Cross-linguistic and cross-scriptal differences in auditory and visual attentional shifts: comparison between native Cantonese and English speakers

Li Wai Lam, Anita

A dissertation submitted in partial fulfilment of the requirements for the Bachelor of Science (Speech and Hearing Sciences), The University of Hong Kong, June 30, 2012.

Cross-linguistic and cross-scriptal differences in auditory and visual attentional shifts: comparison between native Cantonese and English speakers

Li Wai Lam, Anita

Abstract

Lallier and colleagues (2010b) put forward a new hypothesis proposing the role of temporal interval between salient units in ones native language in shaping the speed of attentional shift. The present study investigated the applicability of this hypothesis to Cantonese speakers and English speakers by comparing their speed of attentional shift in auditory and visual stream segregation tasks. Contrary to Lallier et al.'s hypothesis, results of stepwise regressions revealed no group difference in the segregation thresholds in both modalities after controlling the participants' mean reaction time and alerting score in the Flanker task, suggesting that the speed of attentional shift is language-independent. Additionally, this study established the normative data of attentional shift in the typical Cantonese-speaking adults. This information can serve as a basis for evaluating the relevance of "sluggish attentional shift" (SAS) to developmental dyslexia in Chinese with a logographic script, which may provide clinical insights to its diagnosis.

Cross-linguistic and cross-scriptal differences in auditory and visual attentional shifts: comparison between native Cantonese and English speakers

There is a general agreement that attention is crucial for efficient processing of relevant information and attenuation of peripheral information. With respect to language processing, attention plays an important role in mediating the selection of competing candidates such that, for example, the correct word can be activated (Kurland, 2011).

Attention had previously been studied extensively in relation to stages of information processing. Broadbent (1958) postulated that there was a filter to exclude irrelevant information in order to prevent overloading of the mechanism. Broadbent (1958) and Triesman (1960) placed the filter soon after sensory registration, while Deutsch and Deutsch (1963) placed it close to the response end. However, both early and late selection theories faced criticisms for not being able to account for the great variability in experimental results.

Investigators of attention gradually shifted their focus from information approach to neuropsychological approach, which stresses that attention is not a unitary system but an interaction between different mechanisms. Posner and Petersen’s (1990) influential three-network model viewed attention as a multifaceted construct: the alerting network which achieves and maintains an alert state; the orienting network which selects information from sensory input; and the executive control network which resolves conflicts among responses. Particularly related to the present study, the orienting network specifies attentional shift into three sub-steps: disengagement of attention from its present focus, shifting of attention from one stimulus to another, and reengagement of attention to a new stimulus.

Attentional blink task and stream segregation task are the most common tasks for studying attentional shift. Attentional blink task measures the shortest time interval between targets (stimulus onset asynchrony, SOA) in which the second target cannot be identified when it appears close in time to the first. Similarly, stream segregation task measures the shortest SOA when perception of one stream of stimulus changes to two streams. This phenomenon of stream segregation can be observed in both auditory and visual modalities: when sequences of auditory stimuli alternate in frequency, or when sequences of visual stimuli alternate in spatial location. To illustrate, in the auditory modality, when the presentation rate of the tones is slow or when the frequency difference between the tones is small, a single stream of tones alternating in high- and low-pitch is perceived. However, increasing the presentation rate of the tones or the frequency difference between the tones results in perception of two parallel streams of tones, one with higher and the other with lower pitch. Similarly, in the visual modality, when the presentation rate of the dots is slow or when the spatial distance between the dots is small, a single stream of dots appearing above and below the central fixation is perceived. However, increasing the presentation rate of the dots or the spatial distance between the dots results in perception of two continuous streams of dots (Lallier et al., 2009) (Figure 1). Perception of one stream occurs when attention can rapidly disengage from a stimulus then re-engages to process the successive one. Perception of two streams, however, occurs when attention is no longer able to shift fast enough. The shortest SOA to change from one to two streams percept, termed as segregation threshold, is thus an index of the highest speed that attention can shift (Hari, Valta, & Uutela, 1999).

Recently, Lallier and colleagues (2010b) put forward a new hypothesis proposing the role of lexical stress in shaping the speed of auditory attentional shift. In their study, English monolingual adults were found to have significantly lower auditory thresholds (105 ms) than Welsh-English bilinguals (136 ms) matched for age, nonverbal intelligence, vocabulary skills, and general reading and spelling level. However, this effect could not be found in the visual segregation task. Lallier et al. proposed that the speed at which attention can automatically shift mi be constrained by the average length of interval between salient units in ones native language. They argued that lexical stress, defined as higher emphasis given to a particular

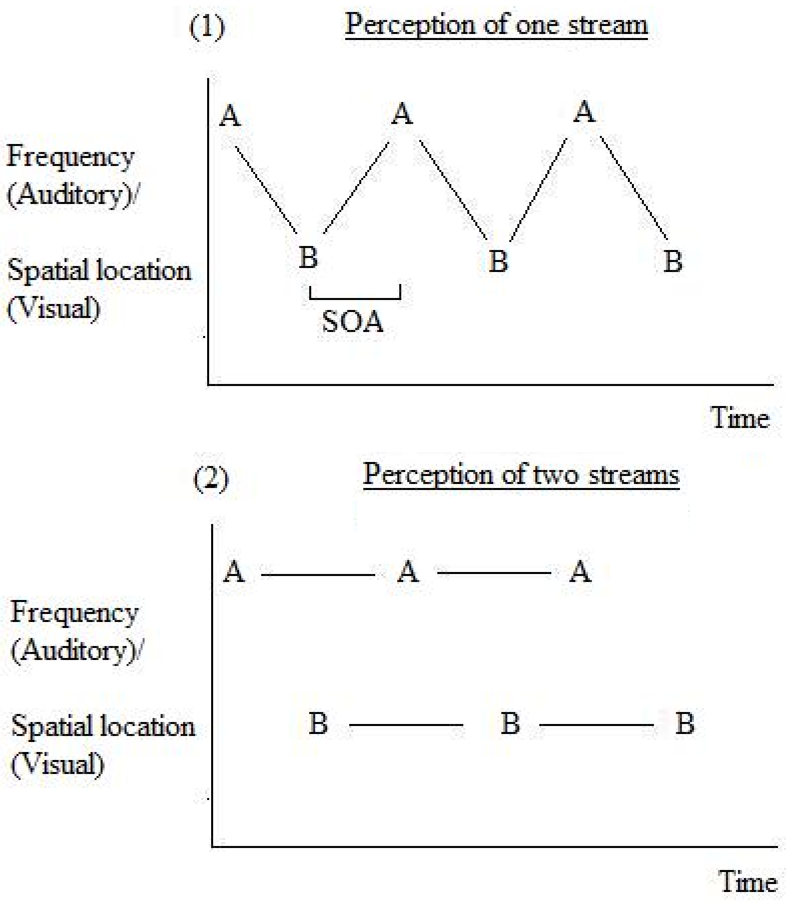

Figure 1. Illustration of the auditory and visual stream segregation phenomenon. One stream (1) or two distinct streams (2) can be perceived.

syllable - by exaggeration of one or more of the phonetic parameters of pitch, loudness, duration, or quality (Laver, 1994), may be the relevant salient unit. This view was supported by Jusczyk (1999) who found that lexical stress act as an important cue for segmentation in speech perception of English-learning infant as early as six months old. According to Williams (1986), post-stress consonant in Welsh is significantly lengthened, but the same phenomenon is not found in English. Therefore, Lallier et al. suggested that the inter-stress temporal interval may be shorter in English than in Welsh, which may shape the auditory attentional shift to be faster in English speakers. Conversely, no group difference was found in the visual task since the two scripts are largely similar.

At present, there have been no other studies that investigated Lallier et al.'s (2010b) hypothesis. Therefore, questions could be raised about the generalizability of Lallier et al.'s findings to populations speaking other languages. Motivated by this research gap, the present study aimed to further examine the role of temporal interval between salient units in ones spoken and written language in shaping the speed of auditory and visual attentional shift respectively. Following Lallier et al.'s logic, Cantonese and English have different stress patterns which may shape the auditory attentional shift. Cantonese is syllable-timed where every syllable receives approximately equal emphasis in emotionally neutral utterances. English is described as stress-timed where only the stressed syllables receive more emphasis (Abercrombie, 1965; Bauer & Benedict, 1997). Mok and Dellwo (2008) reported that the average syllable duration in Cantonese was 178 ms in spontaneously telling of the standard passage "the North Wind and the Sun's story", which can be taken as an estimation of the inter-stress interval in Cantonese. On the other hand, Dauer (1983) found that the mean inter-stress interval in English was 450 ms in reading of a modern novel spoken with everyday language. Since the inter-stress interval is shorter in Cantonese, native Cantonese speakers would be predicted to have lower auditory thresholds than English speakers.

