RECOGNITION OF SPOKEN WORDS IN THE CONTINUOUS SPEECH: EFFECTS OF TRANSITIONAL PROBABILITY

Michael C.W. Yip

Department of Psychology
Chinese University of Hong Kong, Hong Kong SAR

ABSTRACT

Do Cantonese listeners really use phonotactics information in the segmentation process of Cantonese continuous speech? Because some phoneme transitions across Cantonese syllables occur much more often than the others, the transitional probability may cue the locations of possible syllable boundaries in Cantonese speech. Two syllable-spotting experiments were conducted. Results clearly indicated that listeners would find it easier to spot the target syllable in the nonsense sound sequence, which consisted of high transitional probability phoneme combinations than the low transitional probability phoneme combinations across syllables. These results imply that Cantonese listeners are sensitive to the probabilistic phonotactic information across syllable boundaries during spoken language segmentation.

1. INTRODUCTION

One crucial aspect of human spoken language processing is the listener's ability to segment the continuous speech signal into individual perceptual units for lexical analysis. How do we know where one word or syllable ends and another begins as spoken languages normally emerge to the listeners as an unbroken sound sequence? Therefore, it is important to examine which kinds of information the listener utilizes to segment the continuous speech for the sake of getting a more complete picture of spoken language processing.

Previous studies on speech segmentation have revealed that the information of prosodic patterns of each respective language [1] is a useful cue to segment the continuous speech. Recently, some new research also found that segmentation is confronted with the information of phonotactic [2]. Knowledge of phonotactics can not only be used categorically in speech segmentation, it can also be used in a probabilistic way. Some phonotactics can not only be used categorically in speech segmentation, but also be used in a probabilistic way. Some recent findings also confirmed the effective role of the probabilistic phonotactics played in speech segmentation process [3,4,5]. Because these probabilistic cues are phonemically based and obviously embedded in every language, it can be therefore seen as a universal account to the speech segmentation problem.

However, studies on speech segmentation has been mainly examined in English and some other European languages (e.g., Dutch and French) so far; to the best of my knowledge, this question has not yet been systematically examined in Cantonese. As a fact, Cantonese is a language that differs significantly from most Indo-European languages, at least, in its use of lexical tones and its morphemic mono-syllabicity [6,7]. Also, these unique psycholinguistic properties of Cantonese should be quite useful to examine the speech segmentation problem. For example, most of the spoken syllables in Cantonese are in a simple Consonant-Vowel (CV) or Consonant-Vowel-Consonant (CVC) phonotactic structure and no consonant clusters are legal in spoken Cantonese [8]. Moreover, the final consonant of Cantonese monosyllables is limited to two classes, one with stops (ended with p, t, k) and the other one with nasals (ended with m, n, N). From these phonotactic structures, one can anticipate that it is possible to have some phoneme transitions across syllables occur much more frequently than the others in Cantonese continuous speech. Therefore, it appears that Cantonese provides a good testing environment in which to examine the role of probabilistic phonotactic information played in speech segmentation from a cross-linguistic perspective, which is exactly the main goal of the present study.

Two syllable-spotting experiments [9] reported in the present study was designed to address at least the following questions: Do listeners really use the frequency information of phoneme transitions from their lexicon to segment Cantonese continuous speech? Which kinds of phoneme transitions are easier for native listeners to notice the locations of possible syllable boundary? Which kinds of phoneme transitions are weaker?

2. EXPERIMENT 1

2.1. Method
Participants. A group of thirty-six native Cantonese speakers who are students at the Chinese University of Hong Kong took part in the experiment as a laboratory requirement for credit in an introductory psychology course.

Materials and Experimental Design. Two sets of stimuli were constructed, one for high TP and one for low TP. One set of materials belonged to the high transitional probability group which included the phoneme transitions of [i:-kh], [w-ph], [N-N], [N-kh], [N-t] while another set of materials belonged to the low transitional probability group that included the phoneme transitions of [i:-kh], [w-ph], [N-N], [N-kh].

Each of the selected phoneme transitions were used to construct a nonsense disyllabic compound word strings context, and a total of 40 nonsense disyllabic compound word strings were constructed that included all of the eight phoneme transitions, 20 for the high transitional probability group and 20 for the low transitional probability group. All of the 40 nonsense syllables strings were embedded with 40 real Cantonese syllables to make a total of 40 nonsense disyllabic compound word strings with real Cantonese syllables. The target Cantonese syllables were only located at the first syllable position of the nonsense syllable strings context. A group of 20 native Cantonese speakers were asked to judge the degree of nonsenseness of the materials. They were given a simple lexical decision test for all of the nonsense syllables and the real syllable fillers used in the present study (materials for transitional probability). Their response type confirmed the nonsense syllables used in this study were not real syllables in Cantonese.

