Holistic Processing in Speech Perception: Experts' and Novices' Processing of Isolated Cantonese Syllables

Tianyin Liu (kanalty@hku.hk)
Janet Hui-wen Hsiao (jhsiao@hku.hk)

Department of Psychology, University of Hong Kong
627 Jockey Club Tower, Centennial Campus, Pokfulam Road, Hong Kong SAR

Abstract

Holistic processing is an expertise marker in visual perception. Here we tested participants' perception of Cantonese syllables through composite paradigm commonly used for measuring holistic processing in vision research. We found that participants were more holistic in processing syllable initials than finals, possibly because initials are shorter and more easily combined with neighboring segments. More importantly, experts' perception of syllable initials was strongly affected by finals (i.e., holistic processing), even when the initials were separated or disconnected from finals; whereas in novices, holistic processing gradually decreased with segment separation and disconnection. As for syllable final perception, experts were strongly influenced by initials, yet not when the segments were separated/disconnected. In contrast, novices showed little holistic processing in all conditions. These results showed that experts' perception of syllable parts were more influenced by neighboring segments, suggesting that holistic processing may also be an expertise marker in auditory perception.

Keywords: holistic processing, speech perception, isolated Cantonese syllable processing.

Introduction

Holistic processing, i.e., gluing features into a gestalt and perceive them as a whole, has been identified as a hallmark of expertise in the perception of faces (Tanaka & Farah, 1993). The most prominent demonstration of holistic face processing is the composite illusion. Participants were shown two composite faces and asked to do same/different judgments about the top halves of the faces and ignore the irrelevant bottom halves. People reported two identical top halves to look different when they were aligned with two different bottom halves. This illusion disappeared when the top and bottom halves were misaligned (Hole, 1994; Fig. 1). Presumably, holistic processing binds the two halves of the faces together so that experts' flexibility to access the information of individual parts is attenuated (Maurer, LeGrand, & Mondloch, 2002). Thus, they get interference from the irrelevant halves in the same/different judgments of the attended halves, and the amount of interference indicates holistic processing. The same paradigm has been applied to the recognition of many other non-face objects, such as cars (Gauthier, Curran, Curby, & Collins, 2003), fingerprints (Busey & Vanderkolk, 2005), and greebles (an artificial stimulus type, Gauthier & Tarr, 1997), and holistic processing has been found to indicate perceptual expertise with these objects. In experts' eyes, parts are integrated and recognized as a whole (Gauthier & Tanaka, 2002), and the more holistic they are, the better they can recognize/individualize faces or objects of expertise (Richler, Cheung, & Gauthier, 2011).

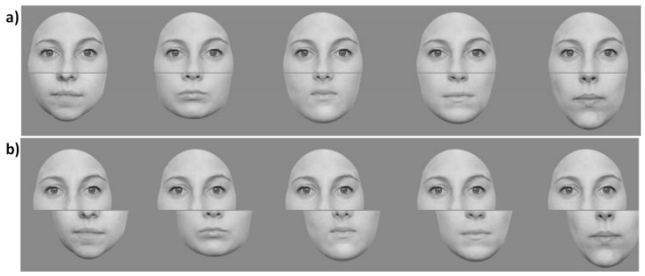

Figure 1. The composite face illusion: a) 5 physically identical top face halves are perceived as different when aligned with 5 different bottom halves; but b) are perceived as the same when misaligned (images taken from Rossion, 2013).