Extending Lallier et al.'s (2010b) hypothesis, Chinese and English also differ in their scripts, which may arguably shape the visual attentional shift. Characters are perceptually salient in Chinese, while words are the relevant salient unit in English. For example, the English word "apple" is written as "蘋果" with two spatially separated characters in Chinese. To study the temporal interval between salient units in reading, Sun and colleagues (1985) compared the reading eye movements for Chinese versus English text using sets of English and their published Chinese translated counterpart articles. Result revealed that readers could read more salient units per minute in Chinese (578 characters) than in English (380 words), suggesting that the temporal distance between two salient units is shorter in the Chinese script. Given this finding, native Cantonese speakers would be predicted to have lower visual thresholds than English speakers.

Interestingly, a considerable amount of studies have found a strong relationship between "sluggish attentional shift" (SAS) and developmental dyslexia in the alphabetical writing system, which suggested the potential clinical importance of segregation thresholds in differentiating between normal children and children with dyslexia. To illustrate, individuals with dyslexia were consistently found to have higher segregation thresholds than controls in studies conducted in the auditory (e.g., Helenius, Uutela, & Hari, 1999) or the visual (e.g., Hari et al., 1999) modality. Few studies have even reported SAS in both modalities in the same group of participants with dyslexia (e.g., Lallier et al., 2010c; Lallier et al., 2009), suggesting that amodal SAS may be the origin of rapid temporal processing deficits in developmental dyslexia. These researchers interpreted the higher thresholds of individuals with dyslexia as implying a problem with shifting of automatic attention, which could affect perception of temporal order of sound in rapidly presented stimuli. SAS in the auditory modality can thus prevent phonemic processing and adequate phonological representations built-up in reading acquisition. The role of SAS in visual modality in reading acquisition is less clear, but Pammer and Vidyasagar (2005) proposed that it may relate to the sequential visual attention processes required to segment the orthographic input before its conversion into a phonological code.

So far, however, this strong relationship between stream segregation deficits and developmental dyslexia was only documented exclusively in British, French and Finnish population with an alphabetical writing system. There were only a few researches investigated the applicability of SAS to a logographic writing system like Chinese; and the findings were rather controversial. For example, a local study by Chak (2008) showed that Chinese dyslexic children and controls did not differ significantly in their auditory segregation thresholds, which did not support the applicability of auditory SAS to developmental dyslexia in Chinese. Another study by Fung (2008) had divergent result on the visual attentional shift: Chinese dyslexic children were found to have visual attentional shifting difficulty as compared with controls in an attentional blink task, which also accounted significantly for the variance in their reading performance. These results seemingly supported the relationship between visual SAS and developmental dyslexia in Chinese. However, only 32% of the participants with dyslexia performed under the level of one standard deviation below the mean of controls, which implied that only a proportion of the individuals with dyslexia exhibited the same visual processing difficulty.

In light of these controversial findings, it was believed that considerably more work would need to be done to explore the relevance of SAS to Chinese dyslexia. Therefore, the second aim of the present study was to collect normative data of auditory and visual attentional shift in the typical Cantonese speaking population in Hong Kong, which will serve as a description of the temporal processing abilities in the typical local population. Further investigation may then focus on measuring the segregation thresholds of the Cantonese dyslexic adults, which can be compared with the current finding to evaluate the applicability of SAS to developmental dyslexia in Chinese. This will enhance our understanding to the possible origin of developmental dyslexia, and may even provide important insights to the development of stream segregation task as a new diagnostic tool of developmental dyslexia.

In sum, the aim of this study was to investigate the influence of linguistic and scriptal differences to the auditory and visual attentional shift ability in Cantonese and English speakers. It tests Lallier et al.'s (2010b) hypothesis proposing the role of temporal interval between salient units in ones native language in shaping the speed of attentional shift, which may stimulate new researches on the theoretical framework of attentional shift in populations speaking other languages. Additionally, due to a lack of research in attentional shift in Chinese, another aim of this study was to establish normative data of the speed of attentional shift for the typical Cantonese-speaking population in Hong Kong. This will serve as a base for investigation of the applicability of SAS in Chinese dyslexia, which may bring important clinical implications to the diagnosis of developmental dyslexia. Following Lallier et al., the speed of attentional shift of the native Cantonese speakers and English speakers were measured using stream segregation tasks in both visual and auditory modalities. It was predicted that native Cantonese speakers would have lower segregation thresholds than English speakers in both modalities.

Method

Participants

Fifty-six adults ranging from 18-33 years old (Lallier et al., 2010b) participated in the present study. Informed consent was obtained. Based on the participants' self report, all of them had normal hearing ability, normal or corrected-to-normal vision, and no history of clinically diagnosed neurological and psychiatric disorders (e.g., Attention Deficit and Hyperactivity Disorder, Autism, and Depression) or developmental dyslexia. All of the participants passed the hearing screening by having thresholds below or equal to 25 dB for the 250, 500, 750, 1000 and 1500 Hz tones. They also did not take any medication or alcohol within 12 hours before the experiment and had slept for more than five hours the night before.

The native Cantonese speakers group consisted of 28 adults (14 males, 1 left-handed and 1 ambidextrous, 26 were born in Hong Kong and 2 in Australia) with a mean chronological age of 21.64 (*SD* = 1.75; Range: 19-27). The native English speakers group consisted of 28 adults (14 males, 2 left-handed, 15 were born in the United States, 11 in the United Kingdom, 1 in South Africa, and 1 in Norway) with a mean chronological age of 21.71 (*SD* = 2.88; Range: 18-29).

Regarding the participants' language background, all native Cantonese speakers acquired Cantonese as first language. However, they were multilingual given the English and

Mandarin language study requirements in local schools. All of them obtained grade D or above for oral, listening, reading and language systems, and writing in the Use of English of the Hong Kong Advanced Level Examination (HKALE). On the other hand, all native English speakers acquired English as first language and were monolingual. They rated the proficiency of their other languages as level two or below in a set of five-level scales (with "one" represents the least and "five" represents the most competent) in the aspects of speaking, listening, reading and writing (See *Appendices A and B* for the self-rating scales in the questionnaires).

Tasks and procedures

Questionnaire. To collect information on inclusion criteria and to rule out the environmental factors that could potentially influence attentional shift, all participants completed a questionnaire inquiring about (1) basic information such as age, gender, handedness, educational level, and place of birth, (2) health conditions such as visual and hearing abilities, psychiatric and neurological conditions, previous diagnosis of developmental dyslexia, duration of sleep, and medication or alcohol use prior to the experiment, (3) language background such as age of acquisition of languages, HKALE Use of English grades, and self rating of language proficiency, and (4) lifestyle factors such as time spent with practicing musical instruments (Bialystok & DePape, 2009), playing ball games (Fontani, Maffei, Cameli, & Polidori, 1999), using computer, and playing video games (Bialystok, 2006). Lifestyle factors included were previously shown to produce global reaction time advantages and reduce interference effect, and thus may affect attentional shift (See *Appendices A and B* for the questionnaires*).*

Hearing screening. To ensure the participants could perceive the 400 and 1000 Hz tones in the stream segregation, their hearing acuity were screened using a pure-tone air conduction hearing test in a sound booth at the University of Hong Kong, which determined the faintest tones that the participants could hear at 250, 500, 750, 1000 and 1500 Hz. The participants wore headphones and were asked to respond to the sounds by raising their hand.

Nonverbal cognitive tasks. All participants completed four nonverbal cognitive tasks: 1) Ravens Standard Progressive Matrices, 2) Flanker task, 3) auditory stream segregation, and 4) visual stream segregation. The order of presentation of these tasks was counterbalanced across participants based on a Latin square design.

Raven's standard Progressive Matrices. To control for possible differences in nonverbal intelligence, the Raven Standard Progressive Matrices (Raven & Summers, 1986) was administered. Participants were asked to select from six to eight alternatives the figure that completed the series best. Raw scores were calculated for each participant.

Flanker task. It was previously reported that multilinguals were better at managing conflicts in more than one active language systems, and this might lead to advantages in general cognitive processing and inhibitory control (Costa, Hernández, Costa-Faidella, & Sebastián-Gallés, 2009). Therefore, the Cantonese multilingual speakers might be able to shift their attention faster than the English monolingual speakers even without account for the participants' native language. In light of this, a Flanker task (Fan, McCandliss, Sommer, Raz, & Posner, 2002) measuring mean reaction time and interference effect was administered to partial out the variance that is related to multilingualism.

Following Fan et al. (2002), the participants were tested individually in a moderately lit sound-proofed booth, and sat 53 cm away from the computer. The participant was required to determine the direction of a central target arrow that pointed to the left or the right, and press the corresponding button of the mouse as quickly and accurately as possible. The target arrow was flanked by two arrows in close spatial proximity on each side, which pointed in the same (congruent) direction as the target arrow or the opposite (incongruent). Before the appearance of the arrows, there were several possible cue conditions: 1) no cue, 2) center cue (warning cue appeared at the central fixation), or 3) spatial cue (warning cue appeared above or below the fixation) (See *Appendix C* for a full description of the Flanker task).

