Speech Prosody 2024
2-5 July 2024, Leiden, The Netherlands

The Role of Auditory and Visual Modality in Perception of English Statements and Echoic Question by Chinese EFL Learners

Shanpeng Li[1], *Yinuo Wang*[2], Shifeng Xia[1], Zhiqiang *Tang*[3], *Ping Tang*[1], *Yan Feng*[1]

[1] School of Foreign Studies, Nanjing University of Science and Technology, China
[2] Faculty of Arts & Social Science, National University of Singapore, Singapore
[3]College of Liberal Arts, Anhui University, China

{shanpeng; shifeng_xia; ping.tang; yanny.feng}@njust.edu.cn, E1289655@u.nus.edu
17031@ahu.edu.cn

Abstract

Previous research underscored the role of auditory and visual cues in perceiving statements and questions, yet with conflicting conclusions regarding their relative significance. It was argued that English speakers relying predominantly on auditory cues, with limited impact from visual cues for intonation. Given the variability observed in language-specific utilization of auditory and visual modalities for interpreting statements and questions, as evidenced in studies involving Dutch and Catalan, the generalizability of findings to other languages remains uncertain. Therefore, the study aimed to investigate the influence of auditory and visual cues on the comprehension of English statements and questions among Chinese EFL learners.

A total of 56 Chinese EFL learners participated in the audiovisual perception study, categorized into three blocks: audio-only (AO), visual-only (VO), and audiovisual (AV) conditions. Additionally, to explore the contribution of specific facial areas, the VO and AV conditions were subdivided into full-face, upper-face only, and lower-face only conditions.

The result revealed that Chinese EFL learners lean towards visual cues, particularly upper face when perceiving English intonations. This inclination could be attributed to factors such as whispered condition and cultural inclinations. Recognizing and incorporating these influences into teaching approaches can significantly enhance the comprehension of intonation among EFL learners.

Index Terms: Intonation comprehension, auditory and visual modality, upper and lower facial actions, Mandarin EFL learners

1. Introduction

Understanding speech involves a complex interplay of perceptual, cognitive, and linguistic abilities [1]. However, this process can face significant challenges when acoustic signals are compromised. Recent studies have demonstrated that communicative functions signalled by acoustic prosody tend to be supported by visual information as well in the form of specific facial expressions[2, 3] The correlation between gestures and speech has been explored in various ways, with differing hypotheses. Some studies posit that gestures are redundant, merely expressing information already evident in the verbal content. Conversely, an alternative hypothesis suggests a trade-off relationship between gestures and speech production, wherein speech and gestures complement each other.

Previous studies in this area of research focused on the recognition of individual segmental sounds, but there is a growing awareness that visual cues may also be used to identification of the statements versus questions [4-7], despite occasionally yielding conflicting result. Most of previous found that even the visual cues can help to distinguish the statements and questions, the auditory modality still contributes the most. For example, Srinivasan and Massaro (2003) conducted a series of five perception experiments aiming to explore the role of auditory and visual cues in differentiating between statements and echo questions in English [8]. Their findings suggested that both auditory and visual cues play a part in this identification, but the influence of auditory cues appeared to be more prominent than visual cues. This trend echoes findings in other languages such as Swedish [6], Brazilian Portuguese [5, 9], indicating a consistent reliance on auditory cues across languages. Their conclusions aligned with the Fuzzy Logical Model of Perception (FLMP) [10, 11], which posits that both auditory and visual modalities contribute to prosodic comprehension. Moreover, the model predicts that when information from one modality is ambiguous, the influence of the other modality becomes more pronounced.

