

The McGurk Effect in Relation to Musicians' Abilities

Claire Luo¹, Olivia Yeroushalmi¹, and Alan Schorn¹

¹John L. Miller Great Neck North High School, Great Neck, NY, US
DOI: <https://doi.org/10.47611/jsrhs.v10i2.1473>

ABSTRACT

The original study of the McGurk Effect, a perceptual phenomenon caused by contradictory audiovisual stimuli fusing together to create the illusion of a third sound, was carried out by psychologists McGurk and MacDonald in 1976. The results of early experiments displayed that observers used both auditory and visual signals while being spoken to, auditory signals being the sound waves entering their ears, and visual signals being how the speaker moved his face while pronouncing a word. When conflicting signals are given, a third sound is perceived, as the brain is disoriented from the different signals. The idea that musicians have superior audiovisual cortexes have led some to speculate if musicians are as susceptible to the McGurk Effect as non-musicians. To research the susceptibility of musicians to the McGurk Effect, the experiment conducted included a total of 40 subjects, 20 musicians and 20 non-musicians. The subjects were played a control video of a speaker saying “ga” and were then presented with four audiovisually incongruent videos, all containing a speaker mouthing the word “ga” with the audio recording of the speaker saying “ba” dubbed on. Two main 2x2 Chi Square tests and fifteen secondary 2x2 Chi Squares tests were run in total. The two main tests, which compared the amount of McGurk interpretations to either audio or visual interpretations, both produced a p-value of <.0005. Upon further research, 25.7% of musicians reported a McGurk interpretation, as opposed to 52.2% of non-musicians, which implied that musicians are less susceptible to the McGurk effect.

Introduction

It is well-known that the human brain uses information from both the auditory and visual cortexes to process words that have been spoken (Gentilucci & Cattaneo, 2005)¹. A popular psychology phenomenon, the McGurk Effect is the end result of two contradictory audio and visual signals sent to the brain. Since the brain uses both auditory and visual cues to determine what an individual perceives, if an individual encounters incongruent audiovisual stimuli, he or she will most likely become confused and hear neither of the two actual words presented in the stimuli. Oftentimes, the brain is so disoriented that subjects end up hearing a mixture of the two words given in the stimuli (McGurk & MacDonald, 1976). For example, a subject seeing the word “da” being mouthed and physically hearing the word “ta” might end up thinking that he or she heard something along the lines of “va”, despite the fact that it was neither the word seen or heard. These response types are largely dependent on the strength of the STS, or Superior Temporal Sulcus, which dictates a substantial amount of speech integration (Nath & Beauchamp, 2012)¹. Since McGurk’s initial experiment, many have investigated different variations of the McGurk effect, like Nath and Beauchamp (2011)¹ on the importance of the STS. Variations of the experiment on language and age are also very popular (Burnham & Dodd, 2004; Sekiyama, 1997). However, the idea that musicians possess more grey matter in their brains and have more complex auditory and visual skills have piqued the interest of select individuals. An experiment regarding the effects of contradictory audiovisual stimuli on musicians and non-musicians has been run to determine if playing an instrument has an effect on an individual’s audiovisual speech perception. Both congruent and incongruent audiovisual stimuli were used in this experiment, in which equal amounts of musicians and non-musicians were tested to see if one population was more or less susceptible to the McGurk effect, or if there was simply no difference. This

experiment was also testing to see if one population was more partial to either auditory or visual information when compared to the other.

Musicians and Brain Distinctions

Professional musicians have been found to possess larger amounts of grey matter compared to amateurs and non-musicians. Grey matter contains a large amount of neuronal cells, and the increased amount of these cells in musicians' brains can be attributed to the repetitive usage of a musician's auditory and visual brain regions. Practicing the memorization, coordination, and improvisation that comes with playing any instrument well has long-term effects in a musician's brain, like his or her remarkable sensorimotor and auditory-spatial skills (Gaser & Schlaug, 2003)¹. These skills can be applied to other everyday activities, specifically activities that require the usage of one's auditory, visual and motor cortexes. Carrying a conversation with someone requires work from all three of these cortexes (Scott, McGettigan & Eisner, 2009).

Audiovisual Integration

The brain has a tendency to fuse auditory and visual information from speech to generate a singular idea of what was being said. This idea is exhibited in experiments using the McGurk effect, in which subjects commonly reported hearing a fusion of the given incongruent stimuli. Although the brain depends more on auditory information, or what is heard, than visual information, or what is seen, subjects still sometimes report hearing the visually presented phonemes given in the stimuli. (Gentilucci & Cattaneo, 2005)².

The brain also has been found to activate a frontal-temporo-parietal region of the brain related to speech production and perception when asked to comprehend stimuli both auditorily and visually. This was found to be important for accurately determining a speaker's words and speech sounds (Dick, Solodkin & Small, 2010).