Reaction time (RT) and accuracy for each trial were recorded. Incorrect responses and outliner responses with RT greater than ± 3 *SD* of the mean were excluded from the analysis. Four measures were then calculated: 1) Mean RT for correct trials taken to indicate general cognitive processing efficiency; network scores including 2) Alerting (difference in RTs between trials with no cue and center cue) taken to indicate ability to achieve and maintain an alert state, 3) Orienting (difference in RTs between trials with center cue and spatial cue) taken to measure ability to select information from sensory input, and 4) Interference effect (difference in RTs between congruent and incongruent trials) taken to indicate ability of inhibitory control.

Auditory and visual stream segregation. Following Helenius et al. (1999), the participants were tested individually in a moderately lit sound-proofed booth, and sat 60 cm away from the computer. The participant was required to determine whether they had perceived “one stream” or “two streams’ of the auditory and visual stimuli respectively in 30 trial sequences. The perception of one stream felt similar to listening to alternating high- and low-pitch tones in the auditory task, and seeing a dot bouncing up and down in the visual task. The perception of two streams felt similar to listening to two tones in parallel in the auditory task, and seeing two dots that appear simultaneously in the visual task.

In order to determine the optimal segregation threshold for each participant, a simple “one-up, one-down” adaptive method was used to estimate the 50% chance in a two forced choice paradigm (Levitt, 1971). As long as the answer was “one stream”, the program would decrease the SOA automatically. In contrast, when the answer was “two streams”, the SOA would increase automatically. The stream segregation thresholds for both modalities were computed by averaging the SOAs of the last 10 sequences (21-30) (See *Appendix D* for a full description of the stream segregation tasks).

Data analysis

To compare the auditory and visual segregation thresholds between the two language groups, a between-subject design was used with language group of the participants as the independent variable and the auditory/visual segregation thresholds as the dependent variables. Independent t-tests were used to compare the chronological age, raw score of the Ravens Standard Progressive Matrices, time spent in various lifestyle factors, and mean RT and network scores of the Flanker task to ensure their equivalence between the groups. For variables that failed to meet the assumption of normality and equality of variance, nonparametric Mann Whitney U test was conducted instead. Stepwise multiple regression analyses were then performed to determine the relative amount of variances in auditory/visual segregation thresholds accounted for by the language group and any of the above variables that were unequal between the two language groups.

Results

Background and lifestyle factors

Table 1 shows a summary of the participants' background and lifestyle factors. Based on the significant Kolmogorov-Smirnov test, chronological age, raw score of the Ravens Standard Progressive Matrices, and time spent on various lifestyle factors were non-normally distributed for both groups (See *Appendix E* for the result of the Kolmogorov-Smirnov test). Therefore, nonparametric Mann Whitney tests were used to assess the equivalence of these variables between the groups. According to Table 2, the groups did not differ significantly in all of the above measures, confirming that participants in the two groups were matched with chronological age, nonverbal intelligence and experience with various lifestyle factors that might covary with attentional shift.

Table 1. *Summary of the participants' background and lifestyle factors*

	Cantonese speakers				English speakers			
	M	*SD*	*Mdn*	Range	*M*	*SD*	*Mdn*	Range
Age (years)	21.64	1.75	21.5	19-27	21.71	2.88	21.0	18-29
Ravens	57.64	2.06	58.0	53-60	56.82	2.54	58.0	52-60
Time spent on (hours per week)								
Musical instruments	1.52	2.35	1.0	0-12	1.32	2.47	0.0	0-8
Ball games	2.36	2.84	2.0	0-12	2.48	3.56	0.0	0-14
Use of computer	36.32	17.44	30.0	0-70	30.36	11.26	30.0	15-55
Videogames	1.39	2.10	0.0	0-7	1.82	3.5	0.0	0-15

Note. *Mdn* = Median; Ravens = raw score of the Ravens Standard Progressive Matrices

Table 2. *Mann Whitney Test results for background and lifestyle factors*

	Mann Whitney U Test			
	U	*z*	*p*	*r*
Age (years)	342	-0.83	.40	-.11
Ravens	329	-1.05	.29	-.14
Time spent on (hours per week)				
Musical instruments	310	-1.47	.14	-.20
Ball games	365	-0.46	.65	-.06
Use of computer	336	-0.94	.35	-.13
Videogames	364	-0.17	.87	-.02

Note. Ravens = raw score of the Ravens Standard Progressive Matrices

Flanker task

For all response time analyses, incorrect responses were excluded. Besides, individual RTs greater than ± 3 *SD* of the mean were considered as outliers and eliminated. After trimming the data, mean RT and network scores were calculated for each participant. According to the non-significant Kolmogorov-Smirnov test and Levene's test, Mean RT and network scores were checked to fulfill the assumptions of normality and equality of variance respectively. Therefore, two-tailed independent t-tests were used to determine whether the groups differ in these measures. To be more conservative to the effect of these confounding variables, correction for multiple comparisons was not applied. As shown in Table 3, the Cantonese speakers had significantly faster mean RT than the English speakers. For the network scores, the Cantonese speakers had significantly higher alerting score than the English speakers. However, orienting and interference effect were not significantly different between the groups. Given these results, the mean RT and alerting score would be taken into consideration when evaluating the effect of native language on attentional shift.

Table 3. *Mean RT (ms) and network scores (ms) in the Flanker task for each language group*

	Cantonese speakers		English speakers		Independent t-test		
	M	*SD*	*M*	*SD*	$t(54)$	*p*	*r*
Mean RT (ms)	472.65	30.82	495.40	35.94	2.54	.01*	.33
Network scores (ms)							
Alerting	28.16	15.90	17.02	21.14	-2.23	.03*	.29
Orienting	45.65	20.29	50.64	24.31	0.83	.41	.11
Interference effect	71.84	22.49	82.48	23.67	1.73	.09	.23

* $p < .05$

Stream segregation tasks

Auditory stream segregation. Figure 2 depicts the mean results for both groups in the auditory task. The Cantonese speakers (n = 28) were found to have mean auditory threshold of 71.23 ± 46.15 ms, and the English speakers (n = 28) were found to have mean threshold of 75.66 ± 43.06 ms. Since the auditory thresholds were non-normally distributed, a square root transformation was performed to improve normality. However, it was still non-normally distributed after the transformation; therefore, three participants from the Cantonese group (number 12, 17, and 21) were excluded from the analysis due to their extremely low thresholds of 21.5 ms. Normality was resumed after excluding the outliners.

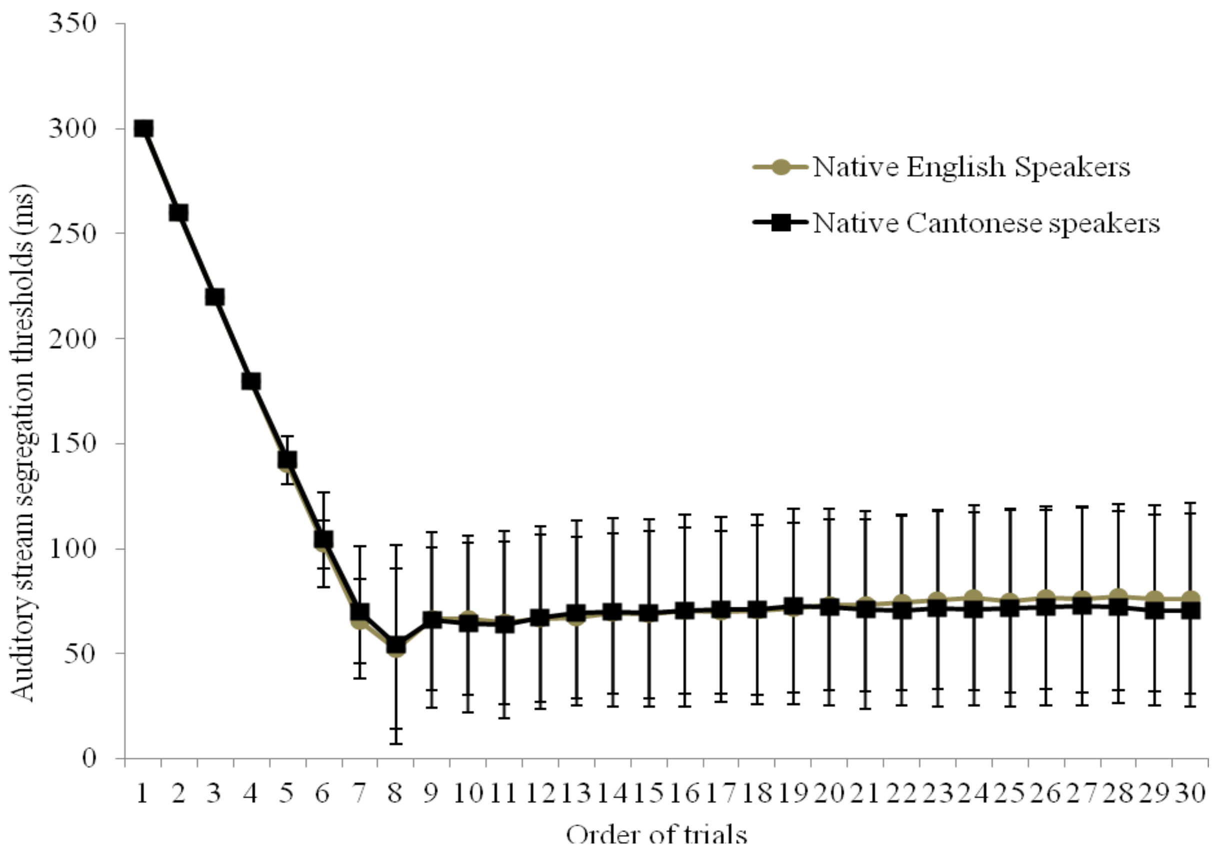

Figure 2. Mean auditory stream segregation threshold (with standard error bars) for the native English speakers (grey circles; n = 28) and native Cantonese speakers (black squares; n = 28) groups. Statistical analyses were performed on the average of trials 21 to 30.