However, an alternative viewpoint is held by other studies. Borràs-Comes and Prieto (2011) examined the audiovisual recognition of echo questions and statements featuring contrastive focus [4]. The study emphasized that for Catalan listeners, the visual cues exerted a greater influence than auditory cues in their decision-making process. The contradictory conclusions drawn from previous studies may be attributed to two potential reasons. Firstly, language-specific variations could play a role. In a subsequent study conducted by Crespo Sendra et al. (2013), a cross-linguistic investigation was carried out involving Catalan and Dutch, two languages employing distinct prosodic strategies to delineate the contrast between these two types of interrogatives. The findings revealed disparities in the perceptual processing of these sentences between Dutch and Catalan listeners. Dutch participants exhibit a higher dependence on auditory cues, whereas Catalan participants rely more extensively on visual facial expression cues [12], providing evidence for the languages specific effects on comprehension and highlighting varying reliance on auditory and visual modalities in different linguistic groups.

Another possible explanation for these discrepancies might stem from the distinctions arising between quiet and degraded environments. Past studies have indicated that in quiet settings, auditory cues tend to take precedence, potentially resulting in a ceiling effect [9]. This effect limits the ability to accurately

10.21437/SpeechProsody.2024-150

gauge the advantages of integrating visual cues, hindering the assessment of their potential benefits in comprehension. One proposed solution is to add noise into auditory channel, like [9]. However, adding noise into a signal reduces its lexical intelligibility without significantly altering the F0 contour, which is crucial for perceiving auditory cues. Miller and Nicely (1955) demonstrated that voicing was the most robust speech feature in noise [13]. Consequently, using a speech-in-noise paradigm to degrade auditory perception of prosodic features may not be efficient. Therefore, in this study, whispered speech was chosen as the degradation technique, since it naturally lacks intonational information (F0) due to the absence of vocal fold vibration.

However, despite the existing research on auditory and visual cues in the comprehension of statements and questions, there is still lack knowledge on how these cues interact with each other, as well as whether the findings from studies conducted on Dutch and Catalan can also be generalized to other languages. Therefore, the primary objective of this study is to undertake a cross-linguistic investigation focusing on English and Chinese, two widely spoken languages in terms of both usage and population.

In the context of audiovisual perception, it is argued that Chinese participants might exhibit a reduced reliance on visual cues using McGurk paradigm, as noted in some previous studies like [14, 15]. This expectation is based on two primary factors. Firstly, the use of tonal language in Chinese may result in a heightened auditory bias, as tone distinctions are primarily conveyed through acoustic properties [16] and second, Chinese culture has a tendency of avoiding direct eye contact, which suggests a decreased dependence on visual information [14].

Based on the findings of Srinivasan and Massaro (2003), who identified a domain-specific auditory cue in distinguishing English statements from questions, our study aims to investigate whether the Chinese English as a Foreign Language (EFL) learners depend on visual facial cues during the comprehension of English statements and questions. Furthermore, we aim to investigate potential disparities in the utilization of different facial areas in this cognitive process. Building upon their research, we hypothesize that Chinese EFL learners will exhibit a less reliance on visual cues in their understanding of English utterances.

2. Method

2.1. Participants

A total of 56 Chinese English learners (31 females and 25 males) from Nanjing University of Science and Technology participated in the current perception experiment. Their ages ranged from 17.9 to 25.5 years, with a mean age of 21.7 years. All participants were native Chinese speakers and had been studying English for a minimum of ten years. None of participants reported any hearing or visual impairment.

To minimize the potential impact of variations in English proficiency among participants, Participants were required to have obtained a score between 520 and 590 on the College English Test Band-6 (CET-6), which is commonly used in China to assess English proficiency. This ensured that all participants possessed a relatively consistent level of English language proficiency. All participants were provided with written informed consent forms and were remunerated for their participation.

2.2. Materials

2.2.1. Target sentence design

A total of 100 English sentence was designed as the experimental stimuli for this study. All stimuli were intentionally designed to be literally neutral in terms of emotions or intonation biases. The words involved in the sentences were chosen to be common and familiar in English, ensuring that they would not introduce any ambiguity or comprehension difficulties.

