อิทธิพลของภาษาต่างประเทศและการลดคุณภาพโสตทัศนสัญญาณต่อการพึ่งพาข้อมูลภาพในการรับรู้เสียงพูดในผู้พูดภาษาไทยเป็นภาษาแม่

Effects of Foreign Language and Audiovisual Degradations on Degrees of Visual Influence in Speech Perception Among Native Speakers of Thai

รัตนสุวรรณ ระวรรณ[1*], และ ศุจิณัฐ จิตวิริยนนท์[2]

Rattansuwan Rawan[1*], and Sujinat Jitwiriyanont[2]

[1*] คณะศิลปศาสตร์, มหาวิทยาลัยเทคโนโลยีราชมงคลกรุงเทพ, กรุงเทพ, ประเทศไทย

[1*] Faculty of Liberal Arts, Rajamangala University of Technology Krungthep, Bangkok, Thailand

[2] คณะอักษรศาสตร์, จุฬาลงกรณ์มหาวิทยาลัย, กรุงเทพ, ประเทศไทย

[2] Faculty of Arts, Chulalongkorn University, Bangkok, Thailand

Received: March 16, 2026 Revised: July 15, 2026 Accepted: July 27, 2026

บทคัดย่อ

การศึกษานี้มีวัตถุประสงค์ ๑) เพื่อสำรวจอิทธิพลของภาษาต่างประเทศและการลดคุณภาพของสัญญาณต่อการพึ่งพาข้อมูลภาพในการรับรู้เสียงพูด และเพื่อสำรวจอิทธิพลของการลดคุณภาพของสัญญาณต่อ ๒) ความแม่นยำในการรับรู้ และ ๓) ระยะเวลาตอบสนอง กลุ่มตัวอย่างเป็นผู้พูดภาษาไทยเป็นภาษาแม่จำนวน ๒๒ คน เข้ารับการทดสอบการรับรู้เสียงพูดด้วยการระบุเสียงพูดผ่านตัวเลือก ๕ ตัวเลือก โดยเปรียบเทียบระหว่างเสียงที่พูดภาษาแม่เป็นภาษาไทยและเสียงพูดภาษาต่างประเทศเป็นภาษาจีนกลาง โดยวัดความแม่นยำในการรับรู้จากรายการทดสอบ ๓ รูปแบบ ได้แก่ รูปแบบเสียงอย่างเดียว รูปแบบเสียงพร้อมภาพที่สอดคล้องกัน และรูปแบบเสียงพร้อมภาพที่ขัดแย้งกัน เพื่อคำนวณอัตราการพึ่งพาข้อมูลภาพและวัดค่าระยะเวลาตอบสนองต่อเสียงพูด วิเคราะห์ค่าเฉลี่ยความแม่นยำและค่าระยะเวลาตอบสนองตามเงื่อนไขคุณภาพของสัญญาณและภาษา สถิติจากผลการทดลองพบว่า ๑) ผู้พูดภาษาไทยเป็นภาษาแม่มีระดับอิทธิพลของข้อมูลภาพในการรับรู้เสียงพูดภาษาจีนกลางสูงกว่าภาษาไทยอย่างมีนัยสำคัญทางสถิติ ๒) เมื่อพิจารณาผลของเสียงสีขาว พบว่าการเพิ่มความเข้มของเสียงสีขาวลดความแม่นยำในการรับรู้อย่างมีนัยสำคัญ ($p < .001$) ในทุกรูปแบบรายการทดสอบในภาษาจีนกลาง แต่มีนัยสำคัญเฉพาะรายการทดสอบรูปแบบเสียงพร้อมภาพที่ขัดแย้งกันในภาษาไทยเท่านั้น เมื่อพิจารณาผลของการลดความคมชัดของภาพ พบว่าไม่ส่งผลต่อความแม่นยำในการรับรู้อย่างมีนัยสำคัญ แต่สัมพันธ์กับระยะเวลาตอบสนองที่สั้นลงอย่างมีนัยสำคัญ ($p = .021$) เฉพาะในรายการทดสอบรูปแบบเสียงพร้อมภาพที่ขัดแย้งกันในภาษาจีนกลางเท่านั้น ผลการศึกษาแสดงให้เห็นว่าการพึ่งพาข้อมูลภาพในการรับรู้เสียงพูดมีลักษณะยืดหยุ่นและแปรผันตามทั้งภาษาที่รับฟังและคุณลักษณะของโสตทัศนสัญญาณที่ได้รับ

คำสำคัญ: การรับรู้เสียงพูด ภาษาต่างประเทศ การลดคุณภาพโสตทัศนสัญญาณ การพึ่งพาข้อมูลภาพในการรับรู้เสียงพูด

