Noise reduction limits the McGurk Effect

by

Justin Deonarine

A thesis

presented to the University of Waterloo

in fulfilment of the

thesis requirement for the degree of

Master of Arts

in

Psychology

Waterloo, Ontario, Canada, 2011

Author's Declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Justin M. Deonarine

Abstract

In the McGurk Effect (McGurk & MacDonald, 1976), a visual depiction of a speaker silently mouthing the syllable [ga]/[ka] is presented concurrently with the auditory input [ba]/[pa], resulting in “fused” [da]/[ta] being heard. Deonarine (2010) found that increasing the intensity (volume) of the auditory input changes the perception of the auditory input from [ga] (at quiet volume levels) to [da], and then to [ba] (at loud volume levels). The present experiments show that reducing both *ambient noise* (additional frequencies in the environment) and *stimulus noise* (excess frequencies in the sound wave which accompany the intended auditory signal) prevents the illusory percept. This suggests that noise is crucial to audiovisual integration and that the McGurk effect depends on the existence of auditory ambiguity.

Acknowledgments

I would like to thank Dr. Derek Besner, Dr. Serje Robidoux, Shannon O'Malley, Stephanie Waechter, Dr. Jonathan Fugelsang and Dr. Katherine White for their help and input with the project, as well as their patience and guidance towards my development as a researcher. I would also like to thank Dr. Karl Borgmann, Shawn Stubitz, Nathaniel Barr and Dr. Myra Fernandes for their feedback and discussion. Special thanks to Jaclyn Fong for her assistance in collecting data.

Table of Contents

List of Figures

List of Tables

Introduction

The "McGurk Effect" is a well known, powerful speech perception illusion, which is often used as an example of audiovisual integration. It was first described by McGurk and MacDonald (1976), who simultaneously presented a visual clip of a speaker silently mouthing the syllable [ga] and an auditory input of [ba] (otherwise known as a "McGurk Stimulus"). The task for the participants was to report what syllable they heard. The "fused percept" of [da] was reported by 98% of the adults in the sample. MacDonald and McGurk would offer an explanation for this phenomenon in a future publication: "By responding to the common information[1] in both modalities, a subject would arrive at the unifying percept /da/..." (MacDonald & McGurk, 1978). Rosenblum and Saldana (1996) described how robust the effect is: "...the effect is quite striking: even when completely aware of the dubbing procedure, observers still report hearing a clear syllable which is (unconsciously) influenced by what they see." Sekiyama, Kanno, Miura and Sugita (2003) also comment on the strength of the effect: "The McGurk effect, speech perception altered by discrepant mouth movements, can be also seen as an example of humans' ubiquitous propensity to bind cross-modal inputs."

MacDonald and McGurk (1978) retested this phenomena using a variety of visual and auditory input combinations. Possible articulations or inputs included [da], [ta], [ga], [ka], [na], [ba], [pa] and [ma], creating 64 possible combinations. They concluded that any non-labial visual articulation (including [ga]) can greatly influence what is heard when a bilabial plosive (including [ba]) is presented as an auditory input simultaneously. This result is supported by the following statement: "When the combinations are of labial sounds with nonlabial lip movements... examples yield highly significant error rates [in identification of the auditory stimulus] from 30% to 100% with a mean rate of 73%...". (MacDonald & McGurk, 1978). In this same experiment, they also found that their original stimulus (visual [ga] and auditory [ba]) did not have the same strength as reported by McGurk and MacDonald (1976). Instead, they found that [da] was only reported by 64% of the sample. In addition, they found that [ba] was reported by 9% of the sample, while [ga] was reported by 27% of the sample. The discrepancy between the 1976 and 1978 papers were never explained or explored, nor why participants could perceive [ga] and [ba] instead of [da].

To better understand McGurk and MacDonald's explanation of their effect (and the common information between the syllables), we need to understand the structure of the speech waveform for each of the three syllables in question ([ga], [da] and [ba]). Liberman, Harris, Hoffman and Griffith (1957) proposed that plosive consonants which differ in place of articulation, such as the ones used in MacDonald and McGurk (1978), are distinguished by the slopes of the second formant (F2) transition of the auditory signal. The term "F2" indicates that it is the 2^{nd} lowest frequency band. It is a brief "transition" between the burst representing the plosive and the beginning of the vowel's F2 frequency. Using artificial CV syllables ending in [a], where the slope of the F2 transition rose ([ba]), fell ([ga]) or remained steady ([da]) over time, Liberman et al. asked participants to identify the ambiguous auditory input as either "ga", "da" or "ba". They found that these syllables were categorically perceived (meaning that there were no half-way or mixed percepts). They confirmed this finding using an ABX task, where stimuli were arranged in groups of three (an A stimulus, a B stimulus, and an X stimulus) and participants had to identify the X stimulus as matching either the A stimulus or the B stimulus. Participants could easily match the X stimulus to the appropriate A/B stimulus, but only if the A and B stimuli were not from the same category. If both A and B were from the same category (even with different F2 transition slopes), participants had difficulty identifying a difference between the two, and therefore struggled to match the X stimulus to either the A or B stimulus.

