Effects of cognitive load on the categorical perception of Mandarin tones

Yan Feng,[a#] Yaru Meng,[b#] Hanfei Li,[c] Gang Peng[a,c*]

[a]*Research Centre for Language, Cognition, and Neuroscience, Department of Chinese and Bilingual Studies, The Hong Kong Polytechnic University, Kowloon, Hong Kong SAR*

[b]*Department of Chinese Language and Literature, East China Normal University, Shanghai, China*

[c]*Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China*

[#]The first two authors contributed equally to this study.

[*]Corresponding author: Gang Peng

E-mail: gpengjack@gmail.com

Address: Department of Chinese and Bilingual Studies, The Hong Kong Polytechnic University, 11 Yuk Choi Road, Kowloon, Hong Kong SAR.

Conflict of Interest Statement:

There are no conflicts of interest.

Funding Statement:

This research was partly supported by grants from the General Research Fund of Research Grants Council of Hong Kong (15607518) and the National Natural Science Foundation of China (11974374) awarded to Gang Peng.

Abstract

Purpose: This study investigated the effect of cognitive load (CL) on the categorical perception (CP) of Mandarin lexical tones to discuss the application of the generalized pulse-skipping hypothesis. This hypothesis assumes that listeners might miss/skip temporal pulses and lose essential speech information due to CL, which consequently affects both the temporal and spectral dimensions of speech perception. Should CL decrease listeners' pitch sensitivity and impair the distinction of tone categories, this study would support the generalized pulse-skipping hypothesis.

Method: Twenty-four native Mandarin-speaking listeners were recruited to complete a dual-task experiment where they were required to identify or discriminate tone stimuli while concurrently memorizing six Chinese characters or graphic symbols. A no-load condition without a memory recall task was also included as a baseline condition. The position of categorical boundary, identification slope, between-/within-category discrimination, and discrimination peakedness were compared across the three conditions to measure the impact of CL on tone perception. The recall accuracy of Chinese characters and graphic symbols was used to assess the difficulty of memory recall.

Results: Compared to the no-load condition, both load conditions showed a boundary shift to Tone 3, shallower identification slope, poorer between-category discrimination, and lower discrimination peakedness. Within-category discrimination was negatively affected by CL in the graphic symbol condition only, not in the Chinese character condition.

Conclusions: CL degraded listeners' sensitivity to subtle F0 changes and impaired CP of Mandarin lexical tones. This provides support for the generalized pulse-skipping hypothesis. Besides, the involvement of lexical information modulated the effect of CL.

Introduction

Speech perception is an important stage in the human speech chain. It enables the speaker to understand his/her speech as well as that of others, which is a foundation of social communication (Denes & Pinson, 2015). As an active cognitive procedure, the rapid processing of continuous speech information places a high demand on resources such as focused attention and the efficient manipulation of information held in working memory (Heald & Nusbaum, 2014). Speech perception in natural daily communication is more cognitively demanding because listeners typically need to complete multiple concurrent processing tasks. The human capacity for cognitive processing is limited. Such additional cognitive load (CL) might hinder speech perception through competition for limited cognitive resources, given that a finite pool of memory and attention is simultaneously shared by different mental processes. Distinct from the perceptual load which arises from signal distortion (e.g., speech in a noisy environment), CL in this context refers to the demand on cognitive resources (e.g., working memory and attention) as a result of dual-task processing (Lavie, 2005; Lavie, 2010; Mattys & Wiget, 2011). The interference effect of CL on speech perception has been well documented in the literature from several viewpoints. For example, researchers have shown that CL impairs not only phoneme monitoring during word identification (Wurm & Samuel, 1997), but also word segmentation in a continuous speech stream (Fernandes et al., 2010). CL has also been observed to interfere with sentence recognition (Hunter & Pisoni, 2018).

Among speech perception, categorical perception (CP) is one of the basic speech processing abilities. It enables the listener to perceive continuous acoustic cues in auditory input to discrete phonological categories (Liberman et al., 1957). Unlike commonly spoken languages such as English and French, Mandarin is a tone language with four lexical tones differing in fundamental frequency (F0). For example, syllable /mɑ/ means "妈 mother" with the high-level tone (Tone 1), "麻 hemp" with the mid-rising tone (Tone 2), "马 horse" with the falling-rising tone (Tone 3), and "骂 scold" with the falling tone (Tone 4) (Chao, 1968). Although the four example words given here share the same consonant and vowel, the different lexical tones determine their different meanings. In tone languages, the tones distinguish different lexical meanings in the same way as vowels and consonants. Thus, F0 in Mandarin not only conveys pitch information in the general acoustic domain but also phonological and lexical information in the language-specific domain. It has been well documented that the perception of lexical tones in native Mandarin speakers is categorical (e.g., Peng et al., 2010; Wang, 1976). In recent decades, many studies have focused on tone perception in adverse conditions prevalent in daily communication, such as the impact of the perceptual load caused by signal degradation (e.g., duration – Wang et al., 2017 and noise – Tang et al., 2018). CL arising from dual tasks has been shown to disrupt speech production and perception in previous research with populations speaking nontonal languages (e.g., Jones et al., 2012; MacPherson, 2019). However, limited attempts have been made to investigate the impact of CL on the perception of lexical tones in listeners speaking tone languages. Thus, this study aims to examine how CL affects Mandarin lexical tone perception in native listeners.