[*] Corresponding author. Tel.: 088 260 0644

Email address: rattanasuwan.r@mail.rmutk.ac.th

DOI: 10.14456/lartsj.2026.24

Abstract

This study examined 1) the effects of foreign language and audiovisual degradation on visual influence in speech perception, and examined the effect of degradation on 2) perception accuracy, and 3) reaction time in speech perception. Participants were 22 native speakers of Thai who completed a speech identification test with 5 options, comparing their perception of speech in their native language and a foreign language (Mandarin Chinese). Perception accuracy from audio-only trials, congruent audiovisual trials, and incongruent audiovisual trials was used to calculate degrees of visual influence. Reaction time to the trials was also measured. Analyses of experimental data were conducted to examine the effect of audiovisual degradations and foreign language. Statistical analyses of the data revealed that 1) participants presented significantly lower degrees of visual influence when listening to speech in Thai compared to Mandarin Chinese. 2) White noise significantly worsened perception across all modalities in Mandarin Chinese language blocks, but did so only in the incongruent audiovisual modality for native language ($p < .001$). In terms of visual degradation, blurry visuals caused a non-significant reduction in perception accuracy and are associated with significantly shorter reaction time ($p = .021$) only in incongruent audiovisual trials in the Mandarin Chinese language blocks. This study provides proofs that a foreign language plays a role in increasing the importance of visual information and the degradation analyses suggests that perception of a foreign language is more effective without noise and visual blur.

Keywords: speech perception, foreign language, audiovisual degradation, visual influence

Introduction

Communication among humans happens countless times in a day with the purpose of establishing shared understanding between speaker and listeners, but unfortunately, when communicating through speech, a noisy environment and lower visibility of the talker's face may make speech perception less accurate. There is also evidence that listening to speech in a foreign language is less accurate than native language, and that listening to degraded speech incurs reduced perception accuracy [2-4,13] (Bidelman et al., 2020; Broersma & Scharenborg, 2010; Bsharat-Maalouf & Karawani, 2022; Schwarz et al., 2022). Speech perception is more effective when accompanied by information from the eyes than solely listening with the ears [10] (Lalonde & Werner, 2019). And with additional visual information, speech perception becomes more accurate in a degraded environment, with the help of visual cues [8-9] (Hazan et al., 2010; Hazan & Li, 2008).

One way to measure how much a person relies on eyes to perceive speech is through degree of visual influence [15] (Sekiyama & Burnham, 2008), and literature shows that this degree of visual influence varies depending on a few factors such as test language [6] (Chen & Hazan, 2009) and degradations [8-9] (Hazan et al., 2010; Hazan & Li, 2008). This means that under circumstances where a speaker of one language listens to speech in another language or is presented with degraded speech or unclear video, how much that person relies on their eyes to help perceive speech is different. The studies have been conducted in

speakers of many major languages, but none has investigated this in native Thai listeners. The findings also suggest that visual influence is flexible and context-dependent; however, it remains unclear whether the same pattern appears in Thai listeners.

To test the effect of a foreign language, Mandarin Chinese was used because it contains slightly more number of consonants than Thai, and its consistent distribution of fricative sounds across different places of articulation allows better understanding of how Thai listeners process the non-existent foreign consonant sounds (/ʂ/ and /x/). To control for the effect of foreign language, the participant reported not having learned, nor being exposed to Mandarin Chinese in their daily life.

Perception accuracy has been a backbone in speech perception studies, until recently more studies have paid increasing attention to reaction time to the speech stimuli. It is an interesting aspect to investigate as it shows whether the participants hesitate or delay time to reach a decision in the identification task. Understanding reaction time will provide additional insight into the relationship between time and language processing in humans. Previous literature found that participants responded more slowly to audiovisual stimuli of speech produced by a speaker with a face mask than to ones without a face mask [13] (Schwarz et al., 2022). Another study found slower reaction time in a categorical perception task with -5SNR condition (noise is 5dB louder than the speech) than 0SNR condition (speech and noise at equal loudness) and the noise-free condition [2] (Bidelman et al., 2020). Results from these two studies showed that reaction time to speech stimuli can be slow under a compromised environment of speech perception.

Literature reviews addressed how speech perception, to some extent, relies on visual information as it can restore accuracy in auditorily degraded speech and when the speech is in a foreign language. The influence of visual information can be measured through degrees of “visual influence”, and it varies depending on the language of the perceived speech. Visual influence studies have been conducted only in major languages, but Thai.

This study presents an elaborate methodology for the development of tools to study the role of information from the eyes on audiovisual speech perception, and it examines the effect of foreign language (Mandarin Chinese) and audiovisual degradations (white noise and a Gaussian blur) on how accurate and how fast human speech perception is under those degradation conditions.

Objectives

This study aims to: 1) examine the degree of visual influence in native Thai listeners when perceiving speech in their native language versus a foreign language (Mandarin Chinese); and 2) investigate how audiovisual degradations (white noise and Gaussian blur) affect speech perception accuracy reaction time in speech perception. The findings are expected to provide insights into how Thai speakers rely on visual information under non-optimal listening conditions.

