Journal of language works

Vol. 8, No. 1, 2023

The Multi-Sensory Training Benefit in Second Language Acquisition

Ida Marie Thomsen Krarup
Aarhus University
202105383@post.au.dk / idamtkrarup@gmail.com

ABSTRACT

This article examines the role of multi-sensory input in the acquisition of non-native phonemes. First, it demonstrates that speech perception is not only auditory and that multi-sensory input such as audiovisual recordings of speech and tactile input from speech production aid the acquisition and comprehension of speech in one's native language. Then, it examines studies that investigate the effectiveness of multimodal training compared to unimodal training when it comes to the identification of non-native phonemes. Based on the evidence presented, I argue that multi-sensory training aids the acquisition of non-native phonemes in some cases. However, when the additional modality provides no additional information, there seems to be no training benefit. There might be further limitations on the multi-sensory training benefit, such as whether it differs across native languages and can occur with other modalities than the audiovisual, but further research is needed to say anything conclusively in this area.

1. INTRODUCTION

The study of speech perception has been described as the study of "how the listener converts a continuous acoustic signal into discrete linguistic units" (Blumstein & Cutler 2003). This definition carries the assumption that speech is primarily auditory which prevails in some of the research on speech perception, where the only modality investigated is audio. This is especially true for much of the research on the perception of non-native phonemes. However, research on cross-modal speech perception shows that senses such as sight and touch also affect speech perception and identification of phonemes. Therefore, it seems likely that these senses would influence the perception of non-native phonemes, which raises the question of whether multi-sensory input can aid the acquisition of non-native phonemes.

In this paper, I will review literature on cross-modal speech perception to examine whether training with multi-sensory stimuli is beneficial for the acquisition of non-native phonemes. First, I will demonstrate that speech is multimodal, and that there is a multimodal training benefit for perceiving and acquiring speech in one's native language. Then, I will argue that there is a multi-sensory training benefit for the perception of non-native phonemes, and lastly, I will examine the limitations of the multi-sensory training benefit as well as gaps in the literature.

2. CROSS-MODAL SPEECH PERCEPTION

The overarching idea behind cross-modal speech perception is that when we perceive speech, we do not just listen to the acoustic signal, we utilize any sensory information that is available to us. One of the most influential findings that supports this is the McGurk effect. It is an illusion created by presenting listeners with incongruent auditory and visual stimuli, e.g., playing the audio of a speaker saying 'ga' and simultaneously playing a video of the speaker saying 'ba'. When the native English speakers were asked what they heard, 92% of the tested adults said they heard 'da' when they saw "ga" and heard 'ba' at the same time (McGurk & MacDonald 1976). Since the participants were not only affected by the acoustic signal in their perception of the syllables, the results demonstrate that visual information influences speech perception.

More recently, similar studies have been carried out with English phonemes that differ in aspiration, such as /p/ and /b/, by using aero-tactile stimuli in combination with auditory or visual stimuli. The participants were presented with the audio or video of a speaker saying a syllable with one of the phonemes and in some cases, a puff of air would be released simultaneously with the video or audio of the syllable. In the cases where a puff of air was released, the participants were significantly more likely to report that the syllable contained /p/ even if /b/ was transmitted through audio or video (Bicevskis et al. 2016, Gick & Derrick 2009). To test whether these findings were simply due to a general response to tactile stimulus, Gick & Derrick (2009) carried out a control study identical to the one with audio and air, except the participants experienced a short tap on the hand instead of a puff of air on the neck. They found that whether the participants experienced a tap did not affect their perception of the speech. This demonstrates that while not all tactile stimuli can affect speech perception, aero-tactile stimuli do.

While the McGurk effect has been replicated many times (e.g., Rosenblum 2019, Gentilucci & Cattaneo 2005, Munhall et al. 1996, Brown et al. 2018), it has not avoided criticism. One such criticism has been voiced by Van Engen et al. (2022) who argue that the McGurk effect simply differs too much from natural speech as participants only hear isolated syllables, rather than full words or sentences where context and lexical knowledge would inform the identification of sounds. Additionally, according to Van Engen et al. (2022), situations where incongruent stimuli come from the same source do not occur in real life. Essentially, the McGurk effect is criticized for lack of ecological validity.

