Interdisciplinary Journal of Linguistics
Volume [17] 2024, Pp. 111-126

ROBUSTNESS OF THE MCGURK EFFECT

Didla Grace Suneetha*

ABSTRACT

The world of audition is intriguing. Although we view the sense of hearing as a routine activity, there are many mysteries involved in the act of hearing which are yet to be explored. Humans perceive speech and language through a complex process that remains only partially understood. It is significant that we comprehend the nuances involved in speech perception as it is very crucial for human existence, survival and social relations. Adding to this complexity is the fact that the sense of hearing is often susceptible to many illusions. Of significance to this paper is the McGurk effect – an auditory-visual illusion that occurs when conflicting visual and auditory speech cues are presented simultaneously, causing the brain to perceive a different sound than what is actually spoken. The discovery of McGurk effect has fundamentally shifted our understanding of speech perception from a purely auditory process to a multimodal one, highlighting the integration of visual and auditory information. This paper gives an overview of the robustness of McGurk effect using a three-fold approach: 1) tracing the beginnings of the McGurk illusion, 2) underscoring the robustness of the McGurk effect, and 3) reviewing its impact and relevance in the domain of speech perception.

Keywords: Speech Perception, Auditory-Visual Illusion, Auditory Phonetics, McGurk effect

1. Introduction

Human audition is a remarkable feat. It is little wonder that human ear is often referred to as the masterpiece of God. The sense of hearing is undeniably crucial and contributes uniquely to our understanding of the world and the ability to interact with it. Its impact on daily life is profound as the acquisition and understanding of speech, the detection of potential danger and our very own existence and survival depend upon the auditory modality. The auditory system consists of peripheral and central components. While the peripheral auditory system (consisting of the outer, the middle, and the inner ear) is responsible for capturing and analysing sound signals and converting

* The English and Foreign Languages University, Hyderabad

them into information that can be experienced as an auditory percept (Brister et. al., 2020), the central auditory system (consisting of a large number of subcortical nuclei) is in control of extraction and analysis of the dimensions of the sound. Not only do the nuclei encode dimensions defining the nature of the sound, but they also extract features of sound location (Masterton, 1992). Perceptual systems have universal function (Fowler, 1996). The auditory systems across different biological species indeed exhibit fascinating diversity, each adapted to their specific ecological niche and survival needs. In many lower animals, for instance, hearing plays a crucial role in the localisation of sound. This ability is particularly critical for survival in environments where predators can approach from any direction. In humans, however, hearing plays a vital role in the perception of speech and language.

The human ear is a remarkably sensitive and complex organ and has an amazing ability to detect a broad range of sound frequencies, intensities and locations, process speech and complex auditory patterns, and adapt to varying environmental conditions. In addition, human ear has amazing powers of discrimination too (Denes and Pinson, 2007). For instance, research indicates that humans have an impressive ability to discern different pitches and can discriminate as many as 3500 different frequency steps (Raphael et.al., 2011). Also, listeners are quite sensitive to changes in formants and can reliably perceive shifts of as little as 4-8% in formant frequencies (Kreiman and Sidtis, 2011). The ear is also highly sensitive to changes in intensities. At any given frequency, the human ear can discriminate 280 different intensities. Additionally, through binaural hearing, our perceptual mechanism is equipped to separate interesting sounds from a background of irrelevant sounds and helps us to perceive speech even in noisy environments – a phenomenon known as the Cocktail party effect (Denes and Pinson, 2007).

The anatomy of hearing is so intricately designed and the physiology of hearing so complex that it is difficult to fathom the finer nuances in the audition process as so much remains unexplained. Although the anatomy of hearing is researched extensively, however, for hearing, the manner in which neurophysiology translates into functional capacities for sound processing remains poorly understood (Kreiman and Sidtis, 2011). For instance, the seemingly simple problem of explaining how a sensation of pitch is produced in the hearing mechanism remains baffling to this day (Denes and Pinson, 2007). Likewise, the perception of intonation requires the ability to track pitch changes. Although this ability is one of the first skills that infants acquire, however, the mechanisms behind how infants perceive pitch remains elusive (Borden and Harris, 2011.

Undoubtedly, hearing is complicated, and the functions of the ear and its connections to the brain are far from understood and … the task of unravelling and interpreting connectivity in audition remains extremely daunting (Kreiman and Sidtis, 2011). As daunting as it is, a comprehensive study of speech perception is essential as it serves

various crucial purposes. Given its significance in our everyday lives, speech perception has attracted the interest of researchers from various domains such as linguistics, psychology, engineering, and speech pathology (Borden and Harris, 2011). It is also intriguing to note that speech perception research is currently branching out in many directions with diverse applications such as clinical (speech therapy), forensic (speaker profiling), AI (machine learning) and so forth.

