Effects of Visualizing Lip Movements on Learning Chinese Orthography

Name: P. H. S. Zhang

Student Number: 2088703

Coördinators: prof. dr. P.C.J. Segers & dr. H. van der Meij

Faculty of Behavioural, Management and Social Sciences

Psychology, Learning Science

University of Twente

Date: August 28th, 2019

Summary

The current study examined what the influence was of adding the visualization of lip movements to the translation, the Chinese character presentation and audio of the pronunciation on the reading skills of Chinese characters. The second question was how the visualization of lip movements affected learning processes and what the role was of learning processes on pronunciation and translation. There were 51 participants without any knowledge of the Chinese language who were asked to learn a total of 40 Chinese characters. We compared learning in two conditions, in a within-subjects design. In the first condition, participants had to learn 20 Chinese characters in a learning module containing a person pronouncing the character next to the translation and the written Chinese character on a screen. In the second condition, participants were asked to learn another 20 Chinese characters where they only heard the pronunciation of the Chinese character and saw the translation and the written Chinese character on a screen. All participants were tested on how well they could pronounce and translate the Chinese characters. The current study showed that visualizing pronunciations influenced recalling pronunciations and translations, but only when the participants could familiarize with the Chinese language first with only the audio clips of the pronunciations. Visualizing pronunciations did not influence learning processes, but learning translations had a negative relation with the amount of times listened to the audio clips. The results suggested that prior knowledge is needed to reduce cognitive load to improve recall and that passively learning by just repetition of audio clips could harm learning translations.

Samenvatting

De huidige studie onderzocht wat de invloed was van de toevoeging van visualisaties van lipbewegingen aan de audio van de uitspraak, vertaling en de geschreven Chinese karakters op de leesvaardigheid van Chinese karakters. De tweede vraag was hoe de visualisatie van lipbewegingen invloed had op leerprocessen en wat de rol was van leerprocessen op leesvaardigheid. Er waren 51 deelnemers zonder kennis van de Chinese taal, die gevraagd werden om in totaal 40 Chinese karakters te leren. We vergeleken het leren in twee condities in een within-subjects design. In de eerste conditie werden deelnemers gevraagd om de uitspraak van 20 karakters in een leermodule te leren waarin zij de uitspraak van een karakter zagen in combinatie met de vertaling en het geschreven Chinese karakter op een computerscherm. In de tweede conditie werden deelnemers gevraagd om 20 andere karakters te leren waarbij zij alleen de audio van de

uitspraak hoorden in combinatie met de vertaling en het geschreven Chinese karakter op een computerscherm. Alle deelnemers werden getoetst op hoeveel zij de uitspraken en vertalingen konden reproduceren. De huidige studie toonde aan dat de visualisatie van de uitspraak invloed heeft op het onthouden van de uitspraak en vertalingen van de Chinese karakters, maar alleen nadat participanten zich eerst met de Chinese taal vertrouwd hebben gemaakt door eerst met audio fragmenten te leren. De visualisatie van lipbewegingen beïnvloedde leerprocessen niet, maar het leren van vertalingen had een negatieve relatie met het aantal keer beluisteren van de audio fragmenten. Dit suggereert dat voorkennis nodig is om cognitieve belasting te verminderen en het leren te bevorderen. Daarnaast kan het passief leren door alleen het herhalen van audio fragmenten leiden tot verminderd leren.

Introduction

Learning to read Chinese characters for Chinese as a foreign language (CFL) learners is difficult. Learning to read in any language is associated with connecting the spoken (phonology), written (orthography) and meaning (semantics) forms of words (Seidenberg & McClelland, 1989; Plaut, McClelland, Seidenberg & Patterson, 1996). These are the three components of the highly influential triangle model of reading, which are required to identify words that are represented in the mental lexicon. The written form is the first step of visual recognition of words. It is necessary to learn to associate the written forms with their meaning and pronunciations, when learning to read. In alphabetic languages, the pronunciation of a word can be accessed indirectly through semantics or directly through the orthography (Seidenberg & McClelland, 1989; Plaut et al., 1996). The phonological component of the Chinese language is not immediately linked to how a word is written, unlike in alphabetic languages like English or Dutch. Evidence shows that specifically seeing the pronunciation of sounds might be helpful in learning new phonological sounds in a new language (Hirata & Kelly, 2010). However, it has not yet been studied if the additional visualization of pronunciations could facilitate learning to read Chinese characters. Therefore, the current research examined whether showing the pronunciation of Chinese characters next to the translation and the written form of the Chinese character could improve learning to read Chinese characters.

Learning to read Chinese as a foreign language

In the triangle model of Seidenberg and McClelland (1989), learning to read Chinese is different from alphabetic languages (Yang, Zevin, Shu, McCandliss & Li, 2006). In Chinese, the semantics of characters is learned more quickly than learning the pronunciation (Yang et al., 2006). This is because the Chinese language is a tone language without an alphabet and is composed of a large number of orthographic units (see Figure 1) and their complex variable combinations (Xu, Chang, Zhang & Perfetti, 2013). There are at least 8,105 different Chinese characters in existence, according to the Table of General Standard Chinese Characters (2013). In alphabetic languages like English and Dutch, the orthographic element is the first step of recognizing written words for pronunciation (Pelli, Burns, Farell & Moore-Page, 2006). In Chinese, however, there is no system from which the reader can deduce the pronunciation when seeing the Chinese character. While orthography and phonology are directly linked to each other in most alphabetic languages, the Chinese language barely has a systematic association between orthography and pronunciation. This makes learning the extensive Chinese orthography very challenging for CFL learners (Everson, 1998).

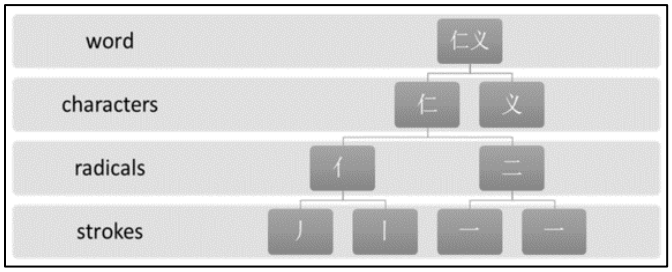

Figure 1. Overview of the Chinese orthography. Chinese words generally consist of two characters. All Chinese characters consist of radicals, which are the smallest meaningful components that can have a semantic or a phonetic purpose in a Chinese character (Shen & Ke, 2007). Lastly, radicals consist of strokes. There are about 24 different types of strokes in the Chinese language (Shen & Ke, 2007).