Each target sentence weas designed to be conveyed in two different intonations, namely statements and echoic questions. For instance, an example pair of target sentences would be “He has been to England.” (statement) and “He has been to England?” (echoic question). To ensure that the different intonations of the target sentences could be produced in a natural and clear way by speakers, the Discourse Completion Task paradigm [5] was employed to design declarative-bias scenarios and interrogative-bias scenarios. These scenarios provided a more natural context for triggering the target sentence. Below are examples illustrating the different scenarios for the aforementioned target sentence:

(1) Declarative-bias scenario: There is a new student in your class, and the teacher introduces him to you: *He has been to England*.

(2) Interrogative-bias scenario: There is a new student in your class, and you ask your friend: *He has been to England?*

2.2.2. Audiovisual Recordings

A female English native speaker, 23 years old, was recruited as the speaker. She was born and raised in London within an English-speaking family. Stimuli were recorded in a professional recording studio at the University of Westminster. A Canon 5D4 video camera was used, coupled with a Philips DLM3541C wireless microphone to enhance high-quality audio track recording. A mini teleprompter was placed on top of the camera, allowing the speaker to view the text materials while recording. During the recording session, the distance between the speaker and the video camera was consistently maintained at approximately 130 cm. The video camera was capable of simultaneously capturing both the auditory signal of the speaker’s voice and the visual signal of her head and upper body movements.

During the recording process, each slide displayed on the teleprompter contained a pair of corresponding statement and question utterance, along with the scenario in which the utterances were suited. The speaker looked through the slides to understand the accompanying scenario and read out the bold target sentences colored in red. To ensure clarity and avoid interference between the production of statements and questions, the speaker was instructed to pause for approximately three seconds after read each statement before continuing to read the corresponding question.

To prevent a potential ceiling effect in the audio-only condition, which has been observed in previous studies, the present study implemented a degraded audio condition by instructing the speaker to articulate the target sentences in a whispered manner. This technique ensured that their vocal cords did not vibrate during the production of the utterances. Prior to the video recording session, the speaker was specifically guided to deliver the target sentences in a whispering manner. They were advised to imagine themselves

engaged in conversation within a library setting where speaking loudly is prohibited. This instruction aimed to encourage a controlled and soft-spoken delivery of the target sentences.

2.2.3. Audiovisual Stimuli Process

Only whispered recordings were set as the intonational perception experiment material. Firstly, these experimental recordings were processed using Adobe Premiere Pro 2018 and exported into 400 (100 utterances × 2 intonations × 2 languages) isolated clips covering each trial. These trails were MPEG-4 encoded, output with a 1024*960 pixels resolution, and stored in AVI format. The audio of the experimental materials was sampled at 44.1 kHz, 16-bit, in dual channel. Secondly, Praat was used to extract the soundtrack of each video clip which was restored as an independent audio file in WAV format. Finally, these video clips were re-edited as audio-only files without the video track and visual-only files without the audio track using the self-written program. In this way, the materials in the present study were separated into three modalities, i.e., audio-only (AO), visual-only (VO), and audio-visual combined (AV).

To explore the relative significance of different facial areas in intonation comprehension, the visual-only and audiovisual conditions were further modified into two versions. In the first version, only the upper part of the face was visible (referred to hereafter as upper-face-only), while the lower portion of the face (referred to hereafter as lower-face-only), including approximately the middle of the nose, was blackened out. Conversely, in the second version, the area above the nose was blackened out, while the mouth area remained visible. Examples of these two modified versions of the original videos can be seen in Figure 1.

