eureka Review article

The McGurk Effect Across Languages

Received: 24 January 2023
Accepted: 03 March 2023
Published: 12 April 2023

Andres F. Dorado Solarte[1]
[1] Department of Psychology, University of Alberta
* Corresponding author: adorados@ualberta.ca

ABSTRACT

The McGurk effect denotes a phenomenon of speech perception where a listener attends to mismatched audio and visual stimuli and perceives an illusory third sound, typically a conflation of the audio-visual stimulus. This multimodal interaction has been exploited in various English-language experiments. The article explores the manifestations of this effect in other languages, such as Japanese and Chinese, as well as considerations for age and keenness (hearing acuity) through a literary review of existing research. The literature confirms the McGurk effect is present in other languages, albeit to differing degrees. The differences in the McGurk effect across languages may be attributed to linguistic and cultural differences. Age differences demonstrate a greater lip-reading reliance as age increases in participants; a similar reliance on visual information is seen in participants as hearing impairment increases. Experimental designs should refine audiovisual stimuli by using immersive technology such as three-dimensional models in virtual reality or ambisonic playback that offers multi-directional sound signals. Future research should also address the influence of audiovisual integration in marketing, foreign language education, and developing better accommodations for the hearing impaired.

KEY WORDS: **McGurk Effect, Linguistics, Speech Perception, Psycholinguistics, Lip-reading**

1 | INTRODUCTION

Speech is typically transmitted audibly, and this audible signal is then perceived by a listener. Some studies have shown that speakers and listeners tend to adapt according to their environments to ensure they are perceived as intended, delivering the speech signal as clearly as possible. For example, individuals tend to involuntarily respond to their environment by adjusting, among other acoustic features, the intensity (volume) of their speech; this reflex is known as the Lombard effect, named after the French otolaryngologist Étienne Lombard, who discovered it (Lombard, 1911). Another phenomenon is Erdner's teacherese, which describes the over-articulated/exaggerated (hyper) speech typically used by teachers in the classroom to ensure their intended speech signal is understood (Erdner, 2017).

Those phenomena demonstrate individual modifications to the audible output (e.g., Lombard speech, and hyperspeech). Accordingly, speech perception could be treated primarily as an auditory event. However, a seminal study found that speech is also perceived visually (McGurk & MacDonald, 1976). They observed that speakers would hear an illusory third sound when presented with mismatched audio and visual stimuli. Thus, they demonstrated that speakers tend to rely on visual information, especially if they need to compensate for any lacking information in auditory signals, usually due to noisy environments that may muddle the speech signal for the listener (McGurk & Macdonald, 1976). This audiovisual integration is known widely as "the McGurk Effect". This effect has been re-examined in many experiments manipulating different environments, varying visual format and auditory stimuli but mostly focused on a similar participant pool: English speakers. Additionally, exploring the McGurk effect has significance in a variety of applications outside of academia, ranging from foreign language acquisition to advertising and the broadcasting industry (Erdner, 2017).

This review will evaluate the McGurk effect across languages, noting the differences in how the phenomenon manifests in English, Japanese and Chinese participants. Furthermore, within language differences will also be presented, such as age and auditory acuity (e.g., hearing impairment, environmental contexts). As we come to understand how the McGurk effect is expressed in different languages, the practical implications derived from studying

the phenomenon in one language may align with comparable applications in other languages (Sekiyama, 1994).

The between- and within-language differences were gathered from a non-systematic literature review to gather a cursory understanding of the existing literature. Thus, the results included are not exhaustive and do not represent an extensive review of the subject matter. The criteria for the articles chosen in the review included prominent phonological differences between English and a foreign language, the effect of culture on lip-reading, and age considerations. Additionally, experiments that compared the McGurk effect in English, Chinese and Japanese did well in demonstrating the nuance of the subject across languages (Sekiyama, 1994; Sekiyama, 1997). Moreover, the literature on the McGurk effect in Japanese and Chinese participants is most comparable in quality and quantity to the body of work done with English participants (Sekiyama, 1994).

2 | MCGURK EFFECT ACROSS LANGUAGES

The McGurk effect is manifested in several languages, albeit to varying degrees. It is especially exhibited in languages that are considered phonologically complex, such as English and Spanish (Tona et al., 2015). The effect is also more easily induced in environments that complicate audition, such as noisy situations or for first-language speakers (or, L1 speakers) while hearing a foreign language (Sekiyama, 1997). Generally, conducting a McGurk-style experiment means displaying a video of someone speaking while playing audio through headphones placed on the participant to isolate the sound stimuli, although there are variations to presenting stimuli, such as playing audio through a speaker (McGurk & Macdonald, 1976). The strength or magnitude of the McGurk effect is determined by the number of auditory confusion errors (an illusory third vowel, usually 't') subtracted from the gross error (the total number of mismatched stimuli, usually 'p' or 'b') (Sekiyama, 1997).

