A Project entitled

The more the better? The modality matters. – Effects of multimodal learning on connected speech in Chinese ESL learners

Submitted by

Wong Sau Ying Isabella

submitted to The Education University of Hong Kong

for the degree of Bachelor of Education (Honours) (English Language)

in *May 2017*

Declaration

I, *Wong Sau Ying Isabella* , declare that this research report represents my own work under the supervision of *Assistant Professor Dr. Wong Wai Lap Simpson*, and that it has not been submitted previously for examination to any tertiary institution.

Signed _____

Student Name

Date

Abstract

Learning with more than a modal benefits perceptual learning in connected speech for English as a Second Language (ESL) learners. Moreover, learning with multimodalities may overload attention and cognitive processing which hinder learning. To examine this hypothesis, a connected speech perceptual dictation is designed and test the effects of learning with bi-modal, listen with either subtitles or shadowing; and multimodal, listen with both subtitles and shadowing. A group of ninety ESL undergraduate and graduate students were recruited from a local university in Hong Kong to participate in the present study. The results challenges the hypothesis that the multiple modalities aid the learning of connected speech. The Wicken's (2007) Multiple Resources Model and the Cognitive Load Theory (CLT) (e.g. Sweller, 2010) may account for the results. The current study sheds light on the pedagogical implication of connected speech learning with multimodalities in ESL classroom context.

Acknowledgement

First of all, I would like to express my gratitude to *Dr. Wong Wai Lap Simpson* who has walked me through the project. From designing the stimuli, operating analysis on SPSS, interpreting results and putting the dissertation together, Dr. Wong has provided me with a lot of insightful ideas to make this paper come into life! Apart from technical and practical research skills, Dr. Wong has inspired me a lot when it comes to personal growth. Here is the most important lesson: Never stop learning because everything in the world is interesting! He always encourages me to give everything a try and explore the meaningfulness. Instead of typing million words of 'Thank you!' here, this is what I want to say 'Dr. Wong, Thank you for always being such a supportive, warm and inspiring supervisor which I cannot thank enough of!'

Also, I am really thankful for the opportunity to carry out my very first research. As an Education major student, I experienced a lot of personal growth which can never be experience in the classroom. For example, sitting in the cold, dark room for 12+ hours a day for just a paragraph of discussion. But most importantly, I have learnt to be my own project manager, data analyst, author, etc. The whole research experience has enriched my life at a whole person level.

Last but not least, I cannot say enough thanks to my amazing nutritionist, *Mom*; my inspiring brother, *Gilbert*; my awesome cheerleaders, the *pure Angel and Mother Teresa*; my dearest girls, *Nicky* and *Jessica*… They stayed by my side and support me physically and emotionally along the journey.

I am now stronger and I believe I can go further!

Introduction

'Why does the English in the classroom sound so different than that of the real world?' This is probably a frequent question asked by a lot of ESL learners. The answer lies the connectedness of native English speech. The sloppy sounds in connected speech imposes challenges on ESL learners, which may lead to miscomprehension, or even communication breakdown. Therefore, there is an urge for ESL learners to acquire connected speech so as to connect to the globalized world.

Thanks to the advanced technology, ESL learners can freely access to the variety of online multimedia resources such as movies and YouTube videos. These multimedia resources then take the learners to the journey of multimodal learning of which engage different channels for perceiving information. However, how well do the learners learn with multimodalities? Specifically, how well do they learn connected speech with subtitles and shadowing? As ESL learners, does the native language influence the processing and learning of connected speech?

Literature review

Challenges of perceiving connected speech in ESL Learners

Connected speech, an everyday phenomenon that occurs in native English speech, is defined as the 'sloppy speech' of which canonical forms that undergo phonological modification (Celce-Murcia et al., 2010). In order to save time and energy, native English speakers tend to economize their articulatory effort which leads to coarticulation (Alameen & Levis, 2015). As a result, listening to connected speech imposes challenge on ESL learners' ability to perceive the

non-salient, connected speech features such as elision, assimilation and juncture. Moreover, the connected speech sounds can in fact, be regarded as non-native contrast due to its non-salience. In some earlier studies on non-native speech perception (e.g. Lado, 1964; Liksker & Abranson, 1970), it has been well documented that the non-native contrasts are of low salience because of native phonological transfer. Particularly, Chinese learners are found less sensitive to English voiced consonants as voicing contrast does not exist in the phonological system in Chinese (Comparison of English and Mandarin (Segmentals), 2014). On top of the ambiguous and non-salient nature of connected speech, limited emphasis has been placed on ESL classroom instruction (Brown & Kondo-Brown, 2006), even though it has been shown that connected speech plays a determinative role in listening comprehension (Kuo, Ting, Chiang & Pierce, 2013), especially for low-proficiency ESL learners (Kato & Tanka, 2015). Along with the lack of instructional resources and teacher training (Brown & Kondo-Brown, 2006; Bradlow & Bent, 2002), the underdeveloped connected speech perceptual ability in ESL learners is inevitable.

Potential benefits of multimodality in language learning

Multimodality learning refers to learning that engages more than one modality. There is abundant convergent evidence in multimodal integration for improving speech perceptual performance due to the interactivity across modalities. According to Shams & Seitz (2008), multimodal learning is more beneficial than unimodal learning when information inputs are congruent as more brain region in both unisensory and multisensory areas are activated. A wealth of neuroimaging studies has also shown the possibility of cross-modality integration (e.g. Ludersdorfer, Wimmer, Richlan, Schurz, Hutzler & Kronbichler, 2016; Scarbel, Beautemps,

Schwartz & Sato, 2014; Millman, Hansen & Cornelissen, 2014). Therefore, it is plausible to infer that the more the modalities are engaged, the better the learning performance. The benefits of multimodality second language learning have been demonstrated in different modality including audiovisual modality (e.g. Hazan, Sennema, Iba & Faulkner, 2005; Li. 2016, Birulés-Muntané & Soto-Faraco, 2016) and audio-motor modality (e.g. Kato & Tanaka, 2015; Linebaugh & Roche, 2015, Nakayama, 2011).

Among the multimodal learning studies, subtitle has been widely adapted as visual input. Subtitle, as defined as the synchronous, intralingual onscreen text along with the audio (Bird & William, 2002; Markham & Peter, 2003), has been generally agreed that it is effective for improving listening comprehension performance (e.g. Hayati & Mohmedi, 2011; Hsu, Hwang, Chang, & Chang, 2013) and vocabulary recognition ability (e.g. Harji, Woods & Alavi, 2010; Yuksel & Tanriverdi, 2009). In other words, subtitles, as a visualized orthographic input, may help learners to process auditory inputs as demonstrated by previous studies (e.g. Petrova, Gaskell & Ferrand, 2011; Roux & Bonin, 2013). Specifically, several studies show that phonological and orthographic information are integrated for processing phonological variants (Bird and William, 2002; Coridun, Ernestus & Bosch, 2015; Ranbom & Connine, 2007). For example, in the study, Bird and William (2002) shows that the textual input along with auditory input was beneficial when new phonological forms are encoded and particularly, it was evidenced that the textual input was used to construct the phonological representation.