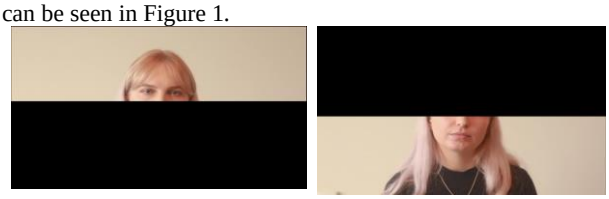

Figure 1: *Examples of the two facial conditions: upper- face only and lower-face only.*

2.3. Experiment Procedures

A perceptual experiment was conducted using PsychoPy 2021.2.3 [17], comprising three blocks: auditory-only (AO), visual-only (VO), and audiovisual (AV). Within the VO and AV modalities, the video stimuli encompassed three versions: original (full-face version), an upper-face-only version, and a lower-face-only version. To minimize the possible confounding effect arising from modalities and sentence selection, a Latin square design was employed, evenly distributing stimuli across 14 lists. The AO and VO blocks were consistently presented before the AV block to mitigate any learning effects. The sequence of presenting the AO and VO blocks varied among participants, resulting in two possible experimental orders: AO-VO-AV or VO-AO-AV. To ensure variability and minimize bias, trial sequences within each block were randomized for each participant.

Each participant was directed to position themselves in front of a desktop computer, utilizing headphones. Preceding the experiment, the instructor delivered a concise verbal briefing on the identification task. Following this instruction, participants engaged in a practice session comprising 12 trials (4 trials for each modality) to acquaint themselves with the experiment's guidelines and procedures. Stimuli were presented on the computer screen, prompting participants to discern the intonation of the specified utterance based on perceived auditory, visual, or audiovisual cues from the stimuli. Participants used a keyboard to record their responses for each trial, selecting either the "F" key for a statement or the "J" key for a question. Each trial allowed a reaction time of 5 seconds; failure to respond within this time resulted in the skipping of that particular trial, with no recorded response. After the practice session, participants had the opportunity to seek clarification on the experiment's procedures from the instructor or proceed directly to the formal experiment, contingent upon their comprehension of the task.

The perception experiment took place in a quiet computer room at Nanjing University of Science and Technology, with an ambient noise level of less than 40 dB. The entirety of the experimental procedure lasted approximately 10 minutes, conducted seamlessly without any intervals or breaks.

2.4. Data Analysis

Participants' responses (categorized as either question or statement) in the intonation identification task were classified as correct (coded as 1) or incorrect (coded as 0). These responses underwent analysis through mixed-effect binomial logistic regression conducted using Jamovi [18]. Two regression models were constructed: The first regression model incorporated fixed effects analysis involving modality (AO, VO and AV) and intonation (statement and question) to assess the influence of modality. The second regression model included fixed effects analysis of Modality (VO and AV), facial condition (upper-face-only and lower-face-only) and intonation to ascertain the impact of specific facial areas. Both regression models employed random intercepts for sentence and participant as the designated random effects.

3. Result

The first mixed-effects logistic model showed a significant interaction effect between modality and intonation ($\chi^2 = 45.7$, $p < 0.001$), signifying the influence of modality on the identification of English statements and questions. Post hoc analysis using Bonferroni-correction indicated a significantly higher identification of questions in the VO modality (85.9%) compared to AV (62.6%, Odds Ratio, OR = 3.633, SE = 1.042, z = 4.498, $p < 0.001$) and AO (13.4%, OR = 0.025, SE = 0.009, z = 10.366, $p < 0.001$) modalities. Furthermore, a significant distinction was observed between AO and AV (OR = 0.092, SE = 0.027, z = 8.028, $p < 0.001$). However, no significant differences were found in the identification of statements between AO (91.9%), VO (93.4%), and AV (97.4%), as depicted in Figure 2.

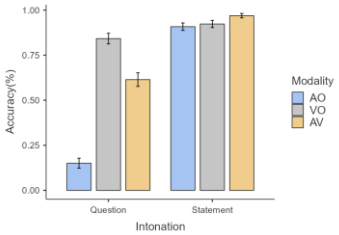

Figure 2: *The bar plot with 95% for accuracy of statement and question in three modalities*

The second mixed-effects logistic model showed a significant interaction effect between facial conditions and intonation (χ^2 = 55.01, p <0.001). However, the three-way interaction effect among modality, facial condition, and intonation was not significant (p = 0.543). This result suggests that both upper and lower face components influence the identification of English statements and questions, irrespective of the presence of an audio track in VO and AV conditions.

