Phonotactic constraints and the segmentation of Cantonese speech

Michael C.W. Yip

Department of Psychological Studies
The Hong Kong Institute of Education, Hong Kong SAR
mcwyip@ied.edu.hk

Abstract

Two word-spotting experiments were conducted to examine the question of whether native Cantonese listeners are constrained by phonotactic information in the segmentation of Cantonese continuous speech. Because there are no legal consonant clusters occurred within individual Cantonese words, so this kind of phonotactic information of words may most likely cue native Cantonese listeners the locations of possible word boundaries in the continuous speech. Finally, the observed results from the two experiments confirmed this prediction. Together with other relevant studies [1,2], we argue that phonotactic constraint is one of the useful sources of information in segmenting Cantonese continuous speech.

1. Introduction

One peculiar capacity of human spoken language processing is listeners' effortless ability to segment the continuous speech signal into individual perceptual units for lexical analysis. However, to uncover the reason(s) why we can easily notice the points where one word ends and another begin during the running sound sequence is intriguingly attractive to many experimental psycholinguists. Therefore, it evokes so many important researches to investigate which kinds of information the listeners may use to segment the continuous speech in order to figure out a more complete and comprehensive picture of spoken language processing.

Many previous studies on speech segmentation have indicated that the prosodic information of each respective language [3] is a useful cue to segment the continuous speech. Furthermore, some other researchers report that segmentation is normally confronted with the information of phonotactics in the language [4,5]. Knowledge of phonotactics can be used categorically and probabilistically in speech segmentation. Some recent findings confirmed the effective role of the probabilistic phonotactics played in speech segmentation process in different languages [6,7,8]. Because the phonotactic cues are phonemically based and embedded in all languages, so it can be seen as a universal cue to the speech segmentation problem.

Studies on speech segmentation have been extensively 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 Chinese. In fact, Cantonese is a language that differs significantly from most Indo-European languages, in terms of topology; in terms of its use of lexical tones and its morphemic mono-syllabicity [9,10]. All these unique psycholinguistic properties of Cantonese Chinese are obviously very useful to examine the speech segmentation problem. For example, almost all of the spoken Cantonese syllables are in a simple Consonant-Vowel (CV) or Consonant-Vowel-Consonant (CVC) phonotactic structure and no consonant clusters are allowed in spoken Cantonese words [11,12]. 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, ŋ). From these phonotactic structures, one can expect that the phonotactic constraint is most likely to be useful to deal with the segmentation task of Cantonese continuous speech. Therefore, it appears that Cantonese Chinese provides an appropriate testing ground in which to examine the role of phonotactic information played in speech segmentation from a cross-linguistic perspective. In the present study, two word-spotting experiments [13] were designed to address the important empirical question: do native Cantonese listeners use the phonotactic information to segment Cantonese continuous speech?

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 legal phonotactic sound sequence and one for illegal phonotactic sound sequence within a word. One set of materials include legal phonotactic sound sequence (CVCVC): phoneme transitions of [i-y], [ei-s], [aau-k], [aan-t] while another set of materials include illegal phonotactic sound sequence (CVCCV): phoneme transitions of [ŋ-y], [n-s], [ŋ-k], [ŋ-t].

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 legal phoneme sequencing group and 20 for the illegal phoneme sequencing 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 evaluate 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. 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 legality of the phoneme sequences (legal vs. illegal). An effort was made to keep the duration of all sound strings as equal as possible. 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.

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 double check for accuracy.

Procedure. Before the experiment began, the experimenter explained the task in Cantonese to each 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 string, if there was any real Cantonese syllable that embedded in the sound string by pressing a response button (Yes/No) in front of them as quickly and accurately as possible, and then speaking aloud the spotted syllable (or no syllable).

All participants did the experiment individually in a quiet room. A computer program called PsyScope [14] 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 20 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 target offset time.

Mean response latencies, error rates and missing rates as a function of legality of phoneme transitions are presented in Table1. Error rates and missing rates were very rare (on the average 3.6% for each conditions), and the error proportions and missing rates were not analyzed in the present experiment.