Literature Review

1. Effect of a foreign language on speech perception

Perception of speech in a foreign language tends to be less accurate than speech in the native language. Studies found that the native phonotactic constraints influence the perception of a foreign language – that certain phonological rules allow combinations of speech sound in a particular language. This has been documented in various speech perception studies in major languages such as English, Mandarin, Spanish, and Japanese [5, 7, 9,14, 17] (Burnham & Dodd, 2018; Davis & Kim, 2006; Hazan & Li, 2008; Sekiyama, 1997; Sekiyama & Tohkura, 1993)

2. Effect of audiovisual degradation on speech perception

Studies have also found that noise, as auditory degradation, and blur, as visual degradation, have a negative effect on perception accuracy. [3] Broersma and Scharenborg (2010) masked different kinds of noise (competing talker, speech-shaped noise, and modulated speech-shaped noise) onto speech stimuli and put those into speech identification tests. They found lower perception accuracy from noisy speech stimuli. Moreover, they found that different types of noise affect perception to different degrees. [8] Sekiyama and Burnham (2008) found that lower perception accuracy is associated with louder noise, meaning that the louder the noise gets, the lower the perception accuracy is.

Apart from noise, visual blur is also associated with lower perception accuracy. [8] Hazan et al. (2010) and [9] Hazan and Li (2008) found decreased perception accuracy from stimuli that have been masked with a Gaussian blur. To be more specific, [9] Hazan and Li (2008) manipulated visual blur using 30.2 and 45 Gaussian blurring on Adobe Premier and developed a speech identification test for speakers of English and speakers of Mandarin Chinese. The results showed decreasing mean accuracy as blur increases in both groups of participants.

3. Role of visual information on speech perception

Extra information brought in by the eyes can either reinforce perception accuracy or alter perception. A seminal study from [11] McGurk and Macdonald (1976) demonstrated that, when information from eyes and ears is incongruent, the perception can be altered. In this work, some native speakers of English participants reported ‘da’ responses that are ‘fused’ when they were presented with a visual video of a person speaking a word /ba/ but an auditory sound of the word /ga/. This phenomenon was then termed “the McGurk effect,” and it was found in speakers of other languages too. This leads to an assumption that visuals and speech are somehow integrated during the process of human speech perception. Moreover, visual information can be helpful to speech perception both when the quality of speech is uncompromised and degraded. [10] Lalonde and Werner (2019) examined the benefits of visual cues on speech perception and found significantly more accurate perception in speech trials with visual cues than speech-only trials in both detection and discrimination tasks. More interestingly, visual cues showing lip movement were contributing to significantly more accurate speech perception [13] (Schwarz et al., 2022). In contrast, visual information can be misleading when it goes against auditory information.

4. Calculation of visual influence and its flexibility

It has been clear that there is integration between auditory information and visual information in speech perception. An effort has been made to quantify how much the eyes are involved in speech perception. [17] Sekiyama and Tohkura (1993) put forward the idea of visual bias as the difference between error from audiovisual perception and error from auditory perception. Another version was proposed later by [15] Sekiyama and Burnham (2008), additionally considering the accuracy when auditory content is different (incongruent) from the visual content. This reflects a finer assessment of how much a listener relies on the eyes, as there is a comparison between helpful visuals and misleading visuals, compared to the baseline auditory content. In this study, visual influence was calculated based on perception accuracy across three modalities: audio-only (AO), congruent audiovisual ($AV_{congruent}$), and incongruent audiovisual ($AV_{incongruent}$). Following [8] Sekiyama & Burnham (2008), visual influence 1* consists of two components: visual augmentation 2* - the benefit gained from visual cues when visual and auditory cues are matched, and visual interference 3* – the perceptual disruption caused when auditory and visual information mismatch.

$$1 * \ Visual\ influence = Visual\ augmentation + Visual\ interference$$
$$2 * \ Visual\ augmentation = \%AVcongruent - \%AO$$
$$3 * \ Visual\ interference = \%AO - \%AVincongruent$$

For example, Participant AA has an accuracy score of 75 from audio-only modality, 90 from congruent audiovisual modality, and 50 from incongruent audiovisual modality. His visual augmentation is 90-75 = 15, and his visual interference is 75-50 = 25. Therefore, his degree of visual influence is 15 + 25 = 40.

Studies also showed that degrees of visual influence are flexible under different circumstances. The first factor to varying degrees of visual influence is a foreign language. [9] Hazan and Li (2008) found a significant effect of stimulus language in their speech perception study, meaning that speech in a foreign language decreases perception accuracy, compared to the native language. [6] Chen and Hazan (2009) also found that native speakers of English had significantly higher degrees of visual influence when presented with speech in Mandarin Chinese. Several precedent studies on speech perception, such as [7, 14, 17] Davis and Kim (2006), Sekiyama (1997), and Sekiyama and Tohkura (1993) also showed lower accuracy from foreign language speech, which lends support to the first factor. The second factor is degradation. Speech stimuli in noise negatively affect perception accuracy in [8] Hazan et al. (2010). The perception accuracy of noisy speech is lower than noiseless speech in both native speakers of Mandarin Chinese and native speakers of English. Consequently, this decreases the degree of visual influence.

