2017-05-01

Exploring McGurk Effect Through Tadoma Method Of Speech Perception

Ashley Moore
Salem State University

EXAMINING THE MCGURK EFFECT ON THE TADOMA METHOD OF SPEECH PERCEPTION

Honors Thesis

**Presented in Partial Fulfillment of the Requirements for the
Degree of Bachelor of Science in Psychology**

In the Department of Psychology at Salem State University

By
Ashley Moore

Dr. David Gow
Faculty Advisor
Department of Psychology

Commonwealth Honors Program
Salem State University
2017

Table of Contents

Abstract

The process of understanding speech perception is one that also poses a handful of questions. The motor theory of speech perception was proposed to resolve some of these issues that arose. While this accounts for things like coarticulation, the evidence for both sides of the fence is mixed. The discovery of mirror neurons and the findings of fMRI studies support the motor theory, whereas work done with Japanese quails tends to contradict this. The current study explores the speech phenomenon the McGurk effect, what happens when an individual fuses the speech sounds they hear with the one they see being articulated. The problem that comes up is whether the McGurk effect is a result of articulatory cues or training. To examine this claim in a different way than in the past, eighteen subjects were taught the speech perception method of Tadoma. After two days of training, the participants were given mismatched sounds that would potentially result in a McGurk effect. The number of fused responses from subjects increased after the training; however the statistic was not significant. Therefore, it can be concluded that while more of a McGurk effect did in fact appear after the two days of training, it was not valuable enough to prove that speech perception is a product of experience and not just articulatory information.

Introduction

How is it that people know exactly what someone else is saying? How is that only humans are capable of interpreting speech sounds? Researchers have set out to resolve many problems that have come up, one of which is the lack of invariance. No two speech sounds will ever sound exactly the same. That is to say, there is no one to one mapping between speech sounds and articulation. The brain's process of coarticulation is a reason for this lack of invariance. Coarticulation refers to how one speech sound is affected by the proceeding speech sounds. Without even consciously realizing it, an individual will generate any given phoneme differently depending on the surrounding phonemes in a word. To get a depiction of this, refer to Figure 1. The image shows formants for both the sound /di/ and /du/. Take notice of the drastic dissimilarity of each syllable if you were to cut off just the /d/ part of the sound. What this exemplifies is how coarticulation results in speech sounds that appear to be the same, but in fact are articulated rather differently. Gestures, however, do not vary. Every time any individual makes a /ba/ sound, their mouth moves in the same way. Anytime someone makes a /da/ sound, the corresponding gesture is exactly the same as it was to produce the previous one, and the next one, and the next one. This paradigm was originally thought to have been solved by the introduction of the motor theory. The motor theory of speech perception, brought to the forefront by Alvin M. Liberman and Ignatius G. Mattingly, suggests that speech perception relies on an individual's knowledge of speech articulation (Liberman & Mattingly, 1985). It's proposed that people perceive spoken words by recognizing the articulatory gestures with which they are pronounced. The underlying question however, and the basis of the current research, is whether or not individuals use these cues to

trigger their own representations of speech articulation. The contrasting side being that individuals merely have become habituated to them over time with experience. The motor theory proposes that it is a person's familiarity with the actual articulatory information that cues them to perceive language. Theoretically though, it could simply be the training humans have with the spoken word that cues them. Therefore, it is possible that with enough training, any arbitrary stimulus could replace articulatory signals and trigger a similar response.

Evidence in support of motor theory is extensive, including work done with mirror neurons and fMRIs. While conducting work on Macaque monkeys, researchers found something they called "mirror neurons" (Rizzolatti et al., 1988). Mirror neurons are important to discuss, as they are the backbone of understanding how an individual can produce articulatory gestures based on the speech they are able to perceive. The critical discovery was found in area F5 of the monkeys, which shares a lot of similarities to Broca's area in the brains of humans. Broca's area remains the part of the brain primarily responsible for language processing in humans. This subsequently suggested that humans have these mirror neurons that function similarly to how they do in monkeys. Rizzolatti and his colleagues' study involved examining the monkeys' control of hand movements, such as grasping and holding. Researchers observed that there were a set of neurons that not only fired when the monkey was actively doing a task (i.e. grasping), but additionally fired when they witnessed the human doing the same thing. If in fact there are such things as mirror neurons in Broca's area of human brains, it can be assumed that they would function correspondingly. Namely, a set of neurons that fire when a person produces any given speech sounds, would hypothetically also fire when they heard those speech

sounds. Mirror neurons could then evidently be the connection between the speech an individual perceives and the gestures through which they produce it. It is presumably the mirror neurons that make it so an individual hears a phoneme and subsequently knows exactly how to make it themselves.

