Title	Cross-linguistic and cross-scriptal differences in auditory and visual attentional shifts : a comparison between native Mandarin and English speakers
Author(s)	Lee, Tsz-chung; 李子聰
Citation	Lee, T. [李子聰]. (2012). Cross-linguistic and cross-scriptal differences in auditory and visual attentional shifts : a comparison between native Mandarin and English speakers. (Thesis). University of Hong Kong, Pokfulam, Hong Kong SAR.
Issued Date	2012
URL	http://hdl.handle.net/10722/237877
Rights	This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.; The author retains all proprietary rights, (such as patent rights) and the right to use in future works.

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

Lee Tsz Chung, Cedric
The University of Hong Kong

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.

Abstract

Lallier (2010) proposed that our attentional shifting speed could be shaped by our native language. In our current study we tested this hypothesis by comparing the attentional shift of native English and native Mandarin speakers using the stream segregation paradigm. English and Mandarin are known to be of two contrastive language systems. The rhythmic and scriptal differences between Mandarin and English are discussed. Despite the differences, results revealed no significant difference between the two groups. We proposed that language difference might not have a direct effect on non-language tasks. Some ambiguities in verbal and written domains of the two languages were also discussed.

Language use in humans involves both the visual and auditory attention systems. Active reading and listening require the use of these two modes of attention, respectively. Apart from the lexical level of decoding words and sentences, the low-level auditory processing of non-linguistic signals is also vital to comprehension of written and verbal language. For instance, in recent decades, research has been continuously showing that dyslexia could be related to a temporal processing disorder – the failure or incapability of processing rapid stimulus sequences (e.g. Helenius, Uutela, & Hari, 1999 ; Hari & Renvall, 2001; Lallier et al., 2009; Facoetti, Lorusso, Cattaneo, Galli, Molteni, 2005 ; Stein & Walsh, 1997). The process of reading a text or listening to a stream of speech, at the sensorial perceptual level, can be regarded as a task embedded with multiple, temporally arranged rapid sensory inputs. And focused (or selective) attention is required to orient our cognitive system to receive the fast and ever-changing stimuli.

Different theories and models were put forward to explain the mechanism of attention in humans. Some of them were based on auditory experiments but a majority of them were developed from results in visual attention studies (Lachter, Forster, & Ruthruff, 2004). Classic example includes Broadbent's (1958) selective filter theory of attention, one of the most influential models that deals with focused auditory attention, proposed that there is a filter that rejects unattended messages at the early stage of processing; on the contrary, other theories such as that of Treisman (1960) or Deutsch and Deutsch's (1963) suggested that analysis of stimuli may occur before any selection of information, which means that even the "unattended stimuli" could have been partially (Treisman) or completely analyzed (Deutsch and Deutsch).

However, the models mentioned above did not deal with attentional shift, which is believed to be an integral part of receiving the rapid and ever-changing stimuli during language perception. Therefore we need to consider an attention model that can address the issue of

attention shift. Posner (1980) developed a three-stage model based on a series of visual field studies. These include 1) orienting to stimulus, 2) detecting it and 3) sustaining alertness (Posner, 1980; Posner & Peterson, 1990). It should be noted that Posner's theory was not used to interpret perception of language and it was based on visual experiment, but we believe that it adequately explain the shift of attention involved in language processing in both modalities.

A spotlight analogy was often used to describe the process of swift disengagement of old stimulus, orienting, detecting and locking onto the new stimulus (Posner, 1980; Posner & Peterson, 1990), though it was also well documented that attention can shift in the absence of any eye-movement - covert attention (Posner, 1980). It is widely believed that selective visual attention consists of two independent but functionally interrelated systems – the endogenous and the exogenous attention systems (Posner, 1980; Theeuwes, 1991; Marius, Dirk, & Jan, 2004). The endogenous (Top-down, goal-directed) system is responsible for directing attention to anticipated stimuli of interest while the exogenous (Bottom-up, stimuli-driven) cater for the unexpected stimuli. This exogenous system is also thought to account for the automatic attention in human's cognitive system (Lambert, Spencer, & Mohindra, 1987; Yantis & Jonides, 1990). We believe that during speech perception and rapid language processing, our exogenous system is engaged since our attentional system will automatically pick up all the incoming signals once our endogenous system has directed our voluntary attention to it. The stream segregation task used in the current study is thought to be engaging the participants' exogenous or automatic attention (Hari & Renvall, 2001)

How can stream segregation task measure attention shift?

The processing speed of attentional shift is believed to be reflected in a auditory or visual stream segregation task, in which the perception of tone streams (auditory) or dot streams

(visual) changes with stimulus presentation rate (Helenius, Uutela, & Hari, 1999 ; Hari & Renvall, 2001). In the auditory stream segregation task, an auditory stream is formed when two tones – one of higher pitch and one of lower pitch – are played alternatively. When the stimulus onset asynchrony (SOA) (i.e. the interval between each two stimuli, in this case, high and low tones) is long enough or the frequency (Hz) difference between two tones is small, the listener will tend to perceive one single alternating stream (Helenius et al., 1999), resembling a "trill" in musical instruments playing. Conversely, if the SOA is short or the frequency (Hz) separation of the two tones is large, stream segregation will occur – the listener will tend to perceive two segregated streams playing simultaneously.

Similarly, for a visual stream segregation task in which two dots – one of higher position and one of lower position with reference to a central cross – are displayed alternatively, the effect was found to be analogous to that in the auditory stream segregation task (Bregman & Achim, 1973). (See Appendix A for an illustration)

Hari and Renvall (2001) proposed that the two possible perceptions reflect the speed at which automatic attention can disengage from one stimulus and engage in the next rapidly presented one. When the automatic attention resources are orienting to capture every single stimulus of the sequence, one stream is perceived. However, segregation of the stream will occur when automatic attention is not fast enough to shift from one stimulus to another subsequent rapidly presented stimulus (Helenius, et al., 1999 ; Hari & Renvall, 2001; Lallier et al., 2009)

Attention shift in written and spoken language processing

Research in recent decades has suggested that dyslexia may be related to a temporal processing disorder, which means that dyslexic individuals may have difficulties processing

rapid stimulus sequences. While auditory stream segregation threshold (i.e. the shortest SOA at which listeners still can perceive one single stream) was generally agreed to be significantly higher in dyslexic individuals (Helenius, et al., 1999 ; Hari & Renvall, 2001; Lallier et al., 2009; Facoetti et al., 2005) -- hence indicating a slower attentional shift; controversy still remains on the issue of whether slow attentional shift is amodal, i.e. affecting both the visual and auditory modalities. (Lallier, 2009; Lailler 2010; King et al., 2008; Heim, Eulitz & Elbert, 2001; Hari & Renvall, 2001)