Since the Chinese language seems to have no direct association between the orthography and phonology, the association between orthography and semantics becomes more important when learning to read Chinese than in alphabetic languages (Zhou, Duff & Hulme, 2015). The research of Zhou et al. (2015) showed that learning both phonology and semantics of Chinese characters helps learning to read Chinese, which means that there is a strong indication that all three components of the triangle model are more applicable to the Chinese language. For learning to

read English, the research of Duff and Hulme (2012) found that only the pronunciation has a strong role in learning to read English and the semantic information has weak or no role in learning to read English. Even though there are differences between alphabetic and non-alphabetic languages on the importance of the components of the triangle model, learning the pronunciation of a language is necessary when learning to read.

Despite the fact that phonology is an important part of learning to read, learning the phonology is constrained for most western CFL learners when learning Chinese. This is because of the many novel phonological sounds of the Chinese characters, which can overload the working memory (Baddeley, 1992; Baddeley, Gathercole & Papagno, 1998). When learning new phonology, these sounds must be encoded from working memory into long-term memory. Difficulties in learning new languages is believed to be caused by overloading this process with too many new phonological sounds. This is because new representations need to be formed for the new phonological sounds. However, when adults have to learn words that consist of familiar phonological sounds, they do not have to rely much on this process. This is because they already have these representations of the sounds in their long-term memory (Baddeley et al., 1998). Thus, when learning a new language, there should be something that can make learning new phonological sounds easier.

The role of visualization of pronunciation in learning to read Chinese

A promising finding on making it easier for people to learn new phonology, is by presenting lip movements of the pronunciations of the phonology (Hirata & Kelly, 2010). The lips have meaningful visual information which creates stronger perceptions of the phonemes (Calvert et al., 1997). Research shows that for learning Japanese, audio and visual presentations of lip movements are more effective for learning new phonological sounds than only presenting the audio of the pronunciation of Japanese phonological sounds (Hirata & Kelly, 2010). This research found that especially the lip movements were most effective for perceiving difficult phonemic contrasts between their native language and Japanese. It has demonstrated that visual presentation of lip movements can strengthen the process of transferring new speech sounds into long-term memory (Hirata & Kelly, 2010).

Other research has also shown improvements of learning and comprehending new languages with visual information of pronunciation (Wang, Behne & Jiang, 2008; Wang, Behne

& Jiang, 2009) and this could be because visual articulatory information in speech and language facilitates the processing of language (Lachs & Pisoni, 2004). This is especially important for non-native speakers for whom similar sounding phonemes in a new language could be more confusing (Broersma & Cutler, 2011). A reason for this might be that adult perceivers tend to increase their gaze on the nose and mouth when other people are speaking (Yi, Wong & Eizenman, 2013). For non-native speakers of a speaker's language, the gaze to the mouth of the speaker is even higher (Barenholtz, Mavica & Lewkowicz, 2016). Aside from perceiving auditory and visual information of speech, learners of new languages have reported that mirroring by repeating words or phonemes is an important strategy for learning new pronunciations (Vitanova & Miller, 2002). Seeing someone pronouncing the new phonemes or words, might help them comprehend and repeat these phonemes or words more correctly.

Facilitating role of technology in learning Chinese

Because of the difficulty of learning to read Chinese, CFL learners often use e-learning modules in which students are presented with many ways of learning that can improve their skills (Chuang & Ku, 2011). Teachers of CFL learners mainly concentrate on improving listening and speaking skills in classes (Chang, Xu, Perfetti, Zhang & Chen, 2014). Learning to read and write Chinese characters is often skipped in class, even though these are important parts of learning a new language (Seidenberg & McClelland, 1989). With the help of technology and online lessons, CFL learners are encouraged to practice Chinese characters in their own time (Chuang & Ku, 2011). An advantage of digital learning environments is the possibility to present auditory and visual information of the pronunciations with the written form of Chinese characters, which has been shown to improve learning a new language (Hirata & Kelly, 2010; Wang et al., 2008; Wang et al., 2009). In this way, information can be processed through different senses (sight and auditory), which can improve learning (Baddeley, 1992).

Even though technology can integrate more stimuli, which could be helpful for the learner, it should be taken into account that the learner has a limitation in visual processing according to the split-attention hypothesis (Ayres & Sweller, 2005). When people are presented with multiple visual stimuli, people need to split their attention between all those sources of information, which means that learning can be more demanding (Ayres & Sweller, 2005). Therefore, it is also

important to know how different multiple visual stimuli would have an effect on the learning processes of the learner.

The present study

In all, the above overview of research shows that learning the phonology and meaning of Chinese characters improves learning to read Chinese characters (Zhou et al., 2015). However, there are difficulties with learning the pronunciation, when the new language contains many phonetically unfamiliar sounds (Everson, 1998). Since the pronunciation of Chinese is so different from languages like English or Dutch, it is difficult to learn these new phonological sounds. Research suggests that presenting lip movements of the pronunciation of a word could help build a stronger memory of the pronunciation of the new sounds (Hirata & Kelly, 2010; Wang et al., 2008; Wang et al., 2009). However, it is not clear how or if this effect of showing lip movements will improve learning to read Chinese orthography.

Therefore, in the current study, it was examined a) what the influence is of visualizing lip movements on the pronunciation and translation of Chinese characters and b) what the effect of the visualization of lip movements is on learning processes and what the relation is between learning processes and learning pronunciation and translation. To answer the research questions, students without knowledge of the Chinese language were asked to learn 40 Chinese characters in a within-subject design. In the experimental condition, the participant received an audio-visual presentation of the pronunciation of the Chinese characters. In the control condition, the participants only heard the audio of the pronunciation of the Chinese characters. Participants were tested on how well they could read the Chinese characters. Reading was tested by how well they could pronounce the Chinese character (phonology) and if they knew the translation of the Chinese characters (semantics).