1.2 Speech Perception

The human ear, undeniably, has been a valuable instrument in the study of the perception of speech. One of the earliest works in speech perception can be traced back to the research carried out by the Psychophysicist, Helmholtz. He used his ears extensively in the study of the acoustics of the human voice and explored how the ear detects and processes sound waves, contributing significantly to our understanding of auditory perception. Consequently, it was the invention of sound spectrograph by Ralph Potter and his colleagues that was instrumental in a drastic increase in our understanding of acoustic of speech. However, speech perception questions remained: what aspects of the complex sound pattern of speech are important in listening to speech, and what aspects are less important? With the invention of 'Pattern playback' - an innovative machine conceived by Potter and further developed by Cooper - the visual patterns created by a sound spectrograph could be reversed and converted back into sound. The use of pattern playback significantly advanced our understanding of how humans process and interpret spoken language (Raphael et. al., 2011). Consequently, Liberman and Cooper extensively used the playback to systematically vary the acoustic parameters of speech to determine the cues used in speech perception (Diehl et. al., 2004).

Based on the initial research, speech perception is usually thought to be dependent solely upon the auditory modality (Wood, 1975 in Dodd, 1977). However, a great deal of further research has demonstrated the interplay between auditory and visual information in speech perception. Most of the early research on the audio-visual (AV) speech perception generally falls into two main categories: 1) vision as an *alternative* mode to hearing in speech perception, and 2) vision as a *complementary* mode to hearing speech perception Dodd (1977).

Research on lip-reading as an alternative to auditory information in speech perception has shown that individuals can understand spoken language using only visual information, such as observing the movements of a speaker's lips, even when auditory cues are absent. For instance, an investigation by Conrad (1977) compared the speech-reading proficiency of hearing and deaf subjects and concluded that the hearing subjects, even without any formal training, performed as well as the deaf subjects, who had several years of training in lip-reading. On similar lines, Simmons (1959) found no significant correlations between measures of hearing impairment and speech-reading ability. Yet another study analysed the perceptual confusions made by hearing subjects on a CV speech-reading task and found that they could identify

five places of articulation nearly perfectly (Binnie et. al., 1974 in Dodd, 1977). Conversely, studies that investigated the role of vision as a complementary source of information to speech perception focused on how visual information can enhance auditory perception, especially in challenging listening environments, such as when there is background noise. One of the earliest investigations (Sumby and Pollack,1954) examined the accuracy of recognizing spoken words presented in a noisy context. They reported that the integration of auditory and visual inputs helps improve our ability to understand speech, especially in noisy environments. Similarly, a study by Dodd (1977) investigated how visual cues can enhance auditory perception, particularly in noisy contexts and concluded that "vision provided a complementary source of information for the perception of speech." Studies thus far have indicated that speech perception in face-to-face communication is primarily unimodal for individuals with normal hearing and that visual information plays an independent and additive role, enhancing the auditory signal without altering it.However, the seminal study by McGurk and MacDonald (1976), unexpectedly revealed the McGurk effect**.** This phenomenon has underscored the significance of the interplay between auditory and visual modes, illustrating how conflicting auditory and visual inputs can combine to create a new integrated perception that reflects the combined influence of both modalities. This effect has undoubtedly elevated the role of vision from being a mere 'additive' to becoming an indispensable component of speech perception. The McGurk effect thus added a new impetus to studies of audio-visual speech (Moore et. al., 2009). In the light of this, this paper aims to provide an overview of the robustness of the McGurk effect and its implications in the emerging multidisciplinary fields.

2. The McGurk Effect – The Beginnings

An interesting paper by John MacDonald (2018) in memory of Harry McGurk, his doctoral supervisor, provides a comprehensive look into the origins, development, and significant impact of the McGurk Effect over the past forty years. The McGurk Effect, discovered in 1976 by Harry McGurk and John MacDonald, is a perceptual phenomenon where what we hear is influenced by what we see. Specifically, when visual information from lip movements contradicts auditory information, our brain integrates these cues in a way that alters our perception of the spoken sound. The accidental discovery of this phenomenon, the reaction of the then scientific community as recounted by John MacDonald himself and, the impact of the illusion on various domains, have been presented below.

What started as a pilot study to probe "if infants had a coordinated audio-visual space while perceiving language," accidentally led to the discovery of this interesting phenomenon called the McGurk effect. The pilot study was designed to see if infants reacted differently to the 'congruent' stimuli (where face and voice would present the same stimuli) from the 'incongruent' stimuli' (where the modalities were presented differently) with the assumption that if the infants had a