Apart from subtitle, shadowing, an active online processing of which learners listen to auditory input and articulate the incoming speech simultaneously (Shiki, Mori, Kadota & Yoshida, 2010), has gained increasing interest in recent years in ESL context (Hamada 2011). It has been widely agreed that different forms of are effective in improving EFL learners' listening

comprehension performance (e.g. Hamada, 2011; Mochizuki, 2006; Tamai, 1997). More recently, studies have shown that the perception of segmental features including phoneme perception (Hamada, 2016) and weak form (i.e. weakened function words) perception (Nakayama, 2011; Nakayama & Iwata, 2012) can be enhanced by shadowing training. Therefore, it is suggested that shadowing may improve learners' speech perception through improving the phonological processing, and hence, the holding capacity of acoustic input in the working memory (Kadota, 2012, Tamai, 1997). Since the acquisition of L2 perception ability can be influenced by the ability to articulate the target language's sound (Kato & Tanaka, 2015), this suggests that there may be a direct link between connected speech perception and shadowing practice for improving the performance of perceiving connected speech.

Possibility of divided attention and cognitive overload

Despite of the above evidence of suggesting the possibilities of multimodality learning on connected speech, the undermining effects such as divided attention and cognitive resources should not be overlooked. As demonstrated in various studies, the additional use of other modalities including visual channel specifically for subtitles processing (Mayer, Lee & Peebles, 2014; Kruger & Steyn, 2014), which causes additional burden to perception. Particularly, Kruger & Steyn (2014) highlighted the divided attention distribution across modalities when processing redundant information from various redundant resources. For these findings, Wickens's Multiple Resource Model (2007) provides a perspective on divided attention in second language learning. Based on the notion that attention is divided when more than one tasks is undergoing at the same time, the Wickens's model predicted the three elements, resource demands, resource similarity

and allocation policy, influence the degree of which task(s) to be 'scarified'. Particularly, the factor, processing modalities, is one of important concerns in the model. It is predicted that dividing attention across modalities for language inputs (visual and auditory, as defined by the model) is challenging for L2 learners.

On top of attention distribution, it is also suspected that the cognitive load as another undermining effect. As demonstrated in the study, Mayer et al. (2014) proposed that the limited cognitive capacity of the non-native English speakers may be one of the factors which lead to the nonsignificant improvement in the comprehension test after the additional use of subtitles. Similar conclusion of which textual inputs impose extra cognitive workload has reached in some previous studies (e.g. Kruger, 2013; Diao, Chandler, & Sweller, 2007; but see Kruger, Hefer & Matthew, 2013) . These evidence posits the possibility for Cognitive load theory (CLT) (e.g. Plass, Moreno & Brünken, 2010; Sweller, van Merriënboer, & Paas, 1998; Sweller, 2010). The theoretical framework is concerned with the limited processing capacity in working memory (WM) and the process of schema construction in long term memory (LTM). Among the three cognitive loads in the theory, namely intrinsic, extraneous, and germane load, the germane load is the major concern for the present study. As according to Sweller (2010), it involves the learners' choice to allocate the use of the resources in WM to work with the interactive elements in the task on hand. Particularly, Paas, van Gog, & Sweller (2010) states the potential problem of learning with multiple sources of information of which learners are overwhelmed by the interactivity between the information sources before the commencement of meaningful, deep learning.

Based on the different perspectives, it is possible to question if the learners in the current study would be benefited by the multimodal learning due to multimodal integration, or overloaded by the multimodal inputs during the connected speech learning process.

Perceptual problems of assimilation in non-native speakers

Among the connected speech processes, assimilation is commonly identified as one of the most challenging type in non-native speaking context (e.g. Abe, 2009; Liang, 2015; Zahedi, Sahragard, & Nasirizadeh, 2007). Assimilation can be further classified as regressive assimilation and progressive assimilation. In the present study, regressive assimilation is chosen as progressive assimilation is relatively uncommon (Cruttenden, 2014). Particularly, Chan & Li (2000) highlighted that as L1 phonological transfer occurs in Cantonese ESL learners, the features in connected speech including assimilation are of low perceptual salience, and thus, often being ignored. As described by Cruttenden (2014), regressive assimilation is 'an influence at word boundaries functions predominantly in an anticipatory direction, features of a sound are anticipated in articulation of preceding sound' (p.308).

Although Gow (2003) demonstrated that place assimilation does not create lexical ambiguity for native English speakers, it is plausible that it would cause perceptual problems in Chinese ESL learners. Therefore, based on Gow's study (2003), English coronal place assimilation of the coronal consonants, word finals with -/t/, -/d/, and -/n/ were chosen for the present study. In place assimilation, segments with coronal place (i.e. /t/, /d/, /n/ in the present study) take place of the subsequent noncoronal segment, such as labial (i.e. /p/, /b/, /m/ in the present study). Thus, right berries, for example, may be perceived as ripe berries.

Based on the view that assimilation is perceive as graded (Gaskell, 2003; Dilley & Pitt, 2007; Martin & Peperkamp, 2011), it is possible that listeners may perceive the assimilated segments as glotalization or elision since there is gradation of acoustic markers in the continuum of assimilation. Specifically, Dilley & Pitt (2007) found that glottal variants were almost always a possible realization of /t/-final words. This suggests the possibility for realizing /t/ and /d/-final words as glottal variants, instead of realizing the assimilated segments as /p/, /b/ or /m/-final word. For example, instead of realizing 'right berries' as 'ripe berries' /raɪp ˈbɛriz/, it is always possible for listeners to hear as its glottal variant, /raɪʔ ˈbɛriz/ or deleted variant, /raɪ ˈbɛriz/.

Apart from the views on different realizations of assimilated segments due to graded assimilation, some studies provide a more specific account for place assimilation in terms of the perceptibility hierarchy of place contrast. Some earlier studies (e.g. Jun, 2004; Mohanan, 1993) shows that oral stops are less likely to assimilate in place when compared with nasals due to the perceptibility difference of place contrasts between nasal consonants and oral stops. In other words, listeners would find it more difficult to perceive the place contrasts in nasals than that of oral stops. For instance, 'green beans' /grin binz/ would therefore, be of lower perceptibility than 'right berries' /raɪt ˈbɛriz/.

Given the different aspects on place assimilation perception, the present study aims at investigating the perceiving and learning process of the chosen types English coronal place assimilation.

Theoretical underpinnings of connected speech learning

To clarify the role of the subtitles and shadowing in connected speech processing and learning, the three connected speech frameworks, the simple abstractionist models, simple exemplar-based models and the hybrid models are proposed for the present study. Although both of the theoretical frameworks are based on the notion that the phonological variants in connected speech can be learnt and retrieved in the mental lexicon, the simple abstractionist models and the simple exemplar-based models hold opposite assumptions while the hybrid model is the combination of the two models (see Ernestus, 2014 for details).

In the simple abstractionist models, the auditory inputs are simultaneously activating and mapping with the corresponding abstract symbols in one's mental lexicon. Before generating the pre-lexical abstract representation as database for later word activation and recognition, all the acoustic details were filtered and the process is known as speaker normalization. In other words, one abstract rule is specific to the metal representation of a single word. Thus, the connected speech learning in the simple abstractionist models would be regarded as the generalization of phonological rules. In contrast, instead of carrying out speaker normalization for generating the word-specific pre-lexical representation, the simple exemplar-based models is composed of the matching between auditory inputs and the exemplars stored the metal lexicon. All the encountered words would be stored in the word cloud with the acoustic details, as well as the contextual cues. Without the single abstract rules, the activation and recognition of a word is weighted on the similarity between the auditory input and the stored exemplars. Hence, the learning of connected speech under the simple exemplar-based models would be considered as experience-driven learning. For the hybrid model, it assumes that both abstract rules generalization and exemplars play different roles in speech perception.