The first hypothesis was that participants would perform better on pronouncing and translating the Chinese characters when they received the Chinese characters in the audio-visual presentation (video condition) than when they received the characters in the audio-only presentation (audio condition). This prediction was based on that the visual presentation of the pronunciation could help learning new phonological sounds (Hirata & Kelly, 2010), which consequently could help learning to read (Zhou et al., 2015).

Secondly, it was expected that there would be differences in learning processes among participants during learning between the video and audio condition. It was hypothesized that participants needed to perceive the presentations of pronunciations more often in audio clips than in video presentations, since there is more information available in video presentations than in audio presentations (Hirata & Kelly, 2010; Wang et al., 2008; Wang et al., 2009).

Method

Participants

There were 51 university students who participated in the current study. The students had normal eyesight and hearing (corrected with glasses/lenses or hearing aids were permitted). The students did not have serious psychological conditions (e.g., depression) or learning disabilities (e.g., dyslexia). There were a total of 41 female and 10 male participants. The age ranged from 18 to 37 years old with a mean age of 20.65 (*SD* = 3.11). None of the participants had any knowledge of the Chinese language. The native language of participants was Dutch (17.6%) and/or German (86.3%) or Bulgarian (2.0%). Of all participants, 27.5% reported being bilingual or multilingual. Specifically, the bilingual Dutch participants reported Aramean (2.0%), English (2.0%) or German (3.9%) as their second native language. Moreover, the bilingual German participants reported English (2.0%), Italian (2.0%), Spanish (2.0%), Tamil (2.0%), Turkish (7.8%) or Russian (2.0%) as their second native language. Another 2.0% reported being a native speaker of Dutch, German, Arabic and Armenian.

Due to technical problems in the recordings of the video material of two participants, the video material of those two participants was not taken into account in the current study when analyzing learning processes. All participants participated voluntarily and gave an informed consent before participation. The participants received credit points for participation.

Materials

Video and audio stimuli. The Chinese characters used in this experiment were all selected from three different study books for beginning CFL learners (Cattsoft Inc., 2005; Lee, Chan & Li, 1998; Wu et al., 1997). For the selection, *pinyin* spelling was used, which is the Romanized alphabet of the pronunciation of Chinese characters (Chung, 2002). In general, pinyin of Chinese characters is composed of initials (onset), finals (rime) and tones (Cattsoft Inc., 2005). For example,

the spelled out version of the Chinese character '马' in Pinyin is 'mǎ' which means 'horse' in English. The letter 'm' is the initial, the letter 'a' is the final and the 'ˇ' above the 'a' is the tone. In Chinese, there 22 different initials (including the combinations of consonants), six different finals with approximately 30 different combinations to form the initials and four different tones (Cattsoft Inc., 2005).

In this experiment, 40 different Chinese characters were used for learning (see Appendix A for the complete list of the Chinese characters). The 40 Chinese characters were split up into two wordlists. In these sets of Chinese characters 20 initials and 26 finals were used. All characters had different pronunciations to differentiate between lip movements of the pronunciations. The selection process of the 40 different Chinese characters was as follows: all characters of three study books for beginning CFL learners were studied thoroughly (Cattsoft Inc., 2005; Lee et al., 1998; Wu et al., 1997). The difficulty of all the Chinese characters was determined by the amount of strokes needed to write the Chinese character (see Figure 2 for an example). The average of strokes needed for Chinese characters are around 10 to 15 strokes (Chang et al., 2014). In this experiment, all the characters exceeding 15 strokes were excluded. Another exclusion criterion was that (parts of) characters would not repeat in other characters. For example, the character '马' also appears in the character '妈'. When this occurred, one of these two characters were excluded.

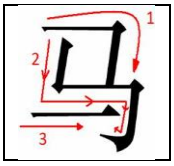

Figure 2. Example of how characters are written with its stroke order, amount of strokes and the direction of the strokes.

When the characters were selected, a native female speaker of the Chinese language was filmed pronouncing the 40 different Chinese characters. The native speaker's face and upper half of the shoulders were filmed to display a humanlike and ordinary appearance, which could promote embodiment of the pronunciation of the Chinese characters and therefore could enhance learning (Tomasino et al., 2018; Desutter & Stieff, 2017). For the audio condition, the videos were covered, so that participants could listen to the audio clip, but could not see the female speaker. All the video and audio clips were presented in PowerPoint.

For every Chinese character in the current experiment, there were two screens presented in which the first screen was presented with only the video or audio clip (see Figure 3). The second screen was presented with the Chinese character, the translation and the video or audio clip. This form of presentation was based on the research of Chung (2002), which showed that learning is enhanced when there is a temporal separation of different stimuli to fully attract the attention of the learner and to minimize overload on the working memory.

The Chinese characters were presented in the DengXian font size 150 and the translations were presented right under the Chinese character in the Calibri font size 60. Presenting the translation right under the Chinese character enhances learning, which has shown to minimize cognitive load (Lee & Kalyuga, 2011). Throughout learning, the participant could go back and forth to look back at the learned Chinese characters by pressing the previous and next arrows on the keyboard of the computer.

Process measures and learning processes. The learning activities of the participant while using the PowerPoint environment were recorded by a screen recorder. The screen recordings were coded by the amount of times the participants had heard the Chinese character. The coding of the screen recordings were coded by the experimenter using the coding program 'Behavioral Observation Research Interactive Software' (BORIS; Friard & Gamba. 2016).

Learning process was defined by the amount of times that a participant listened to the video/audio clips of the Chinese characters. Since some participants did not see all the Chinese characters, 'learning process' was measured by calculating the mean amount of times listened to the video/audio clip per perceived Chinese character.

Reading. In the current study, reading was tested on how well the participant could recall the Chinese characters on both pronunciation and translation. When the participant finished a learning module, the participant was tested on their knowledge of the pronunciation and translation of every Chinese character. The total amount of correct responses was counted. This was done by presenting the Chinese character one by one on a white A4 size paper on the table in front of the participant. The characters on the paper were presented in the DengXian font size 300. Both pronunciation and translation were rated correct or incorrect by the experimenter and a native speaker of Chinese. The interrater reliability for the raters was found to be Kappa = 0.83 ($p < 0.001$), which reflected a high agreement between the raters (Landis & Koch, 1977). Correct responses were given 1 point. When the participant correctly recalled only the initial of a Chinese

character, 0.5 points were given to the participant. When the participant correctly recalled only the final of a Chinese character, 0.5 points were given to the participant. When the participant could not recall the full pronunciation, the initial or the final of a Chinese character, 0 points were given. The total maximum score of pronunciation was a score of 20 per condition and the maximum score of translation was also a score of 20 per condition.

