A BEHAVIORAL INTERPRETATION OF

THE MCGURK EFFECT

A Thesis Presented to the Faculty
of
California State University, Stanislaus

In Partial Fulfillment
of the Requirements for the Degree
of Master of Arts in Psychology

By
David Roth
December 2015

CERTIFICATION OF APPROVAL

A BEHAVIORAL INTERPRETATION OF

THE MCGURK EFFECT

Signed Certification of Approval Page is on file with the University Library

by
David Roth

Dr. William Potter Professor of Psychology	Date
Dr. Bruce Hesse Professor of Psychology	Date
Dr. Gary Williams Professor of Psychology	Date
Dr. David Palmer Professor of Psychology	Date

DEDICATION

This thesis is dedicated to the memory of Tony Scorsune.

ACKNOWLEDGEMENTS

This thesis is a culmination of my educational endeavors that would not have been possible without the help, support, and guidance of many individuals. I would like to thank my family for their persistent emotional support throughout my educational career. I would also like to thank two of my closest friends, Giancarlo and Tony, for each encouraging my return to school in his own unique way. An important shout-out is extended to my dear friend, TJ, for his support and participation in the initial pilot studies for the present research. I would like to acknowledge my analytic-twin, Patrick, for providing me with a reflection of my own behavior analytic passion. Thank you Stacey, Meredith, and Danie for your friendships and individual contributions to my thesis and graduate school experiences. I would like to thank my committee members Drs. Potter, Hesse, and Williams. And lastly, I extend a special thank you to my mentor, Dr. David Palmer, whose guidance throughout my thesis has been just one of his many inspirational contributions to my personal development as a "fire-breathing behaviorist."

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

FIGURE PAGE

ABSTRACT

The McGurk effect is a perceptual phenomenon in which the combination of discrepant visual and auditory speech stimuli (e.g. hear-ba/see-ga) produces reports of hearing a completely novel response form (e.g. “da”). The present study attempted to explain the McGurk effect and related phenomena in terms of principles of behavior. Skinner (1953) proposed that perception itself is *behavior,* and interpreting experimental results within the framework of experimentally validated behavioral principles may help to guide future research on perceptual phenomena. Additionally, the study contributed to the analysis of the McGurk effect by comparing the auditory and visually discrepant isolated syllable conditions (e.g. hear-BA/see-GA) with a second procedure, in which the same stimulus conditions were presented as the initial syllables of rhyming word pairs taken from the average educated adult English speaking repertoire (e.g. hear-BUST/see-GUST). Since the McGurk illusion is produced by auditory and visual stimuli that are discrepant in their corresponding labial or non-labial lip movements, the authors analyzed the effect by selecting two labial speech sounds “ba” and “ma,” and two non-labial speech sounds “ga” and “la.” The results support the authors’ hypothesis that the McGurk effect would be stronger when syllables were presented in isolation than when they were presented in the context of whole words. The second hypothesis, which proposed that the tendency to report hearing the word “dust” in the hear-BUST/See-NonLabial conditions would be stronger than the tendency to report hearing “nust” in the hear-MUST/See-NonLabial conditions was not supported by the present study.

Key words:

B.F. Skinner, classical conditioning, conditioned perceptual behavior, McGurk effect, multiple control, stimulus blocking, verbal behavior

INTRODUCTION

Every day, people are fooled into “seeing” or “hearing” things that are not actually present at the moment they are experienced. Take for example the following anecdote offered by B.F. Skinner (1980):

> I was sitting in the garden listening to music through earphones. A few drops of rain began to fall, and I could see circles forming on the pool at my side. A slight defect in the phonograph record made a click in one ear. I heard it as a drop of water striking the phone on that ear. When the rain stopped falling, I found myself turning to look at the pool to see if it had started again whenever I heard a slight click.” (p. 13-14).

How are we to interpret this illusion and countless others of the sort that we experience throughout our lives? Although the field of psychology has conducted numerous experimental studies on such phenomena, researchers are still at odds about just what processes are responsible for the perceptual events.

In 1976 two researchers, Harry McGurk and John MacDonald, discovered some peculiar results from their study on the role of lip movement in speech perception. Participants were shown videos of an individual repeating the sound “ga,” while the actual sound “ba” was dubbed over the videos in synchrony with the movements of the individual’s mouth. The authors had hypothesized that the discrepant auditory and visual stimuli would either conflict with each other or that one sense mode would simply dominate over the other. However, their discoveries were unexpected: Instead of interpreting the emitted sounds as being either “ba” or “ga,” many participants perceived the sound as “da.” This phenomenon, in which a novel response is made to a complex discrepant auditory-visual speech stimulus, has come to be known as the *McGurk effect.*

Bruner and Postman (1949) demonstrated analogous results in the perception of compound discrepant *visual*-visual stimuli by briefly presenting participants with playing cards whose colors were reversed from those found in a normal deck. For example, when offered a fleeting glimpse of a red six of spades, some participants reported having seen a purple six of hearts or a purple six of spades. Interestingly, these participants did not report the cards as being the color historically conditioned with a spade (black), nor did they report it as being the actual color presented to them (red), instead some of the participants reported “seeing” a *fusion* of the two.

In both the Bruner and Postman (1949) and McGurk and MacDonald (1976) studies, the directly measured behaviors were verbal responses, not private events, and the manipulated variables were objective stimuli. The challenge to the scientist is to explain why such stimuli should evoke such unusual verbal responses. What is the relationship between the particular manipulation of environmental stimuli in these experiments, the histories of the participants, and their recorded verbal behavior? Must we refer to unmeasured behavioral effects of the relevant environmental stimuli? To put it differently, must we invoke “perception” as a controlling variable? If so, how can we offer a scientific account of such a seemingly subjective phenomenon?

Range of Phenomena in the McGurk Effect

Although most people report experiencing the McGurk effect as described above, individual differences are found. Some participants, for example, report the auditory stimuli accurately, despite the discrepant visual stimuli. Children with developmental delays frequently do so (Meronen, Tiipana, Westerholm, & Ahonen 2013), and in a pilot study of the present research, two adult participants, who reported that they had been drinking alcohol, accurately reported the auditory stimuli. Other participants sometimes report "hearing" the sound normally corresponding to the lip movement alone. Moreover, the nature of the effect depends on the specific speech sounds being manipulated. Under some conditions *blends* rather than *fusions* are found. A blend is a mixture of the two relevant phonemes (e.g., *ba* and *ga* might yield *bga*), whereas a fusion is a novel sound that is topographically distinct from, yet controlled by, each of the discrepant stimuli (e.g., *ba* and *ga* might yield *da*).

Blended Responses

As revealed in the present study and many others before it, when the lip movement and acoustic stimuli of a speech sound are artificially dissociated, listeners report "hearing" a different sound altogether. Yet, whether the listener's response is reported as a blend or a fusion of the dissociated stimuli depends upon whether the presented lip movement is produced labially or non-labially, that is, with the lips closed or open. When the visual component of the complex stimulus is the speaker's labial movement the reported response tends to be a blend of the two stimuli. As noted above, when participants are presented with the lip-movement corresponding to the sound “ba” and the acoustic stimulus is the syllable “ga,” participants may report the spoken stimulus as “bga” (McGurk & MacDonald, 1978). This novel response form appears to be a result of the visual stimulus controlling the first part of the blend, while the acoustic stimulus controls the second.

Fused Responses

As we have just seen, the two consonants within a blended response appear to have point-to-point correspondence with the two discrepant acoustic and visual stimuli presented to the listener. However, when the discrepant complex stimulus is composed of a non-labial lip movement and a labially produced speech sound (e.g. hear-BA/see-GA) participants tend to report “hearing” a *fusion* of the two stimuli (“da”). It is this particular effect that has received the most attention in such studies, and has been famously classified as *the McGurk effect*.

Further Research on the McGurk Effect

Over the last three decades, the McGurk effect has been replicated and extended many times. For example, Meronen, Tiipana, Westerholm, and Ahonen (2013) explored the McGurk effect with children diagnosed with Developmental Language Disorder (DLD). They hypothesized that in degraded listening conditions, individuals with DLD would experience poorer auditory discrimination and thus be influenced more strongly by the discrepant visual stimulus, leading to a stronger McGurk effect. The authors established degraded listening conditions by manipulating what is referred to as signal-to-noise ratio (SNR); this is accomplished by either reducing the volume of the auditory output or increasing white noise over the produced auditory stimuli. With each of the degraded listening conditions, responses emitted by individuals with DLD were not influenced by discrepant visual stimulation. For example when the lip movement for “ga” was synchronized with the vocal production of “ba,” individuals with DLD still reported having heard “ba.”