In contrast to the abundant literature on holistic processing in visual perception, it remains unclear whether a similar phenomenon also exists in auditory perception, and whether it is a hallmark in expert auditory processing. Our sense of hearing begins to develop since infancy or even the fetal period, much earlier than vision (Hepper & Shahidullah, 1994). The auditory system enables us not only to hear, but also to perceive and interpret sound information. The ability to recognize a caregiver's voice and understand what he/she says is as important as the ability to recognize his/her face in the early years of development. It is also an essential tool for later learning, language development, and social interaction (Bellis, 2003; Clark, 2008; Sharma, Purdy, & Kelly, 2009). Hence, it is important to investigate what signals the expertise of auditory processing, especially in the perception of speech. Past research in speech perception has been focusing on phonological awareness, i.e., explicit awareness of the phonological structure of the language, and revealed its critical role in children's acquisition of reading, in both alphabetic languages (Goswami & Mead, 1992), and logographic languages such as Chinese (Shu, Peng, & McBride-Chang, 2008; although it remains controversial; see Huang & Hanley, 1995). However the maturation of phonological awareness with age and training does not necessarily lead to a better ability to selectively attend to parts in speech perception without being interfered by irrelevant parts; in other words, phonological awareness does not necessarily signal analytical processing (reduced

holistic processing) defined according to the composite paradigm.

To investigate whether holistic processing also exists in auditory processing, here we examine the perception of Cantonese base syllables (controlled for tonal information) in expert and novice Cantonese listeners. Cantonese is the most commonly spoken Chinese language in Southern China and Hong Kong. There is a one-to-one mapping between written characters and spoken Cantonese syllables (i.e., a logographic language; Ho & Bryant, 1997), and components of a written character do not map to components in the pronunciation, in contrast to words in alphabetic languages. The basic speech unit in Cantonese is the base syllable with a tone, and it is conventional to divide a Cantonese syllable into two parts: the initial and the final segments (Yue-Hashimoto, 1972). The initial segment refers to the beginning consonant, whereas the final segment refers to the ending vowel(s) or vowel(s) plus an ending consonant. There are in total 19 initial segments and 51 final segments in Cantonese, together with nine tones attached to the final segments. In Hong Kong, the majority of children learn to read Chinese characters without the aid of an alphabetic phonetic system, e.g., Pinyin used in Mainland China and Zhu-Yin-Fu-Hao (or Bopomofo, a phonetic notation system consisted of 37 characters and four tone marks) used in Taiwan. Thus, this examination with Cantonese speakers in Hong Kong provides a unique opportunity to examine the perceptual expertise of Cantonese speech without explicit knowledge about the constituent (phonemic) structure of the syllables. Here we aim to examine how native Cantonese speakers process base Cantonese syllables in terms of holistic processing in comparison with novices whose native language is English.

Here we hypothesize that in speech perception/syllable processing, similar to the literature on visual recognition, experts of Cantonese will demonstrate stronger holistic processing than novices. More specifically, Hong Kong Cantonese speakers' perception of Cantonese initial or final segments may be more influenced by neighboring segments as compared with novices who speak alphabetic languages. This study is the first attempt to apply the composite paradigm to auditory perception and identify the hallmark of expertise in speech processing.

Methods

Participants

20 native Cantonese speakers (12 male, 8 female) and 20 novices of Cantonese (12 male, 8 female) were recruited from the University of Hong Kong. All expert Cantonese speakers were born and grew up in Hong Kong and had limited or no experience with Pinyin or Zhu-Yin-Fu-Hao. All novice participants were native English speakers who were exchange students or freshmen at the University of Hong Kong. By the time they were recruited, they had resided in Hong Kong for less than six months, and received no formal education of Cantonese or any other Chinese languages[1]. The two groups were matched in gender, age (novice: *mean* = 22.15, *S.E* = .72; expert: *mean* = 20.75, *S.E* = .50, $F(1, 38) = 2.57$, $p = .12$) and years of education (novice: *mean* = 15.25, *S.E* = .35; expert: *mean* = 14.65, *S.E* = .40, $F(1, 38) = 1.26$, $p = .27$).

Materials

Baseline Auditory Temporal Processing To make sure that participants did not differ in baseline temporal processing of auditory information, a birdsong discrimination task was used. 60 sound clips of birdsongs were downloaded from *New Zealand bird songs and calls* corpus (Te Papa Atawhai, 2013), and each clip was trimmed to 1.5s from the beginning of the birdsong using GoldWave.