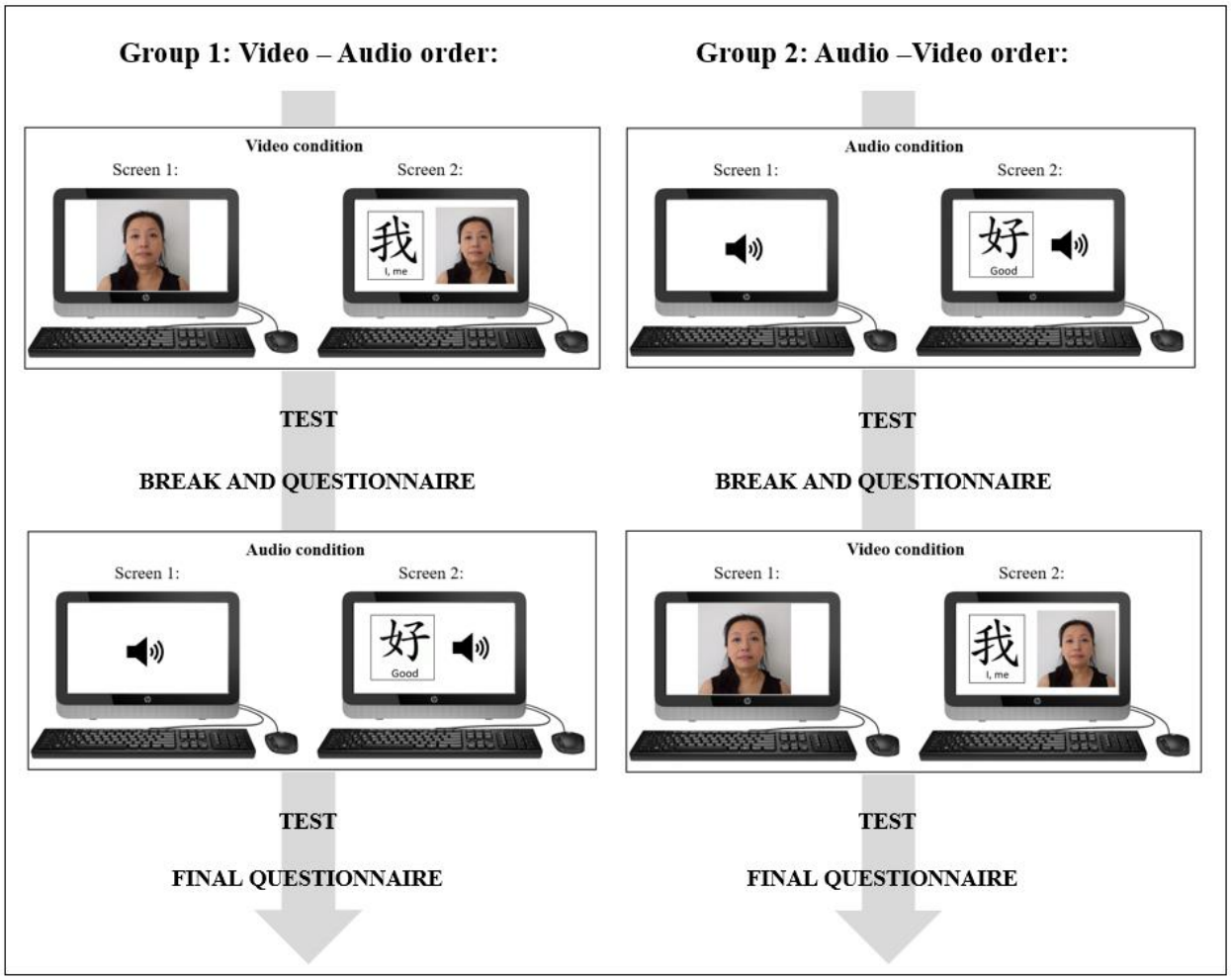

Figure 3. The procedure for participants in a randomized within-subject design, which resulted in two different order possibilities (group 1 and group 2). The participants started the first learning module in either the video (group 1) or audio (group 2) condition, in which they could learn 20 Chinese characters. In each condition, the video clip on screen 1 was square shaped and filled up the screen in length. On screen 2, the Chinese character and the translation of the Chinese character was presented on the left side of the screen and the video or audio clip was presented on the right side of the screen. After learning the words in each learning module, the participants received a test on how well they remembered the words.

Procedure

Participants registered on SONA and enrolled for a timeslot. On the day of the research, participants were welcomed and asked to sit down in front of a computer in a quiet room (see Figure 3 for a full overview of the procedure). Participants were first asked to read and fill in the informed consent. After this, the participant was instructed to use the computer to start the module. All participants received both the video and audio condition. The order of the presentation of the conditions was randomized. This meant that approximately half of the participants started with learning 20 Chinese characters (either from wordlist one or two; see Appendix A for a full list), presented in the video condition and the other half started with learning 20 Chinese characters presented in the audio condition. Both word lists were available in the audio and video condition, which meant that the wordlists were also randomized for the participants.

The learning module started with further instructions on the computer about which buttons to press and what the learning program looked like. After learning a set of 20 Chinese characters for 15 minutes, the participants received a test on pronunciation and translation on this set of 20 Chinese characters. In this test the researcher let the participant see the Chinese characters on paper one by one and asked the participant what the pronunciations and the translations were. After the first module and the first test (either the video or audio condition and either from wordlist one or wordlist two; see Appendix A), there was a short break of 10-15 minutes, in which the participant filled in a short questionnaire about their demographics and which languages they speak.

After the break, the participant started with the second round of learning the other set of 20 words in either the video or audio condition, depending on which condition they received in the first learning module before the break. After learning this set of words for 15 minutes, the participant received a test on pronunciation and translation on this set of 20 Chinese characters. After finishing the test, the participant received a short questionnaire about their experiences with learning the Chinese characters in both the conditions. They were also asked which module they preferred (video, audio, both or neither) and why.