It is certainly a valid point of criticism that the McGurk effect experiments differ from natural speech situations, however there is a type of situation they are very similar to. The training studies on the acquisition of non-native phonemic contrast with audiovisual stimuli presented in this paper use a very similar procedure. That is, the participants hear minimal pairs, isolated syllables, or nonsense words to limit the influence of context and lexical knowledge on the identification of the phonemes. Therefore, while it might open the experiments up to challenges about ecological validity, it also aids the researchers' ability to focus only on the area they are investigating, namely how multimodal stimuli affect the identification of phonemes. As the speech situations in experiments on the McGurk effect are very similar to the cross-modal training studies presented, in this paper I will regard the

McGurk effect as an indication of how much audiovisual integration occurs in these types of situations.

Another possible point of criticism for the McGurk effect is its lack of universality. It is strong among native speakers of languages such as English, where 92% identified the syllable as 'da' when they experienced a visual 'ga' and auditory 'ba' at the same time (McGurk & MacDonald 1976). Similar levels of receptiveness to the McGurk effect have also been found for native speakers of Spanish (Wang et al. 2008). However, studies have been done with native speakers of Japanese who identified the syllable as 'ba' 100% of the time when they heard 'ba' and saw 'ga' (Sekiyama & Tohkura 1991). This means that the McGurk effect is weak to non-existent for native speakers of Japanese, and the same has been found for native speakers of Chinese (Wang et al. 2008).

There are McGurk effect studies that indicate that children are equally attentive to visual cues regardless of their native language. A study that compared the receptiveness to the McGurk effect for 6-year-old and adult native speakers of English or Japanese found that the visual influence did not differ significantly for the two groups of 6-year-olds (Sekiyama et al. 2003). The two groups of adults did however show a significant difference. The adult native speakers of English were influenced much more by visual information than the 6-year-olds. The native speakers of Japanese did not show the same development as the visual influence did not increase with age. The original study on the McGurk effect also tested children and adults with English as their native language and found that the adults were more receptive to the effect (McGurk & MacDonald 1976). This indicates that the extra attention native speakers of English pay to visual cues is a learned behavior.

A possible explanation for why English native speakers adopt this behavior, while native speakers of Japanese do not, might be that in both Japanese and Chinese there are fewer visually contrastive phonemes. Hence the visual input is less informative and therefore less attention is paid to it. For Japanese native speakers, cultural habits might also affect the salience of the McGurk effect, as looking too much at people's faces can be considered impolite (Inceoglu 2016). As the McGurk effect differs across native languages, cultures, and ages, it cannot be considered universal.

Another example that illustrates speech perception's cross-modal nature is the Tadoma technique. It is usually used by deaf-blind individuals and consists of placing a hand on the speaker's face, so it touches the speaker's lips, jaw, and neck and by doing this they are able to comprehend what the speaker is saying. While the method is becoming rarer as medical and technical advances mean that there are fewer deaf-blind individuals (Rosenblum 2019: 47), it does demonstrate something important about speech perception. Namely, that speech can be perceived accurately through touch. Additionally, it demonstrates that neither auditory nor visual stimuli are necessary to perceive speech.

These studies indicate that speech is perceived through any sensory information that is available about speech, and it can be perceived even without access to the acoustic signal. Therefore, speech and speech perception are not only auditory, rather they are cross-modal.

3. Evidence for the multi-sensory training benefit in L1

While the McGurk effect has shown that several senses influence speech perception by presenting incongruent stimuli, the question is whether the presentation of congruent multimodal stimuli can affect perception. In this section, I will examine whether training with audiovisual stimuli can enhance speech perception of one's native language or aid first language acquisition.

Several studies have shown that if native English speakers receive multimodal training, they will improve their ability to hear degraded speech (Schwartz et al. 2004, Montgomery et al. 1984) or distinguish between similar-sounding speakers when they are presented with the audio alone after the training (Schall & von Kriegstein 2014, Rosenblum et al. 2007). This shows that multimodal training can enhance speech perception in one's native language.