The present study

From the above review, connected speech learning plays a predominant role in listening comprehension in an ESL context. In most of the studies, the benefits of multimodal learning were reported and especially, in the use of subtitles for perceiving native speech with accents (Mitterer & McQueen, 2009) and non-native speech (Birulés-Muntané & Soto-Faraco, 2016), as well as shadowing practice for improving weak form perception (Nakayama, 2011; Nakayama & Iwata, 2012). However, there is no existing study which combined both subtitles and shadowing for examining the effects on connected speech learning.

The possible effects of multimodal learning, multimodal integration and cognitive overload when learning connected speech were tested in the present study. To examine the hypothesis, connected speech audio input, subtitles and shadowing were used in the training phase.

The major research questions are as follows:

1. What are the performances of connected speech perception after learning under the three conditions, I. Subtitles-only; II. Shadowing-only; III. Both subtitles and shadowing?

2. Is there any perceptual patterns in Chinese ESL learners when perceiving the three types of place assimilation, namely the segments with word-final alveolar stop (/t/, /d/, /n/) followed by a labial consonant (/p/, /b/, /m/)?

Firstly, it is predicted that learners who learn connected speech with both subtitles and shadowing would outperform among the two groups, while the subtitle-only group would perform better than that of the shadowing-only group in the connected speech perception test.

The hypothesis was tested by comparing the learners' performance across the three learning conditions: I. Subtitle-only; II. Shadowing-only; III. Both subtitles and shadowing.

Secondly, it is hypothesized that learners may perceive the assimilated segments of /t/ and /d/ as glottalization or elision, and therefore, these segments would be better perceive and learn better that the segment with /n/. The postulation was examined by comparing learners' perceptual performance across the three types of assimilated segment, word finals with -/t/, -/d/ and -/n/.

Method

Participants

In the current study, ninety Chinese-speaking undergraduate and graduate students aged over 18 from The Education University of Hong Kong (EdUHK) were recruited. Through random sampling, all the participants were recruited from through announcements on the university intranets. With the age ranged from 18 to 32, all participants in the sample have taken English Language as compulsory subject since Grade 1 and received education with English as the medium of instruction for at least half a year in The Education University of Hong Kong.

Procedure

After reading the information sheet and signing the written consent form (see Appendix A), participants took part in a 50-minutes, one-sitting experiment in individual format. The experiment was conducted in a quiet room in the library of the EdUHK campus Tai Po. The

experiment was administrated by the author. Upon the completion of the experiment, each participant received HKD$ 20 cash as remuneration.

Audios preparation

The audios were recorded with a high quality recorder (Roland R-09 HR) which is digitalized at a sample rate of 44.1 kHz in a quiet room. The stimuli were spoken aloud by a 22-year-old American female speaker who was born and raised in The United States (Ohio for 18 years; Colorado for 4 years).

Measures

1. Connected Speech Dictation

This test assesses the participants' perceptual abilities to perceive connected speech, specifically, place assimilation. Hence, the assimilated segments were deliberately designed with word-final alveolar stop (/t/, /d/, /n/) followed by a labial consonant (/p/, /b/, /m/) based on Gow's (2003) study. All stimuli items were short sentences composed of five to eight word with basic syntactic structure of Noun (subject) + Verb + (Preposition) + Article + Adjective + Noun (object) so as to minimize the effects of grammatical structures in speech processing. The target segments were embedded in the last two words of the sentence in which the adjectives ended with either -/t/, -/d/ and -/n/ and the object noun starts with either -/p/, -/b/ and -/m/. The sentences were designed without specific context to remove the contextual cues for top-down speech processing. There were a total of twenty sentences with target place assimilation patterns

(see Appendix B and C for details). The test was used as the pre-test and pos-test. The reliability of the whole test was moderate (Cronbach's $\alpha = .79$).

All the stimuli items were presented from a laptop computer through headphones. Participants were required to type the sentences in the computer after hearing each item. Upon the completion of typing a sentence, another item is played by the experimenter without providing any feedback.

2. The Multimodal Training Session

To investigate the effects of different forms of multimodal learning on connected speech perception, participants were randomly assigned to the three learning conditions, I. Subtitle-only; II. Shadowing-only; and III. Both subtitle and shadowing for the training session. During the ten-minute training session, all participants were required to listen to the audio materials while trained either with reading the subtitles for the subtitle-only group or shadowing immediately after the audio is played for the shadowing-only group. The participants who were assigned to the group to learn with both subtitles and shadowing were instructed to read the subtitles while listening to the audios and immediately shadow after the audio is played. All participants were informed to practice the items as many times as possible within the given time.

The 20 items in the connected speech dictation were taken as the training material. All audios were delivered with headphones and all subtitles were presented in black words in a white background in PowerPoint over laptop.

3. Phonological Memory Subtest – Memory of digits and Non-word repetition in The Comprehensive Test of Phonological Processing (CTTOP; Wagner, Torgesen & Rashotte, 1999*)*

To control the effects of phonological memory to connected speech learning, memory of digits and non-word repetition of CTTOP phonological memory subtest were conducted. In the subtest of memory of digits, participants were required to repeat a chunk of number after listening to each of the stimulus item. While in the non-word repetition subtest, participants were instructed to repeat the pseudo-words after listening to each of the stimulus item. All items were delivered through headphones connected with a laptop computer.

4. Test of Articulation Rate (Cheung & Kemper, 1993)

To control the effects of articulation rate to the participants who learnt connected speech with shadowing, the test of articulation rate was conducted. In the PowerPoint, participants were shown with a total of six word pairs with two, four and six syllables respectively (see Appendix D for details). They were required to repeat aloud the word pairs for twenty times as accurately and as rapidly as possible. Participants were informed that repetition would be recorded.

To calculate the articulation rate (wps) for each participant, the total seconds needed for repeating the middle ten repetitions of the six items were divided by 120, the total number of words articulated within the middle ten repetitions (6 word pairs X 10 repetition).

Results

A repeated-measures analysis of variance (ANOVA) has been conducted with connected speech listening performance and the three types of place assimilation (i.e. word final with stop consonants, /t/ and /d/, and nasal consonant, /n/) as within-subject factor, and the three connected speech learning conditions as between-subject factor.

Connected speech perception performances in various learning conditions

In the connected speech training phase, three conditions have been manipulated – I) learning with subtitles only, II) learning with shadowing only, and III) learning with both subtitles and shadowing. The means and standard deviations are shown in Table 1. The main effect of training was significant, $F(1, 86) = 20.85$, $p<.001$, $\eta^2_{\text{partial}} = .17$, showing that the connected speech perceptual performance improved significantly after the training phase. The interaction effect of the listening performance in the three learning conditions was significant, $F(2, 86) = 88.75$, $p<.001$, $\eta^2_{\text{partial}} = .67$, indicating that the performances of connected speech perception were better after learning in the three conditions.