There are few researchers who have attempted to test the limits of the McGurk Effect (and audiovisual integrative speech perception as a whole). Many use the established effects to explore another topic, assuming that the fused percept always occurs. This is shown in many McGurk type studies, including those involving the McGurk Effect in children or infants (Rosenblum, Schmuckler & Johnson, 1997; Burnham & Dodd, 2004, Teinonen, Aslin, Alku, & Csibra, 2008), the effects of different primary languages or language development on the McGurk effect (Sekiyama & Tohkura, 1991; Sekiyama, 1997; 2007; Sekiyama & Burnham, 2008) and the effects of disabilities such as schizophrenia (Williams, Light, Braff, & Ramachandran, 2010) or hearing damage on the McGurk Effect (Rouger, Fraysse, Deguine, & Barone, 2008).

However, there are some who have tested the limits of the McGurk effect, intentionally or not. In a brain imaging study, Sekiyama, Kanno, Miura, and Sugita (2003) used two different volume levels in order to overcome the noise produced by a fMRI machine. Their behavioural results revealed that accurate auditory input identification responses ([ba]) became more frequent at the louder volume level (compared to the responses given at the quiet volume level) for both the auditory only and audiovisual conditions, but that the change was greater in the audiovisual condition. This implies that the visual influence had a greater effect at quiet volume level than it did at loud volume level. Jones and Jarick (2006) found that a temporal delay between the visual input and the auditory input affects what is reported by participants. They concluded that if the auditory input is presented more than 60 ms before or 180 ms after the visual articulation, the frequency of accurate auditory input identification responses ([ba]) increases, suggesting a temporal component to audiovisual integration.

Deonarine (2010) expanded on the behavioural methods used by Sekiyama et al. (2003) and attempted to test the limits of the McGurk Effect by more systematically manipulating the intensity of the auditory input (volume) during the presentation of a McGurk Stimulus. He concluded that [ga] is reported at the quiet volume levels, [da] at the intermediate volume levels, and [ba] at the loud volume levels. As mentioned above, all three of these syllables were reported by MacDonald and McGurk (1978) when a McGurk Stimulus was presented. However, they did not address why [ga] and [ba] were heard, whereas Deonarine's findings provide a set of conditions under which [ga] and [ba] are heard. Unfortunately, the volume conditions of the original experiments are not known.

Deonarine (2010) also carefully defined both components of the sound waveform: (1) Signal, the intended auditory stimulus and (2) Noise, the excess frequencies which accompany the Signal. He suggested that audiovisual integration occurs due to the use of noise frequencies. In the McGurk Effect, conflict is present (between the visual input and auditory input). Deonarine suggested that, using the guidance of the visual input, the frequencies from noise can replace the intended signal entirely (resulting in the percept of [ga]), part of the signal can be combined with noise frequencies (resulting in the percept of [da]) or the signal can be used while the conflicting visual input is ignored (resulting in the percept of [ba])[2]. It has been well established that noise can be integrated as part of the signal, often to alter the signal itself, or to fill in for missing information. Warren and Obusek (1971) described the Phonemic Restoration Effect, where a speech sound (either a phoneme or syllable in a sequence) can be replaced by an unrelated sound (such as a cough) without a listener being able to identify the missing or unrelated sound. It can be noted that the effect worked best with fricatives, as this manner of articulation is closer to the sound of noise than other manners. Kennedy, Chattaway and Chattaway's (1974) reanalysis of the Phonemic Restoration Effect suggests that this effect occurs because the frequencies of the noise (such as the cough) replace the frequencies of the missing speech sound. An important observation by both Warren et al., as well as Kennedy et al., is that there is no restoration effect when a silent gap is present (instead of a cough or white noise). Instead, participants were able to report the gap in the auditory stimulus. Syrdal-Lasky (1978) examined the effect of signal intensity and noise on the perception of stop consonants in a similar manner to Liberman, Harris, Hoffman and Griffith (1957). They presented synthesized CV syllables with F2 transitions either rising ([gae]/[kae]), remaining constant ([dae]/[tae]) or falling ([bae]/[pae]) in frequency over time, within 3 different conditions: (1) volume of 75 dB, (2) volume of 92 dB with 60 dB of noise to mask the signal and (3) volume of 92 dB. They report that between-category discrimination was best at 75 dB, less accurate in the 92 dB with noise task and worst at the 92 dB condition, with [b]/[p] being reported more frequently as performance on the between-category discrimination task worsened. These results suggest that (1) even categorical perception is stimulus intensity dependent and (2) noise can oppose the effects of increased stimulus intensity (leading to higher counts of [d]/[t] and [g]/[k] being reported than in the non-noise counterpart). The finding that stimulus intensity (or additional noise) can affect which stop consonant is heard using the same F2 transition slope suggests that stimulus intensity and noise are important for the perception of any ambiguous auditory signal.