Previous studies have assessed how the CP of segments is affected by the interference of CL with visual information. For example, Casini and colleagues (2009) examined the effect of visual load on the auditory perception of French vowel duration (*cache* vs. *cage*) using a dual-task paradigm. Participants were asked to simultaneously identify sound stimuli and make quick responses to the color of a light-emitting diode. Researchers found that the concurrent visual task led to a shift of boundary position to long vowels, indicating that vowels were perceived as shorter under divided attention. Such a "shrinking of time" mechanism is also supported by the metaanalysis of Block and colleagues (2010) which shows that CL leads to an underestimation of duration in speech perception. The basis of this mechanism is the domain-general timer hypothesis (e.g., Coull et al., 2004; Rammsayer & Ulrich, 2005). This assumes that the estimate of signal duration is attention-based and depends on the continuous encoding of accumulated temporal pulses in sensory input. Due to the competition for attentional resources caused by CL, listeners might miss/skip temporal pulses and lose essential speech information, which consequently hinders speech perception. The decreased temporal sampling rate of sensory input caused by CL has also been termed the pulse-skipping hypothesis (Chiu et al., 2019).

As the pulse-skipping hypothesis is specific to the temporal dimension of speech perception, it remains unclear whether the perception of spectral cues would also be influenced by CL. Chiu and colleagues (2019) further examined the pulse-skipping hypothesis in terms of the duration, intensity, and F0 dimensions of English vowel perception, and found CL had a negative impact on all three dimensions. Thus, they proposed a generalized pulse-skipping hypothesis in which CL interferes not only with the perception of the time-dependent dimension of speech directly, but also indirectly with its time-independent dimensions. Compared with the pulse-skipping hypothesis, the generalized pulse-skipping hypothesis extends the scope of the effect of CL to both the temporal and spectral domains of speech perception. The biological basis of the generalized pulse-skipping hypothesis is temporal integration. This denotes the basic ability of the central auditory system to accumulate acoustic signals over time to improve speech perception in both the temporal and spectral domains (Moore, 2007). That is to say, the more temporal pulses of sensory input, the better the perception of temporal and spectral cues. Some literature has also demonstrated better lexical tone perception with longer signal duration (e.g., Chen et al., 2017; Wang et al., 2017). Based on the concept of temporal integration, although CL only affects the detection of temporal pulses, it might impair speech perception in both the temporal and spectral domains. The so-called generalized pulse-skipping hypothesis provides a potential mechanism underlying the effect of CL on speech perception in the spectral domain, by proposing the loss of input pulses may decrease listeners' sensitivity to spectral changes.

However, the F0 manipulated in the work of Chiu and colleagues (2019) was an acoustic cue without phonological information. Many studies have demonstrated a difference in F0 perception at the phonological and acoustic levels (Peng et al., 2010; Wang, 1976): Lexical tone perception in Mandarin is categorical because of the distinctive tone categories in phonological memory, while pure-tone F0 perception is less categorical. In comparison to pure tones, the extra encoding of phonological and lexical information in lexical tones increases cognitive demand, which might also modulate the effect of CL to some extent. Furthermore, Mattys and Wiget (2011) investigated the interference of visual CL in the CP of voiced and voiceless consonants in English (/gɪ/ *vs.* /kɪ/) with different voice onset times. They observed that CL interfered with early auditory processing and increased listeners' reliance on lexical knowledge. To make a better comparison with previous research such as the work of Casini and colleagues (2009), and Mattys and Wiget (2011) focusing on the effect of CL on the CP of segments with different temporal cues, it is necessary to explore the impact of CL on the CP of lexical tones. The latter is a suprasegment differing in spectral cue, which has been less widely studied. Such work may provide further evidence for the generalized pulse-skipping hypothesis.

It is worth noting that the interpretation of empirical research findings should also take methodologies into consideration. Phoneme identification requires listeners to encode one stimulus in the early auditory processing stage and map it to a mental phonological category to make a judgment. Phoneme discrimination, however, means listeners need to encode and maintain several sound stimuli, and then compare them to make a decision. The different manipulation of auditory input in phoneme identification and discrimination may lead to different observations of CL effects in behavioral experiments. For example, Mattys and Wiget (2011) failed to find an effect of CL on the steepness of the identification function, but observed a negative effect of CL on between-category discrimination. The different encoding processes of primary and secondary tasks also modulate the effect of CL. Mitterer and Mattys (2017) manipulated the sequence of primary and secondary tasks and found no significant decrease of consonant discrimination in either “early” or “late” load conditions, indicating that memory load was not sufficient to impair CP. They further asked participants to memorize two faces sequentially as they heard four sounds. A significant effect of CL was observed for discrimination accuracy. This indicated that the CL effect was caused by the continuous and concurrent encoding of auditory and visual new information.

Furthermore, the representation format of CL seems to play an important role in modulating the effect of CL. Chiu and colleagues (2019) contrasted the effects of two types of visual load (i.e., nonwords and images) on speech discrimination. They found a larger negative effect of CL for nonwords due to competition in phonological encoding between primary and secondary tasks. Previous research has also shown that the difficulty of the secondary task might modulate the effect of CL. Jones and colleagues (2012) showed that the disruptive effect of CL on phonological processing varied with the difficulty of the concurrent memory task. Consistent with this, Mattys and colleagues (2014) examined the relationship between the difficulty of visual search task and word/nonword discrimination across nasal (/n/ and /m/) and liquid (/l/ and /r/) phonemes. They found a tradeoff effect, whereby the accuracy of discrimination decreased linearly as the difficulty of the concurrent secondary task increased.

