Evidence for a Generic Process Underlying Multisensory Integration

Corinne Tremblay[1,2], François Champoux[3], Benoit A. Bacon[4] and Hugo Théoret[*,1,2]

[1]*Psychology, University of Montreal, Montreal, Canada*

[2]*Research Center, Sainte-Justine Hospital, Montreal, Canada*

[3]*Speech language pathology and Audiology, University of Montreal, Montreal, Canada*

[4]*Psychology, Bishop's University, Sherbrooke, Canada*

Abstract: It has been shown repeatedly that the various sensory modalities interact with each other and that the integration of incongruent percepts across two modalities, such as vision and audition, can lead to illusions. Different individual cognitive features (i.e., attention, linguistic experience, etc.) have been shown to modulate the level of multisensory integration. As such, it may be hypothesized that an intra-individual generic process underlies parts of illusory perception, irrespective of illusory material. One simple way to address this issue is to assess whether observers experience multisensory integration to a similar degree when the illusory material varies with respect to its sensory features. Here, performance on two distinct audio-visual illusions (McGurk effect, illusory flash effect) was tested in a group of adult observers. Results show a positive within-subject correlation between both illusions indirectly supporting the existence of a generic process for multisensory integration that could include individual differences in attention.

Keywords: McGurk effect, illusory flash effect, multisensory integration.

INTRODUCTION

The ability to combine visual stimuli with the auditory stimuli that are related to them is critical. Indeed, identification of an object or its position in space relies on this integration (or segregation) of multiple audio and visual inputs. When vision and audition deliver incongruent information, the interaction between modalities can lead to illusions. The McGurk effect [1] is a well known speech illusion where vision biases audition. In this classic demonstration, incongruent lip movements induce the misperception of spoken syllables. For example, upon hearing /baba/ but seeing /gaga/, most normal subjects will report hearing the fused percept /dada/ [1]. Subsequent studies have confirmed that the McGurk effect is a very robust illusion [2,3]. Although vision was at first believed to dominate audio-visual interactions, recent findings suggest that auditory inputs can also bias visual perception. Shams and collaborators reported that a single visual flash can be perceived as two flashes if it is accompanied by two (rather than one) closely successive sounds [4]. This illusion, known as the illusory flash effect, has been shown to occur in healthy observers despite important differences in contrast, form and texture, duration of flash and auditory signals, as well as spatial disparity between the sound and the flash [4].

There are numerous reports of considerable inter-individual differences in audio-visual integration. For example, the McGurk and illusory flash effects do not occur in all individuals and their respective strength varies widely across observers. Furthermore, motivation and personality [5], linguistic experience [6], sex [7] and attention [8] have all been shown to modulate the level of multisensory integration occurring at the individual level. These data suggest that a generic process could underlie multisensory integration, where specific individual features modulate the strength of audio-visual integration. Following on this, it may be hypothesized that the level of audio-visual integration in one illusion predicts the strength of integration in another illusion. To gain insight into the rules governing different types of multisensory integration, the degree to which observers experienced two well-known audio-visual illusions was assessed in a within-subjects design. To this end, performance on the McGurk effect (a speech illusion where vision biases audition) and illusory flash effect (a non-speech illusion where audition biases vision) was evaluated and correlated at the individual level. A high correlation across illusions would tend to support the existence of an intra-individual generic process not attributable to specific features of both illusions.

MATERIALS AND METHODS

Participants

Nineteen observers (11 males, 3 left-handed) between 18 and 30 years of age gave written informed consent and participated in the study. All participants had normal pure-tone audiometric thresholds at octave frequencies between 250 and 8000 Hz. They also had normal or corrected-to-normal vision (Snellen chart). The study was approved by the Research Ethics Board of Sainte-Justine Hospital.