To expand on the definition of "noise" in Deonarine (2010), two distinct types can be defined: Ambient Noise (the ubiquitous frequencies available in any natural environment) and Stimulus Noise (excess frequencies within the sound wave). Considering that (1) Signal and

Noise are the only two components of a speech signal and (2) signal intensity and the presence/absence of noise can greatly affect one's ability to perceive any given speech, it makes sense to follow Deonarine's (2010) manipulation of auditory input intensity with an additional manipulation of both forms of Noise. It is hypothesized that reducing noise will increase the frequency of [ba] (auditory input) reports, at the expense of [da] (fused percept) and [ga] (visual input) reports, as the auditory input becomes more salient without additional noise. Assuming that a sufficient amount of noise is removed from the auditory input, [ba] reports should be found even at the quiet volume levels. The first method used to test this hypothesis in the present thesis is reduced Ambient Noise.

Experiment 1

Method

Participants. Twenty-five students, from the University of Waterloo's Department of Psychology subject pool, participated in this study for course credit. They consisted of nine males and sixteen females between the ages of 18 and 24.

All participants reported English as their primary language. However one participant also spoke another language and was removed from the analysis[3]. All others reported English as their sole spoken language. Two participants were removed from the analysis due to technical difficulties during the experiment. Twelve participants were removed from the analysis, as they were unable to correctly identify the visual articulation when asked (see Procedure), confusing [ga] for another CV syllable[4]. These failures to correctly identify the visual articulation are taken up in the General Discussion, as this is not a common issue within the paradigm.

The final sample in the analysis consisted of n = 10.

Apparatus. A McGurk Stimulus was created by the first author using a web camera (producing a recording with an image size of 640 pixels by 480 pixels), with the auditory signal recorded in a room with 36 dBA of ambient noise. The stimulus contained only one presentation (or cycle) of the syllable. The syllable was recorded and edited for duration using Audacity (http://audacity.sourceforge.net/). The movie was compiled using Windows Movie Maker on Windows 7. This video was saved in .wmv format and played during the experiment using a Windows Media Player applet within a video player programmed by the first author in Visual Basic using Microsoft Visual Basic 2010 Express Edition.

A laptop computer (Toshiba Satellite A200 with Windows Vista Home Premium) was used to control the experiment. A monitor (LG Flatron 24" LCD) was used to present the video to the participants. Participants were seated approximately 50 cm away from the screen. The audio signal was presented via two loudspeakers (Adam A5) using the Left/Right stereo option, which were connected to the laptop's headphone output. The speakers were approximately 65 cm apart from each other. Crucially, to reduce ambient noise, all testing occurred in a sound attenuated booth, with a width of 182 cm, a length of 192 cm, and a height of 198 cm. The level of noise in the sound attenuated booth was functionally zero (less than 33 dBA)[5].

To gauge the dBA range for each volume level, a CheckMate SPL Meter (CM-130) was held 110 cm above the floor and 45 cm away from the centre point of the distance between the two speakers (which matched where the average participant's head was positioned). The highest and lowest dBA measurements were taken while the video played.

Trials were randomized without replacement, using a randomization program written in Visual Basic using Microsoft Visual Basic 2010 Express Edition.

Procedure. Prior to greeting the participant, the trial randomization program was run to determine the order in which volume levels would be changed. After being greeted, filling out an Information Sheet (which collected the participants age, gender and spoken languages), and giving consent to participate, participants were taken into a sound attenuated booth. Each stimulus was presented at the default image size, with the Windows Vista System Volume set at 30 out of 100, each speaker's volume set at ¼ of the maximum, and the Media Player's volume set at the level for the first trial. The Media Player's volume was the only volume level manipulated during the experiment. The volume levels were a subset of those used from Deonarine (2010). The volume levels used can be seen in Table 1.