The authors explored the possibility that poor visual lip-reading discrimination might have been a factor in these unanticipated results, which prompted them to conduct an additional study in which isolated lip-movements in the absence of sound were presented to account for visual discriminative abilities. When comparing lip-reading behavior in individuals with DLD versus individuals without DLD, those with DLD performed significantly poorer when visual stimuli were presented in the absence of auditory stimuli. These findings suggest that visual speech detection is a complex skill necessary for even typically developing individuals to be effective listeners within their environment, and that individuals with DLD might benefit from direct treatment in this area. Additionally, when auditory stimuli were presented in isolation, discrimination was nearly perfect across participants with and without DLD. The ambiguous nature of visual speech stimuli raises an important question regarding the strength of its discriminative control when an individual depends on the supplementary stimulation provided by lip movement as a listener.

Lidestam and Beskow (2006) explored the ambiguity of visual speech stimuli (i.e. lip movement) by analyzing the discrimination of specific phonemes each produced by visually similar lip movements. For example, the lip movements corresponding to the consonants “b,” “m,” and “p” are each produced by the labial contact between the upper and lower lips, whereas the visual response products of the emitted sounds “d,” “l,” and “n” are all produced in the absence of labial contact between the upper and lower lips. The authors found that there was weak discrimination of individual phonemes when presented visually. An error analysis of the participants’ responses demonstrates that, although participants had difficulty determining the precise phonemes that correspond to each subtle movement of the articulators, the errors that were made still corresponded to their respective labial and non-labial visual stimulus classes. For example, when presented with the visual “g” consonant, the participants had difficulty discriminating with accuracy that it was the consonant “g,” however the errors that were made corresponded to responses produced without the upper and lower lips making contact with each other. This research demonstrates that although there is strong visual stimulus control for discriminations of speech stimuli, without supplemental acoustic stimulation the fine-grained discrimination of visually produced syllables appears to be weak in normal hearing individuals.

Green, Kuhl, Meltzoff, and Stevens (1991) analyzed the McGurk effect using gender discrepancies with 44 undergraduate students. The authors manipulated the gender of the face (visual stimulus) presenting the articulation with an incongruent gendered voice dubbed over (e.g. a male face with a female voice or a female face with a male voice). The complex stimuli for the McGurk variables were the auditory stimulus “ba” presented with the visual stimulus “ga” and the auditory stimulus “ga” presented with the visual stimulus “ba.” The former compound stimulus historically produces the response “da,” whereas the latter historically produces a blend of the two “bga.” The dependent variable was the spoken response to a selection of five initial-consonant choices of what was “heard”: /b/, /d/, /g/, /th/, or /bga/. Results demonstrated that with gender discrepancies, the findings were consistent with historical McGurk findings (i.e. gender discrepancy had no significant impact on perceived speech sounds). The authors then replicated the study with novel participants and novel speakers dubbed over the presented visual stimuli and found no difference from the above study with novel voices. Finally, 10 new participants were selected to determine how well gender incongruities were discriminable within the presented stimuli from the two previous studies, and the participants were easily able to detect the gender differences in male-female discrepancies of presented speech.

As revealed in the above study, manipulating the auditory speech stimulus properties related to gender of the speaker had no influence over strength or weakness of the illusion. Soto-Faraco and Alsius (2009) explored the strength of the McGurk effect when the auditory and visual stimuli were presented with different temporal arrangements. The results showed that, whether the auditory stimulus was presented before or after the visual stimulus, participants were able to discriminate that there was a discrepancy in temporal contiguity. However, they still reported “hearing” a fusion of the acoustic and visual stimuli when presented within a temporal discrepancy of up to 500 milliseconds. The authors state that these findings “suggest that whatever mechanism serves multisensory integration, it is not a single monolithic process that encompasses all attributes of the stimuli at once" (Soto-Faraco & Alsius, 2009, page 585).

A Behavioral Interpretation of the McGurk Effect

One purpose of the present study is to attempt to explain the McGurk effect and related phenomena in terms of principles of behavior. The reason for doing so is that, irrespective of one's theoretical orientation, the principles of operant and classical conditioning are well established. These principles may shed light on what appears to be a mysterious empirical phenomenon.

The Distinction Between Interpretation and Experimental Analysis

In an experimental analysis, experimental control of all relevant variables is imposed. However, commonplace behavioral phenomena are often too complex for a researcher to control all of the relevant variables within a rigorous experimental design. The shaping of speaker and listener repertoires begins at such an early age of development that it would be impossible to determine, with experimental certainty, the specific learning histories that may underlie the McGurk effect. In such cases researchers must fill the explanatory gaps with scientific interpretation until such control is possible. Scientific interpretation is the plausible extrapolation of empirically validated principles to phenomena that do not permit a complete experimental analysis. According to Palmer (1991):

> The interpretation of complex phenomena in the light of empirically established principles lies in the middle ground between experimental analysis and mere speculation. Speculation is unconstrained, while interpretation is

> constrained by experimental analyses. Interpretation is useful in circumstances too complex or too vast to control experimentally, but where informal or incomplete data are available (p. 261).

There are many examples of scientific interpretation in the other sciences, such as evolutionary biology, astronomy, and geology, however Skinner's (1957) analysis of verbal behavior, may be the most ambitious example within the field of psychology. He tackles the many problems present in a scientific account of language with an "orderly arrangement of well-known facts, in accordance with a formulation of behavior derived from an experimental analysis of a more rigorous sort" (p. 11). That is, everyday observations about verbal behavior were interpreted in light of laboratory principles of behavior. Examples of such laboratory principles include not just operant and classical conditioning, but phenomena such as *multiple control*, which has two implications: "(1) The strength of a single response may be, and usually is, a function of more than one variable and (2) a single variable usually affects more than one response" (Skinner, 1957, p. 227). In the context of verbal behavior, Michael, Sundberg, and Palmer (2011) caution scholars that "if one fails to consider multiple control, one's interpretations of [complex] behavior are likely to be conspicuously inadequate" (p. 4). The principles of operant and classical conditioning and of multiple control will play prominent roles in the interpretation of the results of the present study. Although the independent variables presumably responsible for the conditioning that leads to blended or fused responses were not measured, a behavioral interpretation of the phenomena can suggest plausible variables that might guide future empirical research of related perceptual events.

Skinner's Interpretation of the Bruner & Postman Study

A particularly relevant example of a behavioral interpretation is Skinner's explanation of the results of the Bruner & Postman (1949) study (Skinner, 1953). First, he proposed that perception itself is *behavior*; that is, it is susceptible to the principles of operant conditioning, classical conditioning, generalization, stimulus blocking, and other behavioral processes that have emerged in laboratory studies. Therefore, he distinguishes between unconditioned perception, which is independent of our experience, and conditioned perception, which is modified by experience. He interprets experimental findings of this sort as illustrating conditioned seeing: As the result of a long history of seeing red hearts and diamonds, the shape of the symbol on the card elicits "seeing red," while the unconditioned response is to see a dark color. That is, the behavior is multiply controlled. Individuals' reports of seeing purple can plausibly be interpreted as an interaction of conditioned and unconditioned "seeing."

An analogous case can be made with respect to the McGurk effect. Speech sounds are ordinarily highly correlated with movements of the articulators. Both are reliable antecedents of the "perceptual behavior" of hearing the relevant speech sound. The auditory stimulus is a conditioned stimulus that divergently strengthens one set of responses and the visual stimulus is a conditioned stimulus that strengthens another. Within a Skinnerian framework, this convergence of more than one environmental variable evoking a single response is an example of behavior under multiple sources of control (Skinner, 1957; Michael, Sundberg, & Palmer, 2011). When the visual and auditory stimuli are artificially dissociated, individuals report a fusion of conditioned responses to the discrepant stimuli.

