Cantonese Spoken Word Retention by Speakers with and without Congenital Amusia: Implications from Phonological Similarity and Cognitive Load Effects

Xiao Wang [1], *Gang Peng*[2, 3]

[1] The Chinese University of Hong Kong
[2] The Hong Kong Polytechnic University, [3] Shenzhen Institutes of Advanced Technology
w.joycewang.x@gmail.com, gpengjack@gmail.com

Abstract

Success in spoken word processing relies not only on accurate word recognition but also the veracity with which words are maintained in memory. However, research on word retention is still scarce, especially in tonal languages and phonologically impaired populations. To address these gaps, the present study administered an auditory order recall task to native Cantonese speakers with and without amusia. Stimuli intrinsic (segmental similarity, suprasegmental similarity, and lexicality) and extrinsic (cognitive load) factors were manipulated. As expected, word recall was significantly impaired by increased cognitive load, and, similar to the retention of non-tonal languages, rime similarity exerted opposite effects on Cantonese real and pseudo word recollection. However, no reversal was observed in the suprasegmental dimension: Lexical tone similarity was persistently detrimental to recall accuracy. Tonal similarity Effects were also more robust against increased cognitive load than segmental similarity effects, implying differential weighting of suprasegmental and segmental cues in spoken word retention. These results are consistent with the view that distinct processing mechanisms exist for segments and tones. Besides, this study found it useful to combine accuracy and response time analyses. The potential of the combined analysis to capture the nature of phonological deficits is discussed in relation to amusia.

Index Terms: spoken word retention, phonological similarity, cognitive load, phonological deficit, amusia

1. Introduction

While lots of efforts have been dedicated to studying spoken word recognition, the maintenance of recognized words in memory remains poorly understood in the field of spoken language processing. This is especially the case with Chinese. Nonetheless, effects of cognitive load [1-3] and phonological similarity [4] have made it abundantly clear that memory traces can be confused and distorted. A primary aim of this study is, therefore, to explore Chinese spoken word retention with regard to stimuli intrinsic (e.g., phonological similarity) and extrinsic (e.g., cognitive load) variables.

1.1. The similarity and cognitive load effects

As one of the extensively studied phenomena in working memory (WM) literature, the phonological similarity effect (alternatively: PSE or similarity effect) has a lot to offer on the subject of spoken word processing. Narrowly put, PSE refers to the phenomenon that similar-sounding words are much harder to recall in the order of presentation than dissimilar words [4]. Defined broadly, it may also be used to refer to the benefits similarity brings on pseudoword recollection [7, 8, 10]. But what is particularly noteworthy about PSE is its linguistic basis. Namely, PSE stems from the confusion of phonological codes rather than sensory memory traces which are transient and prone to decay [4]. In line with this claim, PSE has been elicited independently of input modalities (visual, auditory) and modes of recollections (written, oral; but see [5]) [11, 12]. It has also been observed in delayed recall tasks, regardless of whether the delay was silent or noise-filled [7-9]. The fact that even recall errors respect phonological constraints [8, 9] speaks further to the linguistic nature of PSE.

More importantly, being phonologically grounded has afforded PSE the ability to tap into various levels of phonological recoding activities, including that of vowels [4], articulatory features [5], onsets [6, 10], rimes [6-8, 10], and lexical tones [9]. Moreover, since participants' attention is not directed to any particular subsyllabic unit, recoding activities induced by PSE also have the merit of being more implicit and intuitive than those elicited in active monitoring tasks [13-15]. All these make PSE an ideally suited tool for exploring the phonological aspects of spoken word retention.

Nonetheless, theories on PSE are mostly based on non-tonal languages, and existing findings are mixed regarding the effects of suprasegmental similarity (alternatively: tonal similarity). In an early study on Mandarin, for instance, more errors were observed for words carrying identical lexical tones [9]. But in Cantonese, tonal similarity turned out to be facilitative [10]. And with the impact of rime similarity being more pronounced in magnitude than the performance gains brought about by tonal similarity, results in [10] further predicts that the additive effect of rime and tonal similarity is negative but less harmful than rime identity. This is again at odds with what Xu has observed with Mandarin Chinese [9]. Why there is such a discrepancy is not entirely clear. It could be that cross-linguistic differences in lexical tone retention are of relevance. However, available findings are not sufficient to warrant this conclusion, especially when tonal similarity effects have not been examined under directly comparable conditions (e.g., rhyming syllables were used as tone-bearing units in [9], but non-rhyming syllables were used in [10]).