In this study, using a dual-task paradigm, we aim to investigate the effects of visual CL on the CP of Mandarin lexical tones and examine the generalized pulse-skipping hypothesis. If listeners' pitch sensitivity was impaired in load conditions, this study would provide evidence for the application of the generalized pulse-skipping hypothesis to the spectral dimension of speech perception at the phonological level. If not, this hypothesis may be specific to the perception of spectral cues at the acoustic level as shown in Chiu and colleagues (2019). Besides, as previous studies showed several modulators of CL effect such as the representation format of CL, we also tested the differential impact of CL in terms of the involvement of lexical information. We used a graphic symbol load condition that required mainly visual encoding and a Chinese character load condition that required both visual and lexical encoding. Wurm and Samuel (1997) found that fewer cognitive processing resources were required for stimuli involving more lexical information. If so, we would expect to observe a larger interference effect in the graphic symbol load than the character load condition. Should the dual task lead to strong competition for limited phonological and lexical processing resources, as found by Chiu and colleagues (2019), we would expect to find a worse perceptual performance in the character load than in the graphic symbol load condition.

Method

Participants

Twenty-four native speakers of Mandarin (11 males, mean age = 22.58 years, SD = 1.89) were recruited for this study. All the participants came from northern China and were right-handed. None of them had formal musical experience or a history of speech or hearing impairment according to their self-reports. A consent form was obtained from each participant using a protocol approved by the Human and Animal Experiment Ethics Committee of the Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences.

Materials

Two words frequently used in Chinese were chosen for this study: "衣" (/i/ with Tone 1, "clothes") and "以" (/i/ with Tone 3, "with"). There were two reasons for the selection of Tone 1 and Tone 3. Firstly, the categorical perception of Tone 1-3 is less explored than other tone pairs (e.g., Tone 1-2, Tone 1-4, Tone 2-3, or Tone 2-4) in previous studies (Hallé et al., 2004; Peng et al., 2010; Wang, Yang, Liu, 2017; Wang, Yang, Zhang, et al., 2017; Xu et al., 2006). Secondly, compared with Tone 1, Tone 3 is associated with more F0 perturbation. There are large differences in pitch direction, height, and slope between Tone 1 and Tone 3. Especially, the difference of pitch height between Tone 1 and Tone 3 is the largest, compared with all other tone pairs (e.g., Peng, 2006; Tupper et al., 2020). This provided an appropriate platform to explore the CL effect on listeners' pitch sensitivity. The two words ("衣" and "以") were uttered by a male native Mandarin speaker as clearly and naturally as possible. The speech samples were recorded using Praat (Boersma & Weenink, 2019) with a 44.1 kHz sampling rate and 16-bit resolution. The duration of both words was normalized to 300 ms using PSOLA in Praat. There were nine stimuli in the tone continuum which were generated by the TANDEM-STRAIGHT software (Kawahara & Morise, 2011). The acoustic variation in the tone continuum was only F0. All nine stimuli were 300 ms in duration and presented at 70 dB SPL. The F0 contours of the tone continuum from Tone 1 (sound stimulus 1) to Tone 3 (sound stimulus 9) are shown in Figure 1.

[Figure 1]

Two types of visual stimuli, Chinese characters and graphic symbols, were used in the visual recall tasks as the load conditions, as shown in Table 1. The Chinese characters involved both visual and lexical information, while the graphic symbols mainly included visual information. All of the characters and graphic symbols selected are frequently used in daily life.

Procedure

All participants were seated in a soundproofed room. In the no-load condition as the baseline condition, participants were required to complete a tone identification task and a discrimination task. Across two load conditions, they performed a dual-task paradigm, in which they needed to memorize six visual stimuli and simultaneously identify or discriminate sound stimuli. The order of the three conditions was counterbalanced across participants. A laptop with E-Prime 2.0 was used to conduct the experiment. A practice block with feedback was performed before the testing block began. In the practice block, only sound stimulus 1 (typical "衣") and stimulus 9 (typical "以") were presented for the identification task. Feedback on the identification, discrimination, and recall tasks was given to familiarize participants with the experiment.

In the identification task, participants were asked to identify the sound from two alternative forced choices within 2,000 ms by pressing key "1" on a keyboard when they judged the sound they heard as "衣" (Tone 1) and key "2" when they judged the sound as "以" (Tone 3). In the testing block, there were five repetitions of each stimulus in the continuum, and all stimuli were presented randomly. In the discrimination task, participants were instructed to judge whether the two stimuli they heard were different within 2,000 ms and press key "1" for "the same" and key "2" for "different". The interstimulus interval between the two sound stimuli was 500 ms. There were 23 pairs of stimuli for the testing block, including 14 pairs consisting of two different stimuli separated by two steps in both the forward (1-3, 2-4, … 6-8, 7-9) and reverse (3-1, 4-2, … 8-6, 9-

7) order, as well as nine pairs consisting of the same stimuli (1-1, 2-2, … 8-8, 9-9). Each pair was repeated five times and presented randomly.

To match the task complexity of the two visual load conditions, the set sizes of both recall tasks were fixed at six. In each trial, six randomly selected visual stimuli were numbered in order and presented horizontally in the middle of a screen. The visual stimuli appeared for 2,000 ms simultaneously with the onset of a sound stimulus in the identification task or the onset of the first sound stimulus in the discrimination task. Before each block commenced, participants were explicitly told which type of stimuli (i.e., graphic symbol in the symbol condition, or Chinese character in the character condition) they needed to recall. Participants were asked to memorize these visual stimuli while aurally perceiving the sound stimuli. After the tone identification or discrimination task, they were required to recall the order number of a visual stimulus randomly chosen from the middle four of the visual stimuli presented.

