The role of positional probability in the segmentation of Cantonese speech

Michael C. W. Yip

School of Arts & Social Sciences
The Open University of Hong Kong, Hong Kong SAR

ABSTRACT

The present paper examines the question of whether native Cantonese listeners make use of probabilistic phonotactics information of words in the segmentation process of Cantonese continuous speech. Because some sounds appear more frequently at the beginning or ending of Cantonese syllables than the others, these kinds of probabilistic information of syllables may be likely to cue the locations of possible syllable boundaries in Cantonese continuous speech. A syllable-spotting experiment was conducted and the results indicated that native Cantonese listeners indeed made use of the positional probabilities of a syllable's onset but not for the case of a syllable's final in the segmentation process. Along with my previous study [1], I argue that probabilistic phonotactics is one useful source of information in Cantonese speech segmentation.

Index Terms: speech segmentation; probabilistic phonotactics; Cantonese speech

1. INTRODUCTION

Listening to spoken languages seems effortless. However, if we consider the questions of how do we know where one word ends and another begins in a continuous sound sequence; and how to segment the continuous speech into different perceptual units for additional lexical analysis, then it turns out to be a non-trivial matter because this fundamental capacity of human beings involved a basic but complex process in spoken language comprehension: speech segmentation process. Psycholinguists worked so hard on this issue during the past ten years since they thought that it is very important to examine which kinds of information listener will use to solve the speech segmentation problem in order to obtain a more comprehensive picture of spoken language processing [2].

On the basis of current literature, some researchers suggested that the prosodic information of each respective language [3] is a useful cue to segment the continuous speech. Moreover, some recent research found that segmentation process was more or less confronted with the information of phonotactics [4]. Knowledge of phonotactics can not only be used categorically in speech segmentation, it can also be used in a probabilistic way. Recent findings also supported the effective role of the probabilistic phonotactics played in speech segmentation process [5,6,7]. And these probabilistic cues are phonemically based and hence it can be seen as a universal explanation to the speech segmentation problem if they are successfully operated.

So far, examination of this phonotactics cue in speech segmentation has been mainly conducted in English and some other European languages (e.g., Dutch and French); to the best of my knowledge, this question has not yet been systematically examined in Cantonese. As a matter of 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 [8,9]. These unique psycholinguistic properties of Cantonese are quite useful to examine the speech segmentation problem. For example, about 99% of the spoken monosyllables in Cantonese are in a Consonant-Vowel (CV) or Consonant-Vowel-Consonant (CVC) phonotactic structure and no consonant clusters are legal in spoken Cantonese (Kao, 1971). Also, the final consonant of Cantonese monosyllables, if any, is limited to two classes: stops (p, t, k) and nasals (m, n, N). That means, when listeners hear a sound sequence with a voiceless alveolar fricative [s], there must be a syllable boundary before that since there are no voiceless alveolar fricatives occur in the ending position in Cantonese words. Hence, it is easy to discover that some beginning/ending sounds occur more frequently than the others in Cantonese syllables. Consequently, it appears that Cantonese provides a good testing environment in which to examine the role of probabilistic phonotactics information played in speech segmentation from a cross-linguistic perspective. In the present study, a syllable-spotting experiment [11] was conducted to address the following question: Do listeners use the frequency information of phoneme in each Cantonese word to segment continuous speech?

2. EXPERIMENT

2.1. Method

Participants. A group of twenty-eight native Cantonese speakers (12 males and 16 females, mean age = 21.6) who reported no speech or hearing deficits participated in the experiment. They are all undergraduate students at the Chinese University of Hong Kong and took part in the experiment as a laboratory requirement for credit in an introductory psychology course.

Materials and Experimental Design. Two sets of materials were constructed. One set of materials belongs to the high positional probability group, which includes the initial consonants [tS], [s], [tSh]. Another set of materials belongs to the low positional probability group, which includes the

initial consonants [khw], [kw], [N]. The probability difference between the two groups was significant at $p < .05$. Each of the selected initial consonant was used to construct a nonsense syllable, like [tSa:j4], [si:m3], [khwi:p2]. A total of 70 nonsense syllables were constructed that included all of the six initial consonants, 35 for the high positional probability group and 35 for the low positional probability group. Rimes in these 70 nonsense syllables were all the same across each testing group except the selected initial consonants. Afterwards, all of the 70 nonsense syllables were paired with 35 real Cantonese syllables to make a total of 70 nonsense word strings. In addition, another 70 nonsense word strings were constructed as the appropriate fillers which did not include any real Cantonese syllables embedded.

Similar to the case of the initial consonant, another two sets of materials were constructed to the finals. One set of materials belongs to the high positional probability group, which included the finals [n], [N], [i:]. Another set of materials belongs to the low positional probability group, which included the finals [3:], [t], [p]. Although some of these types of final phonemes appeared both at the initial and the final position of Cantonese syllables, they were only voice released at the initial position. So, they had borne with some crucial psychoacoustics differences for their different locations. Likewise, the probability difference between the two groups was significant at $p < .05$. Each of the selected final phoneme was used to construct a nonsense syllable, like [phTn6], [jTt3], [thO:N6] then a total of 70 nonsense syllables were constructed that included all of the six finals, 35 for the high positional probability group and 35 for the low positional probability group. Similarly, the initials and rimes in these 70 nonsense syllables were all the same across each testing group except the selected finals. Afterwards, all of the 70 nonsense syllables were paired with 35 real Cantonese syllables to make a total of 70 nonsense word strings. In addition, another 70 nonsense word strings were constructed as the appropriate fillers which did not include any real Cantonese syllables embedded.