Since neither [9] nor [10] systematically addressed the role of lexicality, the generalizability of the tonal similarity effects is also potentially at issue. Studies on non-tonal languages have already shown that while rhyming impedes the recall of order for real words [7, 8], it can benefit the memorization and recollection of non-words [16]. Now the questions are: Can lexicality likewise reverse the direction of tonal similarity effects and whether this reversal is replicable using a tonal

language as phonologically diverse as Cantonese [17]? In cases where the reversal is not generalizable to lexical tones, questions may also be raised as to why lexicality modulates tonal and segmental similarity effects in different directions, an issue warranting further theoretical and empirical attention.

Another line of research demonstrating the importance of memory integrity in spoken word processing involves cognitive load. A common approach to manipulating cognitive load is to vary the amount of information to be maintained in memory. This can be realized by changing either the number of words to be recalled in a list [3] or the number of syllables comprising each word [1, 2]. However it is defined, cognitive load seems to correlate inversely with memory accuracy, particularly when verbal information is stored in memory in phonological codes [1]. As phonological encoding strategy is also the prerequisite for eliciting PSE, these variables (i.e., phonological similarity and cognitive load) may well interact. However, how they may interact and whether their interaction is identical across the segmental and suprasegmental dimensions is as yet unknown. Also, it remains open as to how these interactions may change in the presence of phonological processing deficits. Would they take on a different form or be abolished entirely? Hence, another aim of this study is to explore the potentially distinct effects of load and similarity among phonologically impaired and unimpaired individuals.

1.2. Phonological deficiency in congenital amusics

Congenitally amusic individuals (alternatively: amusics) – a group of people with mild phonological impairment but no cognitive deficits or clinically important hearing or brain injuries [18, 19] – are well suited for this study. To our knowledge, [20] was the first systematic study to relate congenital amusia (hereinafter: amusia) to phonological deficits. It showed that apart from elevated pitch thresholds and singing difficulties [18-21], English amusics also had significantly weakened phonological and phonemic awareness that could be readily detected by standardized tests.

Although less explicit, similar conclusions hold true for Mandarin amusics. In [23] for example, reduced phonological awareness led to a sharp drop in amusics' discrimination accuracy (from ceiling to a level much worse than controls' mean) when irrelevant variations were introduced into tone-bearing syllables. In a slightly different fashion, phonological deficit caused amusics to experience persistent difficulties in sentence comprehension, regardless of background noise or whether intonation contours were flattened or natural [22]. More compelling evidence came from [24], where amusics showed impaired performance in discriminating nativelike speech but were unimpaired when stimuli were clearly non-native. As non-native speech also abolished the performance gaps between dyslexics (patients with known associations with phonological deficiencies) and controls [25], results in [24] pointed unambiguously to the phonological basis of amusics' speech impairment. However, there is still much to learn about the consequences of phonological deficits, especially when they may differ across the two dimensions of speech.

2. Highlights of the present study

The present study was designed to explore the mechanisms of Cantonese spoken word retention. Effects of cognitive load and phonological similarity were inspected in the segmental and suprasegmental dimensions, respectively, and in relation to words of different lexicality. In line with [3], cognitive load was indexed by list length; and consistent with [9] and [10], segmental and tonal similarity were operationalized as rime and tone identity. Syllable complexity was also controlled for (i.e., all stimuli were of CVC structure), creating a close parallel between [9] and the present study. Hence, if there were no cross-linguistic differences in lexical tone retention, findings in [9] would be replicated. To obtain a comprehensive understanding of Chinese spoken word retention in typically and atypically developed populations, this study also recruited congenital amusics. Given that amusia may be a disconnection syndrome [26], and that response times (RTs) correlate closely with white matter integrity [27], RT analyses were additionally carried out to assess amusics' phonological abilities.