coordinated audio-visual space, then the 'incongruent' stimuli would disrupt their visual attention. The methodology involved the use of dynamic visual stimuli of various consonant-vowel (CV) combinations, which allowed to control both onset and duration in each modality. A young lady was recruited as a subject and she was filmed with the following sets of stimulus combinations: 1) 'congruent' samples where the audio sample was the same as the visual utterance (eg. audio /baba/ dubbed onto a visual /baba/ and 2) 'incongruent' samples where there was a mismatch in the audio and visual presentation (eg., /baba/ dubbed onto visual /gaga/). When eventually the tapes were ready and on viewing them, MacDonald's initial reaction was there was some error carried in the recording and dubbing of the 'incongruent' tapes while the 'congruent' ones were fine. Disappointed as he was, he informed his supervisor McGurk about the anomaly. On listening to the tapes, McGurk also experienced a similar anomaly. However, it was McGurk who "took the crucial step of listening to the tapes with his eyes closed and then open." He was quick to realise that there was nothing wrong with the tapes, but the incongruent audio-visual stimulus enabled them to experience an auditory illusion, where the audio /baba/ when dubbed onto video /gaga/ was heard as /dada/, which never existed. Curious enough, they tried this phenomenon on several listeners and an overwhelming majority of the people reported this auditory illusion. On thorough research, they realised that this phenomenon was never reported in the then existing literature. Having run the experiments and confirmed the 'illusory' response was robust, the paper titled, "Hearing lips and seeing voices," describing this 'phenomenon' was published in 1976 in the most prestigious journal – *Nature*.

This publication was meted with severe criticism: that 'the findings were simply not true and that we perfectly hear well without vision,' that 'the research is showing the obvious and therefore redundant,' and to top it all, that 'the Universities are wasting the taxpayers' money funding it.' However, there was also interest from unexpected directions. This phenomenon was not just discussed in 'Sunday Times,' a popular newspaper, but the authors were also invited to speak about the illusion at several conferences and seminars. This auditory illusion over the years came to be popularly known as the 'McGurk effect,' after one of its authors. Today, with a citation score of more than 4800, the McGurk effect has caught the attention and interest of varied areas.

Overall, MacDonald's paper serves as a tribute to Harry McGurk's legacy while providing a comprehensive overview of how their discovery of the McGurk Effect has shaped and continues to influence our understanding of perception, cognition, and the integration of sensory information. It underscores the interdisciplinary nature of research on speech perception and its implications for various fields of study.

2.2 Robustness of the McGurk Effect

The McGurk effect has undoubtedly percolated and occupied the centerstage in the realm of audio-visual speech perception. This

phenomenon demonstrates that the visual cues by the speaker "is not simply an adjunct of auditory speech but is an intrinsic component of normal speech perception," (MacDonald, 2018). In addition, the effect is robust, in the sense that the listeners continued to experience the illusion even when they were made aware of it (McGurk and MacDonald, 1976). It is no surprise that the McGurk effect has an impressive record of about 4800 citations over the past 40 years, demonstrating its robustness. The McGurk illusion has been studied extensively across various demographics. Research has shown that this illusion affects a wide range of populations, methodologies and languages, although the degree of susceptibility can vary. Research across different age groups has provided insights into how this illusion influences speech perception throughout development. For instance, studies have shown that infants are remarkably perceptive and that even within weeks of birth, infants can recognise the congruence between lip movements and speech sounds and are affected by incongruent auditory and visual information (Burnham and Dodd, 2004). A series of McGurk experiments were conducted on pre-linguistic infants which tested for the McGurk effect in 5-month-old English-exposed infants. Infants were first gaze-habituated to audio-visual stimuli and then were presented two different dishabituation stimuli. These findings suggest that infants can also integrate audio-visual speech in a McGurk-type manner (Rosenblum et.al., 1997). The McGurk effect was also tested on young children and interestingly, evidence suggests that the effect is somewhat weaker in children compared to the adults. McGurk and MacDonald (1976) reported that two groups of English-speaking children (3-5 years and 7-8 years old) displayed a smaller overall visual influence than adult subjects (18-40 years) and the variation was even more pronounced when the range of consonant combinations was expanded (MacDonald and McGurk, 1978). However, for some audio-visual tokens, the younger children displayed as many fused responses as adults. Similarly, Massaro and his colleagues (Massaro et. al., 1986) have shown that preschool children show less of a visual influence of /ba/ on /da/ (and /da/ on /ba/) than adults. However, these researchers also claim that the reduced effects are due to developmental variations in sensitivity to visual information rather than differences in integration processes. Investigations were also carried out to see how aging affects auditory-visual perception of speech. A comparative study of young (19-21 years) and aged (60-65 years) people was designed to see how aging affects the audio-visual speech comprehension. These experiments examined accuracy and response time in syllable identification for auditory-visual congruent and incongruent stimuli in addition to auditory-only and visual-only presentation modes. Results indicate that older adults were more strongly influenced by visual speech than the younger ones for acoustically identical signal-to-noise ratios of auditory speech (Sekiyama et.al., 2014). Research has explored whether the McGurk effect is language-specific, aiming to understand how different linguistic backgrounds influence this phenomenon. While most research has focused on English speakers, a few studies have explored this phenomenon in other languages. These

studies suggest that the McGurk effect is indeed not language-specific, as it has been observed in speakers of diverse native languages. For instance, a study comparing Mandarin speakers and English speakers from USA revealed that the frequencies of the McGurk effect were similar across both groups, with Chinese participants experiencing the effect at a rate of 48% and American participants at 44% (Magnotti et. al., 2015). In yet another study, it has been reported that Japanese speakers also exhibited McGurk effect. However, Japanese speakers showed a smaller McGurk effect, and so a stronger auditory independence, compared with English native speakers (Sekiyama and Tohkura, 1991). This Japanese-English difference can be attributed to the cultural and linguistic differences.

