RICE UNIVERSITY

Factors Affecting Audiovisual Speech Perception as Measured by the McGurk Effect

by

Debshila Basu Mallick

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE Doctor of Philosophy

APPROVED, THESIS COMMITTEE

Michael S. Beauchamp, Research Supervisor, Professor, Departments of Neurosurgery & Neuroscience, Baylor College of Medicine

James L. Dannemiller, Committee Chair, Lynette S. Autrey Professor of Psychology

Margaret E. Beier, Associate Professor of Psychology

Richard E. Grandy, McManis Professor of Philosophy

HOUSTON, TEXAS
March, 2016
ABSTRACT

An Investigation of Audiovisual Speech Perception as Measured by the McGurk Effect

by

Debshila Basu Mallick

Multisensory speech perception occurs when an individual integrates spoken sounds and mouth movements of a talker into a coherent percept, e.g., during face-to-face conversations. Under usual circumstances, spoken sounds and mouth movements match. However, when there is a mismatch between spoken sounds and mouth movements, individuals sometimes perceive a “fused” percept, different from the constituent audiovisual information. This phenomenon, known as the McGurk effect has been used in thousands of papers in the literature as a measure of audiovisual integration in speech.

For my dissertation I attempted to extend the findings of my previous work by investigating the sources of interindividual and interstimulus differences in the McGurk effect. In the first experiment, I attempted to investigate the influence of response-type on individuals’ perception of the McGurk effect. Studies of the McGurk effect have predominantly adopted either an open-choice or a forced-choice response format to record participants’ responses. For my dissertation, I compared open vs. forced choice responses in two groups. To allow me to collect data from large numbers of subjects, I developed an experimental toolkit that uses a web-based crowdsourcing tool called Amazon Mechanical Turk (MTurk) and methods to collect and analyze data using MTurk. I collected data from 110 and 117 participants in the open-choice and forced-
were more likely to report the McGurk effect than the open-choice group (69% vs 42%, p = 10^{-7}). This increase was consistent across all 8 stimuli. I showed that there was large variability in McGurk responses across subjects and stimuli for both open and forced choice conditions, ranging from 0% to 100% for subjects, and 30% to 80% for stimuli.

In the second experiment, I attempted to influence the efficacy of McGurk stimuli by changing the speed of video playback. As technology becomes geared more towards audiovisual communication (e.g. videos on YouTube, Coursera), individuals now have the option of slowing information down or speeding them up to accommodate information processing needs. I modified the playback rate such that the stimuli were presented at .5x, 1x, and 2x speeds (slow, normal, fast) to 2 groups of participants (58 in one group and 60 in another) recruited using MTurk. I found that playback rate does indeed affect frequency of McGurk responses. Under slow speeds, McGurk responses dropped (an estimated 11%), while visual responses increased (12%), whereas, speeding up the video to 2x did not result in responses different from the normal speed (0.7%). The drop in McGurk responses in the slow condition may be explained with increase in onset asynchrony between the visual and auditory cues.
Acknowledgments

It is a wonderful feeling to be so close to a major milestone in my life, the successful completion of my graduate career. I am grateful for all the support and guidance I received from my research mentor, Prof. Michael S. Beauchamp; his brilliance is remarkable and instilled in me great interest and respect for science. The most valuable lesson I have learnt from him is that simplicity is a great virtue in science. Prof. James L. Dannemiller who has been my Rice supervisor has been very understanding and encouraging. I am especially grateful to him for nominating me for the Lodieska Vaughn Stockbridge Fellowship that funded me for my last academic year. I am thankful for having Prof. Margaret E. Beier on both my Masters and Ph.D. supervisory committees. She has made very helpful suggestions to improve my work, and given me novel ideas on future career paths. I would like to express my gratitude for Prof. Richard E. Grandy who despite being on sabbatical, kindly offered to make accommodations to remain on my committee.

Special thanks to Dr. John F. Magnotti who has been a wonderful teacher and collaborator, always open for questions and a constant source of support; he introduced me to R programming which I have used extensively in my work. My lab mates Müge Özker Sertel, Lin L. Zhu, and Johannes Rennig, have always looked out for me, and I am grateful for their friendship.

I am grateful to the Department of Psychology, especially, Lanita K. Martin and Jennifer Gucwa for always being there if I ever needed anything. A big thank you to Dr. Lane for his JavaScript class that I used to script some of my experiments.

My family and friends kept me going and I am grateful for their unconditional support.
# TABLE OF CONTENTS

**BACKGROUND** ........................................................................................................... 1

**STUDY I. FACTORS AFFECTING SPEECH PERCEPTION AS MEASURED BY THE MCGURK EFFECT: INFLUENCE OF RESPONSE TYPE, TALKER IDENTITY, TALKER GENDER, AND PARTICIPANT GENDER**

- Introduction .................................................................................................................. 11
- Methods ......................................................................................................................... 18
- Results ........................................................................................................................... 20
- Discussion ..................................................................................................................... 26
- Future Directions .......................................................................................................... 29

**STUDY II. EFFECT OF VIDEO PLAYBACK RATE ON AUDIOVISUAL SPEECH PERCEPTION AS MEASURED BY THE MCGURK EFFECT**

- Introduction .................................................................................................................. 31
- Methods ......................................................................................................................... 34
- Results ........................................................................................................................... 42
- Discussion ..................................................................................................................... 55
- Future Directions .......................................................................................................... 60

**REFERENCES** .............................................................................................................. 61

**APPENDIX** ................................................................................................................... 72
LIST OF FIGURES

BACKGROUND

1.1. Google scholar citation frequencies of the original McGurk article since its publication .......................................................... 3

STUDY I. FACTORS AFFECTING SPEECH PERCEPTION AS MEASURED BY THE MCGURK EFFECT: INFLUENCE OF RESPONSE TYPE, TALKER IDENTITY, TALKER GENDER, AND PARTICIPANT GENDER

2.1. Comparing congruent audiovisual responses in open-choice and forced-choice response-types .................................................. 20

2.2. Percentage of individuals having extreme (EXT) and middle (MID) fusion percentages ........................................................... 22

2.3. Within group comparison of fusion responses in open-choice condition with the difference in fusion responses between open and forced-choice condition .......... 23

2.4. Comparing congruent audiovisual responses in open-choice and forced-choice response-types .................................................. 25

2.5. Percentage of correct responses for unisensory stimuli ....................... 26

2.6. Two-dimensional syllable feature space illustrating audiovisual integration in McGurk perception ............................................ 28

STUDY II. EFFECT OF VIDEO PLAYBACK RATE ON AUDIOVISUAL SPEECH PERCEPTION AS MEASURED BY THE MCGURK EFFECT

3.1. The components participants had to complete before starting the main experiment on MTurk ..................................................... 34

3.2. The starting page of the task as seen on Amazon Mechanical Turk .............. 36

vii
3.3. The main task page as seen on Amazon Mechanical Turk, showing all 3 response options, and the trial counter above the options .......................... 37

3.4. The ending screen for the task ................................................................. 37

3.5. Analysis of stimulus features ................................................................. 38

3.6. Spectrogram of the auditory “ba” stimulus which is the auditory component of the McGurk stimuli ......................................................... 39

3.7. Change in McGurk frequency across different playback rates ................. 42

3.8. Change in visual responses across different playback rates ...................... 44

3.9. Change in auditory responses across different playback rates ................... 45

3.10. Responses across different playback rates for three participants who either never perceive the McGurk effect (0% average McGurk frequency), perceive the effect 50% of the time, or who always perceive the effect (100% average frequency) ............................................................................................................. 47

3.11. Mean asynchrony values averaged across all stimuli at each playback rate .... 49

3.12. Fundamental frequency (f₀) measured for each stimulus at different playback rates .................................................................................................................. 50

3.13. Percentage of McGurk responses from old study (Basu Mallick et al., 2015) and current study ................................................................. 53

3.14. Possible explanation for the effect of changing playback rates of videos ....... 56
LIST OF TABLES

1.1. Percentage of mean McGurk responses weighted by the total number of participants, reported in studies using open-choice and forced-choice response-type ................................................................. 13

3.1. Summary of the linear mixed effect model ............................................... 49
BACKGROUND

A vast majority of human communication is typically face-to-face conversations. In such face-to-face conversations individuals not only listen to the talker’s voice but also look at the corresponding mouth movements in order to perceive speech. In fact, having access to the visual mouth movements not only make perception faster but also more accurate (Grant & Seitz, 2000; Stein & Meredith, 1993; Sumby & Pollack, 1954), especially if the auditory cues are noisy. However, visual cues in speech can sometimes modify perception of the auditory cues even in the absence of noise. In one such phenomenon called the McGurk effect, an incongruent audiovisual syllable pairing (for e.g.: auditory “ba” + visual “ga”) results in a completely different third percept (e.g.: “da”; McGurk & MacDonald, 1976). This phenomenon is considered a quintessential demonstration of multisensory integration in speech.

Studies of the McGurk effect most generally involved presenting incongruent audiovisual syllables to participants and asking them to report what they perceive. Participants’ responses have most commonly been categorized into auditory (e.g. “ba”), visual (e.g. “ga”), fusion or the canonical McGurk responses (e.g. “da” or “tha”), combination (e.g. “gba” or “bga”), and other syllables (e.g. “fa”) (McGurk & MacDonald, 1976). Due to the unavailability of the original stimuli, different laboratories created their own versions of the McGurk stimuli in order to study the effect.

When it was first reported almost 40 years ago, it was found to be perceived by almost everyone (for e.g.: 98% reported “da” for an auditory “ba” + visual “ga” pairing). This effect has been found to persist even in the most unusual circumstances. For instance, Green, Kuhl, Meltzoff, and, Stevens, paired male talkers with female voices and
female talkers with male voices, and found that individuals still reported the McGurk effect.

Due to its robustness and simplicity, the McGurk effect has been widely adopted as an index of audiovisual integration in speech. The McGurk effect has been used as a measure of audiovisual integration across different healthy populations: in children (Nath, Fava, & Beauchamp, 2011; Tremblay, Champoux, Bacon, Lepore & Théoret, 2007a); comparing audiovisual integration between children and adults (Erdener, Sekiyama & Burnham, 2010; Tremblay, Champoux, Voss, Bacon, Lepore & Théoret, 2007b); individuals with different native languages (Bovo, Ciorba, Prosser, & Martini, 2009; Magnotti et al., 2015; Sekiyama, Braida, Nishino, Hayashi, & Tuyo, 1995; Sekiyama & Tohkura, 1991). McGurk effect has also been used as a diagnostic index across different clinical populations such as: autism spectrum disorder (Irwin, Tornatore, Brancazio, & Whalen, 2011; Woynaroski et al., 2013), schizophrenia (de Gelder, Vroomen, Annen, Masthof, & Hodiamont, 2003; Pearl et al., 2009); hearing deficits and cochlear implant (Rouger, Fraysse, Deguine, & Barone, 2008), amblyopia (Burgmeier et al., 2015; Narinesingh, Wan, Goltz, Chandrakumar, & Wong, 2014). With advances in neuroimaging technology, the McGurk paradigm has been a popular tool to investigate the neural correlates of multisensory integration and speech perception, such as: with functional magnetic resonance imaging (fMRI) (Baum, Martin, Hamilton, & Beauchamp, 2012; Nath & Beauchamp, 2012b), transcranial magnetic stimulation (TMS) (Beauchamp, Nath, & Pasalar, 2010), positron emission tomography (PET) (Sekiyama, Kanno, Miura, & Sugita, 2003), magnetoencephalography (MEG) (Keil, Müller, Ihssen,
& Weisz, 2012), and evoked response potential (ERP) (Pratt, Bleich, & Mittelman, 2015).

Most studies of the McGurk effect focus on the percentage of fusion responses. Some studies have examined response times on the McGurk task. For example, Sekiyama et al. (Sekiyama, Soshi, & Sakamoto, 2014) examined McGurk perception in older and younger adults and found that response times were longer for older adults than younger adults, and for both groups response times were longer for incongruent compared to congruent stimuli.

Since 1976, the original McGurk article (McGurk & MacDonald, 1976) has been cited over 4,400 times (based on Google scholar citations), with about 70% of the works published in the last decade. Based on the pattern of citations of the original paper, it can be considered a sleeping beauty in science (Ke, Ferrara, Radicchi, & Flammini, 2015), i.e., interest in this paper picked up quite some time after its publication.

![](Image)

**Figure 1.1.** Google scholar citation frequencies of the original McGurk article (McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature.*) since its publication.
The McGurk effect was found to persist under a variety of circumstances due to which it was considered a robust effect. For instance, a study by Munhall, Gribble, Sacco, and Ward (1996), found that individuals reported the McGurk effect even when the auditory syllable (“b”) lagged the visual syllable (“g”) by 180 ms.

The McGurk effect prevailed even when the consonants were paired with different vowels (“a”, “i”, “u”; e.g. auditory “bi” + visual “gi”) (Green, Kuhl, & Meltzoff, 1988).

In a study of the McGurk effect, researchers occluded the left and right sides of the face to examine its effect on McGurk perception (Jordan & Thomas, 2011). They found that there was no change in McGurk perception when individuals saw the full face or half of the talkers’ face. However, another study, that investigated the facial asymmetries in conveying information found that the right side of the mouth conveyed more information than the left side of the mouth, resulting in higher rates of McGurk responses when only the right side of the mouth was visible (the left side was occluded) (Nicholls, Searle, & Bradshaw, 2004).

A previous study (Paré, Richler, ten Hove, & Munhall, 2003) found that McGurk effect did not decrease until gaze was directed 10°-20° away from the talker’s mouth; the effect became significantly smaller when direction of the gaze moved eccentrically 60° away from the mouth.