Similar to the finding of mirror neurons, studies using fMRIs found data that demonstrated the overlap of the brain regions involved in speech production and speech perception. In a 2006 study, researchers looked at fMRI activity when subjects spoke and listened to certain sounds (Pulvermuller et al., 2006). They gave the participants phonemes such as /p/ and /t/, in which they would need to move their lips and tongues to produce. Results of the fMRIs showed that when an individual was perceiving speech, the largest amount of activity was in the superior temporal complex. Interestingly though, there was a noticeable activation of the same brain regions when a subject was listening to the speech sounds as there was when they articulated them.

While there is far-reaching support for the motor theory of speech perception, like most of the scientific world, it is not universally supported by evidence. There are other accepted explanations for the interpretation of speech though. Take for instance the Japanese quail. Research has been done on these kinds of birds and was found that they can actually understand speech similar to humans (Liberman & Mattingly, 1985). So here is an animal that has no ability to produce the speech sounds, but can still perceive it just like a human would be able to. In later years, a study was conducted using four Japanese quails that were trained to identify sounds such as /da/ and /ga/ (Lotto, Kluender & Holt, 1997). What they found was that these quails were demonstrating similar data patterns as

humans did, indicating that speech perception may not be limited to species that have the

ability to articulate the sounds themselves, namely humans.

   The McGurk effect is a psycholinguistic phenomenon that demonstrates how non-

acoustic information can affect speech perception. The perceptual effect was first

discussed by Harry McGurk (McGurk & MacDonald, 1976). The first step in the

demonstration is having an audio sound of someone saying "ba" for example. Separately,

the participants have a visual of a one person mouthing "ba" and another mouthing "ga."

Even though, the actual sound has not changed, the participants recall hearing a blended

sound of the two, i.e. "da." These kinds of experiments are of great importance to the

world of psychology and of language in general. What it does is showcase the fact that

speech perception is a multisensory task. We don't just use auditory information when

interpreting speech, but rather we use auditory information integrated with visual

information. This was an extremely significant finding because before that it was

assumed that speech perception was very much a product strictly of auditory input. In

addition to the obvious takeaway that the McGurk effect demonstrates the merging of

information from two vital senses in the process of speech perception, it also has other

noteworthy implications. The findings of McGurk effect studies suggest that even people

that don't hear well, or can't hear at all, still have the cognitive ability to perceive speech

to some degree. "To some degree" being the key phrase here, as there are quite a bunch

of speech signals that do not generate cues that can be distinguished visually. Even when

it comes to reading lips, people still rely heavily on context. Speech perception as a

whole is dependent on visual input as well as other forms of input, such as auditory and

tactile, which is the primary focus of this research.

The McGurk effect looks at how the speech perception process takes into account auditory and visual information, not just one or the other. It has been offered that the blending of auditory and visual information have an impact on the way speech is perceived. What this current study is after, is whether or not the same impact will be seen when the stimuli is non-visual, but rather tactile. The method of speech perception known as Tadoma is one that dates back to the early 20th century (Alcorn, 1932). It was invented to cater to deaf-blind individuals. Tadoma is performed by placing fingers along the speaker's lips, jaw line, and throat to interpret sounds (see Figure 2). If the ideas put forth in the motor theory are in fact correct, then the tactile cues decoded through Tadoma, should also present a McGurk effect. A contrasting point is that the McGurk effect shows that speech perception does not happen because of an individual's articulatory representations, but rather because visual cues that are greatly correlated. If this were true, it would insinuate that using Tadoma should not create a McGurk effect.