A separate group of 20 native Cantonese speakers were asked to judge the degree of nonsenseness for the relevant materials. They were given a simple lexical decision test to all the nonsense syllables and the real syllable fillers used in the study of positional probability. Their responses confirmed that on the average, over 98% of the nonsense syllables in this study were not real syllables in Cantonese.

In the present experiment, we used a mixed design that contained target syllables embedded both in the first and the second position of the nonsense word strings. This kind of mixed design should be useful to reduce the strategic effects of the participants to attend to one single location to the sound sequences. Therefore, there were altogether 280 sound sequences (nonsense word strings) used in the experiment that included the within factor of positional probability (high vs. low) as well as target syllable position (first syllable vs. second syllable), and these 280 sound sequences were divided into two different target-bearing-context versions, each version has 140 sound sequences (70 target-bearing sound sequences and 70 fillers). Therefore, all the target syllables appeared only once in each version.

The 28 participants were randomly assigned to two groups of fourteen. Each group randomly received an equal number of nonsense sound sequences from either one of the two different versions of materials. Each listener received 140 nonsense sound sequences in the experiment. The order of presentation for the target-bearing sound sequences 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 of their verbal responses.

Procedure. Before the experiment began, experimenter explained the task in Cantonese to the listener. Listeners were told that they would hear a series of meaningless Cantonese word strings; each string was two-syllables in length. Their task was to identify, for each nonsense word string, if there was any real Cantonese syllable embedded, once they heard a target, they were asked to press the "Yes" response key and then speak aloud the detected syllable; while if they thought there was no syllable embedded, they were asked to press the "No" response key and then speak aloud the word "No".

All participants did the experiment individually in a quiet room. A computer program, PsyScope [12] controlled the presentation of the materials. Listeners heard each nonsense word string via two amplified JBL speakers connected to the Macintosh G3 computer. The time interval between the sound sequences 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 sound sequences. 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), and missing responses were all excluded from 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 to key press.

Mean response latencies, error rates and missing rates as a function of positional probability and target position are

presented in Table 1. Error rates and missing rates were very rare (on the average 2.5% for each condition), so the error proportions and missing rates were not analyzed in the present experiment. The false alarm rate was 41.2% in this experiment.

	Target position	
Positional probability	Initial	Final
High		
RT (milliseconds)	838.0	928.1
Error rate (%)	1.68%	2.55%
Missing rate (%)	1.94%	4.13%
Example	[fa:1]	[thQw4]
	[fa:1**tS**a:k4]	[ju**N**3thQw4]
Low		
RT (milliseconds)	890.0	952.7
Error rate (%)	1.38%	2.35%
Missing rate (%)	1.53%	4.18%
Example	[fa:1]	[thQw4]
	[fa:1**kw**a:k4]	[jT**t**3thQw4]

Table 1. Mean response latencies, error rates and missing rates of the Experiment

A 2 (high positional probability vs. low positional probability) x 2 (target position: first vs. second syllable) repeated measure ANOVA was conducted on the response latencies of each spotted syllable. Results from the ANOVA revealed that there were significant main effects on positional probability, $F1(1,27) = 8.4$, $p < .05$, $F2(1,34) = 5.89$, $p < .05$; and also the main effect on target position, $F1(1,27) = 7.37$, $p < .05$, $F2(1,34) = 2.76$, $p > .05$, n.s. Also, their interaction was marginally significant at $F1(1,27) = 5.33$, $p = .051$, $F2(1,34) = 0.98$, $p > .05$, n.s. Collapsed over levels of target position, listeners responded 38 milliseconds faster to the high positional probability materials set than the low positional probability materials set (883 vs. 921 milliseconds). These results were again consistent with other related studies [5,6].

Again, collapsed over the levels of two variables, in case of "first syllable as target" condition, listeners responded, on the average, 52 milliseconds faster to the high positional probability materials set than the low positional probability materials set. These findings confirmed that listeners made use of the probabilistic information on a syllable's beginning portion during the Cantonese segmentation and word recognition processes. In case of "second syllable as target" condition, listeners' responded 25 milliseconds, on the average, faster for the high positional probability materials set than for the low positional probability materials set although statistically unreliable. However, the distribution of missing rate and the error proportion in this experiment was also generally compatible with the response latencies data, which still indicated a strong positional probability effect.

3. GENERAL DISCUSSION

The present study has attempted to tackle one of the fundamental problems of spoken language processing: speech segmentation. Much of our knowledge about this problem has come from experimental findings from Indo-European languages. Researchers have generally assumed that their data can be generalized to support a general theory of spoken language processing. Although there may be good reasons to assume so, there are certainly other reasons to examine the same matter beyond the Indo-European language family as some special psycholinguistic features may simply not exist in these languages [2]. 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 Indo-European languages, its phonological and phonotactics properties make the language ideal for examining the speech segmentation issue [13].