The McGurk illusion is said to be robust to the range of methodologies that have been used--- natural and artificial voices and faces, stimulus tokens (CV, CVC etc.), language groups, non-tonal and tonal language speakers (MacDonald, 2018). Interestingly, studies have shown that participants can experience the McGurk effect even when presented with images which involve no identifiable facial features but are comprised only of a few illuminated dots on a darkened face (Rosenblum and Saldaña, 1996), which suggests that our brain is capable of extracting meaningful information from very limited visual cues. Research also indicates that the illusion is also experienced when auditory information lags the visual signal by as much as 300ms (Munhall et. al., 1996).

Finally, the McGurk effect can also occur when the auditory and visual signals are generated by speakers of different genders. For instance, in an experiment by Green and his colleagues (1991), a male talker's voice was dubbed onto a videotape containing a female talker's face (and vice versa) to examine whether a discrepancy in the gender of the talker between the auditory and visual signals would influence the magnitude of the McGurk effect. Surprisingly, results indicate that McGurk illusion is perceived even when different genders provide the visual and auditory input, although it was perceptually apparent. Although there were individual differences reported in the way the subjects were susceptible to the illusion, studies which tested the reliability of the illusion indicate that the intra-speaker reliability is quite reliable (strand et. al., 2014), and that despite the substantial variability across individuals and stimuli, there was little change in the frequency of the McGurk effect between initial testing and a one-year retest (Mallick et.al., 2015). Given its robustness, the McGurk effect is seen as being both significant and mandatory (Rosenblum and Saldana, 1996) and provides strong evidence that hearing people use visual speech perception cues if they are available (Woodhouse et. al., 2009).

2.3 Impact of the McGurk Effect

The McGurk effect has indeed had a profound impact on our understanding of how we perceive speech, and its robust presence has remarkably shifted the focus of audio-visual speech perception from “being the issue of compensatory information in the deaf and hard of

hearing to the realization that speech is a multimodal phenomenon and later as a key demonstration of multisensory processing and integration" (MacDonald, 2018). Consequent to the discovery of the McGurk Effect, innumerable studies on the robustness of the McGurk effect have been carried out cutting across different populations, age groups, gender and languages, incorporating a range of methodologies. The findings are suggestive that these "auditory illusions are not mere transitory phenomena elicited by one or two peculiar auditory-visual combinations. Rather, they appear to be illustrative of a general effect of vision upon speech perception in face-face situations" (MacDonald and McGurk, 1978). More importantly, people continue to experience the illusion even after they are made aware of it. The McGurk effect provides compelling evidence of the multimodal nature of speech perception thus shifting the entire gamut of research in exploring three prominent questions: where, why and how does the McGurk illusion happen?

a) Where in the brain does the McGurk illusion happen?

The integration of auditory and visual cues, as demonstrated in the McGurk effect, is a powerful example of multimodal sensory processing which plays a significant role in the perception of spoken language. Consequent to the discovery of the McGurk effect, several studies have attempted to answer as to where in the brain does the illusion happen. The neural basis for multimodal sensory integration is gaining attention which is indeed vital for understanding and interacting with the world around us. Studies on differential hemispheric contributions comparing normal and clinical subjects have shown that both the hemispheres make significant and distinct contributions to the McGurk effect (Baynes et. al., 1994). Research indicates that the McGurk effect involves activity in the frontal, parietal, and temporal regions of the brain, particularly the posterior superior temporal gyrus, which plays a key role in integrating audio and visual information and this illusion occurs when there is a mismatch between the auditory and visual inputs (Kiel et. al., 2012). However, the precise functions of this region and other brain regions in "correcting" mismatched audio-visual sensory inputs are still not fully understood. To explore this aspect, Erickson et. al. (2014) carried out a functional magnetic resonance imaging (fMRI) study, which aimed to isolate brain areas specifically involved in processing congruent audio-visual (AV) speech and the McGurk effect. Their findings showed distinct brain regions involved in the processing of congruent AV speech and incongruent AV speech when compared to unimodal speech (acoustic-only and visual only).Yet another fascinating study using functional magnetic resonance imaging (FMRI) (Calvert et.al., 1997) has shown that visual linguistic cues are sufficient to activate the auditory cortex in individuals with normal hearing, even in the absence of actual speech sounds. This aspect underscores the brain's ability to process language through multiple sensory channels. What is even more intriguing is that these auditory cortical regions are not activated when a person observes non-linguistic facial movements, but respond

promptly to silent, meaningless speech-like movements. This indicates that the brain differentiates between meaningful linguistic cues and other types of visual input, emphasizing the unique role of language-related gestures in activating auditory processing areas.

b) Why does the McGurk illusion happen?