Since individuals are familiar with their own faces and voices, researchers examined whether the McGurk effect would occur while perceiving one’s own face and voice (Aruzzo & Shore, 2011). Although reduced, McGurk effect persisted even when looking at one’s own face was paired with their own voice.
McDonald et al. (2000) degraded the video of the talkers’ face using a mosaic transformation in order to determine the point at which the McGurk effect becomes negligible. Spatial degradation of the talker’s face varied at five levels from no degradation to the coarsest being 11.2 pixels/face, and 4.4 pixels/mouth. Even when the talkers face was maximally degraded spatially (11.2 pixels per face), the McGurk effect, although substantially reduced still occurred (MacDonald, Andersen, & Bachmann, 2000).

Recent studies however have brought forth strong individual differences in perception of the effect (Nath & Beauchamp, 2012b; Nath, Fava, & Beauchamp, 2011). In a previous study conducted with a 165 participants (Basu Mallick, Magnotti, & Beauchamp, 2015) I found wide interindividual differences in McGurk perception, ranging from no McGurk perception (0%) to McGurk reports on all trials (100%). Some individuals were found to perceive the effect much more frequently than others, whereas others did not perceive the effect at all. We also found large differences in McGurk perception across different stimuli used (Jiang & Bernstein, 2011), ranging from weaker stimuli evoking 17% McGurk perception to stronger stimuli evoking about 3 times more McGurk responses (58%) (Basu Mallick, Magnotti, & Beauchamp, 2015). These findings were interesting as they showed that the McGurk effect was not as robust as previously thought to be. The next step was to examine what factors resulted in such wide individual differences in McGurk perception.

Nath et al. (2012) preselected 14 adults based on their level of McGurk perception varying from no McGurk perception to McGurk perception on all presentations of the McGurk stimuli. They then measured their brain activity while the individuals watched
McGurk stimuli. They found that McGurk perception was positively correlated with brain activity in an established multisensory area, the left posterior superior temporal sulcus (pSTS); the higher the McGurk perception, the higher the brain activity. Nath et al. (2011) studied 17 children and found that their McGurk perception ranged from 0% to 100%, and similar to adults, their brain activity in the pSTS was positively correlated with their percentage of McGurk perception. In order to investigate a causal role of pSTS in perception of the McGurk effect, Beauchamp et al. (2010), recruited individuals who perceived the McGurk effect very strongly, used TMS to temporarily disable the pSTS and recorded the frequency of McGurk perception. They found that after the pSTS had been temporarily disabled, the rate of McGurk perception reduced dramatically (mean ± SEM = 94 ± 2 % to 43 ± 9 %). They were able to replicate their findings when they tested individuals with a different McGurk stimulus (86 ± 5 %). These findings showed that pSTS activity was a possible neural substrate that resulted in the variability in McGurk perception across individuals.

Keil et al. (2012) also reported variability in McGurk perception not only across individuals, but also within individuals when they reported the McGurk effect, relative to when they did not. They studied McGurk perception while examining its corresponding neural correlates using magnetoencephalography (MEG), which allowed more sensitive examination of the time-course of neural activity. They found that brain activity not only during the McGurk effect but also immediately prior to it affects McGurk perception.

Gurler et al. (2015) found that individuals who had higher percentages of McGurk perception (≥50%) frequently fixated the mouth compared to individuals who had lower percentage of McGurk perception. This is another potential contributor to individual
differences in McGurk perception (Gurler, Doyle, Walker, Magnotti, & Beauchamp, 2015). These are some factors that have been attributed to the individual differences in McGurk perception. However, there may still be other factors at play that are yet to be uncovered.

In addition to individual differences, stimulus differences are also crucial in determining McGurk perception. The initial study (McGurk & MacDonald, 1976) had reported that 98% of the adult participants had reported a McGurk percept for the auditory “ba” and visual “ga” stimuli generated from a single female talker. But stimuli from the same talker with different audiovisual syllable combination (auditory “pa” + visual “ka”) resulted in 81% of the participants reporting the effect (“ta” percept). Since the original McGurk stimuli were no longer available, different research groups have developed their own versions of the McGurk stimuli to use in various research projects (Nath & Beauchamp, 2012; Quinto, Thompson, Russo, Trehub, 2010; Skipper & Wassenhove, 2007).

In my previous studies, I had compared McGurk responses across several different McGurk stimuli collected from different research groups and from online presentations of the effect. I found wide interstimulus differences in the strength of the stimuli to evoke the McGurk effect, ranging from weak stimuli resulting in 17% McGurk responses to strong stimuli resulting in 81% McGurk responses averaged across 165 participants. Other studies have forwarded such interstimulus differences as well (Jiang & Bernstein, 2011; MacDonald & McGurk, 1978).

In addition, I also found that there were various procedural differences in the studies investigating the McGurk effect. For instance, some studies used open-choice
response type (e.g. “report what you perceived.”) to collect responses from participants (Fixmer & Hawkins, 1998; Jiang & Bernstein, 2011; J. F. Magnotti et al., 2015; Nath & Beauchamp, 2012; Paré et al., 2003); whereas others have used forced-choice response type in which participants are asked to choose the best alternative from a given set of responses, with the number of response alternatives most commonly ranging from 2 to 4 alternatives (K. P. Green, Kuhl, Meltzoff, & Stevens, 1991; Keil et al., 2012; Quinto, L., Thompson, W.F., Russo, F.A., Trehub, 2010; Sekiyama, Burnham, Tam, & Erdener, 2003; Stevenson, Zentsov, & Wallace, 2012).

Another procedural difference across studies of the McGurk effect is how the researchers operationally defined the McGurk response. Some researchers considered any response other than the auditory component of the McGurk stimulus used as a McGurk response (Munhall, Gribble, Sacco, & Ward, 1996; Rosenblum & Yakel, 2001), while others considered only the canonical “da” or “tha” responses as McGurk responses (Magnotti et al., 2015; Quinto, L., Thompson, W.F., Russo, F.A., Trehub, 2010; Szycik, Stadler, Tempelmann, & Münte, 2012), and some others considered any response other than the auditory and visual syllable constituents to be a “fused” percept (Aruffo & Shore, 2011; Pearl et al., 2009).

In the original McGurk study (McGurk & MacDonald, 1976), the stimuli were two syllables (e.g. auditory “ba-ba” + visual “ga-ga”) repeated thrice (e.g. auditory “ba-ba ba-ba ba-ba” + visual “ga-ga ga-ga ga-ga”). It is unclear how the participants were asked to make responses to the stimuli. It appears that the responses recorded were just bisyllables instead of the bisyllable triads (e.g. “da-da” instead of “da-da da-da da-da”). Other studies (Pearl et al., 2009; Quinto, L., Thompson, W.F., Russo, F.A., Trehub,
2010), have used similar paradigms in which they repeated the McGurk syllables multiple times and instructed the participants, for example, to respond to the last syllable they heard.

Comparing McGurk studies are difficult due to these stimuli and procedural differences. Therefore, my previous studies used a consistent operational definition of the McGurk response while testing the effect with many different stimuli and participants.

The literature includes many models of audiovisual integration that can be categorized as primarily early or late integration models (Dupont & Luettin, 2000) depending on which stage of processing they propose the integration process actually occurs (Calvert & Thesen, 2004). The early integration perspective states that integration occurs in a feature space synthesized from all acoustic features of the cues, combined across modalities (Braida, 1991; Dupont & Luettin, 2000; Ma, Zhou, Ross, Foxe, & Parra, 2009). The space includes representations of various speech elements and their prototypical units. Whichever representation the encoded speech most likely resembles, individuals perceive that speech unit. Example of early integration models include the prelabelling model of speech (Braida, 1991). These models consider that the auditory and visual modalities are conditionally dependent. The late perspective suggests that inputs from each modality are categorized independent of each other, and result in integration based on estimates of their occurrence together (Dupont & Luettin, 2000). This late approach assumes conditional independence which makes it more restricted relative to the early models. Example of late models include the Fuzzy Logical Model of Perception (Oden & Massaro, 1978). Recent neuroimaging studies suggest that multisensory integration may be a combination of the early and late approaches (Calvert & Thesen,
Aims of the Current Dissertation

My previous studies revealed wide interindividual and interstimulus differences in McGurk perception. A key unanswered question is the sources of the interindividual and interstimulus differences. First, I attempted to influence individuals’ responses to McGurk stimuli by manipulating response-type: either by allowing people to independently generate a response to the stimuli (open-choice responding), or forcing them choose to a response from the options provided (forced-choice responding). Second, I attempted to influence the efficacy of a given McGurk stimulus by manipulating the speed of the speech presented in the stimulus.

The McGurk effect is one of the most important tools for understanding speech perception, which in turn is the most important form of human communication. Improving our knowledge about the McGurk effect is therefore important for understanding communication in the normal human condition and in clinical populations.
STUDY I. FACTORS AFFECTING SPEECH PERCEPTION AS MEASURED BY
THE MCGURK EFFECT: INFLUENCE OF RESPONSE TYPE, TALKER
IDENTITY, TALKER GENDER, AND PARTICIPANT GENDER


Introduction

In the previous chapter, I discussed the wide variability in estimates of the McGurk effect across the literature. Our next step was to determine factors that potentially resulted in this variability across different studies. From a careful literature review, I found that across different studies, there were widespread methodological differences including the response type used to collect reports from the participants; specific stimulus differences, such as the gender of the talkers; and finally participant differences including gender of the participants. I will discuss each of these factors in the following sections.

Effect of Response Type on Speech Perception

Studies of speech perception using the McGurk effect have predominantly adopted two different ways in which individuals are asked to report what they perceived. These response types include open-choice and forced-choice responses. Open-choice responses allow individuals to independently produce a response to a question, and forced-choice responses give preselected response alternatives out of which individuals are instructed to select the best (or correct) option. In experiments investigating the McGurk effect, in case of open choice responses, participants are instructed to report
what syllable they perceived after presentation of the McGurk stimulus. In case of forced-choice responses, participants are instructed to choose among given response alternatives; the numbers of response alternatives commonly vary between 2 (‘ba’ and ‘da’) to 4. While forced-choice responses deliver cues that may not have spontaneously been considered, since participants generate open-choice responses, no cues are made available. However, open-choice responses can sometimes be difficult to code, whereas forced-choice responses are easier to code and work with (Cassels & Birch, 2014). Therefore, different strategies may be employed in case of these two response types. For example, in case of forced-choice responses individuals may adopt an eliminative strategy, whereby, individuals continue to eliminate least likely alternative to hone in on the alternative they believe to be more likely (Cassels & Birch, 2014).

Expression of researcher’s expectations, or demand characteristics, during interaction with participants is a problematic confound. Forced-choice responses are often influenced by demand characteristics. However, in case of open-choice responses expression of demand characteristics are easily avoided with non-directed questions. A procedure that incorporates experimenter expectations (“did the stimulus sound like da?”) might give different results than one that does not (“what did the stimulus sound like?”) (Orne, 1962).

Colin, Radeau and, Deltenre (2005), examined how sensory and cognitive factors regulate mechanisms of speech perception using the McGurk effect. They measured percentage of McGurk perception while manipulating the intensity of auditory speech, size of talker’s face, and the instructions given to participants to make responses aligned with a forced-choice or an open-choice format. However, like many other studies, they
instructed their participants to report what they heard. They found a main effect of instruction used, with higher percentage of McGurk perception for forced choice responses. They also found an interaction between instructions used and the intensity of auditory speech. Massaro (1998, p184-188) found more combination responses for forced-choice responses compared with open-choice responses.

Table 1. Percentage of mean McGurk responses weighted by the total number of participants, reported in studies using open-choice and forced-choice response types. Across the 8 studies that used open-choice responses, the average McGurk response is 51%, whereas for the 9 studies using forced-choice responses, it is 68%.

In the present study, I examined the influence of forced vs. open-choice response type on perception of audiovisual speech as assessed with the McGurk effect with modality neutral instructions, while holding other relevant factors (e.g. intensity of auditory speech, and stimuli used) constant.
For the eight previous studies (table 1) that tested participants with open-choice responding I found a mean frequency of 51% weighted by the number of participants in each study. While for the nine studies that tested participants with forced choice responding I found a weighted mean frequency of 68%. The findings of Colin et al. (2005) and the literature review suggest that response choice may be a significant contributor to variability in speech perception. In case of forced-choice responses, participants can compare their percept with available response alternatives. However, in case of open-choice responses, participants attempt to retrieve which syllable closely matches their percept from an unrestricted number of possible syllables. Therefore, I hypothesize that forced-choice response type will result in greater frequency of McGurk reports. To test this hypothesis, I tested one group of participants with open-choice and another group with forced-choice responding, with the same set of stimuli for both groups.

Green and Swets (1966) used the forced-choice paradigm in the context of the signal detection theory in which they present to their participants 2 intervals, one including the signal and noise, and the other including just noise. After the participants are presented with these intervals, they are asked to choose which interval includes the signal. The paradigm I use is more akin to forced-choice questions used in surveys or questionnaires. A key difference between Green and Swets and my experiments is that there is a correct answer for Green and Swets, but there are no correct responses in my experiments for the incongruent stimuli that I use.
Effect of Talker Gender on Speech Perception

Studies of speech perception have used different stimuli generated often by a single male or female speaker. For e.g., the original McGurk article (McGurk and MacDonald, 1976) included stimuli generated by a single female talker, whereas, Keil et al. (2012) used stimuli from a single male talker. However, studies have shown that talker variability results in the decline in speech intelligibility (Mullennix et al., 1989; Nygaard et al., 1995; Pisoni, 1993; Sommers et al., 1994). Due to differences in physical qualities, words and syllables uttered by a male and female talker will sound different (Peterson and Barney, 1952). Generally, male talkers produce lower vowel formant frequencies, wider bandwidths, and lower fundamental frequency which may be attributed to the male vocal cords being larger, and reshaping of their larynx during puberty (Mendoza-Denton & Strand, 1998; Strand, 1999). Bradlow, Toretta, and Pisoni (1996) found that male talkers were less intelligible than female talkers. However, researchers propose the idea of a perceptual normalization process, in which differences among talkers are compensated while listeners perceive the same word or syllables as belonging to same linguistic class (Bladon, Henton, & Pickering, 1984; Johnson, 1991; Ladefoged & Broadbent, 1957; Strand, 1999). This perceptual normalization process has been illustrated in McGurk perception (Green et al., 1991) in which the researchers compared McGurk perception in a congruent (male face – male voice; female face – female voice) and incongruent condition (male face – female voice; female face – male voice) and found no significant difference in the rate of McGurk perception between these two conditions. Green et al. (1991) stated that the McGurk effect is not vulnerable to the apparent incompatibility of the male – female face and voice pairings, and attributed the
findings of this study to a perceptual normalization process. Green et al. (1991) state that the normalization process occurs early on in processing the audiovisual cues, which results in the neutral signal then being forwarded for further phonetic processing (Mullennix & Pisoni, 1990; Mullennix, Pisoni, & Martin, 1989; Nusbaum, 1990).