Phoneme Transitions	Legal /jaau^{2}kit^{1}/	Illegal /biŋ^{1}tan^{6}/
RT (milliseconds)	578.3	498.1
Error rate (%)	4.1%	2.3%
Missing rate (%)	1.7%	0.6%

Table 1. Mean reaction times, error rates and missing rates of Experiment 1

A paired-samples *t*-test (legal phoneme transitions vs. illegal phoneme transitions) was conducted on the response latencies of each spotted syllable. The main effect was significant, $t(35) = 5.76$, $p < .05$.

In this experiment, listeners were instructed to monitor exclusively for initially embedded target syllables in the nonsense sound sequence with either a legal phoneme transitions or an illegal phoneme transitions. The observed legality effects were consistent with the prediction. First, with respect to the reaction time data, listeners detected the target syllables in the nonsense sound sequence with an illegal phoneme transitions, on the average, 80 milliseconds faster than the sound sequence with a legal phoneme transitions (498 vs. 578 milliseconds). In addition, both the missing rates and the error proportion were generally consistent with the reaction time data. Listeners made fewer missing target items and mistakes in the illegal phoneme transitions materials set than the legal phoneme transitions materials set. Therefore, the results confirmed the prediction that listeners used the phonotactic information of phoneme transition across syllables' boundary to segment the Cantonese

continuous speech. These results are also consistent with other relevant studies on phonotactic constraints [4,5]. In the first experiment, listeners were asked to focus on a single location (the first syllable position) to each nonsense sound string, so in an attempt to make a stronger conclusion on the usage of phonotactic information across phoneme transitions in Cantonese speech segmentation, a parallel experiment ,that changed the target position to the second syllable in the nonsense word strings, was followed up on to further verify the phonotactic legality effects in the speech segmentation 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. 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

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 target offset time.

Mean response latencies, error rates and missing rates as a function of legality of phoneme transitions are presented in Table 2. Error rates and missing rates were small again (on the average 5.7% for each conditions), and the error proportions and missing rates were not analyzed in the present experiment.

Phoneme Transitions	Legal /nei^{1}yuk^{6}/	Illegal /dan^{1}sek^{6}/
RT (milliseconds)	784.5	722.6
Error rate (%)	5.1%	4.7%
Missing rate (%)	7.4%	6.9%

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

A paired-samples *t*-test (legal phoneme transitions vs. illegal phoneme transitions) was conducted on the response latencies of each spotted syllable. Again, there was a significant main effect, $t(35) = 3.15$, $p < .05$.

In this experiment, listeners were instructed to monitor exclusively for finally-embedded target syllables in the nonsense sound sequence. The observed legality effects show a similar pattern to the first experiment and are also in line with the research prediction. First, with reference to the reaction time data, listeners actually spotted the target syllables in the nonsense sound sequence with an illegal phoneme transitions, on the average, 62 milliseconds faster than the sound sequence with a legal phoneme transitions (722 vs. 784 milliseconds). In addition, the missing rates and the error proportion were generally consistent with the reaction time data. Listeners produced fewer missing target items and a comparable error rate to the illegal phoneme transitional materials set than the legal phoneme transitions materials set.

Overall, the observed results in both experiments confirmed the prediction that listeners actually used the phonotactic information of phoneme transitions across syllables' boundaries to solve the segmentation task for Cantonese continuous speech.

Moreover, from comparing the reaction time between experiments 1 and 2, it is 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 215 milliseconds faster to the initially embedded target materials than to the finally embedded target materials. This is a simple time-course effects of spoken word recognition.

4. General Discussion

The present study extended the investigation to the problem of speech segmentation in Cantonese Chinese. Most of our knowledge about this psycholinguistics problem has come from the experimental results collected from those European languages. With a view to investigating further this important question across languages, we used Cantonese continuous speech as a crucial test case. Since Cantonese Chinese represents a significantly different language from other Indo-European languages, its phonological and lexical properties make the language interesting to examine the speech segmentation issue [15] by this approach.