The participant was presented with the McGurk Stimulus and asked to report what they heard. They were presented the McGurk Stimulus once on each trial. Responses were collected verbally and input into a Microsoft Excel spreadsheet. After each response, the experimenter advanced to the next trial (as determined by the randomization program).

After 25 trials had been completed, participants were presented with a single trial in which they were asked to identify the visual articulation. They were only presented the visual articulation (the volume was muted), but were allowed to see the visual articulation as many times as they wished. This did not occur at any other point during the experiment (only after the 25th trial).

Upon completion of all 50 trials (2 trials for each of the 25 volume levels), participants were debriefed, thanked again for their participation and were provided with time to ask any questions they may have had.

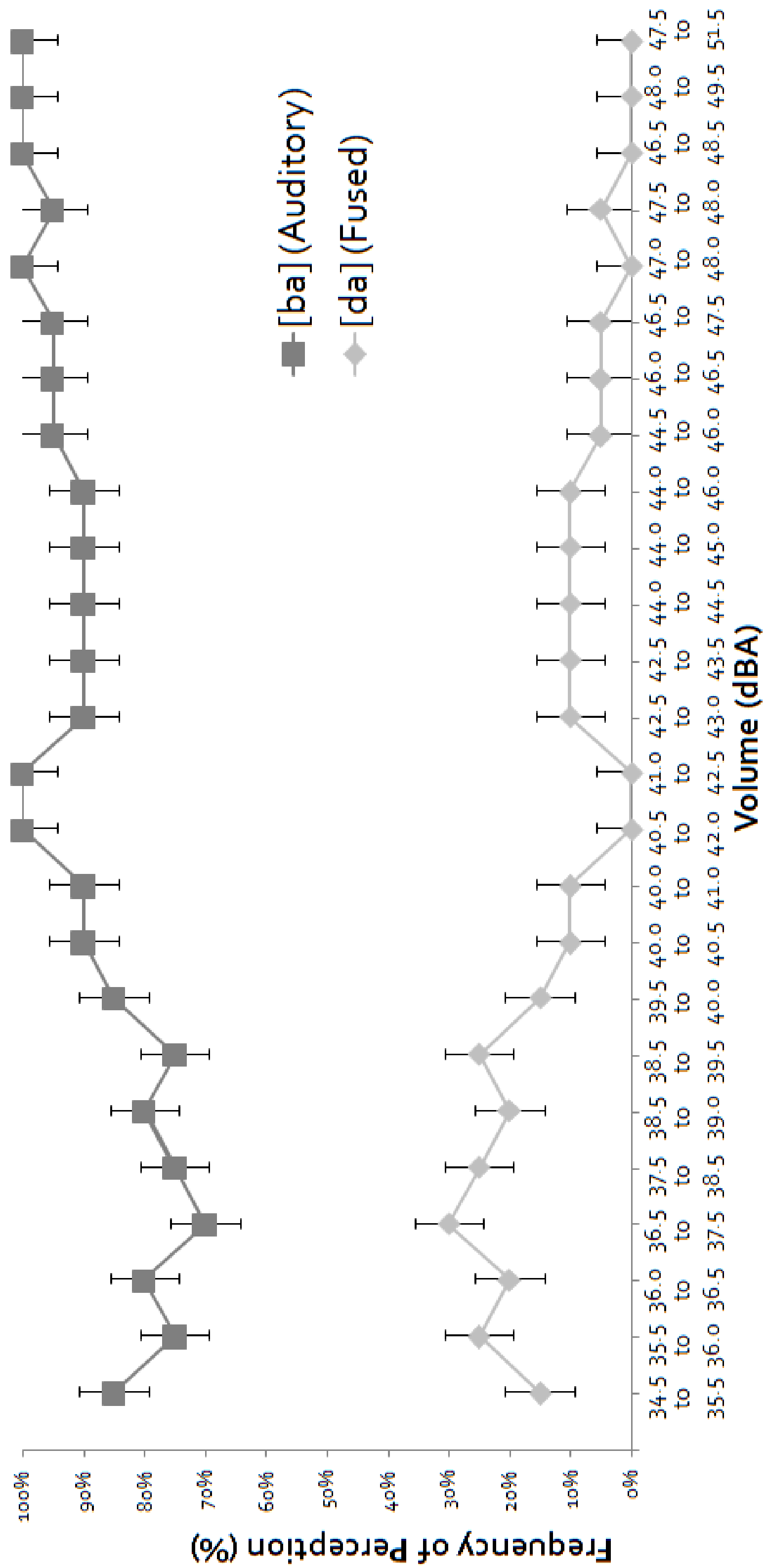
Table 1. Volume levels used in Experiment 1.