With that being said, there have been a handful of studies done trying to establish a McGurk effect using Tadoma. Not much evidence though has been found to support the claims that are currently being put forth; however, most of the studies that have been conducted have used subjects that have been trained in the Tadoma method. One of the few experiments did find a McGurk effect though in a population of individuals that were normal seeing and hearing individuals (Fowler & Dekle, 1991). This is contrary to much of the other research that has been done. Minimal evidence has been made apparent that confirms a McGurk effect occurs with Tadoma. The proposed reason for why most researchers do not see a McGurk effect is the lack of training, despite Fowler and Dekle's findings. To remedy this problem, the objective of this current study is to provide more

Tadoma training for the subjects. Because of this, every participant will receive two days of training before being tested. Previous studies have come up short when it comes to finding a McGurk effect, in part because of the fact that they have not allowed memory processes like sleep consolidation to occur. It is believed that individuals who are better trained in Tadoma will be more likely to display the typical speech perception characteristics, which one would see with normal audio-visual information. This is going to be the ambition of this study being conducted, something that is presumed to make a difference in the results of Tadoma/McGurk experiments.

The goal of the current research at hand is to use previous findings on Tadoma generated speech perception to examine the McGurk effect in a new light. This study first requires training of Tadoma for all participants. Every participant will go through two days of training before the day of testing. This is to ensure that sleep consolidation of the material occurs. Day one of training presents the participants with strings of sounds that they must identify. Each trial will be repeated with a different background noise level, including one with no noise at all. Day two of training will consist of the same procedures as day one, with the exception of the stimuli with no noise. It is the goal that between day one and day two, the participants' accuracy will improve. It's not until the testing that mismatched sounds will be presented. The participant must quickly choose what he or she heard when given the audio of one sound and the tactile information of another. If a participant recalls hearing a blended response, it is indicative of a McGurk effect. It is hypothesized that a McGurk effect will be seen when using Tadoma, which in turn exemplifies how speech perception is a response to experience and learning, not merely articulatory action. When using Tadoma, if a McGurk effect appears it can be

assumed that it is not due to a subject's knowledge of it. Because none of the subjects

have prior understanding of Tadoma, it is not a like they would have been able have that

as a representation of articulation. Instead, if they perceive a fused response, it is likely

from the training they will be given.

<div align="center">Methods</div>

Subjects

Participants were gathered from a variety of outlets, including the Honors

Program at Salem State University and through the use of Facebook. Of the eighteen

subjects in the study, all had "normal" hearing and vision capabilities. Not one of the

participants reported having any prior knowledge or training in the method of Tadoma.

All were learning for the first time simultaneously during this study. Participants were not

limited to Salem State University students, but were required to be over the age of

eighteen. The ages of all of the subjects ranged fairly wide from nineteen to fifty-one.

The mean age though came out to be approximately twenty-two years of age. Out of the

eighteen participants, there were thirteen females and five males that took part in this

research. No coercion occurred to gather participants, rather every individual agreed to

join the study on their own free will. The sample of subjects examined is not perfectly

representative of the entire population, as it was more a matter of accessibility.

Furthermore, it is important to note that all research was approved by the Institutional

Review Board at Salem State University in December of 2016.

Stimuli

All stimuli for this study were generated on a Windows computer using the Praat

program for phonetics. Participants heard speech sounds that were made into longer

strings of sounds. The sequence of speech sounds, which were personally recorded on

Praat, were then overlaid with a recording of white noise gathered off of iTunes. The

phonemes were as follows: /ba/, /da/, /ga/, and /pa/, /ta/, /ka/. Each subject was given a set

of headphones to be able to listen. The person verbalizing the sounds heard through the

headphones, was the same person who delivered the tactile stimuli. There were four

different ways the strings of stimuli could be presented, depending on the signal-to-noise-

ratio that was assigned. A lower signal-to-noise ratio, the harder it is to make out the

sounds. The first group of stimuli had no background white noise at all. The second

group was at an SNR level of 1.0. Another group was at a level of 0.5. The last group,

which was the set of stimuli used in the testing condition, was fixed at an SNR of 0.1.