On the basis of accumulating amount of research on attentional shifting ability of dyslexic individuals, Lallier (2010) further documented that difference in stream threshold among groups of individuals who speak different languages can also be attributed to relevant characteristics of one's mother tongue. Lallier (2010) reported that Welsh-English bilingual speakers, when compared with English monolingual speakers, show slower attentional shift in the auditory modality. She argued that it was because the stress pattern, which is one of the cues used in segmenting connected speech (Culter & Norris, 1988; Jusczyk, 1999), is principally different in Welsh and in English. Although both languages are considered to have a "strong-weak" stress pattern (e.g. "Ap- ple"), the second part of a Welsh word is usually more salient than its initial part, due to the lengthening of the consonant and vowel of that second syllable (Vihman et. al., 2007). The reverse is true for English, which usually has its first part of the word being the most salient. Lallier (2010) thus claimed that attention of Welsh speakers will be delayed as their attention tends to pull towards to end of word. As a result, Welsh speakers generally have slower attentional shift than English speakers.

Our present study aims at testing the hypothesis of Lallier (2010) by comparing the attentional shifting speed of native English and native Mandarin speakers in both the auditory

and visual modalities.

Prosodic differences between English and Mandarin.

English and Mandarin have very different prosodic patterns. There is stress in virtually every English word. Although a variety of stress patterns exist in English words, the predominant pattern has stress in the initial syllable (e.g. “Apple”, “Baby”), following a strong-weak pattern. When only those high-frequency words occur in daily conversation are taken into consideration, over 85% of lexical words were found to begin with strong syllables (Cutler & Carter, 1987). This implies that for listeners of English, their attentional resources must be allocated to the first part of most of the words. The ability to identify the “Strong-weak” pattern as a cue to mark word boundaries is believed to have stemmed from infancy (Cutler & Norris, 1988; Jusczyk, Cutler & Redanz, 1993).

The prosodic features of Mandarin Chinese, in contrast, are more complicated and different approaches have been used to study the rhythm and stress of Chinese (曹剑芬, 2003). Our hypothesis is that, since Chinese is generally considered a syllable-timed language (i.e. the duration between each syllable is equal), listeners to Chinese must shift their attention more quickly compared with listeners of English, which is a stress-timed language.

Scriptal difference between Mandarin (Chinese) and English

As for the visual modality, Lallier (2010) found that there is no significant difference between the visual stream segregation threshold of Native English speakers and that of English-Welsh bilinguals, possibly due to the fact that Welsh and English written form do not vary a lot in relevant aspects. In spite of that, difference in attention shifting speed in the visual respect may arise from the very different nature of Chinese and English. Chinese differs from English in a way that the Chinese writing system is logographic while English writing system is

alphabetic. Hoosain and Osgood (1983) concluded that there is clear evidence showing processing of some aspects of meaning of Chinese words is faster than that of English words in terms of reading. They suggested that since Chinese syllables represent morphemes (characters), the processing of meanings of morphemic symbols could be more direct and thus faster, while English, on the other hand, comprises alphabetic symbols that do not contain meaning and would thus require longer time for semantic comprehension and processing. Another study by Sun, Morita, and Stark (1985) had tried to compare the eye-movements and reading rates whilst reading Chinese and English. The reading rates of Chinese (390 words/min) and English words (380 words/min) were found to be similar. However, we would like to take a more straight-forward stance by taking into account the number of Chinese characters in each word. That is, attentional shift may occur between all or some characters. In this sense, the frequency of attention shift could actually be higher when reading Chinese (reading rate by characters = 580 /min).

Due to the aforementioned differences, we hypothesize that (1) native Mandarin speakers will possess faster auditory attention shift ability as compared to native English speaker due to the differences in the prosodic patterns of their respective mother tongue while (2) in visual modality, the different processing mechanisms of two writing systems and reading rates will lead to a faster visual shifting ability in native Mandarin learners. It was hoped that our study could provide more insight towards language influence on general cognitive ability.

Method

Participants

Twenty-eight native speakers of English and 28 native speakers of Mandarin were recruited randomly in the University of Hong Kong, with equal number of males and females in each group. Participation was on a voluntary basis. The age range was set at 18 - 33 to eliminate any possible effects of maturity or aging on cognitive function. The mean chronological age of Mandarin group was 24.54 (SD= 4.06) while that of the English group is 21.71 (SD=2.54). All participants were having undergraduate study or above. All of them have normal or corrected-to-normal vision and passed hearing screening test. In addition, all of them reported to have no history of neurological or psychiatric disorders, nor any learning impairment for reading and spelling, which is known to be associated with lower attentional shifting speed (Hari & Renvall, 2001; Lallier et al., 2009.)

All participants completed a questionnaire regarding general health condition, language background, expertise in music, sports or computer game experience, prior to the experiment. The reasons for collecting these information is that video game play (Bialystok, 2006), sports training (Nakamoto & Mori, 2008) and musical expertise (Bialystok & DePape, 2009) may lead to advantages on global reaction time, although effects on attentional shift remain uncertain. In addition, Beauvois and Meddis (1997) discovered that musicians possess a greater persistence of auditory stream biasing, which means musicians are more likely to continue to hear segregated streams even after the SOA is increased. Beauvois and Meddis (1997) proposed that this may be caused by the superior auditory-grouping abilities gained through musical experience. In short, we want to control for all these possible confounding variables.

Apparatus and Measures

All participants completed four tasks – (1) Standard Raven Progressive Matrices, (2) Visual stream segregation task, (3) Auditory stream segregation task and (4) Attention Network Test (ANT), which is a variation of the Flanker test originally designed by Eriksen and Ericksen (1974) to test an individual's reaction time, orienting, alerting and inhibitory control (Fan, McCandliss, Sommer, Raz, Posner, 2001). The test was administered to eliminate the possible advantage of bilingualism. One major limitation of our study is that all Mandarin-speaking participants recruited in the University of Hong Kong will necessarily possess a certain degree of English proficiency due to the requirement of the curriculum. Bilingualism was reported to give advantages to global reaction time and possibly interference effect (see a review by Hilchey & Klein, 2011), which could in turn affect the attention shift ability. Given this finding, the auditory and visual stream segregation threshold could be related to the attentional parameters measured in the ANT and we would like to control for this. On the other hand, the Raven's Standard Progressive Matrices is used to match the non-verbal IQ of the participants in two groups. Each participant completed the four tasks in one of the four possible orders rotated across participants. The four orders were arranged by Latin square so that each task is only preceded by another with equal chance.