Table 1: Descriptive Statistics for Various Learning Conditions (N=90)

	Pretest	Posttest
Learning conditions	*M (SD)*	*M (SD)*
I. Subtitle-only (N=30)	5.37 (3.54)	15.90 (4.20)
II. Shadowing-only (N=30)	5.93 (2.89)	8.13 (3.26)
III. Both subtitles and shadowing (N=30)	5.57 (3.43)	16.70 (3.47)

To compare the difference between the performance after learning with the three learning conditions, contrast analysis was conducted and the results is shown in Table 2. The results showed that participants' performance in the learning conditions of 'subtitles-only' and 'both subtitles and shadowing' were significantly higher compared to the performance in the learning condition of 'shadowing-only' (ps<.001), showing that learning connected speech without subtitles improved the listening performance minimally. Surprisingly, the listening performance in the learning condition of 'both subtitles and shadowing' was not significant when compared to the condition of 'subtitles-only' (p>.05), indicating that the additional use of shadowing in the training phase did not improve the listening performance significantly.

Table 2: Comparison across the Three Learning Conditions (N=90)

Learning condition comparison	MD	Std. Err.
Subtitle-only V.S. Shadowing-only	3.39*	.73
V.S. Both subtitle and shadowing	-.68	.73
Shadowing-only V.S. Subtitle-only	-3.39	.73
V.S. Both subtitle and shadowing	-4.07*	.73
Both subtitle and shadowing V.S. Subtitle-only	.68	.73
V.S. Shadowing -only	4.07*	.73

**p<.001*

Listening performances across the types of place assimilation

Further analysis was conducted to compare the listening performances across the three types of place assimilation, namely the word final with stop consonants, -/t/ and -/d/, and nasal

consonant, -/n/. The results are indicated in Table 3. A significant main effect was found between the three types of placed assimilation, $F(2, 85) = 7.53$, $p<.001$, $\eta^2_{partial} = .15$, showing that there is significant difference between the respective types of assimilation. The interaction effect of learning and the three types of assimilation was significant, $F(2, 85) = 23.86$, $p<.001$, $\eta^2_{partial} = .36$, indicating that the learning responsiveness of the three types of assimilation are significantly different.

Table 3: Descriptive Statistics across the Three Types of Place Assimilation (N=90)

	Pretest	Posttest
Assimilation type	*M (SD)*	*M (SD)*
I. word final -/t/	2.83 (1.04)	5.18 (1.65)
II. word final -/d/	1.13 (1.22)	4.34 (2.33)
III. word final -/n/	1.66 (1.37)	4.06 (1.69)

Contrast analysis of the three types of place assimilation was then conducted and the results are shown in Table 4. When compared with the voiced word finals, -/d/ and -/n/, the voiceless stop consonant, -/t/ was the most responsive to learning (ps<.001). Meanwhile, the voiced stop consonant, -/d/ and the voiced nasal consonant, -/n/, were nonsignificant in terms of learning responsiveness (ps>.05). This shows that learners are more responsive to learn the word finals with the voiced consonant -/t/ than the voiceless consonants, -/d/ and -/n/.

Table 4: Comparison across the Three Types of Place Assimilation (N=90)

Types of place assimilation	MD	Std. Err.
Word final -/t/ V.S. -/d/	1.27*	.11
V.S. -/n/	1.15*	.11
Word final -/d/ V.S. -/t/	-1.27*	.11
V.S. -/n/	-0.12	.09
Word final -/n/ V.S. -/t/	-1.15*	.11
V.S. -/d/	0.12	.09

**p<.001*

Discussion

The first purpose of the present study was to clarify the role of multimodality in connected speech learning in ESL learners. The findings indicated that the additional use of shadowing did not contribute significantly to connected speech learning. The results disconfirmed the hypotheses that multimodality learning aids connected speech perception among the adult Chinese ESL learners. It was assumed that cross-modality effects is likely to occur in multisensory learning and promotes perceptual learning (Shams & Seitz, 2008), however, the present results challenge this view. The undermining effects of multimodal learning including divided attention and cognitive overload may account for the current results.

Another aim of the study is to look for the pattern of perceptual performance of the three types of assimilation in Chinese ESL learners. The current results shows that word final with /t/ can be perceived and learnt better than -/d/ and -/n/. This challenges the hypothesis of graded

assimilation (Gaskell, 2003; Dilley & Pitt, 2007; Martin & Peperkamp, 2011) of which assumes word finals with /t/ and /d/ can perceived as glottalizaed or deleted variants. The perceptibility hierarchy of place contrast and native language transfer is possible to account for the present result.

Divided attention across modalities

In the present study, audio, subtitles and shadowing entered as spoken words and textual inputs into the visual, auditory and motor sensory memory respectively. In line with the results of the study of Kruger & Steyn (2014), the present results shows that processing the same information from redundant resources affects attention distribution across modalities. Based on the literature review, the view on divided attention in second language learning, the Multiple Resource Model (Wickens, 2007) may be the first possibility of the nonsignificant use of an extra modality.

On the one hand, based on the notion that the different tasks undergoing at the same time interfere each other, when the learners in the multimodal group attempt to engage the three modalities, the auditory channels for the connected speech audio input, the visual channels to read the subtitles while concurrently engaging the vocal channel for shadowing, it is likely that the attention is divided in a greater extent than those in the bi-modal groups who only engage either the visual channels for subtitles or the vocal channel for shadowing while listening to the audio input with the auditory channel, even when the information across modalities are redundant. In some recent studies, selective attention in multimodality redundancy has been documented in the context of infant and early childhood learning (e.g. Bahrick & Lickliter, 2014;

Talsma, Senkowski, Soto-Faraco, & Woldorff, 2010). These may put forward the possibility that similar phenomenon might occur in adult ESL learners on connected speech learning.

On the other hand, based on the idea of attention policy in the model of which determining which information source to be protected and scarified, the current results suggest that the learners in the multimodal group chose to protect the information from the visual and auditory channel while scarify the information from the vocal channel. Interestingly, due to the direct link between attention allocation and individual difference (Hede, 2002; Wickens, McCarley, Alexander, Thomas, Ambinder & Zheng, 2007), it is possible to propose that the learners in the multimodal group are visualizer who respond better to visual learning style, however, learning preference does not necessarily lead to successful learning outcome (Kollöffel, 2012). If this is the case, it can be concluded that the learners deliberately forgo the shadowing information from the vocal channel and allocate more attention to the visual channel for reading the subtitles. Then this view on learning style may account for the nonsignificant use of additional shadowing. Yet, further studies are needed to warrant the significance of learning style in multimodality connected speech learning.

Cognitive overload in multimodal learning

Apart from divided attention during multimodality learning, another possibility is the cognitive overload. Thus cognitive load theory may explain the results in terms of the limited capacity in working memory. The interconnectedness in between the information sources may be the key which lead to the nonsignificant results between the multimodal groups who learnt with

both subtitles and shadowing, and the bimodal group who learnt with either subtitles or shadowing.

As mentioned in the literature review, germane load is the major concern of type of cognitive load in the current study. When handling the multiple resources from the modals, auditory, visual and motor memory are activated for knowledge manipulation. In the present study, the three processing centers in WM, namely the two slave-system in phonological loop - acoustics stores and articulatory loop, as well as the visuo-spatial sketchpad are responsible for processing the auditory connected speech inputs, shadowing and reading subtitles. As assumed in the theory that working memory is limited in processing capacity, it is possible to infer that the multimodal group experienced a more vigorous competition between the three components in the WM when compared with the bimodal groups. Not only did the learners in the multimodal group store the subtitles as textual image temporarily into the visuo-spatial sketchpad, but also further engage the acoustics store and articulatory loop for holding the auditory connected speech input and shadowing respectively. In contrast, the bi-modal groups only engage either the visuo-spatial sketchpad for reading subtitles or the articulatory loop for shadowing. Therefore, the engagement of all of the three centers in the WM may potentially increase the cognitive load which counteracts with the benefits of multimodal learning. Similar to the present findings, some previous studies (e.g. Kruger & Doherty, 2016 ; Kalyuga, 2000) demonstrated that cognitive overload may occurs when receiving identical information from more than one components of WM is involved. Therefore, the limited processing capacity in WM which lead to cognitive overload may explain the nonsignificant connected speech perceptual improvement with the use of an additional modality.