Data analysis

For the identification curve, identification score was defined as the percentage of responses for Tone 1. A logistic regression equation was used to calculate the slope of the identification function (Xu et al., 2006):

$$log_e\left(\frac{P_I}{1-P_I}\right) = b_0 + b_1 x \quad (1)$$

In the Equation (1), P_I refers to the identification score, and b_0 the intercept of the identification function. The coefficient b_1 denotes the slope of the identification function indicating the sharpness of the categorical boundary. The steeper the slope, the clearer the listeners' distinction between tone categories. The x denotes stimulus number (1-9) in the tone continuum. The category boundary position (x_{cb}) was estimated when P_I = 0.5 (Xu et al., 2006). The logistic regression in Equtation (1) becomes the Equation (2) when P_I = 0.5 (Chen et al., 2017):

$$x_{cb} = \frac{-b_0}{b_1} \tag{2}$$

For the tone discrimination, all pairs of stimuli were divided into seven comparison units consisting of four types of pairs (AA, BB, AB, and BA). We calculated the d-prime score for each comparison unit as an index of discrimination sensitivity (Macmillan & Creelman, 1991):

$$d' = z(H) - z(F) \tag{3}$$

where d' refers to the d-prime score, $z(H)$ the z-transformed hit rate, and $z(F)$ the z-transformed false alarm rate. The hit rate was defined as the percentage of different responses to different pairs (i.e., AB and BA), and the false alarm rate as the percentage of different responses to the same pairs (i.e., AA and BB). According to the category boundary position estimated by equation (2) (e.g., x_{cb} = 5.5), between-category discrimination sensitivity was defined as the d-prime score of the comparison units corresponding to the category boundary (e.g., the average $d'$of comparison unit 4-6 and 5-7). Within-category discrimination sensitivity was defined as the d-prime score of the remaining comparison units (e.g., the average $d'$of comparison unit 1-3, 2-4, 3-5, 6-8, and 7-9). Peakedness of discrimination was calculated as the difference between within- and between-category discrimination. The peakedness illustrates the categorical discrimination ability to enhance between-category and inhibit within-category discrimination (Xu et al., 2006). Furthermore, in the two load conditions, the recall accuracies of the Chinese characters and graphic symbols during the identification and discrimination tasks were calculated.

Calculations of the identification score, identification slope, category boundary position, between-/within-category discrimination, peakedness of discrimination, and recall accuracy were all conducted on an individual basis. One outlier (labeled with an asterisk in Figure 3) was excluded from all statistical analyses. A repeated-measures ANOVA was conducted using SPSS v20.0 to test the effect of CL on perceptual performance. The Greenhouse-Geisser correction was used when appropriate.

Results

Tone identification

Figure 2 (A) shows the identification score for /i/ with Tone 1 across the three conditions. The category boundary position and the slope of the identification function in the three conditions are shown in Table 2. To investigate whether the CL effect would impair listeners' sensitivity to F0 perturbation, a repeated-measures ANOVA with *condition* as a within-subject factor was conducted on the position of category boundary. The results demonstrated a significant main effect of *condition* on the position of category boundary (F (2,44) = 6.173, p = .004). Pairwise comparison showed that the category boundary in the graphic symbol load condition significantly shifted to Tone 3 compared with that in the no-load condition (p = .003). The category boundary in the character load condition also showed a marginal shift to Tone 3 in comparison to the no-load condition (p = .056). These results indicated that more sound stimuli had been perceived as Tone 1 with the dual-task interference. No significant difference in category boundary was found between the graphic symbol and character load conditions (p = 1.000).

[Figure 2]

To examine the effect of CL on the sharpness of category boundary, a repeated-measures ANOVA with *condition* as a within-subject factor was calculated on identification slope. The results showed a significant main effect of *condition* on identification slope (F (2,44) = 5.211, p = .009). There was a shallower slope for the two load conditions (character: p = .044; symbol: p = .015) than in the no-load condition. This indicated a more ambiguous distinction between the two contrasting tones with the interference of the dual task. However, no significant difference was found in the slope of tone identification between the graphic symbol and Chinese character load conditions (p = 1.000).

Tone discrimination

Figure 2 (B) presents the d-prime scores for the between- and within-category discrimination in the three conditions. To explore the effect of CL on discrimination sensitivity, a two-way repeated-measures ANOVA was conducted with *category type* and *condition* as the within-subject factors. The results showed significant main effects of *condition* (F(2, 44) = 13.453, p < .001) and *category type* (F(1, 22) = 53.231, p < .001), and a significant interaction effect between *category type* and *condition* (F(2, 44) = 6.000, p = .005). A simple main effect analysis indicated that in each condition, there was a significant difference between within- and between-category discrimination (all p < .001; see Figure 2 (B)). There was also a significant difference in between-category discrimination between the no-load and the Chinese character load conditions (p = .033), and a significant difference between the no-load and graphic symbol load conditions (p < .001). These results indicated that both visual loads negatively affected between-category discrimination. No significant difference was found between the two load conditions (p = .265). On the other hand, there was a significant difference in within-category discrimination between the no-load and graphic symbol load conditions (p = .031). A marginal difference was also observed between the two load conditions (p = .068). No significant difference was found between the no-load and character load conditions (p = 1.000).