Holistic Processing of Cantonese Syllables The complete composite paradigm (Gauthier & Bukach, 2007) commonly used in vision research was adopted to examine holistic processing in the perception of isolated Cantonese syllables. In each trial, participants were presented with a pair of Cantonese syllables sequentially, and told to attend only to a particular segment of the syllables (either the initial or the final), and judge whether they were the same or different. In congruent trials, the attended and irrelevant segments led to the same response, i.e., both were "same" or both were "different" (i.e., no interference from irrelevant segments; see Table 2 and Fig.1 for examples); whereas in incongruent trials, they led to different responses (i.e., there was interference from irrelevant segments). Holistic processing was assessed by the performance difference between the congruent and the incongruent trials.

240 pairs of original Cantonese syllables recorded from a native female Cantonese speaker from a sound corpus *Chinese Character Database* (Kwan, 2001) were used. They all had the same tone (first tone), and were equally divided into four trial types: same congruent, different congruent, same incongruent, and different incongruent (see Table 1 and Fig. 2a for examples). Note that in a Cantonese syllable, the initial and final segments can influence each other due to co-articulation, and thus the same initial may sound a little different when combined with different finals. In other words, the segments in an original syllable are already integrated to form a holistic representation. To examine how participants perceived Cantonese segments when they were not integrated, we created artificial syllables by concatenating standardized isolated initials and finals corresponding to the original syllables with no time delay (separated syllables; Fig. 2b). All standardized isolated initials and finals were downloaded from the same database (Kwan, 2001) recorded by the same female. In addition, we created disconnected syllables (Fig.2c) by adding a 0.5s gap between the initial and the final of the separated syllables. This disconnected condition is analogous to the misaligned

[1] Note that the official language for instruction at University of Hong Kong is English; all our native Cantonese speaker participants were Chinese-English bilinguals.

condition in the composite paradigm used in vision research. The original and separated stimuli were about 1s long, whereas the disconnected stimuli were about 1.5s long.

		Attend to Initial	Attend to Final
Congruent	Same	[p]in$^{(1)}$ – [p]in$^{(1)}$	m[aau$^{(1)}$] – m[aau$^{(1)}$]
	Different	[j]in$^{(1)}$ – [t]eng$^{(1)}$	h[ang$^{(1)}$] – s[ing$^{(1)}$]
Incongruent	Same	[p]an$^{(1)}$ – [p]ou$^{(1)}$	l[aau$^{(1)}$] – ng[aau$^{(1)}$]
	Different	[j]yun$^{(1)}$ – [t]yun$^{(1)}$	gw[ang$^{(1)}$] – gw[ing$^{(1)}$]

Table 1. Examples of the Cantonese syllable stimuli. The number in the small bracket indicates the tone of the syllable.

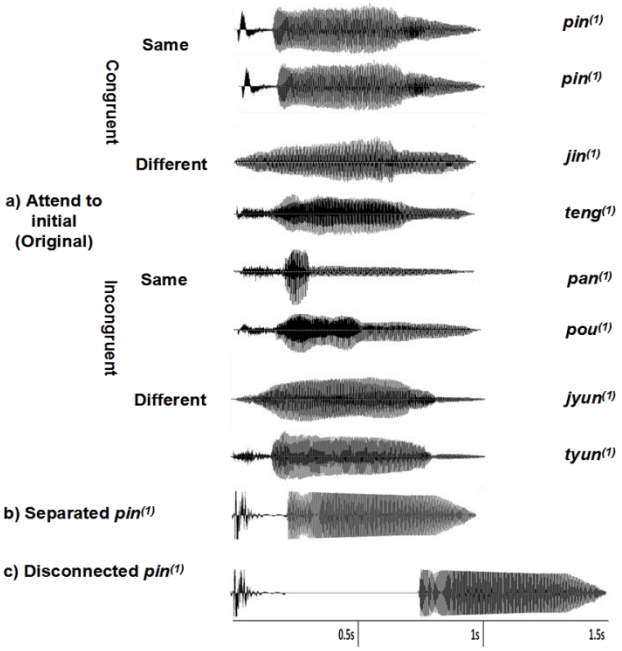

Figure 2. a) Complete composite paradigm with original Cantonese syllables in the initial judgment task, and examples of b) separated and c) disconnected syllables.

