

PHONOTACTIC CONSTRAINTS IN MCGURK EFFECT

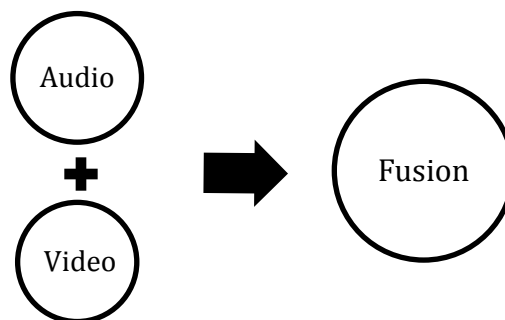
Shubham Atreja, Enayat Ullah
Indian Institute of Technology Kanpur
Mentor- Prof. Amitabha Mukherjee

ABSTRACT

McGurk Effect demonstrates an interaction between audio and visual sensory inputs and how an incongruency between the two can lead to a completely new percept. Since this audio-visual fusion takes place unconsciously, we used it to verify how Phonotactics is unconsciously used during speech processing. An experiment was carried out, where the subjects were presented with meaningless words containing McGurk effect. Phonotactic constraints in English language were used to classify the effect as either legal or illegal depending upon the resultant fusion. Results show that the strength of McGurk effect was significantly influenced by these constraints. While there was a strong effect in words where the fusion was legal, the effect was significantly suppressed in cases where the fusion was illegal. This also proves that McGurk Effect is not completely autonomous and is altered by cognitive intervention

INTRODUCTION

Usually, when we perceive speech, the auditory and visual components of speech are coherent. What happens when there is an incoherence between the information in the visual channel and the audio channel?



McGurk Effect is a perceptual phenomenon that demonstrates the interaction between hearing and vision during speech perception. It was discovered by McGurk and Macdonald in 1976^[1]. The effect occurs when the auditory component of one sound is paired with the visual component of another sound, leading to the perception of a third sound. This effect clearly establishes that speech perception is not at all an auditory phenomenon but an audio-visual phenomenon. The syllables prone to McGurk fusion are governed by certain Phonological rules, like the fusion of a palatal syllable and a dorsal syllable gives a velar syllable.

$$\boxed{/pa/} + \boxed{/ka/} = \boxed{/ta/}$$

However, it is not very clear as to whether the effect is autonomous or can it be influenced by other cognitive processes. A number of attempts have been made to study this. Windmann, in 2004 discovered that sentence context and expectation affected the McGurk illusion in German language^[2]. Azra N. Ali gave similar results for sentences in English^[3]. For our work, we use phonotactic constraints in English language.

Phonotactics is the branch of phonology that deals with restrictions on the permissible combinations of phonemes. The rules of phonotactics vary from language to language. For example: /st/ is valid in English, but is invalid in Hindi and Japanese. Moreover, the validity of phonemic cluster also depends on its location in the word. For example: /sp/ is valid in the onset of a word, but restricted in coda^[5].

Further it has been established that that humans unconsciously use phonotactic constraints during online speech perception and these constraints can significantly affect the final perception^[4]. There we introduce phonotactic constraints in McGurk effect to introduce a bias, based on whether the consonant clusters in the spoken word are Phonotactically constrained or not.

In the present study, an experiment is carried out where an English speaking subject is presented with meaningless words. Phonotactic constraints are used to classify the effect as either legal or illegal depending upon the resultant fusion. If the results show a marked difference in the strength of the effect, it would suggest that fusion is not robust to cognitive intervention. It will further verify the fact that humans unconsciously use phonotactic constraints during speech perception, even though they pertain to meaningless units^[4].

EXPERIMENTATION

Stimuli

Audio-video stimuli consisted of English syllables /pa/ and /ka/ respectively, embedded in arbitrary words.. According to McGurk effect, the auditory /pa/ and visual /ka/ syllables, when dubbed together give rise to /ta/ perception in most of the cases. Similar results are expected when the syllables were embedded in different words. All the words used were meaningless and are listed in the Appendix.

For example- when auditory **Praan** is dubbed with visual **Kraan** the expected perception should be **Traan**.