Another finding that supports this conclusion is that children who are visually impaired from birth acquire some contrasts later than their peers with normal sight. For example, they have been found to acquire the /m/-/n/ contrast, which is audibly quite similar but visually distinct, later than children with normal sight (Mills 1983). A study by de Lima & Nunes (2015) interviewed 20 speakers of Brazilian Portuguese between the ages of 6 and 9 with impaired vision to index their phonemic inventories of consonants. They found that the visually impaired children demonstrated a more limited phonetic inventory than their sighted peers. This indicates that there is a multi-sensory training benefit, as people who lack access to one modality showcased delayed phonemic development in some areas when compared to their normally sighted peers. To summarize, people seem to experience a multimodal training benefit in their native language.

4. Evidence for multi-sensory training benefit in L2 acquisition

While evidence suggests there is a multi-sensory training benefit in one's native language, the question is whether learners of a second language also benefit from this kind of training. In this section, I will examine a few of the training studies that investigate the effect of audiovisual training compared to auditory training for the identification of non-native phonemic contrasts.

One example is Hazan et al.'s (2005) experiment on the identification of English /b/, /p/, and /v/ for native Japanese speakers. This labial/labiodental contrast often causes problems for native speakers of Japanese, as there are no labiodental phonemes in Japanese (Hazan et al. 2006: 1742). This experiment's aim was to test whether receiving training with multi-sensory stimuli would aid the identification of this non-native phonemic contrast more than purely auditory training would. The participants in this experiment were tested twice (pre- and post-test) in nonsense words with the consonants in different vowel environments in three test conditions: audio only, video only, as well as an audiovisual (AV) condition. They were divided into two groups based on pre-test performance so the groups' average performance would be similar. Between the pre- and post-test one group received auditory training and the other received audiovisual training for the consonant contrasts. The training was High Variability Phonetic Training, meaning the training material included multiple speakers in multiple phonetic contexts rather than a single speaker in a single context (Colantoni et al. 2015), and they received feedback on their answers in the training. The results of the two groups'

tests showed that both groups improved after training, but as is shown in Figure 1, they did not improve equally. The points on the graph show the change in correct identification of the phonemic contrast from the pre- to post-test in each condition for each of the groups. As is visible on the graph, the audiovisual training group improved significantly more in every test condition (Hazan et al. 2005). Therefore, this study suggests that multi-sensory training can be more effective than unimodal training for the acquisition of non-native phonemes.

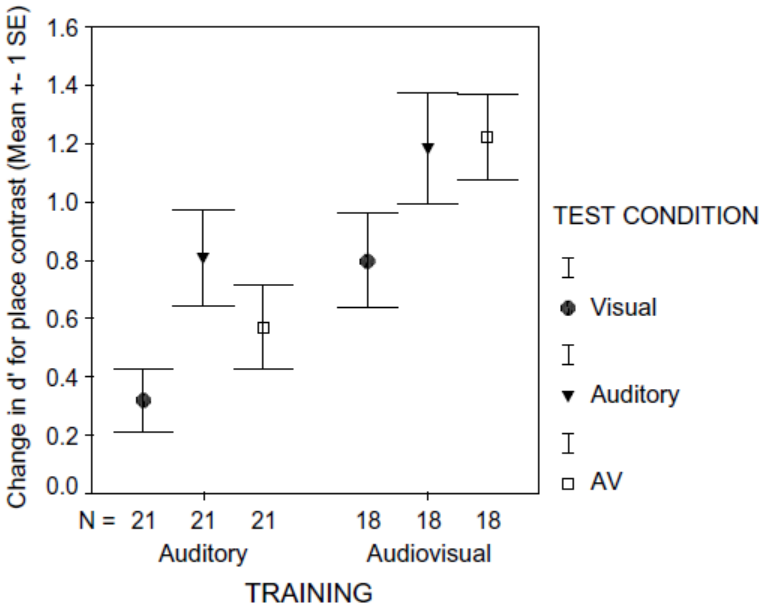

Figure 1: Results for the auditory and audiovisual training groups. The y-axis shows change in d' (a measure of how often the labial/labiodental contrasted was correctly identified) from pre- to post-test. The results of the auditory (left) and the audiovisual (right) training groups in each of the test conditions are shown on the x-axis. N is the number of participants (Hazan et al. 2005: 366).

