the same morphological structure, and for the aloud task it was 14%. This compares to error rates of around 27% in the No Competing Tone condition for both tasks in Mandarin.

An attempt will now be made to explain the difficulty experienced by Chinese subjects in making homophone judgements in Experiment 3. After this, we will try to home in on the source of the tone effect found in all three experiments.

The first idea that comes to mind in attempting to account for the poor performance observed in the homophone decision task is that the link between orthography and phonology is not as strong in Chinese as it is in English, owing to the greater arbitrariness of the relationship for the former compared to the latter. It has been shown, for example, that phonological recognition is slower and less accurate than graphemic recognition for Chinese characters, while the reverse is true for English words (Chen & Juola, 1982). In the homophone decision task, a weak link between orthography and phonology could manifest itself at two loci; when the pronunciation of the presented character is generated, and when an alternative orthographic representation for that pronunciation is sought. Given that characters can typically be given their correct pronunciation, it is more likely that difficulties arise in the homophone decision task at the latter stage, when alternative orthographic forms are being sought.

The problem with this account, however, is that the link between phonology and orthography may be weak, but it must be strong enough to account for the fact that Chinese language users are nevertheless able to generate a correct graphemic representation for a spoken syllable, even out of context. One could envisage that it takes longer to think of the graphemic representation of a monosyllabic word in Chinese than in English, but it is difficult to see why there would be such a dramatic difference in error rates.

Instead of using such vague concepts as strong or weak links between orthography and phonology, we should look more precisely at how homophone judgements might be made. Consider the single item task used in Experiment 3 and by Taft in 1984. Intuitively, it has been assumed that one is able to make one's decision by determining the pronunciation of the item and then ascertaining whether that pronunciation can be graphemically represented in any other way. This involves access on a visual basis to the lexical entry for the presented item, extraction of the pronunciation
associated with that entry, and then an attempt to access the lexicon using this pronunciation in order to discover a lexical entry which is associated with a different orthographic representation to the one that has been presented.

While this may seem to be an intuitively obvious way to perform the homophone decision task, the fact is that in English (or any alphabetic language) there is actually another way in which the task could be successfully carried out. Having determined the pronunciation of the letter string, (for example, after establishing that PLAIN is pronounced /plein/), one could then generate other ways of spelling that pronunciation, (e.g., PLANE, PLEIN, PLEIGN, PLAYNE) and attempt to access these on a visual basis. If a lexical entry is so accessed, as it would be in the case of PLANE, a "yes" response could be made. If not, a "no" response could be made. Evidence for the use of this approach would be if a word like HEED were sometimes mistakenly thought to be homophonic with HEAD, since the pronunciation of HEED (namely, /hiːd/) can be spelled HEAD. Treiman, Baron, and Freyd (1983) provide some indication that such errors do occur, though they did not directly address this issue, while Taft (1982) found that such errors occurred using letter strings which were nonwords (like THREED being thought to be homophonic with THREAD).

Now, this second way of determining homophony is impossible in Chinese. One cannot reliably generate potential orthographic representations of specific pronunciations since the relationship between pronunciation and orthography is quite arbitrary. Instead, one can only use the first method. But what if this method were extremely inefficient (at least in Chinese and perhaps in English also)? If it were, we would witness a great deal of difficulty in performing homophone decisions on single characters in Chinese compared to English, since there would be no alternative approach to fall back on.

So, why might the attempt to access other characters on the basis of the pronunciation of the presented character be inefficient? One explanation could rest on the fact that a character in Chinese frequently combines with a second character to form a word. Therefore, when determining whether different characters may be used to represent a spoken syllable, all of these two character words must be considered. For example, when presented with '肉', one establishes that it is pronounced rū, but then must think of all words that include that syllable in order to determine whether any of these use a different character to '肉'. Given that subjects can never be sure that they have thought of all possible words containing a particular syllable, they
can never be sure that there is no other way of orthographically representing that syllable if they cannot actually think of one. This problem is compounded by the fact that there are no constraints on what that alternative orthographic representation might be, given the arbitrary mapping of pronunciation and orthography. In such circumstances, they might take a guess and perform at around chance level. This would result in around 50% errors for non-homophones, but a rather better performance for homophones, since subjects would occasionally actually think of the alternative orthographic representation of the syllable. In this way, the general difficulty with the task used in Experiment 3 can be explained. The relative ease of performing the task used in Experiments 1 and 2 is also explained. A phonological-orthographic association is not necessary in these experiments at all, since the homophony judgement is based on two given characters. All that is required in this task is for the pronunciation of the two characters to be generated and then for these two pronunciations to be compared.