Results

Descriptive Analysis

The distributions of pronunciation and translation in both the audio and video conditions were moderately normally distributed with Skewness and Kurtosis values between -1 and 1 (Ryu,

2011). The means and standard deviations for the video and audio conditions on the scores of pronunciation and translation are provided in Table 1. As can be seen in Table 1, on average the participants learned the pronunciation of about 25% of the Chinese characters and translation of over 50% of the Chinese characters. Table 1 also shows the bivariate correlations between the amount of times listened and pronunciation and translations in the video and audio condition.

Table 1. *Statistics for Scores on Pronunciation and Translations in the Video and Audio Condition, including Bivariate Correlations (N = 49) with Amount Listened.*

	Condition	*M (SD)*	*r* Amount listened
Pronunciation	Video	5.16 (3.25)	-.18
	Audio	5.23 (3.20)	-.24
Translation	Video	10.75 (4.76)	-.27
	Audio	11.22 (4.54)	-.35 *
Amount listened	Video	6.51 (1.72)	
	Audio	6.43 (2.13)	

* $p < .05$

Effect of Video on Pronunciation

The effect of condition (video and audio) on pronunciation was examined with a Repeated Measures ANOVA. The RM ANOVA was conducted to control for the effect of the order of the conditions on pronunciation between the video and the audio condition. Half of the participants were randomly assigned to the order of first learning a set of 20 Chinese characters in the video condition and secondly learning a new set of 20 Chinese characters in the audio condition. The other half started learning in the audio condition and secondly in the video condition. The within-subject factor was condition (video and audio), the between-subject factor was the order of the conditions (video-audio or audio-video order) and the dependent variable was pronunciation (with a score of 0-20; quantitative).

The RM ANOVA showed that there was no significant main effect of condition on pronunciation after controlling for order of conditions ($F(1, 49) = .06$, $p = .816$, $\eta_p^2 = .001$), with the mean score of pronunciation of 5.23 ($SD = 3.20$) in the audio condition and 5.16 ($SD = 3.25$) in the video condition. In addition, the analysis showed no significant between-subject effect of order on pronunciation ($F(1, 49) = 1.15$, $p = .290$, $\eta_p^2 = .023$), with the mean score of pronunciation

of 4.79 (*SD* = .53) in the video-audio order and 5.58 (*SD* = .52) in the audio-video order. However, there was a large significant interaction effect between condition and order of the condition on pronunciation ($F(1, 49) = 17.84$, $p < .001$, $\eta_p^2 = .27$; Cohen, 1988). Table 2 shows the amount of correct responses on pronunciations in the audio and video conditions in both the video-audio and audio-video order. These results showed that there was no overall effect of condition, but the effect of condition on pronunciation was dependent on the order of conditions.

Table 2. *Mean Scores and Standard Deviations of Correct Pronunciations of the Conditions and the Order of the Conditions.*

Condition	Order	*M*	*SD*	*N*
Audio	Audio-Video	4.67	2.38	26
	Video-Audio	5.80	3.51	25
Video	Audio-Video	6.48	3.58	26
	Video-Audio	3.78	2.18	25

To disentangle this interaction, further one-way ANOVA analyses were done to compare the simple main effects of the order of conditions on pronunciation separately in the audio and video conditions. In the analysis, the order of conditions (audio-video or video-audio) was the between-subject factor and the performances of pronunciation of the video and audio conditions were the dependent variables. As can be seen in Table 2, the performance of pronunciation in the audio condition was higher in the video-audio order than in the audio-video order. However, the analysis showed no significant difference in performance on pronunciation in the audio condition between the audio-video and video-audio order ($F(1, 49) = 1.60$, $p = .212$). Table 2 also shows that the performance of pronunciation in the video condition was higher in the audio-video order than in the video-audio order. The analysis showed a significant difference in performance on pronunciation in the video condition between the audio-video and the video-audio order ($F(1, 49) = 10.47$, $p = .002$). The analysis of the simple main effects showed that there was one simple main effect found for the interaction in the RM ANOVA. A significant difference of pronunciation was only found in the video condition between the audio-video and video-audio order, with higher performance on pronunciation in the audio-video order. This meant that participants performed

better on pronunciation in the video condition after the participants had to learn the Chinese characters in the audio condition.

Effect of Video on Translation

Similarly to the analysis of pronunciation, the effect of condition (video and audio) on translation was examined with a RM ANOVA. The RM ANOVA was conducted to control for the effect of the order of the conditions on translation between the video and the audio condition. The within-subject factor was condition (video and audio), the between-subject factor was the order of the conditions (video-audio or audio-video) and the dependent variable was translation (with a score of 0-20; quantitative). The RM ANOVA on translation showed that there was no significant main effect of condition on translation after controlling for order of conditions ($F(1, 49) = 1.00$, $p = .322$, $\eta_p^2 = .02$), with the mean score of translation of 11.22 ($SD = 4.54$) in the audio condition and 10.75 ($SD = 4.76$) in the video condition. In addition, the analysis showed no significant between-subject effect of order on translation ($F(1, 49) = 3.01$, $p = .089$, $\eta_p^2 = .06$), with the mean score of translation of 10.02 ($SD = .78$) in the video-audio order and 11.90 ($SD = .76$) in the audio-video order. However, there was a large significant interaction effect between condition and order of the condition on translation ($F(1, 49) = 34.18$, $p < .001$, $\eta_p^2 = .41$; Cohen, 1988). This meant that there was no overall effect of condition, but the effect of condition on translation was dependent on the order of conditions.