In addition, another 40 nonsense disyllabic compound word strings were constructed as the appropriate fillers, which did not include any real Cantonese syllables, embedded. Altogether, there were 80 sound strings used in the experiment that already included the within factor of transitional probability (high vs. low). These 80 sound strings were divided into two different target-bearing-context versions. Each version had 40 sound strings (20 target-bearing sound strings and 20 fillers). Therefore, all the target Cantonese syllables appeared only once in each version.

Note that the transition of syllables would had either Consonant-to-Consonant or Vowel-to-Consonant combinations and generally included at least four major transitional types: CVCCVC, CVVCV, CVCV, CVCCV. These four types of transitions were nested within the 40 target-bearing contexts. The 36 participants were randomly assigned to two groups of eighteen. Each group randomly received an equal number of nonsense strings from one of the two different versions of materials. Each listeners received 40 nonsense sound strings in the experiment. The order of presentation for the target-bearing sound strings and the fillers was pseudorandomly arranged.

Experimental Apparatus. All the materials were recorded by a male native Cantonese speaker at a normal speaking rate, and then tape-recorded in a SONY DAT deck and then digitized into a Macintosh G3 computer. A sampling rate of 44.1kHZ with a 16-bit sound format was used for digitizing. The acoustic boundary of each Cantonese syllable was located as accurately as possible by inspecting speech waveforms and using auditory feedback. A unidirectional microphone to register listeners’ vocal response was connected to a remote-controlled SONY tape-recorder by the experimenter in another partition of the experimental room to check for accuracy.

Procedure. Before the experiment began, the experimenter explained the task in Cantonese to the listener. Listeners were told that they would hear a series of meaningless Cantonese syllable strings, each string was of a two-syllable length. Their task was to identify, for each piece of the nonsense strings, if there was any real Cantonese syllable that would occur in between by pressing a response button in front of them as quickly and accurately as possible, and then speaking aloud the detected syllable.

All participants did the experiment individually in a quiet room. A computer program called PsyScope [10] controlled the presentation of the materials. Listeners heard each nonsense string via two amplified JBL speakers connected to the Macintosh G3 computer. The time interval between the two strings was set at about 5 seconds. Before the test began, listeners were given a practice session in which they heard a set of separate but similar strings. The whole experiment lasted for 30 minutes.

2.2. Results and Discussion

False alarms, error responses (listeners named a syllable that was different from the target syllable), missing responses were all excluded for the analysis. Responses of duration that were over three standard deviations were also treated as missing values. All the response latencies were measured from the offset time of the nonsense sound sequences.

Mean response latencies, error rates and missing rates as a function of transitional probability are presented in Table1. Error rates and missing rates were very rare (on the average 1.37% for each conditions), and the error proportions and missing rates were not analyzed in the present experiment. The false alarm rate was 39.6% in this experiment.

<table>
<thead>
<tr>
<th>Transitional probability</th>
<th>High</th>
<th>Low</th>
</tr>
</thead>
</table>
A paired-samples t-test (high transitional probability vs. low transitional probability) was conducted on the response latencies of each spotted syllable. The main effect on transitional probability was significant at both the subject-analysis, t1(35) = 2.45, p < .05; and item-analysis, t2(39) = 2.13, p < .05 (one tailed).

In this experiment, listeners were instructed to monitor exclusively for initially embedded target syllables in the nonsense sound sequence with either a high transitional syllable boundary phonemic sequence or a low transitional syllable boundary phonemic sequence. Clearly, the observed transitional probability effects were in line with the basic prediction. First, with respect to the reaction time data, listeners detected the target syllables in the nonsense sound sequence with a higher likelihood of phoneme transition, on the average, 61 milliseconds faster than the sound sequence with a low probability of phoneme transition (520 vs. 581 milliseconds). In addition, the missing rates and the error proportion were generally consistent with the reaction time data. Listeners made many fewer missing target items and mistakes in the high transitional probability materials set than the low transitional probability materials set. Therefore, the results confirmed the prediction that listeners used the information of phoneme transition across syllables' boundary to segment the Cantonese continuous speech. These results are also consistent with other relevant studies on transitional probability [3,4,5]. In this experiment, listeners only focused on a single location (the first syllable position) to each nonsense sound strings, so in an attempt to make a stronger conclusion on the usage of transitional probability in Cantonese speech segmentation, I changed the target position at the following experiment to confirm that whether the probabilistic effects are genuine regardless of any positional specificity influence.

3. EXPERIMENT 2

3.1. Method

Participants. Another group of thirty-six native Cantonese speakers who are students at the Chinese University of Hong Kong took part in the experiment as a laboratory requirement for credit in an introductory psychology course.

Materials and Experimental Design. Two sets of stimuli were constructed again, one for high TP and one for low TP. One set of materials belong to the high transitional probability group which included the phoneme transitions of [i:-j], [w-s], [N-k], [N-t] while another set of materials belong to the low transitional probability group that included the phoneme transitions of [3:-j], [a:-s], [ph-k], [m-t]. The materials construction and the experimental design were the same as Experiment 1.