Similar studies have been done on e.g. French vowels that differ in terms of roundedness for English native speakers (Inceoglu 2016), English /r/ and /l/ for native speakers of Japanese and Korean (Hardison 2003), and English consonants such as /p/ and /f/ for native speakers of Korean (Lee et al. 2018). The methods of these studies are similar with small variations. They all have a group that receives auditory training and one that receives audiovisual training, two identification tests in different modalities with training in between, but the amount of training differs, as does the type of stimuli used and the participants' proficiency in the target language. While there are minor methodological differences between them and Hazan et al. (2005), they do all point toward the same conclusion. All these studies found that the group trained with audiovisual stimuli improved significantly more than the group trained with only auditory stimuli. Together, these studies provide a strong indication that there is a multi-sensory training benefit in the acquisition of non-native phonemes.

5. LIMITATIONS AND TOPICS FOR FUTURE RESEARCH

There is evidence that supports the hypothesis that learners of a foreign language can benefit from multi-sensory training. However, it is not the case that every study with multimodal training reaches this result. In this section, studies that did not find a multi-sensory training benefit will be examined and possible explanations for this will be discussed. Additionally, the universality of the multi-sensory training benefit will be called into question both in terms of participants' sensitivity to speech cues in certain modalities, as well as whether this benefit occurs with all modalities.

5.1 Salience of sensory cues

One of the training studies whose findings do not support the existence of a multi-sensory training benefit is an experiment by Hazan et al. (2005) on English /r/ and /l/ for native Japanese speakers. As I mentioned earlier, there is another study on the same phonemic contrast with native speakers of Japanese, where a multimodal training benefit was shown (Hardison 2003). At first glance, the studies seem nearly identical except for their outcome. However, there is an important difference between these two studies: the variety of English that was used for the training and test material.

The study by Hardison (2003) used speakers of American English who produced an [ɻ], while the study by Hazan et al. (2005) used speakers of Southern British English (SBE) who realized the phoneme as [ɹ]. As Hazan et al. (2005: 372) point out in their article, there are differences in how much visual information is conveyed by the two regiolects, as [ɻ] is slightly rounded, while [ɹ] and /l/ are not. This means that there is a larger visual difference between [ɻ] and /l/ than [ɹ] and /l/, so the additional modality was more informational for American English. This could explain why the participants trained with American English experienced a multi-sensory training benefit while those trained with SBE did not.

Another study that did not find a multi-sensory training benefit was done by Lee et al. (2018). As mentioned in the previous section, in one of their experiments Lee et al. (2018) found that there was a multimodal training benefit for native speakers of Korean with English consonant contrasts such as /p/ and /f/. In their second experiment, they used the same procedure with vowel contrasts. All but one of the consonant contrasts were significantly visually contrastive. The vowel contrasts, however, were vowel pairs that are close to each other in the vowel space and do not differ in roundedness, such as /ɛ/ and /æ/. Therefore, the tested vowel contrasts were not as visually distinctive as the consonant contrasts. The experiment with the vowel contrasts found that the audiovisual training group did not improve more than the audio-only training group in their identification of the vowels.

This seems to point toward the same conclusion as Hazan et al. (2005), namely, that for audiovisual training to be more effective than purely auditory training, the phonemes need to be visually contrastive. To put it more generally, this indicates the multi-sensory training benefit only occurs when the phonemic contrast is contrastive in the additional modality.

5.2 The universality of the McGurk effect and the multi-sensory training benefit

The McGurk effect's lack of universality has significant implications for research on the multi-sensory training benefit. It raises the question of whether the salience of the audiovisual McGurk

effect in a participant's native language is correlated to how receptive they are to audiovisual training. Most of the studies on the multi-sensory training benefit in one's native language, including the ones presented in this article, have been done on native English speakers. This is not sufficient evidence to convincingly claim that the multi-sensory training benefit for L1 is universal. However, there are studies that have shown that native Japanese speakers experience a McGurk effect almost as strong as native English speakers when noise is added to the acoustic signal (Sekiyama & Tohkura 1991). Therefore, a topic for further research could be whether native Japanese speakers experience the same multi-sensory training benefit for the perception of degraded speech as native English speakers.

This question of universality undoubtedly applies to multi-sensory training in foreign language speech as well. Therefore, I will now investigate whether the receptiveness to visual cues in one's native language carries over to non-native speech perception. This was the central focus of Hazan et al. (2006). In the experiment, they tested native speakers of Spanish and Japanese once in identification of the English phonemes /b/, /p/, and /v/. Figure 2 shows a measure of each group's number of correct answers in each condition. As the graph shows, the native Spanish speakers did significantly better in every test condition than the native speakers of Japanese (Hazan et al. 2006). At a glance, this could indicate that Spanish speakers, who are more receptive to the McGurk effect than native speakers of Japanese, are also more receptive to visual speech cues when identifying non-native phonemes.