This leads us then to an attempt at explaining the source of the tone effect. Since Experiments 1 and 2 did not require the generation of orthographic representations from phonological ones, the observed tone effect must have arisen at some locus other than this, a locus which was common to the pronunciation-matching task of the first two experiments and the homophone-generation task of the third experiment.

One might consider the stage at which pronunciation of the characters is determined and suggest that the confusion between different tones results from an inaccuracy in pronouncing tones. However, the confusion between syllables with different tones was just as great when the task was performed silently as when it was performed aloud, and Chinese readers have not been observed to make a large number of tonal errors in reading aloud, at least not with characters of the frequency of occurrence that were used in the present experiments. As mentioned earlier, in the Cantonese study, reading errors were monitored and if anything, there were fewer errors with the Different Tone items. Similarly, the pronunciations given in the aloud task of Experiment 3 were monitored and inappropriate tones were found to be rarely given. So it seems that an alternative explanation is required.

Another possibility might reside in the fact that both tasks require working memory. In the pronunciation-matching task, subjects must hold the pronunciation of the first character in memory while they determine the pronunciation of the second and compare it to the first. In the homophone-
generation task, the pronunciation of the presented character must be held in working memory while alternative orthographic representations are being sought. It has been amply shown that working memory involves an articulatory component (see Baddeley, 1990). To illustrate, the immediate recall of a list of digits (digit-span) is correlated with the speed with which the digits can be uttered, such that the shorter the names of the digits, the greater the digit-span. For example, digit-span in Cantonese is greater than in English (Hoosain & Salili, 1988), presumably because of the fact that the mean articulation rate of Cantonese digits is considerably faster than that of English.

Now, in order to explain the tone confusions observed in the homophone decision tasks, what can be suggested is that pitch characteristics are hard to represent in the articulatory coding of the syllable in working memory. It would not be surprising if such concretely defined articulatory features as the position of the tongue and lips (which contribute to a description of the vowels and consonants) are easier to represent in a memory code than the frequency of laryngeal vibration (which describes the tone). If so, we can say that the difficulty in deciding that two syllables with different tones are pronounced differently arises from the traces laid down in working memory.

There are several points that can be raised as being potential problems for this claim, however. First, it has recently been demonstrated (Xu, 1991) that immediate memory for the ordering of a visually presented list of characters whose pronunciations rhyme (e.g., bân, gân, ân, tân) is poorer than that for a list of characters whose pronunciations rhyme except for having different tones (e.g., bân, gân, ân, tân). This implies that tones are able to be represented in short-term memory, since different tones seem to be used as cues for discriminating between stored syllables. In response to this point though, the claim being challenged is not that tones cannot be represented, but rather that the representation is imperfect. Thus, it would presumably be the case that the list of rhyming characters with different tones would not be as well remembered as a list of characters which did not rhyme but which had the same tone (e.g., bân, gông, âng, tă), and this would be equivalent to the tone effect observed in Experiment 1.

A second possible argument against the suggestion that the source of the tone effect is working memory, is that one might have expected that the effect would disappear when the characters were pronounced aloud. In other words, one might expect that the implicit pronunciation of the character in the silent task, where the tone is imperfectly represented, would be replaced
by an explicit pronunciation where the tone is correctly produced. If so, it might be supposed that the echoic trace of the explicit pronunciation would take the place of working memory, and that therefore, different tones would not be confused. However, this assumption does not necessarily follow. It may be the case that a simple echoic trace is inadequate when a decision is required about the relationship of that trace to a second trace (Experiments 1 and 2) or to other pronunciations stored in lexical memory (Experiment 3). Under such circumstances, an articulatory-based working memory may still be required.