The visual stream segregation and the Flanker test were administered on a PC with 15-inch Philip CRT monitor at a refresh rate of 60 Hz. Participants sat at 60 cm from the monitor in the visual stream segregation task and 53 cm in the Flanker test as specified by the respective tests. The fixation point was a cross located at the center of the screen. The auditory stream segregation task was conducted on a Lenovo laptop, which was connected to an external audio

interface device (M-Audio, Fast Track Pro.). The output level of the device was set at 60dB and stimuli were delivered via headphones (Sennheiser, HD280). The standard progressive matrices was a pencil-and-paper test. Administration of the test was done according to the test manual (Raven, Court & Raven, 1998). Participants were given time to complete 60 multiple choice questions in total, all of which required choosing one out of eight answers that can most suitably complete the pattern given in the questions.

Procedures

In the auditory stream segregation task, participant were instructed to give response after hearing high tones and low tones played alternatively for about 5 seconds through the headphone. He/she had to tell whether a connected or segregated stream was perceived by pressing the corresponding buttons: 1 for hearing the two tones playing alternatively (connected) and 2 for hearing the tones playing simultaneously (segregated). The first trial started with a SOA of 300ms. If the response was a perception of one single stream, the SOA would be reduced by 40ms each time. This process would be repeated until the participant gave a response of perceiving two streams, after which the SOA would be increased by 20ms. After that, the SOA would be increased or decreased by 10 ms depending on the participant's response, then all the subsequent change would be in 5ms intervals. Practice trials were given to the participant to give them an idea of what a single stream and segregated streams should be like. For each participant, 30 trials were done in total.

The visual stream segregation task followed a similar procedure. The only difference is that participants would not hear the two tones, but would see two dots displayed alternatively on the screen, one of higher position and one of lower position. Participants had to indicate whether

he/she had seen the dots bouncing up and down or two separate streams.

Apart from the stream segregation task, the participants had to finish a variation of the flanker task, originally designed by Eriksen and Ericksen (1974). At the center of the computer screen there is a cross on which the participants were told to keep their focus. Participants would then see in each trial five arrows on the screen - lining horizontally- above or below the arrows. The five arrows except the central one always points to the same direction, either to the left or right. Meanwhile, the central one might point in the same direction as the others (congruent) or the opposite direction (incongruent). Participants had to decide which direction the central arrow is pointing. There are four possible cueing conditions before the arrows appear: (1) no cue (2) center cue, (3) spatial cue – up and (4) spatial cue – down. A total of 144 trials were done for each participant.

Global reaction time (RT) was calculated by taking the average of all the congruent and incongruent trials. All outliers with z-scores greater than 3 were excluded. Scores for orienting, alerting and conflict were calculated from the results. The alerting score for each participant was calculated by subtracting the mean RT of the center-cue condition from the no-cue condition. It represents the effect of giving prior warning to an imminent stimulus. The orienting score was calculated by subtracting the mean RT of the spatial-cue conditions (both up & down) from the mean RT of center-cue condition. It was believed that both the center cue and spatial cues produce an alerting effect but only the spatial cues offer predictive spatial information, thus orienting the attention towards the upcoming site of stimulus. Finally, the conflict score were calculating by subtracting the mean RT of all congruent trials, regardless of the cueing conditions, from the mean RT of the incongruent trials.

Results

We were interested to find out whether there were differences in the stream segregation threshold or on any of the variables across the two language groups and if there were, how they might affect the visual or auditory stream segregation threshold. Table 1 shows the respective means for each language group for all measures, including (a) age, (b) raw scores on the Raven's standard progressive matrices, (c) numbers of hours per week on ball games, computer use, video games, music play (d) mean RTs on the flanker test for the alerting, orienting and conflict conditions, and (e) visual and auditory stream segregation thresholds. Figure 1 and 2 illustrates the mean stream segregation thresholds of the two groups across the thirty trials.

Table 1. *Mean scores and standard deviations across language group on measure of Age, Nonverbal IQ, Lifestyle factors, Flanker mean RT and Visual & Auditory stream segregation thereshold.*

	English			Mandarin		
Measures	*n*	*M*	*SD*	*n*	*M*	*SD*
Age	28	21.71	2.88	28	24.54	4.06
Raven Score	28	56.82	2.53	28	57.36	2.71
Ball games (hrs/wk)	28	2.48	3.56	28	1.09	1.50
Comp. use (hrs/wk)	28	30.36	11.26	28	41.46	18.86
Video games (hrs/wk)	28	1.82	3.49	28	1.05	2.16
Music play (hrs/wk)	28	1.32	2.47	28	0.82	2.18
Mean reaction time (ms)	28	495.36	35.94	28	514.21	41.80
Alerting score	28	17.02	21.14	28	15.99	14.67
Orienting score	28	50.64	24.31	28	53.29	23.44
Conflict score	28	82.49	23.67	28	73.3	14.11
Visual stream segregation threshold (ms)	28	130.68	45.951	28	137.02	41.93
Auditory stream segregation threshold (ms)	28	75.66	43.06	28	80.16	41.48

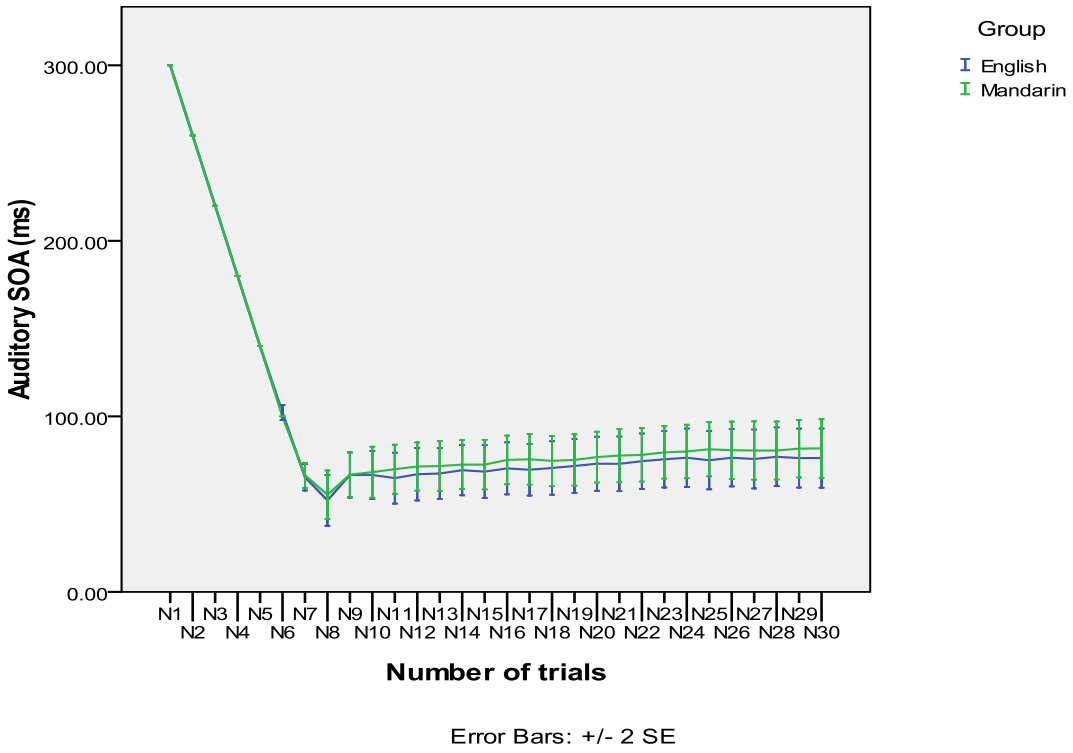

Figure 1. Mean Auditory SOA for each language group across 30 trials. Auditory stream Segregation threshold was taken by averaging the last 10 trials.