Proposed Extension to the McGurk Effect

Skinner (1953) functionally analyzed conditioned perceptual behavior "on the pattern of the conditioned reflex" (p. 266). Utilizing the principles of the classical conditioning paradigm, the behavioral interpretation begins by addressing the conditioned response to the visual stimulus (i.e. non-labial lip movement). When a normal-hearing individual responds to a speaker's production of the unconditioned acoustic stimulus "da" (as in "doctor," "doll," or "dog") the unconditioned response is *hearing* the sound "da." Since, throughout the lifetime of an experienced listener, the visual movement of the articulators is precisely and reliably paired with the sound "da," we can presume that seeing the movement of the articulators in isolation should elicit, or at least strengthen, the conditioned response "hearing da." Through this same process there are a number of different acoustic sounds paired with members of the non-labial stimulus class, including, but not limited to, "ga," "la," "na," "tha," and "ta" (see Figure 1). Therefore, the visual response products of a speaker's non-labial mouth shape will divergently strengthen at least all of these responses concurrently in an experienced listener. "If we now begin to introduce conditioned stimuli of a different form… we may be able to show the fusion of two distinct effects" (Skinner, 1953, p. 267).

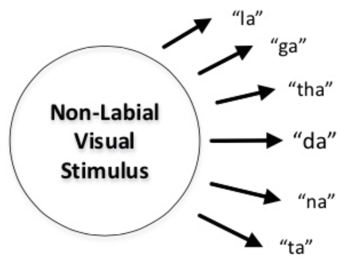

Figure 1. Divergent multiple control of a number of different acoustic sounds paired with members of the non-labial stimulus class.

A speaker's acoustic response of the syllable "ba" (as in "ball," "box," or "bomb") is the product of at least two fine-grained articulatory behaviors. The first is classified as a *stop* consonant, defined as the "complete obstruction of the outgoing airstream by the articulators, a build up of intraoral air pressure, and a release" (http://soundsofspeech.uiowa.edu). The second acoustic property is classified as *voiced*, which implies that the "sound is produced with vibration of the adducted vocal folds in the larynx" (http://soundsofspeech.uiowa.edu). There are only three English consonants that share the stop-voice properties in their vocal productions: /b/, /g/, and /d/. The isolated onset of a *stop-voiced* produced consonant may divergently strengthen "hearing" all three of these consonants, but since "da" is acoustically closer to "ba" than "ga" (Donahoe & Palmer, 1994), the strengthened response "hearing da" to the non-labial lip-movement, and the strengthened response of "hearing da" to the stop-voice acoustic stimulus converge to produce the so-called McGurk effect (see Figure 2).

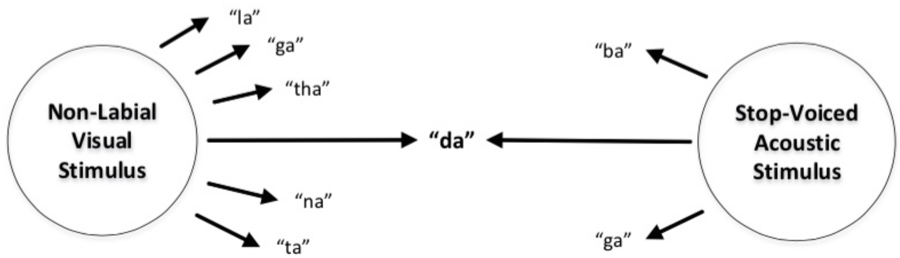

Figure 2. Convergent multiple control in which the strengthened response "hearing da" to the non-labial lip-movement, and the strengthened response of "hearing da" to the stop-voice acoustic stimulus converge to produce the McGurk effect.

This explanation of the McGurk effect concludes with one final piece of the interpretive puzzle. In ideal listening conditions (i.e. environments with no competing acoustic stimuli), normal-hearing individuals have no difficulty discriminating between the stop-voiced syllables "ba" and "da." However, the additive effect of a discrepant visual stimulus appears to obscure the finer grained acoustic properties responsible for the discrimination between the two sounds. In the phenomenon known as stimulus blocking, the most reliable predictor of reinforcement acquires control of behavior at the expense of competing stimuli (e.g., Kamin, 1969, vom Saal & Jenkins, 1970). In a noisy environment, the visual stimuli arising from a speaker's vocal apparatus are likely to be more highly correlated with the actual behavior of the speaker than the auditory signal, which may be mingled with irrelevant speech sounds. Due to the extensive conditioning history involved in the discrimination of lip movement, the discrepant visual speech syllable appears to "block" (Kamin, 1969) the listener's discrimination between the actual acoustic syllables, and leaves only the vague stop-voiced stimulus properties as the most salient acoustic variables available to the listener.

Skinner (1980) offers a description of how the sight of lip movements develop this potency throughout the lifetime of a competent listener:

> To some extent vocal verbal behavior affects the listener visually. We are less intelligible in the dark or on the telephone, given the same level of articulation. When auditory conditions are difficult, we amplify the visual properties of our speech. If there is no chance of being heard – if we are speaking above a loud noise, through thick windows, seen from a distance through binoculars, or to a friend at a distance when we must remain quiet, we may speak silently, with exaggerated movements of lips, jaw, and tongue. The exaggeration greatly exceeds any level previously established by differential reinforcement, but this is not surprising, since we also strike with a wholly unique force when weaker blows are ineffective, provided some differential reinforcement of more powerful blows has occurred. (p. 306).

The Present Study

The present study contributed to the analysis of the McGurk effect by comparing the acoustically and visually discrepant isolated syllable conditions (e.g. hear-BA/see-GA) with a second procedure, in which the same stimulus conditions were presented as the initial syllables of rhyming word pairs taken from the average educated adult English speaking repertoire (e.g. hear-BUST/see-GUST). Since the McGurk illusion is produced by auditory and visual stimuli that are discrepant in their corresponding labial or non-labial lip movements, the authors analyzed the effect by selecting two labial speech sounds “ba” and “ma,” and two non-labial speech sounds “ga” and “la.”

The authors hypothesized that the McGurk effect would be strongest in the Isolated Syllable condition when compared with the Whole-Word stimulus condition, since the Whole-Word condition introduces more restricting contextual variables related to the participants’ verbal histories. Additionally, the authors hypothesized that within the Whole-Word condition, the predicted illusion of perceiving “dust” in response to hearing “bust” dubbed over a non-labial lip movement was hypothesized to be stronger than perceiving “nust” in response to hearing “must” dubbed over a non-labial stimulus. Since “must” is a familiar word and “nust” a non-word, the former is likely to have greater strength in a verbal context. Analyzing the topographies of the participants’ responses with the lawful principles derived from an experimental analysis of behavior, the authors expand upon the behavioral interpretation of what is currently an “unknown level of processing” (McGurk & MacDonald 1978, p. 254) regarding the conditioned perceptual responses demonstrated in the McGurk effect.

METHOD

Participants

Twelve college students, 4 males and 8 females, between the ages of 18 and 30 were recruited for the study from California State University, Stanislaus. All participants had normal hearing abilities and either normal vision or vision corrected by prescription eyewear. Informed consent was obtained from each participant prior to running the experimental procedures. Each participant was randomly assigned to one of the two randomized sequences of video clips.

Materials and Setting

Each participant was seated approximately 44.5 cm. in front of a 33.02 cm. MacBook Pro computer screen in an office setting that was free of salient competing visual and auditory stimuli. Each participant was provided with a pencil and a sheet of paper with spaces in which they transcribed the trial number and their responses. Participants came into contact with auditory stimuli through built-in speakers that were connected to the MacBook Pro, and the participants contacted visual stimuli by orienting their eyes toward the computer monitor during each presented trial. The auditory-visual complex stimuli were constructed through the video editing software: iMovie.

Interobserver Agreement

All sessions produced permanent products of the participants' responses. The experimenters analyzed the topography of each response corresponding to each presented auditory-visual complex stimulus as they corresponded to the auditory stimulus, the visual stimulus, or some "other" stimulus independent of the auditory and visual stimuli presented. The experimenters also measured the frequencies of each topographical response. To determine interobserver agreement, one-third of the data were scored by an independent Board Certified Behavior Analyst (BCBA). The second observer was trained by the experimenter in the operational definitions of each variable within the current study and how to determine if a response corresponded to the McGurk effect or corresponded to the presented visual or auditory stimulus. IOA was calculated using the trial-by-trial method, in which the number of trials on which observers agreed was divided by the total number of trials. This proportion was multiplied by 100 in order to convert it to the percentage of agreements between the two observers. IOA was collected for 33% of all trials and the data collectors achieved 99.7% agreement on the calculations.