In order to determine whether talker gender affects McGurk perception, I tested the same participants with McGurk stimuli created with 4 male and 4 female talkers, and compare McGurk perception between male and female talkers.

*Effect of Perceiver Sex on Speech Perception*

In addition to talker gender, perceiver gender may also affect speech perception. The research on sex differences in speech perception is contentious. Some studies have found that women tend to be more influenced by the visual cues than men when presented discrepant audiovisual stimuli (Aloufy, Lapidot, & Myslobodsky, 1996 for American English; Öhrström & Traunmüller, 2004 for Swedish). Examining the neural correlates of speech perception shows that the superior temporal sulcus (Calvert, 2001; Nath & Beauchamp, 2012) and the inferior frontal gyrus (IFG, Calvert & Campbell, 2003) are critical for speech perception. Several studies have found sex differences in functional magnetic resonance imaging (fMRI) related brain activity in the IFG, with men exhibiting more left lateralized activity in this area, relative to females who show more bilateral activation in auditory tasks (Pugh et al., 1996, Shaywitz et al., 1995), semantic tasks (Baxter et al., 2003). Brain activity recorded using other techniques like positron emission tomography (PET) have echoed these sex differences, for e.g., past tense production tasks (Jaeger et al., 1998). Bilateral processing in case of females are
considered to result in speedier and more accurate phonological processing (Coney, 2002).

However, sex differences in speech perception are not found in some instances (Aloufy et al., 1996, in case of Hebrew). No sex differences were found in language comprehension (Frost, 1999), and verbal reasoning tasks (Gur et al., 2000). Irwin et al. (2006) conducted a study examining the effect of sex differences and task difficulty on McGurk perception (using stimuli recorded from a single talker), and found no sex differences in McGurk perception in the baseline condition. In order to determine if the sex of the perceiver affects McGurk perception, I used similarly sized groups of males and females. In addition to being able to examine the effects of response type, talker gender and perceiver sex, I also examined whether there are any interactions between these factors.

All data were collected using Amazon Mechanical Turk (MTurk) which is a crowdsourcing platform enabling large volumes of data collection from all over the globe, economically, and with quick turnaround time, relative to laboratory studies. Although using MTurk does not allow for as much control over the experimental study as under laboratory conditions, several studies found high accuracy in speech transcription tasks. A study that examined the reliability of speech transcription on MTurk found that recognition error was low, ranging from 4.19% to 6.53%, with a 95% agreement among transcribers (Marge, Banerjee, & Rudnicky, 2010). Other studies echo similar findings, with near expert level recognition accuracy of transcription on MTurk (Gruenstein, McGraw, & Sutherland, 2009; McGraw, Gruenstein, & Sutherland, 2009).
Materials and Methods

Stimuli

Because the stimuli in Basu Mallick (2014) were selected from previous studies, they had been created in many different laboratories and varied along a number of dimensions, including auditory and visual quality and the size of the face within the video frame. To minimize the effect of these potential confounds, I created an additional stimulus set consisting of eight McGurk stimuli (labeled 2.1–2.8, auditory “ba” + visual “ga”) recorded from four male and four female talkers. In order to avoid retention and later comparison of congruent and incongruent syllables from the same talker, I included congruent stimuli (congruent “ba”, “da”, and “ga”, audiovisual control stimuli) from a different talker. The stimuli in Experiment 1 are available for download at http://openwetware.org/wiki/Beauchamp:McGurkStimuli.

Participants were instructed to adjust the volume of their speakers to a comfortable level, and adjust the screen to be able to see the full face of the talker. There are individual differences in perception of physical intensity of auditory stimuli (Langers, van Dijk, Schoenmaker, & Backes, 2007). Therefore, instead of attempting to control the physical intensity of the auditory stimuli, I controlled for the perceived intensity by instructing the individuals to set their system volumes at a level that is comfortable for them.

Experimental Procedure and Participants

Experiment 1. Data were collected using an online data collection service (Amazon Mechanical Turk) as approved by a Rice University IRB. The eight McGurk stimuli were each presented ten times, and the control stimuli were presented five times, all randomly
interleaved. One group of participants (N = 110: 38 female, 72 male) used an open-choice response format, typing their percept following the presentation of each stimulus; these responses were coded as in Experiment 1. Responses in this condition were scored as follows: “ba” was scored as an auditory, “ga” as visual, “da” or “tha” as a McGurk fusion response, and any other responses as an other response.

A second group of participants (N = 117: 55 female, 62 male) made forced-choice responses, selecting among three possibilities corresponding to the auditory (“ba”), visual (“ga”), or McGurk fusion percept (“da” or “tha”) with a mouse click.

Six participants who did not report their gender were used for all analyses except gender comparisons, six participants were excluded for failing to complete the task as instructed, and 33 participants completed both experiments (total unique N = 195).

**Experiment 2.** To determine the accuracy of unisensory syllable perception, I extracted the auditory and visual components from the eight McGurk stimuli used in Experiment 2 and tested them using the same forced-choice procedure as in Experiment 2. Participants (N = 50: 21 female, 29 male; seven had also participated in Experiment 2) were presented with each of auditory-only and visual-only “ba”, “da”, and “ga” from eight talkers (48 unique stimuli). Each stimulus was repeated twice (96 trials) and randomly interleaved.

**Data analysis**

In the open-choice condition, responses to the McGurk stimuli were divided into four mutually exclusive categories (McGurk & MacDonald, 1976): fusion responses (“da” or “tha”), auditory responses (“ba”), visual responses (“ga”), and other (e.g., “va”). The forced-choice responses were segregated into auditory (“ba”), visual (“ga”), and fusion responses (“da”). All data were analyzed
using R statistical software (R Development Core Team, 2014).

**Results**

In Experiment 1, I presented a set of eight new McGurk stimuli to two different groups of participants. The first group used the same open-choice design as in experiment 1 of Basu Mallick et al. (2015), but the second group made a three-alternative forced choice (corresponding to the auditory component of the stimulus, the visual component of the stimulus, or the illusory McGurk percept). The forced-choice group was much more likely to report the McGurk effect than the open-choice group (69 % vs. 42 %, Kolmogorov–Smirnov D = 0.36, p = 10^{-7}; see Figure 2.1a). This increase was consistent across all eight stimuli (Figure 2.1b). Replicating the results from Basu Mallick et al. (2015), I found high variability across participants (range from 0 % to 100 %) in both the open-choice (Figure 2.1c) and forced-choice (Figure 2.1d) groups. When restricting the scoring for the open-choice condition to be similar to the forced-choice condition, there were 16% more auditory, 2% more visual and 8% other responses in open-choice condition relative to the forced-choice condition.
A critical finding from Basu Mallick et al. (2015) is that McGurk responses are not normally distributed due to the restricted range imposed in the data (0-100). A histogram showing the percentage of participants in 10% bins for 0-100% fusion responses shows highest percentage of individuals scoring either less than 10% or greater than 90% fusion responses across all trials for each stimulus (e.g. Stimulus #2.5, figure 2.2.a). In order to summarize this finding, I averaged the percentages of participants having extreme scores (<10% or >90% fusion responses) and those who have scores in between the extremes (>10% and <90% fusion responses). Replicating Basu Mallick et al. (2015), most participants were found in the extremes of the distribution, both averaged across stimuli and for each individual stimulus (Figure 2.2. c & d; data combined across choice groups).
If McGurk perception were normally distributed, then fewer participants would have scores at the extremes of the distribution compared to the middle of the distribution (figure 2.2.b). If the mean is 42% and the standard deviation is 39%, then ~20% of the data would be expected to be < 0 and > 100 for a normal distribution. However, since the range of the fusion scores is restricted between 0 -100%, the distribution of individuals across fusion scores, resembles a truncated normal at best (figure 2.2.b compared with figure 2.2.c). In fact, a Shapiro–Wilk (1965) test of normality rejected normality for each stimulus distribution (all ps < 10\(^{-13}\)), and a dip test (Hartigan, & Hartigan, 1985) rejected the hypothesis that any of the distributions were unimodal (all ps < 10\(^{-16}\)).

Figure 2.2. Percentage of individuals having extreme (EXT) and middle (MID) fusion percentages.
a. Histogram for stimulus #2.5 showing percentage of participants within each 10% frequency bin.
b. In a hypothetical normal distribution, average percentage of individuals who have extreme scores (EXT, dark bar: <10% fusion responses and >90% fusion responses) and those who have scores in the middle of these extremes (MID, light bar: >10% and <90% fusion responses).
c. Average percentages of participants across stimuli at the extremes (EXT, dark bar) and in the middle (MID, light bar: >10% and <90%) of the actual distribution.
d. Percentages of participants in the extremes and in the middle of the distribution for each individual stimulus.

I examined if there were changes in responses to the same stimuli from open-choice to forced-choice response conditions (fig 2.3). There were 33 participants who did both the open-choice and forced-choice experiments. For 14 subjects who had either 0% or 100% McGurk responses, could not decrease or increase any further due to floor and ceiling effects respectively. For the remaining 19 subjects, 11 had increased fusion responses in forced-choice condition which aligns with my finding for the overall study. Only 3 participants had 10% or more decrease in fusion in forced-choice condition. I found that responses between these two conditions were highly correlated ($r = 0.82, p = 10^{-9}$), and a paired t-test shows about 7% increase in fusion responses in forced-choice condition [$t(32) = 1.9, p = 0.07$].

![Figure 2.3](image_url)

**Figure 2.3.** Within group comparison of fusion responses in open-choice condition with the difference in fusion responses between open and forced-choice condition. The horizontal line marks the point of no difference.
Replicating the results from Basu Mallick (2014), I also found high variability across stimuli, ranging from 30% to 52% for open-choice and 57% to 80% for forced-choice conditions. The McGurk stimulus set for this experiment was created with four female and four male talkers, allowing us to measure the effects of talker gender and its interaction with the participant and task factors, and holding other stimulus factors constant. We fit a linear mixed-effects model to the behavioral data, with choice type (open vs. forced), talker gender, participant gender, and their interactions as fixed effects, and participant and stimulus as random effects. Using Satterthwaite approximations to test the significance of the model coefficients, the only large effect was the main effect of choice type [estimated 18.0% higher for forced choice, SE = 3.2%; t(1754) = 5.6, p = 10^{-8}]. There was no main effect of participant gender [8.1% lower for males, SE = 5.1%; t(292.3) = -1.6, p = .12] or talker gender [2.6% higher for male talkers, SE = 6.0%; t(7.8) = 0.43, p = 0.68], and only weak interactions between participant gender and choice type [5.8% lower for male participants in forced choice, SE = 4.1%; t(1753) = -1.4, p = .15], talker gender and choice type [6.8% lower for male talkers in forced choice, SE = 3.0%; t(1562) = -2.3, p = 0.02], and talker gender and participant gender [3.4% lower for male participants viewing male talkers, SE = 2.9%; t(1562) = -1.2, p = 0.24]. The three-way interaction was also weak [3.2% lower for male participants viewing male talkers in forced choice, SE = 3.9%; t(1562) = -0.08, p = 0.94]. Eliminating participant gender and stimulus gender from the model did not significantly change its predictive accuracy [full model, root mean squared error
(RMSE) = 19.1 %; reduced model, RMSE = 19.3 %; mean difference = 0.05 %;
paired t-test: t(1767) = -1.14, p = .25].

Figure 2.4. Comparing congruent audiovisual responses in open-choice and forced-choice response-types.

 Participants were presented with congruent stimuli interleaved with the McGurk stimuli. Responses on the congruent stimuli were found to be near ceiling for all 3 syllables averaged across all participants and trials (open-choice vs. forced-choice; 94% ± 1% vs. 92% ± 1%; figure 2.4). This shows that participants were not just responding randomly, otherwise they would not be able to have such high accuracies for the congruent stimuli.
While creating the McGurk stimuli for this experiment, I recorded the same eight talkers speaking “ba”, “ga”, and “da”. I presented the auditory and visual components of these stimuli in Experiment 1 using the three-alternative forced-choice design used in Experiment 1. Identification of the auditory-only syllables was at ceiling (fig 2.5; mean accuracy = 97 %, $SD = 4 \%$), whereas identification of the visual-only syllables was significantly worse (80 %, $SD = 10 \%$) [paired t-test: $t(49) = 13, p < 10^{-16}$]. Accuracy for the “ga” visual-only syllable was especially low (58 %, as compared with 96 % for “ba” and 86 % for “da”), with high variability across talkers (range from 5 % to 77 %) and participants (range from 6 % to 88 %).