3. Method

This study followed a 3 × 3 × 2 × 2 mixed factorial design, with *Similarity (low, medium, high)*, *Lexicality (real, pseudo, reversed)*, and *Load (high, low)* as the within-subject variables, and *Group (amusic, control)* as the between-subject variable. To contrast lexicality, three kinds of speech stimuli were used: real, pseudo, and reversed. Pseudowords were novel combinations of Cantonese onsets and rimes, while reversed stimuli were generated by time-reversing the spectral and temporal features of real word stimuli. Real and pseudo stimuli were all read and recorded by an adult speaker of Hong Kong Cantonese. For each lexicality condition, three types of lists were further created to contrast similarity. Words in high-similarity lists had both rimes and lexical tones in common (e.g., "fung6, zung6, dung6, gung6"), while words in medium-similarity lists shared nothing but rimes (e.g., "toi4, zoi1, hoi6, goi2"). As for a low-similarity list (e.g., "giu1, sim4, ging1, siu6"), the nucleus was the only common element shared by all words. Tests on a separate set of rating data confirmed the statistical significance of the differences across the three similarity conditions (Tukey HSD: all ps < .001). Finally, list length was manipulated to vary cognitive load, with short lists (4 words in length) imposing a much lower load than long lists (6 words). The same methods were used to create pseudo and reversed lists for different load and similarity conditions. Also worth mentioning is that in addition to test lists (i.e., lists to be submitted to statistical analysis), there are 16 filler lists in each lexicality condition. The probe always corresponded to the first or the last stimuli in the filler sequence, but this was never the case in test lists. The aim was to minimize the potentially confounding effects of recency and primacy on order memory.

Thirty-four (amusics = 13) native Cantonese speakers agreed to participate (mean age = 23.45). Amusics scored significantly lower than controls in the online diagnostic test [28] (controls' mean = 87.71%, amusics' mean = 70.08%, $p < .001$). During the experiment, participants were seated in front of a desktop computer in a sound-attenuated booth, listening to stimuli presented binaurally over insert earphones. In each trial, participants were asked to remember a sequence of four or six words. Each word lasted for 500ms, followed by a random ISI of 300-500ms. Afterward, 2-second reversed speech mask was presented, followed by a probe. Participants were asked to recall the serial position of the probe and to respond by pressing number buttons on the keyboard. The next trial automatically began after the response was detected. While no time limit was imposed, participants were instructed to respond as accurately and as quickly as possible. Trials were blocked by lexicality and were randomized for each

participant. Block order was counterbalanced in each group. E-prime was used for stimuli presentation and data logging.

4. Results

4.1. Accuracy-based analysis

Recall data were first submitted to a four-way mixed-design ANOVA, with *Group (amusic, control)* as the between-subject variable, while *Lexicality (real, pseudo, reversed), Similarity (low, medium, high)*, and *Load (low, high)* as the within-subject variables. A significant three-way (*Lexicality* × *Similarity* × *Load*) interaction was observed [$F(4, 128) = 3.757$, $p < .01$]. Given this, separate three-way mix-design ANOVAs were carried out for each lexicality condition.

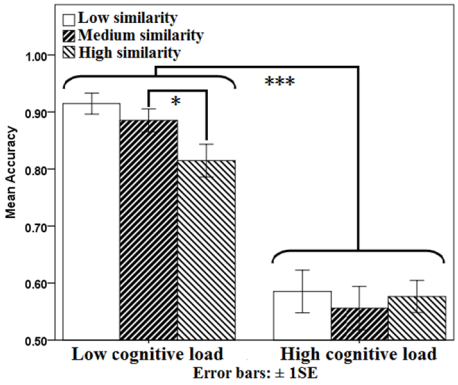

Figure 1: *Recall of real speech words as a function of similarity and load. ***: p < .001, *: p < .05*

Figure 1 summarizes participants' recall performance as a function of load and similarity. Apparently, recall was more accurate for short lists than for long lists, which led to a 30.2% difference in global accuracy. Three-way (*Similarity* × *Load* × *Group*) mixed-design ANOVA confirmed the statistical significance of this *Load* effect, $F(1, 32) = 167.875$, $p < .001$. Meanwhile, a *Similarity* × *Load* interaction was observed [$F(2, 64) = 3.612$, $p < .05$]. It turned out that while increased similarity caused accuracy to decrease significantly for short lists [$F(2, 64) = 8.229$, $p < .001$], the recall of long lists was not affected ($p = .595$), pointing to an abolishment of the narrowly-defined similarity effect under high cognitive load.