Furthermore, a repeated-measures ANOVA with *condition* as the within-subject factor revealed a significant effect of CL on peakedness ($F(2, 44) = 6.000$, $p = .005$). Pairwise comparison indicated a significantly higher peakedness for the no-load condition (mean = 1.634) than for either the Chinese character (mean = 0.912) ($p = .042$) or the graphic symbol (mean = 0.902) ($p = .034$) load conditions. This indicated that discrimination ability was impaired under both load conditions, with no significant difference between them ($p = 1.000$)

Recall accuracy

The distributions of the recall accuracies in the two load conditions during the identification and discrimination tasks are presented in Figure 3. To compare the effect of stimulus type (character or graphic symbol), a two-way repeated-measures ANOVA was calculated with *task* and *condition* set as the within-subject factors. The results showed a significant main effect of *condition* ($F(1, 22) = 128.009$, $p < .001$) and a marginally significant main effect of *task* ($F(1, 22) = 4.267$, $p = .051$). There was no interaction effect between *task* and *condition* ($F(1, 22) = 1.844$, $p = .188$). During both the identification and discrimination tasks, the recall accuracies of Chinese characters were higher than those of graphic symbols (both $p < .001$). This indicated that the former was much easier than the latter because of the involvement of lexical information.

[Figure 3]

To assess the association between tone perception and character/symbol retention, we also conducted a Pearson correlation analysis. The results demonstrated a marginally significant negative correlation between the recall accuracy of Chinese characters during the identification task and the identification slope ($r = -.411$, $p = .051$). This revealed that the participants who had achieved higher recall accuracy also showed a clearer distinction between Tone 1 and Tone 3. There was also a positive correlation between the recall accuracy of Chinese characters during the discrimination task and both the peakedness of discrimination (r = .511, p = .013) and between-category discrimination (r = .546, p = .007), indicating that those who performed better in character recall also had a better between-category discrimination ability. However, no significant correlation was observed between the recall accuracy of graphic symbols and perceptual performance (all $p > .1$).

Discussion

This study examined the different effects of the CL derived from visually recalling Chinese characters and graphic symbols on the identification and discrimination of Mandarin lexical tones. The central question of this study is whether the generalized pulse-skipping hypothesis could be applicable to the perception of spectral cues at the phonological level. Our results demonstrated an effect of CL on boundary shift, identification slope, between-category discrimination, and discrimination peakedness across two load conditions, providing support for the generalized pulse-skipping hypothesis. This hypothesis suggests that due to the competition for limited attentional and memory resources between two concurrent tasks, a loss of input pulses degrades listeners' sensitivity to subtle F0 changes (Chiu et al., 2019). Therefore, larger F0 changes were required for sound stimuli to be perceived as Tone 3 (it is noteworthy that Mandarin Tone 1 is associated with less F0 perturbation than Tone 3). Moreover, we found a significant CL effect on within-category discrimination only in the graphic symbol condition. The more significant interference effect in the graphic symbol condition suggested that the involvement of lexical information might modulate the CL effect.

Graphic symbol load significantly impaired both tone identification and discrimination, while the character load only marginally shifted the category boundary to Tone 3 and had no effect on within-category discrimination. It seems that the graphic symbol condition showed a more significant interference effect than the character load condition. This is inconsistent with the results of Chiu and colleagues (2019), where nonword processing elicited a larger CL effect on F0 perception than did image recall. Firstly, according to Wurm and Samuel (1997), the involvement of lexical information in the character recall task led to less cognitive resource recruitment than for the graphic symbol recall task. Thus, tone perception in the character load condition could utilize more cognitive processing resources than in the graphic symbol load condition. This may be one of the reasons why tone perception was less affected by character load. Secondly, the involvement of lexical information modulated the difficulty of the two recall tasks, and further influenced the CL effect on tone perception. A significantly higher recall accuracy was found in the Chinese character condition than the graphic symbol condition. This suggested that visual stimuli with lexical information are memorized much more easily. Consistent with the results of Jones and colleagues (2012) and Mattys and colleagues (2014) where a tradeoff effect was observed, we found that increased task difficulty degraded tone perception. However, Chiu and colleagues (2019) showed that the difficulty of nonword and image recall tasks was comparable. This may explain the inconsistency of CL effects involving different encoding modalities between our study and that of Chiu and colleagues (2019). Thus, it is necessary to take into consideration the impact of both representation format and task difficulty of CL when interpreting the generalized pulse-skipping hypothesis.

Significant correlations were observed only between the recall accuracy of the Chinese character and the performance of lexical tone perception. This emphasizes the need to consider the impact of representation format of CL in future research. Previous research on Chinese character recognition (e.g., Perfetti & Tan, 1998; Tzeng et al., 1977) has found that characters in written form can be transferred to the phonological store through the articulatory control process (Baddeley, 2007; Baddeley & Hitch, 1974). The character retention interfered with lexical tone perception via competition for the limited capacity of the phonological store in the phonological loop. The higher recall accuracy might indicate a greater phonological store capacity, thus explaining the correlation with the better performance in the CP task. Moreover, those with better recall performance might be better able to utilize language knowledge. Tone perception also requires the involvement of language knowledge, and CL enhances listeners' reliance on such knowledge (Mattys & Wiget, 2011). Therefore, those with better recall accuray had better CP performance.