Several theories have attempted to answer as to how and why the McGurk effect happens. Around the same period that the McGurk effect was discovered, cutting (1976) reported a corresponding phenomenon involving the fusion of conflicting auditory stimuli. In his dichotic listening experiment, participants were presented with the sound /ba/ in one ear and /ga/ in the other. Many subjects perceived a centrally located /da/, leading Cutting to propose the *acoustic averaging hypothesis* to explain this integrated perception. However, this proposition could not account for other fused responses (McGurk and Mac Donald, 1976). From the findings by Binnie et.al. (1974, in Dodd, 1977), it may be understood that the feature, place of articulation, is more efficiently detected by vision than is the manner of articulation; while voicing and nasality of consonantal utterances are readily perceived auditorily (Miller and Nicely, 1955 in MacDonald and McGurk, 1978). There were also other hypotheses where proponents have argued the dominance of one modality over the other, which did not stand the test of time either. Given these anomalies, a *manner-place hypothesis* has been proposed which says that "in face-to-face communication between normally hearing people, manner of articulation of consonantal utterances is detected by ear and place of articulation, on the other hand, is detected by the eye" (MacDonald and McGurk, 1978). Moreover, the results of their study also confirm the predictive validity of the manner-place hypothesis with respect to the nature of the illusions.

The three dominant theories to speech perception: 1) auditory theories, 2) motor theories and 3) theories based on psychology of perception, tried to explain the McGurk effect. The auditory theories (Blumstein, 1986; Diehl and Kluender, 1989) claim that speech is predominantly auditory in nature. Using the same hearing mechanism and perceptual processing used for any other types of sound, the listener simply identifies acoustic pattern or features and matches them directly to the learned and stored acoustic-phonetic features of the language (Raphael et. al., 2011). In other words, the objects of speech perception are found at the acoustic surface thus relating it to the physics of the speech waves (MacDonald, 2018). The Motor theories (Liberman and Mattingly, 1985; Mattingly and Studdert-Kennedy, 1991), on the other hand, claim that there is a biological link between perception and production. This specialization prevents listeners from hearing the signal as an ordinary sound, but enables them to use the systematic, yet special, relation between signal and gesture to perceive the speech (Liberman and Mattingly, 1985). Since McGurk illusion is an outcome of the visual influence on facial movements on speech perception, the motor theorists have demonstrated this phenomenon to

be a strong support of the motor theorists. The third perspective that has been applied to audiovisual speech perception is that of 'Direct perception' (Fowler, 1996). Here, it is proposed that the function of our sensory systems is to perceive the causes of the sensory input we receive. For example, in the case of speech, the cause of the sensory stimulus is the vocal tract activity of the speaker. This theory also has no problem with audio-visual speech perception. (MacDonald, 2018). More recent research by Corballis (2002) provides an alternative view which suggests that speech and language evolved from gestures (both manual and facial) and not from sounds. With the passage of time, the gestures moved from manual (hands) to face to allow the hands to be used for other purposes. Consequently, these visual gestures were coupled with sounds which led to the complex system of language. Sound, undoubtedly dominated as it has its distinctive advantages over gestures- sound can be used for distant communication and in the dark. Consequently, our sensitivity to visual speech is a residue of the evolutionary process and this strongly lends support to the McGurk effect. This theory convincingly provides an explanation of why we process and are affected by visible speech information and why we experience this illusion (MacDonald, 2018).

c) How does the McGurk illusion happen?

While the role of visual component in hearing is established, another pertinent question in the understanding of speech perception is to probe how the listeners combine the auditory and visual input that is simultaneously available during communication. If there is only one mechanism for audio-visual integration, then any task that involves these integrative processes, like the McGurk effect, would be directly related to other audio-visual tasks that are routinely carried out. Conversely, if multiple mechanisms exist for combining information across modalities, it is essential to determine whether two tasks rely on the same integrative mechanism or not. One approach to categorizing different integrative mechanisms is based on whether the integration happens at an early stage or late stage (Peelle and Sommers, 2015). In other words, it is crucial to understand if the fusion of the auditory and visual signals happens at an early stage of receiving them or at a later stage.