Superior Temporal Sulcus

The STS is considered an essential part of speech perception, as neuroscientists have found that it has strong connections with the auditory cortex and helps to accurately distinguish words in speech-in-noise experiments (Nath & Beauchamp, 2011)². The STS has also been found to have a key role in the perception of the McGurk effect, as the strength of the response in the STS was what determined whether or not a subject was susceptible to the McGurk effect. The weaker the response from the STS, the less likely a subject would have been able to perceive the McGurk effect (Nath & Beauchamp, 2012)².

The enhanced auditory and visual skills of a musician and the brain's natural process of speech integration using the STS and various cortexes are all incorporated into this experiment, which seeks to determine if playing an instrument has any impact on audiovisual speech perception.

Methods

Recruitment and Participation

A total of 40 participants were chosen from a public high school in the New York Metropolitan area. Subjects were informally recruited for participation, and the subject's consent was given before they were subjected to the stimuli. In all, 20 males and 20 females in between the ages of 14 and 16 took part in this experiment. Twenty of these subjects were musicians with at least two years of experience and the rest were non-musicians with no experience with any

type of instruments. Subjects participated voluntarily and were informed that they were free to withdraw from the experiment at any chosen time.

Instruments

One control video and four stock videos (*Appendix A*) found on a popular video-sharing website were used as the audiovisual stimuli for this experiment. The control video, or congruent stimulus, was filmed by the experimenters, while the four stock videos, or the incongruent stimuli, were recorded by professional speakers. The control video contained a speaker saying the word “ga”, with no differing audiovisual stimuli. The incongruent stimuli contained four different professional speakers mouthing the word “ga” with the audio recording of the speaker saying “ba” dubbed onto the video. After individually giving his or her consent, each subject was presented with an iPad containing the stimuli and a pair of noise-cancelling headphones to ensure that there were no distractions. The subjects were not informed about the McGurk effect. Prior to the presentation of the stimuli, the subjects were instructed to answer three open-answer questions on the top of the page regarding their musical abilities and practice habits (*Appendix B*). They were also instructed to state their gender and age as a precautionary measure, in case those two variables affected the results. After the videos started playing, the subjects were given time, in between stimuli, to answer five forced-choice questions about their perception of the videos (*Appendix C*). These questions were similar to the question seen in Figure 1. The name of the subject was of no importance, so the surveys were filled out anonymously.

After watching video 3, answer the question below as accurately as possible:

Which choice is the closest to the sound you heard?

Ba

Ga

Fa

Va

Figure 1. Survey Question Example. An example of a question in the survey given to subjects. The question in the figure was not used in the real survey. (Figure produced by author).

There was no scoring method for this survey. Rather, the results were put into multiple 2x2 Chi Square tests for further analysis and evaluation. This was done to find if musicians were less or more susceptible to the McGurk effect, or partial one of the two senses (auditory and visual) relating to the McGurk effect.

Procedure

Script

Prior to watching the video, subjects were read a script (*Appendix D*). The script informed the subjects that their participation was voluntary and withdrawal from the experiment was allowed at any time, as well as the fact that their grade was in no way affected by their participation. Instructions about the procedure of the experiment were given after the script.

Instructions

The subject was given a device containing the stimuli but were not allowed to watch the stimuli prior to answering the pre-stimuli questions on the survey, which helped determined the musicianship and skill of the subject.

Analysis

The subjects' responses were analyzed through seventeen 2x2 Chi Squares. Each Chi Square compared the ratio of one specific type of response to another in both musicians and non-musicians. No attempt to establish validity or reliability was made. The surveys that were either incomplete or incorrectly filled out were not included in the analysis. After data collection, SPSS was used to analyze the Chi Squares.

Results

Chi-Square Analysis

Because the statistical test of choice, the Chi-Square test, required frequency data rather than simple numerical values, the post-stimuli questions were forced-choice questions, with one choice containing the correct auditory interpretation, another choice containing the correct audio interpretation, and two other choices containing McGurk interpretations. Means and error bars were not part of the results because of the analysis of choice. Since there were five videos and forty subjects, each subject's survey responses were entered five times into SPSS, once for his or her response to the control video, once for his or her response to the first incongruent stimuli, once for the second, and so on. Therefore, in terms of the experiment, every subject was counted as five subjects, and figuratively speaking, there were a total of 200 subjects. The response data was then categorized into three groups: auditory interpretations (a), visual interpretations (v), and McGurk interpretations (m). 2x2 Chi-Squares were then set up to compare the ratio of McGurk responses to visual responses, McGurk responses to auditory responses, and visual responses to auditory responses in musicians and non-musicians in each individual stimulus, which accounted for 15 out of 17 tests (*Appendix E*). The other two were set up to compare the ratio of McGurk responses to either auditory or visual responses in all stimuli. These two tests were considered the main tests because they compared the total quantities of each response, while the other fifteen were considered secondary tests because they compared smaller amounts of data. The two main tests both produced a p-value of less than .0005, meaning that musicians did perceive incongruent data differently than non-musicians. The secondary tests produced a wide range of p-values, from less than .0005 to more than .9555. Table 1 is an example of a secondary Chi-Square.