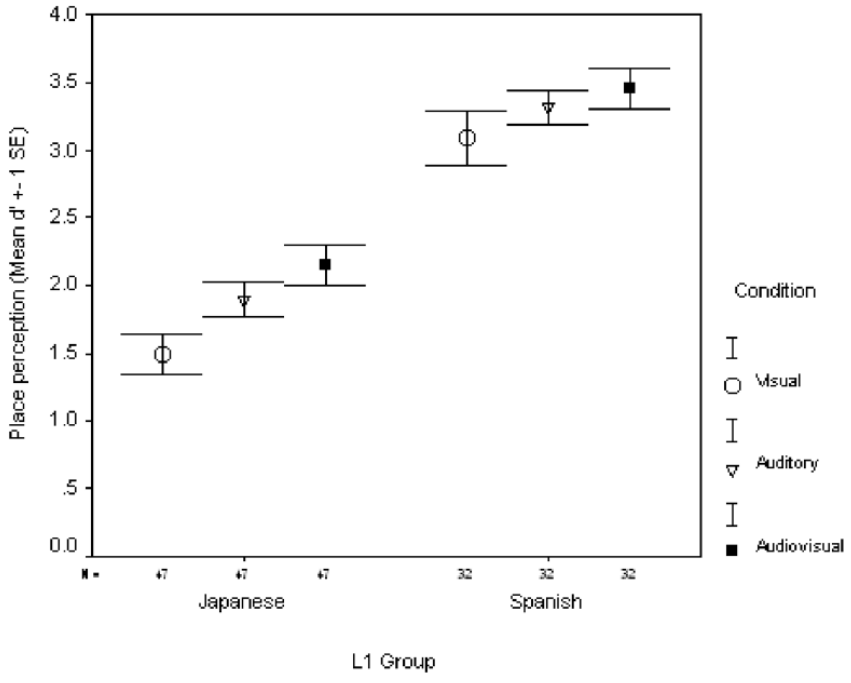

Figure 2: Results for native speakers of Japanese and Spanish. The y-axis indicates d' (a measure of how often they correctly identified the labial/labiodental phonemes in the tests. The result of the L1 Japanese (left) and the L1 Spanish (right) groups in each test condition are shown on the x-axis. N is the number of participants (Hazan et al. 2006: 1745).

However, it seems that there could be another explanation for why the two groups did differently. Namely, that the phoneme inventories of Spanish and Japanese differ when it comes to the tested contrasts. [v] exists as an allophone in Spanish, and there is a labial/labiodental contrast in Spanish between /p/ and /f/. In Japanese, however, there are no labiodental sounds. As Spanish speakers were more familiar with the [v] sound as well as the labial/labiodental place of articulation contrast, this might also be a factor in why they outperformed the native Japanese speakers (Hazan et al. 2006: 1742). Therefore, while it might be tempting to claim that the weighing of visual cues in one's native language affects one's receptiveness to audiovisual speech cues in non-native speech, this study does not provide a solid enough basis to conclude whether this is the case.

As I have attempted to demonstrate in this section, it is difficult to conclusively say whether the relative weighing of sensory cues in one's native language is related to one's receptiveness to multi-sensory speech cues and training. Since children with different language backgrounds do not differ in their attentiveness to different senses while adults do, this is a learned behavior. It might be possible that adults can also learn to pay attention to these cues if they are trained. Future research could investigate whether native speakers of e.g., Japanese and English who receive the exact same training and the same tests experience the same multi-sensory training benefit. To limit how much any differences could be explained away by differences in phoneme inventories of the participant's native language, such a study could use a model such the PAM-L2 to attempt to find a contrast that is assimilated similarly by all participants. The Perceptual Assimilation Model for of Second Language Speech Learning (PAM-L2) is a theory that proposes that a listener has formed a set of phonemic categories based on the articulatory gestures of the sounds in their native language. According to the model, the phonemes of the L2 will either be assimilated to existing categories, some with a better fit than others, or remain uncategorized and these assimilation patterns in theory predict a listener's ability to discriminate non-native phonemic contrasts (Colantoni et al. 2015). Yet, however, there seems to be no research that has attempted to answer this question with this approach and therefore, it remains unanswered whether the audiovisual multi-sensory training benefit is experienced equally regardless of language background or depends on the importance of visual information in the participants' native language.