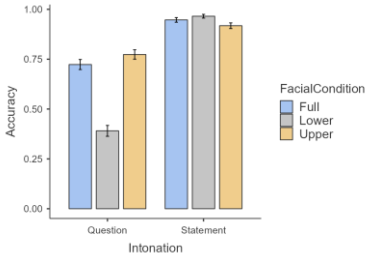

Figure 3: *The bar plot with 95% for accuracy of statement and question in full-face, upper-face only, and lower-face only condition.*

The Bonferroni-corrected post hoc analysis examining the interaction between facial conditions and intonation highlighted that the identification rate of questions in the upper-face-only condition (81.7%) was significantly higher than in the lower-face-only condition (37.6%, OR = 0.135, SE = 0.027, z = 9.89, p < 0.001), as shown in Figure 3. Yet, the difference between the full-face (76.2%) and upper-face-only conditions was not significant (p = 0.999). However, there were no significant differences in the identification of statements among the full-face (89.1%), upper-face-only (86.2%), and lower-face-only conditions (80%).

4. Discussion

The current study explored how auditory and visual modality influence the comprehension of English statements and questions among Chinese EFL learners. Specifically, the study aimed to determine whether facial expressions conveyed through visual modality assist L2 learners in perceiving English intonations. Furthermore, the research sought to identify which specific facial areas contribute more significantly to this comprehension.

Regarding the audio-visual comprehension of English intonation, previous reports indicated that English native speakers predominantly rely on the auditory modality. Meanwhile, visual cues like eyebrow movement or slow vertical head tilting were not strong indicators of question intonation [19]. Similar results were reported in studies involving Swedish [6] and Brazilian Portuguese [5, 9]. However, to the best of our knowledge, no prior research has explored this issue specifically among L2 learners of English. Our findings represent the initial indication that Chinese L2 learners place a heavier reliance on the visual modality when perceiving English question intonations.

The potential reasons for the dominant influence of the visual modality may stem from two main factors. First, according to the FLMP model, both auditory and visual modalities will influence the comprehension of prosody, with the one modality will be greater to the extent when the other is ambiguous [11]. In the case of auditory comprehension of statements and questions, it tends to be highly accurate [19, 20] , leading to a ceiling effect. This ceiling effect signifies that it becomes nearly impossible to measure a potentially significant improvement when the visual modality is introduced. One possible solution lies in designing degraded speech conditions involving noise or whispered speech. Therefore, the absence of fundamental frequency (F_0) information in whispered utterances, as in the current study, amplifies the reliance on the visual modality, especially as participants struggled to perceive questions solely through the auditory modality.

Additionally, the impact of foreign language effects could contribute to this outcome. Specifically, listeners have a tendency to rely more on visual cues when processing a non-native language as compared to their native language using McGurk paradigm [21, 22]. These enhancing effects have been observed across various languages, encompassing Japanese, American, Chinese, Spanish, and German listeners when perceiving foreign languages [23-25].

In regards to the second question regarding the role of specific facial areas, the findings indicate a preference for actions in the upper face, such as eye and eyebrow movements. A similar result was also observed in a previous study [26], which revealed that participants devoted a significantly greater amount of time to looking at the upper part of the face (i.e., eyes and forehead) and made more gazes in that area when making judgments about intonation patterns (statement vs. question). Visual attention toward upper face regions may be related to events such as wrinkling of the forehead, raising of the eyebrows, or eye-widening when intonation production. The visual attention directed towards the upper face regions may be attributed to events such as forehead wrinkling, eyebrow raising, or widening of the eyes during intonation production. It is worth noting that English native speakers tend to employ significant eyebrow raises and head tilts in conjunction with question intonation, but little or no eyebrow raise and negligible head movement when using statement intonation [26].