Design

The design had three within-subject variables: congruency (congruent vs. incongruent), stimulus type (original vs. separated vs. disconnected), and attended syllable segment (initial vs. final); and a between-subject variable: group (expert vs. novice). The dependent variable was discrimination sensitivity measured by A', which is a bias-free nonparametric measure of sensitivity[2]. The degree of holistic processing was indicated by the performance difference between congruent and incongruent trials.

Procedure

Baseline Auditory Temporal Processing 30 trials of birdsong discrimination were carried out before the holistic processing tasks. In each trial, participants first saw a fixation cross for 500 ms at the center of the screen to signal the start of the trial. They then heard two birdsongs sequentially, each of which lasted for 1.5s, with a 1s interval between the two stimuli. They were asked to judge whether the two birdsongs were the same or different with a response box. Half of the trials were "same" trials, and the other half were "different" trials. All birdsongs were drawn from the 60 birdsong clips; the presentation of a birdsong in a "same" or "different" trial was counterbalanced across participants.

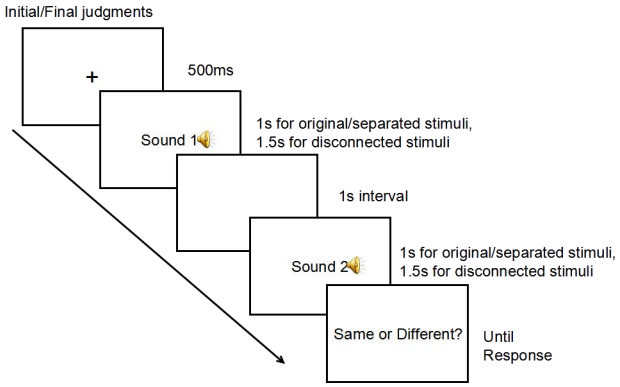

Figure 3. Procedure of the holistic processing task

Holistic Processing of Cantonese Syllables There were 720 trials in total, evenly divided into the original, separated, and disconnected conditions. For each condition, there were six blocks, two for syllable initial discrimination and four for final discrimination; each block contained 40 trials. The sequence of presenting the initial and final discrimination blocks was counterbalanced across participants. In each trial, after a 500 ms central fixation, they listened to two sequentially presented Cantonese syllables with a 1s interval in between, and then judge whether the initials (or the finals in the final discrimination blocks) of the two syllables were the same or different with a response box (Fig. 3). They performed a practice session with 24 trials consisting of all types of stimuli (8 trials for each type shown in Figure 2, with 4 initial and 4 final discriminations respectively) not used in the materials before the experiment.

Results

Bird Song

In the birdsong discrimination task, the expert and novice groups had similar performance in both accuracy (experts, *mean* = .89, *S.E* = .03; novices, *mean* = .92, *S.E* = .02) and response time (experts, *mean* = 1.72sec, *S.E* = .13sec; novices, *mean* = 1.83sec, *S.E* = .23sec), suggesting that they

[2] $A' = .5 + \left[sign(H-F)\frac{(H-F)^2 + |H-F|}{4\max(H,F) - 4HF} \right]$, where H and F present hit rate and false alarm rate respectively. Better performance is indicated by a higher *A'*. Here we used *A'* instead of *D'* because it can be calculated when the hit rate or the false alarm rate is 1 or 0, which was present in the data we collected.

generally could discriminate sounds accurately and there was no difference in discrimination ability of auditory information between the two groups (accuracy, F (1,38) = .35, p = .56; response time, F (1, 38) = .17, p = .68).