Research Methodology

Participants

Recruitment of the participants was approved by the research ethics committee of Chulalongkorn University (Project Number 670463). Inclusion criteria include a) age of over 18, b) not diagnosed with having hearing or vision disorder, c) having normal or corrected vision, d) no report of color-blindness, e) being a

native speaker of Thai, and must not have learned Mandarin Chinese or used Mandarin Chinese on a daily basis (to control for their exposure to Mandarin Chinese language). After passing the eligibility screening, they were provided the information about the test and have given consent to take the test. There were 22 participants recruited through snowball sampling. They met the first author by individual appointment to participate in the study.

Recording of speech stimuli

Audiovisual recording of speech stimuli was performed in a quiet, well-lit room. A speaker was seated in front of a condenser microphone, a phone camera, a spotlight, and the researcher. The audio recording output had a sampling frequency of 41 kHz, and the video output was 1080HD with 30 fps. There were two speakers, one was a native speaker of Thai, and another was a native speaker of Mandarin Chinese. Each speaker read words from the list and was instructed to keep their eyes on the camera and to start uttering the word with their lips closed and finish with their lips closed.

Word list

Only stops and fricative consonants were included in the word list as testing phonemes. It is important to note that fricatives in Mandarin Chinese are distributed across different places of articulation, including where Thai lacks. In contrast, both languages share the same number of stops. These make Mandarin Chinese an effective foreign language test as it captures similarity and reflects differences in the phonology. Distractors were also added to the test to prevent participants from being able to guess what was being tested. They amounted to 25% of the trials in the study. The syllable structure of the stimuli is CV, comprising of one initial consonant and one vowel. Table 1 below presents words used in the test for both languages.

Language	**Manner**	Bilabial	Labiodental	Alveolar	Retroflex	Velar
Thai	**Stops**	[p^ha]	-	[t^ha]	-	[k^ha]
	Fricatives	-	[fa]	[sa]	-	-
	Distractors	[wa] [la]				
Mandarin Chinese	**Stops**	[p^ha]	-	[t^ha]	-	[k^ha]
	Fricatives	-	[fa]	[sa]	[ʂa]	[xa]
	Distractors	[la] [ma]				

Table 1 Testing and distractor word list in Thai and Mandarin

Stimuli composition and manipulation

Recorded words were made into testing trials and distractor trials. Each trial starts with a 500 milliseconds of fixation point in the center of the media frame to guide the participants to where the stimuli will be presented. The stimuli are played for the maximum duration of 1500 milliseconds after the fixation point disappears. Participants have 3500 milliseconds for the response period. The last part of each trial is the inter-trial interval of 500 milliseconds. Figure 1 illustrates the composition of the trials in the test.

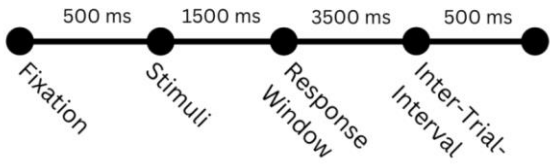


Figure 1 Composition of testing and distractor trials

Before manipulation, the auditory content of each stimulus was assured to have the same loudness. Audiovisual manipulation was adapted from [9] Hazan and Li (2008). Auditory manipulation was masking white noise of SNR -8 (noise being 8 dB louder than the speech - N1) and SNR -16 (noise being 16 dB louder than the speech - N2) to the original sound. Visual manipulation was masking the Gaussian blur of 25% (B1) and of 50% (B2) to the original video. Therefore, there are three levels for each degradation: Noise N0, N1, and N2, and Blur B0, B1, and B2. Figure 2 shows an example of a visual component under three blur conditions. Figure 3 features an example of spectrograms of speech sounds under three levels of noise conditions. Table 2 shows different degradation conditions in the test blocks.