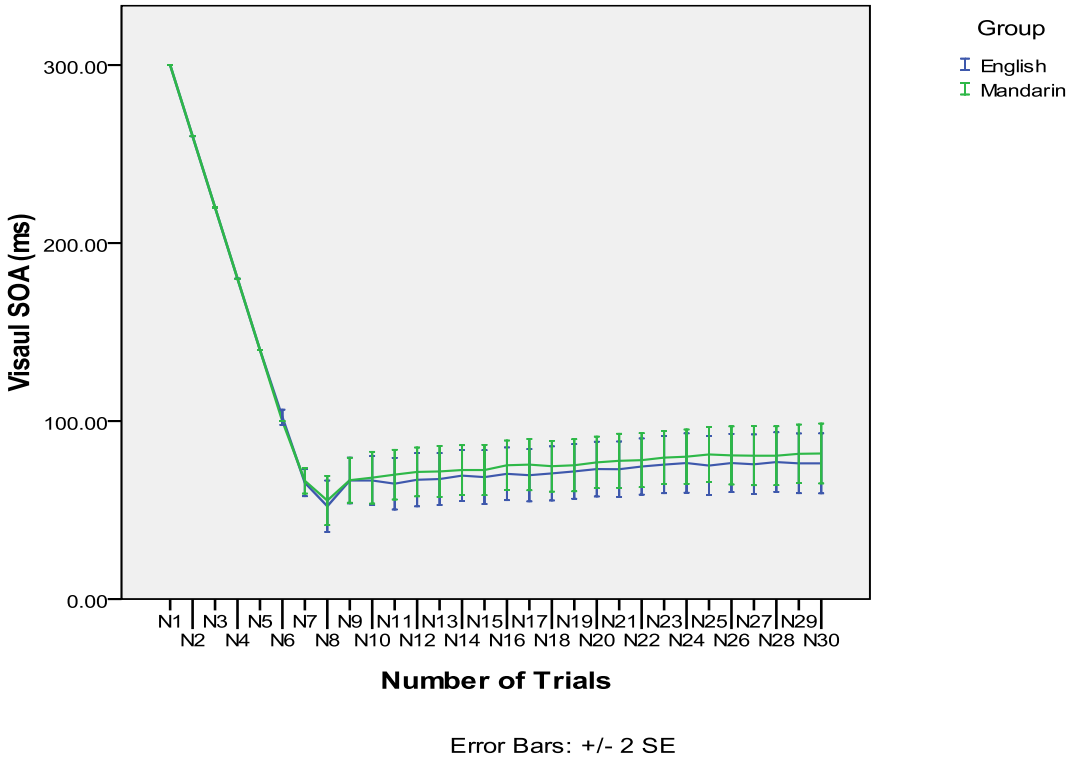

Figure 2. Mean Visual SOA for each language group across 30 trials. Visual stream Segregation threshold was taken by averaging the last 10 trials.

Note. Stream Segregation threshold was taken by averaging the last 10 trials.

Table 2 shows the results of the Kolmogorov-Smirnov normality test of all measures. It revealed non-normal distribution ($p<.05$) for both language groups in data of (a) age, (b) non-verbal IQ (as measured by Raven Standard Progressive Matrices), and (c) number of hours per week on ball games, computer use, video games, and music play. In addition, mean reaction time of the Mandarin group and visual stream segregation threshold of the English group also violated the assumptions of normal distribution.

Table 2.
Normality Test for various variables of the two language groups

	English			Mandarin		
	D-statistic	Df	*p*	D-statistic	Df	*p*
Age	0.28	28	<.001	0.20	28	<.01
Non-verbal IQ	0.20	28	<.01	0.20	28	<.05
Ball games (hrs/wk)	0.26	28	<.001	0.26	28	<.001
Comp. use (hrs/wk)	0.19	28	<.05	0.19	28	<.01
Video games (hrs/wk)	0.31	28	<.001	0.300	28	<.001
Music play (hrs/wk)	0.38	28	<.001	0.38	28	<.001
Mean reaction time	0.12	28	n.s.[b]	0.22	28	<.005
Alerting score	0.93	28	n.s.	0.12	28	n.s.
Orienting score	0.99	28	n.s.	0.16	28	n.s.
Conflict score	0.87	28	n.s.	0.12	28	n.s.
Visual stream segregation threshold	0.23	28	<.05	0.15	28	n.s.
Auditory stream segregation threshold	0.15	28	n.s.	0.11	28	n.s.

Note. A significance level <.05 indicates non-normal distribution.
[b]*n.s.* = not significant, $p > .05$

Mean reaction time (Mean RT) data was transformed using the reciprocal transformation method (i.e. 1/Mean reaction time) which restored the assumptions of normality as shown by the Shapiro-Wilk normality test (English group, $W(28) = .973$, $p = .67$; Mandarin group, $W(28) = .945$, $p = .15$). Independent t-test was then used and found no significant difference between the transformed mean RT of the two language groups (see Table 3). Independent t-test also showed no significant differences among the two language groups in measures of alerting, orienting, and conflict effect.

Table 3.
Independent t-test comparing diff. variables across the two language groups

Measures	*t*	*df*	*p*	*M diff.*	*SE Diff.*	*Effect size (r)*
Mean reaction time [a]	1.849	54	.07	18.84 (ms)	10.42	.18
Alerting score	.212	54	.83	1.03	4.86	.06
Orienting score	-.416	54	.68	2.65	6.38	.09
Conflict score	1.765	54	.08	9.19	5.21	.18
Auditory stream segregation threshold	-.398	54	.69	4.5 (ms)	11.30	0.09

[a] t-test was run after Reciprocal Transformation

For other measures that violated the normality assumption, none of the transformation methods could produce normal distribution within both groups. Hence, the non-parametric Mann-Whitney U test was used for these variables (see Table 4). The test revealed significant differences in Age and Computer use across the two language groups, indicating that the Mandarin speaking group was significantly older and used the computer for longer hours than the English speaking group. Other variables, i.e. hours in ball games, video games, and music play, did not show any significant differences.