Perceptual performance of the types of place assimilation

In the present study, the segments with word-final alveolar stop (/t/, /d/, /n/) followed by a labial consonant (/p/, /b/, /m/) was chosen. The results indicated that the voiceless, word-final consonant /t/ can be better discriminated than the other two voiced consonant, /d/ and /n/ after learning. Surprisingly, the present results disconfirmed the hypothesis of graded assimilation in perceiving place assimilation. Instead, the perceptibility hierarchy of place contrast is compatible to explain the current result.

The distinctive features of word-final consonants may impose challenges on listeners In terms of the manner of articulation, both /t/ and /d/ are categorized as stops while /n/ is nasal. Based on the early findings on asymmetrical assimilation pattern (e.g. Jun, 2004; Steriade, 2001), it has been observed that the perceptibility of nasal consonant (/n/) is less salient when compared to oral stops (/t/, /d/) because the place contrast in nasal is more likely to assimilate in place. In other words, the asymmetry of different types of assimilation may lead to perceptibility difference. In the similarity judgement experiments and identification experiments, Kawahara & Garvey (2014) confirmed that the place contrasts in nasal consonants in syllable boundaries are of lower perceptibility while voiceless stops is of higher perceptibility than voiced stops and nasals. Hence, rather than the grounded view of perceiving assimilation as other types of phonological variants, the place contrast in place assimilation may cause ambiguity for Chinese ESL learners at a phonetic featural level.

In addition to place contrast, voicing contrast may be another possible account for the present results. As indicated in the results, the perceptual performance of /t/-final word is significantly better after learning when compared to the /d/- and /n/- final word. /t/, the voiceless

consonant, is of higher perceptual salience to Chinese learners as no voiced consonant can be found in the Chinese phonological system. Thus, without the voiced feature, Chinese ESL learners may be therefore, more able to discriminate voiceless consonants than the voiced counterparts. This echoes the previous findings which stress the role of language specific knowledge of native language in processing in both phonetic features (Kuo, Uchikoshi, Kim & Yang, 2016) and phonological modification (e.g. Cho & Lee, 2016; Darcy, Damus, Christophe, Kinzler & Dupoux, 2009). Specifically, Darcy et al. (2009) demonstrated that the French participants used more of the phonological knowledge in native language, voicing rule than that of the non-native language, namely place assimilation when compensating phonological assimilation. In other words, in the current study, Chinese learners may have higher sensitivity to discriminate the voiceless consonant /t/ than that of the voiced ones, /d/ and /n/ due to the native phonological knowledge. However, it is noteworthy that there is a plausibility for acquiring the non-native contrasts within a short period of time (Tamminen, Peltola, Kujala & Näätänenm, 2015), regardless the influence of native language phonological knowledge. This suggests the learnability of the three types of place assimilation, regardless the place contrast of oral stops and nasal or the voicing contrast between /t/-final segment or the voiced /d/ and /n/-final segment.

Theoretical implication for connected speech learning

Among the three postulated phonological underpinnings of connected speech learning, the simple exemplar-based models provides a better account for the use of subtitles and shadowing in the present study. For subtitles, it serves as a tool to visualize the lexical representation of the canonical forms of connected speech along with the auditory input. Both of

the auditory and visual inputs entered learners' mental lexicon and stored as exemplars in the word cloud for later activation and recognition, regardless the divided attention and cognition overload brought by multimodal learning. Yet, it is unknown if learners would further map the sound units in the auditory input and the segments in the textual image correspondingly. While for the shadowing practice, it enables learners to produce the encountered exemplars apart from perceiving them. This may aid the speed of recognition if the multimodal integration does not counteracted by divided attention and cognition overload. Moreover, learning repeatedly during the training session increases the weight of the node of the assimilated segments and thus, further facilitates later discrimination. Therefore, the simple exemplar-based model for connected speech learning is compatible with the present study.

While for the simple abstractionist models and the hybrid models, as there is a lack of evidence to suggest that transfer of learning place assimilation phonological rules with subtitles and shadowing, further researches are needed to warrant the possibility of learning transfer in connected speech.

Limitation & further research

There are two limitations in the present study. The first limitation lies in the design of the stimuli in the audio input. In connected speech perception, there are some confounding factors such as frequency effects of words and phase, which significantly affect learners' recognition. For example, it might be easier for learners' to recognize 'last match' than 'wet bench'. Therefore, the word and phrase frequency effect should be concerned in the further research.

Another limitation concerns the use of subtitles and shadowing in the training phase. For the subtitle presentation, instead of using PowerPoint slides to show the subtitles, it can be shown in a video form so as to control the time of subtitles shown on screen which resemble with those in the movies. For the form of shadowing, learners should be better instructed to carry out complete shadowing of which whole sentence is shadowed, rather than selective shadowing which shadow certain selected words only. As the variations of shadowing may lead to different learning effects (Hamada, 2011), the uniformity of the shadowing practice should also be put into consideration in the further studies.

Based on the existing findings on the effects of multimodal learning, it is suggested that learners' learning style may also put into measure. Therefore, learning style inventories may be included for the further research.

Pedagogical implication

The present results have a direct implication for the use of multimodality connected speech teaching and learning. In the classroom context, learning to perceive connected speech with multiple modalities may be too cognitively demanding for ESL learners, especially when the hearing inputs with various accents and dialects. Therefore, practitioners should be cautious with the sequence of using subtitles and shadowing for connected speech instruction.

As proposed by the previous findings (e.g. Ahmadian & Matour, 2014; Jia & Fu, 2011, Khaghaninezhad & Jafarzadeh, 2013), explicit connected speech instruction significantly enhance ESL learners' perception. Based on the present study, it is suggested that teachers can use video with subtitles before practicing shadowing to train learners to perceive connected

speech sounds. For example, teachers may firstly use subtitles for promoting overall comprehension and vocabulary recognition. Then, with the same video, only the audio is played and students are directed to focus on perceiving and processing the connected speech features for purely listening and shadowing purpose. The suggested example deposits to remove the potential barriers before learning to perceive connected speech sounds and free the working memory for auditory processing to discriminate the features.

The current study also shed light on teaching connected speech with the concern of non-native contrast. It has been well-documented that the non-salient features of a second language, for instance, constants with voicing contrast or an assimilated sound at the word boundary, are often neglected (Ellis 2006; Schmidt 2012). This implies an urge to train up ESL learners' perceptual sensitivity before 'connected speech deafness'. As suggested in the current results, learning connected speech is language-specific. Therefore, teachers are recommended to provide remedial-based listening and shadowing practice on the non-salient connected speech features on top of the explicit instruction so as to further highlight the language-specific, non-salient features to ESL learners.