Many traditional frameworks for audio-visual processing have emphasized on late integration (Grant et. al., 1998). They claim that auditory and visual information are processed independently (unimodal) and combined late in a region of the brain that has anatomical and functional connections to the relevant sensory cortices. In other words, in the late fusion, two separate classifiers make decisions, respectively, on the Audio and Video channels, and the decisions are then fused (Bernstein et. al., 2002). However, this perspective has been challenged by early integration models of speech perception, which argue that visual influences on auditory processing do not merely result from feedback from later multisensory regions but actively affect processing in real time. (Atligan et al., 2018).

3. Limitations of the McGurk Effect

Undoubtedly, the McGurk effect is a fascinating phenomenon which has made significant contributions to our understanding of speech perception, particularly in illustrating the interplay between auditory and visual information. However, The McGurk effect does have its limitations and constraints in its application. For instance, it tends to be most pronounced in controlled experimental settings using unnatural stimuli like that of syllables. However, a more productive line of research to understand the everyday speech perception requires that we focus on audiovisual speech comprehension in situations that are closer to real life communication, which involve sentences and conversations. If these results are consistent with those provided by McGurk stimuli, we can be more certain that both are indexing the same phenomena. (Van Engen et.al., 2022). Research also indicates that not all individuals experience the McGurk effect to the same extent. Factors such as gender (Irwin et. al., 2006), age (McGurk and MacDonald, 1976) and familiarity with the speaker (Walker et. al., 1995) can affect the strength and presence of the McGurk effect. Research also indicates cross-linguistic variability in the way the illusion is perceived and may not be equally robust across all languages. Various clinical conditions such as Autism (Mongilo et. al., 2008) and language learning disabilities (Norrix et. al., 2007) can have a bearing on the McGurk effect. In addition, individuals with enhanced sensory integration have been shown to be more susceptible to the effect (Calvert et. al., 2004).

As Van Engen and his colleagues (2022) rightly put, “although it is clear that sighted listeners use both auditory and visual cues during speech perception, the manner in which multisensory information is combined is a matter of debate.” While the McGurk effect is a powerful demonstration of multisensory integration, it represents just one aspect of the complex process of speech perception. It may not fully capture all the nuances of how people perceive and interpret speech in naturalistic settings, where other sensory modalities and cognitive processes also play a role. Despite these limitations, the McGurk effect remains a valuable tool in cognitive science and psychology for studying multisensory integration and the complex interactions between auditory and visual information in speech perception.

4. Conclusion

Undoubtedly, the McGurk effect is a fascinating phenomenon which has made significant contributions to our understanding of speech perception, particularly in illustrating the interplay between auditory and visual information. It is the McGurk effect that has established that hearing people invariably use visual information when available, which is a clear demonstration of how our senses work together to shape our perception of the world. The McGurk effect has indeed transcended many boundaries and has significantly impacted

many aspects of speech communication. Given its wide range of applications, one can only anticipate that future research will further explore the significance of McGurk effect which will not only enhance our understanding of speech perception but also have implications for various fields, including communication technology, language learning, rehabilitation for those with hearing impairments and so forth.

References

Atligan, H., Town, S. M., Wood, K. C., Jones, G. P., Maddox, R. K., Lee, A. K. C., & Bizley, J. K. (2018). Integration of Visual Information in Auditory Cortex Promotes Auditory Scene Analysis Through Multisensory Binding. *Neuron*, 97(3):640-655.

Baynes, K., Fummell, M. & Fowler, C. (1994). Hemispheric Contributions to the Integration of Visual and Auditory Information in Speech Perception. *Perception and Psychophysics*, **55** (6): 633–641. doi:10.3758/bf03211678. PMID 8058451.

Bernstein L. E., Burnham D. & Schwartz J. L. (2002). Special Session: Issues in Audiovisual Spoken Language Processing (When, Where, and How?), in: *Proceedings of the 7th International Conference on Spoken Language*, Denver, CO, USA, pp. 1445–1448.

Binnie, C. A., Montgomery, A. A. & Jackson, P. L. (1974). Auditory and Visual Contributions to the Perception of Consonants. *Journal of Speech and Hearing Research,* 17:619-630 as cited in Dodd, 1977.

Blumstein, S. E. (1986). *On Acoustic Invariance in Speech, in: Invariance and Variability in Speech Processe.* J. S. Perkell and D. H. Klatt (Eds), pp. 178–197. Erlbaum, Hillsdale, NJ, USA.

Brister, E., Agarwal, A. & Richter, C. (2020). *The Senses: A Comprehensive Reference* 2nd ed., 2:18-31, Elsevier. https://doi.org/10.1016/B978-0-12-809324-5.24241-7.

Burnham, D. & Dodd, B. (2004). Auditory-visual Speech Integration by Prelinguistic Infants: Perception of an Emergent Consonant in the McGurk Effect. *Developmental Psychobiology*, **45**:204-22.