Design and Procedures

In this study there was one independent variable (stimulus type) with two levels, isolated syllables (IS) and whole-words (WW), across "complex pairs." *IS* refers to consonant-vowel speech stimuli that are presented outside of the context of a word (e.g. "BA" and "GA"). *WW* refers to consonant-vowel speech stimuli that are presented as the initial syllables of whole words (e.g. "BUst" and "GUst"). Previous research has indicated that the McGurk effect occurs only when the auditory and visual stimuli are discrepant in their labial or non-labial lip formations. *Complex pairs* refers to stimulus presentations, in which the auditory syllable and the visual syllable are discrepant in their corresponding lip movements. An example of a complex pair would be when the labially produced auditory syllable “ba” is produced in synchrony with the non-labially produced visual lip movement for the syllable “ga.” The primary variable of interest is the effect that each complex pair has on a listener’s behavior.

Dependent Variable

Using the pencil and paper provided, each participant was instructed to engage in the observing response of transcribing the trial number when presented on the screen, and then to write down what they heard. Each trial corresponded to a blank space for the trial number and a blank space for the recorded response on the sheet of paper provided. Each presented stimulus was separated by an inter-trial interval of about eight seconds in order to provide ample time for the recording of the responses. The dependent variable of interest was the number of responses that lacked correspondence to either the visual or the acoustic stimulus in the presented complex stimulus presentation. Each complex stimulus (16 in total) was presented 10 times, so the number could range from 0 to 10 for each stimulus.

IS Condition

In the IS condition, the stimulus presentations were analogous to the complex stimulus presentations from the original McGurk studies (McGurk & MacDonald, 1976; MacDonald & McGurk, 1978). The isolated labial syllables “ba” or “ma” were presented in synchrony with either the isolated non-labial syllables “ga” or “la,” each as either visual or auditory stimulus presentations. See Table 1 for all of the *IS* complex stimulus pairings.

Each participant was randomly assigned to one of two orderings of randomized video clips in order to counterbalance the order in which the two conditions were presented. Each participant was exposed to both the WW and the IS conditions and the order in which the two conditions was presented was also randomized and counterbalanced across participants. Each of the eight complex stimulus pairings was presented 10 times for a total of 80 complex stimulus trials. Additionally, 10 trials of each of the four non-discrepant pairs (e.g. hear-BA/see-BA) were presented intermittently throughout the sequences in order to reduce the participants' habituation to the discrepant pairs for a total of 40 non-discrepant pairs. Each trial was separated by approximately eight seconds of a black screen with white text that indicated the trial number for the proceeding stimulus presentation.

WW Condition

The WW condition targeted the same syllable combinations as the *IS* condition, except that the isolated syllables were presented as the initial syllables in rhyming word pairs that already existed in the verbal repertoires of the participants. Each word began with one of the four syllables targeted in the *IS* condition, but ended in "-ust" producing the words: *bust*, *gust*, *lust*, and *must*. See Table 2 for all of the WW compound stimulus pairings.

The procedures in the *WW* condition were identical to those in the *IS* condition, with the exception that instead of the isolated syllables presented, the previously discussed words ending in –ust were used instead.

Data Analysis

Statistical analyses were conducted utilizing R statistical software to conduct a correlated-measures t-test for all comparisons. Topographical analyses of the dependent variables were conducted using frequency and percentage measurements, and were recorded descriptively as well, in order to support a behavioral analytic interpretation of the results.

RESULTS

The raw data in Table 1 is a compilation of all 12 participants' results for each of the discrepant Isolated Syllable stimulus conditions. The far left column displays each participant's number and to its right is a column that is subdivided into three dependent variable types: (A) refers to the number of reported responses that correspond only to the acoustic stimulus, (V) refers to the number of reported responses that correspond only to the visual stimulus, and (O) refers to the number of reported responses corresponding to some stimulus *other* than the presented auditory or visual stimulus. The top horizontal bar is categorized into the specific conditions in which each participant was exposed. Note that there were 10 trials for each complex stimulus presentation. For example, participant #1 was exposed to 10 trials of the stimulus presentation "hear-ma/see-ga," and of those 10 trials, four responses corresponded to the acoustic stimulus "ma," zero trials corresponded to the visual stimulus "ga," and six responses corresponded to some other stimulus (e.g. "na"). Table 2 is organized in an identical format but contains the raw data for the participants' responses in the Whole-Word condition.

Figure 3 depicts the data from each participant's frequency of errors discriminating the non-discrepant pairs (e.g. hear-BA/see-BA or hear-GA/see-GA). There were 10 trials of each non-discrepant pair, resulting in a total of 40 possible responses for each participant. There was minimal variability in errors across participants and Participant 2 was the only one who demonstrated a relatively large number of errors.

Table 1
Compiled data of all 12 participants' results for each of the discrepant Isolated Syllable stimulus conditions

Isolated Syllable Data for All Participants

Participant #	Response Type	Hear-Labial/See-Non				TOTALS	Hear-Non/See-Labial				TOTALS
		MA/GA	MA/LA	BA/GA	BA/LA		GA/MA	GA/BA	LA/MA	LA/BA	
1	A	4	6	1	0	11	10	10	0	0	20
	V	0	4	0	10	14	0	0	10	0	10
	O	6	0	9	0	15	0	0	0	10	10
2	A	0	0	0	0	0	8	7	0	0	15
	V	0	0	0	0	0	0	3	0	0	3
	O	10	10	10	10	40	2	0	10	10	22
3	A	0	0	0	0	0	0	1	0	0	1
	V	0	0	0	0	0	0	0	0	0	0
	O	10	10	10	10	40	10	9	10	10	39
4	A	9	10	8	0	27	10	10	10	9	39
	V	0	0	0	10	10	0	0	0	0	0
	O	1	0	2	0	3	0	0	0	1	1
5	A	0	0	1	0	1	10	10	0	0	20
	V	0	0	0	0	0	0	0	0	0	0
	O	10	10	9	10	39	0	0	10	10	20
6	A	0	0	0	0	0	10	10	0	0	20
	V	0	1	0	0	1	0	0	0	0	0
	O	10	9	10	10	39	0	0	10	10	20
7	A	0	0	0	0	0	0	0	0	0	0
	V	0	6	7	10	23	0	10	0	10	20
	O	10	4	3	0	17	10	0	10	0	20
8	A	0	0	0	0	0	10	10	0	0	20
	V	0	0	0	0	0	0	0	0	0	0
	O	10	10	10	10	40	0	0	10	10	20
9	A	8	10	9	6	33	10	10	3	0	23
	V	0	0	0	2	2	0	0	0	0	0
	O	2	0	1	2	5	0	0	7	10	17
10	A	0	0	1	0	1	10	10	0	0	20
	V	0	0	0	0	0	0	0	0	10	10
	O	10	10	9	10	39	0	0	10	0	10
11	A	0	1	4	1	6	10	10	2	1	23
	V	0	0	0	0	0	0	0	1	0	1
	O	10	9	6	9	34	0	0	7	9	16
12	A	0	3	0	0	3	10	10	0	0	20
	V	0	0	0	0	0	0	0	0	10	10
	O	10	7	10	10	37	0	0	10	0	10

Note - Response Type: (A) = topography corresponding to auditory stimulus; (V) = topography corresponding to visual stimulus, (O) topography neither corresponds to auditory or visual stimulus

Table 2

Compiled data of all 12 participants' results for each of the discrepant Whole-Word stimulus conditions