Volume Level (dBA)	**Comparison to Normal Conversation Volume**
34.5 - 35.5	1/8 volume level
35.5 - 36.0	
36.0 - 36.5	
36.5 - 37.5	
37.5 - 38.5	
38.5 - 39.0	
38.5 - 39.5	
39.5 - 40.0	1/4 volume level
40.0 - 40.5	
40.0 - 41.0	
40.5 - 42.0	
40.5 - 42.5	
42.5 - 43.0	
42.5 - 43.5	
44.0 - 44.5	
44.0 - 45.0	
44.0 - 46.0	
44.5 - 46.0	
46.0 - 46.5	
46.5 - 47.5	
47.0 - 48.0	
47.5 - 48.0	
46.5 - 48.5	
48.0 - 49.5	1/2 volume level
47.5 - 51.5	

Results

All ten participants reported either [da] or [ba] for each trial. Figure 1 graphically shows the distribution of responses across each volume level. It is seen that participants reported [da] more frequently at the quiet volume trials, and [ba] more frequently as the volume increased. Figure 1 also shows the 95% confidence intervals for the frequency of each response. Masson and Loftus (2003) explain that, in the context of a within-subject design, for each pair of points whose 95% confidence intervals do not overlap, there is a significant difference ($p < 0.05$). [ba] is heard significantly more often than [da] at all volume levels, as the confidence intervals for [ba] replies and [da] replies do not overlap, suggesting that the absence of ambient noise decreases the likelihood that speech perception will be influenced by seen lip movements.

Figure 1. Category judgment as a function of volume level (Error Bars: ± 95% CI)



Discussion

As anticipated, [ba] is reported more frequently than [da] at all volume levels. This suggests that auditory noise is required for speech perception to be influenced by seen lip movements. This is further supported by the fact that [ba] is reported more frequently at each volume level compared to Deonarine (2010), where ambient noise was included. It is worth noting that any further volume reduction would make the stimulus too difficult to hear and participants would not be able to give reliable responses.

It should be noted that [da] still represents a considerable proportion of the replies (10.6%). This suggests that reducing Ambient Noise alone is not sufficient to prevent the McGurk Effect from occurring. As proposed, noise exists in two domains: In the environment (Ambient Noise) and interlaced with the signal (Stimulus Noise). Experiment 2 further reduced noise by reducing Stimulus Noise.

Experiment 2

Method

Participants. 19 students, from the University of Waterloo's Department of Psychology subject pool, participated in this study for course credit. They consisted of seven males and twelve females between the ages of 18 and 24.

All participants reported having English as their sole spoken language. Nine participants were unable to identify the visual articulation correctly (see Experiment 1) and were therefore removed from the analysis. Again, this problem is taken up further in the General Discussion.

The final sample in the analysis consisted of n = 10.

Apparatus. The same equipment as found in Experiment 1 was used, with minor changes. The first is that the auditory signal was recorded in the sound attenuated booth. As described earlier, the level of noise in the sound attenuated booth was functionally zero (less than 33 dBA). The second was that the volume level was decreased while compiling the movie in Windows Movie Maker, in order to maintain a range close to that in Experiment 1.

Procedure. The procedure is the same as in Experiment 1. Table 2 shows the volume levels used in this experiment.

Table 2. Volume levels used in Experiment 2.