Although the tone identification and discrimination tasks required different manipulation of auditory input, the interference effect was found on both of the primary tasks. Inconsistent with the work of Mattys and Wiget (2011), the slope of tone identification in this study was significantly shallower when CL was present. A possible reason for this is that recall tasks in this study were more difficult than those used by Mattys and Wiget (2011), causing a larger interference effect on tone identification. In this study, for example, the recall accuracy of graphic symbols (mean = 47.39%) was much lower than that in Mattys and Wiget (2011) where the accuracy of the visual-search task was around 80%. Another potential explanation is the different phonemes used in the two studies. This study focused on the perception of suprasegment, while Mattys and Wiget (2011) investigated the perception of segment. More work is necessary to figure out the controversy about the CL effect on speech perception in future research.

Conclusion

In this paper, we have examined the effects of dual-task induced CL on the CP of Mandarin lexical tones to identify whether the generalized pulse-skipping hypothesis is applicable to the perception of spectral cues. The results have shown that CL impaired listeners' pitch sensitivity and negatively affected the CP of Mandarin Tone 1 and Tone 3 due to competition for attention and memory resources, providing support for the generalized pulse-skipping hypothesis. Across two load conditions, impaired tone identification, discrimination peakedness, and between-category discrimination were observed in comparison to the no-load condition. However, the CL effect was only found for within-category discrimination in the graphic symbol condition, not the Chinese character load condition. This indicates that the involvement of lexical information in character recall lessened the CL effect. The correlations between tone perception and character recall indicate that those with higher character recall accuracy had better phonological store capacity and performed better in lexical tone perception.

Acknowledgments

This research was partly supported by grants from the General Research Fund of Research Grants Council of Hong Kong (15607518) and the National Natural Science Foundation of China (11974374) awarded to Gang Peng.

References

Baddeley, A. (2007). *Working Memory, Thought, and Action*. New York, NY: Oxford University Press.

Baddeley, A. D., & Hitch, G. (1974). Working memory. In G. H. Bower (Eds), *Psychology of Learning and Motivation* (Vol. 8, pp. 47-89). Cambridge, MA: Academic Press.

Block, R. A., Hancock, P. A., & Zakay, D. (2010). How cognitive load affects duration judgments: A meta-analytic review. *Acta Psychologica, 134*(3), 330-343.

Boersma, P., & Weenink, D. (2019). Praat: Doing phonetics by computer (version 6.0.49). http://www.praat.org (Last viewed February 24, 2019).

Casini, L., Burle, B., & Nguyen, N. (2009). Speech perception engages a general timer: Evidence from a divided attention word identification task. *Cognition, 112*(2), 318–322.

Chao, Y. R. (1968). *A grammar of spoken Chinese*. Berkeley, CA: University of California Press.

Chen, S., Zhu, Y., & Wayland, R. (2017). Effects of stimulus duration and vowel quality in cross-linguistic categorical perception of pitch directions. *PloS One, 12*(7), e0180656.

Chiu, F., Rakusen, L. L., & Mattys, S. L. (2019). Cognitive load elevates discrimination thresholds of duration, intensity, and f_0 for a synthesized vowel. *The Journal of the Acoustical Society of America, 146*(2), 1077–1084.

Coull, J. T., Vidal, F., Nazarian, B., & Macar, F. (2004). Functional anatomy of the attentional modulation of time estimation. *Science, 303*(5663), 1506-1508.

Denes, P., & Pinson, E. (2015). *The speech chain : The physics and biology of spoken language* (2nd edition). Long Grove, Illinois: Waveland Press.

Fernandes, T., Kolinsky, R., & Ventura, P. (2010). The impact of attention load on the use of statistical information and coarticulation as speech segmentation cues. *Attention, Perception, & Psychophysics, 72*(6), 1522-1532.

Hallé, P. A., Chang, Y. C., & Best, C. T. (2004). Identification and discrimination of Mandarin Chinese tones by Mandarin Chinese vs. French listeners. *Journal of Phonetics, 32*(3), 395-421.

Heald, S., & Nusbaum, H. C. (2014). Speech perception as an active cognitive process. *Frontiers in Systems Neuroscience, 8*, 35.

Hunter, C. R., & Pisoni, D. B. (2018). Extrinsic cognitive load impairs spoken word recognition in high-and low-predictability sentences. *Ear and Hearing, 39*(2), 378-389.

Jones, R. M., Fox, R. A., & Jacewicz, E. (2012). The effects of concurrent cognitive load on phonological processing in adults who stutter. *Journal of Speech, Language, and Hearing Research, 55*(6), 1862-1875.

Kawahara, H., & Morise, M. (2011). Technical foundations of TANDEM-STRAIGHT, a speech analysis, modification and synthesis framework. *Sadhana, 36*(5), 713-727.

Lavie, N. (2005). Distracted and confused?: Selective attention under load. *Trends in Cognitive Sciences, 9*(2), 75-82.

Lavie, N. (2010). Attention, distraction, and cognitive control under load. *Current Directions in Psychological Science, 19*(3), 143-148.

Liberman, A. M., Harris, K. S., Hoffman, H. S., & Griffith, B. C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology, 54*(5), 358–368.

Macmillan, N. A., & Creelman, C. D. (1991). *Detection Theory: A User's Guide*. Cambridge: Cambridge University Press.

MacPherson, M. K. (2019). Cognitive load affects speech motor performance differently in older and younger adults. *Journal of Speech, Language, and Hearing Research, 62*(5), 1258-1277.

Mattys, S. L., Barden, K., & Samuel, A. G. (2014). Extrinsic cognitive load impairs low-level speech perception. *Psychonomic Bulletin & Review, 21*(3), 748–754.