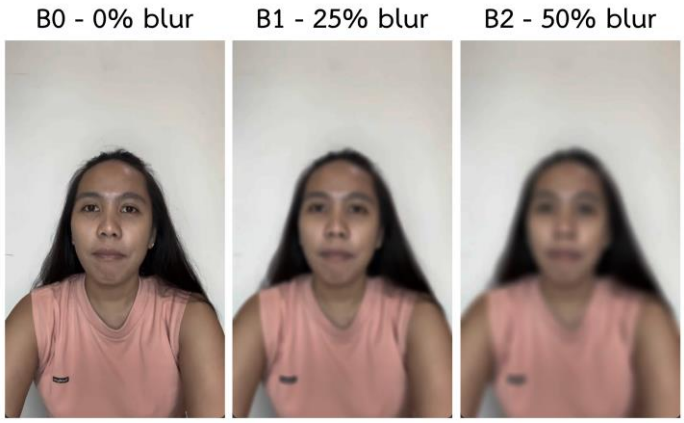


Figure 2 Example of 3 degradation levels in visual component V/p^{h}a/

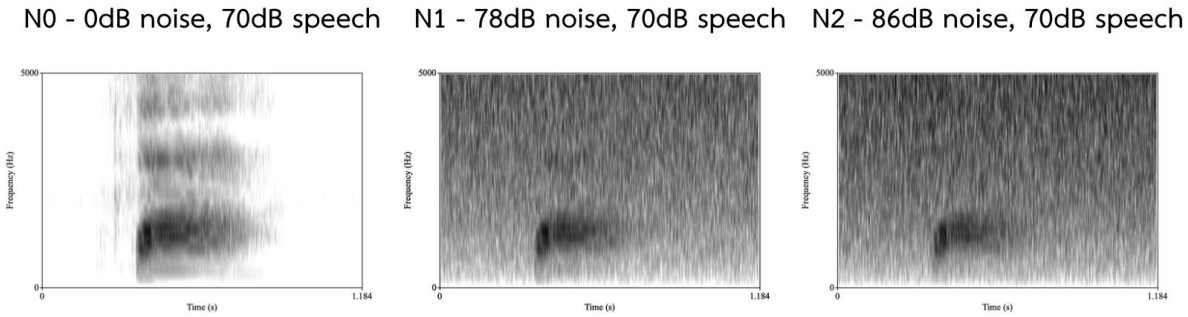


Figure 3 Example of spectrograms from 3 degradation levels in component Auditory /fa/

Degradations	Quiet (N0)	Low (N1)	High (N2)
Clear (B0)	0 SNR, 0% blur	-8 SNR, 0% blur	-16 SNR, 0% blur
Low (B1)	0 SNR, 25% blur	-8 SNR, 25% blur	-16 SNR, 25% blur
High (B2)	0 SNR, 50% blur	-8 SNR, 50% blur	-16 SNR, 50% blur

Table 2 Details on different audiovisual degradation conditions in trials in the tests

After manipulation, stimuli were re-checked to have their noise and blur in their respective conditions. Then the stimuli were incorporated into trials in the test. Audio-only trials were visually presented with a big black circle fixation point and auditorily presented with a speech component. Figure 4 illustrates the interface of the test trials from both audio-only and auditory test clocks. Congruent audiovisual trials were visually presented with a video component that matched the speech component. Incongruent audiovisual trials were visually presented with a video component that differed from the speech component in terms of labial features. There was a total of 147 trials (5 testing + 2 distractor phonemes: 63 congruent + 63 incongruent audiovisual trial + 21 audio-only trials) for the test in the Thai language and 189 trials (7 testing + 2 distractor phonemes: 81 congruent + 81 incongruent audiovisual trial + 27 audio-only trials) for the test in Mandarin Chinese language.

Test task

The test was developed in Psychopy Builder and run using Psychopy Runner [12] (Peirce et al., 2019), desktop applications commonly used in psycholinguistic studies. Speech identification test used in the study required the participants to look and listen to the trials, then identify the sound they had “heard”, according to the instructions, by clicking the answer among the options boxes provided. Options in each trial follow these criteria.

- same place of articulation, same manner of articulation (actual congruent sound)
- same place of articulation, different manner of articulation
- different place of articulation, different manner of articulation
- different place of articulation, same manner of articulation
- Other “อื่นๆ”– if options above are not what they have heard

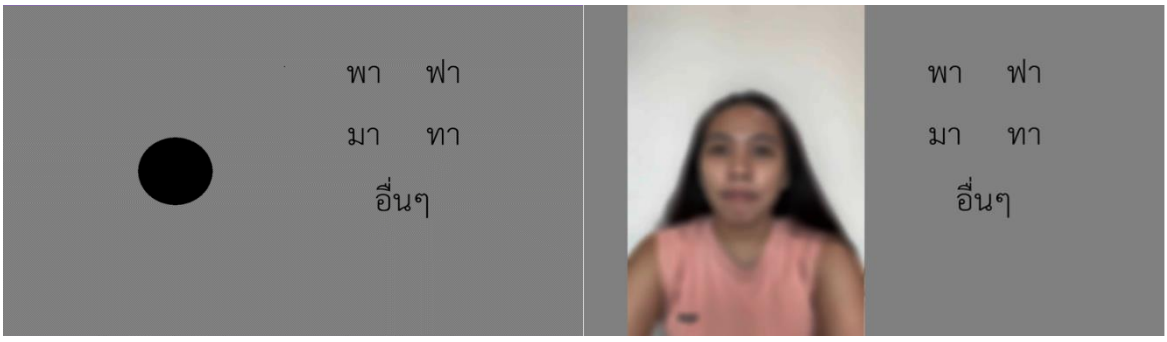


Figure 4 Interfaces of the test trial in the test blocks; Left – Audio-only, Right – Audiovisual

Orders of the test trials are randomized. Trials in the test, both testing and distractors, for the two language blocks follow the same structure mentioned earlier in **Stimuli composition and manipulation**.