Conclusion

The findings of the present study imply the potential problems of learning connected speech with multimodalities. In other words, 'the more the better' is not always the case. Instead of overwhelming students with excessive inputs across modalities, ESL teachers and learners are reminded to be aware if the modalities engaged would lead to the success of perceiving connected speech sounds. Since the benefits of learning with more than one single modal cannot

be denied, as shown in the current results, the success of connected speech acquisition with bimodalities or even multimodalities may lies in the sequence of engaging the auditory, visual and motor channel during the learning process.

Reference

Abe, H. (2009). The effect of interactive input enhancement on the acquisition of the English connected speech by Japanese college students. *Phonetics Teaching & Learning Proceedings, University College London.*

Ahmadian, M., & Matour, R. (2014). The Effect of Explicit Instruction of Connected Speech Features on Iranian EFL Learners'. *International Journal of Applied Linguistics and English Literature*, *3*(2), 227-236.

Alameen, G., & Levis, J. M. (2015). 9 Connected Speech. *The Handbook of English Pronunciation*, 159.

Bahrick, L. E., & Lickliter, R. (2014). Learning to attend selectively: The dual role of intersensory redundancy. *Current directions in psychological science*, *23*(6), 414-420.

Bird, S. A., & Williams, J. N. (2002). The effect of bimodal input on implicit and explicit memory: An investigation into the benefits of within-language subtitling. *Applied Psycholinguistics, 23(4),* 509-533.

Birulés-Muntané, J., & Soto-Faraco, S. (2016). Watching Subtitled Films Can Help Learning Foreign Languages. *PloS one, 11(6),* e0158409.

Bradlow, A. R., & Bent, T. (2002). The clear speech effect for non-native listeners. *The Journal of the Acoustical Society of America*, *112*(1), 272-284.

Brown, J. D., & Kondo-Brown, K. (2006). Introducing connected speech. In J. D. Brown & K. Kondo-Brown (Eds.), Perspectives on teaching connected speech to second language speakers (pp. 1–15). Honolulu: University of Hawaii, National Foreign Language Resource Center

Celce-Murcia, M., Brinton, d.M., Goodwin, J.M., and Griner, B. 2010. Teaching Pronunciation Paperback with Audio CDs (2): A Course Book and Reference Guide, 2nd edition, Cambridge University Press.

Chan, A. Y., & Li, D. C. (2000). English and Cantonese phonology in contrast: Explaining Cantonese ESL learners' English pronunciation problems. *Language Culture and Curriculum*, *13*(1), 67-85.

Cheung, H., & Kemper, S. (1993). Recall and articulation of English and Chinese words by Chinese-English bilinguals. *Memory & cognition*, *21*(5), 666-670.

Cho, M. H., & Lee, S. (2016). The impact of different L1 and L2 learning experience in the acquisition of L1 phonological processes. *Language Sciences*, *56*, 30-44.

Comparison of English and Mandarin (Segmentals). (2014) In Pronunciation Learning Website, FHM, HKIEd. Retrieved from http://ecconcord.ied.edu.hk/phonetics_and_phonology/wordpress/?page_id=328

Coridun, S., Ernestus, M., & Ten Bosch, L. (2015, August). Learning pronunciation variants in a second language: Orthographic effects. In Scottish consortium for ICPhS 2015, M.

Wolters, J. Livingstone, B. Beattie, R. Smith, M. MacMahon, et al. (Eds.), Proceedings of the 18th International Congress of Phonetic Sciences (ICPhS 2015). Glasgow: University of Glasgow.

Cruttenden, A. (2014). Gimson's pronunciation of English. Oxon: Routledge.

Darcy, I., Ramus, F., Christophe, A., Kinzler, K., & Dupoux, E. (2009). Phonological knowledge in compensation for native and non-native assimilation. In F. Ku¨gler, C. Fe´ry, & R. van de Vijver (Eds),Variation and gradience in phonetics and phonology. Berlin: Mouton De Gruyter.

Diao, Y., Chandler, P., & Sweller, J. (2007). The effect of written text on comprehension of spoken English as a foreign language. *The American Journal of Psychology*, *120*, 237-261.

Dilley, L. C., & Pitt, M. A. (2007). A study of regressive place assimilation in spontaneous speech and its implications for spoken word recognition. *The Journal of the Acoustical Society of America*, *122*(4), 2340-2353.

Ellis, N. C. (2006). Selective attention and transfer phenomena in L2 acquisition: Contingency, cue competition, salience, interference, overshadowing, blocking, and perceptual learning. *Applied Linguistics*, *27*(2), 164-194.

Ernestus, M. (2014). Acoustic reduction and the roles of abstractions and exemplars in speech processing. *Lingua*, *142*, 27-41.

Gaskell, M. G. (2003). Modelling regressive and progressive effects of assimilation in speech perception. *Journal of Phonetics, 31(3),* 447-463.

Gow, D. W. (2003). Feature parsing: Feature cue mapping in spoken word recognition. *Perception & Psychophysics*, *65*(4), 575-590.

Hamada, Y. (2011). Improvement of listening comprehension skills through shadowing with difficult materials. *The Journal of AsiaTEFL, 8(1),* 139-162.

Harji, M. B., Woods, P. C., & Alavi, Z. K. (2010). The effect of viewing subtitled videos on vocabulary learning. *Journal of College Teaching and Learning, 7(9),* 37.

Hayati, A., & Mohmedi, F. (2011). The effect of films with and without subtitles on listening comprehension of EFL learners. *British Journal of Educational Technology, 42(1),* 181-192

Hazan, V., Sennema, A., Iba, M., & Faulkner, A. (2005). Effect of audiovisual perceptual training on the perception and production of consonants by Japanese learners of English. *Speech communication, 47(3),* 360-378.

Hede, A. (2002). An integrated model of multimedia effects on learning. *Journal of Educational Multimedia and Hypermedia, 11(2),* 177-191.

Hsu, C. K., Hwang, G. J., Chang, Y. T., & Chang, C. K. (2013). Effects of Video Caption Modes on English Listening Comprehension and Vocabulary Acquisition Using Handheld Devices. *Educational Technology & Society, 16(1),* 403-414.

Jia, X., & Fu, G. (2011). Strategies to Overcome Listening Obstacles and Improve the Listening Abilities. *US-China Foreign Language, 9(5),* 315-323.

Jun, J. (2004). Place assimilation. In Bruce Hayes, Robert Kirchner & Donca Steriade (Eds.) Phonetically based phonology (pp. 58–86) Cambridge: Cambridge University Press.

Kadota, S. (2012). Shadoingu to ondoku to eigoshutoku no kagaku [Science of shadowing, oral reading, and English acquisition]. Tokyo: Cosmopier Publishing Company.

Kato, S., & Tanaka, K. (2015). Reading Aloud Performance and Listening Ability in an L2: The Case of College-Level Japanese EFL Users. *Open Journal of Modern Linguistics, 5(02),* 187.

Kawahara, S., & Garvey, K. (2014). Nasal place assimilation and the perceptibility of place contrasts. *Open Linguistics*, *1*(1).

Khaghaninezhad, M. S., & Jafarzadeh, G. (2014). Investigating the Effect of Reduced Forms Instruction on EFL Learners' Listening and Speaking Abilities. *English Language Teaching*, *7*(1), 159.

Klein, M., Grainger, J., Wheat, K. L., Millman, R. E., Simpson, M. I., Hansen, P. C., & Cornelissen, P. L. (2014). Early activity in Broca's area during reading reflects fast access to articulatory codes from print. *Cerebral Cortex*, bht350.