Stimuli and Design

For the McGurk effect task, a male speaker was videotaped saying the consonant-vowel syllables /ba/ and /va/. Production began and ended in a neutral, closed mouth position. One utterance of /ba/ and one utterance of /va/, of the same duration, were selected for inclusion in the study. Two congruent conditions were set from these audio-visual utterances. In the unimodal condition, the audio sequence of the

*Address correspondence to this author at the Psychology, University of Montreal, CP 6128, Succ. Centre-Ville, Montréal, QC, H3C 3J7, Canada; Tel: 514-343-6362; Fax: 514-343-5787; E-mail: hugo.theoret@umontreal.ca

1874-2300/07 2007 Bentham Science Publishers Ltd.

syllable /ba/ was used without the video sequence. In the congruent bimodal condition, the video sequence of the syllable /ba/ was paired with the audio sequence of the same utterance. In the incongruent bimodal condition, the video sequence of the syllable /va/ was paired with the audio sequence of the syllable /ba/. The temporal synchrony of the visual /va/ and the auditory /ba/ was achieved by aligning the burst corresponding to the beginning of the /b/ in the auditory stimulus with the beginning of the /v/ in the video sequence. In this version of the McGurk illusion, the fusion of the incongruent auditory and visual stimuli typically gives rise to the percept /va/ (i.e., vision dominates) [9].

The characteristics of the stimuli used in the illusory flash effect task were similar to those used in the original experiments [4,10,11]. The flash was a white circle subtending 2 degrees of visual angle. It had a luminance of 0.02 cd/m2 and it appeared for 67 ms, either once or twice. When it appeared twice, an interval of 67 ms separated the two flashes. The auditory signal consisted of one or two 7 ms beeps that had a frequency of 3500 Hz. When a single beep was presented, it occurred 20 ms before the first flash. When there were two beeps, the first occurred 23 ms before the first flash (or before the single flash), and the second occurred 67 ms later. Between trials, participants fixated on a cross in the center of the screen.

Procedure

For both effects, the auditory stimuli (/ba/ utterance and beeps) were always projected *via* two loudspeakers positioned at ear level and located on each side of a 17" video monitor at 60 dB HL. In each task, stimuli were presented at the participant's eye level. The McGurk effect and the illusory flash effect tasks were performed in a single session, in a counterbalanced order. Testing took place in a semi-dark room with participants sitting 57 cm away from the computer monitor. The entire procedure took approximately 30 minutes.

For the McGurk effect, each condition (unimodal, bimodal congruent, bimodal incongruent) was presented 10 times in random order. Participants were asked to watch the screen, listen to the speakers and report whatever they had heard. Performance was determined as the percentage of "auditory-based" responses (/ba/) reported by the participant (out of 10) in the bimodal incongruent condition. This percentage was used to calculate the proportion of audio-visual integration responses. More specifically, an audio-visual integration response was deemed to have occurred whenever the participant reported hearing anything other than a /ba/. For the illusory flash effect, all congruent stimuli (one flash-one beep or two flashes-two beeps) and incongruent stimuli (one flash-two beeps) were presented in randomized order, with ten trials per condition. Subjects were asked to watch, listen and report the number of flashes that they had seen on the screen (one or two). The average number of reported flashes for each participant was used as the dependent variable.

RESULTS

McGurk Effect

Participants were very accurate in non-illusory trials. In the unimodal trials, participants correctly identified the syllable /ba/ in 86% of the trials. Performance increased to 99% correct answers when the congruent visual stimulus /ba/ was added. A one way repeated-measures ANOVA revealed a main effect of CONDITION (unimodal, bimodal congruent, bimodal incongruent; $F_{2,18} = 57.64$, $p < 0.001$). As shown in Fig. (**1A**), participants reported significantly less "auditory-based" /ba/ responses in the bimodal incongruent condition than during the unimodal ($t_{18} = 8.14$, $p < 0.001$) or bimodal congruent ($t_{18} = 8.02$, $p < 0.001$) conditions. This replicates the original findings of McGurk and McDonald [1].