Table 1. Chi Square Analysis of Stimuli 2. The auditory responses were excluded in this test in order to create a 2x2 Chi-Square. Picking the auditory response meant that one heard what was said, picking the visual response meant that one heard what was mouthed, and picking the McGurk response meant that one heard a mix. Since $p = .350$, the differences were not significant. (Table by author).

	Musical Status		
Response Type	Musician	Non-Musician	Total
McGurk	2	9	11
Visual	2	3	5
Total	4	12	16
$p = .350$			

Main Analyses

Additionally, musicians were found to be less susceptible to the McGurk effect, with only 25.7% of musicians' responses being McGurk interpretations, in comparison to 52.2% of non-musicians' responses being McGurk interpretations. Figure 2 displays the staggering difference between the amount of musicians' McGurk responses and non-musicians' McGurk responses.

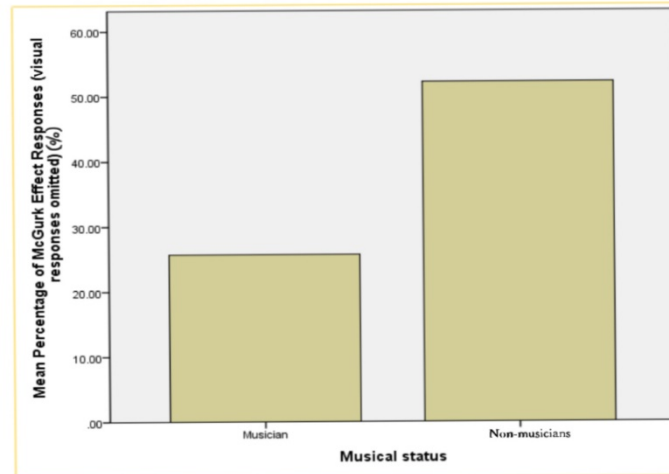


Figure 2. Bar Graph of Musicians' and Non-Musicians' McGurk Responses. It is apparent that there is a substantial difference between the quantity of non-musicians' McGurk responses and the quantity of musicians' McGurk responses. (Graph by author).

Both groups were heavily reliant on auditory information, but musicians were found to be much more reliant on auditory information than non-musicians, with audio responses (when subjects heard what was being said) accounting for 68.8% of all musicians' responses and only 40% of all non-musicians' answers, and while visual responses (when subjects heard what was mouthed) were quite rare, it can be deduced that non-musicians were slightly more dependent on visual information, with 7.8% of non-musicians' responses being visual interpretations, and only 5.5% of all non-musicians' responses being visual interpretations. Given that the p-value for both main tests was less than .0005, it can be concluded that these differences were most likely not due to chance (Table 2).

Table 2. The two main Chi-Squares. The two main Chi-Squares were the ones comparing the ratios of McGurk responses to auditory responses and McGurk responses to visual responses in all of the stimuli. 2a compares McGurk responses to auditory responses, while 2b compares McGurk responses to visual responses. The p-values of <.0005 mean that the differences between the perceptions of musicians and the perceptions of non-musicians were statistically significant. (Tables by author).

(Auditory)	Musical Status		
Response Type	Musician	Non-Musician	Total
McGurk	19	55	74
Visual	35	32	67
Total	54	87	142
P<0.0005			

2a

(Visual)	Musical Status		
Response Type	Musician	Non-Musician	Total
McGurk	19	6	25
Visual	35	13	48
Total	54	19	74
p<0.0005			

2b

Discussion

Determinations

With a p-value of less than .0005 on both main tests as well as several secondary ones, it can be inferred that musicians perceive the McGurk effect differently than non-musicians, and have a different susceptibility to the McGurk effect than non-musicians. It can also be deduced from the data that musicians are less susceptible to the McGurk effect and more reliant on auditory information than non-musicians are, with ratios of 25.7(musicians):52.2(non-musicians) for the percentage of McGurk responses in all stimuli and 68.8(musicians):40(non-musicians) for the percentage of audio responses in all stimuli. Non-musicians are also somewhat more dependent on visual information, even if there were quite few visual responses, with a ratio of 5.5(musicians):7.8(non-musicians) in all stimuli.

Relation to Previous Research

The conclusions of this experiment go hand in hand with the results of previous experiments. Subjects did tend to rely more on auditory information, as stated by Gentillucci and Cattaneo (2005)³, and musicians did demonstrate superior hearing skills, as stated by Gaser and Schlaug (2003)². However, musicians depended less on visual information than non-musicians, which went against an aspect of Gaser and Schlaug's results; they stated that musicians had both superior hearing *and* vision, whereas the musicians in the experiment only had superior hearing.