A stepwise multiple regression was then conducted with (square root of) the auditory segregation threshold entered as the dependent variable, and mean RT, alerting score and language group entered as predictors. The assumptions of nonmulticollinearity, homoscedasticity, normally distributed error, and linearity were met. According to Table 4, the combination of the three predictors only accounted for non-significant percentage (8.9%) of variance in (square root of) the auditory threshold. All of the three predictors, including the language group, did not significantly predict the (square root of) the auditory threshold.

Table 4. *Stepwise regression predicting (square root of) the auditory segregation threshold*

	Variables	R^2	Adjusted R^2	R^2 change	*p*
Step 1	Mean RT	0.041	0.023	/	.144
Step 2	Mean RT	0.071	0.034	0.03	.157
	Alerting score				
Step 3	Mean RT	0.089	0.034	0.02	.200
	Alerting score				
	Language group				

Visual stream segregation. Figure 3 depicts the mean results for both groups in the visual task. The Cantonese speakers (*n* = 28) were found to have mean visual threshold of 103.55 ± 46.96 ms, and the English speakers (*n* = 28) were found to have mean threshold of 130.68 ± 45.95 ms. Since the visual thresholds were negatively skewed and non-normally distributed, a reflect and log transformation was performed to improve normality. However, it was still non-normally distributed after the transformation; therefore, one participant from the English group (number 22) was excluded from the analysis due to her extremely high threshold of 221.5 ms. Normality was resumed after excluding the outliner.

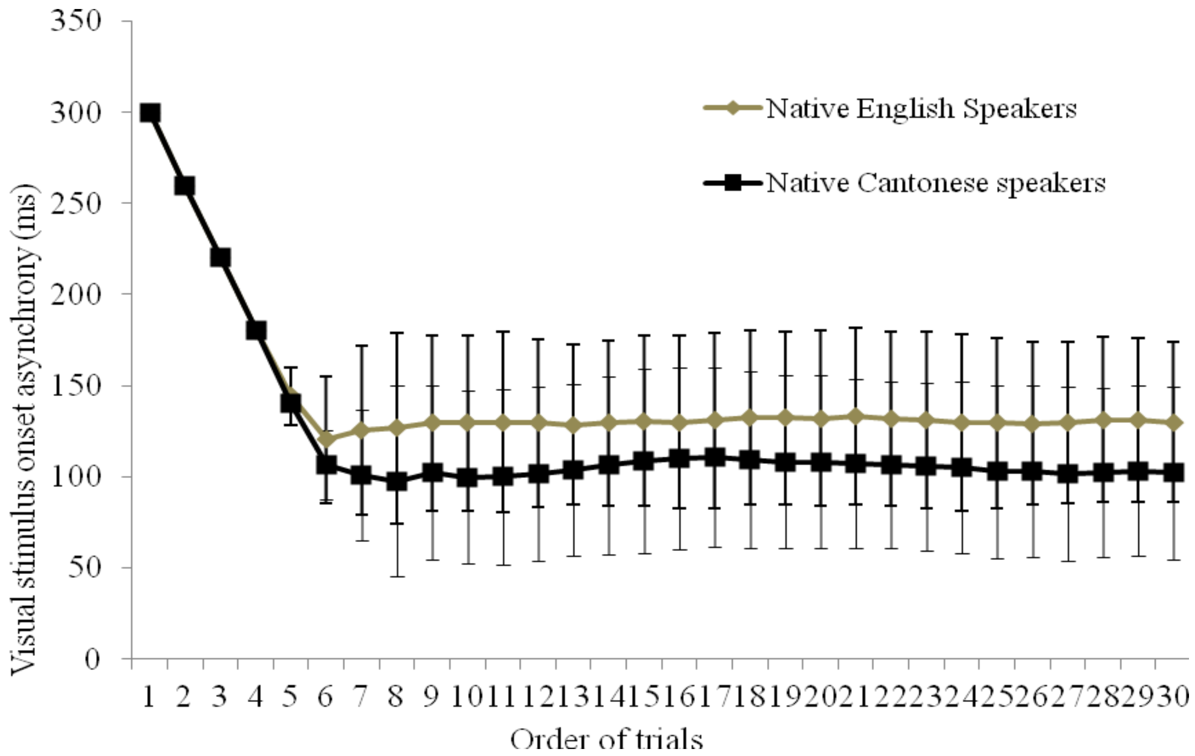

Figure 3. Mean visual stream segregation threshold (with standard error bars) for the native English speakers (grey circles; n = 28) and native Cantonese speakers (black squares; n = 28) groups. Statistical analyses were performed on the average of trials 21 to 30.

A stepwise multiple regression was then conducted with (reflect and log of) the visual segregation threshold entered as the dependent variable, and mean RT, alerting score, and language group entered as predictors. The assumptions of nonmulticollinearity, homoscedasticity, normally distributed error, and linearity were met. As shown in Table 5, the three predictors variables together could predict 13.9% of the variances in the (reflect and log of) visual threshold, with the model approaching borderline significance (p = .052). After step 1, with mean RT entered as predictor, 9% the variance could be significantly predicted. After step 2, with alerting score added as predictor, an additional 3.4% of the variance could be significantly predicted. After step 3, the addition of language group to the model did not reliably improve R^2. Therefore, the language group of the participant did not have significant predictive power in (reflect and log of) the visual threshold beyond that contributed by mean RT and alerting score. In contrast, mean RT and alerting score together significantly accounted for 12.5% of the variance in (reflect and log of) the visual threshold.

Table 5. *Stepwise regression predicting (reflect and log of) the visual segregation threshold*

	Variables	R^2	Adjusted R^2	R^2 change	*p*
Step 1	Mean RT	0.090	0.073	/	.026*
Step 2	Mean RT	0.125	0.091	0.034	.031*
	Alerting score				
Step 3	Mean RT	0.139	0.088	0.014	.052
	Alerting score				
	Language group				

* $p < .05$

Standardized betas of the regression are presented in Table 6, which showed that only the mean RT made a significant unique contribution to the prediction of (reflect and log of) the visual segregation threshold, but not the language group or the alerting score. Although the beta of the mean RT appeared to be negative, the values of the visual segregation thresholds were changed after the data transformation. Figure 4 shows a scatterplot of the original value of visual segregation thresholds against mean RT, which indicated a clearer positive association between mean RT and visual segregation thresholds, particularly when the mean RT was low: if the participant had significantly low mean RT, the participant tended to have low visual segregation threshold as well. However, this positive association between mean RT and visual segregation threshold became weaker with higher mean RT.