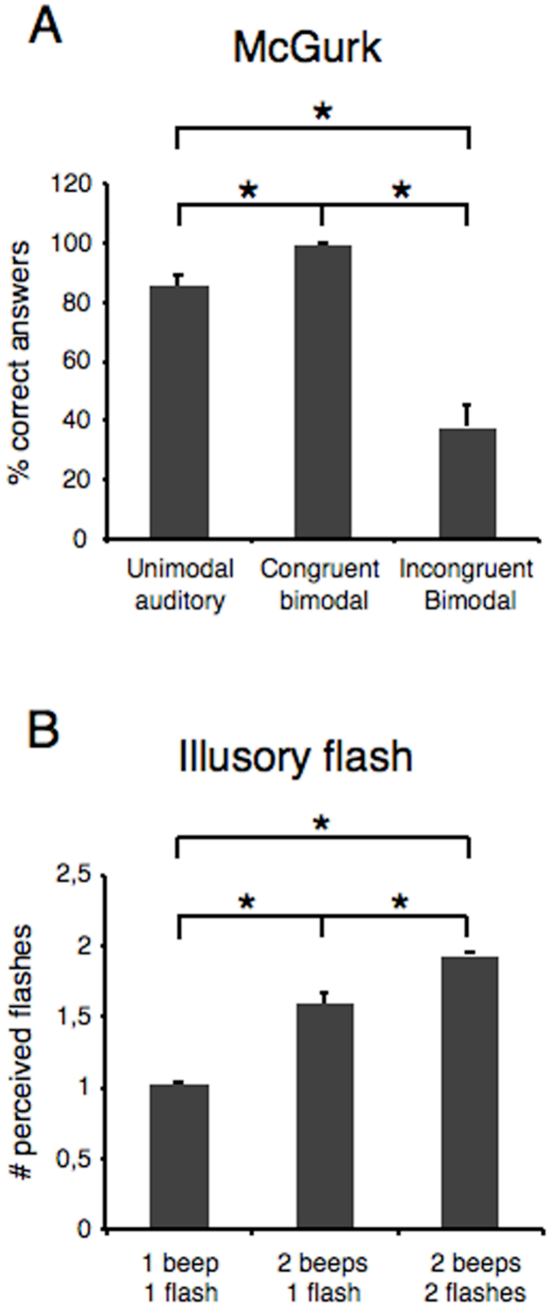

Fig. (1). (**A**) Percentage of "auditory-based" /ba/ responses stimuli on the McGurk task. (**B**) Number of perceived flashes in the three experimental conditions. * $p < 0.001$.

Illusory Flash Effect

In both control conditions, participants were very accurate in reporting the number of perceived flashes. In the one flash-one beep condition, participants reported an average of 1.02 flashes. In the two flashes-two beeps condition, participants reported an average of 1.94 flashes. A one way repeated-measures ANOVA revealed a main effect of CONDITION (one flash-one beep, two flashes-two beeps, one flash-two beeps; $F_{2,18} = 136.49$, $p < 0.001$). As shown in Fig. (**1B**), the illusory condition (one flash-two beeps) yielded significantly more flash reports than the one flash-one beep condition ($t_{18} = -8.13$, $p < 0.001$) and significantly less than the two flashes-two beeps condition. This replicates the original findings of Shams and collaborators [10].

Correlation Analysis

Within-subject correlational analysis (Pearson's r) revealed that perception of both illusions was significantly and positively correlated (Fig. (**2**); $r = 0.6$; $p < 0.05$). In other words, participants that were more susceptible to the McGurk speech illusion (reporting less "auditory-based" /ba/ responses) were also more susceptible to the non-speech illusory flash illusion (reporting more flashes in the one flash-two beeps condition).

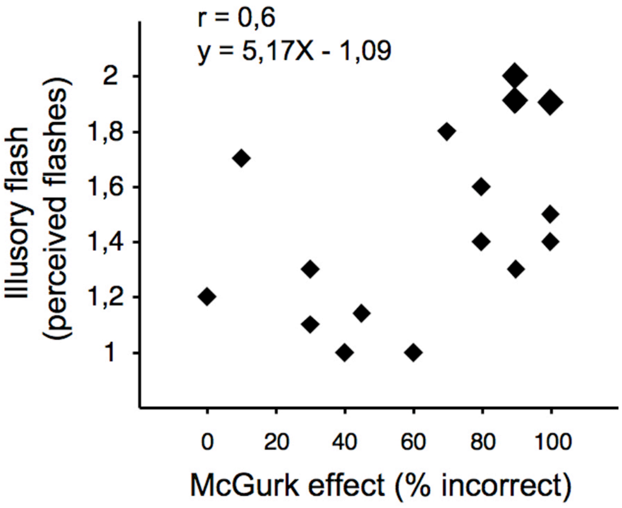

Fig. (2). Correlation plot of the strength of the illusory flash effect (number of perceived flashes in the *one flash-two beeps* condition) and the McGurk effect (percentage of visual dominant responses in the *incongruent bimodal* condition). Larger symbols reflect two participants with the same *x* and *y* values.