5.3 Other modalities

It is important to note that almost every study on cross-modal speech perception that I have referenced in this article have used audiovisual stimuli. The only studies on other combinations of modalities are the studies that demonstrated the McGurk effect also occurs when puffs of air are combined with audio or video. Therefore, while I have concluded that there is in some cases a multi-sensory training benefit, it would be more accurate to refer to it as an audiovisual training benefit. It seems that there is little to no research on other combinations of senses so far, even though speech is perceived through other senses such as touch. As an example of a study that would not use audiovisual stimuli, it could be investigated whether training with audio and puffs of air would aid identification of the Thai phonemes /b/, /p/, and /pʰ/ for native speakers of e.g. English who do not have this three-way aspiration contrast (Colantoni et al. 2015: 196). As Danish has no voicing contrast between /s/ and /z/ (Grønnum 1998: 99) a topic for future research could be whether training with audio and touch

would improve native Danish speakers' ability to distinguish between English /s/ and /z/. However, it seems that the question of whether the multi-sensory training benefit occurs with more than just audiovisual stimuli remains unanswered.

All the cross-modal speech perception studies referenced in this paper have only combined two modalities at a time. I have been unable to find any research that investigates whether training with three modalities, such as sight, sound, and touch, would be more effective than training with two modalities, which could be a topic for future research.

6. Conclusion

In this paper, I have argued that speech perception is not only auditory, and that multi-sensory training can enhance the perception of audio-only speech as well as aid the acquisition of native phonemic distinctions. There is evidence to support the hypothesis that there is also a multi-sensory training benefit for learners of a foreign language. However, this seems to only be the case when the phonemes are contrastive in the additional modality.

Further research is needed to better understand the effects of multi-sensory training on speech perception. It is uncertain whether the multi-sensory training benefit differs across languages, cultures, and ages as the McGurk effect does and more research is needed in this area. The existing research is primarily focused on audiovisual training studies. However, as I have argued, speech is also perceived through other senses such as touch. It remains unanswered whether the multi-sensory training benefit occurs only with audiovisual stimuli or if it could occur with other modalities as well. In conclusion, the evidence presented suggests that while there is a multi-sensory training benefit in some cases, there are many questions about the extent and limitations of this effect that have yet to be answered.

REFERENCES

Bicevskis, K., D. Derrick & B. Gick. 2016. Visual-tactile integration in speech perception: Evidence for modality neutral speech primitives. *The Journal of the Acoustical Society of America* 140(5). 3531-3539. DOI: 10.1121/1.4965968.

Blumstein, S. & A. Cutler. 2003. Speech Perception. W. Frawley (ed.), *International Encyclopedia of Linguistics*. 2nd ed. Oxford: Oxford University Press. DOI: 10.1093/acref/9780195139778.001.0001.

Brown, V. A., M. Hedayati, A. Zanger, S. Mayn, L. Ray, N. Dillman-Hasso & J. F. Strand. 2018. What Accounts for Individual Differences in Susceptibility to the McGurk Effect? *PLoS ONE* 13(11). DOI: 10.1371/journal.pone.0207160

Colantoni, L., J. Steele & P. R. E. Neyra. 2015. *Second Language Speech: Theory and Practice*. Cambridge, United Kingdom: Cambridge University Press.

de Lima, A. L. & R. T. D. Nunes. 2015. Phonological profile of children with low vision from 6 to 9 years old at an institution for the blind in the city of Salvador. *Revista CEFAC* 17(5). 1490-1498. DOI: 10.1590/1982-0216201517521114.

Gentilucci, M. & L. Cattaneo. Automatic Audiovisual Integration in Speech Perception. 2005. *Experimental Brain Research* 167 (1). 66-75. DOI: 10.1007/s00221-005-0008-z.

Gick, B. & D. Derrick. 2009. Aero-tactile integration in speech perception. *Nature* 462(7272). 502-504. DOI: 10.1038/nature08572.