Two word-spotting experiments provided convergent evidence to support the following points. First, the phonotactic information of each Cantonese syllable is in fact used by native Cantonese listeners to segment the continuous sound sequences. The present study also demonstrated a positional specificity effect that response times for initially embedded target syllables were 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 aids listeners to detect

an initially embedded target in the nonsense sound string much faster and easier than the finally embedded target [16,17,18].

Ongoing experiments are being designed to further examine the effects of phonotactic information (both categorical and probabilistic) operate in Cantonese speech segmentation by extending the continuity nature of the sound sequences from disyllables to multi-syllables.

5. Acknowledgements

I would like to thank 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 grant from the Department of Psychology, The Chinese University of Hong Kong as well as the Research Support Fund from the Hong Kong Institute of Education.

6. References

1. Yip, M. C. W. (2000). Recognition of spoken words in continuous speech: Effects of transitional probability. *Proceedings of the ICSLP'2000,* 758-761.
2. Yip, M. C. W. (2006). The role of positional probability in the segmentation of Cantonese speech. *Proceedings of ICSLP'2006,* 865-868.
3. Cutler, A., Dahan, D. & van Donselaar, W. (1997). Prosody in the comprehension of spoken language: A literature review. *Language and Speech, 40*, 141-201
4. McQueen, J. M. (1998). Segmentation of continuous speech using phonotactics. *Journal of Memory & Language, 39,* 21-46.
5. Warner, N., Kim, J., Davis, C., & Cutler, A. (2005). Use of complex phonological patterns in speech processing: evidence from Korean. *Journal of Linguistics*, *41*, 353-387.
6. Gaygen, D. E. (1999). *Effects of phonotactic probability on the recognition of words in continuous speech*. Unpublished doctoral Dissertation, State University of New York at Buffalo, New York.
7. van der Lugt, A. (2001). The use of sequential probabilities in the segmentation of speech. *Perception & Psychophysics, 63,* 811-823.
8. McQueen, J. M., & Pitt, M. A. (1996). Transitional probability and phoneme monitoring. *Proceedings of ICSLP'96,* 2502-2505.
9. Li, P. (1998). Crosslinguistic variation and sentence processing: The case of Chinese. In D. Hillert (ed.), *Sentence processing: A crosslinguistic perspective*. San Diego, CA: Academic Press, pp. 33-51.
10. Zhang, Y., Wu, N., & Yip, M. C. W. (2006). Lexical ambiguity resolution in Chinese sentence processing. In P. Li, L. H. Tan, E. Bates, & O. J. L. Tzeng. (eds.), *Handbook of East Asian Psycholinguistics (Vol. 1: Chinese)*, pp. 268-278. Cambridge, UK: Cambridge University Press.
11. Kao, D. (1971). *Structure of the syllable in Cantonese*. The Hague: Mouton.
12. Matthews, S. & Yip, V. (1994). *Cantonese: A Comprehensive Grammar*. London: Routledge.
13. McQueen, J. M. (1996). Word spotting. *Language and Cognitive Processes, 11*, 695-699.
14. Cohen, J. D., MacWhinney, B., Flatt, M., and Provost, J. (1993). PsyScope: A new graphic interactive environment for designing psychology experiments. *Behavior Research Methods, Instruments, and Computers, 25*, 257-271.
15. Yip, M. C. W. (2000). Probabilistic phonotactics and the segmentation of Cantonese continuous speech. Dissertation Abstracts International, 61 (8-B), 4444.
16. Li, P. & Yip, M. C. W. (1998). Context effects and the processing of spoken homophones. *Reading and Writing, 10*, 223-243.
17. Yip, M. C. W. (2000). Spoken word recognition of Chinese homophones: The role of context and tone neighbors. *Psychologia, 43*, 135-143.
18. Connine, C. M., Blasko, D. G., & Titone, D. (1993). Do the beginnings of words have a special status in auditory word recognition? *Journal of Memory and Language, 32*, 193-210.