Holistic Processing

The assumption of sphericity was met for all independent variables, and repeated measures ANOVA revealed main effects of stimulus type (F(2, 76) = 3.01, p < .05, η_p^2 = .08), congruency (F(1, 38) = 41.45, p < .001, η_p^2 = .52), and attended syllable segment (F(1, 38) = 28.86, p < .001, η_p^2 = .43) in A'. Participants performed the worst in perceiving original Cantonese syllables (original vs. separated: t(39) = 2.85, p < .01; original vs. disconnected, t(39) = 4.11, p < .001), and had similar performance in the separated and disconnected conditions (t(39) = 1.54, p = .13). They were better in congruent than incongruent trials, and in processing finals than initials. There was a significant two-way interaction between stimulus type and congruency (F(2, 76) = 9.52, p < .001, η_p^2 = .20), a significant three-way interaction between stimulus type, congruency and attended segment (F(2, 76) = 3.30, p < .05, η_p^2 = .08), and a marginal four-way interaction between stimulus type, congruency, attended segment, and group (F(2, 76) = 2.74, p = .07, η_p^2 = .08). The marginal four-way interaction indicated that the level of holistic processing in the perception of the syllable initials and finals might differ between the two groups. To further investigate this effect, we examined the data in the syllable initial and syllable final discrimination tasks separately.

Syllable Initials A significant main effect of congruency was found (F(1, 38) = 34.77, p < .001, η_p^2 = .48): participants showed holistic processing in syllable initial perception. There was a significant interaction between congruency and stimulus type (F(2, 76) = 6.99, p < .01, η_p^2 = .16), indicating that the level of holistic processing differed across different types of stimuli. The three-way interaction between congruency, stimulus type, and group was not significant (F(2, 76) = 1.81, p = .17, n.s.). However, when we separated the data by groups, we found that among experts, there was no significant interaction between congruency and stimulus type (F(2, 38) = .96, p = .40, n.s.): experts showed a similar level of holistic processing in the original (t(19) = 3.78, p < .01), separated (t(19) = 4.08, p < .01), and disconnected (t(19) = 2.17, p < .05) conditions. In contrast, there was a significant interaction between congruency and stimulus type in novices (F(2, 38) = 7.94, p < .05, η_p^2 = .30): the holistic processing effect decreased gradually as the initial and final segments were separated and then disconnected; a significant congruency effect was observed only in the original (t(19) = 3.77, p <.01) and separated (t(19) = 2.94, p <.01) syllables, but not in the disconnected condition (Fig. 4)[3].

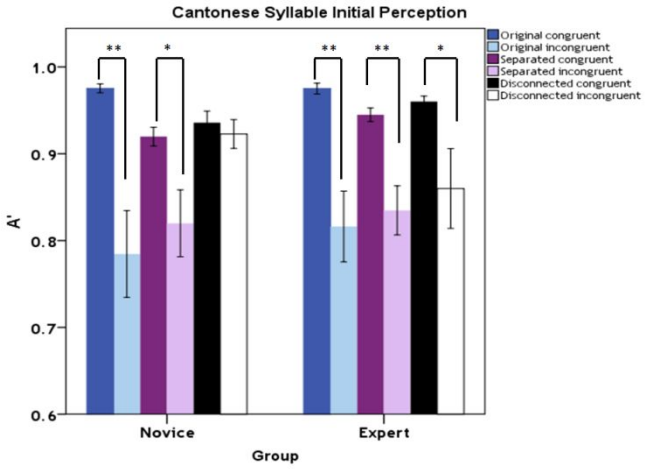

Figure 4: Performance of novices and experts of Cantonese in the holistic processing task with Cantonese syllable initials (* p < 0.05; ** p < 0.01)