Table 4.

Mann-Whitney U test comparing diff. variables across the two language groups

Measures	*Mean Rank*		*U*	*z*	*Exact CI (2-tailed)*
	English	*Mandarin*			
Age	22.14	34.86	214	-2.936	<.005
Raven Score	26.41	30.59	333.5	-.97	.34
Ball games (hrs/wk)	30.73	26.27	329.5	-1.109	.27
Comp. use (hrs/wk)	23.79	33.21	260	-2.174	<.05
Video games (hrs/wk)	30.30	26.70	341.5	-.998	.32
Music play (hrs/wk)	30.41	26.59	338.5	-1.153	.27
Visual stream segregation threshold	27.52	29.48	364.5	-.451	.65

Critically, the results from Table 3 and Table 4 indicated that the visual stream segregation threshold (Mann-Whitney U test) and auditory stream segregation threshold (independent t-test) did not differ between the English and Mandarin speakers. As mentioned, age and the time spent on computer use were significantly different amongst the two language groups. In order to see whether age and computer use could influence the stream segregation threshold, participants were arranged in ascending order according to their age and divided into two age groups. Mann-Whitney U test revealed no difference in both visual stream segregation threshold, *U*= 364.5, p= .658 (two-tailed) and auditory stream segregation threshold, *U*= 388.5, p=.958 (two-tailed). Likewise, a high-computer-use group and a low-computer-use group were created using the same method. Mann-Whitney U test was run again and found no difference across the two groups in terms of visual stream segregation threshold, *U*=378, p= .823 (two-tailed) and auditory stream segregation threshold, U=372.5, p=.754 (two-tailed). Results showed that in our experiment the different ages and diff. amount of computer use should not affect the experiment outcome.

Discussion

The main goal of our study is to investigate whether native language might have influence on individuals' attentional shifting speed. Our hypothesis is that due to the contrastive difference of English and Mandarin in both the auditory and written context, and the apparent need of Mandarin speakers to process faster language stimuli, Mandarin native speakers would have faster attentional shifting speed than English native speakers in both auditory and visual modalities.

Our results did not support the hypothesis and indeed, the two groups had similar attentional shifting speed given that other possible confounding variables are controlled for. If Lallier's (2011) claim was correct, i.e., one's attention shifting speed is linked to his/her native language, then the possible reason for this result is that despite the apparent contrast of the two language systems, the intrinsic mechanism and speed for perceiving rapid language stimuli in both languages are largely identical. For the ease of discussion, we will first look at the suprasegmental features of the two languages, before we proceed to discuss the differences between the Chinese and English scripts. In addition, during the course of discussion, we will also reconsider the rationales for our original predictions by reviewing other relevant literature and propose possible explanations for our null finding.

Rhythmic structure of Mandarin and English.

To make any comparison meaningful, first we have to unravel what the basic rhythmic unit of English and Mandarin are and what we are attending to when we listen to a stream of speech. Lallier (2010) proposed that our attention is oriented towards stresses when we listen to English. Traditionally, Languages were thought to fall into two genres according to their rhythmic

features - “stress-timed” and “syllable-timed”. English was an exemplar of stress-timed language. This theory was first put forward by Pike (1945), and further elaborated by Abercrombie (1967) and others. A major essence in this classification is that *isochrony* exists in both types of language. This refers to the belief that in stress-timed language, duration between each stress (inter-stress intervals) is the same while in syllable-timed language, isochrony refers to the approximate equal duration each syllable lasts(Abercrombie , 1967; Couper-Khlen, 1993). However, over the course of scientific research in recent decades, it has been shown that in fact isochrony does not exist acoustically. For example, Roach (1982) discovered that the variance (SD) of syllable duration in traditional stress-timed and syllable-timed language roughly lie within the same range; on the other hand, contradictory to prediction, “stress-timed languages” actually possess more variance in terms of inter-stress intervals than “syllable-timed languages”. This means that when using instrumental method, the notion of equal inter-stress intervals in so-called “stress-timed” language and equal syllable time in syllable-timed languages” are without its basis. Other researches also had similar findings and therefore they believed that the stress-timed and syllable-timed categories simply do not exist (Dauer, 1983; Dauer, 1987; Roach, 1982; Bertran, 1999).

Nevertheless, some researchers still believed that rhythm class do exist and have tried to use different acoustic measures to establish the rhythm class, instead of focusing on the “isochrony theory” (Grabe & Low, 2002; Ramus, Nespor, & Mehler, 1999; Ramus, Dupoux, & Mehler, 2003). For instance, percentage of vowel time (%V) and consonant time (%C) in sentences, standard deviation of consonantal (Δ C) and vowel duration (Δ V) were the measures used in the study of Ramus et al. (2003). Results of these studies showed that languages are more or less stress-timed or syllable timed, which implies that they should belong to a

continuum in terms of rhythmic characteristics, rather than falling neatly into categories. It is worthwhile to note that although one may instinctively classify Mandarin as syllable-timed language, Mandarin was not included in any preliminary studies until the research by Grabe and Low (2002), in which Singaporean Mandarin was one of the languages studied. In their study, they used the *Pairwise Variability Index* (PVI) to investigate the variability in (1) duration of vowels and (2) duration of intervals between vowels (i.e. consonantal intervals) in each successive pairs. This approach was reinforced by Lin and Wang (2005) who adopted the same methodology to study Chinese Mandarin. Both studies showed that Mandarin is closer to French – which is viewed as a syllable-timed language – in terms of rhythmic structure. Although Lin and Wang (2005) claimed that their study has proven that Mandarin is a syllable-timed language, their data (e.g. PVI, %V, Δ C) actually showed that Mandarin lies somewhere between French and English. This echoed with the results of Grabe and Low (2002), who found that Mandarin is different from both French and English, but resemble more closely to French.

Based on the evidences we have mentioned, we can conclude that in terms of the rhythmic structure, Mandarin and English at least differs to a certain extent. It is reasonable to claim then, if language had any effect on attentional shift due to its rhythmic pattern, native English speakers and native Mandarin speakers should have varied speed of attentional shift in auditory modality. Nonetheless, our result showed otherwise. In fact, Lallier's proposal may seem plausible at first thought, but close examination revealed problems underneath. She claimed that auditory attentional orientation to words would be delayed, since it would be pulled towards the end of Welsh words because of the salient syllable being at the latter part of most Welsh words. However, it is obvious that as long as the inter-stress intervals remain the same, the attentional

shift constraint on both languages will be roughly the same regardless of the position of the salient unit in words, if our attention truly orient to stressed syllables and shift among them. To date there such relevant information on Welsh (i.e. the "average length of inter-stress intervals") is not available. Research on Welsh has been restricted to other respects of prosody (Gibbon & Williams, 2007).