2.1 Between-Language Differences

2.1.1 English

The McGurk effect was originally observed with English-speaking participants; thus, it has been primarily studied within English-language experiments (Sekiyama, 1994). Although the phenomenon is present outside of English contexts, it is not experienced homogenously in every language (Erdner, 2017). Several researchers have pioneered studies that examined these differences, usually referring to English-speaker control groups as a baseline for comparing across-language effects; many of these studies have found that English speakers have the highest reliance on visual information; they also demonstrate an even greater degree of audiovisual integration when tested on foreign language stimuli (Erdner, 2017; Sekiyama, 1997; Sekiyama & Burnham, 2008; Sekiyama & Tohkura, 1993; Sekiyama & Tohkura, 1991; Tona et al., 2015).

2.1.2 Japanese.

Much of the research on the McGurk effect in Japanese participants was pioneered by Sekiyama (Sekiyama, 1997; Sekiyama & Burnham, 2008; Sekiyama & Tohkura, 1993; Sekiyama & Tohkura, 1991). She found that native Japanese participants relied significantly less on lip-reading information than English L1 participants, even when Japanese participants were presented with foreign language stimuli. Some suggest that linguistic factors are at the root of these differences in reliance on visual information, citing phonological complexity: Japanese is considered less phonologically complex than English because it has fewer vowels, fewer consonants, and no phonological consonant clusters (where two or more consonants are clustered together, e.g., the 'lk' in hulk and 'str' in strong) (Sekiyama & Burnham, 2008; Sekiyama & Tohkura, 1993; Tona et al., 2015).

Moreover, Japanese does not have any labiodentals (sounds articulated with the lips and the teeth, e.g., 'f' and 'v') in its consonant inventory. This may explain the lesser degree of lip-reading in Japanese participants, as the McGurk effect tends to show fused responses between labials (sounds articulated with the lips, e.g., 'f', 'p', 'm', and 'w') and other consonants. Such fusion may not be so prevalent in Japanese given its simpler structure, and thus a lesser need to compensate the speech signals with visual information obtained from lip-reading to discriminate between syllables (Sekiyama & Tohkura, 1993). Others have posited that these differences in audiovisual integration may be due to cultural factors, as there is a tendency in Japanese culture to generally avoid looking directly at the speaker (Sekiyama, 1997; Sekiyama & Tohkura, 1993; Sekiyama & Tohkura, 1991).
Considering this cultural note, some researchers have compared Japanese participants with Chinese participants, due to cultural similarities in gaze-avoidance; linguistic similarities also reinforce the validity of their comparison (Sekiyama, 1997; Tona et al., 2015).

2.1.3 Chinese

Studies on Chinese audiovisual integration showed a weaker McGurk effect, and despite cultural and some

linguistic similarities the extent of lip-reading dependence was even more limited than that exhibited by Japanese L1 participants (Prieto et al., 2015). This reduced integration of visual information is largely due to the various lexical tones (the use of pitch to express differences in meaning, e.g., in Chinese, 'mā' mother and 'má' hemp) present in Mandarin and Cantonese, further facilitating discrimination, insomuch that compensating the speech signal with lip-reading is not as necessary (Erdner, 2017).

2.2. Within-Language Differences

2.1.1 Hearing impairment

Within-language factors may also affect the presence of the McGurk effect and the influence of visual information across modalities in speech perception. One factor that may have an effect is auditory acuity. A study compared profoundly deafened Japanese children with cochlear implants against a group of normal-hearing children (Tona et al., 2015). A McGurk-style experiment using mismatching audio and visual stimuli was conducted to measure the level of reliance on lip-reading and visual information in participants. The study revealed that, despite nearly equal speech perception skills, prelingually deafened children generally relied more on visual input than normal-hearing participants, perhaps due to insufficient speech signals provided by their cochlear implant.

Moreover, it was found that the McGurk effect was more easily induced and to a stronger degree in deafened children over six years old than it was their younger counterparts (Tona et al., 2015). Other studies confirmed that these findings carried through into adulthood, where adult participants maintained strong lip-reading abilities; postlingually deafened adults with cochlear implants also tended to provide more visual responses and fewer auditory responses, than normal-hearing control groups (Tona et al., 2015).