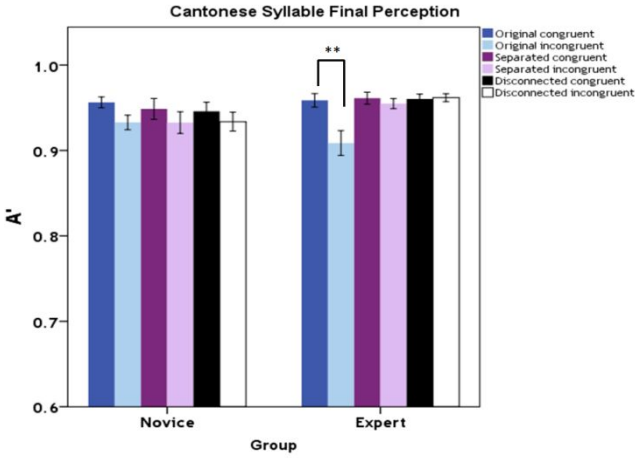

Figure 5: Performance of experts and novices of Cantonese in the holistic processing task with Cantonese syllable finals (** p < 0.01)

Syllable Finals A main effect of congruency was observed (F(1, 38) = 30.68, p < .001, η_p^2 = .45), suggesting that participants showed holistic processing in perceiving syllable finals. A significant interaction between congruency and stimulus type (F(2, 76) = 9.23, p < .001, η_p^2 = .20) indicated different levels of holistic processing across different stimulus types. A significant three-way interaction between congruency, stimulus type, and group (F(2, 76) = 4.03, p < .05, η_p^2 = .10) further revealed a between group difference in the level of holistic processing across stimulus types. We then examined their performance separately by groups. Experts exhibited a significant interaction between

[3] Note however that in the response time data, both groups showed decreased congruency effects in the separated and disconnected conditions as compared with the original condition.

congruency and stimulus type ($F(2, 38) = 13.68$, $p < .01$, $\eta_p^2 = .42$): they showed a strong congruency effect only in the original condition ($t(19) = 4.51$, $p < .01$) but not in the separated or disconnected conditions. In contrast, there was no interaction between congruency and stimulus type in novices ($F(2, 30) = .52$, $p = .60$, n.s.): they showed a small congruency effect, which was unaffected by separating or disconnecting the segments (Fig. 5)[4].

Discussion

In the current study, we adopted the composite paradigm to examine the perception of Cantonese base syllables in native (expert) Cantonese speakers and novices of Cantonese who speak only English.

First, the results showed that both participant groups processed syllable initials more holistically than syllable finals. In Cantonese, syllable initials are single consonants, while syllable finals are composed of vowel(s) or vowel(s) with an ending consonant. As can be visualized in Fig. 2, the duration of syllable initials are significantly shorter than that of syllable finals, and thus syllable initials may be more easily perceived as combined with other segments than syllable finals. Indeed, developmental studies showed that children develop awareness of large sound segments first and gradually become sensitive to smaller ones at later stages (Goswami & Mead, 1992; Ho, & Bryant, 1997), suggesting that the awareness of syllable initials is harder than finals and emerges at a later stage. Hence, when asked to identify initials while disregarding finals, participants had larger interference from irrelevant finals, as compared with the interference from the irrelevant initials when identifying syllable finals. In addition, Cantonese syllable initials and finals also differ in acoustic properties due to the fundamental differences between consonants and vowels. For example, consonants typically have lower amplitude and more high frequency energy than vowels (Ladefoged & Disner, 2012). These acoustic differences may also be related to the observed differences between initial and final syllable perception.

Secondly, we found that Cantonese experts were in general more holistic than novices in Cantonese syllable perception, yet the manifestation differed between syllable initial and final perception. In perceiving Cantonese syllable initials, novices showed decreased holistic processing when the final segment was separated or disconnected from the initial segment, reflecting the baseline behavior in which features were less likely to be perceived as a whole when they were separate/disconnected (note that the initial and final segments in an original syllable already formed a holistic representation due to co-articulation). In contrast, experts' holistic processing was uninfluenced by segment separation or disconnection; their perception of syllable initials was affected by syllable finals even when the finals were separated or disconnected from the initials. This effect suggests that experts combine syllable initials with finals even when the continuity between the segments is disrupted due to their experience with Cantonese syllables. We speculate that this effect might be because initials are short and difficult to be recognized alone, and thus through experience expert Cantonese speakers have learned to combine them with finals for processing, even when there is a gap between the initial and the final in a syllable.