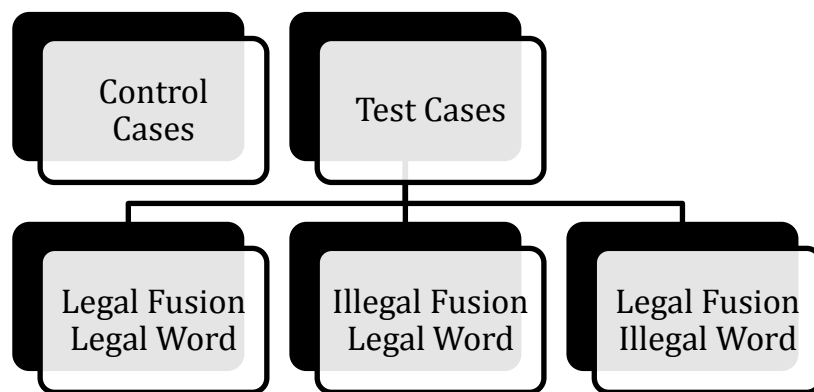
A speaker producing the required audio and video stimuli was filmed using a Panasonic HDC-HS300 Camcorder. The video was filmed in a closed room with minimum noise, against a white background. The entire frame was occupied by the face of the speaker to prevent any kind of distraction. The audio tracks of different words were then dubbed with the required video tracks using Adobe Premiere. These individual words were then joined together to form a video sequence of duration 51 seconds. The frame size was 1902*1080 pixels at a frame rate of 25fps and a bitrate of 32209 kbps. The audio was sampled at 48Khz with a bitrate of 190kbps.

Procedure

The experiment involved 30 English speaking subjects with normal hearing and vision. The subjects were tested individually. They were made to watch the video sequence of words and were instructed to report the word immediately after they heard it, as it is. The instructions were further embedded in the video with a gap of around 3-4 seconds to allow subjects to report the word they just heard. The videos were played on a standard 15.6" Laptop screen (1366*768) with external speakers.

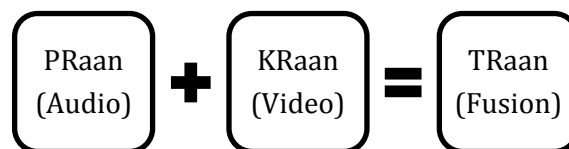
Design

21 words were created at random to introduce the McGurk effect and were put in the video sequence. Out of these 21 words, 6 words were coherent; i.e. the audio and the video stimuli corresponded to the same utterance. These constituted the control cases. These words did not have any McGurk effect. The control cases constituted phonotactically legal as well as restricted words. They were included to ensure that the subject was reporting the accurate perception, even in case of illegal. The remaining 15 words constituted the test cases where the stimuli was incoherent; i.e. different audio and video tracks were dubbed together to introduce the effect. The test cases were further divided into three categories-



- 1) Legal Fusion-Legal Word - The audio track of a phonotactically legal word was dubbed with the video track of another phonotactically legal word. The words were chosen such that the fused word, perceived due to McGurk effect, was also legal.

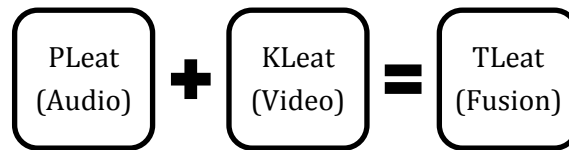
For example-



Here the effect is present in the consonant cluster */pr/*, */kr/* and */tr/* and all three of them are legal.

- II) Illegal Fusion-Legal Word - The audio track of a phonotactically legal word was dubbed with the video track of another phonotactically legal word. In this case, the words were such that the fused word was restricted due to the phonotactic constraints.

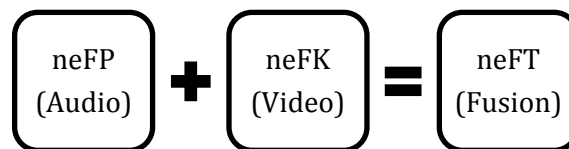
For example-



Here the effect is present in the consonant cluster of /pl/, /kl/ and /tl/. While /pl/ and /kl/ are legal in English, /tl/ is not a legal cluster.