Table 3. *Mean Scores and Standard Deviations of Correct Translations of the Conditions and the Order of the Conditions*

Condition	Order	*M*	*SD*	*N*
Audio	Audio-Video	10.62	5.04	26
	Video-Audio	11.84	3.96	25
Video	Audio-Video	13.19	4.28	26
	Video-Audio	8.20	3.85	25

To disentangle this interaction, one-way ANOVA analyses were done to compare the simple main effects of the order of conditions on translation in the audio and video conditions separately. In this analysis, the order of conditions (audio-video or video-audio) was the between-

subject factor and the performances of translation of the video and audio conditions were the dependent variables. As can be seen in Table 3, the performance of translation in the audio condition was higher in the video-audio order than in the audio-video order. However, the analysis showed no significant difference of performance on translation in the audio condition between the audio-video and video-audio order ($F(1, 49) = .93$, $p = .340$). Table 2 also shows that the performance of translation in the video condition was higher in the audio-video order than in the video-audio order. The analysis showed a significant difference of performance on translation in the video condition between the audio-video and the video-audio order ($F(1, 49) = 19.12$, $p < .001$). The analysis of the simple main effects showed that there was one simple main effect found for the interaction in the RM ANOVA. A significant difference of translation was only found in the video condition between the audio-video and video-audio order, with higher performance on translation in the audio-video order. This meant that participants performed better on translation in the video condition after the participants had to learn the Chinese characters in the audio condition.

Role of Learning Processes on Learning Chinese characters

To answer if there was a difference in learning processes between video and audio conditions, a t-test was conducted with the condition (video and audio) as the independent within-subject factor and the 'mean of the amount of times listened' to the audio/video clips as the dependent variable (quantitative). The analysis showed that there was no significant difference between the video and audio condition on amount of times listened (t (48) = -.32, $p = .748$, $d = -0.04$, *two-tailed*).

To answer what the relation was of learning processes with pronunciation and translation in the video and the audio condition, a bivariate Pearson correlational analysis was conducted with the amount of times listened, pronunciation and translation as the variables. The analysis showed that there was only a significant negative correlation between amount of times listened and translation in the audio condition ($p = .014$). This result indicated that as the amount of times listened increased, learning translation decreased in the audio condition and vice versa. The negative correlations between amount of times listened and translation in the video condition ($p = .065$) and pronunciation in the audio ($p = .094$) and video ($p = .214$) conditions were non-significant.

Discussion

The purpose of the current study was to examine a) what the effect was of visualizing lip movements on learning the pronunciation and translation of Chinese characters and b) what the effect was of visualizing lip movements on learning processes and what the relation was between learning processes and performance on pronunciation and translation. Results showed that visualizing lip movements only influenced recall on pronunciation and translation depending on the order of conditions. To be exact, the current research showed a learning effect in which participants performed better on pronunciation and translation in the video condition after the participants had to learn the Chinese characters in the audio condition. This meant that participants who received the video condition in the audio-video order had significantly higher performances on pronunciation and translation in the video condition than in the video-audio order. Furthermore, there was no influence of visualization of lip movements on learning processes. However, there was a significant negative relation found between amount of times listened and performance on translation in only the audio condition, which indicated that the more a participant listened to the audio clips, the lower the scores on translation was in the audio condition.

These findings did support the first hypothesis that participants would perform better on recalling pronunciations and translations when they saw the lip movements of the pronunciations, but only after the participants had to learn the Chinese characters in the audio condition. The current study supports the results of the importance of visual information of pronunciation found in the studies of Hirata and Kelly (2010), Wang et al., (2008) and Wang et al. (2009). However, the current study showed a form of a learning effect, in which participants performed better at the second time they had to learn a new set of Chinese characters. It is possible that learning Chinese characters first in the audio condition helps to familiarize and attain prior knowledge of the Chinese language, which seemed to help participants to optimally perceive and learn Chinese characters in the video condition. The advantage of prior knowledge has indeed been shown to be effective for learning in previous studies and it reduces cognitive load (Hailikari, Katajavuori & Lindblom-Ylänne, 2008).

A reason why lip movements did not have a main effect on pronunciation and translation could be that perceiving the novel phonological sounds of a new language can be more demanding than when only hearing the audio of the pronunciation. Even though the current study tried to minimize the overload on the working memory by separating the stimuli based on the research of

Chung (2002), this might have not been enough. When participants received the video first in the video-audio order, the video might still have caused cognitive overload. As mentioned in the introduction, in learning new languages, it is known that learning novel phonological sounds can overload the working memory (Baddeley et al., 1998). The extra visual information of pronunciation could have overloaded the working memory even more and could have impaired learning. This is in line with the split attention hypothesis, which states that people have to split their attention and integrate between separated sources of information, which could increase cognitive load (Ayres & Sweller, 2005). In the current study, participants had to split their attention between visual information of the Chinese orthography, translation and the video of the pronunciation. All this information could have demanded a great portion of the executive functioning when participants immediately started with the video condition.

The second hypothesis was that participants would need to listen to audio clips more often than to video material. However, the current study did not confirm this. The results showed that there was no influence of visualization of lip movements on learning processes. This result can also be related to overloading the working memory and the split attention hypothesis, which could have impaired learning (Ayres & Sweller, 2005; Baddeley et al., 1998). There is the possibility that participants needed to see the video clips more often than expected, because the learning environment contained more visual information. To study all the visual information presented in the video condition, the participant needed to look more often at the videos to compensate for splitting their attention on the video and Chinese character.

Even though visualization of lip movements did not influence learning processes, further results indicated that learning translation was negatively correlated with the amount of times listened to the audio clips. This could be the result of participants passively learning the Chinese characters in the audio condition. Participants in the current study could not actively use materials such as paper and pen during learning. Participants could only go back and forth to see all the Chinese characters. Although participants were encouraged to repeat the pronunciation out loud, which enhances learning (Vitanova & Miller, 2002), participants did not always follow this instruction. Previous research has shown that passively learning foreign words, by only visual repetition, is the worst learning strategy to use and does not improve learning (Gu & Johnson, 1996). It seemed that participants in the current study who repeatedly only listened to the audio clips tended to also learn less translations.

One of limitations of the current study was that video and audio clips were used of only one native speaker. Since pronunciation could also be dependent on the speaker (Wang et al., 2016), the influence of the speaker itself could not be excluded in the current research. In the current research, the majority of the participants reported that they preferred seeing a person pronouncing the Chinese character. However, some people preferred only listening to the audio and indicated that seeing someone pronouncing the characters was distracting. Future research could look at how gender or different accents could affect learning to read Chinese characters and what the role is of the preference or the learning style of the learner.