In syllable final perception, experts showed strong holistic processing in the original condition but not in the separated or disconnected conditions, suggesting that experts were able to perceive the final as an isolated segment when it was separated/disconnected from the initial. In contrast, novices showed a small holistic processing effect in the original condition, and separation/disconnection did not change this effect. Taken together, these results suggest that experts' experience with original Cantonese syllables may have encouraged them to integrate the initial into the perception of the final. However, in contrast to their perception of the initials, separation/disconnection was able to disrupt this integration. This difference between syllable initial and final perception may be due to the longer duration in the final segment, making it easier to be perceived alone when being separated/disconnected from the initial segment.

Thus, our results suggest that experts' perception of Cantonese syllable initials and finals seems to be more influenced by neighboring segments than novices'. This result is consistent with the perceptual expertise literature in vision research, which shows that experts in visual recognition tends to combine features together to form a holistic representation, resulting in an inability to selectively attend to parts (Maurer, LeGrand, & Mondloch, 2002).

Note that in the literature on auditory perception, the term "holistic processing" has often been defined as processing at the level of syllable or a global unit of speech (Charles-Luce & Luce, 1990; Walley, 1988). This is in contrast to the definition used in vision research with the composite paradigm, in which holistic processing refers to feature combination that results in an inability to selectively attend to parts. In the development of speech recognition, processing at the syllable level is considered as a less efficient and immature form of speech processing. On the contrary, incremental/analytic processing, i.e., processing a word as a combination of individual sounds from the beginning to the end (Byrd, Conture, & Ohde, 2007), emerges at a later stage and marks the maturation of phonological encoding. The typical developmental trend of syllable perception from the syllable level to incremental processing (e.g., Brooks & MacWhinney, 2000; Bryd, Conture, & Ohde, 2007) has been demonstrated through a series of tasks such as onset/rhyme detection or deletion, phoneme deletion, etc. Note however that being aware of the internal structure of a syllable does not necessarily mean a better ability to selectively attend to parts in syllable perception. Indeed, Ho and Bryant (1997) found that Cantonese children in Hong Kong have developed awareness of onset/rhyme, and can perform onset/rhyme detection tasks by age eight, suggesting that sensitivity to

[4] Similar effects were observed in the response time data.

the internal structure of syllables can emerge without explicit teaching. Nevertheless, our results using the composite paradigm suggest that their perception of onset (initial) and rhymes (final) seems to be more influenced by neighboring segments than novices', regardless of their awareness of the internal structure of the syllables. Future work will investigate whether similar effects can also be observed in Chinese speakers who learned speech sounds through a phonetic alphabet such as Pinyin, or in expert processing of speech sounds in alphabetic languages.

In conclusion, the current study revealed that experts demonstrated stronger holistic processing over novices in Cantonese syllable perception as assessed by the composite paradigm. The results mirrored the findings in vision research, indicating that holistic processing may be an expertise marker in the recognition of both visual and auditory stimuli.

Acknowledgments

We are grateful to the Research Grant Council of Hong Kong (project code: HKU 745210H to J.H. Hsiao), and the help from our research assistant Masaya Okamoto.

References

Bellis, T. J. (2003). *Assessment and management of central auditory processing disorders in the educational setting: From science to practice (2nd ed.).* Clifton Park, NY: Delmar Learning.

Brooks, P. J., & MacWhinney, B. (2000). Phonological priming in children's picture naming. *Journal of Child Language, 27,* 335 - 366.

Busey, T. A. & Vanderkolk, J. R. (2005). Behavioral and electrophysiological evidence for configural processing in fingerprint experts. *Vision Research, 45,* 431-448.

Byrd, C. T., Conture, E. G., & Ohde, R. N. (2007). Phonological Priming in Young Children Who Stutter: Holistic Versus Incremental Processing. *American Journal of Speech-Language Pathology, 16,* 43-53.