Whole Word Data for All Participants

Participant #	Response Type	Hear-Labial/See-Non				TOTALS		Hear-Non/See-Labial				TOTALS
		MUST/GUST	MUST/LUST	BUST/GUST	BUST/LUST			GUST/MUST	GUST/BUST	LUST/MUST	LUST/BUST	
1	A	9	1	1	0	11		9	10	0	1	20
	V	0	9	0	10	19		0	0	10	1	11
	O	1	0	9	0	10		1	0	0	8	9
2	A	5	4	7	3	19		8	5	2	1	16
	V	0	2	0	0	2		0	3	0	0	3
	O	5	4	3	7	19		2	2	8	9	21
3	A	1	1	0	1	3		5	4	0	0	9
	V	0	0	0	6	6		0	0	0	0	0
	O	9	9	10	3	31		5	6	10	10	31
4	A	10	8	8	10	36		10	10	9	8	37
	V	0	2	0	0	2		0	0	1	0	1
	O	0	0	2	0	2		0	0	0	2	2
5	A	9	5	8	9	31		10	10	0	0	20
	V	0	0	0	0	0		0	0	0	0	0
	O	1	5	2	1	9		0	0	10	10	20
6	A	0	0	1	0	1		10	10	0	0	20
	V	0	2	0	0	2		0	0	0	0	0
	O	10	8	9	10	37		0	0	10	10	20
7	A	0	0	0	0	0		0	0	0	0	0
	V	9	10	9	9	37		0	10	1	9	20
	O	1	0	1	1	3		10	0	9	1	20
8	A	10	8	2	1	21		10	9	8	10	37
	V	0	2	0	9	11		0	1	1	0	2
	O	0	0	8	0	8		0	0	1	0	1
9	A	9	5	7	8	29		10	10	7	9	36
	V	0	5	2	2	9		0	0	2	1	3
	O	1	0	1	0	2		0	0	1	0	1
10	A	10	9	4	5	28		10	10	2	2	24
	V	0	1	0	5	6		0	0	0	7	7
	O	0	0	6	0	6		0	0	8	1	9
11	A	10	10	9	10	39		10	10	8	6	34
	V	0	0	0	0	0		0	0	2	0	2
	O	0	0	1	0	1		0	0	0	4	4
12	A	10	10	9	10	39		10	10	0	0	20
	V	0	0	0	0	0		0	0	0	0	0
	O	0	0	1	0	1		0	0	10	10	20

Note - Response Type: (A) = topography corresponding to auditory stimulus; (V) = topography corresponding to visual stimulus, (O) topography neither corresponds to auditory or visual stimulus

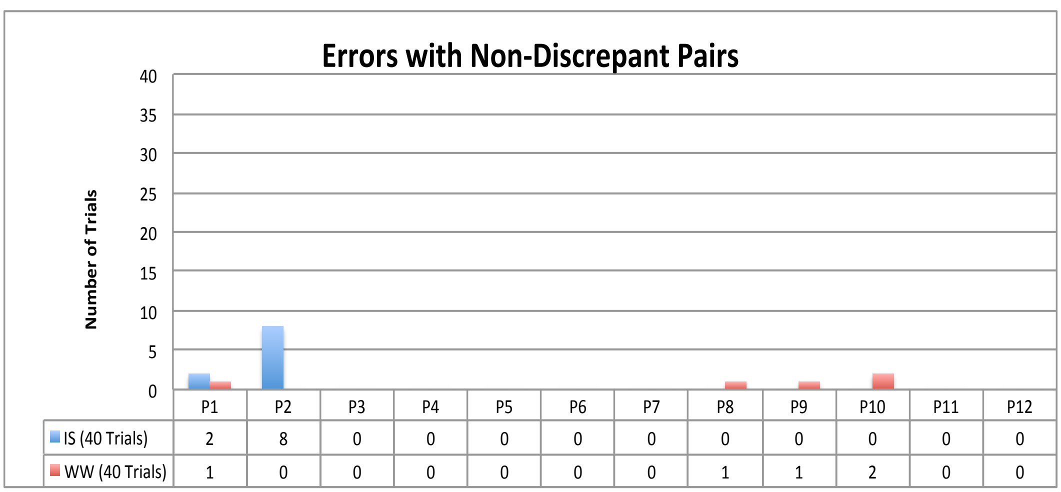

Figure 3

Results from each participant's errors with non-discrepant pairs (e.g. hear-BA/see-BA or hear-GA/see-GA).

Statistical Analyses

The statistical analyses comparing the conditions for the two hypotheses can be found in Table 3. These tests were conducted using correlated-measures t-tests, with an alpha level of .05. Each comparison is summarized below. For the individual statistical tests, please refer to the appendix.

Table 3
Eight statistical analyses of responses lacking visual or auditory control conducted with correlated-measures t-tests

Conditions Compared	# of Trials per Condition	Mean Difference (Condition 1 - Condition 2)	Standard Deviation	t-value	*p* value	Statistically significant (S) or not-significant (N) at an alpha level of .05
Isolated Syllable vs. Whole-Word	80	22.1666	16.0444	4.7859	0.000566	S
Hear-MUST/See-Nonlabial vs. Hear-BUST/See-Nonlabial	20	1.7500	3.9341	1.5409	0.1516	N

Isolated Syllables vs. Whole Words

This statistical test analyzed the total number of responses lacking correspondence with acoustic or visual control in the Isolated Syllable conditions as compared with those in the Whole-Word condition. The difference between the Isolated-Syllable condition (M = 46.08) and the Whole-Word condition (M = 23.92) was evaluated with a Correlated-groups t test ($t_{(11)} = 4.79$, $p = .00056$). The confidence interval for the mean difference between conditions was 11.97 – 32.36 and the effect size for the independent variable, using Cohen's d for paired scores, was 1.38. See Figure 4 for individual participant data comparing results in the Isolated-Syllable condition with the results from the Whole-Word condition.

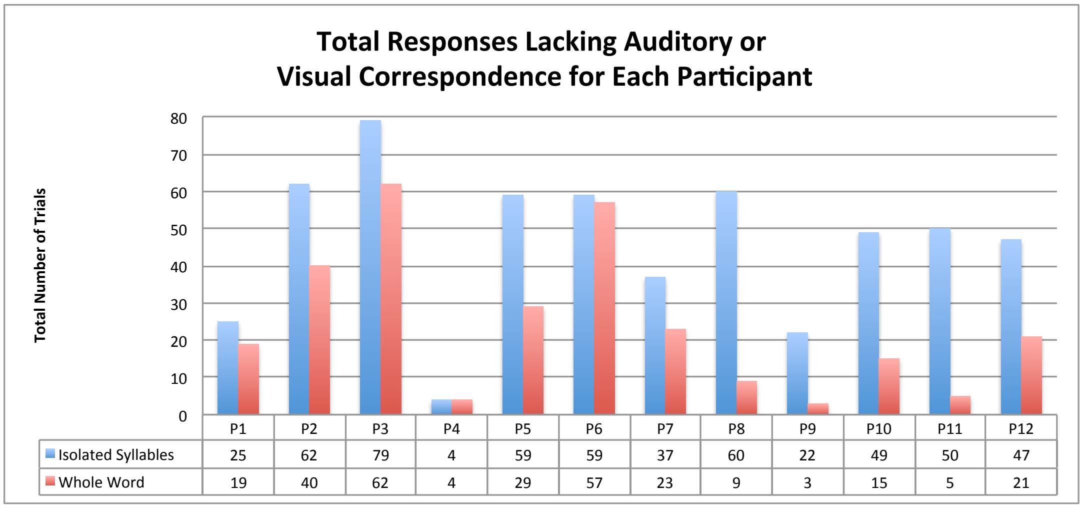

Figure 4
Individual participant data comparing results in the Isolated-Syllable condition with the results in the Whole-Word condition.

Hear-MUST/See-NonLabial vs. Hear-BUST/See-NonLabial

This statistical test compared the total number of responses lacking visual or auditory control in the conditions whose complex stimuli consisted of the acoustic whole word “must” and non-labially produced visual stimuli corresponding to “gust” or “lust” with the conditions whose complex stimuli consisted of the acoustic whole word “bust” and non-labially produced lip movement corresponding to “gust” or “lust”. The relevance of this test is that it compares the predicted McGurk effect as a nonsense word (e.g. “nust”) with the predicted effect of a response within the verbal repertoires of the participants (e.g. “dust”). No statistically significant difference was found between these conditions: $t_{(11)} = 1.54$ ($p > .05$). The difference between the Hear-MUST/See-NonLabial condition ($M = 6.25$) and the Hear-BUST/See-NonLabial condition ($M = 4.5$) was evaluated with a Correlated-groups t test ($t_{(11)} = 1.54$, $p = .15$). The confidence interval for the mean difference between conditions was -0.75 – 4.25 and the effect size for the independent variable, using Cohen's d for paired scores, was 0.4448267. See Figure 5 for the individual participant data comparing results in the Hear-MUST/See-NonLabial condition with the results in the Hear-BUST/See-NonLabial condition.