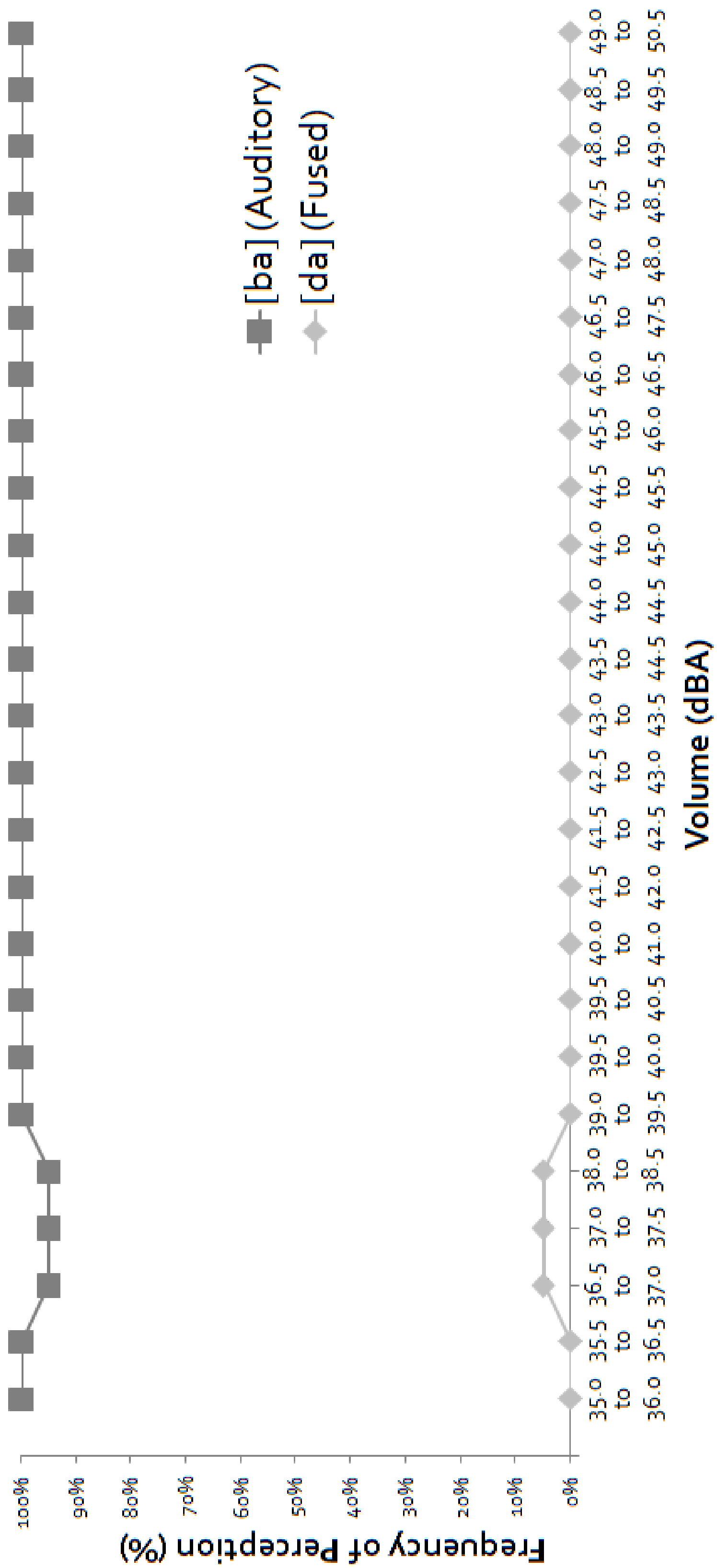
Volume Level (dBA)	Comparison to Normal Conversation Volume
35.0 - 36.0	1/8 volume level
35.5 - 36.5	
36.5 - 37.0	
37.0 - 37.5	
38.0 - 38.5	
39.0 - 39.5	
39.5 - 40.0	
39.5 - 40.5	1/4 volume level
40.0 - 41.0	
41.5 - 42.0	
41.5 - 42.5	
42.5 - 43.0	
43.0 - 43.5	
43.5 - 44.5	
44.0 - 44.5	
44.0 - 45.0	
44.5 - 45.5	
45.5 - 46.0	
46.0 - 46.5	
46.5 - 47.5	
47.0 - 48.0	
47.5 - 48.5	
48.0 - 49.0	
48.5 - 49.5	1/2 volume level
49.0 - 50.5	

Results

Nine of the ten participants heard only [ba] at every volume level presented. One participant heard a combination of [da] and [ba], with [da] only occurring at low volume levels.

Figure 2 shows the distribution of replies across each volume level graphically, along with the 95% confidence intervals for the frequency of each response.

Figure 2. Category judgment as a function of volume level (Error Bars: ± 95% CI)



Discussion

As anticipated, [ba] is reported more frequently than [da] at all volume levels, much like Experiment 1. The major difference is that [ba] was reported for an even greater percentage of trials in Experiment 2 (99.4% of the replies), while [da] was reported by only 1 participant and accounted for 0.6% of the replies. Again, especially when compared to the results of Deonarine (2010), this suggests that noise is required to allow audiovisual integrative speech perception (as evidenced by a "da" response) to occur. When noise is reduced, the auditory input dominates the perceived syllable.

The one participant who heard [da] at the lower volume levels may be understood according to the results of Schwartz (2010). He suggested that audiovisual integration can be subject dependent, with some subjects attending to the auditory input more than the visual input, while others attend to the visual input more than the auditory input. However, Schwartz does not suggest a mechanism which accounts for why participants attend to one source of information over the other. Alternatively, the data of this participant may also be understood in terms of the present noise hypothesis. This participant's Internal Noise (Bennett, Sekuler & Ozin, 1999) may be higher than that of other participants, therefore there was additional noise available (allowing the fused percept to emerge at quiet volume levels). This alternative hypothesis is, of course, completely post hoc.