Procedure. Same as Experiment 1. The only difference to the procedure of Experiment 1 was the location of the target syllable. In this experiment, all the target syllables were located at the second syllable position in the sound syllable string.

3.2. Results and Discussion

Mean response latencies, error rates and missing rates as a function of transitional probability are presented in Table 2. Error rates and missing rates were very rare (on the average 5.5% for each conditions), and were not analyzed in the present experiment. The false alarm rate was 44.2% in this experiment.

<table>
<thead>
<tr>
<th>Transitional probability</th>
<th>High</th>
<th>Low</th>
</tr>
</thead>
<tbody>
<tr>
<td>RT (milliseconds)</td>
<td>680.3</td>
<td>721.9</td>
</tr>
<tr>
<td>Error rate (%)</td>
<td>3.61%</td>
<td>3.52%</td>
</tr>
<tr>
<td>Missing rate (%)</td>
<td>6.7%</td>
<td>8.21%</td>
</tr>
</tbody>
</table>

Table 2. Mean reaction times, error rates and missing rates of Experiment 2

A paired-samples t-test (high transitional probability vs. low transitional probability) was conducted again on the response latencies of each spotted syllables. There was a significant main effect on the variable of transitional probability, t1(35) = 1.92, p < .05; and t2(39) = 1.88, p < .05 (one tailed).

Unlike the first experiment, in this experiment, listeners were instructed to monitor exclusively for finally-embedded target syllables in the nonsense sound sequence with again either a high transitional syllable boundary phonemic sequence or a low transitional syllable boundary phonemic sequence. Clearly, the observed transitional probability effects show a similar pattern to the first experiment and are also in line with the research hypothesis. First, with reference to the reaction time data, listeners actually spotted the target syllables in the nonsense sound sequence with a higher phoneme transition probability, on the average, 42 milliseconds faster than the sound sequence with a low probability on phoneme transition (680 vs. 722 milliseconds).
milliseconds). In addition, the missing rates and the error proportion were generally consistent with the reaction time data although a bit higher than the results obtained from the Experiment 1. Listeners produced fewer missing target items and a comparable error rate to the high transitional probability materials set than the low transitional probability materials set.

Once again, the observed results confirmed the prediction that listeners actually used the probabilistic information of phoneme transition across syllables' boundaries to handle the segmentation task for Cantonese continuous speech.

Moreover, from comparing the reaction time between Experiments 1 and 2, it clearly shown that to spot an initially embedded target syllable in the nonsense sound sequence is much easier than to spot a finally embedded target syllable. On the average, listeners responded 150 milliseconds faster to the initially embedded target materials than to the finally embedded target materials. This is a clear time-course effects of spoken word recognition.

4. GENERAL DISCUSSION

The present study extended the investigation to the problem of speech segmentation in Cantonese. Much of our knowledge about this problem has come from results of those European languages. With a view to investigating further this important question across languages, I used Cantonese continuous speech as a crucial test case. Since Cantonese represents a significantly different language from Indo-European languages, its phonological and lexical properties make the language ideal for examining the speech segmentation issue [6,7] by this new approach.

Two syllable-spotting experiments provided convergent evidence to support the following points. First, the relative frequency information of phoneme transition across syllable is actually used by Cantonese listeners to segment the continuous sound sequence. Also, the present study also demonstrated a positional specificity effect that response times for initially embedded target syllables are generally faster than the times for finally embedded target syllables. These results are consistent with the time course study of spoken word recognition due to lexical influences. Lexical information from the meaningful preceding context to a certain extent aids listeners to detect an initially embedded target in the nonsense sound string much faster and more easy than the finally embedded target [11].

One thing should be noted that is the relatively high false alarm rate. From the follow-up analysis to the types of false alarm error, over 80% came from the tone-mismatch cases; 15% came from the onset-mismatch cases; and only 3% came from the vowel-mismatch cases. These patterns of results are also in line with the other relevant findings of Cantonese spoken word processing [12].

Ongoing experiments are being designed to further examine the effects of probabilistic phonotactics operate in Cantonese speech segmentation by extending the continuity nature of the sound sequence from disyllables to multi-syllables.

5. ACKNOWLEDGEMENTS

I would like to thank Hsuan-Chih Chen, Anne Cutler, and James McQueen for their valuable suggestions to the present study. Thanks also go to David Kwan and Iris Wong for their excellent help in the preparation and running of these experiments. In addition, the present study was supported by a postgraduate fellowship from the Chinese University of Hong Kong and a dissertation grant from the Department of Psychology, Chinese University of Hong Kong. Email: mcwyip@psy.cuhk.edu.hk

6. REFERENCES