Table 6. *Standardized coefficients for regression equation predicting (reflect and log of) the visual segregation threshold*

	Variables	β	*t*	*p*
Step 1	Mean RT	-.301	-2.296	.026*
Step 2	Mean RT	-.296	-2.284	.027*
	Alerting score	.186	1.430	.159
Step 3	Mean RT	-.252	-1.821	.074
	Alerting score	.150	1.105	.274
	Language group	.132	0.915	.365

* $p < .05$

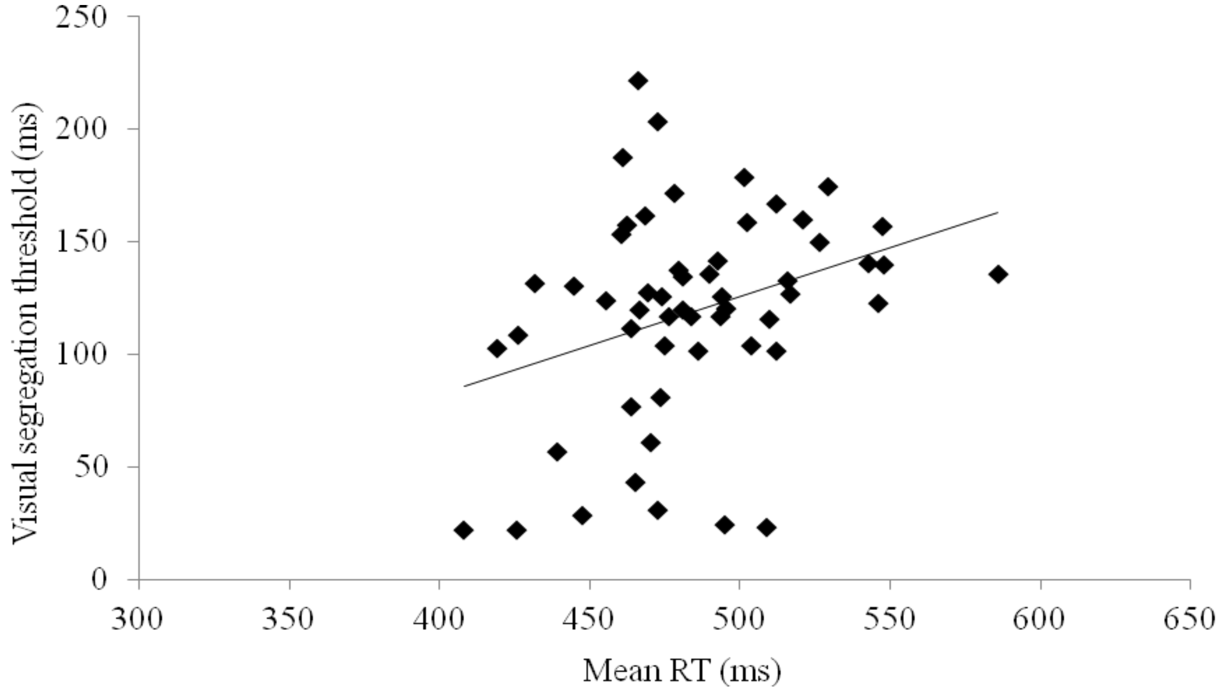

Figure 4. Scatterplot of visual segregation threshold against mean RT in the Flanker task (with a line of best fit). A positive association can be observed between mean RT and visual segregation thresholds, particularly when the mean RT is low.

Discussion

The present study was designed to examine if temporal interval between salient units in ones spoken and written language can shape the speed of auditory and visual attentional shift respectively, by comparing the auditory and visual segregation thresholds of native Cantonese speakers and native English speakers. Lallier et al. (2010b) showed that English speakers had significantly lower auditory thresholds than Welsh speakers, while their visual thresholds were largely similar. Lallier et al. attributed such findings to the shorter inter-stress interval in English, which may shape the speed of auditory attentional shift to be faster. On the other hand, since the English script and the Welsh script are largely similar, no group difference was found as expected.

Extending Lallier et al.'s (2010b) hypothesis, it was predicted that Cantonese speakers would have lower segregation thresholds than English speakers in both modalities, given the shorter temporal interval between stress in Cantonese, and between orthographic salient units (i.e., characters) in Chinese script. Contrary to our predictions, Cantonese speakers and English speakers did not differ significantly in either auditory or visual thresholds after controlling for their mean RT and alerting score in the Flanker task. Therefore, it would appear that Lallier et al.'s hypothesis was not supported. Rather, the results suggest that the speed of attentional shift is language-independent.

The inconsistency between our findings and those obtained by Lallier et al. (2010b), particularly in the auditory thresholds, might be accounted for by three possible reasons. Firstly, although the duration of post-stress consonant is often lengthened in Welsh but not in English (Williams, 1986); more evidence, such as the average number of syllables between stresses, is required to support Lallier et al.'s claim of shorter inter-stress temporal interval in English. This challenged the basis of Lallier et al.'s hypothesis, which suggested the role of inter-stress interval in shaping the speed of auditory attentional shift. Additionally, it is possible that Lallier et al.'s hypothesis was not applicable to the Cantonese speakers. Given the abundant evidence of the importance of lexical stress in cueing for word segmentation in English-learning infants (e.g., Jusczky, 1999), Lallier et al. hypothesized that temporal interval between the lexical stress may shape the speed of attentional shift in early infancy. It should be noted that, however, Cantonese-learning infants may not use the same stress-based strategy in segmenting words, since stress occurs in every syllable of Cantonese and may not act as a useful cue for word segmentation. Following this logic, inter-stress interval in Cantonese may not be a relevant unit in shaping the speed of auditory attentional shift, and this may justify the different finding between Lallier et al.'s study and the present study.

Secondly, in Lallier et al.'s (2010b) study, the English speakers and Welsh speakers were only matched for age, nonverbal intelligence, vocabulary skills, general reading level and spelling level. No attempt was made to measure and control for the participants' overall RT and network scores in the Flanker task, which could be affected by multilingualism (Costa et al., 2009) and amount of time spent in different lifestyle factors (Bialystok, 2006; Bialystok & DePape, 2009; Fontani et al., 1999), and may in turn covary with attentional shift. In the present study, the mean RT and alerting score, taken to measure general cognitive processing efficiency and ability to maintain an alert state respectively, were found to be significantly different between the two language groups. Therefore, these two variables were entered as predictors in the stepwise regressions to control for their effects. Result of the regression revealed that none of the variables significantly predicted the variance in the auditory threshold. On the other hand, among the predictor variables, only the mean RT significantly accounted for the variance in the visual threshold. This suggested that the mean RT, but not the native language, played a role in shaping the speed of visual attentional shift.

To illustrate the effect of the uncontrolled confounding variables, an additional analysis was conducted to compare the visual thresholds between the two groups with an nonparametric Mann Whitney U test, without adjusting for mean RT and alerting score. An entirely different result emerged from this analysis: the visual thresholds were significantly lower for the Cantonese speakers (*M* = 103.55 ms; *SD* = 46.96 ms) than the English speakers (*M* = 130.68 ms; *SD* = 45.95 ms), $U = 249$, $z = -2.35$, $p < .05$, $r = -.32$, which would have led to the false conclusion of the presence of scriptal influence in predicting the visual attentional shift. Although the present study did not find any effect of the mean RT and network score on the auditory attentional shift, one may still question whether failure to control for these variables in Lallier et al.'s (2010b) study might have resulted in an apparently significant difference in auditory thresholds between the two language groups.

Thirdly, Lallier et al. (2010b) recruited only 12 English speakers and 12 Welsh speakers for comparison of their auditory thresholds. Given this small sample size, the generalizability of the comparison to the population was questionable. In contrast, the present study compared the auditory thresholds of 28 Cantonese speakers and 25 English speakers, which was about twice as large as the sample size in Lallier et al.'s study. Therefore, it is believed that the comparison of segregation thresholds in the present study had better generalizability.

Having accounted for the difference between Lallier et al.'s (2010b) finding and the current finding in the auditory task, we would then turn our attention to the visual task. Incompatible with our prediction, the present study demonstrated that Cantonese speakers and English speakers did not differ significantly in the visual thresholds after adjusting for their mean RT and alerting score in the Flanker task, whereas the mean RT significantly accounted for 9% of the variance in the (reflect and log of) the visual threshold. This seems to indicate that the mean RT, a nonverbal measure of the general cognitive processing efficiency, has a stronger predictive power on the visual attentional shift ability than the participants' native language. This result was not too surprising, because both the visual stream segregation and the Flanker task measured nonverbal cognitive abilities in the visual modality. It is reasonable that if a participant has higher general cognitive processing efficiency in the Flanker task, the participant may also have advantage in shifting visual attention. However, interestingly, this positive association between mean RT and visual thresholds was only strong in low mean RT condition. This finding provides important implications for further investigation of the nature of association between mean RT and the speed of visual attentional shift.

Taken together, the current findings in visual and auditory segregation suggest that one's native language, a verbal factor, does not play a role in predicting the speed of auditory and visual attentional shift, which are measures of nonverbal abilities. In contrast, the nonverbal measure of mean RT has a stronger link with the visual attentional shift. So far, however, there has been little investigation on this topic to adequately evaluate Lallier et al.'s (2010b) hypothesis. It would be interesting to compare the current findings with populations speaking other languages, such as Mandarin. Information on segregation thresholds from different populations would help us to evaluate Lallier et al.'s hypothesis more fully.

The second purpose of the present study was to establish normative data of the auditory and visual attentional shift for the typical Cantonese-speaking population in Hong Kong. It was found that the mean segregation thresholds for Cantonese speakers were 71.23 ± 46.15 ms and 103.55 ± 46.96 ms for the auditory and visual modalities, respectively. Future studies may measure the auditory and visual segregation thresholds of the Chinese dyslexic adults, which may then be compared with the current findings to evaluate the applicability of SAS in Chinese dyslexia. This may contribute to the research of the origin of developmental dyslexia, and may motivate the development of early diagnostic tools for developmental dyslexia.