General Discussion

The data can be summarized as follows. The removal of noise (both Ambient and Stimulus) in large part serves to eliminate the McGurk Effect. This supports the idea that noise is required for audiovisual integration to occur, as some ambiguity is required before the visual information is used. In Experiment 1, a lower volume level provided an opportunity for the auditory input to become ambiguous (compared to the level of Stimulus Noise), while in Experiment 2, the absence of ambient and stimulus noise meant that even the lowest volume level could not provide the same opportunity.

This suggests that visual information is only integrated if sufficient ambiguity is present. If the auditory stimulus isn’t intense enough (volume) or clear enough (noise present in the environment or sound wave), then the visual information affects the construction of the auditory percept. When the visual articulation does not match with the auditory information and the auditory signal is not clear enough to be understood on its own, then either the two sources of information are fused (resulting in [da]) or the visual articulation dominates the auditory input entirely (resulting in [ga]), which matches the concept of “visual dominance” (Rosenblum, 2008).

One possible future study to elaborate on the findings here could compare four different conditions in a Within-Subject design, in order determine the power of the effects created by removing Ambient Noise alone, removing Stimulus Noise alone and removing both Ambient and Stimulus Noise together. This paper addresses the first and third conditions in a Between-Subject analysis. These three conditions would have to be compared to a fourth, baseline condition where both Ambient and Stimulus Noise are present. This baseline condition was addressed by Deonarine (2010), and should be replicated in this new within-subjects experiment. Using this method, it can be determined whether removing both Ambient and Stimulus Noise together creates an additive, underadditive or overadditive effect compared to removing either type of noise separately.

Upon reflection of the methodology used in these experiments, improvements to future studies can be made (such as the proposed study above). The first would be to include multiple visual alone trials (which are not looped, but presented once as the other trials were in the two experiments above), as well as congruent audiovisual presentations, allowing the incongruent audiovisual condition to be compared to other conditions.

It was noted that twelve participants were removed from Experiment 1 and nine participants were removed from Experiment 2, as they were unable to correctly identify which syllable was being visually articulated. Most audiovisual integration research assumes that participants are accurately identifying the visual articulation in audiovisual conditions. However, this was not the case in this particular experiment. One possible cause for this occurrence is that the visual stimulus was unclear. However, while formal testing was not performed, the stimulus was extensively pre-tested by independent observers to ensure accuracy. Examples of misidentified visual articulations from both experiments include: [a], [da], [la] or [ða] (pronounced as "tha"). All of the participants followed a particular pattern in reply: They reported the mistaken visual articulation in the quiet half of volume levels, while [ba] was reported in the loud half of the volumes. Further, those who fail to correctly identify the visual articulation show an effect similar to "Visual Dominance" (Rosenblum, 2008), where the visual articulation overrides the auditory percept completely, and the heard syllable matches the visual articulation (rather than providing a fused percept). Because these participants had no opportunity to provide a fused response (in the absence of a perceived velar visual), they were not included in the analysis.

In conclusion, reducing the noise in the auditory input of a McGurk stimulus limits or prevents the McGurk illusion. Following the theoretical proposal of Deonarine (2010), this result is not surprising, as removing noise removes the ability for the perceptual system to expropriate auditory noise to construct a visually influenced or dominated percept.

References

Anderson, S., Skoe, E., Chandrasekaran, B., & Kraus, N. (2010). Neural Timing Is Linked to Speech Perception in Noise. *The Journal of Neuroscience, 30(14)*, 4922–4926.

Bennett, P. J., Sekuler, A. B., & Ozin, L. (1999). Effects of aging on calculation efficiency and equivalent noise. *Journal of the Optical Society of America, 16(3)*, 654-668.

Burnham, D. & Dodd, B. (2004). Auditory–Visual Speech Integration by Prelinguistic Infants: Perception of an Emergent Consonant in the McGurk Effect. *Developmental Psychobiology, 45(4)*, 204-220.

Deonarine, J.M. (2010). *Increase in volume changes [ga] to [da] to [ba] in the McGurk Effect*. Unpublished undergraduate dissertation, University of Toronto at Scarborough, Scarborough, Ontario, Canada.

Kennedy, J.M., Chattaway, L.D., & Chattaway, E. (1974). Restoration and selection processes in listening to sentences with phonemes deleted. Presented at the meeting of the Canadian Acoustical Association, Edmonton.