A second limitation of the current study was the design of the study. In contrast of the current study, the research of Hirata and Kelly (2010) used a learning module consisting of multiple sessions spread over two weeks. Such a learning module would resemble an ordinary learning module, in which participants can learn in their own pace and spread over a longer period of time. Having control over the pace of a learning module has proven to be an important and beneficial factor in learning (Tullis & Benjamin, 2011). In the current study, the participants were asked to learn as many Chinese characters in just 15 minutes. Future research can take this into account and can also look at the long-term effects of a full learning module.

In spite of the limitations, the current study did contribute to the problem that there was not much research on how or if there was an effect of showing lip movements on learning to read Chinese orthography. More knowledge on learning Chinese orthography is important, because it is becoming more popular for people to learn Chinese orthography digitally (Zhu, Fung & Wang, 2012). Therefore, it is important to organize and present learning material in a way that corresponds to the learners ability to maximize learning (Shen, 2013).

The present study focused on the influence of visualizing lip movements of the pronunciation of Chinese characters on the recall of pronunciation and translation. The current study also looked at what the influence was on learning processes. The experiment showed that visualizing pronunciations did influence recalling pronunciations and translations, but only when participants got to familiarize with the Chinese language first with only the audio clips of the pronunciation. Furthermore, the current study found that visualizing pronunciations did not influence learning processes, but learning translations was found to have a negative relation with the amount of times listened to the audio clips.

Reference List

Ayres, P., & Sweller, J. (2005). The split-attention principle in multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 206–226). New York: Cambridge University Press.

Baddeley, A. (1992). Working memory. *Science 255*, 556-559. doi: 10.1126/science.1736359

Baddeley, A. D., Gathercole, S. E., & Papagno, C. (1998). The phonological loop as a language learning device. *Psychological Review, 105*, 158–173. doi: 10.1037/0033-295X.105.1.158

Barenholtz, E., Mavica, L., & Lewkowicz, D. J. (2016). Language familiarity modulates relative attention to the eyes and mouth of a talker. *Cognition, 147*, 100–105. doi: 10.1016/j.cognition.2015.11.013

Broersma, M., & Cutler, A. (2011). Competition dynamics of second-language listening. *Quarterly Journal of Experimental Psychology, 64*(1), 74–95. doi: 10.1080/17470218.2010.499174

Calvert, G. A., Bullmore, E. T., Brammer, M. J., Campbell, R., Williams, S. C. R., McGuire, P. K., Woodruff, P. W., Iversen, S. D., & David, A. S. (1997). Activation of auditory cortex during silent lipreading. *Science, 276*, 593–596. doi: 10.1126/science.276.5312.593

Cattsoft, Inc. (Eds.). (2005). *Basic Chinese Volume 1.* Beijing: Beijing Language and Culture University Press.

Chang, L. Y., Xu, Y., Perfetti, C. A., Zhang, J., & Chen, H. C. (2014). Supporting Orthographic Learning at the Beginning Stage of Learning to Read Chinese as a Second Language. *International Journal of Disability, Development and Education, 61*, 288-305. doi: 10.1080/1034912X.2014.934016

Chuang, H., & Ku, H. (2011). The effect of computer-based multimedia instruction with Chinese character recognition. *Educational Media International, 48,* 27-41. doi: 10.1080/09523987.2011.549676

Chung, K. H. (2002). Effective use of Hanyu pinyin and English translations as extra stimulus prompts on learning of Chinese characters. *Educational Psychology, 22,* 149-164. doi: 10.1080/01443410120115238

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences.* Hillsdale, NJ: Erlbaum.

DeSutter, D., & Stieff, M. (2017). Teaching students to think spatially through embodied actions:

Design principles for learning environments in science, technology, engineering, and mathematics. *Cognitive Research, 2,* 1-20. doi: 10.1186/s41235-016-0039-y

Duff, F. J., & Hulme, C. (2012). The role of children's phonological and semantic knowledge in learning to read words. *Scientific Studies of Reading, 16*, 504–525. doi: 10.1080/10888438.2011.598199

Everson, M. (1998). Word recognition among learners of Chinese as a foreign language: Investigating the relationship between naming and knowing. *Modern Language Journal, 82,* 194–204. doi: 10.1111/j.1540-4781.1998.tb01192.x

Friard, O., & Gamba, M. (2016). BORIS: a free, versatile open-source event-logging software for video/audio coding and live observations. *Methods in Ecology and Evolution, 7*, 1325–1330. doi: 10.1111/2041-210X.12584

Gu, Y., & Johnson, R. K. (1996). Vocabulary learning strategies and language learning outcomes. *Language Learning, 46,* 643-679. doi: 10.1111/j.1467-1770.1996.tb01355.x

Hirata, Y., & Kelly, S. D. (2010). Effects of lips and hands on auditory learning of second-language speech sounds. *Journal of Speech, Language, and Hearing Research, 53*, 298–310. doi: 10.1044/1092-4388(2009/08-0243)

Lachs, L., & Pisoni D. B. (2004). Specification of cross-modal source information in isolated kinematic displays of speech. *The Journal of the Acoustical Society of America*, *116*(1), 507–518. doi: 10.1121/1.1757454

Landis, J. R., Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33,*159-174. doi: 10.2307/2529310

Lee, S. H., Chan, S. W., & Li, P. Y. (1998). *Chinese Language*. Amsterdam: Project Chinees Onderwijs in Eigen Taal Amsterdam

Lee, C. H., & Kalyuga, S. (2011). Effectiveness of different pinyin presentation formats in learning Chinese characters: A cognitive load perspective. *Language Learning, 6,* 1099-1118. doi: 10.1111/j.1467-9922.2011.00666.x

Myklebust, H. R. (1956). Language Disorders in Children. *Exceptional Children, 22,* 163-166. doi: 10.1177/001440295602200406

Pelli, D. G., Burns, C. W., Farell, B., & Moore-Page, D. C. (2006). Feature detection and letter identification. *Vision Research, 46,* 4646–4674. doi:10.1016/j.visres.2006.04.023

Plaut, D. C., McClelland, J. L., Seidenberg, M. S., & Patterson, K. (1996). Understanding normal

and impaired word reading: computational principles in quasi-regular domains. *Psychological Review, 103,* 56-115. doi: 10.1037//0033-295X.103.1.56