- III) Legal Fusion-Illegal Word- The audio track of a phonotactically restricted word was dubbed with the video track of another phonotactically restricted word. However, the words were such that the fused word was phonotactically legal.

For example-



Here the effect is present in the consonant cluster of /fp/, /fk/ and /ft/. While /fk/ and /fp/ are not legal in English, /ft/ is a legal cluster.

RESULTS

I. CONTROL CASES

All the subjects correctly reported the phonotactically legal words belonging to control cases, thereby certifying that they were reporting the actual perception as it is. However in case of restricted words, some of the subjects did not report the word as it is. Some of them simply dropped the consonant that was restricted.

For ex- In case of the word TLAPE, some subjects reported LAPE, dropping the /t/ since /tl/ is not a legal consonant cluster

This suggests that phonotactic constraints were unconsciously altering the subject's perceptions. To take this into account, the mean probability of perceiving a word was normalised by the probability with which subjects reported the correct word in the control cases.

II. TEST CASES

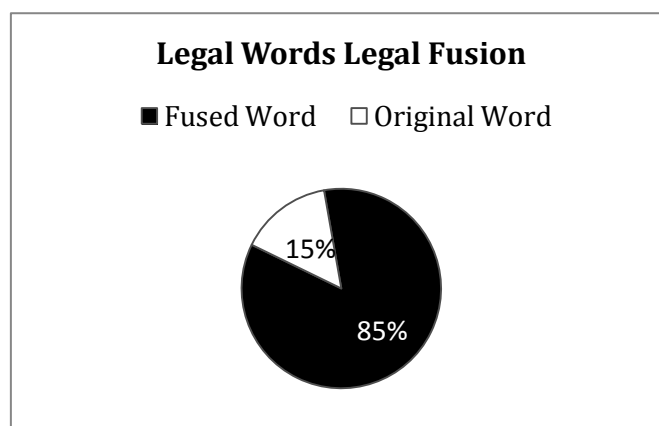
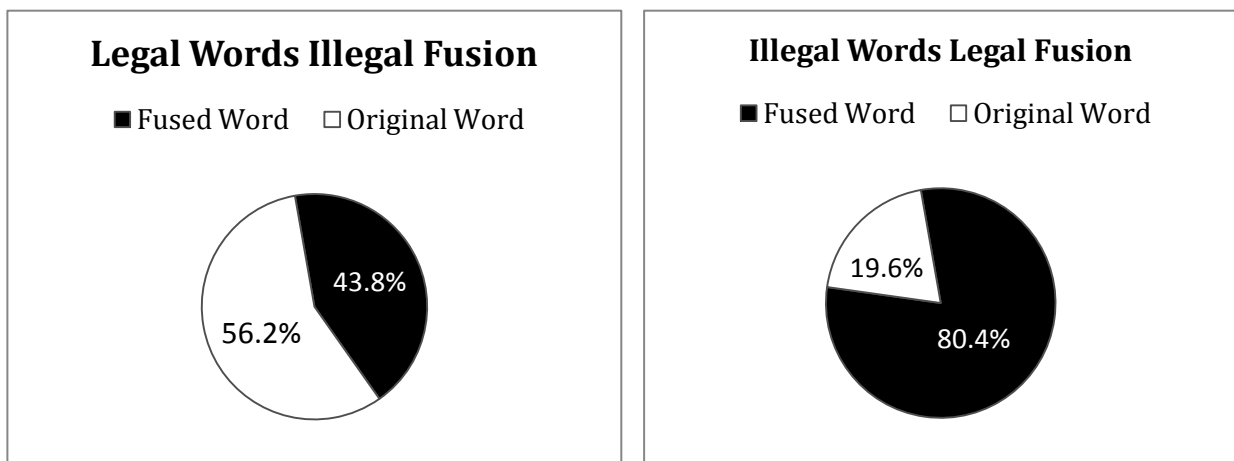
The probability of McGurk effect was determined by counting the relative number of times the subjects gave the expected fusion response. This probability was calculated for the three test cases. Results show that a strong McGurk effect exists in the chosen words, with a probability of **80.4%**.

The table presents the difference between the mean values of the probability of the effect in the three test cases.