Mattys, S. L., & Wiget, L. (2011). Effects of cognitive load on speech recognition. *Journal of Memory and Language, 65*(2), 145–160.

Mitterer, H., & Mattys, S. L. (2017). How does cognitive load influence speech perception? An encoding hypothesis. *Attention, Perception, & Psychophysics, 79*(1), 344–351.

Moore, B. C. J. (2007). Temporal Resolution and Temporal Integration. In *Cochlear Hearing Loss: Physiological, Psychological and Technical Issues* (2nd edition, pp. 117-141). West Sussex, England: John Wiley & Sons.

Peng, G. (2006). Temporal and tonal aspects of Chinese syllables: A corpus-based comparative study of Mandarin and Cantonese. *Journal of Chinese Linguistics, 34*(1), 134-154.

Peng, G., Zheng, H. Y., Gong, T., Yang, R. X., Kong, J. P., & Wang, S. Y. (2010). The influence of language experience on categorical perception of pitch contours. *Journal of Phonetics, 38*(4), 616–624.

Perfetti, C. A., & Tan, L. H. (1998). The time course of graphic, phonological, and semantic activation in Chinese character identification. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 24*(1), 101-118.

Rammsayer, T. & Ulrich R. (2005). No evidence for qualitative differences in the processing of short and long temporal intervals. *Acta Psychologica, 120*, 141-171.

Tang, W., Wang, X. J., Li, J. Q., Liu, C., Dong, Q., & Nan, Y. (2018). Vowel and tone recognition in quiet and in noise among Mandarin-speaking amusics. *Hearing Research, 363*, 62-69.

Tupper, P., Leung, K., Wang, Y., Jongman, A., & Sereno, J. A. (2020). Characterizing the distinctive acoustic cues of Mandarin tones. *The Journal of the Acoustical Society of America, 147*(4), 2570-2580.

Tzeng, O. J., Hung, D. L., & Wang, W. S.-Y. (1977). Speech recoding in reading Chinese characters. *Journal of Experimental Psychology: Human Learning and Memory, 3*(6), 621-630.

Wang, W. S.-Y. (1976). Language change. *Annals of the New York Academy of Science, 208*, 61-72.

Wang, Y., Yang, X., & Liu, C. (2017). Categorical perception of Mandarin Chinese tones 1–2 and tones 1–4: effects of aging and signal duration. *Journal of Speech, Language, and Hearing Research, 60*(12), 3667-3677.

Wang, Y., Yang, X., Zhang, H., Xu, L., Xu, C., & Liu, C. (2017). Aging effect on categorical perception of Mandarin tones 2 and 3 and thresholds of pitch contour discrimination. *American Journal of Audiology, 26*(1), 18-26.

Wurm, L. H., & Samuel, A. G. (1997). Lexical inhibition and attentional allocation during speech perception: Evidence from phoneme monitoring. *Journal of Memory and Language*, *36*(2), 165–187.

Xu, Y., Gandour, J. T., & Francis, A. L. (2006). Effects of language experience and stimulus complexity on the categorical perception of pitch direction. *The Journal of the Acoustical Society of America, 120*(2), 1063–1074.

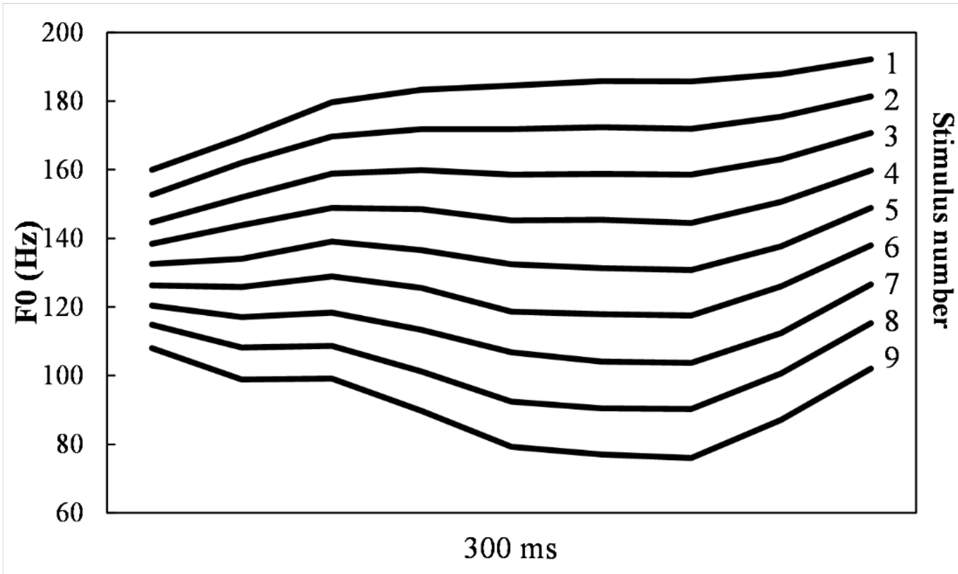

Figure 1. F0 contours in the tone continuum.