2.1.2 Age

As noted by Tona's findings, another factor that may affect the degree to which a participant relies on audiovisual integration is the age of an individual, as they also reported in their publication that the McGurk effect can be more easily induced with increased age (Tona et al., 2015). This is further reinforced as adults show a greater dependence on visual information than children when presented with foreign language stimuli, perhaps due to a loss in foreign phonemic recognition as age increases among participants (Erdner, 2017). An increase of dependence on visual information can be seen as early as five years old with notable differences between 5-year-olds and participants ranging six to eight years old, and again between 14–15-year-olds (Sekiyama & Burnham, 2008; Tona et al., 2015).

3 | DISCUSSION

The literature evaluated in this review suggests that the McGurk effect and, generally, audiovisual integration is present globally in the participants and the languages studied at varying degrees of magnitude. Phoneme space, perceptual difficulty in labial/nonlabial discrimination and cultural traditions determine how easily the McGurk effect is induced, however environmental aspects such as noise or quality of the auditory signal also influenced the need for relying on visual information (Sekiyama & Tohkura, 1991). Auditory ambiguity produced by the native-foreign language effect also increased lip-reading reliance, regardless of L1 (Erdner, 2017; Sekiyama, 1997). That cross-modal influence can be measured by simply evaluating the strength of the McGurk effect in a participant and that these results are replicated in several languages means that dependency on visual information can be assessed and applied accordingly. For example, results can be applied in marketing, by helping vendors to determine how much visual information they must provide to compensate for a lacking speech signal in an advertisement.

However, the research could go further than just examining how visual information is integrated to compensate the auditory signal and aid in understating articulation. Effectively, one of several limitations of a McGurk-style experiment resides in its use of pre-recorded video and sound stimuli to examine the effect. Though the effect was conceived by the mismatch of sound and video, not all speech signals perceived in our day-to-day conversations happen through a two-dimensional representation of our three-dimensional selves (as ubiquitous as online videoconferencing is becoming in today's world). Future research could further examine the McGurk effect in a 3D space using virtual reality technology in more phonologically complex languages like those mentioned in this review, building on the studies that have already been done in the more phonologically simple English and French languages (Siddig, Sun, Sun, Parker & Hines, 2019; Thézé et al., 2020). Another limitation is the use of monophonic or stereophonic playback (using one or two audio channels) through speakers or headphones, which makes it difficult to mimic true ambient sound as it is easier to localize the origin of the signal, whereas in a noisy environment, the disruptive signals that confuse a listener can be projected from any direction (Snow, 1953; Yao et al., 2020). This is addressed in another study, where a

McGurk-style experiment was realized using ambisonic playback, meaning audio was played back through multiple channels to simulate sounds coming from all directions (Siddig, Ragano, Jahromi & Hines, 2019); however, like much of the work on the McGurk effect, the ambisonic study was done in English. Thus, future research could explore the McGurk effect using ambisonic sound in more phonologically and tonally diverse languages.

Future research on this topic should also move beyond examining the integration of visual information and explore how cross-modal information influences prosody (the use of stress or emphasis and intonation in languages to express meaning) and how these visual gestures that listeners integrate contribute to "filling in the blanks" as a whole. Aside from its significance in marketing and advertising, it would be interesting to explore how the McGurk effect can be applied generally, to accommodate for hearing impairment in media (beyond providing closed captions). Moreover, the results discussed surrounding the cultural considerations noted in Chinese between-language differences could lead to an interesting discussion of whether traditions of face-avoidance facilitated a need for tonal diversity and phonological simplification to ease discernibility and perceptual discrimination in tonal languages, e.g., Mandarin, Cantonese, Japanese, along with other cultural considerations for other tonal languages with cultures where face or gaze-avoidance is not a factor, e.g., Swedish, Norwegian (Argyle et al., 1994). Finally, considering a speaker's involuntary and voluntary adaptations to their auditory output according to their environment (e.g., the Lombard effect and teacherese, respectively), future experiments could analyse whether speakers also make adaptations to their visual output (e.g., exaggerated facial articulation) and whether the adaptation is voluntarily or involuntarily used to make visual cues more reliable for bimodal integration.

4 | CONCLUSION

Evaluating the McGurk effect across languages provides a lot of information on a language (e.g., the discrimination of consonant clusters, discerning lexical tones) and manipulating the different stimuli within a McGurk-style experiment demonstrates to researchers how listeners "see" language (Erdner, 2017). Using mismatched audio/visual stimuli to induce the McGurk effect is a common method of testing how much participants rely on visual information to compensate for inadequate auditory signals (Sekiyama, 1997). This seems to work well across languages; however, each language differs in its degree of audiovisual integration. Comparing reliance on cross-modal information across English L1, Japanese L1 and Chinese L1 participants demonstrates that the McGurk effect manifests differently depending on linguistic features, (e.g., the presence or absence of certain consonants) and cultural traditions (e.g., whether looking at the face of someone who is speaking is respectful or not) (Prieto et al., 2015; Sekiyama & Burnham, 2008).