Another potential factor contributing to this finding could be cultural differences. Research demonstrates that individuals from diverse cultures process human facial information differently [27]. For example, studies suggest that Eastern participants, such as the Japanese, tend to prioritize cues from the eyes, whereas Western participants, such as Americans, place greater emphasis on cues presented by the mouth when evaluating emotions [28].

In conclusion, this study sheds light on the nuanced interplay between auditory and visual cues in English intonation comprehension among Chinese EFL learners. Understanding the reliance on visual modalities and specific facial cues enriches language pedagogy, facilitating more effective methods to enhance intonation comprehension for second language learners.

5. Conclusions

The current study explored the audiovisual perception by examining the respective impacts of auditory and visual cues on the comprehension of English statements and questions among Chinese EFL learners. Additionally, it aimed to elucidate the significance of specific facial regions in this process. The findings from the perception analysis highlighted the predominant influence of visual cues over auditory cues in the interpretation of question intonation. Specifically, there was a substantial reliance on upper facial expressions—such as eye and eyebrow movements—when discerning English question intonations from statements. This finding underscores the significance of integrating visual cues, notably facial expressions, into speech prosody.

6. Acknowledgements

This research was supported by the Ministry of Education of Humanities and Social Science Project (No. 23YJC740012).

7. References

[1] K. Lalonde and R. W. McCreery, "Audiovisual Enhancement of Speech Perception in Noise by School-Age Children Who Are Hard of Hearing," *Ear & Hearing*, vol. 41, pp. 705-719, 2020-1-31 2020.

[2] E. Krahmer and M. Swerts, "How children and adults produce and perceive uncertainty in audiovisual speech," Lang Speech, vol. 48, pp. 29-53, 2005-1-20 2005.

[3] C. Dijkstra, E. Krahmer and M. Swerts, "Manipulating uncertainty: the contribution of different audiovisual prosodic cues to the perception of confidence," in Proceedings of the Third International Conference on Speech Prosody Dresden, 2006.

[4] J. Borràs-Comes and P. Prieto, "'Seeing tunes.' The role of visual gestures in tune interpretation," Laboratory Phonology, vol. 2, 2011.

[5] M. Cruz, M. Swerts and S. Frota, "The role of intonation and visual cues in the perception of sentence types: Evidence from European Portuguese varieties," Laboratory Phonology: Journal of the Association for Laboratory Phonology, vol. 8, p. 23, 2017-9-11 2017.

[6] D. House, "Intonational and visual cues in the perception of interrogative mode in Swedish," in Proceedings of Seventh International Conference on Spoken Language Processing Denver, Colorado, USA, 2002, pp. 1957–1960.

[7] K. G. Nicholson, S. Baum, A. Kilgour, C. K. Koh, K. G. Munhall, and L. L. Cuddy, "Impaired processing of prosodic and musical patterns after right hemisphere damage," Brain and Cognition, vol. 52, pp. 382-389, 2003-1-1 2003.

[8] R. J. Srinivasan and D. W. Massaro, "Perceiving Prosody from the Face and Voice: Distinguishing Statements from Echoic Questions in English," Language and speech, vol. 46, pp. 1-22, 2003-1-1 2003.

[9] L. Miranda, M. Swerts, J. Moraes, and A. Rilliard, "The Role of the Auditory and Visual Modalities in the Perceptual Identification of Brazilian Portuguese Statements and Echo Questions," Language and Speech, vol. 64, pp. 3-23, 2020-1-20 2021.

[10] D. W. Massaro, "Testing between the TRACE model and the fuzzy logical model of speech perception," Cognitive Psychology, vol. 21, pp. 398-421, 1989-1-1 1989.