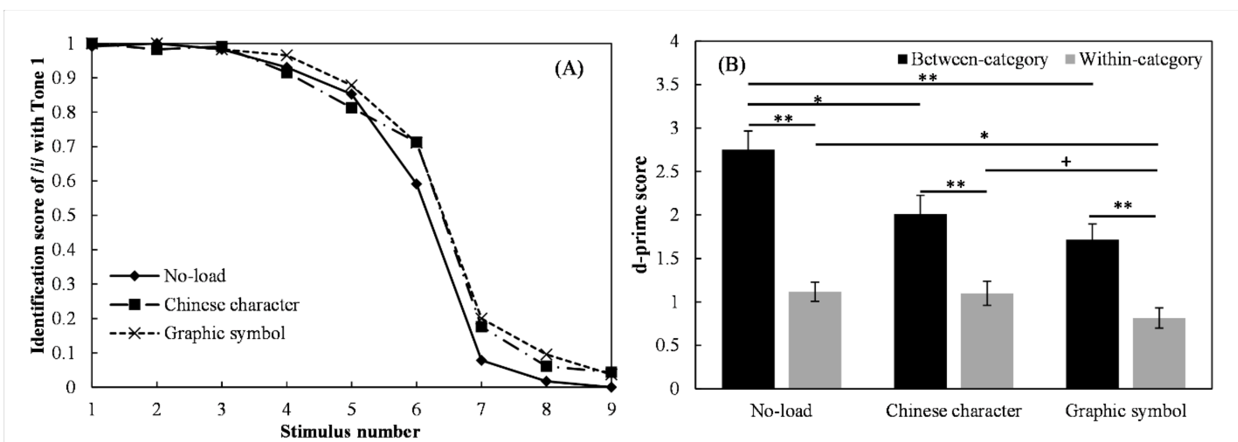

Figure 2. (A) Identification curves of tone perception across the three conditions. (B) The d-prime scores for within- and between-category discrimination of tone across the three conditions. Error bars: ±1 standard error. +: $0.05 < p < 0.07$; *: $0.01 < p < 0.05$; **: $p < 0.001$, two-tailed.

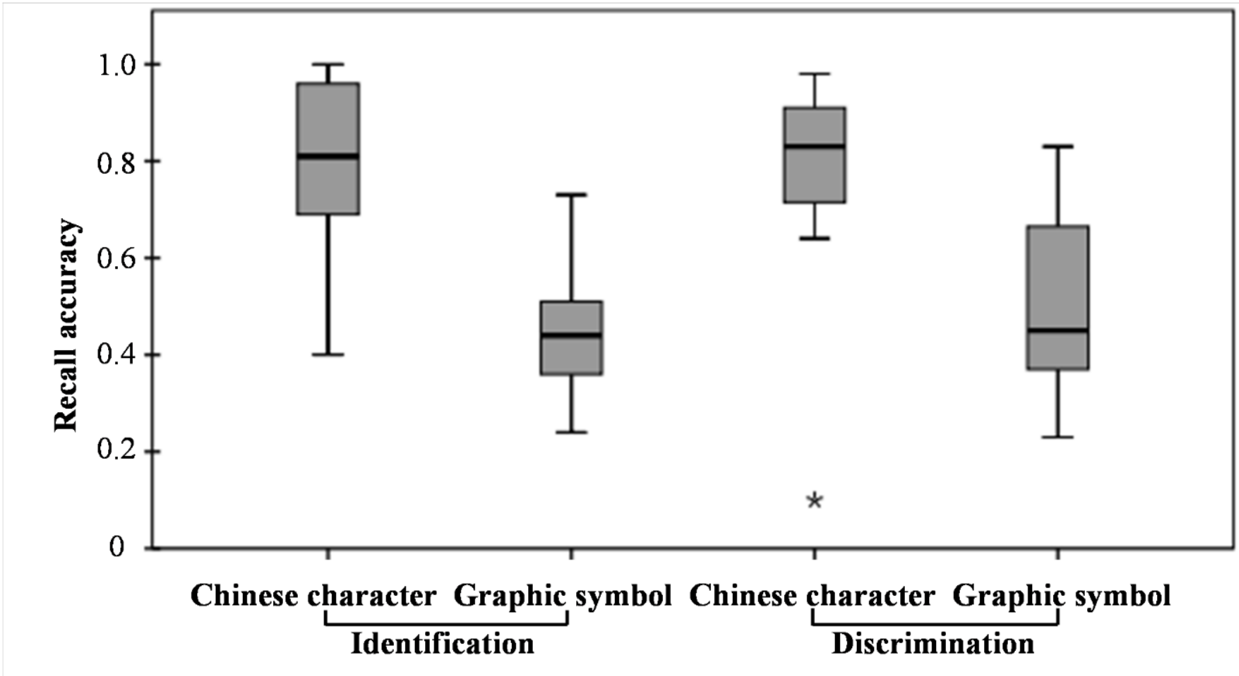

Figure 3. Recall accuracy for Chinese characters and graphic symbols during the identification and discrimination tasks.

Table 1. Chinese characters and graphic symbols in the two load conditions.

Type of CL	**Stimuli**	**Load information**	
		Visual	Lexical
Chinese character	八, 高, 金, 当, 朱, 迷, 和, 亭, 柔, 才, 水, 冷, 产, 孔, 雪, 算, 份, 破, 去, 念	+	+
Graphic symbol	☆, 卍, ♂, ○, ◎, ⊿, ♫, ♯, ¤, ⊙, ♪, ♀, ※, ☺, Ю, ©, ▽, ◇, △, □	+	-

Table 2. Position of the category boundary and slope of the identification function across the three conditions.

Condition	**Boundary position (SD)**	**Slope (SD)**
No-load	5.702 (0.531)	-1.378 (0.139)
Chinese character	6.136 (0.945)	-1.241 (0.222)
Graphic symbol	6.279 (0.903)	-1.254 (0.207)