Results indicated that listeners are sensitive to a syllable's beginning sound to notice where the boundary of a new syllable begins during the continuous speech. They responded much faster to those items with a very likely beginning sound than to those items with a very rare beginning sound. However, results did not support that listeners could make use of the probabilistic information provided by a syllable's ending sound to notice the exact completion boundary of a syllable during the continuous speech.

The present results also reveal that the relative likelihood of a syllable's onset seems to be more important than the relative likelihood of a syllable's offset. Basically, many studies have confirmed the importance of a syllable's onset in the fast recognition of words in continuous speech. For example, gating studies [14,15,16] have indicated that listeners could recognize a spoken word after the onset and before the acoustic offset. Cross-modal priming studies also suggest that the information of a word's beginning is crucial for efficient spoken word recognition [17,18,19]. These effects are robust in many relevant studies on speech perception and speech segmentation [4,5,6].

In addition, the present study shows that the response times for initially-embedded target syllables are generally faster than the times for finally-embedded target syllables. These positional specificity effects illuminated the weak positional probability effect on a syllable's ending sound. The null positional probability effects on syllables' ending sound may be related to the low perceptual sensitivity to the ending portion of Cantonese syllables, which is in line with the findings of weak coda effect in Chinese speech analysis [20].

Similar to my previous work [1], there is a relatively high false alarm rate in the present study. These high false alarm rates may imply that the obtained results may also be due to the effects of response set rather than the true probabilistic effects. So, we double checked for the data and analyzed the

pattern of false alarm. Finally, there were no significant differences between the high and low positional probability in each respective filler, $p > .05$. However, another interesting phenomenon was observed. That is, over 80% of the false alarm is due to the tone mismatch case, about 15% is due to the onset mismatch and only 3% is due to the change of vowel. These results are consistent with the findings of Cutler and Chen [21], that tonal information is perceptually more vulnerable than the segmental information in Cantonese Chinese.

Ongoing experiments are being designed to further examine the effects of probabilistic phonotactics and the prosodic structure of Cantonese words operated in Cantonese speech segmentation.

4. ACKNOWLEDGEMENTS

I would like to thank Hsuan-Chih Chen, Anne Cutler, James McQueen and Arie van der Lugt for their comments to this study.

5. REFERENCES

1. Yip, M. C. W. (2000). Recognition of spoken words in continuous speech: Effects of transitional probability. In B. Yuan, T. Huang, & X. Tang. (Eds.), *Proceedings of the ICSLP'2000,* 758-761. Beijing: China Military Friendship Publish.
2. Cutler, A. (1997). The comparative perspective on spoken-language processing, *Speech Communication, 21*, 3-15.
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. 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.
6. van der Lugt, A. (2001). The use of sequential probabilities in the segmentation of speech. *Perception & Psychophysics, 63,* 811-823.
7. McQueen, J. M., & Pitt, M. A. (1996). Transitional probability and phoneme monitoring. In *Proceedings of ICSLP 96, Vol.4*, pp. 2502-2505. Philadelphia, USA.
8. Chen, H-C. (1996). Chinese reading and comprehension: A cognitive psychology perspective. In M. H. Bond (Eds.), *Handbook of Chinese psychology* (pp. 43-62). Hong Kong: Oxford University Press.
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. Kao, D. (1971). *Structure of the syllable in Cantonese*. The Hague: Mouton.
11. McQueen, J. M. (1996). Word spotting. *Language and Cognitive Processes, 11*, 695-699.
12. 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.
13. Chen, H-C. & Tzeng, O. J-L. (1992), *Language processing in Chinese (Eds.)*. Amsterdam: North-Holland.
14. Grosjean, F. (1980). Spoken word recognition processes and the gating paradigm. *Perception and Psychophysics, 28*, 267-283.
15. Li, P. & Yip, M. C. W. (1998). Context effects and the processing of spoken homophones. *Reading and Writing, 10*, 223-243.
16. Yip, M. C. W. (2000). Spoken word recognition of Chinese homophones: The role of context and tone neighbors. *Psychologia, 43*, 135-143.
17. Li, P. & Yip, M. C. W. (1996). Lexical Ambiguity and context effects in spoken word recognition: Evidence from Chinese. In G. Cottrell. (ed.). *Proceedings of the Eighteenth Annual Conference of the Cognitive Science Society*, 228-232.
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.
19. Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word-recognition, *Cognition, 25,* 71-102.
20. Bertelson, P., Chen, H-C., & de Gelder, B. (1997). Explicit speech analysis and orthographic experience in Chinese readers. In H-C. Chen (Ed.)., *Cognitive Processing of Chinese and related Asian languages* (pp.27-46). Hong Kong: Chinese University Press.
21. Cutler, A. & Chen, H-C. (1997). Lexical tone in Cantonese spoken-word processing. *Perception and Psychophysics, 59*, 165-179.