Test Procedure

The test was conducted in a quiet room with the participant sitting facing the laptop running the test. They were instructed to pay attention to the laptop screen throughout the session. There were 4 main parts of the test: 1) Familiarization block, 2) Practice block, 3) Test blocks in Thai, and 4) Test blocks in Mandarin Chinese. In each part of the test, participants were accompanied by instructions and were told to carefully read the instructions before starting the block in each part. Only data from test blocks in the Thai language and in Mandarin Chinese language were analyzed. The total duration of the test was 40 minutes.

Statistical analysis

Two categories of experimental data were collected. First, perception accuracy was collected based on the response received from each participant in each test trial, and then the accuracy percentage was calculated. Second, reaction time only from a correct response was collected. This is to reflect only the successful perception. To satisfy the first objective in examining degrees of visual influence in speech perception, data on perception accuracy and reaction time from both language blocks were compared. The second objective is answered by comparing the experimental data across the three noise conditions and the three blur conditions. These comparisons were tested with an appropriate inferential statistical test, depending on statistical assumptions about their distribution and homogeneity of variance. The statistical analyses were run on RStudio. Statistical tests were performed and validated by the authors.

Results

Median, interquartile range, quartile 1 and quartile 3 of the perception accuracy percentage and reaction time are reported in Table 3. Key features of the finding include 1) lower median of accuracy percentage from congruent audiovisual trials than audio-only in native language, and 2) higher median of accuracy percentage from congruent audiovisual trials than audio-only in the foreign language.

TESTLANG	MODALITY	median	IQR	Q1	Q3
NATIVE	AO	85.00	8.52	80.95	89.47
NATIVE	AVc	44.26	27.88	28.57	56.45
NATIVE	AVi	46.77	30.48	27.59	58.06
FOREIGN	AO	52.00	15.75	46.15	61.90
FOREIGN	AVc	65.00	7.93	60.49	68.42
FOREIGN	AVi	37.66	16.13	26.92	43.06

Table 3 *Median, interquartile range, quartile 1, and quartile 3 of the accuracy percentage and reaction time, grouped by stimulus language and test modalities*

To examine the effect of foreign language on degrees of visual influence, accuracy from three modalities (audio-only, congruent audiovisual, and incongruent audiovisual) of the test was calculated according to the equation from [15] Sekiyama & Burnham (2008). It is worth noting that the distribution of data is not normal and the number of observations is limited; therefore non-parametric Wilcoxon Rank Sum test was used. Comparison between degrees of visual influence from test blocks in Thai and Mandarin Chinese (N =22) indicated that degrees of visual influence in native speakers of Thai were significantly lower when listening to speech in Thai (Mdn = -0.84, IQR = 7.51) than in Mandarin Chinese (Mdn = 25.3, IQR = 11.3), $p < .001$, $r = 0.87$.

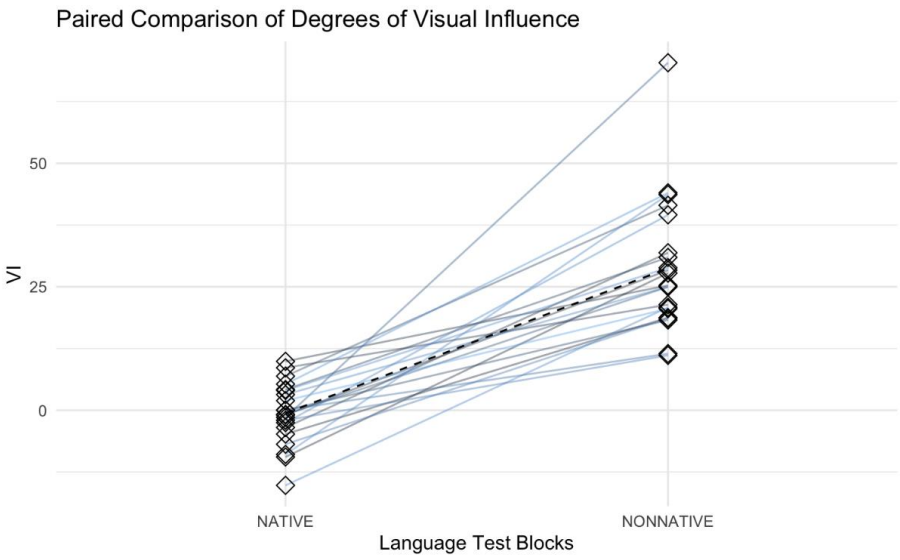


Figure 5 Paired comparison of degrees of visual influence between native speakers of Thai listening to Thai (NATIVE) and Mandarin (NONNATIVE)

This means that when listening to speech in Thai language, participants relied on their eyes significantly less than when they listened to speech in Mandarin Chinese. Additional analysis on how fast participants gave a correct response indicated that participants reacted to speech in Thai faster than in Mandarin Chinese. However, the difference is significant in audiovisual incongruent trials ($p = .011$). Despite a marginal difference ($p=.09$) being found for audio-only and a non-significant difference ($p = .48$) from congruent audiovisual trials, average reaction time in Thai language blocks is lower than in Mandarin Chinese language blocks, lending support to the prediction in the foreign/native language hypothesis.