The fact that the SNR was set at lower and lower levels each trial, indicates that the

power of the white noise was being increased, and consequently made it more difficult to

distinguish the speech sounds.

Procedures

      The Tadoma technique was taught by the researchers of the experiment, who were

familiar with the method. It was imperative that the syllables were mouthed

simultaneously to the sound of them being produced over the headphones. This was to

present the illusion that the talker was emitting those sounds at that moment. Those that

exhibited knowledge on how to properly do it, would continue to go on with the trials.

Eventually they graduated to distinguishing speech sounds with the presence of white

noise in the background. This is to ensure that the participants were relying on the tactile

training rather than anything else. As the training goes on, the signal to noise ratio

increased. The last step to the experiment was the introduction of mismatched sounds and

gestures with the presence of the noise. Participants were given a sound simultaneously with an articulatory gesture that was not in accordance to the former. The first day of the training regimen goes as follows. The individual will initially hear a string of sounds in a random order, with no noise in the background, and will be asked to elicit what they hear. This will continue with another string of sounds in random order, with slight noise in the background. The same will be done for three additional sets of stimuli, which present an increased noise level until it is so loud, the participants can barely make out the sounds.

<div align="center">Results</div>

It was presumed that a McGurk effect would present itself with enough practice of Tadoma. A total of five fused responses were recorded prior to any Tadoma training at all. More instances of fused responses were seen however within in the testing condition, after two days of training. The total number of occurrences after the two days of training was nine. It is important to note that both the pre-training trial and post-training trial consisted of the same amount of mismatched cues that could potentially result in a McGurk effect. This was to ensure that the frequency of fused responses was based solely on the difference of the amount of practice each subject had, rather than more opportunities of doing so. Throughout each of the trainings, accuracy levels were also being recorded. As the progression of the SNR went down, so wouldn't the accuracy levels. The goal was to find an SNR level that would make it so participants would struggle to hear the speech sounds clearly, having them rely more heavily on Tadoma. As Table 1 and Table 2, located in the appendix, display, accuracy levels for each condition decreased, indicating to an observer that the subjects were beginning to depend more on

the tactile information given to them through Tadoma rather than just the auditory information they were hearing.

Some mismatched combinations provided more fused responses than others. Table 3 depicts each of the different auditory and tactile groupings that were given to participants in the testing condition. It is clear that while most subjects perceived hearing the auditory speech sound, there is a contradiction between the velar and labial phonemes. Results show that tactile information seemed to play more of a role with labial speech sounds, i.e. /b/ and /g/.

While all of these results provide pertinent information, arguably the most important thing to look at is whether or not a McGurk effect is more likely to occur with advanced training. This is the one thing that sets apart this study from most others. To examine this claim, a T test was performed. It was found that after two days of training, the frequency of a McGurk effect was in fact higher than no training at all, $t(17) = -1.11$, $p = .282$, not significant. The outcome of the T test does show that occurrences of the McGurk effect would increase with experience; however, it cannot be concluded that it is that experience that is responsible for speech perception matters such as the McGurk Effect. With that being said, all of the results indicate that more research needs to be done to provide evidence that supports the hypothesis.

Stimuli		Responses			
Auditory	*Tactile*	*Auditory*	*Tactile*	*Fused*	*Other*
/ba/	/ga/	13	1	4	0
/ga/	/ba/	11	4	3	0
/pa/	/ka/	11	3	2	2