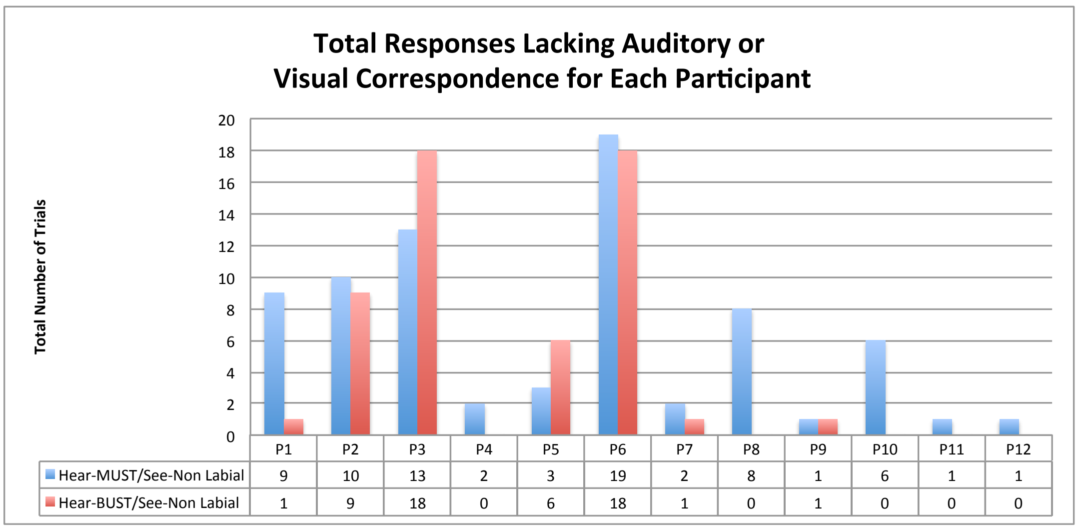

Figure 5
Individual participant data comparing results in the Hear-MUST/See-NonLabial condition with the results in the Hear-BUST/See-NonLabial condition.

Response Topography, Frequency, and Percentage Analysis

Tables 4 and 5 summarize the frequency and percentage of each isolated syllable complex stimulus condition for all participants. The vertical bar on the far left depicts the acoustic stimuli presented for each complex stimulus and the top horizontal bar displays the visual stimulus presented for each complex stimulus. Within the block for each complex stimulus arrangement there are three columns. The left column corresponds to the topography of the response recorded by the participants. The middle column displays the frequency of the responses corresponding to each of the topographies. The right column displays the percentage of participant responses corresponding to each of the displayed topographies. Since each complex stimulus trial was presented 10 times to each participant and there were 12 participants, there are a total of 120 possible responses for each condition. Tables 6 and 7 are arranged identically to Tables 4 and 5, but display the data for the Whole-Word conditions.

Table 4

Topography, frequency and percentage of each Hear-Labial/See-NonLabial Isolated Syllable condition for all participants

AUDITORY STIMULUS	VISUAL STIMULUS					
	GA			LA		
	Topography	Total	Percentage	Topography	Total	Percentage
	"na" (O)	**90**	**75.0%**	**"na" (O)**	**79**	**65.8%**
	"la" (O)	**9**	**7.5%**			
				"ma" (A)	30	25.0%
	"ma" (A)	21	17.5%	"la" (V)	11	9.2%
MA	"ga" (V)	0	0.0%			
	"da" (O)	**57**	**47.5%**	**"da" (O)**	**60**	**50.0%**
	"tha" (O)	**17**	**14.2%**	**"tha" (O)**	**20**	**16.7%**
	"la" (O)	**13**	**10.8%**	**"bla" (O)**	**1**	**0.8%**
	"ta" (O)	**1**	**0.8%**			
BA	**"pa" (O)**	**1**	**0.8%**	"ba" (A)	7	5.8%
				"la" (V)	32	26.7%
	"ba" (A)	24	20.0%			
	"ga" (V)	7	5.8%			

Note - "X" (A) = topography corresponding to auditory stimulus; "X" (V) = topography corresponding to visual stimulus; "X" (O) = topography lacking correspondence with auditory or visual stimulus

Table 5

Topography, frequency and percentage of each Hear-NonLabial/See-Labial Isolated Syllable condition for all participants

AUDITORY STIMULUS	VISUAL STIMULUS					
	MA			BA		
	Topography	Total	Percentage	Topography	Total	Percentage
GA	**"ba" (O)**	**12**	**10.0%**	**"bga" (O)**	**9**	**7.5%**
	"bga" (O)	**10**	**8.3%**			
				"ga" (A)	98	81.7%
	"ga" (A)	98	81.7%	"ba" (V)	13	10.8%
	"ma" (V)	0	0.0%			
LA	**"bla" (O)**	**94**	**78.3%**	**"bla" (O)**	**69**	**57.5%**
				"ma" (O)	**11**	**9.2%**
	"la" (A)	15	12.5%			
	"ma" (V)	11	9.2%	"la" (A)	10	8.3%
				"ba" (V)	30	25.0%

Note - "X" (A) = topography corresponding to auditory stimulus; "X" (V) = topography corresponding to visual stimulus; "X" (O) = topography lacking correspondence with auditory or visual stimulus

Hear-MA/See-GA

This condition in which the complex stimulus was composed of the acoustic syllable “ma” and the visual stimulus corresponding to the syllable “ga” had two different response types that lacked correspondence with either the acoustic stimulus or the visual stimulus. The fused response “na” had the highest frequency of responses with a total of 90 occurrences, which is 75% of all 120 responses. Acoustic correspondence occurred 21 times, or 17.5% of the trials. There were no occurrences of visual correspondence for this condition.

Hear-MA/See-LA

This condition in which the complex stimulus was composed of the acoustic syllable “ma” and the visual stimulus corresponding to the syllable “la” had only one response type that lacked correspondence with either the acoustic stimulus or the visual stimulus. The fused response “na” had the highest frequency of responses with a total of 79 occurrences, which is 65.8% of all 120 responses. Acoustic correspondence occurred 30 times, or 25% of the trials. Visual correspondence occurred 11 times, or 9.2% of the trials.

Hear-BA/See-GA

This condition in which the complex stimulus was composed of the acoustic syllable “ba” and the visual stimulus corresponding to the syllable “ga” had five response types that lacked correspondence with either the acoustic stimulus or the visual stimulus. The fused response “da” had the highest frequency of responses with a total of 57 occurrences, which is 47.5% of all 120 responses. Acoustic correspondence occurred 24 times, or 20% of the trials. Visual correspondence occurred 7 times, or 5.8% of the trials.

Hear-BA/See-LA

This condition in which the complex stimulus was composed of the acoustic syllable “ba” and the visual stimulus corresponding to the syllable “la” had three response types that lacked correspondence with either the acoustic stimulus or the visual stimulus. The fused response “da” had the highest frequency of responses with a total of 60 occurrences, which is 50% of all 120 responses. Acoustic correspondence occurred 7 times, or 5.8% of the trials. Visual correspondence occurred 32 times, or 26.7% of the trials.

Hear-GA/See-MA

This condition in which the complex stimulus was composed of the acoustic syllable “ga” and the visual stimulus corresponding to the syllable “ma” had two response types that lacked correspondence with either the acoustic stimulus or the visual stimulus. Responses with acoustic correspondence had the highest frequency of responses with a total of 98 occurrences, which is 81.7% of all 120 responses. The blended response “bga” only occurred 10 times, or 8.3% of the trials. Visual correspondence occurred zero times. There were no occurrences of the pure blend “mga” in which the response would have had perfect point-to-point correspondence with both the visual and the acoustic stimuli.

Hear-GA/See-BA

This condition in which the complex stimulus was composed of the acoustic syllable “ga” and the visual stimulus corresponding to the syllable “ba” had two response types that lacked correspondence with either the acoustic stimulus or the visual stimulus. Responses with acoustic correspondence had the highest frequency of responses with a total of 98 occurrences, which is 81.7% of all 120 responses. The blended response “bga” occurred 9 times, which is a percentage of 7.5% of the trials. Visual correspondence occurred 13 times, or 10.8% of the trials.

Hear-LA/See-MA

This condition in which the complex stimulus was composed of the acoustic syllable “la” and the visual stimulus corresponding to the syllable “ma” had one response type that lacked correspondence with either the acoustic stimulus or the visual stimulus. The blended response “bla” had the highest frequency of responses with a total of 94 occurrences, or 78.3% of the trials. Acoustic correspondence occurred a total of 15 times, which is 12.5% of the total responses. Visual correspondence occurred 11 times, which is 9.2 % of the trials. There were no occurrences of the pure blend “mla” in which the response would have had perfect point-to-point correspondence with both the visual and the acoustic stimuli.