Grønnum, N. 1998. Danish. *Journal of the International Phonetic Association* 28(1-2). 99-105. DOI: 10.1017/S0025100300006290.

Hardison, D. M. 2003. Acquisition of second-language speech: Effects of visual cues, context, and talker variability. *Applied Psycholinguistics* 24. 495-522.

Hazan, V., A. Sennema, A. Faulkner, M. Ortega-Llebaria, M. Iba & Hyunsong Chung. 2006. The use of visual cues in the perception of non-native consonant contrasts. *The Journal of the Acoustical Society of America* 119(3). 1740-1751. DOI: doi.org/10.1121/1.2166611.

Hazan, V., A. Sennema, M. Iba & A. Faulkner. 2005. Effect of audiovisual perceptual training on the perception and production of consonants by Japanese learners of English. *Speech Communication* 47(3). 360-378. DOI: 10.1016/j.specom.2005.04.007.

Inceoglu, S. 2016. Effects of perceptual training on second language vowel perception and production. *Applied Psycholinguistics* 37(5). 1175-1199. DOI: 10.1017/S0142716415000533.

Lee, H., J. Lee, H. Hwang & J. Y. Lim. 2018. Effects of audiovisual perceptual training with corrective feedback on the perception and production of English sounds by Korean learners. *The Journal of the Acoustical Society of America* 144(3). 1867-1867. DOI: 10.1121/1.5068210.

McGurk, H. & J. MacDonald. 1976. Hearing lips and seeing voices. *Nature* 264(5588). 746-748. DOI: 10.1038/264746a0.

Mills, A. E. 1983. Acquisition of speech sounds in the visually-handicapped child. Mills, A.E. (ed.), *Language acquisition in the blind child: Normal and deficient,* 45-56. San Diego: College Hill.

Montgomery, A. A., B. E. Walden, D. M. Schwartz & R. A. Prosek. 1984. Training Auditory-Visual Speech Reception in Adults with Moderate Sensorineural Hearing Loss. *Ear and Hearing* 5. 30-36.

Munhall, K. G., P. Gribble, L. Sacco & M. Ward. 1996. Temporal constraints on the McGurk effect. *Perception & Psychophysics* 58(3). 351-62. DOI: 10.3758/bf03206811.

Rosenblum, L. D. 2019. Does Sensory Modality Matter? Not for Speech Perception. *Acoustical Society of America* 15(3). 47-54.

Rosenblum, L.D., R. M. Miller & K. Sanchez. 2007. Lip-Read Me Now, Hear Me Better Later: Cross-Modal Transfer of Talker-Familiarity Effects. *Psychological Science* 18(5). 392-396. DOI: 10.1111/j.1467-9280.2007.01911.x.

Sekiyama, K., D. Burnham, H. Tam & D. Erdener. 2003. Auditory-visual speech perception development in Japanese and English talkers. *Proceedings of the International Conference on Auditory-Visual Speech Processing, St. Jorioz, France*. 61-66. https://www.researchgate.net/publication/237903326_Auditory-visual_speech_perception_development_in_Japanese_and_English_talkers (Accessed 16 Apr 2023).

Sekiyama, K. & Y. Tohkura. 1991. McGurk Effect in Non-English Listeners: Few Visual Effects for Japanese Subjects Hearing Japanese Syllables of High Auditory Intelligibility. *The Journal of the Acoustical Society of America* 90(4). 1797-1805. DOI: 10.1121/1.401660.

Schall, S. & K. von Kriegstein. 2014. Functional connectivity between face-movement and speech-intelligibility areas during auditory-only speech perception. *PLoS ONE 9.* DOI: 10.1371/journal.pone.0086325.

Schwartz, J., F. Berthommier & C. Savariaux. 2004. Seeing to Hear Better: Evidence for Early Audio-Visual Interactions in Speech Identification. *Cognition* 93(2). B69-B78. DOI: 10.1016/j.cognition.2004.01.006.

Van Engen, K. J., A. Dey, M. S. Sommers & J. E. Peelle. 2022. Audiovisual speech perception: Moving beyond McGurk. *The Journal of the Acoustical Society of America* 152(6). 3216-3225. DOI: 10.1121/10.0015262.

Wang, Y., D. M. Behne & H. Jiang. 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.