Charles-Luce, J., & Luce, P. A. (1990). Similarity neighborhoods of words in young children's lexicon. *Journal of Child Language, 17,* 205 - 215.

Clark, J. G. (1980). Central auditory dysfunction in school children. *Language, Speech, and Hearing Services in Schools, 11*, 208–213.

de Heering, A., Houthuys, S., & Rossion, B. (2007). Holistic face processing is mature at 4 years of age: Evidence from the composite face effect. *Journal of Experimental Child Psychology, 96,* 57–70.

Gauthier, I., & Bukach, C. (2007). Should we reject the expertise hypothesis? *Cognition, 103*, 322-330.

Gauthier, I., Curran, T., Curby, K.M., & Collins, D. (2003). Perceptual interference supports a non-modular account of face processing. *Nature Neuroscience, 6,* 428–432.

Gauthier, I., & Tarr, M.J. (1997). Becoming a 'Greeble' expert: Exploring the face recognition mechanism. *Vision Research, 37,* 1673–1682.

Gauthier, I., & Tanaka, J. W. (2002). Configural and holistic face processing: The Whole story. *Journal of Vision, 2* (7), 616.

Goswami, U., & Bryant, P. (1990). *Phonological skills and learning to read.* Hove, UK: Lawrence Erlbaum.

Goswami, U., & Mead, F. (1992). Onset and rime awareness and analogies in reading. *Reading Research Quarterly*, *27* (2), 153-162.

Hepper, P. G., & Shahidullah, B. S. (1994). Development of fetal hearing. *Archives of Disease in Childhood – Fetal and Neonatal Edition, 71* (2), F81 – F87.

Ho, C. S-H., & Bryant, P. (1997). Development of Phonological Awareness of Chinese Children in Hong Kong. *Journal of Psycholinguistic Research, 26* (1), 109 – 126.

Hole, G. J. (1994). Configurational factors in the perception of unfamiliar faces. *Perception, 23,* 65–74.

Huang, H. S., & Hanley, J. R. (1995). Phonological awareness and visual skills in learning to read Chinese and English. *Cognition, 54*(1), 73-98.

Kwan, T. W. (2001). *A Chinese Character Database: With Word-formations phonologically disambiguated according to the Cantonese Dialect.* Retrieved on October 1st, 2012 from http://humanum.arts.cuhk.edu.hk/Lexis/lexi-can/.

Ladefoged, P., & Disner, S. F. (2012). *Vowels and Consonants, 3rd Edition*. Oxford, UK: Blackwell Publishing.

Maurer, D., Le Grand, R., & Mondloch, C. J. (2002). The many faces of configural processing. *Trends in Cognitive Sciences, 6,* 255–260.

Richler, J. J., Cheung, O. S., & Gauthier, I. (2011). Holistic Processing Predicts Face Recognition. *Psychological Science, 22* (4), 464-471.

Rossion, B. (2013). The composite face illusion: A whole window into our understanding of holistic face perception. *Visual Cognition, 21* (2), 139 - 253.

Sharma, M., Purdy, S., & Kelly, A. (2009). Comorbidity of auditory processing, language, and reading disorders. *Journal of Speech, Language, and Hearing Research, 52,* 706–722.

Shu, H., Peng, H. & McBride-Chang (2008). Phonological awareness in young Chinese children. *Developmental Science, 11*(1), 171–181.

Tanaka, J., & Farah, M. (1993). Parts and wholes in face recognition. *The Quarterly Journal of Experimental Psychology, 46,* 225–245.

Te Papa Atawhai. (2013). *New Zealand bird songs and calls.* Retrieved on February 27th, 2013 from http://www.doc.govt.nz/conservation/native-animals/birds/new-zealand-bird-songs-and-calls/..

Walley, A. C. (1988). Spoken word recognition by young children and adults. *Cognitive Development, 3,* 137 - 165.

Yue-Hashimoto, A. O.-K. (1972). *Studies in Yue Dialects: Phonology of Cantonese.* Cambridge, UK: Cambridge University Press.