Liberman, A.M., Harris, K.S., Hoffman, H.S., & Griffith, B.C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology, 54(5)*, 358-368.

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

Masson, M.E.J., & Loftus, G.R. (2003). Using confidence intervals for graphically based data interpretation. *Canadian Journal of Experimental Psychology, 57(3)*, 203-220.

McGurk, H., & MacDonald, J. (1976). Hearing lips and seeing voices. *Nature, 264*, 746-748.

Rosenblum, L.D. (2008). Speech Perception as a Multimodal Phenomenon. *Current Directions in Psychological Science*, *17*, 405-409.

Rosenblum, L.D. and Saldaña, H.M. (1996). An audiovisual test of kinematic primitives for visual speech perception. *Journal of Experimental Psychology: Human Perception and Performance, 22(2),* 318-331.

Rosenblum, L.D., Schmuckler, M.A., & Johnson, J.A. (1997). The McGurk effect in infants. \ *Perception & Psychophysics, 59(3)*, 347-357.

Rouger, J., Fraysse, B., Deguine, O., & Barone, P. (2008). McGurk effects in cochlear-implanted deaf subjects. *Brain Research, 1188*, 87-99.

Schwartz, J. (2010). A reanalysis of McGurk data suggests that audiovisual fusion in speech perception is subject-dependent. *Journal of the Acoustical Society of America, 127(3),* 1584-1594.

Sekiyama, K. (1997) Cultural and linguistic factors in audiovisual speech processing: The McGurk effect in Chinese subjects. *Perception & Psychophysics, 59(1)*, 73-80.

Sekiyama, K. (2007) Audiovisual speech perception and its inter-language differences. *The Japanese Journal of Psychonomic Science, 15(2)*, 122-127.

Sekiyama, K. & Burnham, D. (2008). Impact of language on development of auditory-visual speech perception. *Developmental Science, 11(2)*, 306-320.

Sekiyama, K., Kanno, I., Miura, S., & Sugita, Y. (2003). Auditory-visual speech perception examined by fMRI and PET. *Neuroscience Research, 47*, 277-287.

Sekiyama, K. & Tohkura, Y. (1991). McGurk effect in non-English listeners: Few visual effects for Japanese subjects hearing Japanese syllables of high auditory intelligibility. *Journal of the Acoustical Society of America, 90(4, Pt 1),* 1797-1805.

Syrdal-Lasky, A. (1978). Effects of intensity on the categorical perception of stop consonants and isolated second formant transitions. *Perception and Psychophysics*, *23(5)*, 420-432.

Teinonen, T., Aslin, R.N., Alku, P., & Csibra, G. (2008). Visual speech contributes to phonetic learning in 6-month-old infants. *Cognition, 108(3)*, 850–855.

Torgesen J.K., Wagner R.K., & Rashotte C.A. (1999) Test of Word Reading Efficiency (TOWRE). Austin, TX: Pro-Ed.

Warren, R.M., & Obusek, C.J. (1971). Speech perception and phonemic restorations. *Perception and Psychophysics, 9(3)*, 358-362.

Williams, L.E., Light, G.A., Braff, D.L., & Ramachandran, V.S. (2010) Reduced multisensory integration in patients with schizophrenia on a target detection task. *Neuropsychologia*, *48(10)*, 3128-3136.

Footnotes

1. While McGurk and MacDonald never theorized on what the common information consists of, similarities between [ga], [da] and [ba] include voicing and manner of articulation (all are voiced, plosive consonants). [da] is considered phonetically "between" [ga] and [ba], having a place of articulation and a F2 Transition between those for [ga] and [ba].

2. It has been established that consonants are particularly noise-vulnerable, as they consist of rapidly moving frequencies, as opposed to the longer, steady-state frequencies found in vowels (see Anderson, Skoe, Chandrasekaran, & Kraus, 2010).

3. It has been established that a participant's primary language can change the syllable heard in a McGurk Effect paradigm (see Sekiyama & Burnham, 2008). Due to this issue, it is not feasible to compare participants with multiple primary languages to participants with English as their sole spoken language.

4. These participants were removed due to the failure to correctly identify the visual articulation. Their pattern of reply suggests that they are subject to an effect which is not the McGurk Effect, but rather Visual Dominance. See the General Discussion for further information.

5. It is noteworthy to mention that not all sound-attenuated booths reduce the noise level to these extremes, hence the use of the term "functionally zero". This sound-attenuated

booth used was constructed with one booth inside another, therefore reducing the noise further than a standard sound-attenuated booth would otherwise.