Another point we have to bear in mind is that the notion of shifting attention towards stressed syllables was not well established, particularly in Mandarin. In English it was shown that infants have learnt to locate word boundaries by attending to stresses (Polka & Sundara, 2003). However, for Mandarin the situation is more complicated. One problem is that Mandarin is a tonal language. Wang, Chu, and He (2003) argued that while syllables with the four normal tones can be said to be all stressed from the viewpoint of phonology, the stress degrees of syllables in a polysyllabic word are not equal from the viewpoint of phonetics 'even if they are all "phonologically" stressed' (p.1827). To make things more complex, there is tone Sandhi in play and tones are reduced or accentuated under different contexts. Furthermore, other factors such as speaking rate, context of speaking (e.g. formal vs casual) are all in play to add in more variability (Yuan, Limberman & Cieri, 2006). It is likely that Lallier's suggestion has over-simplified underlying contributing factors of language to attention, if any is present.

Scriptal differences of Chinese and English.

Our present finding also showed no difference in attentional shift among native Mandarin and English speakers in the visual modality. As from what we have discussed for speech processing, the same problem arises here. In order to make the two languages comparable, it is ideal to have the same model applied to both languages. Most of the research of cognitive neuroscience has focused on single lexical retrieval, which may not be applicable to our present

study which require rapid attentional shift. Eye movements during reading were studied extensively by Rayner in the recent decades (See Rayner, 1998; Reicle, Rayner & Pollastek, 2003 for reviews and comparison among different reading models), but most of the studies were based on English scripts. In an attempt to test the model using the two extremely varied written languages, Rayner, Li and Pollastek (2007) extended their E-Z reader model to Chinese reading. It is widely accepted that during reading, basic eye movements involves series of saccades and fixation . The core of the E-Z reader model is that our attention is oriented to one word at a time, and strictly in serial manner. During each fixation, one word is processed and the completion of one lexical access acts as a signal to start a saccade (Reicle, Pollastek & Rayner, 2006). However, since this model was developed on English scripts, and English words boundaries are signalled by spaces, which is radically different from the closely-packed words in Chinese, how do we know Chinese readers do not process by characters, but by words? Bai, Yan, Liversedge, Zang, and Rayner (2008) manipulated spacing between characters to investigate the effects on reading rate. They found that if spaces are inserted between each character or randomly, reading rate will be lower than normal (no space inserted), indicating that reading is interrupted; on the other hand, when spaces are inserted between words, reading rate would not be affected. This revealed that when Chinese readers read, the basic processing unit is words. On the basis of this, Rayner, Li and Pollastek (2007) tested their reading model and concluded that the EZ model is applicable to Mandarin as well, which implied that the underlying mechanism of reading Chinese and English are literally similar. It appears that our original and relatively straight-forward character-based hypothesis was not correct.

Lallier (2010) did not find any difference in speed among visual attentional shift of native English and Welsh speakers, nor did our result showed any differences between Mandarin and

English. One may argue that it might be because the reading mechanism for three types of text is similar, or due to the limitation that our participants are all proficient language users exposed to a lot of English reading materials which virtually make them in no way differ from any native English readers. Indeed, it is questionable whether the attentional shift measured in the stream segregation tasks shares the same principles and uses the same cognitive resources during reading. According to the model of Reicle et al. (2006), reading utilised covert attention from word to word, and this covert attention shift involves eye movements as well. However, though the visual stream segregation task in our experiment is related to covert attentional shift also, it does not involve any eye movements. Indeed, some studies have failed to establish a strong link between reading and visual temporal processing in normal readers (Au & Lovegrove, 2001). In the following discussion, we will further examine Lallier's and our initial hypothesis, also to investigate possible explanation for our null findings.

Language as a unique module vs. non-modular view

It is inevitable for us to lead to the discussion of whether there are modules unique in human for language processing. A possible explanation for our null results in our experiment might be due to the fact that habitual mode of perceiving one's native language might not shape how we attend to and perceive non-speech sounds. Or to simply put, there might be two mechanisms for perceiving verbal and non-verbal stimuli. A modular view contends that language is so special that it cannot be reduced to explain simply in terms of other cognitive processes. On the other hand, the non-modular view holds that language is merely the collective product of the cognitive processes that were involved in all other human activities (B. Robinson-Riegler & G. Robinson-Riegler, 2012). According to a review by Barrett and Kurzban (2006), the intense debate on modularity has lasted for more than a few decades and has not

seemed to end. For example, a very influential modular view of speech perception is the motor theory of speech perception (Liberman & Mattingly, 1985). It contends that speech perception is only made possible by the linkage between speech perception and speech production. It is our inherent articulatory mechanism that assists us to recognize the speech sounds (Liberman & Mattingly, 1985). According to the advocates of this theory, our perception of language cannot be possibly based on direct interpretation of acoustic or physical signals. This is because even for the same phonemes we perceive, their acoustic signals can vary a lot (this phenomenon was termed as "coarticulation"). Moreover, human is thought to be unable to process (not only to attend to) 10-15 phonemes per second (Liberman & Mattingly, 1985). Therefore it was argued that there must be a unique mechanism devoted to language perception solely.

Hence, from the standpoint of modular view, it would not be surprising that null result was found in our study, since we are trying to generalise the result of a non-speech task to the language domain. In fact, cases and studies on auditory agnosia (i.e. the inability to discriminate sound) and pure word deafness (i.e. comprehension of language is severely impaired while the ability of processing non-speech auditory information remains relatively intact) may provide some evidence for separate language and non-speech sounds processing. For instance, Taniwaki, Tagawa, Sato and Lino (2000) reported a case in which a patient, who suffered from bilateral subcortical hemorrhage, was found to have auditory agnosia restricted to environmental sounds only. Meanwhile, Poeppel (2001) reviewed cases of pure word deafness and suggested that speech perception is separable from other aspects of auditory cognition in a modular sense. In addition, there are studies discovering functional dissociation in different auditory domains such as music, speech and environmental sounds (Peretz et al., 1994). In fact, neuroimaging results also revealed that verbal and non-verbal sounds may involve different neural pathways. Belin,

Zatorre, Lafaille, Ahad and Pike (2000) studied the voice-selective areas in human auditory cortex. They used functional MRI and discovered that when participants listened to verbal and non-verbal sounds, neuronal activities were observed in different cortical areas. Dehaene-Lambertz (2009) has even demonstrated this functional specialization or specialized modules of speech and non-speech stimuli could have been present within the auditory cortex as early as 4-month-old.