Additionally, it would be worthwhile to compare the segregation thresholds of the English speakers found in the present study with those reported from other studies. Several studies have reported that the mean auditory threshold of normal English-speaking adults ranged from 105 to 131 ms, and the mean visual threshold ranged from 104 to 129 ms (Lallier et al., 2010b; Lallier et al., 2009, Lallier et al., 2010c). From these results, it was evident that the mean visual threshold found in the current study for the English monolingual adults (130.68 ± 45.95 ms) corresponded closely with other studies. However, unexpectedly, the mean auditory threshold found (75.66 ± 43.06 ms) was around 30 to 55 ms lower. Further investigation is warranted to account for this surprising finding.

Another interesting finding that is worth our attention was that the two language groups significantly differ in the overall mean RT, but not in the interference effect measured by the Flanker task. This indicated that multilingualism promoted general cognitive processing efficiency in multilingual Cantonese speakers, but had no effect on their inhibitory control ability. A considerable amount of literature with controversial results on this issue has been published. Several studies have demonstrated that multilingual speakers had smaller interference effect in the Flanker task than monolingual speakers (e.g., Costa et al., 2009), suggesting that multilingual speakers have a higher ability in suppressing irrelevant information. This might be attributed to their frequent use of the inhibitory processes in managing conflicts between multiple languages. However, as reviewed by Hilchey & Klein (2011), this inhibitory control advantage was rather rare and only few studies have reported this effect. On the other hand, there were clearer evidences to suggest that multilingual speakers enjoy general cognitive processing advantages, as indexed by the faster overall mean RT in different experimental conditions (e.g., Bialystok, Craik, Klein, & Viswanathan, 2004; Costa et al., 2009). Therefore, the current finding further supported the notion of presence of general cognitive processing advantages in multilingualism, instead of the rather sporadic phenomenon of inhibitory control advantage.

Apart from the mean RT, it was also found in the present study that multilingual Cantonese speakers had significantly larger alerting effect (28.16 ± 15.90 ms) than monolingual English speakers (17.02 ± 21.14 ms). This type of relationship was previously found in several studies (e.g., Costa et al., 2009). Costa and colleagues proposed that this larger alerting effect might help multilingual speakers to reach and maintain a state of alertness, which prepared the system for orienting attention or resolving conflict. However, the current finding showed no significant difference in the orienting score and interference effect between the multilingual Cantonese speakers and the monolingual English speakers. Therefore, at present we do not have adequate evidence to give an account of what advantages are brought by the improved alerting network in multilingual speakers. Considerably more work will need to be done to explore this issue.

The most important limitation of this study lies in the fact that it was impossible to recruit Cantonese monolinguals in Hong Kong, given the English and Mandarin learning requirements in local schools. However, considering the multilingual advantages in general cognitive processing, the Cantonese multilingual speakers may have advantages in shifting their attention, and this would result in uncertainty of the true effect of native language on attentional shift. To address this limitation, the present study measured the mean RT of the participants from both groups in a Flanker task, and then regressed out its variance in predicting the segregation thresholds. For further study, this limitation may be addressed by recruiting Cantonese monolinguals in the Guangdong Province of the mainland China.

Finally, the present study provided some important insights for future research. Firstly, it would be interesting to replicate this study with populations speaking other languages, such as Mandarin, which will provide more evidence in evaluating Lallier et al.'s (2010b) hypothesis. Secondly, further investigation would need to be conducted in measuring the auditory and visual segregation thresholds of Cantonese dyslexic adults. The comparison of their thresholds with the normative data obtained from the present study would then help to evaluate the applicability of SAS in Chinese dyslexia. If SAS is found in Chinese dyslexia, stream segregation task may serve as a new and cost-effective screening tool in differentiating between normal children and children with dyslexia. Thirdly, despite the absence of evidence for linguistic influence on attentional shift, there has been a large body of literature reporting the strong link of SAS with developmental dyslexia in alphabetic scripts, suggesting a positive relationship between attentional shift and language ability. In a similar vein, the present study can be replicated in children with specific language impairment (SLI) or other language problems, so as to investigate if SAS is a common origin of these language impairments. Fourthly, this study has also raised questions of multilingual advantages in alerting network, which is in need of further investigation.

Conclusion

To sum up, incompatible with Lallier et al.'s (2010b) hypothesis, the present study did not find any evidence of linguistic and scriptal influences in shaping the speed of attentional shift. Rather, the mean RT in the Flanker task, which represented the participants' general cognitive processing efficiency, was found to have significant predictive power on the visual attentional shift. This suggested that the speed of visual and auditory attentional shift, which are measures of nonverbal abilities, were better predicted by the nonverbal measure of general cognitive processing efficiency instead of the native language of the participants. In addition, the present study also established the normative data of the auditory and visual attentional shift in the typical Cantonese-speaking population in Hong Kong. This information can then serve as a valuable basis for evaluating the relevance of SAS to Chinese dyslexia.

Acknowledgement

The author would like to take this opportunity express her deepest and sincere gratitude to her supervisors, Dr. Law Sam-Po and Dr. Su I-Fan, for their enormous guidance, valuable advice, and unconditional support throughout the year. Thanks are also extended to all the fifty-five participants who kindly joined the study. Lastly, a word of appreciation should also go to the author's dear friends and family for their love, patience, and encouragement.

References

Abercrombie, D. (1965). *Studies in phonetics and linguistics*. London: Oxford University Press.

Bauer, R. S., & Benedict, P. K. (1997). *Modern Cantonese phonology*. New York: Mouton de Gruyter.

Bialystok, E. (2006). Effect of bilingualism and computer video game experience on the Simon task. *Canadian Journal of Experimental Psychology, 60*(1), 68-79.

Bialystok, E., Craik, F. I. M., Klein, R., & Viswanathan, M. (2004). Bilingualism, aging, and cognitive control: Evidence from the Simon task. *Psychology and Aging, 19*, 290-303.

Bialystok, E., & DePape, A. M. (2009). Musical expertise, bilingualism, and executive functioning. *Journal of Experimental Psychology: Human Perception and Performance, 35*(2), 565-574.

Broadbent, D. E. (1958). *Perception and communication*. London: Pergamon Press.

Chak, K. W. (2008). *Do Chinese children with developmental dyslexia have deficit in auditory temporal processing?* **(**Unpublished master's thesis). University of Hong Kong, Hong Kong.

Costa, A., Hernández, M., Costa-Faidella, J., & Sebastián-Gallés, N. (2009). On the bilingual advantage in conflict processing: Now you see it, now you don't. *Cognition, 113*, 135-149.

Dauer, R. M. (1983). Stress-timing and syllable-timing reanalyzed. *Journal of Phonetics, 11*, 51-62.

Deutsch, J., & Deutsch, D. (1963). Attention: Some theoretical considerations. *Psychological Review, 70*, 80-90.

Fan, J., McCandliss, B. D., Sommer, T., Raz, A., & Posner, I. (2002). Testing the efficiency and independence of attentional networks, *Journal of Cognitive Neuroscience, 14*(3), 340-347.

Fontani, G., Maffei, D., Cameli, S., & Polidori, F. (1999). Reactivity and event-related potentials during attentional tests in athletes. *European Journal of Applied Physiology, 80*, 308-317.

Fung, K. S. (2008). *Do Chinese children with developmental dyslexia have deficit in visual temporal processing?* **(**Unpublished master's thesis). University of Hong Kong, Hong Kong.

Hari, R., Valta, M., & Uutela, K. (1999). Prolonged attentional dwell time in dyslexic adults. *Neuroscience Letters, 271*(3), 202-204.

Helenius, P., Uutela, K., & Hari, R. (1999). Auditory stream segregation in dyslexic adults. *Brain, 122*, 907-913.

Hilchey, M. D., & Klein, R. M. (2011). Are there bilingual advantages on non-linguistic interference tasks? Implications for the plasticity of executive control processes. *Psychonomic Bulletin & Review, 18*, 625-658.

Jusczyk, P. W. (1999). How infants begin to extract words from speech. *Trends in Cognitive Sciences, 3*(9), 323-328.

Kurland, J. (2011). The role that attention plays in language processing. *Perspectives on Neurophysiology and Neurogenic Speech and Language Disorders, 21*, 47-54.

Lallier, M., Berger, C., Donnadieu, S., & Valdois, S. (2010a). A case study of developmental phonological dyslexia: Is the attentional deficit in the perception of rapid stimuli sequences amodal? *Cortex, 46*(2), 231-241.

Lallier, M., Carreiras, M., Tainturier, M. J., & Thierry, G. (2010b, November). *Does the specific acoustic structure of a language shape auditor attention underlying speech perception? Behavioral and ERP evidence in Welsh-English bilinguals.* Symposium conducted at The 2nd Neurobiology of Language Conference, San Diego, California, USA.