**Discussion**

Our results confirm and extend a number of results from previous studies that have used smaller sample sizes. We showed that manipulating response type also significantly alters the frequency of the McGurk effect, with forced-choice responding increasing the frequency of McGurk perception by an estimated 18%, as compared with open choice for identical stimuli, similar to the findings of Colin et al. (2005). Colin et al. (2005) concluded that the reduction in the number of McGurk responses in the open-

![Figure 2.5. Percentage of correct responses for unisensory stimuli.](image)
choice condition was because the participants were primed to report percepts similar to their reports in a unisensory auditory condition. Since, the participants were not provided with any response alternatives in the open-choice response condition, their responses were diverse. Since the participants were not exposed to any unisensory stimulus condition prior to this task, I reject Colin et al.’s inference of the unisensory condition priming the multisensory responses. This may be attributed to the participants being more conservative about their responses so that they could report exactly what they perceived, thereby resulting in a drop in the number of McGurk responses.

Contrarily, since they had three response alternatives in the forced-choice response condition, they reported the option that had the highest likelihood based on the sensory information. Colin et al. (2005) inferred that the increase in the number of McGurk responses in forced-choice response condition was due to individuals attempting to balance the responses across all the different alternatives. However, in this study I saw that there was a disproportionately large number of McGurk responses, and low number of auditory or visual responses.

Although I found a large effect of response type, I did not examine the effect of the task instructions. Our instructions were designed to be modality neutral, so as not to bias participants toward any particular response. It is possible that different task instructions could change the frequency of the McGurk effect, adding an additional source of variability across studies.

Along the lines of Ma, Zhou, Foxe, and Parra (2009), I consider all the syllables in an individual’s dictionary to be represented in a space characterized by different syllable features. Each syllable has a prototypical representation, and several other
variations due to noise, talker, and articulatory differences. The syllables with similar features are neighbors. A McGurk syllable pair consists of an auditory “ba” and visual “ga”. Since these are feature wise different syllables, sometimes individuals presented with the incongruent stimulus pairing go with the percept that is midway between the audiovisual constituents, i.e., “da”. In other cases, they might consider the stimulus to be closer to either “ba” or “ga” or other syllables and respond accordingly.

Figure 2.6. Two-dimensional syllable feature space illustrating audiovisual integration in McGurk perception.

In forced-choice condition, subjects compare their percept with the given response options and go with the most likely option. In case of open-choice condition, they compare the percept with their existing syllable representations, and go with the most likely representation.

We did not find an effect of talker gender or of participant gender, consistent with previous reports for McGurk syllables (Irwin, Whalen, & Fowler, 2006) and visual-only phonemes (Strelnikov et al., 2009). It is possible that individuals compensate for talker differences by perceiving the McGurk syllables as belonging to the same linguistic class, or perceptually normalize the stimuli (Bladon, Henton, & Pickering, 1984; Johnson, 1991; Kuhl, et al., 1991; Ladefoged & Broadbent, 1957; Strand, 1999), but it is beyond the purview of this study. No participant gender effects on perception of McGurk
syllables echo the findings of previous studies by Aloufy et al. (1996) and Irwin et al. (2006).

Our visual-only results confirm that visual “da” and visual “ga” are easily confusable, whereas visual “ba” is distinct (Binnie, Montgomery & Jackson, 1974; Erber, 1975; Fisher, 1968; Lucey, Martin, & Sriradharan, 2004).

Despite the increase in McGurk perception in forced-choice responses, the variability across subjects and stimuli are still persistent, similar to the findings of Basu Mallick et al. (2015). Together, these results demonstrate that differences in participants, stimuli, and experimental paradigms all contribute to the wide range of published estimates of the frequency of the McGurk effect. The high variability in the effect suggests that caution is necessary when comparing McGurk frequencies across groups or across studies in which any of these factors vary.

**Future Directions**

Since the McGurk effect is becoming more and more accepted as a universal index of audiovisual speech integration, it is critical that researchers are able to standardize methodology in studying the phenomenon such that they do not incorrectly accept differences in McGurk perception that might occur due to chance. In order to harness the vast literature studying the effect, it is critical to account for the procedural differences across studies which may affect the results. The present study is just one example that shows that McGurk perception can be affected by the response type decided upon by the researcher to collect participants’ responses. Other procedural differences, such as the operational definition of a McGurk response, and the effect of repetition of McGurk syllable needs to be investigated as well.
STUDY II. EFFECT OF VIDEO PLAYBACK RATE ON AUDIOVISUAL SPEECH PERCEPTION AS MEASURED BY THE MCGURK EFFECT

Introduction

In the technology driven world of today, face-to-face communication is also evolving. We are constantly having “face-to-face conversations” using applications such as Microsoft Skype, Apple FaceTime, and Google Hangouts. In 2011, on average there were 300 million minutes per month of video calling on Skype (Rao, 2011) amounting to 50% of all of Skype’s communication. We are using more and more remote education platforms such as Coursera, YouTube, and EdX, which deliver video lectures. In just 2 years’ time Coursera had about 5,266,200 students across the globe and 48,784,829 hours of videos watched (Coursera, 2013). Some applications have lags in transmission and are hence slowed down, and in others, individuals have the option to speed up videos or slow them down as they like to accommodate processing needs. Therefore, speed of communication has become a crucial component in such technology.

These video lectures and remote conversations, like any other face-to-face conversation includes the talker’s voice as well as the corresponding mouth movements. These visual cues correspond to the place of articulation, which are the only externally visible physical articulatory features (e.g. both lips touching to articulate a “ba” syllable, see https://s3.amazonaws.com/mcgurkstimuli/3.1.mp4) (Fisher, 1969). Integrating these audiovisual cues results in coherent speech perception. Previous studies have found that having access to the visual information in addition to the auditory information improves speed and accuracy of speech perception (Erber, 1969; Grant & Seitz, 2000; Stein & Meredith, 1993; Sumby & Pollack, 1954), especially if the auditory information is noisy.
However, visual cues can sometimes influence auditory perception even in the absence of noise. McGurk effect is one such phenomenon (McGurk & MacDonald, 1976), in which perception of an auditory syllable (e.g. “ba”) is altered (e.g. “perceived as “da”) when it is paired with a different visual syllable (e.g. “ga”) (refer to https://s3.amazonaws.com/mcgurkstimuli/2.8.mp4 for a demo of the McGurk effect).

Previous studies investigating the speed of audiovisual speech on perception, focused on either the auditory, or visual, or, the incongruity in speed between the two modalities (Brancazio & Miller, 2005; Munhall, Gribble, Sacco, & Ward, 1996). They had modified the speeds of each modality separately. For example, in one such study (Brancazio & Miller, 2005) selectively modified the playback rate of the visual component of audiovisual syllables by removing frames, or duplicating frames to create fast, and slow visual speech stimuli, that the researchers paired with auditory speech at normal speed at varying voice-onset times (duration of the beginning of a consonant and vocal fold oscillation, Kewley-Port & Preston, 1974). They found that when the slow visual speech stimuli paired with auditory speech at normal speed, there was a slower shift in the voicing boundary (from voiced “d” to voiceless “t”) relative to the fast visual speech stimuli even if the responses were unaffected by the McGurk illusion (from voiced “b” to voiceless “p”). This effect was previously shown to occur for congruent syllables is referred to as the visual rate effect (Green & Miller, 1985).

Another previous study (Munhall et al., 1996) examined the effect of varying speaking rates on perception of the McGurk effect. They wanted their stimuli to be ecologically valid and therefore created their stimuli by instructing talkers to speak fast, normally, or clearly. For the clear speech condition they instructed their talkers to speak
as though someone was having difficulty understanding them. This resulted in the syllable being uttered for a longer duration and yet close to natural speech. They found that the percentage of McGurk perception increased from fast to normal and clear speech conditions. Since Munhall et al.’s main focus was to investigate temporal asynchronies (the difference between the onset of the mouth opening and the voice onset time), and incongruities (e.g. fast audio paired with slow visual, and vice versa) on audiovisual speech perception, they did not delve into the details of the differences in speed of speech.

**Objective of the study**

In this study I aimed to investigate, how changing a stimulus feature affects audiovisual speech perception, specifically the speed of both audiovisual components of speech. When people speak too quickly, sometimes it is difficult to comprehend what they are saying. It is possible to circumvent this problem by speaking at a slower speed to clearly convey the message. I artificially changed the playback rate of a normally articulated syllable (1x) to half the normal speed (.5x, slow) or twice the normal speed (2x, fast) in order to investigate how changing speed affects audiovisual speech perception. Changing the playback rate in this manner affected both the auditory and the visual speech components. I used McGurk stimuli from 8 different talkers adopted from Basu Mallick et al., (2015) for this study. Previous studies have shown that there are wide inter-individual differences in McGurk perception (Basu Mallick et al, 2015; Nath & Beauchamp, 2012), and when individuals do not perceive the McGurk effect they generally tend to report the auditory syllable “ba” (Basu Mallick et al., 2015).
The main hypothesis for this study was that slowing the audiovisual stimuli down would result in an increase in the visual influence in speech perception, resulting in higher McGurk responses; and as the stimuli are speeded up, the speech cues will be compressed resulting in less McGurk perception.

**Materials and Methods**

*Participant recruitment*

This study was conducted using the crowdsourcing platform called Amazon Mechanical Turk (MTurk), which allowed for quick turnaround time for data collection. In Basu Mallick et al. (2015), the authors used MTurk to collect speech perception data and found the data to be of similar quality to a lab based McGurk perception study (Experiment 2, Basu Mallick et al., 2015). All participants consented to participate in the study following a Rice Institutional Review Board (IRB) approved protocol, and completed a demographic questionnaire prior to beginning the task. Since all participants were recruited online, there was no way of getting paper consents. Therefore, they were directed to an online consent form with all the details of the study, potential risks and benefits, and relevant contact information of the researchers and the IRB authorities.
Figure 3.1. The components participants had to complete before starting the main experiment on MTurk.

a. This is the demographic questionnaire as seen during the task. The highlighted text shows the link to the consent form.

b. The online consent form that popped up in a new tab once the participants clicked on the link on the main task page.
I released the task blocks in 3 different waves, and it resulted in recruitment of 110 unique participants, and repeat recruitment of 10 participants. Out of these 10 participants, 2 did the same task twice and therefore, their second performance on the task was excluded. The 8 remaining repeat participants did the two separate task blocks (please refer to stimulus section for more details). All participants who completed the task were given monetary compensation for their participation.

**Stimuli**

I used 8 McGurk stimuli (auditory “ba” + visual “ga”) created with eight different talkers from Basu Mallick et al. (2015), and grouped them into odd and even stimuli. These stimuli were previously ordered by average McGurk perception on open-choice responses. Each McGurk stimulus was presented 10 times. We aimed to recruit 60 people each in the odd and even task blocks. In addition to the 8 McGurk stimuli, 3 congruent audiovisual stimuli from another talker were included as a control measure to ensure that participants were not randomly responding to each stimuli. These congruent audiovisual stimuli (“ba”, “da”, and, “ga”) were each presented twice at each playback rate condition. The entire task was scripted using JavaScript and HTML5 (the script is available from https://github.com/debshila/msi/blob/master/OddMcg.html). The task was tested on different browsers including the latest versions of Google Chrome, Safari, Mozilla Firefox, and, Internet Explorer on both Mac and Windows platforms. The task however did not work properly on Mozilla, and therefore, the task only allowed individuals with Chrome, Safari, and Internet Explorer. In order to screen out Mozilla users for the task, I used a browser detection script adapted from http://www.javascripter.net/faq/browsern.htm. A demo video of a congruent stimulus was
used in order to allow participants to adjust the volume to be at a comfortable level and make sure that they scrolled to ensure that they could see the entire face of the talker (Fig 3.2). Similar to study I, this would control for the perceived intensity of the auditory stimuli. This demo video could be played multiple times. Participants had to play the demo video at least once in order to start the task.

Figure 3.2. The starting page of the task as seen on Amazon Mechanical Turk.

The playback rate of each stimulus was either reduced to 0.5x, or sped up to 2x, or remained as is (1x). At each playback rate, each McGurk stimulus was repeated 10 times, whereas each congruent stimulus was repeated twice. The stimuli were randomly interleaved and presented to each participant. Together with the congruent stimuli, this resulted in 2 task blocks consisting of the odd and even McGurk stimuli.

A 3-alternative forced-choice response format was adopted for this study, with options corresponding to the auditory (“ba”), McGurk (“da”), or visual (“ga”) percepts (Figure 3.3). These response options appeared as three buttons to the right of the video. In
order to control for participants prematurely responding to the video prior to watching it fully, the response options were deactivated for the duration of the video. This resulted in the participants responding to the stimuli only at the end of each video with a mouse click. Above the response options, a counter counted up to the number of trials in the task to indicate to the subject how many more trials they should expect before the end of the task. At the end of the task, the participants pressed the “Submit” button to end the task (see Figure 3.4).

Figure 3.3. The main task page as seen on Amazon Mechanical Turk, showing all 3 response options, and the trial counter above that.

Figure 3.4. The ending screen for the task.
Figure 3.5. Analysis of stimulus features.

a. Determination of the onset of the mouth movement using Final Cut Pro.
b. Determination of the onset of the auditory burst using PRAAT. The inset shows the point that marked the onset of the auditory burst.
Figure 3.6. Spectrogram of the auditory “ba” stimulus which is the auditory component of the McGurk stimuli. The horizontal axis corresponds to the time ranging from the onset of the stimulus to the end of the stimulus, while the vertical axis corresponds to the frequency (Hz) of the auditory waveform.
The McGurk stimuli were further analyzed to determine which properties may have been affected by the change in playback rate. The asynchrony between onset of the mouth movement and onset of the voice for normal, slow and fast stimuli was determined (Figure 3.5 a & b). For the normal videos, onset of the mouth movement was determined by inspecting the videos frame by frame using Final Cut Pro (version 10.2.2, Apple Inc.) in order to determine the frame at which the mouth and jaw get ready to articulate “ga”. Since the videos were shot at 30 frames/ second, this frame number was converted into seconds by dividing the frame number by 30. This was followed by the determination of the onset of the voice burst by examining the auditory component of the stimuli using PRAAT (version 6.0.08, Figure 3.6). The onset of the sound burst was determined by inspecting the point at which the auditory waveform deviates from the the 0 Hz. frequency. The asynchrony was measured by the difference between the mouth movement onset and the beginning of the sound burst (Miller & D’Esposito, 2005). The same process was repeated for the slowed down and speeded up videos, the only difference being that video rate for these videos was detected as 60 frames/ second.