On the other hand, we need to acknowledge that there are other researchers who oppose to the notion of language-specific domains (Barrett and Kurzban,2006). For example, Holt and Lotto (2008) proposed that the uniqueness of speech signals need not to be denied, but should be placed within a general cognitive-perceptual framework to be studied. Indeed, it is possible that despite the functional specialization of speech and non-speech stimuli, there is a certain degree of interaction between the two. However, the results of our present study did not support this claim since there is not an apparent link between native languages and attentional shifting speed in both the visual and auditory modalities after other factors were controlled.

To conclude, our current study failed to validate Lallier's (2010) hypothesis. Our point of view is that languages may not be the determining factor of the speed of attentional shift. It is also questionable whether performance in non-language tasks can be attributed to language domains. Finally, since the "sluggish attention shift" was claimed to be the underlying deficit of individuals with developmental dyslexia, we hope our research can contribute to the discussion in a cross-language sense. Also, we hope our study can encourage more research aiming to establish the link between languages to other general cognitive abilities of human.

References

Abercrombie, D. (1967). *Elements of General Phonetics.* Edinburgh: University Press.

Au, A., Lovegrove, B., 2001. Temporal processing ability in above average and average readers. *Perception & Psychophysics 63*, 48-55.

Barrett, H. C., & Kurzban, R. (2006). Modularity in cognition: Framing the debate. *Psychological Review, 113*, 628-647.

Bai, X., Yan, G., Liversedge, S. P., Zang, C., & Rayner, K. (2008). *Reading spaced and unspaced Chinese text: Evidence from eye movements.* Journal of Experimental Psychology. *Human Perception and Performance, 34*, 1277-1287.

Beauvois, M. W., & Meddis, R. (1997). *Time decay of auditory stream biasing. Perception & Psychophysics, 59,* (1), 81-86.

Belin, P., Zatorre, R. J., Lafaille, P., Ahad, P., & Pike, B. (2000). Vocie-selective areas in human auditory cortex. Nature, 403, 309-312

Bertran, A. P. (1999) Prosodic Typology: On the Dichotomy between *Stress-Timed* and *Syllable-Timed* Languages. *Language Design, 2,* 103-130.

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

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

Bregman, A. S., & Achim, A. (1973). Visual stream segregation. *Perception and Psychophysics, 13*(3), 451-454.

Broadbent, D.E. (1958). Perception and communication. Oxford: Pergamon.

Couper-Kuhlen, E. (1993). English speech Rhythm. Amsteradam: John Benjamins Publishing Company.

Cutler, A. and Carter, D. M. (1987). The predominance of strong initial syllables in the English vocabulary. *Computer Speech and Language, 2,* 133-142.

Cutler, A., & Norris, D. (1988). The role of strong syllables in segmentation for lexical access. *Journal of Experimental Psychology: Human Perception and Performance, 14*, 113--121.

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

Dauer, R.M. (1987). Phonetic and phonological components of language rhythm. Proceeding, *XIth International Congress of Phonetic Sciences,* 5, 447-450.

Dehaene-Lambertz. (2000). Cerebral Specialization for speech and non-speech stimuli in infants. *Journal of Neuroscience, 12*(3), 449-460

Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception and Psychophysics, 16*, 143-179.

Facoetti, A. Lorusso, M. L., Cattaneo C., Galli, R. & Molteni, M. (2005). Visual and auditory attentional captures are both sluggish in children with developmental dyslexia. *Acta Neurobiologiae Experimentalis 65*, 61-72

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

Grabe, E. & Low, E. L. (2002). Durational variability in speech and the rhythm class hypothesis. In Carlos Gussenhoven & Natasha Warner (eds.) *Laboratory Phonology 7.* New York: Moutonde Gruyter. 515-546.

Hari, R., & Renvall, H. (2001). Impaired processing of rapid stimulus sequences in dyslexia. *Trends of Cognitive Sciences*, 5, 525-532.

Heim, S., FreemanJr., R. B., Eulitz, C., & Elbert, T., (2001). Auditory temporal processing deficit is associated with enhanced sensitivity in the visual modality. *NeuroReport, 12*, 507-510

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 nonlinguistic interference tasks? *Implications for the plasticity of executive control processes. Psychon Bull Rev, 18*, 625-658.

Hoosain, R., & Osgood, C. E. (1983). Processing times for English and Chinese words. *Perception & Psychophysics, 34*(6), 573-577

Holt, L. L., & Lotto, A.J. (2008). Speech perception within an auditory cognitive science framework. *Current Directions in Psychological Science. 17*, 42-46

Jusczyk, P. W., Cutler, A. and Redanz, N. (1993). Infants' Preference for the predominant stress patterns of English words. *Child Development, 64*, 675-687

Jusczyk, P. (1999). How infants begin to extract words from speech. *Trends in Cognitive Science, 3* ,323-331

King, B., Wood, C., & Faulkner, D., (2008). Sensitivity to visual and auditory stimuli in children with developmental dyslexia. *Dyslexia, 14*, 116-141

Lachter, J., Forster, K. I. & Ruthruf, E. (2004). Forty-five Years After Broadbent: Still No Identification Without Attention. Psychological Review 111(4), 880-913.

Lallier, M., Thierry, G., Tainturier, M., 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.

Lallier, M., Carreiras, M., Tainturier, M. J., & Thierry, G. (2010). Does the specific acousticstructure of a language shape auditory attention underlying speech perception? Behavioural and ERP evidence in Welsh-English bilinguals. Symposium conducted at The 2nd Neruobiology of Language Conference, San Diego, California, USA.

Lambert, A., Spencer, E. & Mohindra, N. (1987). Automaticity and theCapture of Attention by a Peripheral Display Change, *Current Psychological Research & Reviews*, 6, 136-147

Lin, H. & Wang, Q. (2005). Mandarin Rhythm: An Acoustic Study. *Journal of Chinese Language and Computing 17 (3):* 127-140

Liberman, A. M. & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition, 21*, 1-36

Marius, V. P., Dirk, J. H., & Jan, T. (2004). Endogenous and exogenous attention shifts are mediated by the same large-scale neural network. *NeuroImage 22*, 822-830.

Nakamoto, H. and S. Mori. 2008. Sport-specific decision-making in a go/no go reaction task: difference among nonathletes and baseball and basketball players. *Perceptual and Motor Skills* 106(1): 163-171.

Peretz, I., Kolisky, R., Tramo, M., Labrecque, R., Hublet, C., Demeurisse, G., & Belleville, S. (1994). Functional dissociations following bilateral lesions of auditory cortex. *Brain, 117*, 1283-1301.