Calvert, G. A., Bullmore, E. T., Brammer, M. J., Campbell, R., Williams, S. C., McGuire, P. K., Woodruff, P. W., Iversen, S. D. & David, A. S. (1997). Activation of Auditory Cortex During Silent Lipreading. *Science*, 276(5312):593-6. doi: 10.1126/science. 276.5312.593. PMID: 9110978.

Calvert, G. A., Spence, C. & Stein, B. E. (2004). *The Handbook of Multisensory Processes*. Cambridge, Mass.: MIT Press. ISBN 978-0-262-03321-3. OCLC 54677752.

Conrad, R. (1977). Lip-reading by Deaf and Hearing Children. *British Journal of Educational Psychology,* 47(1): 60–65. *https://doi.org/10.1111/j.2044-8279.1977.tb03001.x*

Corballis, M. C. (2002). *From Hand to Mouth: The Origins of Language*. Princeton University Press, Princeton, MA, USA.

Cutting, J. E. (1976). Auditory and Linguistic Processes in Speech Perception: Inferences from Six Fusions in Dichotic Listening. *Psychological Review,* 83:114-140.

Denes, P. B. & Pinson, E. N. (2007). *The Speech Chain,* 2nd ed. New York: W. H. Freeman and Company.

Diehl, R. L. & Kluender, K. R. (1989). On objects of Speech Perception. *Ecol. Psychol.,* 1:121-144.

Diehl, R. L., Lotto, A. J. & Holt, L. L. (2004). *Speech Perception. Annu Rev Psychol.,* 55:149-79. *doi: 10.1146/annurev.psy ch.55.090902. 142028. PMID: 14744213.*

Dodd, B. (1977). The Role of Vision in the Perception of Speech. *Perception,* 6:31–40. doi: 10.1068/p060031. PMID: 840618.

Erickson, L. C., Zielinski, B. A., Zielinski, J. E., Liu, G., Turkeltaub, P. E., Leaver, A. M. & Rauschecker, J. P. (2014). Distinct Cortical Locations for Integration of Audiovisual Speech and the McGurk effect. *Front Psychol.,* 2(5):534.

Fowler, C. A. (1996). Listeners do Hear Sounds, not Tongues, *J. Acoust. Soc. Am*. 99: 1730–1741.

Grant, K. W., Walden, B. E. & Seitz, P. F. (1998). Auditory-visual Speech Recognition by Hearing-Impaired Subjects: Consonant Recognition, Sentence Recognition, and Auditory-Visual Integration. *Journal of Acoustical Society of America*, 103: 2677-2690.

Green, K., Kuhl, P., Meltzoff, A. & Stevens, E. (1991). Integrating Speech Information Across Talkers, Gender, and Sensory Modalities: Female Faces and Male Voices in the McGurk Effect. *Perception and Psychophysics,* 50: 524.

Irwin, J. R., Whalen, D. H., Fowler, C. A. (2006). A Sex Difference in Visual Influence on Heard Speech. *Perception and Psychophysics*, 68 (4):582–592. doi:10.37 58/bf03208760. PMID 16933423.

Keil, J., Müller, N., Ihssen, N. & Weisz, N. (2012). On the Variability of the McGurk Effect: Audiovisual Integration depends on Prestimulus Brain States. *Cereb Cortex*. 22(1):221-31. doi: 10.1093/cercor/bhr125. Epub 2011 May 30. PMID: 21625011.

Krieman, J. & Sidtis, D. (2011). *Foundations of Voice Studies: An Interdisciplinary Approach to Voice Production and Perception.* Wiley-Blackwell.

Liberman, A. M. & Mattingly, I. G. (1985). The Motor Theory of Speech Perception. *Cognition,* 21:1–33.

MacDonald, J. (2018). Hearing Lips and Seeing Voices: The Origins and Development of the 'MCGURK Effect' And Reflections on Audio–Visual Speech Perception over the Last 40 Years. *Multisensory Research, 31*(1-2):7-18. https://doi.org/10.1163/22134808-00002548

MacDonald, J. & McGurk, H. (1978). Visual Influences on Speech Perception Processes. *Perception and Psychophysic.,* 24(3):253-257.

Magnotti, J.F., Mallick, B. D., Feng, G., Zhou, W., & Beauchamp, M.S. (2015). Similar Frequency of the McGurk Effect in Large Samples of Native Mandarin Chinese and American English Speakers. *Exp. Brain Res.*, 233(9): 2581-2586.

Mallick, D. B., Magnotti, J. F. & Beauchamp, M. S. (2015). Variability and Stability in the McGurk Effect: Contributions of Participants, Stimuli, Time, and Response Type. *Psychon Bull Rev.*, 22(5):1299-307. doi: 10.3758/s13423-015-0817-4. PMID: 25802068; PMCID: PMC4580505.