Since I used a JavaScript attribute to manipulate the stimulus playback rate, I did not control the fundamental frequency for each stimulus at the different playback rates. Therefore, I examined the fundamental frequency (F0) for the auditory component of each stimulus at each playback rate using PRAAT with a script adapted from http://www.fon.hum.uva.nl/praat/manual/Script_for_listing_F0_statistics.html.

Data analysis

All data were analyzed using R statistical software (R Development Core Team, 2014). After each stimulus was presented, participants could report whether they
perceived either the auditory (‘ba’), visual (‘ga’), or McGurk percept (‘da’). For each
stimulus, the sum of auditory, visual and McGurk responses, summed to 100. Each
McGurk stimulus was presented 10 times at each playback rate. Estimates of McGurk
perception was averaged across all 10 trials.

Results

This study aimed to determine how different playback rates affect McGurk
perception. I had 3 different playback rates: slow (.5x), normal (1x), and fast (2x). I had
hypothesized that for slow playback rates, the visual influence will be greater, thereby
resulting in more McGurk perception. On the other hand, as the audiovisual cues get
more compressed in the fast condition, individuals would have less time to process the
audiovisual cues, and it would result in lesser McGurk perception.

Effect of playback rate

Contradictory to my hypotheses, I found that the slow playback condition
resulted in a reduction in McGurk perception from the normal condition [paired t-test,
t(109) = -6.87, p = 10^{-10}], whereas, there was no significant difference between the
normal and the fast conditions [paired t-test, t(109) = 0.54, p = 0.59] (Figure 3.7a). I
proceeded to examine if this effect was consistent across all participants. I found that the
majority of the participants had lower percentages of McGurk responses compared with
their responses in the slow condition \(r = 0.82, p = 10^{-16}\) (Figure 5b). This was also true
at the stimulus level, i.e., the average McGurk response for each stimulus was lower in
the slow condition compared with the normal condition \(r = 0.89, p = 0.003\); red X’s seen
in figure 3.7b]. When comparing consistency of responses in the fast and normal
conditions, I found that in the fast condition, average percentage of McGurk responses
for participants’ \( r = 0.89, p = 10^{-16} \), as well as stimuli \( r = 0.94, p = 0.0005 \), was strongly correlated to their responses in the normal condition. However, this high correlation between fast playback rate condition and normal stimuli merely indicates that individuals’ responses on the fast and normal playback conditions are not significantly different as evident from the previous reported t-value.

Figure 3.7. Change in McGurk frequency across different playback rates.
a. Each bar represents the percentage of McGurk responses across a different playback rate, averaged across participants, stimuli and trials. The error bar denotes the SEM across subjects. The p-values correspond to the paired-t test comparing the slow and fast playback rates with the normal playback rates.

b. The leftward scatterplot shows the relation between McGurk responses in the normal playback condition and the slow playback condition across individuals. Each black dot represents each individuals’ responses averaged across stimuli. The diagonal is the line of equality. The rightward plot shows the same for McGurk responses in the fast and normal playback rates.

c. The leftward scatterplot shows the relation between McGurk responses in the normal playback condition and the slow playback condition across stimuli. Each red ‘X’ represents average McGurk responses for each stimulus averaged across participants. The diagonal is the line of equality. The rightward plot shows the same for McGurk responses in the fast and normal playback rates.

Interestingly, I found more visual responses in the slow playback rate condition [paired t-test, $t(109) = 7.7, p = 10^{-11}$], and declined through the normal, and fast playback rate conditions (Figure 3.8). This pattern of increased visual responses in the slow relative to the normal playback rate condition was consistent across most participants, and across all stimuli (Figure 3.8). When examining the fast condition relative to the normal condition, I saw that the rates of visual responses are low for both normal and fast playback conditions for majority of the participants and the stimuli.
Figure 3.8. Change in visual responses across different playback rates.

a. The bar plot shows the average percentage of visual responses across slow, normal, and fast playback rates (error bars are SEM). The p-values correspond to paired t-tests between visual responses in slow and normal conditions, and fast and normal conditions.

b. The leftward scatterplot shows the relationship between visual responses in the normal and slow conditions, in which each transparent black dot corresponds to each participant’s average visual responses. The rightward plot shows the same for visual responses in the fast and normal playback rates.
c. The leftward scatterplot shows the relation between visual responses in the normal and slow playback condition across stimuli. Each red ‘X’ represents average visual responses for each stimulus, averaged across participants. The diagonal is the line of equality. The rightward plot shows the same for McGurk responses in the fast and normal playback rates.

Figure 3.9. Change in auditory responses across different playback rates.

a. The bar plot shows the average percentage of visual responses across slow, normal, and fast playback rates (error bars are SEM). The p-values correspond to paired t-tests between visual responses in slow and normal conditions, and fast and normal conditions.

b. The leftward scatterplot shows the relationship between visual responses in the normal and slow conditions, in which each transparent black dot corresponds to each
participant’s average visual responses, and each red ‘X’ is the average visual response of each stimulus. The diagonal line is the line of equality.

c. The leftward scatterplot shows the relation between auditory responses in the normal playback condition and the slow playback condition across stimuli. Each red ‘X’ represents average McGurk responses for each stimulus averaged across participants. The diagonal is the line of equality. The rightward plot shows the same for auditory responses in the fast and normal playback rates.

I found that the percentage of auditory responses remains unchanged across slow (paired t-test, $t(109) = -0.97$, $p = 0.34$), and fast playback rate conditions (paired t-test, $t(109) = 1.84$, $p = 0.07$) (figure 3.9).

When I examined the individual participants’ data sorted by their percentage of McGurk responses in the normal, condition I found that for participants in the extremes of the distribution (0%, or 100% McGurk responses) there was only one direction of movement in the other conditions because they were either at the floor or the ceiling. However, for individuals in the middle of the distribution, a clear pattern emerged in which McGurk responses increased, and visual responses decreased from slow to normal playback rate conditions, whereas auditory responses remained more or less stable across all three conditions (Figure 3.10 a).

An exploratory k-means clustering revealed four distinct patterns of responses across the three playback rates (Figure 3.10 b). Clusters less than 4, explained less than 80% of the variability. Therefore, I chose 4 clusters because it explained over 80% of the variability in responses, and yet the clusters returned interpretable results. In the first cluster, the percentage of fusion responses is high at the normal playback rate, and drops in the slow condition, while the visual responses increase from normal to slow; auditory responses are at the floor and do not change across playback rate. In cluster 2, the percentage of fusion and visual responses are similar in the normal condition, in the slow
condition, fusion decreases while visual increases; similar to cluster 1, auditory remains at the floor and unchanged across playback rates. In cluster 3, all three types of responses changes only little across playback rates, however, fusion is lesser than cluster 2 in normal condition, while auditory responses are highest and visual responses are least. In cluster 4 across playback rates, fusion responses are at floor and auditory responses at ceiling level. From this cluster analyses I found that individuals who make visual responses in the normal playback condition are likely to have more visual responses and less fusion in the slow relative to the normal playback condition.
Figure 3.10. Responses across different playback rates for the three different type of responses.

a. Three participants who either always perceive the McGurk effect (100% average McGurk frequency), never perceive the effect, or who perceive the effect 50% of the time in the normal playback rate condition are shown. Individuals who are at 100% McGurk perception at the normal speed condition, cannot have an increase in McGurk responses (ceiling effects), and those at 0% McGurk perception cannot have a decrease in their McGurk responses (floor effects).

b. An exploratory k-means cluster analysis showed 4 distinct patterns of responses across playback rates. As denoted by the n, clusters 1 through 4 include 44, 27, 21, and 18 participants respectively. The shaded region denotes the within cluster error (SEM).

The behavioral data were fit with a linear mixed effects model, with playback rate (normal, slow, and fast) as a fixed effect, stimuli, and the playback rate varying across participants, as random effects (table 3.1). The significance of the model coefficients was
tested with Satterthwaite approximations, showing clear effects of slow speech on McGurk responses [estimated 11% lesser in case of slow speech, $SE = 2\%$; $t(170.4) = -6.015, p = 10^{-8}$].

Summary of Overall Model

<table>
<thead>
<tr>
<th>Proportion of McGurk responses</th>
<th>Estimate (SE)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Slow</td>
<td>-11*** (2)</td>
</tr>
<tr>
<td>Fast</td>
<td>1 (2)</td>
</tr>
<tr>
<td>Constant</td>
<td>53*** (6)</td>
</tr>
</tbody>
</table>

Observations 2,640
Log Likelihood -163.56
Akaike Inf. Crit. 349.13
Bayesian Inf. Crit. 406.94

Note: ***p<10^{-6}

Table 3.1. Summary of the linear mixed effect model. This model includes McGurk responses as the outcome variable and playback rate as fixed effects variable, and stimuli, and the effect of playback rate varying across participants as random effects. The slow playback rate condition is estimated to result in 11% less McGurk responses relative to the normal playback rate.

Figure 3.11. Mean asynchrony values averaged across all stimuli at each playback rate. Error bars correspond to the SEM.
I wanted to examine what features of the stimuli changed when I changed their playback rate. A strong contender that may have resulted in the differences in the McGurk responses across the different playback rate conditions was the asynchrony between the visual mouth movements and the voice onset (van Wassenhove, Grant, & Poeppel, 2007). When stimuli were slowed down, the asynchrony between the auditory and visual cues were extended, resulting in the visual cues appearing perceptibly earlier than the auditory cues [mean = 294.5 ms; SD = 154 ms; figure 3.11.]. The average difference between the asynchrony in the slow and normal condition was 171 ms [paired t-test, \( t(7) = 3.36, p = 0.01 \)]. Since most of the onset asynchronies for the slowed down stimuli were near or over the 225 ms, this may have resulted in the stimuli not being integrated. On the other hand, there was not a strong perceptible difference in the asynchrony between the fast and the normal playback rate condition [normal vs. fast (M ± SD): 123±34 ms vs. 70 ms ± 20ms]. The average difference between the asynchrony in the fast and normal condition was smaller 54 ms [paired t-test, \( t(7) = -5, p = 0.002 \)]. Therefore, the responses in the normal and slow conditions did not change noticeably.

![Figure 3.12. Fundamental frequency (f_0) measured for each stimulus at different playback rates.](image-url)
I found that there was no perceptible difference among the fundamental frequencies at each playback rate level for each stimuli (figure 3.12), the maximum absolute difference being 50Hz which is much smaller than the minimum 100Hz needed to be detected by humans (Klapuri, 2003).

Responses to the congruent stimuli were near ceiling with the audiovisual “ba”, “da”, and “ga” stimuli resulting in (mean ± SEM) 97% ± 0.07%, 89% ± 1.5%, and 98% ± 0.5% accurate responses averaged across the three playback rates.

*Replicating previous findings*

In a previous study examining the effect of response type (open vs. forced-choice) on percentage of McGurk responses, I had used the same stimuli as in this study (Basu Mallick, et al., 2015). I compared the McGurk frequencies from the previous study (Basu Mallick et al., 2015) with the findings of this study to ensure that the average McGurk responses on the same stimuli were consistent between the two studies.

I found that there was wide variability in McGurk perception across individuals in the normal condition ranging from 0% McGurk response to 100% McGurk responses. This finding replicates the finding from a previous study (Basu Mallick et al., 2015).

I found that the average percentages of McGurk responses were lower in the current study relative to the previous study (previous study vs. current study: 69% vs. 51%). Examining the average percentage of McGurk responses revealed that the reduced percentage from the old to the current study was true in case of all the 8 stimuli that were tested (see figure 3.11a).
Possible sources of this reduction in McGurk responses in the current study can be attributed to several procedural differences between the two studies including the fact that in the present study, the same stimulus was presented at different playback rates, whereas, in the previous study, the stimuli were presented at the normal playback rate. To examine if the presentation of the same stimulus at different playback rates affected rate of McGurk responses, I compared the difference in McGurk responses between the old and new study with the t-values obtained from comparing McGurk responses at slow and normal playback rates (since the fast playback rate did not result in responses different from the normal condition, the difference between slow and normal conditions was used) and found a high correlation between them (see figure 3.11b; \( r = 0.51, p = 0.2 \)). This high correlation helps establish that when presented with the same stimulus in different playback rates affects the percentage of McGurk responses.
Figure 3.13. Percentage of McGurk responses from previous study (Basu Mallick et al., 2015) and current study.

a. Each bar represents the percentage of McGurk responses for each stimulus averaged across participants and trials. The dark grey bars are estimates from the old study, and the light grey bars are estimates from the current study. The error bars are the standard error of mean.

b. Relationship between the difference in previous and current percentages of McGurk responses and the t-values comparing the difference between McGurk percentages in the slow and normal playback speeds. Each dot corresponds to each of the stimulus. The dashed line corresponds to the line of best fit.
Discussion

As we move toward a world predominated by technology, our social interactions are becoming more dependent on advanced technology, and talking to others on the computer is becoming a norm. Understanding how humans process audiovisual cues mediated by technology will make communication technology more user centered and accommodate human mental models.

The main focus of this study was to determine how a change in a specific stimulus feature, viz., speed of speech (by changing playback rate) affects audiovisual speech perception as measured using the McGurk effect. Initially, I had hypothesized that McGurk responses would increase when speech is slowed down. My findings were completely different from my hypothesis. I found that McGurk perception decreased when speech is slowed down, and it remained unchanged from normal speech when speech was sped up.