Pike, K. L. (1945). The Intonation of American English. Ann Arbor, Mich: University of Michigan Publications.

Poeppel, D. (2001). Pure word deafness and the bilateral processing of the speech code. *Cognitive Science, 25*, 679-693.

Polka L., & Sundara, M. (2003). Word segmentation in monolingual and bilingual infant learners of English and French. Proceedings of the 15th International Congress of Phonetic Sciences (1021–24)

Posner, M.I. (1980). Orienting of attention. The 7th Sir Frederic Bartlett lecture. *Quarterly Journal of ExperimentalPsychology, 32*A, 3–25.

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

Ramus, F., Dupoux, E. & Mehler, J. (2003). The psychological reality of rhythm classes:Perceptual studies. *Proceedings of the 15th International Congress of Phonetic Sciences*, Barcelona, 337-342.

Ramus, F., Nespor, M. & Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal, *Cognition,* 73, 265-292

Rayner, K. (1998). Eye movements in reading and information processing: Twenty years of research. *Psychological Bulletin, 124*, 374-422.

Rayner, K., Li, X., & Pollatsek, A. (2007). Extending the E-Z Reader Model of Eye Movement Control to Chinese Readers. Cognitive Science 31 (2007) 1021–1033

Reichle, E. D., Rayner, K., & Pollatsek, A. (2003). The E-Z Reader model of eye-movement control in reading: Comparisons to other models. *Behavioral and Brain Sciences, 26*, 445–476.

Reichle, E. D., Pollatsek, A., & Rayner, K. (2006). E-Z Reader: A cognitive-control, serial-attention model of eye-movement behavior during reading. *Cognitive Systems Research, 7*, 4–22.

Robinson-Riegler, B. & Robinson-Riegler, G. (2012). *Cognitive psychology: Applying the science of mind.* Boston: Allyn & Bacon.

Roach, P. (1982). On the distinction between 'stress-timed' and 'syllable-timed' languages. In *Linguistics controversies, Essays in linguistic theory and practice*, D. Crystal (ed.) London: Edward Arnold, 73-79

Saygin, A. P., Dick, F., Wilson, S. W., Dronkers, N. F., & Bates, E. (2003). Neural resources for processing language and environmental sounds, Evidence from aphasia. *Brain, 126*, 928-945.

Stein, J., & Walsh, V., (1997). To see but not to read; the magnocellular theory of dyslexia. *Trends of Neuroscience 20*, 147-152

Sun, F., Morita, M., & Stark, L. W. (1985). Comparative patterns of reading eye movement in Chinese and English. *Perception & Psycholphysics, 37*, 502-506.

Taniwaki, T., Tagawa, K., Sato, F. & Lino, K. (2000). Auditory agnosia restricted to environmental sounds following cortical deafness and generalized auditory agnosia. *Clinical Neurology and Neurosurgery, 102*, 156-162.

Theeuwes, J. (1991). Exogenous and endogenous control of attention - the effect of visual onsets and offsets. *Perception & Psychophysics 49(1)*, 83-90.

Vihman, M.M., Thierry, G., Lum, J., Portnoy, T., & Martin, P., (2007). Onset of word form recognition in Enlish, Welsh, and English-Welsh bilingual infants. *Applied Psycholinguistics, 28*, 475-493.

Wang, Y., Chu. M., & He, L. (2003). *Location of Sentence Stresses within Disyllabic Words in Mandarin.* Beijing Language and Culture University, Beijing.

Williams, B. (2007) .Timing patterns in Welsh . *Proceedings of ICPhS XVI* (1249-1252.)

Yantis, S. & Jonides, J. (1990). Abrupt Visual Onsets and Selective Attention: Voluntary Versus Automatic Allocation. *Journal of Experimental Psychology: Human Perception and Performance 16*(1), 121-134

Yuan, J., Liberman, M. & Cieri, C., (2006), Towards an Integrated Understanding of Speaking Rate in Conversation. in *Proceedings of the Conference on Spoken Language Processing.*

曹剑芬 (2003). *Rhythmic structure of Mandarin Chinese*. 中国社会科学院语言研究所.

Appendix A Illustration of the visual stream segregation paradigm

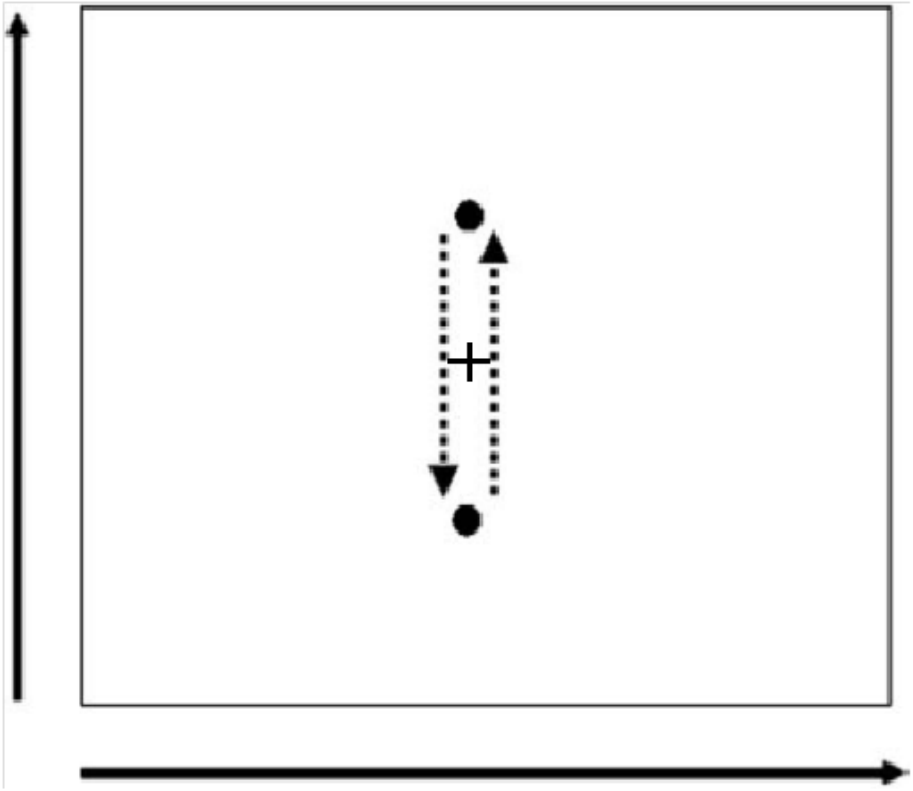

Figure A1 The vertical arrows represent the spatial movement of the dots. The horizontal axis represents the time elapsed.