Phonetics/cognitive linguistics/psycholinguistics

Perceptual Discretization during Speech Identification: A Study on Intelligibility of Locally Time-Reversed Speech in Mandarin and Cantonese

It has been hypothesized that speech signals are analyzed in a discrete fashion in the human brain [1]. A previous study used a method of “local time-reversal” (slice speech signals into segments of equal duration and then time-reverse each segment) to generate discrete time frames one after another in English spoken sentences [2]. They found that reversed segment durations of less than 40 ms resulted in modest degradation of word intelligibility, and when the duration is up to 40~50 ms, the word intelligibility can still reach 70~80%. In the current study, we further investigated this perceptual discretization in Mandarin and Cantonese speech identification.

In this experiment, we used Mandarin and Cantonese Semantically Unpredictable Sentences (SUSs, which are syntactically acceptable but semantically anomalous [3]) naturally pronounced at 4, 6 and 8 Hz syllabic rates. The sentences were then locally reversed with segments of various lengths, including 0 (namely unreversed condition), 5, 10, 20, 30, 40, 50, 60, 80 and 120ms. 18 Mandarin-Cantonese bilinguals were recruited in Guangdong Province and they were instructed to listen to both Mandarin and Cantonese SUSs and write down the words or syllables they heard. Syllable identification accuracies were calculated for different reversed length conditions. The results are quite consistent in both Mandarin and Cantonese SUSs: (1) The syllable accuracies reach over 80% at reversed segment durations of 20~50 ms for 4-Hz-syllabic-rate SUSs, at 20 and 30 ms for 6-Hz-syllabic-rate SUSs and at 20 ms for 8-Hz-syllabic-rate SUSs, respectively; (2) for other locally reversed durations, the syllable accuracies drop to lower than 80%.

The results show that the range of reversed segment durations is not the same for SUSs at different syllabic rates when syllable accuracies are over 80%. However, some common grounds are shared: the lower limit (20 ms) of SUSs at all the three syllabic rates is slightly longer than twice of the fundamental frequency (F_0) period of the utterances (7~8 ms in average), while the upper limits are consistent with 1/5~1/6 of the average syllable durations. In other words, subjects were still able to achieve very high syllable accuracies when listening to sentences in discrete frames with the duration between the above two boundaries with respect to both F_0 and syllable duration respectively. These observations thus connect to the intrinsic characteristics of both Mandarin and Cantonese (tonal and largely syllable-timed) and deepen our understanding about the cognitive processes of perceptual integration of Chinese spoken languages: Consistent with the previous hypothesis [1] and study [2], we found that Mandarin and Cantonese speech signals can be perceived in a discrete fashion. However, this perceptual discretization is found to be influenced by the inherent features of Chinese spoken languages, compared to various European languages (like English) which are non-tonal and stress-timed.

[1] Poeppel, D. *et al*. (2008) Speech Perception at the Interface of Neurobiology and Linguistics. *Philosophical Transactions of the Royal Society, Biological Sciences* 363: 1071-1086.
[2] Saberi, K. & Perrott, D.R. (1999) Cognitive Restoration of Reversed Speech. *Nature* 398: 760.
[3] Benoit, C. *et al*. (1996) The SUS Test: A method for the assessment of text-to-speech synthesis intelligibility using Semantically Unpredictable Sentences, *Speech Communication* 18: 381-392.

Key Words: perceptual discretization; Mandarin; Cantonese; locally time-reversed speech; speech intelligibility