Lallier, M., Tainturier, M. J., Dering, B. R., Donnadieu, S., Valdois, S., & Thierry, G. (2010c) Behavioral and ERP evidence for amodal sluggish attentional shifting in developmental dyslexia. *Neuropsychologia, 48*(14), 4125-4135.

Lallier, M., Thierry, G., Tainturier, M. J., Donnadieu, S., Peyrin, C., Billard, C., & Valdois, S. (2009). Auditory and visual stream segregation in children and adults: An assessment of the amodality assumption of the 'sluggish attentional shifting' theory of dyslexia. *Brain Research, 1302*, 132-147.

Laver, J. (1994). *Principles of phonetics.* Cambridge: Cambridge University Press.

Levitt, H. (1971). Transformed up-down methods in psychoacoustics. *The Journal of the Acoustical S*ociety *of America*, *49*, 467-477.

Mok, P., & Dellwo, V. (2008) Comparing native and non-native speech rhythm using acoustic rhythmic measures: Cantonese, Beijing Mandarin and English. *Proceedings of the 4th Speech Prosody* (pp. 423-426). Campinas, Brazil.

Pammer, K., & Vidyasagar, T. R. (2005). Integration of the visual and auditory networks in dyslexia: a theoretical perspective. *Journal of Research in Reading, 28*, 320-331.

Posner, M. I., & Petersen, S. E. (1990). The attention systems of the human brain. *Annual Review of Neuroscience*, *13*, 25-42.

Raven, J. C., & Summers, B. (1986). *Manual for Raven's Progressive Matrices and Vocabulary Scales - Research supplement no. 3*. London: Lewis.

Sun, F., Morita, M., & Stark, L. A. (1985). Comparative patterns of reading eye movement in Chinese and English. *Perception and Psychophysics*, *27*, 502-507.

Treisman, A. (1960). Contextual cues in selective listening. *Quarterly Journal of Experimental Psychology, 12*, 242-248.

Williams, B. (1986). An acoustic study of some features of Welsh prosody. In C. Johns-Lewis (Ed.), *Intonation is discourse* (pp. 35-51). London: Croom-Helm.

Appendix A: Questionnaire for native Cantonese speakers

The University of Hong Kong
Faculty of Education
Division of Speech and Hearing Sciences
Health, Language, and Lifestyle Questionnaire

A. <u>Basic information</u>
Participant No. (Filled by researcher): ________________________
Age: _____ Gender: _____
Place of Birth: _______________ Handedness: Left / Right / Both
Current/Completed Education Level: High School / Undergraduate / Postgraduate

B. <u>Health condition</u>
Q1. Have you had any alcoholic beverages within12 hours prior to answering this questionnaire?
☐Yes ☐No

Q2. Have you taken any drugs or medication within 12 hours prior to answering this questionnaire?
☐Yes ☐No

Q3. How many hours of sleep did you have last night? _____ hours

Q4. Do you have normal or corrected-to-normal (e.g. by wearing glasses/ contact lens) vision?

☐Yes ☐No, please specify: ____________________

Q5. Have you had history of hearing loss/problems?
☐Yes, please specify: ____________________ ☐No

Q6. Have you had history of neurological and psychiatric conditions? (e.g. Attention Deficit and Hyperactivity Disorder, Autism, Depression)
☐Yes, please specify: ____________________ ☐No

Q7. Have you had history of reading and writing difficulties?
☐Yes, please specify: ____________________ ☐No

C. <u>Language background</u>
Q1. What is your native language? ____________________ (e.g. Cantonese, English, Mandarin etc)

Q2. If you have taken the **<u>Hong Kong A-level Examination</u>**, what were the grades you achieved in the **<u>Use of English</u>** examination?
(a) Oral: _____ ; (b) Listening: _____ ;
(c) Reading and Language Systems: _____; (d) Writing: _____

Q3. Apart from Cantonese and English, what other language(s) do you use in daily life? How old were you when you began acquiring the language(s)?

Language A: _________; _____ years old; Language B: _________; _____ years old
Language C: _________; _____ years old

Q4. Below are descriptive statements which represent a wide range of abilities in speaking, listening, reading and writing. Place a √ in the box of the level that you think best represents your ability in each of the languages you wrote down in Q3.

(a) Speaking:

Level	**Descriptive statements**	**Language (s)**		
		A	**B**	**C**
1	I can speak only simple words and phrases (*e.g. "Hello", "Thank you").*			
2	I can participate in conversations on everyday topics (*e.g. when ordering food in a restaurant),* without the help of any complementary communication *(e.g. gestures, writing).*			
3	I can explain my point of view freely in some formal conversations (*e.g. academic discussion).*			
4	I can speak fluently in most conversations on practical, social and professional topics, but I may need to pause quite a lot when expressing more complex idea.			
5	I can speak the language as well as an educated native.			

(b) Listening:

Level	**Descriptive statements**	**Language (s)**		
		A	**B**	**C**
1	I know only common words and phrases, (*e.g. "Hello", "Thank you").*			
2	I can understand essential points in a *face-to-face* conversation on everyday topics (e.g. *when buying food)*; but I have *some difficulties* in understanding messages heard through *radio or telephone.*			
3	I can understand and infer other's point of view adequately in some formal conversations (*e.g. academic discussion).* I can understand messages heard through *radio and telephone with minor difficulties.*			
4	I can understand what is said in most conversations on practical, social and professional topics in *both face-to-face and telephone conversations*; but I cannot understand idioms and humors.			
5	I can understand the language as well as an educated native.			

You may omit the following parts for "reading" and "writing" if a writing system is not applicable to the language.

(c) Reading:

Level	Descriptive statements	Language (s)		
		A	B	C
1	I can only understand simple written texts or directions for familiar things (*e.g. "Hello", "Thank you")*.			
2	I can understand the meaning of simple texts about familiar subjects (*e.g. when reading menu in a restaurant).*			
3	I can understand the main content of some newspaper articles but not all of the details (e.g. *news about a car accident*).			
4	I can understand most things written in the language, but I need to look for dictionary for less common words.			
5	I can read and understand the language as well as an educated native.			

(d) Writing:

Level	Descriptive statements	Language (s)		
		A	B	C
1	I can only write simple words and phrases e.g. *"(your name)"*			
2	I can write simple texts connected with my own life and my needs (*e.g. writing a shopping checklist)*			
3	I can write a simple article about familiar subjects (*e.g. description of my favorite leisure time activities)*			
4	I can write about most topics in the language, but I need to look for dictionary for less common words			
5	I can write the language as well as an educated native			

D. Lifestyle

Q1a. In general, how much time do you spend on playing ball games(s)? _____ hours per week. *(If you do not play any, please skip and move to Q2)*

b. What kind of ball game(s) do you usually play?

1. ________________ 2.__________________ 3._________________

Q2. In general, how often do you use computer? _____ hours per week

O3a.In general, how often do you play video games? _____ hours per week (*If you do not play any, please skip and move to Q4)*

b: What kind of video games do you usually play? *(You can tick more than one box)*

- ☐ Adventure
- ☐ Shooting
- ☐ RPG
- ☐ Car Racing
- ☐ Simulation
- ☐ Sports
- ☐ Real time strategy game
- ☐ Sports
- ☐ Turn-based strategy game
- ☐ Action
- ☐ Others

Q4a. Do you play any musical instrument(s)? If yes, what do you play?

1. ________________ 2.__________________ 3._________________

b. How old were you when you began learning the instrument(s)?

1. _____ years old 2. _____ years old 3. _____ years old

c. In general, how much time do you spend on musical practice/performance? ___ hours per week

d. Did you take any Graded exam (e.g. ABRSM) on instruments and obtain a grade?

If yes, please specify: ________________________________

e. What is the context of performing your instruments? (e.g. leisure, orchestra, recital) __________

Thank you very much for your kind participation in this research.

Appendix B: Questionnaire for native English speakers

The University of Hong Kong
Faculty of Education
Division of Speech and Hearing Sciences
Health, Language, and Lifestyle Questionnaire

A. Basic information
Participant No. (Filled by researcher): ________________________
Age: _____ Gender: _____
Place of Birth: ________________ Handedness: Left / Right / Both
Current/Completed Education Level: High School / Undergraduate / Postgraduate

B. Health condition
Q1. Have you had any alcoholic beverages within12 hours prior to answering this questionnaire?
☐Yes ☐No

Q2. Have you taken any drugs or medication within 12 hours prior to answering this questionnaire?
☐Yes ☐No

Q3. How many hours of sleep did you have last night? _____ hours

Q4. Do you have normal or corrected-to-normal (e.g. by wearing glasses/ contact lens) vision?