[11] D. W. Massaro and G. C. Oden, "Evaluation and Integration of Acoustic Features in Speech Perception," The Journal of the Acoustical Society of America, vol. 67, pp. 996-1013, 1980-1-1 1980.

[12] V. Crespo Sendra, C. Kaland, M. Swerts, and P. Prieto, "Perceiving incredulity: The role of intonation and facial gestures," Journal of Pragmatics, vol. 47, pp. 1-13, 2013-1-1 2013.

[13] G. A. Miller and P. E. Nicely, "An Analysis of Perceptual Confusions Among Some English Consonants," The Journal of the Acoustical Society of America, vol. 27, pp. 338-352, 1955-1-1 1955.

[14] K. Sekiyama, "Cultural and linguistic factors in audiovisual speech processing: the McGurk effect in Chinese subjects," Percept Psychophys, vol. 59, pp. 73-80, 1997-1-1 1997.

[15] Y. Hayashi and K. Sekiyama, "Native-foreign language effect in the McGurk effect: a test with Chinese and Japanese," in Auditory-Visual Speech Processing (AVSP' 98) Sydney, Australia, 1998.

[16] K. Sekiyama and D. Burnham, "Impact of language on development of auditory-visual speech perception," Dev Sci, vol. 11, pp. 306-20, 2008-3-1 2008.

[17] J. Peirce, J. R. Gray, S. Simpson, M. MacAskill, R. Hochenberger, H. Sogo, E. Kastman, and J. K. Lindelov, "PsychoPy2: Experiments in behavior made easy," Behav Res Methods, vol. 51, pp. 195-203, 2019-2-1 2019.

[18] The-jamovi-project, "jamovi,", 2.3.0 ed, 2022.

[19] R. J. Srinivasan and D. W. Massaro, "Perceiving prosody from the face and voice: distinguishing statements from echoic questions in English," Lang Speech, vol. 46, pp. 1-22, 2003-3-1 2003.

[20] P. Lieberman, "Intonation, perception, and language.,". vol. Doctoral dissertation: Massachusetts Institute of Technology, 1966.

[21] D. M. Hardison, "Bimodal Speech Perception by Native and Nonnative Speakers of English: Factors Influencing the McGurk Effect," Language Learning, vol. 46, pp. 3-73, 1996.

[22] Y. Zhang, X. Chen, S. Chen, Y. Meng, and A. Lee, "Visual-auditory perception of prosodic focus in Japanese by native and non-native speakers," Frontiers in Human Neuroscience, vol. 17, 2023-9-21 2023.

[23] K. Sekiyama and D. Burnham, "Impact of language on development of auditory-visual speech perception," Developmental science, vol. 11, pp. 306-320, 2008-1-1 2008.

[24] Y. Chen and V. Hazan, "Developmental factors and the non-native speaker effect in auditory-visual speech perception," The Journal of the Acoustical Society of America, vol. 126, pp. 858-865, 2009-8-1 2009.

[25] Y. Hayashi and K. Sekiyama, "Native-foreign language effect in the McGurk effect: a test with Chinese and Japanese," in Auditory-Visual speech procesing Terrigal-Sydndy, Australia, 1998.

[26] C. R. Lansing and G. W. McConkie, "Attention to facial regions in segmental and prosodic visual speech perception tasks," J Speech Lang Hear Res, vol. 42, pp. 526-39, 1999-6-1 1999.

[27] C. Blais, R. E. Jack, C. Scheepers, D. Fiset, R. Caldara, and A. O. Holcombe, "Culture shapes how we look at faces," PloS one, vol. 3, p. e3022-e3022, 2008-1-1 2008.

[28] M. Yuki, W. W. Maddux and T. Masuda, "Are the windows to the soul the same in the East and West? Cultural differences in using the eyes and mouth as cues to recognize emotions in Japan and the United States," Journal of Experimental Social Psychology, vol. 43, pp. 303-311, 2007-1-1 2007.