Post hoc analyses further revealed a statistically highly significant difference in the low-load condition, between the medium- and high-similarity lists (7.6% decline in accuracy, $p < .05$). However, the difference between the low- and medium-similarity conditions did not reach statistical significance ($p = .883$). As such, the presence of identical lexical tones appeared to be more deleterious to order memory than the identity of codas. Throughout the analyses, however, *Group* remained statistically non-significant, suggesting parallel performance between the amusic and control groups.

Figure 2 illustrates the recall of pseudowords. As can be seen, the significance of *Similarity* persisted in pseudo speech, regardless of low- or high-load conditions [short lists: $F(2, 64) = 14.068$, $p < .001$; long lists: $F(1.718, 54.978) = 4.625$, $p < .05$]. Moreover, different from what has been seen in real speech, similarity variations in the low-load condition resulted in a highly significant rhyming advantage: Word order was much easier to retrieve for rhyming lists than for the rest (all ps < .001). However, no rhyming advantage was observed in the high-load condition. Post hoc analysis revealed that while accuracies were significantly different between the medium- and high-similarity conditions ($p = .019$), performance was virtually indistinguishable between the low- and medium-similarity conditions ($p = .958$). Hence, somewhat similar to real speech (see Figure 1, left panel), having additional similarity in lexical tones seemed to be more consequential than having additional similarity in codas.

As for reversed speech, recall accuracy was on average much higher for short lists (41.20%) than for long lists (23.22%), and high-similarity lists (26.6%) were harder to recall compared to medium- (32.90%) and low-similarity lists (37.16%). Three-way (*Similarity* × *Load* × *Group*) ANOVA analysis confirmed the significant role of *Similarity* [$F(2, 64) = 10.827$, $p < .001$] and *Load* [$F(1, 32) = 52.961$, $p < .001$] in Cantonese spoken word retention. The remaining statistics were non-significant, including the *Group* effect ($p = .212$) and the *Similarity* by *Load* interaction ($p = .296$).

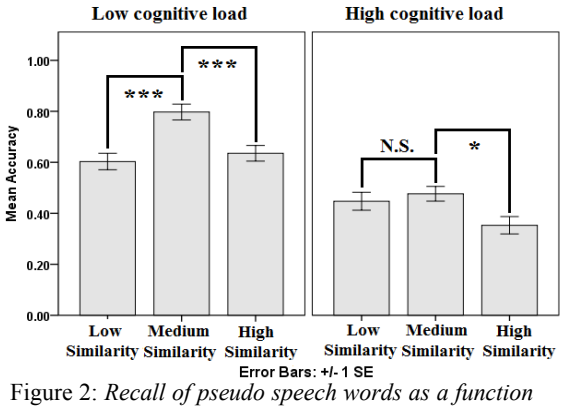

Figure 2: *Recall of pseudo speech words as a function of similarity and load. ***: p < .001, *: p < .05.*

4.2. Response time (RT) analysis

RT data were log-transformed before being submitted to the *Similarity* × *Load* × *Group* mixed-design ANOVA tests. There was a significant *Group* effect in real speech [$F(1, 334) = 22.094$, $p < .001$], which was then found to persist across cognitive load conditions [low-load: $F(1, 337) = 19.760$, $p < .001$; high-load: $F(1, 335) = 11.722$, $p < .001$]. *Group* effect likewise persisted in pseudo speech, regardless of cognitive load variations (all ps < .05). In reversed speech, however, null effect was observed for *Group* ($p = .960$), forming a contrast with the remaining two lexicality conditions.

Additionally, to assess amusics' phonological awareness, planned comparisons (one-way ANOVAs) were carried out for the real and pseudo speech conditions. Analyses showed that it was controls' performance that was driving the *Similarity* effects in real speech [low-load: $F(2, 627) = 5.801$, $p < .01$; high-load: $F(2, 624) = 6.172$, $p < .01$]. Amusics' performance, albeit following a qualitatively similar trend, was far from statistical significance (all ps > .5). Similar results were obtained for pseudo speech. Not only were controls much faster in recall, but RT analyses again identified them to be the more sensitive party to phonological manipulations [low-load: $F(2, 627) = 6.887$, $p < .01$; high-load: $F(2, 626) = 6.141$, $p < .01$], not amusics ($p = .253$ and $p = .331$ for low- and high-load conditions respectively).