/ka/	/pa/	10	7	0	1
/ba/	/ta/	13	5	0	0
/da/	/pa/	12	3	0	3

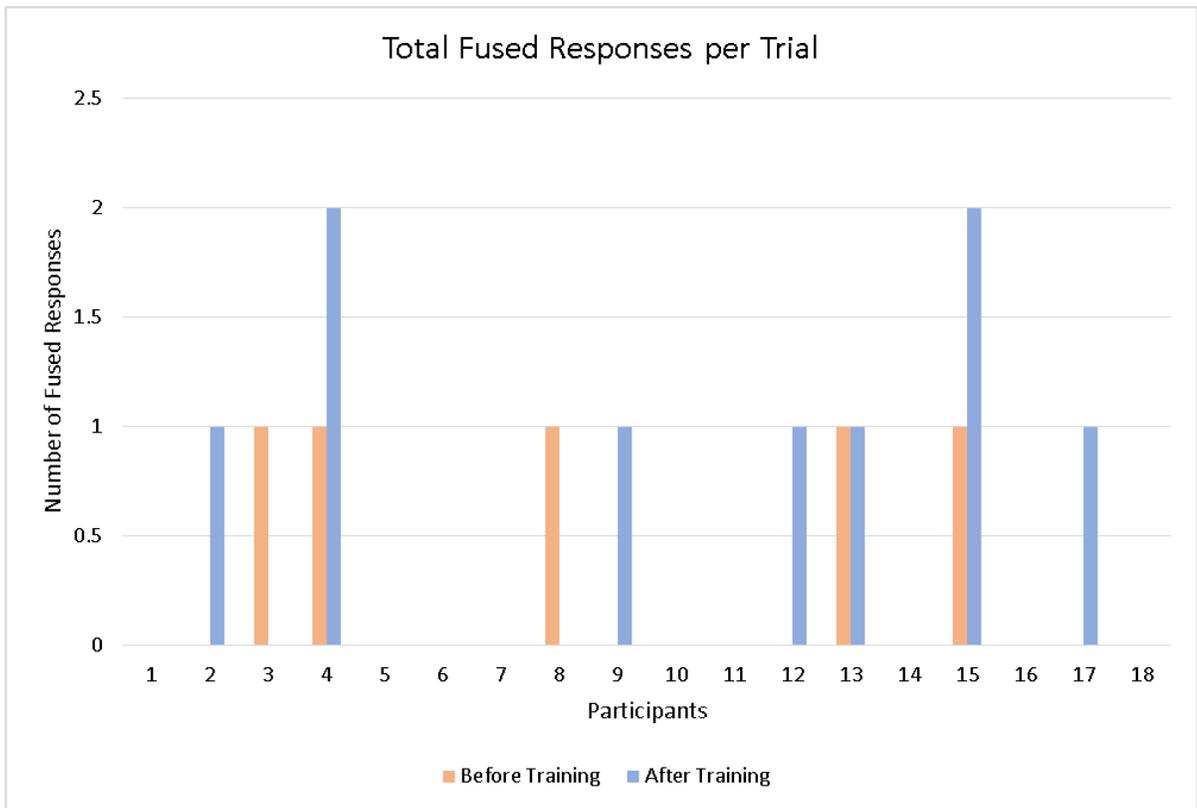**Table 3.** Subjects' responses for testing condition

Figure 3. McGurk effect pre and post-training

Discussion

Unlike previous Tadoma studies, participants in this study were given two days of training before being tested. The hypothesized outcome was not necessarily found to be true; however, there are some possible explanations for why this was the case. It is important to understand the reasoning behind any and all findings that were brought to light throughout the analysis of the data.

The most conspicuous result of this study comes from the T test that was performed. It was discovered that while there was in fact an increase in amount of fused responses after training, it was not significant enough to be of value. Thus, it cannot be concluded at this time that speech perception, manifested through the McGurk effect, is solely a direct byproduct of an individual's amount of experience. This does contradict some prior thinking, but it falls more in line with the belief that speech perception is accomplished through articulatory information. There are a number of reasons as to why this was the result found and not what was previously expected.

One explanation could be that the participants were not given enough training for an effect to be visible. Subjects in this study had a requirement of two days of training on Tadoma prior to testing for a McGurk effect again. Although two days of training is exceedingly more than similar studies that were previously conducted, there is a possibility it was not enough. In order for the subjects to have a firm grasp of tactile methods of speech perception, perhaps more training should have been provided.

Another shortcoming within this study that certainly could have affected the direction of the results found, is the sheer number of participants. Eighteen individuals were examined, which does not give this research a weighty *N,* number of people in the sample. As *N* increases, so does the significance of any statistical value that may be calculated. This is something that is applicable here. Having a greater number of participants likely makes the sample more representative, and in turn breeding more accurate statistics.