In natural speech, the onset of the mouth movements occurs prior to the onset of the auditory speech (Smeele, Sittig, & Van Heuven, 1994). When speech is slowed down, the visual place of articulation is shown for a prolonged period of time, and on the other hand, the auditory onset is further delayed. Since the participants had a relatively longer exposure to the visual cue, long before the auditory stimulus began, they reported the visual stimulus. Our findings differ from the findings of Munhall et al. (1998), who found that McGurk responses decreased from clear to normal to fast speech. However, they modified the speed of speech by instructing their talkers to speak clearly or, speak fast.
One explanation of these findings can be based on the causal inference framework. According to the causal inference perspective, when individuals infer that there is a unitary source for both the auditory and visual cues, they are more likely to integrate. On the other hand, if individuals infer that the auditory and visual cues emerge from 2 disparate sources, then they are less likely to integrate. van Wassenhove et al. (van Wassenhove et al., 2007) found that for speech stimuli, integration was most likely to occur within a window ranging from 35 ms when auditory led visual to 225 ms when visual led auditory speech. Based on the causal inference framework and van Wassenhove et al.’s (2007) finding, if the asynchrony between the auditory and the visual cues fall within 0 - 225 ms of the visual cue leading the auditory (or “common cause window”), they will be inferred to have a common cause, and therefore, more likely to be integrated (Magnotti, Ma, & Beauchamp, 2013). In the normal speed condition, the natural asynchrony between visual and auditory cues is maintained. Therefore, the audiovisual cues are within the common cause window, and result in audiovisual integration, or a “da” percept (fig 3.12a). In the slowed down condition, the auditory component occurs outside the integration window; and the place of articulation is displayed for a prolonged period of time, resulting in the participants inferring that the audiovisual cues originate from two different sources, and making a decision in favor of the visual cue (figure 3.12b). For individuals who do not show this effect as seen in the last 2 clusters (figure 3.10b), they may be giving little to no weight to the visual information at the normal playback condition to begin with and therefore are unaffected in the slow playback rate condition.
Figure 3.14. Possible explanation for the effect of changing playback rates of videos.

a. When the visual and auditory cues occur well within the integration window it is more likely to result in an audiovisual percept.

b. When the auditory cues occur beyond the integration window, it will more likely to result in a visual percept.

c. Illustration of asynchrony values for one stimulus (#2.5) depicted with respect to the asynchrony window framework.

A previous study from my lab (Magnotti et al., 2013) attempted to answer this question in the causal inference framework. In this study, the authors manipulated the reliability of the visual cues, and the onset asynchrony of the audiovisual cues, in a synchrony judgment task. When visual cues were less reliable, the authors found that the sensory noise parameter for subjects increased, and they were more likely to report synchronous for a broader range of onset asynchronies (0 to 267 ms visual leading auditory) relative to when the visual cue is reliable (67 ms to 133 ms visual leading auditory).
auditory). In the low visual reliability condition, the onset of the visual cue is harder to estimate, due to which they are more likely to consider that the audiovisual cues were generated by the same talker, and are more likely to integrate over a broader range of asynchronies.

McGurk responses were not found to change significantly from normal to fast playback rate. This may be attributed to the fact that 2x speed is not fast enough to be perceived as sufficiently sped up. Perhaps, if I was able to speed up the speech to higher levels, for e.g. 4x, then there may be perceptible changes in speech. The stimuli used for this study were all created while the talkers were asked to clearly articulate the syllable they were saying, thus the speed of speech may have been slightly slow at the normal state itself. A recent study found that there were greater articulatory movement for clear speech (Tang, Hannah, Jongman, Sereno, & Wang, 2015), which may have resulted in another advantage for the visual cues, especially when slowed down.

Aligned with the findings of Basu Mallick et al. (2015), I found large variability in the McGurk perception across individuals in case of stimuli with natural speed, and when individuals did not perceive the McGurk effect, they perceived the auditory “ba” component. This finding is similar to a previous study (Basu Mallick et al., 2015) that found that when individuals did not perceive the McGurk illusion, they primarily reported the auditory stimulus, and very rarely reported the visual response. In this study I find that auditory responses remain stable across different playback rates. A similar pattern was observed when individuals saw the sped up stimuli. However, when stimuli are slowed down, individuals reported the visual component more often when they did not perceive the McGurk effect. I inferred that during the slow stimuli, the visual cues
appeared sufficiently before the auditory cues such that individuals’ decided that the stimulus was closer to the visual component of the McGurk stimulus. This increased asynchrony falls beyond the temporal binding window (Stevenson et al., 2012); a window of time discrepancies for which individuals still integrate the audiovisual cues and perceive them to be simultaneous (Noel, Wallace, Orchard-Mills, Alais, & Van Der Burg, 2015).

Presently, YouTube, Coursera and other video sharing services allow the videos to be slowed down to .5x and speeded up to 2x speed. Our findings may be adopted by these services to determine the appropriate upper and lower limits to changing the playback rates. A slow playback rate of .5x should not be used for videos of individuals talking because viewers may be distracted by the lag between the audiovisual cues.

An interesting finding is that seeing the same stimulus in multiple conditions appears to have an effect on the percentage of McGurk responses. Participants had fewer McGurk reports for slowed down stimuli, and exposure to the slowed down stimuli was found to affect their responses for stimuli played at normal speed. This reduced percentage of McGurk responses alludes to the contextual effects in McGurk perception (Nahorna, Chandrashekara, Berthommier, & Schwartz, 2013), in which preceding context affects integration of audiovisual cues. In the Nahorna et al. studies (Nahorna, Attigodu, Berthommier, & Schwartz, 2013; Nahorna, Berthommier, & Schwartz, 2011, 2012), exposure to an “incoherent” context in which auditory syllables are paired with the video of a random sentence, results in a decrease in McGurk perception. Presentation of the slow stimuli results in a breakdown of the integration process, which is analogous to the
incoherent context used in the Nahorna et al. studies (2011, 2012), resulting in the decrease in McGurk responses.

**Future Directions**

The next phase in this study will be to extend it while including a playback speed of 4x to determine if that helps to increase the contrast between speech at normal and fast playback rate conditions. Also, I would like to include intermediate speed levels and determine the psychophysical functions for the three response types at each playback rate. This will allow me to determine at what points perception shifts from visual to McGurk and then auditory for McGurk audiovisual pairings.

Some studies of the McGurk effect use VCV (vowel-consonant-vowel) syllables such as and auditory “aba” + visual “aga” pairing (Gentilucci & Cattaneo, 2005; Keil et al., 2012), instead of the CV (consonant-vowel) syllable pairings used in this study. It is possible that a playback rate manipulation for these VCV syllable pairs would result in reduced any potential asynchrony effects because the consonant is placed between the vowels, than observed with the CV syllables in this study. Future investigation is necessary to observe any potential differences between these different syllable pairs.

An interesting finding is that seeing the same stimulus in multiple conditions appears to have an effect on the percentage of McGurk responses. Since the slowed down stimuli are perceptibly different from the normal stimuli, and participants have fewer McGurk reports for this type of stimulus, affects their responses in the normal condition. Perhaps separating the stimuli with different playback rates into task separate task blocks can help overcome this issue in future experiments.
REFERENCES


Gruenstein, a, Mcgraw, I., & Sutherland, a. (2009). A self-transcribing Speech Corpus: Collecting Continuous Speech with an Online Educational Game. Speech and Language Technology in Education (SLaTE).


12th International Conference on Auditory-Visual Speech Processing, Annecy, France. Retrieved from http://hal.archives-ouvertes.fr/hal-00941306/


APPENDIX

To view the scripts for the Experiment 3 Task please check: https://github.com/debshila/msi/blob/master/OddMcg.html

**R Data Analysis scripts**

*Functions needed to analyze the data*

library(reshape2)

#--------Functions----------

# builds matrix with mean and sd
m_sd = function(mat) {
  mat = as.matrix(mat)
  res = cbind(colMeans(mat, na.rm = T), apply(mat, 2, sd, na.rm = T))

  colnames(res) = rep(c("mean", "sd"), length.out = ncol(res))

  return(res)
}

# To organize the respective columns into rows
searchCol = function(pattern, dataset, ignoreCase) {
  require(reshape2)
  matched.col = list()
  idx = grep(pattern, colnames(dataset), ignore.case = ignoreCase)
  matched.col = as.list(rbind(dataset[,idx]))
  return(melt(matched.col))
}

# Process individual data to organize by pattern into columns
ind.process = function(df, pattern = list('trial', 'stimulus', 'PlaybackRate', 'Response', 'ReactionTime')){
  subj.dat = lapply(pattern, searchCol, dataset = df, ignoreCase = TRUE) # df = dataframe
  ind.sub = do.call(cbind, subj.dat)

  # Taking the value columns out
  ind.sub = ind.sub[,grep('value',colnames(ind.sub))]
  colnames(ind.sub) = c('trial', 'sorted_stimuli_names', 'sorted_factor', 'response', 'rt') # rt = reaction time; sorted_factor = playback rate
ind.sub$sorted_stimuli_names = gsub('https://s3.amazonaws.com/mcgurkstimuli/", ",
ind.sub$sorted_stimuli_names)

#Ordering by trial number
ind.sub = ind.sub[order(ind.sub$trial),]
ind.sub$id = rep(df$AssignmentId, length(unique(ind.sub$trial)))
ind.sub$wid = rep(df$WorkerId, length(unique(ind.sub$trial)))
return(ind.sub)
}

#Remove NAs from the subject matrices
num_not_na = function(x) sum(!is.na(x))

get_mcg_movie_proportions = function(resp.id=2, pbr=1.0)
  lapply(as.numeric(substr(mcg.movies,1,3)), function(movies)
    sapply(sub.resp.count, function(wid)
      idx = wid[,1] == movies & wid[,2] == pbr
      if(any(idx)){
        return(wid[idx,(resp.idx + 2)]/ sum(wid[idx,3:5]))
      } else {
        return(NA)
      }
    )
  )

get_mean_mcg_movie_proportions = function(resp.idx=2, pbr=1.0) {
  rowMeans(matrix(unlist(get_mcg_movie_proportions(resp.idx, pbr)), nrow=110),
  na.rm=TRUE)
}

get_cong_movie_proportions = function(
  resp.idx=2, pbr=1.0)
  lapply(as.numeric(substr(cong.movies,1,3)), function(movies)
    sapply(sub.resp.count, function(wid)
      idx = wid[,1] == movies & wid[,2] == pbr
      if(any(idx)){
        return(wid[idx,(resp.idx + 2)]/ sum(wid[idx,3:5]))
      } else {
        return(NA)
      }
    )
  )
}
get_congruent_by_pbr = function(movie = 1, resp.idx = 1){
  movie.idx = as.numeric(paste0('3.', movie))
#lapply(as.numeric(substr(cong.movies, 1,3)), function(movies){
  lapply(pbrate, function(pbr){
    sapply(sub.resp.count, function(wid){
      idx = wid[,1] == movie.idx & wid[,2] == pbr
      if(any(idx)){
        return(wid[idx, resp.idx+2]/ sum(wid[idx, 3:5]))
      } else{
        return(NA)
      }
    })
  })
#  })
}

get_mean_congruent_by_pbr = function(movie = 1, resp.idx = 1){
  data = apply(matrix(unlist(get_congruent_by_pbr(movie, resp.idx)), nrow = 110), 2, m_sem)
  colnames(data) = pbrate
  return(data)
}

sem = function(x) sd(x, na.rm = TRUE)/ sqrt(num_not_na(x))

m_sem = function(x) c(mean = mean(x, na.rm = TRUE), sem = sd(x, na.rm = TRUE)/sqrt(num_not_na(x)))

p.adj.diff = function(m1, m2, padj.method = 'fdr', num.trials = 40){
  p.adjust(mapply(function(f1, f2) {
    prop.test(c(f1*num.trials, f2*num.trials), c(num.trials, num.trials))$p.value
  }, m1, m2), method = padj.method)
}

binom.p.adj = function(m1,m2, padj.method = 'fdr') {
  p.adjust(mapply(function(f1, f2) {
    binom.test(c(round(f1 * 40),round(f2 * 40)), c(40,40), alternative = 'two.sided', conf.level = .95)$p.value
  }, m1, m2), method = padj.method)
}

is_within = function(x, range) {
  l = min(range)
  u = max(range)
  sapply(x, function(xi) (xi >= l) & (xi <= u))
}
```r
compare = function(x, range) {
  if (x > max(range)) return (3)
  if (x < min(range)) return (1)
  return (2)
}

fix_range = function(x, l, u) {
  x[x<l] = l
  x[x>u] = u
  return(x)
}

oldData = function() {
  oc = read.csv("~/Dropbox/large_scale_mcgurk/mTurk_results/openchoice_r1r2_unequalN_Dec12.csv", row.names = 1)

  # Load up Forced Choice mTurk data
  fc = read.csv("~/Dropbox/large_scale_mcgurk/mTurk_results/forcedchoice_r1r2_unequalN_Dec12.csv", row.names = 1)

  dupes = which(sapply(row.names(oc), `%in%`, row.names(fc)))
  length(dupes)

  unique_names = c(row.names(oc), row.names(fc)[-dupes])
  length(unique_names)

  oc.sbj = colMeans(oc[, 1:8])
  fc.sbj = colMeans(fc[, 1:8])

  oc.msem = m_sem(oc[, 1:8])
  fc.msem = apply(fc[,1:8], 2, m_sem)
  colnames(fc.msem) = c(paste0('2.',c(6, 1, 8, 5, 7, 2, 4, 3)))
  (fc.msem = fc.msem[,order(colnames(fc.msem))])
}

#Converting frame number to seconds. Frame rate is 30 fps
convertFrame2time = function(x){
  x/30
}
```
cor2R = function(v1, v2) {
  (cor.test(v1, v2)$estimate)**2
}

# Calculate RMSE
rmse = function(dat, mod) {
  stop('don’t use this')
}

# Change column names for VOT or other variables that need to have the pbr as suffix
change_colname = function(x, pbr) {
  colnames(x) = paste0(names(x), '_', pbr)
  return(x)
}