5. Discussion

To understand the processing of spoken words following their initial recognition, this study investigated the veracity of verbal memory as a function of *Group*, *Lexicality*, *Load*, and *Similarity*. Results showed that the cognitive load effects and PSE apply to tonal language retention as much as that of non-tonal languages. The same could be said of the rhyming effect: For while rhyming improved the retrieval of pseudowords in the order of presentation, it hindered the recall of serial order for Cantonese real words.

Interestingly, changes in lexicality did not appear to affect the direction of suprasegmental similarity effects: Participants' performance decreased irrespective of the stimuli's lexical status. The corruptive effects of tonal identity are in good agreement with the narrowly-defined PSE phenomenon. They are also reminiscent of what Xu observed with Mandarin [9]. However, such a result clearly contradicted the additive effect predicted by [10]. In light of these, we may reasonably rule out the conjecture that discrepant findings between Xu [9] and Yip [10] reflect cross-linguistic differences in lexical tone retention. Rather, these inconsistencies might be attributed to the fact that in [9] syllable complexity was controlled, and that primacy effect was reduced due to the exclusion of words in the 1st serial position from data analyses. As this study made similar methodological decisions, these may also be the reason why results acquired here align better with [9] than with [10].

Moreover, in support of the psycholinguistic accounts [8], the similarity effects varied as a function of subsyllabic phonological components. In real speech, for instance, rime identity was more detrimental than nucleus identity, in that accuracy was consistently lower for medium-similarity lists than for low-similarity lists. Similarly, tone identity exerted a stronger interference effect than coda identity. The adverse impact of tonal similarity was even robust enough to abolish the rhyming advantage observed for pseudo speech words. Nonetheless, the concept of robustness may be operationalized differently. One such alternative is to define robustness as the ability to persist in cognitively demanding situations. Take this study as an example, unlike in real speech where heightened cognitive demands led to an across-the-board elimination of the effects (thus the absence of the stair-step decline in the right panel of Figure 1), PSE's influence remained discernable in pseudo speech context. Although with segmental similarity effects diminishing below significance, it was tonal similarity that was driving the statistical significance. Taken together with the differential effects of lexicality in the segmental and suprasegmental dimensions, findings of this study are strongly indicative of the existence of distinct processing mechanisms for tones and segments.

Although not unprecedented, the elimination of the similarity effects is also worth commenting on. For different from previous literature, PSE was abolished by increased cognitive load. In light of the attenuation of the similarity effect by the induction of semantic encoding strategies [1], one might be attempted to conclude that listeners have the propensity to switch to semantic strategies under loaded conditions. It is also tempting to argue that high load prompted listeners to use semantic knowledge to enlarge the distinctness of words in memory [16], thereby weakening the magnitudes of PSE. However, semantics cannot justify why similar patterns could occur in the pseudo speech condition of this study. To comprehensively account for the data, a third variable has to be considered, such as the differential weightings of subsyllabic units. As tonal similarity effects were more robust against the high cognitive load, results suggested a heavier weighting of lexical tones than segments. But whether this differential weighting can be taken to indicate that segments, in general, are weighed differently from tones remains unclear, as is the question of whether effects of cognitive load vary as a function of subsyllabic components. Further research is needed to address these issues.

The rhyming advantage surfaced during pseudo word recollection is another important finding of this study. As similar patterns failed to hold for reversed speech (where phonological recoding is impossible), results obtained here provided further support to PSE's linguistic basis. Still open is the question of why rhyming advantage more reliably occurs during pseudoword recollection but not the recall of real words. One possibility is to consider listeners' tendency to encode speech holistically, using either syllable-sized units [14, 15, 29] or onsets and rimes [6-9]. Due to holistic encoding, for example, Cantonese subjects of this study would transform the real word stimuli into syllable-sized units for temporary storage in working memory. But when presented with pseudowords (i.e., novel onsets and rimes combinations), syllable-sized holistic representations were no longer available in listeners' mental lexicon. This forced the participants to recode pseudoword stimuli using onsets and rimes. Since rhyming was implemented at the same subsyllabic level (i.e., onset-rime), listeners could readily detect the regularities in phonemic makeup and use that to their advantage during order recollection. By comparison, structural regularities were much less salient in the real word condition, due largely to the misalignment of phonological units used for retention and those shared at the subsyllabic level (e.g., nucleus, coda). This is presumably why "rhyming advantage" (or the "taxonomic" and "cueing" effects of rhyming lists) [7, 8] tends to influence the recall of pseudowords but not real words.