Ryu, E. (2011). Effects of skewness and kurtosis on normal-theory based maximum likelihood test statistic in multilevel structural equation modeling. *Behavior Research Methods, 43*, 1066-1074. doi: 10.3758/s13428-011-0115-7

Seidenberg, M. S., & McClelland, J. L. (1989). A distributed, developmental model of word recognition and naming. *Psychological Review, 96,* 523–568. doi: 10.1037//0033-295X.96.4.523

Shen, H. H. (2013). Chinese L2 literacy development: cognitive characteristics, learning strategies, and pedagogical interventions. *Language and Linguistics Compass, 7,* 371-387. doi: 10.1111/lnc3.12034

Shen, H. H., & Ke, C. (2007). Radical awareness and word acquisition among nonnative learners of Chinese. *Modern Language Journal, 91*, 97–111. doi: 10.1111/j.1540-4781.2007.00511.x

Table of general standard Chinese characters. (2013). *Ministry of Education of the People's Republic of China.*

Tomasino, B., Nobile, M., Re, M., Bellina, M., Garzitto, M., Arrigoni, F., Molteni, M., Fabbro, F., & Brambilla, P. (2018). The mental simulation of state/psychological verbs in the adolescent brain: An fMRI study. *Brain Cognition, 123*, 34–46. doi: 10.1016/j.bandc.2018.02.010

Tullis, J. G., & Benjamin, A. S. (2011). On the effectiveness of self-paced learning. *Journal of Memory and Language, 64,* 109–118. doi: 10.1016/j.jml.2010.11.002

Vitanova, G., & Miller, A. (2002). Reflective practice in pronunciation learning. *The Internet TESL Journal, 8*(1). Retrieved from: http://iteslj.org/

Wang, Y., Behne, D. M., & Jiang, H. (2008). Linguistic experience and audio-visual perception of non-native fricatives. *The Journal of the Acoustical Society of America, 124*(3), 1716–1726. doi: 10.1121/1.2956483

Wang, Y., Behne, D. M., & Jiang, H. (2009). Influence of native language phonetic system on audio-visual speech perception. *Journal of Phonetics, 37*(3), 344–356. doi: 10.1016/j.wocn.2009.04.002

Wang, J., Samal, A., Rong, P. & Green, J. R. (2016). An optimal set of flesh points on tongue

and lips for speech-movement classification. *Journal of Speech, Language, and Hearing Research, 59*, 15–26. doi: 10.1044/2015_JSLHR-S-14-0112

Wu, X. M., Zheng, L. G., Jia, Y. M., Peng, L. M., Lai, G. X. & Fan, P. X. (1997). *Chinese Volume 1*. Jinan: Jinan University Press.

Xu, Y., Chang, L., Zhang, J., & Perfetti, C. A. (2013). Reading, writing, and animation in character learning in Chinese as a foreign language. *Foreign Language Annals, 46*, 423–444. doi: 10.1111/flan.12040

Yang, J., Zevin, J. D., Shu, H., McCandliss, B. D., & Li, P. (2006). *A "triangle model" of Chinese reading.* Paper presented at 28th annual conference of the cognitive science society, Mahwah, NJ.

Yi, A., Wong, W., & Eizenman, M. (2013). Gaze patterns and audiovisual speech enhancement. *Journal of Speech, Language, and Hearing Research, 56*(2), 471–480. doi: 10.1044/1092-4388

Zhou, L., Duff, F. J., Hulme, C. (2015). Phonological and semantic knowledge are causal influences on learning to read words in Chinese. *Scientific Studies of Reading, 19*, 409-418. doi: 10.1080/10888438.2015.1068317

Zhu, Y., Fung, A. S. L., & Wang. H. (2012). Memorization effects of pronunciation and stroke order animation in digital flashcards. *Computer Assisted Language Instruction Consortium Journal, 29*, 563-577.

Appendix A

	List 1					
	Chinese	**Pinyin**	**English**	**Dutch**	**German**	**Amount of strokes**
1	八	bā	eight	acht	acht	2
2	跑	pǎo	to run	rennen	rennen	12
3	名	míng	name	naam	Name	6
4	风	fēng	wind	wind	Wind	4
5	对	duì	correct	correct	korrekt	5
6	头	tóu	head	hoofd	Kopf	5
7	女	nǚ	woman	vrouw	Frau	3
8	流	liú	to flow	vloeien	fließen	10
9	坐	zuò	to sit	zitten	setzen	7
10	草	cǎo	grass	gras	Gras	9
11	真	zhēn	true	echt	echt	10
12	吃	chī	to eat	eten	essen	6
13	手	shǒu	hand	hand	Hand	4
14	今	jīn	today	vandaag	heute	4
15	秋	qiū	autumn	herfst	Herbst	9
16	下	xià	down	onder	unten	3
17	关	guān	to close	dichtdoen	schließen	6
18	哭	kū	To cry	Huilen	weinen	10
19	花	huā	flower	bloem	Blume	7
20	云	yún	cloud	wolk	Wolke	4
						Average strokes: 6.40

	List 2					
	Chinese	**Pinyin**	**English**	**Dutch**	**German**	**Amount of strokes**
1	北	běi	north	noord	Norden	5
2	朋	péng	friend	vriend	Freund	8
3	门	mén	door	deur	Tür	3
4	飞	fēi	to fly	vliegen	fliegen	3
5	电	diàn	electricity	elektriciteit	Elektrizität	5
6	停	tíng	to stop	stoppen	stoppen	11
7	年	nián	year	jaar	Jahr	6
8	劳	láo	to work	werken	arbeiten	7
9	再	zài	again	opnieuw	schon wieder	6
10	从	cóng	from	van	von	4
11	竹	zhú	bamboo	bamboe	Bambus	6
12	车	chē	vehicle	voertuig	Fahrzeug	4
13	山	shān	mountain	berg	Berg	3

14	讲	jiǎng	to speak	spreken	sprechen	6
15	去	qù	to go	gaan	gehen	4
16	雪	xuě	snow	sneeuw	Schnee	11
17	国	guó	country	land	Land	8
18	看	kàn	to watch	kijken	schauen	9
19	河	hé	river	rivier	Fluss	8
20	羊	yáng	sheep	schaap	Schaf	6
						Average strokes: 6.15