Kollöffel, B. (2012). Exploring the relation between visualizer–verbalizer cognitive styles and performance with visual or verbal learning material. *Computers & Education, 58(2),* 697-706.

Kuo, F. L., Ting, W. Y., Chiang, H. K., & Pierce, B. (2013). Effectiveness of Connected Speech-Focused Instruction and Stress-Focused Instruction on Taiwanese EFL Learners. *SPECTRUM: NCUE Studies in Language, Literature, Translation*, (11), 57-69.

Kuo, L. J., Uchikoshi, Y., Kim, T. J., & Yang, X. (2016). Bilingualism and phonological awareness: Re-examining theories of cross-language transfer and structural sensitivity. *Contemporary Educational Psychology, 46,* 1-9.

Kruger, J. L. (2013). Subtitles in the classroom: balancing the benefits of dual coding with the cost of increased cognitive load. *Journal for Language Teaching, 47(1),* 29-53.

Kruger, J. L., & Doherty, S. (2016). Measuring cognitive load in the presence of educational video: Towards a multimodal methodology. *Australasian Journal of Educational Technology, 32(6),* 19-31.

Kruger, J. L., Hefer, E., & Matthew, G. (2013, August). Measuring the impact of subtitles on cognitive load: Eye tracking and dynamic audiovisual texts. In Proceedings of the 2013 Conference on Eye Tracking South Africa (pp. 62-66). ACM.

Kruger, J. L., & Steyn, F. (2014). Subtitles and eye tracking: Reading and performance. *Reading Research Quarterly*, *49*(1), 105-120.

Lado, R. (1964) Linguistics Across Cultures: Applied Linguistics for Language Teachers . Ann Arbor: The University of Michigan Press.

Li, Y. (2016). Audiovisual Training Effects on L2 Speech Perception and Production. *International Journal of English Language Teaching, 3(2),* 14-36.

Liang, D. (2015). Chinese Learners' Pronunciation Problems and Listening Difficulties in English Connected Speech. *Asian Social Science, 11(16),* 98-106.

Linebaugh, G., & Roche, T. B. (2015). Evidence that L2 production training can enhance perception. *Journal of Academic Language and Learning, 9(1),* A1-A17.

Lisker, L., & Abramson, A. S. (1970). The voicing dimension: Some experiments in comparative phonetics. In B. Hala, M. Romportl, & P. Janota (Eds.), Proceedings of the Sixth International Congress of Phonetic Sciences (pp.563–567). Prague: Academia.

Ludersdorfer, P., Wimmer, H., Richlan, F., Schurz, M., Hutzler, F., & Kronbichler, M. (2016). Left ventral occipitotemporal activation during orthographic and semantic processing of auditory words. *NeuroImage, 124,* 834-842.

Markham, P. L., & Peter, L. (2003). The influence of English language and Spanish language captions on foreign language listening/reading comprehension. *Journal of Educational Technology Systems, 31(3),* 331-341.

Martin, A., & Peperkamp, S. (2011). Speech perception and phonology. *The Blackwell companion to phonology, 4,* 2334-2356.

Mayer, R. E., Lee, H., & Peebles, A. (2014). Multimedia learning in a second language: A Cognitive load perspective. *Applied Cognitive Psychology, 28(5),* 653-660.

Mitterer, H., & McQueen, J. M. (2009). Foreign subtitles help but native-language subtitles harm foreign speech perception. *PloS one, 4*(11), e7785.

Mochizuki, H. (2006). Application of shadowing to TEFL in Japan: The case of junior high school students. *Studies in English Language Teaching, 29,* 29–44.

Mohanan, K. P. (1993) Fields of attraction in phonology. In The Last Phonological Rule: Reflections on Constraints and Derivations, John Goldsmith, ed., Chicago: University of Chicago Press, 61–116.

Nakayama, T. (2011). Weak Forms in Shadowing: How can Japanese EFL learners perform better in shadowing tasks? *The Society of English Studies, 41*, 17-31.

Nakayama, T., & Iwata, A. (2012). Differences in comprehension: visual stimulus vs. auditory stimulus. *城西大学語学教育センター研究年報, (6)*, 1-8.

Petrova, A., Gaskell, G., & Ferrand, L. (2011). Orthographic consistency and word-frequency effects in auditory word recognition: New evidence from lexical decision and rime detection. *Frontiers in Psychology, 2*:263. doi:10.3389/fpsyg. 2011.00263

Plass, J. L., Moreno, R., & Brünken, R. (2010). Cognitive load theory. New York, NY: Cambridge University Press.

Ranbom, L. J., & Connine, C. M. (2007). Lexical representation of phonological variation in spoken word recognition. *Journal of Memory and Language*, *57*(2), 273-298.

Roux, S., & Bonin, P. (2013). "With a little help from my friends": Orthographic influences in spoken word recognition. *L'Année psychologique, 113(01),* 35-48.

Schmidt, R. (2012). Attention, awareness, and individual differences in language learning. In W. M. Chan, K. N. Chin, S. Bhatt, & I. Walker (Eds.), Perspectives on individual characteristics and foreign language education (pp. 27–50). Boston, MA: Mouton de Gruyter.

Scarbel, L., Beautemps, D., Schwartz, J-L., & Sato, M. (2014). The shadow of a doubt? Evidence for perceptuo-motor linkage during auditory and audiovisual close-shadowing. *Frontiers in Psychology*, *Frontiers, 5*, 568.

Shams, L., & Seitz, A. R. (2008). Benefits of multisensory learning. *Trends in cognitive sciences, 12(11),* 411-417.

Shiki, O., Mori, Y., Kadota, S., & Yoshida, S. (2010). Exploring differences between shadowing and repeating practices: An analysis of reproduction rate and types of reproduced words. *Annual Review of English Language Education in Japan, 21,* 81-90.

Steriade, D. 2001 "Directional asymmetries in place assimilation" in Elizabeth Hume and Keith Johnson (eds.) The role of speech perception phenomena in phonology, Academic Press.

Sweller, J. (2010). Element interactivity and intrinsic, extraneous, and germane cognitive load. *Educational Psychology Review*, *22*(2), 123-138.

Sweller, J., Van Merrienboer, J. J., & Paas, F. G. (1998). Cognitive architecture and instructional design. *Educational Psychology Review*, *10*(3), 251-296.

Talsma, D., Senkowski, D., Soto-Faraco, S., & Woldorff, M. G. (2010). The multifaceted interplay between attention and multisensory integration. *Trends in Cognitive Sciences, 14(9),* 400-410.

Tamai, K (1997). Shadowing no koka to chokai process ni okeru ichizuke [The effectiveness of shadowing and its position in the listening process]. *Current English Studies, 36,* 105–116.

Tamminen, H., Peltola, M. S., Kujala, T., & Näätänen, R. (2015). Phonetic training and non-native speech perception—New memory traces evolve in just three days as indexed by the mismatch negativity (MMN) and behavioural measures. *International Journal of Psychophysiology, 97(1),* 23-29.

Wagner, R.K., Torgesen, J. & Rashotte, C.A. (1999). Comprehensive Test of Phonological Processing. Austin, TX: PROED.

Wickens, C. D. (2007). Attention to the second language. *IRAL-International Review of Applied Linguistics in Language Teaching*, *45*(3), 177-191.