Cases	Mean	Standard Deviation
Legal Fusion-Legal Words	0.804	0.09
Legal Fusion-Illegal Words	0.850	0.10
Illegal Fusion-Legal Words	0.428	0.24

The data suggests that there is a slight difference in the strength of the McGurk effect, when the spoken words are illegal compared to the neutral case, where all the words are legal; from a mean probability of **80.4%** it has gone up to **85%**. Although the difference is not huge, we can conclude that a person is less likely to report a phonotactically restricted word than a legal word which has led to the increase in the strength of this effect.

However there is a marked difference in the case when the fused word is illegal. The probability of McGurk effect has reduced to **42.8%** from **80.4%**. On further analyzing, **31%** of the time, the reported word corresponded to the video track, while **16.4%** of the time, the subjects dropped the restricted consonant and another **9.8%** constituted the cases when the reported word corresponded to the audio track. This suggests a significant variation in the results largely depending upon the subject. However, what can be observed is that the McGurk effect has suppressed significantly. We can conclude that the possible reason for this drop is the fact that fused words are phonotactically restricted and a person is more likely report a legal word than an illegal one, which in this case is more likely to be the word in either audio track or video track.



On computing the one way ANOVA of the three test cases, we get $F(2,57) = 42.48$; $p < 0.001$. This again indicates a marked difference between the three test cases. Therefore the results show that Phonotactic constraints can significantly alter the strength of the McGurk Effect; especially suppressing it when the fused word is illegal. This also verifies the fact that phonotactics is actively used by humans during online speech processing.

FUTURE WORK

Since most of the subjects were bilinguals (Hindi as well as English speakers), the knowledge of phonotactic constraints in Hindi can affect their perception. Hence, we would like to carry out this study for constraints in Hindi as well, and then test them on Hindi speaking subjects.

ACKNOWLEDGEMENT

The work has been done as a part of the course SE367A. We would like to thank Prof. Amitabha Mukerjee for his useful insights and continuous support and guidance throughout the project. We would also like to thank Media Center, IIT Kanpur for providing us with the resources for filming the sequence.

REFERENCES

- [1] McGurk, H., & MacDonald, J. (1976); Hearing lips and seeing voices. *Nature*, 264
- [2] Windmann, S., "Effects of sentence context and expectation on the McGurk illusion", *J. Memory and Language*, Vol 50, 2004.
- [3] Ali A. N.. Exploring semantic cueing effects using McGurk fusion, in *Auditory-Visual Speech Processing* (Hilvarenbeek: Kasteel Groenendaal) (2007)
- [4] Keren B. Shatzman & René Kager, "A role for phonotactic constraints in speech perception"; *ICPhS 2007*
- [5] Lecture Notes- CS626: Speech, NLP and the Web- Pushpak Bhattacharyya, IIT Bombay

APPENDIX

		AUDIO	VIDEO	FUSION
Coherent	Legal Word	Trape	Trape	-
Coherent	Legal Word	Puft	Puft	-
Coherent	Illegal Words	Tluf	Tluf	-
Incoherent	Words Illegal Fusion Legal	Nefp	Nefk	Neft
Incoherent	Words Legal Fusion Legal	Proot	Kroot	Troot
Coherent	Illegal Word	Tafp	Tafp	-
Incoherent	Words Legal Fusion Illegal	Plape	Klape	Tlape
Coherent	Illegal Word	Tlore	Tlore	-
Coherent	Legal Word	Plub	Plub	-
Incoherent	Words Illegal Fusion Legal	Mifk	Mifp	Mift
Incoherent	Words Legal Fusion Legal	Praan	Kraan	Traan
Incoherent	Words Legal Fusion Illegal	Pleet	Kleet	Tleet
Coherent	Legal Word	Treab	Treab	-
Coherent	Illegal Word	Tlube	Tlube	-
Incoherent	Words Legal Fusion Legal	Prud	Krud	Trud
Incoherent	Words Legal Fusion Illegal	Plime	Klime	Tlime
Coherent	Illegal Word	Roft	Roft	Roft
Incoherent	Words Illegal Fusion Legal	Gafp	Gafk	Gaft