To examine the effect of noise and blur on perception accuracy and reaction time from trials in three modalities across both languages, the Kruskal-Wallis test was administered due to the non-normal distribution of the data.

Effect of Noise on speech perception

Prediction on the effect of noise is that higher levels of noise are associated with lower perception accuracy and longer reaction time. According to the data from Thai language blocks, noise was found to

contribute to significantly lower perception accuracy ($p < .001$) in audio-only, congruent audiovisual, and incongruent audiovisual modalities. The difference was more subtle for reaction time because only the audio-only condition presented with significant difference ($p = .043$), while it was non-significant in the other two audiovisual conditions. When native speakers of Thai were presented with trials in Mandarin Chinese, noise had a significant effect ($p < .001$) on perception accuracy in all modalities. Reaction time was also found to be significantly affected by noise only in congruent audiovisual trials ($p < .001$).

Effect of Blur on speech perception

Prediction on the effect of blur is that the blurrier the visual information, the lower the accuracy percentage and the longer the reaction time it is. Data indicated no main effect of blur for accuracy in most combinations of modality and language block. The only significant difference was found in congruent audiovisual trials for Mandarin Chinese language blocks ($p = .021$). There was also no main effect of blur on reaction time, meaning that participants spent a similar amount of time no matter how blurry the visual cue was.

Discussion

On the issue of foreign language, the results indicated that when Thai participants listened to speech in Thai, they depended on visual information significantly less than when they listened to speech in Mandarin Chinese. The results are in line with other studies found in speakers of different languages [6] (Chen & Hazan, 2009). The explanation behind this can be behavioral, as when auditory information does not match what the participant has in their understanding, they resort to visual information, which is the only other available information. The results add an unprecedented finding in the area of speech perception of native Thai listeners and contribute the evidence for increased degrees of visual influence in a foreign speech to the existing literature. In this particular study, the behavior is constant across 22 participants, although not statistically powerful enough to account for the entire population of native speakers of Thai. In conclusion, the evidence supports a heightened degree of visual influence when listening to foreign speech, consistent to the previous literature.

On the issue of audiovisual degradations, white noise worsened perception accuracy in all modalities across both languages. This supports the prediction, and the results are in line with other studies [3, 8, 15] (Broersma & Scharenborg, 2010; Hazan et al., 2010; Sekiyama & Burnham, 2008).

While findings on noise aligned with the prediction for perception accuracy, findings on blur only did so in a very limited aspect. Although there is existing evidence of visual blur increasing reliance on auditory cues [7] (Hazan & Li, 2008), and the rate of the McGurk effect [16] (Sekiyama & Tohkura, 1991), there is no such strong evidence from this study. There are a few explanations for why blur does not cause a significant difference. First, it is very likely that Thai-speaking participants were relying heavily on auditory information, and it was informative enough to decide without additional help from visual information. Large significance values ($p = .517$ for native, $p = .68$ for foreign) from Thai language and Mandarin Chinese language blocks imply that visual information of any blur does not affect how accurate the perception was. Second, it could be due to very short visual input of lip movement. Studies in the past used VCV syllable

structure (such as “aba” or “aga”), while this study used CV syllable structure. This limited visual cue was probably deemed not to significantly contribute to the perception, thus not making a significant difference, regardless of how clear or how blurry that visual cue is.

The lack of statistical significance in reaction time may be due to the nature of the test. The task may be simple and repetitive enough for participants to become familiar with and ultimately spend a similar amount of time on the trials. Another explanation is that they were instructed to answer as fast as they could; as a result, the reaction time difference became very similar across testing conditions.

It is also worth highlighting that the only significant difference in reaction time by blur was found in congruent audiovisual trials in Mandarin Chinese ($p = .021$). This means that the blurrier the visual cues are, the longer the time it takes to respond. The explanation behind this could be the acoustic difference of Mandarin Chinese speech sounds. Hearing something similar to a native sound but with a slight difference can trigger the listener to have the native sound representation, but since it is not ideal enough, it could take a longer time to decide on what sound is before finally thinking that it is a sound in their native language, according to the Perceptual Assimilation Model [1] (Best, 1995). This explanation is sensible when considering that no significant difference in reaction time was found for stimuli in native language blocks, as all speech stimuli are Thai.

Conclusion

This study investigates the role of eyes in speech perception across native and foreign languages. In line with the first objective, aiming to investigate the effect of native and foreign languages on degrees of visual influence during speech perception, the results show that native speakers of Thai rely more on visual information when perceiving foreign speech (Mandarin Chinese) than their native language. This is reflected in significantly higher degrees of visual influence when listening to Mandarin Chinese. Addressing the second objective, which focuses on the effect of white noise and Gaussian blur on perception accuracy and reaction time, the findings reveal that noise makes a significant difference in perception accuracy in both languages, while blur does not. For reaction time, the difference was not significant, possibly due to the nature of the participants. Analysis of blur does not indicate a significant difference in perception accuracy or reaction time. It shows that low variability in reaction time from both language blocks aligns with lower reliance on visual information during speech perception. Moreover, the acoustic similarity of a foreign language gives support to why the reaction time is only significant when listening to speech in Mandarin Chinese.