Implications

The results imply that musical training *does* improve auditory senses more than visual senses, and that non-musicians' auditory cortexes are weaker than those of musicians. The results also imply that non-musicians are forced to rely more on visual information, when compared to musicians, due to their subpar hearing.

Conclusion

The McGurk effect is a perceptual phenomenon that occurs when a person is subjected to incongruent audiovisual information. Musicians, who have been thought to have more grey matter in their brains as well as stronger auditory cortexes when compared to other populations, were subjected to incongruent stimuli. Their interpretations of incongruent stimuli were compared to the interpretations of non-musicians, and a test was run to see if there was a difference in susceptibility to the McGurk effect between musicians and non-musicians. There was indeed a difference in susceptibility; musicians were found to be less susceptible to the McGurk effect. Musicians were found to be more reliant on auditory information (one non-McGurk interpretation) when compared to non-musicians, while non-musicians were found to be more reliant on visual information (the other non-McGurk interpretation) when compared to musicians.

Limitations

There were several limitations in this experiment. The lack of variety in the stimuli might have negatively impacted the accuracy of the results, because the usage of only one audiovisual pairing (“ba/ga”) in the stimuli meant that the results of this experiment might not be applicable to other audiovisual pairings. The smaller age range might also impact results, since the hearing and sight of a 35-year-old, for example, might not be as pristine as that of a tenth-grader’s, meaning that the results might also not be applicable for those in a different age group. A third limitation would be the fact that there was no attempt to establish validity or reliability. A fourth limitation may be the lack of professional training of the musicians in the utilized population, since the minimum number of years of musical training, two years, is hardly enough time to become somewhat of a decent musician, and the presence of professional musicians in the experiment would have made the results seem much more concrete.

Acknowledgments

I would like to thank Dr. Alan Schorn for his assistance in carrying out this experiment and analyzing the data.

References

- Burnham, D. & Dodd, B. (2004). Auditory-Visual Speech Integration by Prelinguistic Infants: Perception of an Emergent Constant in the McGurk Effect. *Developmental Psychobiology*, 45, 204-220. Retrieved February 6, 2019 from Google Scholar Database. <https://onlinelibrary.wiley.com/doi/abs/10.1002/dev.20032>
- Dick, A.S., Solodkin, A. & Small, S. L. (2010). Neural Development of Networks for Audiovisual Speech Comprehension. *Brain Language*, 113, 101-114. Retrieved February 6, 2019 from Google Scholar Database. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2891225/>
- Gaser, C. & Schlaug, G. (2003). Brain Structures Differ between Musicians and Non-Musicians. *Journal of Neuroscience*, 23, 9240-9245. Retrieved February 6, 2019 from Google Scholar database. <https://www.jneurosci.org/content/23/27/9240>
- Gentilucci, M., & Cattaneo, L. (2005). Automatic Audiovisual Integration in Speech Perception. *Experimental Brain Research*, 167, 66-75. Retrieved February 6, 2019 from Google Scholar database. <https://psycnet.apa.org/record/2008-04784-007>

- McGurk, H. & MacDonald, J. (1976). Hearing Lips and Seeing Voices. *Nature*, 264, 746-748. Retrieved October 28, 2018 from Google Scholar database. <https://www.nature.com/articles/264746a0>
- Nath, A.R. & Beauchamp, M.S. (2011). Dynamic Changes in Superior Temporal Sulcus Connectivity during Perception of Noisy Audiovisual Speech. *Journal of Neuroscience*, 31, 1704-1714. Retrieved February 6, 2019 from Google Scholar database. <https://www.jneurosci.org/content/31/5/1704>
- Nath, A.R. & Beauchamp, M.S. (2012). A Neural Basis for Interindividual Differences in the McGurk Effect, a Multisensory Speech Illusion. *Neuroimage*, 59, 781-787. Retrieved February 6, 2019 from Google Scholar database. <https://pubmed.ncbi.nlm.nih.gov/21787869/>
- Scott, S.K., McGettigan, C. & Eisner, F. (2009). A Little More Conversation, A Little Less Action – Candidate Roles for the Motor Cortex in Speech Perception. *Nature Reviews Neuroscience*, 10, 295-302. Retrieved February 6, 2019 from Google Scholar database. <https://pubmed.ncbi.nlm.nih.gov/19277052/>
- Sekiyama, K. (1997). Cultural and Linguistic Factors in Audiovisual Speech Processing: The McGurk Effect in Chinese Subjects. *Perception and Psychophysics*, 59, 73-80. Retrieved February 6, 2019 from Google Scholar database. <https://link.springer.com/article/10.3758/BF03206849>