Wickens, C. D., McCarley, J., Alexander, A., Thomas, L., Ambinder, M. & Zheng, S. (2007). Attention-Situation Awareness (A-SA) model of pilot error. In Pilot Performance Models, David Foyl and Becky Hooey (Eds.), Mahwah, NJ: Lawrence Erlbaum

Yuksel, D., & Tanriverdi, B. (2009). Effects of watching captioned movie clip on vocabulary development of EFL learners. *TOJET: The Turkish Online Journal of Educational Technology, 8(2),* 48–54

Zahedi, H., Sahragard, R., & Nasirizadeh, Z. (2007). The Effects of Phonological Features on Iranian EFL Learners Listening Comprehension. *Journal of Pan-Pacific Association of Applied Linguistics, 11*(2), 115-130.

Appendix A

Consent form and Information Sheet

香港教育大學
心理研究學系

參與研究同意書

< 字幕及影子跟讀法對於英語為第二語言的學習者在英語連音聽力之學習效果 >

本人同意參加由黃緯立博士負責監督，黃琇盈小姐負責執行的研究計劃。他們分別是香港教育大學的教員和學生。

本人理解此研究所獲得的資料可用於未來的研究和學術發表。然而本人有權保護本人的隱私，本人的個人資料將不能洩漏。

研究者已將所附資料的有關步驟向本人作了充分的解釋。本人理解可能會出現的風險。本人是自願參與這項研究。

本人理解我有權在研究過程中提出問題，並在任何時候決定退出研究，更不會因此而對研究工作產生的影響負有任何責任。

參加者姓名: _____

參加者簽名: _____

年齡: _____

實驗日期: _____

有關資料

< 字幕及影子跟讀法對於英語為第二語言的學習者在英語連音聽力之學習效果 >

誠邀閣下參加黃緯立博士負責監督，黃琇盈小姐負責執行的研究計劃。他們分別是香港教育大學的教員和學生。

本研究計劃的目的為: 一) 探討字幕、影子跟讀法、字幕及影子跟讀法對於以中文為母語、英語為第二語言的成年人在英語連音聽力之學習效果。為探究過了語言關鍵期的學習者能否透過交叉形式(即字幕及影子跟讀法)學習接收英語連音，本研究選擇以成年人為研究對象, 計劃將於本地大學透過電郵, 向年齡在 18 歲或以上、以英語作為第二語言的成年人作出邀請, 並透過方便抽樣法, 邀請九十位同意參與研究的人士參與實驗。

研究方法

本研究計劃於本地的大學透過隨機抽樣法將邀請九十位以英語為第二語言的成年人（年齡為 18 歲或以上）參與。本研究的實驗（約一小時）將以組別形式進行。實驗的流程如下:

實驗概要
1. 英語連音默寫前測 (10-15 分鐘) ⬇ 2. 連音短句訓練 (5-10 分鐘) ⬇ 3. 英語連音默寫後測 (10-15 分鐘) ⬇ 4. 英語語音記憶測試 (10-15 分鐘)

參與的學生將會被隨機而平均地分成三個組別，一）聆聽錄音及看字幕、二）聆聽錄音及跟讀、三）聆聽錄音、看字幕及跟讀，並按照表內的流程進行。本研究的日期及時間將按照參與者的時間彈性處理，地點為香港教育大學大埔校園。

實驗進行期間將不會涉及任何風險及不適。閣下的參與純屬自願性質。閣下享有充分的權利，在任何時候決定退出這項研究，而不會因此引致任何不良後果。凡有關閣下的資料將會高度保密，一切資料的編碼只有研究人員得悉。所有已收集的數據將於研究完成後銷毀。

本研究的成果將於香港教育大學內匯報，我們亦希望取得閣下的同意，讓此研究結果用於未來的研究和學術發表，如於會議、期刊論文等分享成果。

如閣下想獲得更多有關這項研究的資料,請以電郵 或電話 與本人，或以電郵 與本人的導師黃緯立博士聯絡。

如閣下對這項研究的操守有任何意見,可隨時與香港教育學院人類實驗對象操守委員會聯絡(電郵: ; 地址: 香港教育大學研究與發展事務處)

謝謝閣下有興趣參與這項研究。

黃琇盈

Appendix B

The Speech Stimuli in the Connected Speech Dictation

word-finals with /t/

1. He doesn't like *swee**t p**eas*.
2. They're sitting on the *we**t b**ench*.
3. They're staring at the *brigh**t m**oon*.
4. I'm looking for the *bes**t b**ar*.
5. It's going to be their *las**t m**atch*.
6. This is such a *tar**t p**each*.
7. I'll do it in a *quie**t p**lace*.

word-finals with /d/

8. He's eating a *col**d m**eal*.
9. They're stacking the *re**d b**ricks*.
10. She just bought a *sue**de b**ag*.
11. You should watch out for the *dea**d m**ice*.
12. She's buying the *gol**d p**urse*.
13. She takes good care of her *blin**d p**ets*.
14. He always reads *sa**d p**oems*.

word-finals with /n/

15. We received some *fa**n m**ails*.
16. I'm gonna pay my *ow**n b**ills*.

17. These're the little *twi**n** **m**ouse*.

18. She's asking for a *clea**n** **b**owl*.

19. We're sitting under the *gree**n** **p**alm*.

20. They're eating some *lea**n** **p**ork*.

Appendix C

The Assimilated Speech Stimuli and Their IPA Transcription.

	Target assimilated segment	IPA transcription of the citation forms	IPA transcription of the assimilated forms
word-finals with /t/			
1.	Sweet peas	/ swit piz /	[swip piz]
2.	Wet bench	/ wɛt bɛnʧ /	[wɛb bɛnʧ]
3.	Bright moon	/ braɪt mun /	[braɪm mun]
4.	Best bar	/ bɛst bɑr /	[bɛsb bɑr]
5.	Last match	/ læst mæʧ /	[læsb mæʧ]
6.	Tart peach	/ waɪt blaʊs /	[waɪb blaʊs]
7.	Quiet place	/ ˈkwaɪət pleɪs /	[ˈkwaɪəp pleɪs]
word-finals with /d/			
8.	Cold meal	/ koʊld mil /	[koʊlp mil]
9.	Red bricks	/ rɛd brɪks /	[rɛb brɪks]

10. Suede bag	/ sweɪd bæg /	[sweɪb bæg]
11. Dead mice	/ dɛd maɪs /	[dɛm maɪs]
12. Gold purse	/ goʊld pɜrs /	[goʊlp pɜrs]
13. Blind pets	/ blaɪnd pɛts /	[blaɪnp pɛts]
14. Sad poems	/ sæd ˈpoʊəmz /	[sæp ˈpoʊəmz]

word-finals with /n/

15. Fan mails	/ fæn meɪlz /	[fæm meɪlz]
16. Own bills	/ oʊn bɪlz /	[oʊm bɪlz]
17. Twin mouse	/ twɪn maʊs /	[twɪm maʊs]
18. Clean bowl	/ klin boʊl /	[klim boʊl]
19. Green palm	/ grin pɑm /	[grim pɑm]
20. Lean pork	/ lin pɔrk /	[lim pɔrk]

Appendix D

The Speech Stimuli in the Test of Articulation Rate (Cheung & Kemper, 1993)

1. Pit fur

2. Soup hint

3. Coming defect

4. Rabbit posture

5. Foreigner conjunction

6. Following explosion