☐Yes ☐No, please specify: ____________________

Q5. Have you had history of hearing loss/problems?
☐Yes, please specify: ____________________ ☐No

Q6. Have you had history of neurological and psychiatric conditions? (e.g. Attention Deficit and Hyperactivity Disorder, Autism, Depression)
☐Yes, please specify: ____________________ ☐No

Q7. Have you had history of reading and writing difficulties?
☐Yes, please specify: ____________________ ☐No

C. Language background
Q1. What is your native language? ____________________ (e.g. Cantonese, English, Mandarin etc)

Q2. Apart from English, what other language(s) do you use in daily life? How old were you when you began acquiring the language(s)?
Language A: _________; _____ years old; Language B: _________; _____ years old
Language C: _________; _____ years old

Q3. Below are descriptive statements which represent a wide range of abilities in speaking, listening, reading and writing. Place a √ in the box of the level that you think best represents your ability in each of the languages you wrote down in Q2.

(a) Speaking:

Level	**Descriptive statements**	**Language (s)**		
		A	**B**	**C**
1	I can speak only simple words and phrases (*e.g. "Hello", "Thank you").*			
2	I can participate in conversations on everyday topics (*e.g. when ordering food in a restaurant),* without the help of any complementary communication *(e.g. gestures, writing).*			
3	I can explain my point of view freely in some formal conversations (*e.g. academic discussion).*			
4	I can speak fluently in most conversations on practical, social and professional topics, but I may need to pause quite a lot when expressing more complex idea.			
5	I can speak the language as well as an educated native.			

(b) Listening:

Level	**Descriptive statements**	**Language (s)**		
		A	**B**	**C**
1	I know only common words and phrases, (*e.g. "Hello", "Thank you").*			
2	I can understand essential points in a *face-to-face* conversation on everyday topics (e.g. *when buying food)*; but I have *some difficulties* in understanding messages heard through *radio or telephone.*			
3	I can understand and infer other's point of view adequately in some formal conversations (*e.g. academic discussion).* I can understand messages heard through *radio and telephone with minor difficulties.*			
4	I can understand what is said in most conversations on practical, social and professional topics in *both face-to-face and telephone conversations*; but I cannot understand idioms and humors.			
5	I can understand the language as well as an educated native.			

You may omit the following parts for "reading" and "writing" if a writing system is not applicable to the language.

(c) Reading:

Level	Descriptive statements	Language (s)		
		A	B	C
1	I can only understand simple written texts or directions for familiar things (*e.g. "Hello", "Thank you")*.			
2	I can understand the meaning of simple texts about familiar subjects (*e.g. when reading menu in a restaurant).*			
3	I can understand the main content of some newspaper articles but not all of the details (e.g. *news about a car accident*).			
4	I can understand most things written in the language, but I need to look for dictionary for less common words.			
5	I can read and understand the language as well as an educated native.			

(d) Writing:

Level	Descriptive statements	Language (s)		
		A	B	C
1	I can only write simple words and phrases e.g. *"(your name)"*			
2	I can write simple texts connected with my own life and my needs (*e.g. writing a shopping checklist)*			
3	I can write a simple article about familiar subjects (*e.g. description of my favorite leisure time activities)*			
4	I can write about most topics in the language, but I need to look for dictionary for less common words			
5	I can write the language as well as an educated native			

D. Lifestyle

Q1a. In general, how much time do you spend on playing ball games(s)? _____ hours per week. *(If you do not play any, please skip and move to Q2)*

b. What kind of ball game(s) do you usually play?

1. ________________ 2.__________________ 3._________________

Q2. In general, how often do you use computer? _____ hours per week

O3a.In general, how often do you play video games? _____ hours per week (*If you do not play any, please skip and move to Q4)*

b: What kind of video games do you usually play? *(You can tick more than one box)*

- ☐ Adventure
- ☐ Shooting
- ☐ RPG
- ☐ Car Racing
- ☐ Simulation
- ☐ Sports
- ☐ Real time strategy game
- ☐ Sports
- ☐ Turn-based strategy game
- ☐ Action
- ☐ Others

Q4a. Do you play any musical instrument(s)? If yes, what do you play?

1. ________________ 2.__________________ 3._________________

b. How old were you when you began learning the instrument(s)?

1. _____ years old 2. _____ years old 3. _____ years old

c. In general, how much time do you spend on musical practice/performance? ___ hours per week

d. Did you take any Graded exam (e.g. ABRSM) on instruments and obtain a grade?

If yes, please specify: ________________________________

e. What is the context of performing your instruments? (e.g. leisure, orchestra, recital) __________

Thank you very much for your kind participation in this research.

Appendix C: Stimuli and procedures of the Flanker task

Following Fan et al. (2002), the Flanker task was presented using the E-Prime 2.1 application software package (Psychology Software Tools, 2001) on a 15-inch CRT monitor with a refresh rate of 60 Hz. Participants sat at a distance of 53 cm from the computer screen. Stimuli consisted of a row of five visually presented horizontal black lines, with arrowheads pointing to the left or to the right. Participants were asked to hold the mouse with their both thumbs on the left and right buttons of the mouse, respectively. They were instructed to focus on the central fixation, and to respond, by pressing the button of the mouse, as quickly and accurately as possible with their left thumb when the arrow pointed to the left and with the right thumb when it pointed to the right.

The task was composed of 162 trials in four blocks. The first block of 18 trials was a practice block. The subsequent three blocks consisted of 48 trials each, for a total of 144 trials. Each trial began with a central fixation. The target consisted of a row of 5 arrows that may point left or right; and the row of arrow appeared just above or below the central fixation. Cues in the form of asterisks appeared at random before the target in 75% of the trials either at the central fixation (central), or in location where the targeted row of arrow would appear next (up or down). The target arrows were flanked by two arrows in close spatial proximity on each side. The flanking arrows point in the same (congruent) direction as the target for half of the trials or the opposite (incongruent). There were 24 test trials for each Cue (central, no, spatial) x Flanker (congruent, incongruent) combination. Within each block the 8 Cue x Flanker conditions were presented in random order. This prevented participants from developing any automatic response to a particular Cue x Flanker combination occurring in sequence.

Appendix D: Auditory and visual stream segregation: Stimuli and procedures

Auditory Stream segregation: stimuli

Following Helenius et al. (1999), the auditory sequences were composed of high (1000 Hz) and low (400 Hz) pitch pure tones presented alternatively through headphones to both ears simultaneously. Sounds were digitally edited to a 16-bit resolution at a sampling rate of 44 kHz using Sound Forge 8.0 (Sony Creative Software, Inc). All tones lasted 40 ms (including 5 ms linear onset/offset amplitude ramps in order to prevent onset and offset clicks). Stimuli were presented using E-prime (Psychology Software Tools, Inc) software on a PC computer running the rapid serial auditory sequences binaurally through headphone (Sennheiser, HD280) at approximately 65 dB sound pressure level through the USB audio interface (M-Audio, Fast Track Pro.).

Visual Stream segregation: stimuli

The visual sequences were composed of black dots subtending 1° x 1° of visual angle, displayed on a white background screen. The dots were displayed alternatively 2° above and below the fixation cross along the vertical median line of the screen. The participants were asked to fixate the central cross at all times. The dots were thus foveally presented, and could be perceived accurately without eye movements. Stimuli were presented using E-prime software on a PC computer with a refresh rate of 60 Hz.

Auditory and visual stream segregation: procedures

The participants were tested individually in a moderately lit sound-proofed booth, and sat 60 cm away from the computer. A fixation cross, subtending 0.5° x 0.5° of visual angel appeared at the centre of the screen. After 500 ms, the auditory or visual sequence was displayed. The participant was asked to keep fixating the crosshair which remained on the screen all along the sequence display. At the end of the sequence, participants had to report whether they had perceived “one stream” or “two streams’ according to a forced choice paradigm. A brief training phase with unambiguous one- (400 ms) and two-stream stimuli (50 ms) were run to illustrate the choices.

In the testing phase, participants were requested to report the percept of one or two streams in 30 trial sequences. In order to determine the optimal segregation threshold for each participant, a simple “one-up, one-down” adaptive method was used to estimate the 50% chance in a two forced choice paradigm (Levitt, 1971). As long as the answer was “one stream”, the program would decrease the SOA automatically. In contrast, when the answer was “two streams”, the SOA increased automatically. The SOA of the first sequence was 300 ms for both modalities to unambiguously yield the same initial perception context of “one stream” for all participants. In the first trials, the SOA was decreased by steps of 40 ms. After the first change in response type, the step was set to 20 ms, then to 10 ms for the second change and finally to 5 ms for the third change.

Appendix E: Results of the Kolmogorov-Smirnov Test for background and lifestyle factors for each language group

	Cantonese speakers		English speakers	
	D(28)	*p*	*D*(28)	*p*
Age (years)	0.28	< .001*	0.24	< .001*
Ravens	0.20	.002*	0.21	.007*
Time spent in (hours per week)				
Musical instruments	0.38	< .001*	0.26	< .001*
Ball games	0.27	< .001*	0.26	< .001*
Use of computer	0.21	.003*	0.19	.011*
Videogames	0.39	< .001*	0.31	< .001*

Note. Ravens = raw score of the Ravens Standard Progressive Matrices

* $p < .05$