Finally, results obtained here showed that while recall accuracies were virtually indistinguishable between controls and amusics, the two groups differed in their processing speed and phonological sensitivity. These are indicative of a verbal memory deficit in congenital amusics, contrary to what has been suggested in [30]. Results also identified RTs as the more sensitive measure in detecting phonological impairment. Our working hypothesis is that this has to do with the nature of amusics' phonological deficit. According to [31] and [32], phonological impairment has two main sources, one being impaired phonological access and the other degraded mental representations. In the case of an access-based deficit, normal performance may be expected if subjects are afforded abundant time and resources; but if long-term representations are implicated, performance restoration is less likely to occur [31]. In the present study, amusics were found to achieve the same recall accuracies (hence the same levels of elaborative phonological recoding of verbal stimuli) as did controls, but at the cost of RTs. This aligns well with the predictions of the phonological access deficit. As congenital amusia is not the only developmental disorder accompanied by phonological impairment, future studies may use a similar logic to explore the nature and source of phonological deficits in atypically developing populations.

6. Acknowledgements

This study is supported by the Direct Grant for Research of The Chinese University of Hong Kong (2014-2015).

7. References

[1] G. Campoy, and A. D. Baddeley, “Phonological and semantic strategies in immediate serial recall,” *Memory*, vol. 16, no. 4, pp.329–340, 2008.

[2] I. Neath, and J. S. Nairne, “Word-length effects in immediate memory: Overwriting trace decay theory,” *Psychonomic Bulletin & Review*, vol. 2, no. 4, pp. 429–441, 1995.

[3] K. M. Visscher, E. Kaplan, M. J. Kahana, and R. Sekuler, "Auditory short-term memory behaves like visual short-term memory," *PLoS Biol*, vol. 5, no. 3, pp. e56, 2007.

[4] A. D. Baddeley, “Short-term memory for word sequences as a function of acoustic, semantic and formal similarity,” *Quarterly Journal of Experimental Psychology*, vol. 18, pp. 362–365, 1966.

[5] J. Schweppe, M. Grice, R. Rummer, “What models of verbal working memory can learn from phonological theory: Decomposing the phonological similarity effect,” *Journal of Memory and Language*, vol. 64, no. 3, pp. 256–269, 2011.

[6] O. Tzeng, D. Hung, W. Wang, and L. E. Bourne, “Speech recoding in reading Chinese characters,” *Journal of Experimental Psychology: Human Learning and Memory*, vol. 3, no. 6, pp. 621–630, 1977.

[7] A. B. Fallon, K. Groves, and G. Tehan, “Phonological similarity and trace degradation in the serial recall task: when CAT helps RAT, but not MAN,” *International Journal of Psychology*, vol. 34, pp. 301–307, 1999.

[8] L. M. Nimmo, and S. Roodenrys, “Investigating the phonological similarity effect: Syllable structure and the position of common phonemes,” *Journal of Memory and Language*, vol. 50, no. 3, pp. 245–58, 2004.

[9] Y. Xu, “Depth of phonological recoding in short-term memory,” *Memory & Cognition*, vol. 19, no. 3, pp. 263–273, 1991.

[10] M. C. Yip, “What is the role of tone in the phonological similarity effect?” *The Psychological Record*, vol. 64, no. 1, pp. 115–122, 2014.

[11] A. D. Baddeley, “Working memory: Theories, models, and controversies,” *Annual review of psychology*, vol. 63, pp. 1–29, 2012.

[12] A. D. Baddeley, and R. H. Logie, “Working memory: The multiple-component model,” in *Models of Working Memory*, A. Miyake and P. Shah, Eds. Cambridge: Cambridge University Press, pp. 28–61, 1999.

[13] D. B. Pisoni, and P. A. Luce, “Acoustic-phonetic representations in word recognition,” *Cognition*, vol. 25, no. 1, pp. 21–52, 1987.

[14] A. F. Healy, and J. E. Cutting, “Units of speech perception: Phoneme and syllable,” *Journal of Verbal Learning and Verbal Behavior*, vol. 15, no. 1, pp. 73–83, 1976.