Conclusions

The only thing that can be definitely concluded from the research reported here is that tonal information is difficult to use when making a homophone decision. Several loci for this difficulty with tones were considered:

First, tones are possibly poorly represented in the underlying phonological representation of a syllable. If this were the case, though, one might have expected a reduction in difficulty when the characters were overtly pronounced, but this did not happen.

Second, subjects possibly mispronounced the characters, either because of dialect variations or because of ignorance. Yet no systematic mispronunciations were detected when subjects' overt responses were monitored. In addition, since the difficulty observed in the pronunciation-matching task of Experiments 1 and 2 arose from the non-homophones, the mispronunciation would have had to have fortuitously coincided with the pronunciation of the other character.

Third, tonal information may be hard to represent in working memory. In the pronunciation-matching task, the match would be attempted on traces of the pronunciation of the two syllables held in working memory. In the homophone-generation task, a representation of the pronunciation of the presented character must be held in working memory while a syllable that is homophonic with it is sought within lexical memory. While this is the favoured explanation (through lack of alternatives), one must make the assumption that such a working memory is used even when a self-generated overt phonetic representation is available.

Two other issues emerged from the research. The first was that it is very difficult to generate a homophone in Mandarin compared to English,
and more particularly, to decide that a singly presented character is not homophonic with any other character. The explanation for this was thought to reside in the fact that one must often generate a character in conjunction with another character since they form a single word, and one cannot be sure that one has thought of all the words that contain a character which is homophonic with but different to the presented character. Furthermore, unlike in English, there is no orthographic basis for generating homophones if the phonological mechanism is difficult to implement.

The other issue concerned the difference between the performance of Cantonese and Mandarin readers in the pronunciation-matching task. While both groups were relatively slow in discriminating two characters which had different tones as opposed to different segments, only the Mandarin speakers made more errors when so doing. In fact, Mandarin speakers made more errors overall than Cantonese speakers. Two possible causes for this were considered, though neither seemed greatly appealing: The structure of the two dialects and the nature of childhood instruction. It would be interesting to ascertain whether Cantonese speakers have a similar difficulty with homophone-generation as Mandarin speakers do, though this would be unlikely to provide further clues about the source of the difference. An examination of Cantonese speakers from Canton rather than Hong Kong might prove illuminating.

Acknowledgments

The first experiment was conducted at the University of Arizona while the first author was on study leave. The authors wish to thank Ken Forster for the use of his facilities, and Xing Kongliang for generously giving up his time to conduct the experiment and perform the analyses. The authors also acknowledge the help provided by Nancy Chen in designing the items.

The second experiment was supported by a UPGC Direct Grant for Research from the Social Science and Education Panel of the Chinese University of Hong Kong. The authors wish to thank Fong Kong for constructing the stimuli, conducting the experiment, and analyzing the data. We are also grateful to Ng Ching Man for assistance in computer programming.
The third experiment was carried out at the University of New South Wales with support from a grant provided by the Australian Research Council. The authors wish to thank Nancy Chen and Xu Ming Yang for constructing the items, Huang Jishan for conducting the experiment, and Bruce Russell for analyzing the data.

Notes

1. The alphabetic Pinyin script will be used to indicate pronunciation.

2. In English, but not in Chinese, it would also be possible to determine the pronunciation of the item on the basis of subcomponents of the item. For example, the pronunciation of PLAIN could be determined from the combined pronunciations of P, L, AI and N. The pronunciation of a nonword, like PRAIN, would need to be determined in this way, since there is no lexical entry for a nonword.

3. In an alphabetic language there are considerable constraints. For example, the pronunciation of the word PLAIN could never be alternatively spelt GRONK.

References


Educational Research Establishment (1986). *A study on Chinese lexical units for Hong Kong junior secondary students*. Hong Kong: Hong Kong Education Department.