The literature included in this review primarily focuses on establishing and documenting differences in linguistic features and cultural traditions to highlight the different factors of variability in the audiovisual integration of speech. Some studies have even covered within-language factors like deafness/hearing impairment, age, or disorders like Cerebral Palsy (Sekiyama, 1997; Sekiyama & Tohkura, 1991). Others evaluated language competencies that may be indicators for visual compensation (e.g., a child who makes substitution errors in articulation typically rely less on visual information) (Sekiyama, 1997). The bimodal integration of audiovisual stimuli evaluated through McGurk effect experiments offer greater insight into language acquisition and language processing that can inform our everyday human experience of communication. Increasing our understanding of the different processes involved in speech perception, can lead to improving our means of relaying information. Whether this means offering more pronounced visual cues or exaggerating our articulation of speech signals, the adaptations may vary depending on the presence of the McGurk effect across languages.

REFERENCES

Argyle, M., Cook, M., & Cramer, D. (1994). Gaze and mutual gaze. *The British Journal of Psychiatry*, 165(6), 848-850. https://doi.org/10.1017/S0007125000073980

Erdner, D. (2017). Teaching Turkish as a Foreign Language: Extrapolating from Experimental Psychology. *Journal of Language and Linguistic Studies*, 13(1), 156–165. https://search.informit.org/doi/10.3316/informit.059860494688059

Lombard, E. (1911). Le signe de l'élévation de la voix. Annuaire Maladies Oreille Larynx Nez Pharynx, 37, 101-119.

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

Prieto, P., Puglesi, C., Borras-Comes, J., Arroyo, E., Blat, J. (2015). Exploring the contribution of prosody and gesture to the perception of focus using an animated agent. Journal of Phonetics, 49, 41-54. https://doi.org/10.1016/j.wocn.2014.10.005

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

Sekiyama, K. (1994). Differences in auditory-visual speech perception between Japanese and Americans: McGurk effect as a function of incompatibility. *Journal of the Acoustical Society of Japan (E)*, 15(3), 143-158. https://doi.org/10.1250/ast.15.143

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

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

Sekiyama, K., Tohkura, Y. (1991). McGurk effect in non-English listeners: Few visual effects for Japanese subjects hearing Japanese syllables of high auditory intelligibility. *Journal of the Acoustical Society of America*, 90, 1797-1805. https://doi.org/10.1121/1.401660

Siddig, A., Ragano, A., Jahromi, H. Z., & Hines, A. (2019). Fusion confusion: Exploring ambisonic spatial localisation for audio-visual immersion using the McGurk effect. In Proceedings of the 11th ACM Workshop on Immersive Mixed and Virtual Environment Systems (pp. 28-33). https://doi.org/10.1145/3304113.3326112

Siddig, A., Sun, P. W., Parker, M., & Hines, A. (2019). Perception deception: Audio-visual mismatch in virtual reality using the McGurk effect. *AICS*, 2019, 176-187. http://aics2019.datascienceinstitute.ie/papers/aics_18.pdf

Snow, W. B. (1953). Basic principles of stereophonic sound. Journal of the Society of Motion Picture and Television Engineers, 61(5), 567-589. https://doi.org/10.5594/J00963

Thézé, R., Gadiri, M. A., Albert, L., Provost, A., Giraud, A. L., & Mégevand, P. (2020). Animated virtual characters to explore audio-visual speech in controlled and naturalistic environments. Scientific reports, 10(1), 15540. https://doi.org/10.1038/s41598-020-72375-y

Tona, R., Naito, Y., Moroto, S., Yamamoto, R., Fujiwara, K., Yamazaki, H., Shinohara, S., Kikuchi, M. (2015). Audio–visual integration during speech perception in prelingually deafened Japanese children revealed by the McGurk effect. *International Journal of Pediatric Otorhinolaryngology*, 79, 2072-2078. https://doi.org/10.1016/j.ijporl.2015.09.016

Yao, S. N., Chen, J. H., Ke, C. T., & Chang, Y. H. (2020). Smartphone-Controlled Multi-Channel Surround Sound System. In Proceedings of the 10th International Conference on Information Communication and Management (pp. 66-70). https://doi.org/10.1145/3418981.3418991

How to cite this article:

Dorado Solarte, A. F. (2023). The McGurk Effect Across Languages. *Eureka. 7* (1). https://doi.org/10.29173/eureka28785