Considering the findings from this study, this means that the speech perception of Thai listeners is affected by foreign speech in the same manner as speakers of other major languages in past studies. This study also contributes additional evidence that auditory degradation has a tremendous effect on perception accuracy, but visual degradation might not play as important a role in audiovisual speech perception.

Recommendation for Further Studies

1. Given a small number of participants in the study, the results could not account for the entire population of native speakers of Thai. A larger scale of study, including equally collecting data in participants with a balanced demographic background, is needed.
2. This study uses a single syllable in a trial, which partly reflects the nature of human speech perception. More sophisticated materials that involve more semantic content can better examine the issue as they better reflect the nature of conversation. Moreover, speech babble or ambient noise can be used to simulate a more realistic listening environment.

List of References

[1] Best, C. (1995). A direct realist view of cross-language speech perception. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 171–204). York Press.

[2] Bidelman, G. M., Bush, L. C., & Boudreaux, A. M. (2020). Effects of noise on the behavioral and neural categorization of speech. Frontiers in Neuroscience, 14, 153. https://doi.org/10.3389/fnins.2020.00153

[3] Broersma, M., & Scharenborg, O. (2010). Native and non-native listeners' perception of English consonants in different types of noise. *Speech Communication*, *52*(11), 980-995. https://doi.org/10.1016/j.specom.2010.08.010

[4] Bsharat-Maalouf, D., & Karawani, H. (2022). Bilinguals' speech perception in noise: Perceptual and neural associations. *PLOS ONE*, *17*(2), e0264282. https://doi.org/10.1371/journal.pone.0264282

[5] Burnham, D., & Dodd, B. (2018). Language-general auditory-visual speech perception: Thai-English and Japanese-English McGurk effects. *Multisensory Research*, *31*(1–2), 79–110. https://doi.org/10.1163/22134808-00002590

[6] Chen, Y., & Hazan, V. (2009). Developmental factors and the non-native speaker effect in auditory-visual speech perception. *The Journal of the Acoustical Society of America*, *126*, 858-865. https://doi.org/10.1121/1.3158823

[7] Davis, C., & Kim, J. (2006). Audio-visual speech perception off the top of the head. *Cognition*, *100*(3), B21-31. https://doi.org/10.1016/j.cognition.2005.09.002

[8] Hazan, V., Kim, J., & Chen, Y. (2010). Audiovisual perception in adverse conditions: Language, speaker and listener effects. *Speech Communication*, *52*(11), 996-1009. https://doi.org/10.1016/j.specom.2010.05.003

[9] Hazan, V., & Li, E. (2008, September). *The effect of auditory and visual degradation on audiovisual perception of native and non-native speakers* [Paper presentation]. INTERSPEECH 2008, Brisbane, Australia.

[10] Lalonde, K., & Werner, L. A. (2019). Infants and Adults Use Visual Cues to Improve Detection and Discrimination of Speech in Noise. *J Speech Lang Hear Res*, *62*(10), 3860-3875. https://doi.org/10.1044/2019_jslhr-h-19-0106

[11] McGurk, H., & Macdonald, J. (1976). Hearing lips and seeing voices. *Nature*, *264*(5588), 746-748. https://doi.org/10.1038/264746a0

[12] Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., Kastman, E., & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*(1), 195-203. https://doi.org/10.3758/s13428-018-01193-y

[13] Schwarz, J., Li, K. K., Sim, J. H., Zhang, Y., Buchanan-Worster, E., Post, B., Gibson, J. L., & McDougall, K. (2022). Semantic cues modulate children's and adults' processing of audio-visual face mask speech. *Frontiers in Psychology*, *13*, 879156. https://doi.org/10.3389/fpsyg.2022.879156

[14] Sekiyama, K. (1997). Cultural and linguistic factors in audiovisual speech processing: The McGurk effect in Chinese subjects. *Perception & Psychophysics*, *59*(1), 73-80. https://doi.org/10.3758/BF03206849

[15] Sekiyama, K., & Burnham, D. (2008). Impact of language on development of auditory-visual speech perception. *Dev Sci*, *11*(2), 306-320. https://doi.org/10.1111/j.1467-7687.2008.00677.x

[16] Sekiyama, K., & Tohkura, Y. (1991). McGurk effect in non-English listeners: few visual effects for Japanese subjects hearing Japanese syllables of high auditory intelligibility. *J Acoust Soc Am*, *90*(4 Pt 1), 1797-1805. https://doi.org/10.1121/1.401660

[17] Sekiyama, K., & Tohkura, Y. i. (1993). Inter-language differences in the influence of visual cues in speech perception. *Journal of Phonetics*, *21*(4), 427-444. https://doi.org/10.1016/S0095-4470(19)30229-3