# To assign worker IDs to the responses
assignId = function(data, pbr = 'a.normal') {
  lapply(1:8, function(stim) {
    data.frame(id = 1:110, pbr = pbr, stimuli = names(data)[stim], prop = data[[stim]])
  })
}

# Match async values
get_async = function(dat) {
  pbr = unlist(str_split(dat[2], '\.\.'))[2]
  idx = (async.melt$stimuli == dat[3]) & (async.melt$pbr == pbr)

  async.melt$asyncs[idx]
}

# This function gives the matching async for the new stimulus dataset
# Using the
get_async2 = function(dat) {
  idx = (stim.feat2$stimulus == dat[3]) & (stim.feat2$char.pbr == dat[2])
  stim.feat2$async_secs[idx]
}

get_VOT = function(dat) {
  idx = (stim.feat2$stimulus == dat[3]) & (stim.feat2$char.pbr == dat[2])
  stim.feat2$VOT[idx]
}

# To get data matching col and row id
get_dat = function(data, rowidx, colidx) {data[rowidx, colidx]}
# Get individual subjects mean plotted by

def get.ind.pbr.mean(idx):
    sub.resp.prop[idx]
    ind.sbj.data = subset(sub.resp.prop[idx], sub.resp.prop[idx]['moviePRFX'] %in% paste0('2.', 1:8))
    data = sapply(unique(ind.sbj.data[, 'pbr']), function(pbr){
        sapply(1:3,  function(resp.idx){
            mean(subset(ind.sbj.data[, 2 + resp.idx], ind.sbj.data[, 'pbr'] == pbr), na.rm = TRUE)
        })
    })
    colnames(data) = unique(ind.sbj.data[, 'pbr'])
    rownames(data) = c('auditory', 'mcgurk', 'visual')  # 'auditory'= ba, 'mcgurk'=da, 'visual' = ga
    return(data)

get.color = function(xy, criterion){
    col = 'gray'
    if(all(xy>0))
        col = 'orange'
    if(all(xy < 0))
        col = 'steelblue'
    return(col)
}

# Converts string to variable names
convert.string2var = function(st){
    eval(parse(text = st))  # st is string
}

gives the data for all playback rates based on the response type

def get.all.pbr.long(resp = 'da'){
    st = paste0('all_', resp)
    normal = assignId(convert.string2var(st), pbr = 'a.normal')
    slow = assignId(convert.string2var(paste0(st, '.slow')), pbr = 'b.slow')
    fast = assignId(convert.string2var(paste0(st, '.fast')), pbr = 'c.fast')
    unlink(st)
    normal = do.call(rbind, normal)
    slow = do.call(rbind, slow)
    fast = do.call(rbind, fast)
    all.pbr.data = rbind(normal, slow, fast)
    all.pbr.data$async = apply(all.pbr.data, 1, get_async2)
    all.pbr.data$resp = resp
    all.pbr.data$VOT = apply(all.pbr.data, 1, get_VOT)
    return(all.pbr.data)
}
#Add stim names to list elements
name.stim = function(dat) {
    names(dat) = paste0('2.', 1:8)
    return(dat)
}

labelByPbr = function(dat) {
    for (i in 1:nrow(dat)) {
        if (dat$pbr[i] == .5) {
            dat$char.pbr[i] = 'b.slow'
        } else if (dat$pbr[i] == 1) {
            dat$char.pbr[i] = 'a.normal'
        } else if (dat$pbr[i] == 2) {
            dat$char.pbr[i] = 'c.fast'
        }
    }
    return(dat)
}

#Get asynchrony values by stimulus and pbr
get.async.stim.pbr = function(data = async_details) {
    sapply(unique(data$stimulus), function(stim) {
        lapply(c(.5, 1, 2), function(pbrate) {
            idx = (data$stimulus == stim) & (data$pbr == pbrate)
            data$async_sec[idx]
        })
    })
}

convert.list2mat = function(data, col.num = 8, byrow = FALSE) {
    matrix(unlist(data), ncol = col.num, byrow = byrow)
}

Functions needed to plot the data

#--------Plot functions--------

#Get hexadecimal value based on colour name and alpha value
getAlphaRGB = function(colname, alpha) {
    c = col2rgb(colname)
    rgb(t(c), alpha = alpha, maxColorValue = 255)
}
# Plot correlation

```r
define_jplot = function(x, y, xlim, ylim, axes = TRUE, xat = c(xlim[1], 0, xlim[2]), ...) {
  #pdf(file = paste0(fig.path, name, '.pdf'), width = 2.5, height = 1.875)
  par(mar = rep(2.4, 4))
  plot(x, y, pch = 19, col = getAlphaRGB('black', 100), axes = FALSE, xlim = xlim, ylim = ylim, 
       ...) # runif(length(x), x-.01, x+.01), runif(length(y), y-.01, y+.01)
  max.x = round(max(abs(x)), 3)
  max.y = round(max(abs(y)), 3)
  axis(1, at = xat, las = 1)
  axis(2, at = c(ylim[1], 0, ylim[2]), las = 1)
  if(axes){
    abline(h = 0)
    abline(v = 0)
  }
}
```

# Compare the difference between slow and normal for each participant based on their adjusted p-values

```r
make_compare_plot_adjp = function(means, colors = c('steelblue', 'gray60', 'orangered'),
                                    adjp, pcrit = 0.05, yat = c(-.6, 0, .6), Ylim = c(-.6, .6)) {
  adjp[is.na(adjp)] = 1
  idx.sort = sort(means, index.return = T)$ix
  means = means[idx.sort]
  adjp = adjp[idx.sort]

  cols = rep(colors[2], length(means))

  cols[adjp < pcrit & means > 0] = colors[3]
  cols[adjp < pcrit & means < 0] = colors[1]

  idx.sort = sort(sapply(cols, function(cl) which(cl == colors)), index.return = T)$ix

  #pdf(file = paste0(fig.path, fname, '.pdf'), width = 3, height = 1.875)
  par(mar = rep(0.2, 4))
  plot(jitter(means[idx.sort]), col = cols[idx.sort], lwd = 3, type = 'h', axes = F, ylab = '', xlab = '', ylim = Ylim)
  abline(h = 0)
  axis(2, tcl = -.2, las = 1, at = yat, labels = F)

  print(table(cols))
}
```

# Compare the difference between slow and normal for each participant

```r
make_compare_plot = function(means, colors = c('steelblue', 'gray60', 'orangered')) {
```
col = colors[sapply(1:length(means), function(ii) compare(means[ii], subjCuts[ii]))]

idx.sort = sort(sapply(col, function(m) which(m == colors)), index.return=T)$ix

par(mar = rep(0.2,4))
plot(means[idx.sort], col=col[idx.sort], lwd=2, type='h', axes=F, ylab="", xlab="",
ylim=c(-.6, .6))
abline(h=0)
axis(2, tcl=-.2, las=1, at=c(-.6, 0, .6), labels=F)

print(table(col))

#Draws error bars on barplots
ebars.y = function(x, y, upper, lower = upper, length = 0.01, ...) {
arrows(x, y + upper, x, y - lower, angle = 90, code = 3, length = length, ...)
}

###Plots the mean and error bars for the difference between slow and normal, and fast
and normal pbr
make_speed_comparison_plot = function(sn.mse, fn.mse) {
#pdf(file = paste0(fig.path, name, '.pdf'), width = 1.5, height = 1.875)
par(mar = rep(0.2, 4))
plot(1, type='n', ylim=c(-.3, .3), xlim=c(0.125, 2.475), axes=F)  #
xpos = barplot(c(sn.mse[1], fn.mse[1]), col='gray20', border=NA, axes=F, add=TRUE, width = .5, space = c(1.25))
ebars.y(xpos, c(sn.mse[1], fn.mse[1]), c(sn.mse[2], fn.mse[2]), col='gray20')
lines(c(.05, 2.475), rep(0,2), col='gray50')
axis(2, tcl=-.2, las=1, at=c(-.3, 0, .3), labels=F)
}

##Plot lines and points based on response categories
make.pbr_resp_plot = function(means, colors=c('grey20', 'orangered', 'steelblue')) {
#sems,
par(mar = rep(1,4))
plot(1:3, 0:2/2, type='n', xlab = '', ylab = '', axes = FALSE, ylim = c(0,1))
for(ii in 1:3) {
#ebars.y(1:3, means[ii], sems[ii], col=colors[ii], length = 0.05)
points(means[ii], pch=20, col=colors[ii], type='o')
}
axis(2, at = c(0,.5,1), las = 1, tcl = -.2)
axis(1, at = 1:3, labels = c('slow', 'normal', 'fast'))
}
# Plot mean responses for individual subjects by different response category, auditory = grey30, visual = steelblue and mcgurk = orangered

```r
single.sub.meanplot = function(mat = by.pbr){
  par(mar=rep(0.2,4))
  plot(x = c(.5,1,2), mat[,3], lwd = 1.5, type = 'b', axes = FALSE, xlab = '', ylab = '', ylim = c(0,1), col = 'orangered')
  ebars.y(x= c(.5,1,2), mat[,3], mat[,4], lwd = 1.5, mat[,4], col='orangered')
  points(x = c(.5,1,2), mat[,1], type = 'b', col = 'grey30',lwd = 1.5)
  ebars.y(x= c(.5,1,2), mat[,1], mat[,2], mat[,2], col='gray30', lwd = 1.5)
  points(x = c(.5,1,2), mat[,5], type = 'b', col = 'steelblue', lwd = 1.5)
  ebars.y(x= c(.5,1,2), mat[,5], mat[,6], mat[,6], col='steelblue', lwd = 1.5)
  axis(1, las = 1, c(0.5,1,2))
  axis(2, las = 1, c(0,.5, 1))
}
```

# This function plots the correlation between the fusion rate for each movie and the difference between normal and fast/slow pbr, with jitter

```r
plot.stim.data = function(normal, mat, mat_diff, jitter = TRUE){
  par(mfrow=c(2,4))
  rownames(normal) = 1:nrow(normal)
  rownames(mat) = 1:nrow(mat)
  rownames(mat_diff) = 1:nrow(mat_diff)
  normal.subset = vector()
  mat_normal.subset = vector()
  sapply(1:ncol(normal), function(mov){
    normal.subset = subset(normal[,mov], normal[,mov] > 0)#& normal[,i] < 1
    mat_normal.subset = mat_diff[rownames(mat_diff)%in%names(normal.subset), mov]
    if(jitter)
      plot(jitter(normal.subset), jitter(mat_normal.subset), pch = 16, col = 'grey30', axes = FALSE, xlim = c(0,1), ylim = c(-1,1), xlab = 'Slow - Normal', ylab = 'McGurk perception')
    else
      plot(normal.subset, mat_normal.subset, pch = 16, col = 'grey30', axes = FALSE, xlab = c(0,1), ylim = c(-1,1), xlab = 'Slow - Normal', ylab = 'McGurk perception')
      abline(lm(mat_normal.subset~normal.subset), lty = 2)
      axis(1, las = 1, at = c(0,.5,1))
      axis(2, las = 1, at = c(-1,0,1))
      print(cor.test(normal.subset, mat_normal.subset))
  })
}
```

# Plot overall mean and sem averaged across stimuli and subjects

```r
barplot_overall_msem = function(ebars = TRUE, mean,sem, fname, width = 1.5, limy = c(0,1), height = 1.875, cols = c('purple', 'grey80', 'yellowgreen'), yat = c(0,.5,1), beside = TRUE, border = FALSE,...){
```
as_pdf(file = fname, h = height, w = width, {
  par(mar = rep(.1, 4))
  x = barplot(mean, border = border, beside = beside, yaxt = 'n', ylim = limy, col = cols, ...
  if(ebars)
    ebars.y(x, mean, sem, col = cols)
  axis(2, las = 1, at = yat)
})
}

#---------------------Individual subject data---------------------
ind.pbr_resp_plot = function(means, colors=c('purple', 'grey80', 'yellowgreen')) {  #sems,
  plot(1:3, 0:2/2, type='n', xlab = '', ylab = '', axes = FALSE, ylim = c(0,1))
  for(ii in 1:3) {
    points(means1[,ii], pch=19, col=colors[ii], cex = 2.5, type = 'o', lwd = 1.5)
  }
  axis(2, at = c(0,.5,1), las = 1, tcl = -.2)
  axis(1, at = 1:3, labels = c('ba', 'da', 'ga'))
}

##
corplot = function(x,y, color = getAlphaRGB('black', 100), ...)  {
  #pdf(file = paste0(fig.path, name, '.pdf'), width = 2.5, height = 1.875)
  par(mar=rep(.2,4))
  plot(x,y, pch=19, col = color, axes = FALSE, xlim = c(0,1), ylim = c(0,1), ...
  )#runif(length(x), x-.01, x+.01), runif(length(y), y-.01, y+.01)
  max.x = round(max(abs(x)),3)
  max.y = round(max(abs(y)),3)
  axis(1, at = c(0,0.5, 1), las = 1)
  axis(2, at = c(0,0.5, 1), las = 1)
}

#---------------------Wrapper function---------------------

#generate pdf of figures
as_pdf = function(file, w, h, expr) {
  fig.path = paste0(getwd(),'figures/')
  pdf(paste0(fig.path,file,'.pdf'), width=w, height=h)
  res = eval(expr)
  dev.off()

  return (invisible(res))
}

#-------Old--------
plot_stim_feat = function(x, y, lim_x = c(round(-max(abs(x)),2), round(max(abs(x)),2))){
plot(x, y, axes = FALSE, pch = 16, ylim = c(0,1), xlim = lim_x, main = names(x))
mtext(text = names(x), side = 3)
axis(2, at = c(0,5,1), las = 1)
axis(1, at = c(lim_x[1],0,lim_x[2]))
abline(lm(y ~ x), lty = 2)
}

corplot_stim_feat = function(pred = stim_feat_normal, y = all_async$da_normal_mean){
lapply(as.data.frame(1:ncol(pred)), function(feat){
plot_stim_feat(pred[feat], y)
})
}

plot.corr.stim.mean = function(x, y, fname, height = 1.5, width = 1.5){
as_pdf(file = fname, w = 1.5, h = 1.5, {
corplot(get.mean.stim(all_ga), get.mean.stim(all_ga.fast))
abline(c(0,1), lwd = 1)
})
}