Massaro, D. W., Thompson, L. A., Barron, B. & Laren, E. (1986). Developmental Changes in Visual and Auditory Contributions to Speech Perception. *Journal. Exp. Child Psychol.*, 41(1):93-113. doi: 10.1016/0022-0965(86)90053-6. PMID: 3950540.

Masterton, R. B. (1992). Role of the Central Auditory System in Hearing: The New Direction. *Trends Neurosci*., 15(8):280-5. doi: 10.1016/0166-2236(92)90077-l. PMID: 1384196.

Mattingly, I. G. & Studdert-Kennedy, M. (1991). *Modularity and the Motor Theory of Speech Perception.* Erlbaum, Hillsdale, NJ, USA.

McGurk, H. & MacDonald, J. (1976). Hearing Lips and Seeing Voices. *Nature,* 264:746-748. doi: 10.1038/264746a0. PMID: 1012311.

Miller, G. A., & Nicely, P. E. (1955). An Analysis of Perceptual Confusions among some English consonants. *Journal of the Acoustical Society of America*, 27:338-353 as cited in MacDonald and McGurk, 1978.

Mongillo, E., Irwin, J., Whalen, D. & Klaiman, C. (2008). Audiovisual Processing in Children With and Without Autism Spectrum Disorders. *Journal of Autism and Developmental Disorders*. 38 (7):1349-1358. doi:10.1007/s10803-007-0521-y. PMID 18307027.

Moore, B. C. J., Tyler, L. K. & Marslen-Wilson, W. D. (2009). *The Perception of Speech: From Sound to Meaning.* Oxford, UK.

Munhall, K, G., Gribble, P., Sacco, L., & Ward, M. (1996). Temporal Constraints on the McGurk effect. *Percept. Psychophys*., 58(3): 351-362.

Norrix, L., Plante, E., Vance, R. & Boliek, C. (2007). Auditory-visual Integration for Speech by Children with and without Specific Language Impairment. *Journal of Speech, Language, and Hearing Research*, 50 (6):1639-1651. doi:10.1044/1092-4388(2007/111).

Peelle, J. E., & Sommers, K. S. (2015). Prediction and Constraint in Audiovisual Speech Perception. *Cortex*, 68: 169-181.

Raphael, L. J., Borden, G. J., & Harris, K. S. (2011). *Speech Science Primer: Physiology, Acoustics, and Perception of Speech*. Lippincott Williams & Wilkins.

Rosenblum, L. D. & Saldaña, H. (1996). An Audio-Visual Test of Kinematic Primitives for Visual Speech Perception. *Journal of Experimental Psychology: Human Perception and Performance,* **22**: 318–331.

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

Sekiyama, K., Soshi, T. & Sakamoto, S. (2014). Enhanced Audiovisual Integration with Aging in Speech Perception: A Heightened McGurk effect in Older Adults. *Front Psychol.*, 14(5):323. doi: 10.3389/fpsyg.2014.00323. PMID: 24782815; PMCID: PMC3995044.

Sekiyama, K. & Tohkura, Y. (1991). Japanese Subjects Hearing Japanese Syllables of High Auditory Intelligibility. *Journal of the Acoustical Society of America*, 90:1797–1805.

Simmons, A. A. (1959). Factors Related to Lipreading. *Journal of Speech and Hearing Research,* 2:340-352.

Strand, J. F., Cooperman, A., Rowe, J., & Simenstad, A. (2014). Individual Differences in Susceptibility to the McGurk Effect: Links with Lipreading and Detecting Audiovisual Incongruity. *J. Speech Lang. Hear. Res.*, 57(6):2322-31. doi: 10.1044/2014_JSLHR-H-14-0059. PMID: 25296272.

Sumby, W. H. & Pollack, I. (1954). Visual Contribution to Speech Intelligibility in Noise. *Journal of Acoustical Society of America,* 26:212-215.

Van Engen, K. J., Dey, A., Sommers, M. S. & Peelle, J. E. (2022). Audiovisual Speech Perception: Moving Beyond McGurk. *J Acoust Soc Am.,* 152(6):3216. doi: 10.1121/10.0015262. PMID: 36586857; PMCID: PMC9894660.

Walker, S., Bruce, V. & O'Malley, C. (1995). Facial Identity and Facial Speech Processing: Familiar Faces and Voices in the Mcgurk Effect. *Perception & Psychophysics*, 57 (8):1124-1133. doi:10.3758/bf03208369.

Wood, C. C. (1975). Auditory and Phonetic Levels of Processing in Speech Perception: Neurophysiological and Information Processing *Analyses. Journal of Experimental Psychology (Human Perception and Performance),* 104:3-20 as cited in Dodd, 1977.

Woodhouse, L., Hickson, L. & Dodd, B. (2009). Review of Visual Speech Perception by Hearing and Hearing-Impaired People: Clinical Implications. *Int J Lang Commun Disord.*, 44(3):253-70. doi: 10.1080/13682820802090281. PMID: 18821117.