Hear-LA/See-BA

This condition in which the complex stimulus was composed of the acoustic syllable “la” and the visual stimulus corresponding to the syllable “ba” had two response types that lacked correspondence with either the acoustic stimulus or the visual stimulus. The blended response “bla” had the highest frequency of responses with a total of 69 occurrences, which is 57.5% of the total responses. Acoustic correspondence occurred a total of 10 times, which is 8.3% of the total responses. Visual correspondence occurred 30 times, which is 25% of the trials.

Table 6

Topography, frequency and percentage of each Hear-NonLabial/See-Labial Whole-Word complex stimulus condition for all participants

AUDITORY STIMULUS	VISUAL STIMULUS					
	GUST			LUST		
	Topography	Total	Percentage	Topography	Total	Percentage
MUST	**"nust" (O)** **"lust" (O)** "must" (A) "gust" (V)	**26** **2** 83 9	**21.7%** **1.7%** 69.2% 7.5%	**"nust" (O)** "must" (A) "lust" (V)	**26** 61 33	**21.7%** 50.8% 27.5%
BUST	**"lust" (O)** **"glust" (O)** **"thust" (O)** **"must" (O)** **"nust" (O)** **"dust" (O)** "bust" (A) "gust" (V"	**28** **9** **8** **4** **3** **1** 56 11	**23.3%** **7.5%** **6.7%** **3.3%** **2.5%** **0.8%** 46.7% 9.2%	**"thust" (O)** **"dust" (O)** **"glust" (O)** **"gust" (O)** "bust" (A) "lust" (V)	**12** **6** **3** **1** 57 41	**10.0%** **5.0%** **2.5%** **0.8%** 47.5% 34.2%

Note - "X" (A) = topography corresponding to auditory stimulus; "X" (V) = topography corresponding to visual stimulus; "X" (O) = topography lacking correspondence with auditory or visual stimulus

Table 7
Topography, frequency and percentage of each Hear-Labial/See-NonLabial Whole-Word complex stimulus condition for all participants

AUDITORY STIMULUS	VISUAL STIMULUS					
	MUST			BUST		
	Topography	Total	Percentage	Topography	Total	Percentage
GUST	**"bust" (O)** **"bgust" (O)** **"glust" (O)** "gust" (A) "must" (V)	**11** **6** **1** 102 0	**9.2%** **5.0%** **0.8%** 85.0% 0.0%	**"bgust" (O)** "gust" (A) "bust" (V)	**8** 98 14	**6.7%** 81.7% 11.7%
LUST	**"blust" (O)** **"bust" (O)** "lust" (A) "must" (V)	**48** **19** 36 17	**40.0%** **15.8%** 30.0% 14.2%	**"blust" (O)** **"must" (O)** **"plust" (O)** **"nust" (O)** **"gust" (O)** "lust" (A) "bust" (V)	**46** **17** **1** **1** **1** 37 17	**38.3%** **14.2%** **0.8%** **0.8%** **0.8%** 30.8% 14.2%

Note - "X" (A) = topography corresponding to auditory stimulus; "X" (V) = topography corresponding to visual stimulus; "X" (O) = topography lacking correspondence with auditory or visual stimulus

Hear-MUST/See-GUST

This condition in which the complex stimulus was composed of the acoustic word “must” and the visual stimulus corresponding to the word “gust” had two different response types that lacked correspondence with either the acoustic stimulus or the visual stimulus. Responses with acoustic correspondence had the highest frequency with a total of 83 occurrences, which is 69.2% of the trials. The fused response “nust” had a total of 26 responses, which is 21.7% of the total responses. There were nine occurrences of visual correspondence, which is 7.5% of the trials.

Hear-MUST/See-LUST

This condition in which the complex stimulus was composed of the acoustic word “must” and the visual stimulus corresponding to the word “lust” had one response type that lacked correspondence with either the acoustic stimulus or the visual stimulus. Responses with acoustic correspondence had the highest frequency with a total of 61 occurrences, which is 50.8% of the trials. The fused response “nust” had a total of 26 responses, which is 21.7% of the trials. There were 33 occurrences of visual correspondence, which is 27.5% of the trials.

Hear-BUST/See-GUST

This condition in which the complex stimulus was composed of the acoustic word “bust” and the visual stimulus corresponding to the word “gust” had six response types that lacked correspondence with either the acoustic stimulus or the visual stimulus. Responses with acoustic correspondence had the highest frequency with a total of 56 occurrences, which is 46.7% of the trials. The fused response “dust” had the lowest frequency of responses with a total of 1 occurrence. Visual correspondence occurred 11 times, which is 9.2% of the trials.

Hear-BUST/See-LUST

This condition in which the complex stimulus was composed of the acoustic word “bust” and the visual stimulus corresponding to the word “lust” had four response types that lacked correspondence with either the acoustic stimulus or the visual stimulus. Responses with acoustic correspondence had the highest frequency with a total of 57 occurrences, which is 47.5% of the trials. The fused response “dust” occurred 6 times, which is 5% of the trials. Visual correspondence occurred 41 times, which is 34.2% of the trials.

Hear-GUST/See-MUST

This condition in which the complex stimulus was composed of the acoustic word “gust” and the visual stimulus corresponding to the word “must” had three response types that lacked correspondence with either the acoustic stimulus or the visual stimulus. Responses with acoustic correspondence had the highest frequency of responses with a total of 102 occurrences, which is 85% of the trials. The blended response “bgust” occurred six times, which is 5% of the trials. Visual correspondence occurred zero times. There were no occurrences of the pure blend “mgust” in which the response would have had perfect point-to-point correspondence with both the visual and the acoustic stimuli.

Hear-GUST/See-BUST

This condition in which the complex stimulus was composed of the acoustic word “gust” and the visual stimulus corresponding to the word “bust” had two response types that lacked correspondence with either the acoustic stimulus or the visual stimulus. Responses with acoustic correspondence had the highest frequency of responses with a total of 98 occurrences, which is 81.7% of the trials. The blended response “bgust” occurred 8 times, which is 6.7% of all the trials. Visual correspondence occurred 14 times, which is 11.7% of all the trials.

Hear-LUST/See-MUST

This condition in which the complex stimulus was composed of the acoustic word “lust” and the visual stimulus corresponding to the word “must” had two response types that lacked correspondence with either the acoustic stimulus or the visual stimulus. The blended response “blust” had the highest frequency of responses with a total of 48 occurrences, which is 40% of the trials. Acoustic correspondence occurred a total of 15 times, which is 12.5% of the trials. Visual correspondence occurred 17 times, which is 14.2% of the trials. There were no occurrences of the pure blend “mlust” in which the response would have had perfect point-to-point correspondence with both the visual and the acoustic stimuli.

Hear-LUST/See-BUST

This condition in which the complex stimulus was composed of the acoustic syllable “lust” and the visual stimulus corresponding to the syllable “bust” had five response types that lacked correspondence with either the acoustic stimulus or the visual stimulus. The blended response “blust” had the highest frequency of responses with a total of 46 occurrences, which is 38.3% of the trials. Acoustic correspondence occurred a total of 37 times, which is 30.8% of the trials. Visual correspondence occurred 17 times, which is 14.2% of the trials.

DISCUSSION

These findings support the authors' initial hypothesis that the McGurk effect would be stronger when syllables were presented in isolation than when they were presented in the context of whole words. The total number of responses that lacked correspondence with the visual or auditory stimuli was statistically greater in the Isolated Syllable condition than in the Whole-Word condition and the effect size for these comparisons was quite large. This suggests that the strength of the conditioned perceptual responses can be affected by the participants' verbal histories of reinforcement. The second hypothesis, which proposed that the tendency to report hearing the word "dust" in the hear-BUST/See-NonLabial conditions would be stronger than the tendency to report hearing "nust" in the hear-MUST/See-NonLabial conditions was not supported by the present study. There was no statistically significant difference between the total number of responses lacking visual or auditory correspondence in the two conditions. Additionally, with the exception of participants #3 and #5, all of the participants experienced higher frequencies of responses lacking acoustic or visual correspondence in the Hear-MUST/See-NonLabial conditions when compared with those in the Hear-BUST/See-NonLabial conditions. Interpretations supported by the data may point to some possible explanations for this outcome.