Functions used to read data into R

#--------Reading data---------

###Getting the data
fig.path = '~/Google Drive/research/behavior/code/dataAnalysis/figures/
dat.path = '~/Google Drive/research/behavior/data/
files = list.files(path = dat.path, full.names = TRUE, pattern = 'Batch+')
all.files = lapply(files, read.csv, header = TRUE, stringsAsFactors = FALSE)

# remove people that were in the pilot and the even run
duplicates = sapply(all.files[[1]]$WorkerId, function(wid) {
x = which(wid == all.files[[2]]$WorkerId)
if(length(x) > 0)
  return(x)
return(NA)
})
duplicates = duplicates[!is.na(duplicates)]
all.files[[2]] = all.files[[2]][-duplicates,]
sapply(all.files, dim)

#Merging all files
all.batch = do.call(rbind, all.files)
# For some reason it generates blank rows from 61 to 130
# all.batch[all.batch == ""] = NA  # Converting all blank values to NA
# idx = apply(all.batch, 1, function(x) all(is.na(x)))  # Indexing all rows with all NA values
# all.batch = all.batch[idx,]  # removing rows with all NA values
ind.data = ind.process(all.batch)

# Components of df ind.data
movies = unique(sort(ind.data$sorted_stimuli_names))
pbrate = unique(sort(ind.data$sorted_factor))
resp = unique(sort(ind.data$response))
workerId = unique(ind.data$wid)
assignmentId = unique(ind.data$id)
mcg.movies = movies[grepl('2.+.mp4', movies)]
cong.movies = movies[grepl('3.+.mp4', movies)]

# I need means of each category of responses across trials for each subj
sub.resp.count = lapply(workerId, function(wid) {
  # for each unique subject
  sbj.dat = ind.data[ind.data$wid == wid,]
  # response per video
  smat = array(NA, c(length(movies)*length(pbrate), length(resp) + 2))
  colnames(smat) = c('moviePRFX', 'pbr', resp)
  for(i in 1:length(movies)){
    all_movies = sbj.dat[movies[i] == sbj.dat$sorted_stimuli_names,]
    for(j in 1:length(pbrate)){
      pbr = pbrate[j]
      q = length(pbrate)*(i-1) + j
      responses = all_movies$response[pbr == all_movies$sorted_factor]
      if(length(responses)>1)
        {smat[q, 3:5] = table(responses)}
      smat[q,1:2] = c(as.numeric(substr(movies, 1,3)[i]), pbr)
    }
  }
  return(smat)
})

# all.prop = get_mcg_movie_proportions(resp.idx = 1:3, pbr = 1)
sub.resp.prop = lapply(sub.resp.count, function(sub){
  cbind(sub[,1:2], (sub[,3:5]/ rowSums(sub[,3:5], na.rm = TRUE)))
})
Functions used for descriptive data analysis

#--------Setup----------

setwd '~/Google Drive/research/behavior/code/dataAnalysis/

#Packages to load
packages = c('stringr', 'ggplot2', 'reshape2', 'lme4')
lapply(packages, library, character.only = TRUE, logical.return = TRUE)

#Packages to load
files_to_source = paste0('salience_', c('functions', 'readData'), '.R')
lapply(files_to_source, source)

#--------Analysis--------

### compare rates of auditory ('da') across different playback rates

### Fusion responses
all_da = get_mcg_movie_proportions() #default resp.idx = 2, pbr = 1
all_da.fast = get_mcg_movie_proportions(pbr=2.0)
all_da.slow = get_mcg_movie_proportions(pbr=0.5)

names(all_da) = paste0('2.', 1:8)
names(all_da.slow) = paste0('2.', 1:8)
names(all_da.fast) = paste0('2.', 1:8)

#--------Take out individuals who have less than .05 mcgurk on all stimuli--------
keep_da_mat = keep_da(do.call(cbind, all_da))
slow_normal_red = slow_normal
sub.effect = rep(NA, nrow(slow_normal_red))
for(i in 1: nrow(slow_normal_red)){
  sub.effect[i] = mean(slow_normal_red[i,keep_da_mat[i,]])
}
hist(sub.effect)

sum(rowMeans(all_da_mat, na.rm=T) < 0.01)

keep_da = function(mat, criterion = 0.05) {
  mat> criterion & !is.na(mat)
}

slow_stim_msem = lapply(all_da.slow, function(stim) cbind(m_sem(stim), pbr = 0.5))
normal_stim_msem = lapply(all_da, function(stim) cbind(m_sem(stim), pbr = 1))
fast_stim_msem = lapply(all_da.fast, function(stim) cbind(m_sem(stim), pbr = 2))

stim_sems = sapply(1:8, function(stim){
  sems =
  c(slow_stim_msem[[stim]][2,1],normal_stim_msem[[stim]][2,1],fast_stim_msem[[stim]][2,1])
})
stim_means = sapply(1:8, function(stim){
  means =
  c(slow_stim_msem[[stim]][1,1],normal_stim_msem[[stim]][1,1],fast_stim_msem[[stim]][1,1])
})
sbj.da.slow = get_mean_mcg_movie_proportions(resp.idx = 2, pbr=0.5)
sbj.da.normal = get_mean_mcg_movie_proportions(resp.idx = 2, pbr=1.0)
sbj.da.fast = get_mean_mcg_movie_proportions(resp.idx = 2, pbr=2.0)

names(sbj.da.normal) = 1:110
names(sbj.da.fast) = 1:110
names(sbj.da.slow) = 1:110

sbj.da.normal.excl = sbj.da.normal[sbj.da.normal<=0.05]

#Calculating mean and sem for barplot
sbj.da.mean = c(mean(sbj.da.slow), mean(sbj.da.normal), mean(sbj.da.fast))
sbj.da.sem = c(sem(sbj.da.slow), sem(sbj.da.normal), sem(sbj.da.fast))

##All movies
normal.movie.msem = sapply(get_mcg_movie_proportions(), function(mov) m_sem(mov))
normal.stim = do.call(cbind, get_mcg_movie_proportions())
slow.stim = do.call(cbind, get_mcg_movie_proportions(pbr = .5))
fast.stim = do.call(cbind, get_mcg_movie_proportions(pbr = 2))

rownames(normal.stim) = 1:110

#------------------Previous study------------------
# Load up Open Choice mTurk data
older.msem = oldData()
old.new.mean = rbind(older.msem[1,], normal.movie.msem[1,])
old.new.sem = rbind(older.msem[2,], normal.movie.msem[2,])

#Comparing forced choice mcgurk from old study with current study
#t-values between slow and normal stimuli

slow_normal.t.val = sapply(1:8, function(stim){
  t.test(slow.stim[,stim], normal.stim[,stim], paired = TRUE)$statistic
})

oldData_msem = oldData()
normal_means = colMeans(normal.stim, na.rm = TRUE)
old_new_diff = oldData_msem[1,] - normal_means
#correlation between old_new difference and slow_normalT values to indicate the slow
#condition has a role in reducing
#mcgurk reports for current study
cor.test(old_new_diff, slow_normal.t.val)

#Normal vs fast and slow conditions
fast_normal = mapply(`-`, get_mcg_movie_proportions(pbr = 2.0),
get_mcg_movie_proportions(pbr = 1.0))
rownames(fast_normal) = rownames(normal.stim)
slow_normal = mapply(`-`, get_mcg_movie_proportions(pbr = 0.5),
get_mcg_movie_proportions(pbr = 1.0))
rownames(slow_normal) = rownames(normal.stim)

#t.test(rowMeans(fast_normal, na.rm=T))
#t.test(rowMeans(slow_normal, na.rm=T))
# t.test(rowMeans(slow_normal, na.rm = TRUE), rowMeans(fast_normal, na.rm =
# TRUE), paired = TRUE)

# t.test(rowMeans(na=T,fast_normal[idx.sorted_fusion,]))
# cor.test(1:110,rowMeans(na=T,fast_normal[idx.sorted_fusion,]))

da_slow_normal = sbj.da.slow - sbj.da.normal
da_fast_normal = sbj.da.fast - sbj.da.normal
t.test(da_fast_normal, na.rm=T)
t.test(da_slow_normal, na.rm=T)
t.test(da_slow_normal, da_fast_normal, na.rm = TRUE, paired = TRUE)
t.test(sbj.da.fast, sbj.da.normal, paired = TRUE)
### compare rates of auditory ('ba') across different playback rates

```r
sbj.ba.slow = get_mean_mcg_movie_proportions(resp.idx = 1, pbr=0.5)
sbj.ba.normal = get_mean_mcg_movie_proportions(resp.idx = 1, pbr=1.0)
sbj.ba.fast = get_mean_mcg_movie_proportions(resp.idx = 1, pbr=2.0)

sbj.ba.mean = c(mean(sbj.ba.slow), mean(sbj.ba.normal), mean(sbj.ba.fast))
sbj.ba.sem = c(sem(sbj.ba.slow), sem(sbj.ba.normal), sem(sbj.ba.fast))
```

#By stimuli and participants
```r
fast_normal.ba = mapply(`-`, get_mcg_movie_proportions(pbr = 2.0, resp.idx = 1),
                        get_mcg_movie_proportions(pbr = 1.0, resp.idx = 1))
rownames(fast_normal.ba) = rownames(normal.stim)
slow_normal.ba = mapply(`-`, get_mcg_movie_proportions(pbr = 0.5, resp.idx = 1),
                        get_mcg_movie_proportions(pbr = 1.0, resp.idx = 1))
rownames(slow_normal.ba) = rownames(normal.stim)
```

#By participants
```r
ba_slow_normal = sbj.ba.slow - sbj.ba.normal
ba_fast_normal = sbj.ba.fast - sbj.ba.normal
```

```r
t.test(ba_slow_normal)
t.test(ba_fast_normal)
t.test(ba_slow_normal - ba_fast_normal)
```

#Stimuli wise auditory responses
```r
all.ba = get_mcg_movie_proportions(resp.idx = 1, pbr = 1) #ba resp.idx = 1, pbr = 1
all.ba.slow = get_mcg_movie_proportions(resp.idx = 1, pbr = .5)
all.ba.fast = get_mcg_movie_proportions(resp.idx = 1, pbr = 2)

all.ba = name.stim(all.ba)
all.ba.slow = name.stim(all.ba.slow)
all.ba.fast = name.stim(all.ba.fast)
```

```r
ba_stim_means = rbind(sapply(all.ba.slow, mean, na.rm = TRUE),
                      sapply(all.ba, mean, na.rm = TRUE), sapply(all.ba.fast, mean, na.rm = TRUE))
ba_stim_sems = rbind(sapply(all.ba.slow, sem),
                     sapply(all.ba, sem), sapply(all.ba.fast, sem))
```

### compare rates of visual ('ga') across different playback rates

```r
sbj.ga.slow = get_mean_mcg_movie_proportions(resp.idx = 3, pbr=0.5)
sbj.ga.normal = get_mean_mcg_movie_proportions(resp.idx = 3, pbr=1.0)
sbj.ga.fast = get_mean_mcg_movie_proportions(resp.idx = 3, pbr=2.0)
```
suj.ga.mean = c(mean(suj.ga.slow), mean(suj.ga.normal), mean(suj.ga.fast))
suj.ga.sem = c(sem(suj.ga.slow), sem(suj.ga.normal), sem(suj.ga.fast))

t.test(suj.ga.slow, suj.ga.fast)

gga_slow_normal = suj.ga.slow - suj.ga.normal
gga_fast_normal = suj.ga.fast - suj.ga.normal

t.test(gga_slow_normal)
t.test(gga_fast_normal)
t.test(gga_slow_normal - gga_fast_normal)

#Stimuli wise auditory responses
all_ga = get_mcg_movie_proportions(resp.idx = 3, pbr = 1) #ba resp.idx = 1, pbr = 1
all_ga.slow = get_mcg_movie_proportions(resp.idx = 3, pbr = .5)
all_ga.fast = get_mcg_movie_proportions(resp.idx = 3, pbr = 2)

#Adding stim names to list elements
all_ga = name.stim(all_ga)
all_ga.slow = name.stim(all_ga.slow)
all_ga.fast = name.stim(all_ga.fast)

ga_stim_means = rbind(sapply(all_ga.slow, mean, na.rm = TRUE), sapply(all_ga, mean, na.rm = TRUE), sapply(all_ga.fast, mean, na.rm = TRUE))
ga_stim_sems = rbind(sapply(all_ga.slow, sem), sapply(all_ga, sem), sapply(all_ga.fast, sem))

#By stimuli and participants
fast_normal.ga = mapply('-', get_mcg_movie_proportions(pbr = 2.0, resp.idx = 3), get_mcg_movie_proportions(pbr = 1.0, resp.idx = 3))
rownames(fast_normal.ga) = rownames(normal.stim)
slow_normal.ga = mapply('-', get_mcg_movie_proportions(pbr = 0.5, resp.idx = 3), get_mcg_movie_proportions(pbr = 1.0, resp.idx = 3))
rownames(slow_normal.ga) = rownames(normal.stim)

#-------------------Stimulus features-----------------------------
# New asynchrony calculations with voice burst onset and mouth movement onset difference values
async.details = read.csv(paste0(path, 'async_details.csv'), header = TRUE)

allpbr.async = convert.list2mat(get.async.stim.pbr(data = async.details))
allpbr.async.msem = apply(allpbr.async, 1, mSem)