[15] H. C. Nusbaum, and J. DeGroot, “The role of syllables in speech perception,” in *Papers from the parasession on the syllable in phonetics and phonology*, M. S. Ziolkowski, M. Noske, and K. Deaton, Eds. Chicago: Chicago Linguistic Society, 1991.

[16] A. Lian, P. J. Karlsen, and B. Winsvold, “A re-evaluation of the phonological similarity effect in adults' short-term memory of words and nonwords,” *Memory*, vol. 9, no. 4–6, pp. 281–299, 2001.

[17] G. Peng, “Temporal and tonal aspects of Chinese syllables: A corpus-based comparative study of Mandarin and Cantonese,” *Journal of Chinese Linguistics*, vol. 34, pp. 134–154, 2006.

[18] I. Peretz, J. Ayotte, R. J. Zatorre, J. Mehler, P. Ahad, V. B. Penhune, et al., “Congenital amusia: A disorder of fine-grained pitch discrimination,” *Neuron*, vol. 33, no. 2, pp. 185–191, 2002.

[19] J. Ayotte, I. Peretz, and K. L. Hyde, “Congenital amusia: A group study of adults afflicted with a music-specific disorder,” *Brain*, vol. 125, no. 2, pp. 238–251, 2002.

[20] J. Jones, J. Lucker, C. Zalewski, C. Brewer, and D. Drayna, “Phonological processing in adults with deficits in musical pitch recognition,” *Journal of Communication Disorders*, vol. 42, pp. 226–234, 2009.

[21] F. Liu, A. D. Patel, A. Fourcin, and L. Stewart, “Intonation processing in congenital amusia: Discrimination, identification and imitation,” *Brain*, vol. 133, no. 6, pp. 1682–1693, 2010.

[22] F. Liu, C. Jiang, B. Wang, Y. Xu, and A. D. Patel, “A music perception disorder (congenital amusia) influences speech comprehension,” *Neuropsychologia*, vol. 66, pp.111–118, 2015.

[23] Y. Nan, Y. Sun, and I. Peretz, “Congenital amusia in speakers of a tone language: Association with lexical tone agnosia,” *Brain*, vol. 133, no. 9, pp. 2635–2642, 2010.

[24] X. Wang, and G. Peng, “Phonological processing in Mandarin speakers with congenital amusia,” *The Journal of the Acoustical Society of America*, vol. 136, no. 6, pp. 3360, 2014.

[25] T. K. Perrachione, S. N. Del Tufo, and J. D. E. Gabrieli, “Human voice recognition depends on language ability,” *Science*, vol. 333, pp. 595, 2011.

[26] P. Loui, D. Alsop, and G. Schlaug, “Tone deafness: A new disconnection syndrome?” *The Journal of Neuroscience*, vol. 29, pp. 10215–10220, 2009.

[27] S. N. Niogi, P. Mukherjee, J. Ghajar, C. Johnson, R. A. Kolster, R. Sarkar, H. Lee et al., "Extent of microstructural white matter injury in postconcussive syndrome correlates with impaired cognitive reaction time: A 3T diffusion tensor imaging study of mild traumatic brain injury," *American Journal of Neuroradiology*, vol. 29, no. 5, pp. 967–973, 2008.

[28] I. Peretz, N. Gosselin, B. Tillmann, L. L. Cuddy, B. Gagnon, C. G. Trimmer, S. Paquette, and B. Bouchard, “On-line identification of congenital amusia,” *Music Perception: An Interdisciplinary Journal*, vol. 25, pp. 331–343, 2008.

[29] T. Liu, and J. H. W. Hsiao, “Holistic processing in speech perception: Experts' and novices' processing of isolated Cantonese syllables,” *Proceedings of the 36th Annual Conference of the Cognitive Science Society*, pp. 869-874, 2014.

[30] B. Tillmann, K. Schulze, and J. M. Foxton, “Congenital amusia: A short-term memory deficit for non-verbal, but not verbal sounds,” *Brain and Cognition*, vol. 71, no. 3, pp.259–264, 2009.

[31] F. Ramus, and G. Szenkovits, “What phonological deficit?” *Quarterly Journal of Experimental Psychology*, vol. 61, pp. 129–141, 2008.

[32] F. Ramus, and M. Ahissar, “Developmental dyslexia: The difficulties of interpreting poor performance, and the importance of normal performance,” *Cognitive Neuropsychology*, vol. 29, no. 1-2, pp. 104–122, 2012.