Since the participants were exposed to a high rate of rhyming discrepant stimulus presentations it is possible that there was some habituation to the stimulus discrepancies, thereby weakening the tendency to hear “real” words (i.e. a word within the subject’s repertoire). This may also be supported by the fact that the predicted effect of “hearing” the word *dust* in response to the hear-BUST/ See-NonLabial conditions was the only illusion corresponding to an actual word within the English language. Future studies might benefit from comparing the hear-BUST/ See-NonLabial condition with the hear-MUST/ See-NonLabial condition when presented within the context of hearing a high frequency of non-discrepant and non-rhyming words. For example, after being primed by a recorded individual saying words, such as *ball, cat, sweet, pencil*, etc. and then presented with the complex stimuli hear-MUST/see-LUST or hear-BUST/see-GUST, the experimenter could then test the tendency for the participant to report having heard *nust* or *dust*. Additionally, a limitation within the present study is that the effect size for these comparisons was moderate and future research analyzing these conditions would benefit from a larger sample.

Another factor that possibly contributed to the strength of hearing *nust* over *dust* might have to do with the differences between the acoustic properties themselves. The stop-voiced properties within the vocally produced consonant /b/, as we have seen, are shared by two other non-labially produced consonants: /d/ and /g/. The acoustic properties within the vocally produced consonant /n/, however, have a unique and more restricting effect on the listener. The consonant /n/ is produced by the speaker’s “complete closure in the oral cavity along with a lowered velum to allow airflow through the nasal cavity” (http://soundsofspeech.uiowa.edu/). The only other non-labially produced English phoneme that shares this acoustic property is the consonant /ng/, which is not emitted as an initial consonant for any word within the English language. A similar analysis can be extended to the increased tendency of participants to report “bl” blends over “bg” blends.

Within each of the eight Whole-Word conditions, responses corresponding to the acoustic stimuli had the highest frequency, with the exception of the responses reported in both of the Hear-LUST/See-NonLabial conditions. The reported blend *blust* was higher than any other alternative response form in these conditions. The prepotency of the response to hear “bl” blends over alternative responses might be better understood when we consider the conditioning history of a typical English-speaking adult. Seeking support for this interpretation, the authors turned to Webster’s Dictionary and discovered a total of 187 words for which the “bl” blend is the initial syllable. In contrast, there is not a single word whose initial sound is a “bg,” “mg,” or “ml” blend. Although the role of reinforcement histories on these findings is largely interpretive, it is plausible, and it points to empirical variables that may be responsible for conditioned perceptual phenomena as a whole.

One limitation of the study is that the researcher used video and audio recordings of his own speech sounds and lip movements and the idiosyncrasies of the individual speaker might have played a role in some of the variability in responding. Mallick, Magnotti, and Beauchamp (2015) explored variability in responding to McGurk stimuli and found a wide range of variability in responding across 360 individuals and across a number of different recorded speakers. However, the results from the responses to non-discrepant stimuli in the present study demonstrated minimal variability across the participants. Only the frequency of Participant 2's errors was noteworthy, in that she had eight errors within the isolated syllable condition. Each of her errors was a misinterpretation of the syllable "ba" as "pa," however she did not demonstrate any corresponding errors in the Whole-Word condition (e.g. interpreting the stimulus "bust" as "pust"). The comparison of Participant 2's errors in the non-discrepant IS trials with the complete lack of errors in the corresponding non-discrepant WW trials may add further evidence to the role of reinforcement histories in the listener-discrimination of speech sounds, considering the fact that her errors to the hear-BA/see-BA trials failed to persist when they would have led the perception of the nonsense word "pust."

The McGurk effect is only one example of the incalculable possibilities for future research on conditioned perceptual behavior. As discussed earlier, conditioning the verbal histories of the participants in order to validate the interpretations asserted in this paper may not be possible due to the limitations of our current technology. However, a principle of behavior is an empirically validated "fact in the bag" (Skinner, 1969, p. 84), and utilizing the handful of behavioral principles to interpret perplexing complex phenomena, such as the McGurk effect, may help to guide future behavior analytic research on perception, and perhaps help to stimulate the refinements of the scientific technology required to overcome such experimental obstacles.

REFERENCES

REFERENCES

Bruner, J. S., & Postman, L. (1949). On The Perception Of Incongruity: A Paradigm. *J Personality Journal of Personality, 18*(2), 206-223. doi:10.1111/j.1467-6494.1949.tb01241.x

Donahoe, J. W., & Palmer, D. C. (1994). *Learning and complex behavior*. Boston, MA: Allyn and Bacon.

Fisher, C. (1968). Confusions among visually perceived consonants. *Journal of Speech and Hearing Research, 11*, 796-804.

Green, K. P., Kuhl, P. K., Meltzoff, A. N., & Stevens, E. B. (1991). Integrating speech information across talkers, gender, and sensory modality: Female faces and male voices in the mcgurk effect. *Perception and Psychophysics, 50*(6), 524-536.

Kamin, L.J. (1969) Predictability, surprise, attention and conditioning. In B.A. Campbell & R.M. Church (Eds.), *Punishment and aversive behavior* (pp. 279-296). New York: Appleton-Century-Crofts.

Lidestam, B., & Beskow, J. (2006). Visual Phonemic Ambiguity and Speechreading. *Journal of Speech, Language, and Hearing Research, 49*(4), 835-847. doi: 10.1044/1092-4388(2006/059)

Mallick, D. B., Magnotti, J. F., & Beauchamp, M. S. (2015). Variability and stability in the McGurk effect: Contributions of participants, stimuli, time, and response type. *Psychonomic Bulletin & Review Psychon Bull Rev*. doi:10.3758/s13423-015-0817-4

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

MacDonald, J., & McGurk, H. (1978). Visual influences on speech perception processes. *Perception and Psychophysics, 24*(3), 253-257.

Meronen, A., Tiipana, K., Westerholm, J., & Ahonen, T. (2013). Audiovisual speech perception in children with developmental language disorder in degraded listening conditions. *Journal of Speech, Language, and Hearing Research, 56*, 211-221.

Michael, J., Palmer, D. C., & Sundberg, M. L. (2011). The multiple control of verbal behavior. *The Analysis of Verbal Behavior, 27*, 3-22.

Palmer, D. C. (1991). A behavioral interpretation of memory. In L. J. Hayes & P. N. Chase (Eds.) *Dialogues on verbal behavior* (pp. 261-279). Reno, NV: Context Press.

Skinner, B. F. (1953). *Science and human behavior.* New York, NY: Macmillan.

Skinner, B. F. (1957). *Verbal behavior.* New York, NY: Appleton-Century-Crofts.

Skinner, B. F. (1969). *Contingencies of reinforcement; a theoretical analysis* (2nd ed.). New York, NY: Appleton-Century-Crofts.

Skinner, B. F., & Epstein, R. (1980). *Notebooks*. Englewood Cliffs, NJ: Prentice-Hall.

Soto-Faraco, S., & Alsius, A. (2009). Deconstructing the mcgurk-macdonald illusion.

Journal of Experimental Psychology, 35(2), 580-587.

Sounds of Speech*. (n.d.). Retrieved from http://soundsofspeech.uiowa.edu/

vom Saal, W., & Jenkins, H.M. (1970). Blocking the development of stimulus control. Learning and Motivation, 1, 52-64.

APPENDICES

APPENDIX A

ISOLATED-SYLLABLES VS. WHOLE-WORDS

	Isolated Syllables (80 Trials)	Whole Word (80 Trials)	Difference
P1	25	19	6
P2	62	40	22
P3	79	62	17
P4	4	4	0
P5	59	29	30
P6	59	57	2
P7	37	23	14
P8	60	9	51
P9	22	3	19
P10	49	15	34
P11	50	5	45
P12	47	21	26

APPENDIX B

HEAR-MUST/SEE-NONLABIAL VS. HEAR-BUST/SEE-NONLABIAL

	Hear-MUST/See-Non Labial	Hear-BUST/See-Non Labial	Difference
P1	9	1	8
P2	10	9	1
P3	13	18	-5
P4	2	0	2
P5	3	6	-3
P6	19	18	1
P7	2	1	1
P8	8	0	8
P9	1	1	0
P10	6	0	6
P11	1	0	1
P12	1	0	1