As stated, in regards to Tadoma training and the McGurk effect, the null hypothesis must be accepted for the time being. "For the time being," serves as the key

message of that statement. The lack of a meaningful outcome only means more research needs to be done. It is imperative that steps are taken in the future to continue to explore these topics. In addition to replicating this same study again, hopefully with more subjects, a good follow up that will be proposed is a study essentially mirroring this one, only with arbitrary non-articulatory cues. To further test whether speech perception is a byproduct of articulatory information or of the training and experience an individual has with speech, non-articulatory cues, such as colors, will take the place of Tadoma. The methodology should remain; however, every time a subject hears /ga/ he might see the color blue. Every time he hears /ba/, the color red could appear. The goal would be to see if the same sort of McGurk effect would present itself. If so, it would be support for the notion that speech perception is actually derived from an individual's experience and not their knowledge of articulatory cues. If this was found to be true, it would first off, oppose the result of the current study, but would also challenge the motor theory.

References

Alcorn S. (1932). The Tadoma method, *Volta Rev. 34*, 195–198.

Fowler, C.A., Dekle, D.J. (1991). Listening with eye and hand: cross-modal contributions

to speech perception. *Journal of Experimental Psychology, Human Perception*

and Performance, 3, 816-828.

Liberman, A. M., Mattingly, I. G. (1985). The motor theory of speech perception

revised. *Cognition, 21*(1), 1-36. doi:10.1016/0010-0277(85)90021-6

Lotto, A. J., Kluender, K. R., & Holt, L. L. (1997). Perceptual compensation for

coarticulation by Japanese quail (Coturnix coturnix japonica). *Journal of the*

Acoustical Society of America.

McGurk, H., MacDonald, J. (1976). Hearing lips and seeing voices. *Nature, 264,* 746-

748.

Pulvermuller, F., Huss, M., Kherif, F., Moscoso del Prado Martin, F., Hauk, O., &

Shtyrov, Y. (2006). Motor cortex maps articulatory features of speech sounds.

Proceedings of the National Academy of Sciences of the United States of

America,

103(20), 7865–7870.

Rizzolatti G, Camarda R, Fogassi L, Gentilucci M, Luppino G, Matelli M. Functional

organization of inferior area 6 in the macaque monkey. II. Area F5 and the control

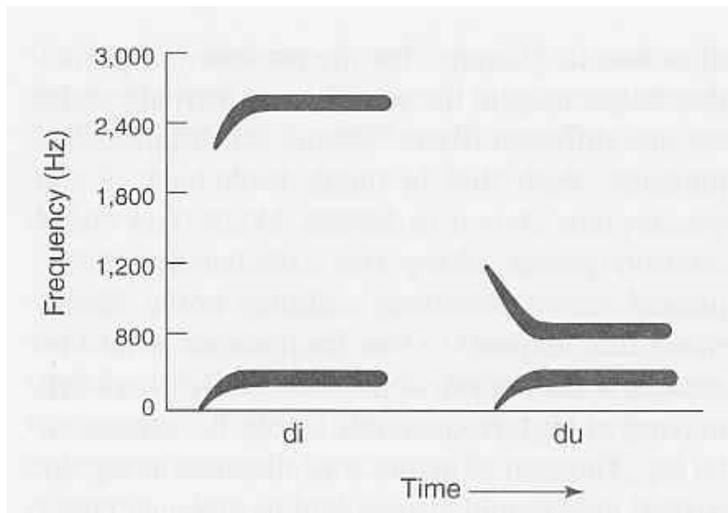of distal movements. Exp Brain Res. 1988;71(3):491–507.

Appendix

Figure 1. Spectrogram image of sound formants.

Figure 2. Tadoma method of speech perception

Auditory	*Tactile*	*Accuracy Level*
/ba/	/ba/	100%
/ga/	/ga/	100%
/pa/	/pa/	100%
/ka/	/ka/	94.44%

Table 2. Subjects' accuracy levels for training condition SNR 1.0

Auditory	*Tactile*	*Accuracy Level*
/ba/	/ga/	77.78%
/ga/	/ba/	83.33%
/pa/	/ka/	66.67%
/ka/	/pa/	77.78%

Table 3. Subjects' accuracy levels for training condition SNR 0.5