DISCUSSION

The results of the present experiment reveal a positive within-subject correlation between the McGurk effect and the illusory flash effect. This suggests that participants susceptible to one illusion are also prone to process other multisensory percepts similarly. Interestingly, our results suggest that within-subject sensitivity is similar in a task where vision biases audition (e.g., McGurk effect) as well as in a task where auditory inputs bias visual perception (e.g., illusory flash effect). Finally, our results indicate that speech and non-speech audio-visual integration may share a common substrate.

Tuomainen and collaborators have argued in favor of a specialized process underlying audio-visual speech integration [12]. This is partly based on the fact that presentation of nonwords in a McGurk-like fashion (auditory-visual incongruence) leads to multisensory integration only when participants are trained to perceive the auditory material as speech. They argue that attentional mechanisms may explain the predominance of integration effects in the 'speech mode', where speech aspects of the material may have enhanced attention to features associated with the phonetic nature of the audio-visual 'object'. Our data are not incompatible with the existence of specific modes of integration. Indeed, multisensory integration relies on complex networks distributed throughout the brain. Many areas have been identified, including the superior colliculus, the insula/claustrum, the thalamus, the superior temporal sulsus, the intraparietal sulsus, the frontal cortex and even sensory-specific cortices (for a review, see [13]). Parts of these networks have been show to be differentially implicated in the multisensory integrative processes. As an example, the intraparietal sulcus appears to have a more prominent role in determining spatial location of an object [14,15]. In contrast, the superior temporal sulcus has been repeatedly shown to play an important role in the synthesizing audio-visual speech information [14,16,17]. The presence of distinct multisensory processes underlying specific aspects of audio-visual integration does not, however, preclude the existence of common mechanisms subserving certain aspects of the multisensory experience. Indeed, specialized areas of integration in the human brain reflect a predominance of activation, whereas multiple brain regions are inevitably activated whatever the nature of the multisensory task at hand. It is thus reasonable to assume that in addition to processes dedicated to specific aspects of multisensory integration, general mechanisms are also necessary to combine auditory and visual inputs in a meaningful manner. One may thus wonder what brain mechanisms underlie specific and generic modes of multisensory integration. Recent neuroimaging studies have investigated the neural basis of the McGurk and illusory flash effects. McGurk-type stimuli appear to recruit predominantly superior temporal and posterior parietal cortex during incongruent trials [18-20]. For the illsuory flash effect, fMRI and EEG studies suggest that a complex interplay between auditory, visual and polymodal areas underlies the illusory perception of a second flash. Early responses to the illusory flash have been found in early visual areas [10,21], which are folllowed by activations in superior temporal cortex [21]. Importantly, early modulation of visual cortex by illusory perception of the second flash has been shown to be stronger in participants that are more susceptible to the flash illusion [21]. Furthermore, early visual cortex activity differences between participants who frequently see the illusory flash illusion and those who do not are also present in a variety of audio-visual stimulus combinations [21]. Mishra and collaborators [21] have suggested that individual differences in functional connectivity between sensory areas may explain a general pattern of behavioral response to multiple combinations of auditory and visual material. Further studies are required to determine the neural mechanism underlying individual difference in mulstisensory integration.

An important issue pertains to the validity of correlational approaches in establishing a direct link between two processes. The fact that performance on both illusory tasks was correlated within subjects does not necessarily imply a common mechanism directly related to multisensory integration. Indeed, as was previously shown, a wide variety of factors such as motivation and personality can influence the strength of multisensory integration [5]. It can therefore be assumed that a general mechanism, which varies at the individual level, is involved in the process linking auditory and visual inputs into a meaningful percept. It may be argued that the efficiency with which individual participants focus attention on the different illusions explains the pattern of responses reported here. Indeed, it has been shown that the McGurk illusion can be practically abolished by having subjects perform a concurrent, unrelated task [8]. If one assumes, then, a constant individual level of attention across tasks, the significant correlation between illusions could reflect interindividual differences in attention rather than multisensory integration. A possible way to investigate the role of attentional mechanisms in the pattern of behavioral responses reported here would be to ask subjects to perform a concurrent, attention-demanding task, during the presentation of both illusions. Individual differences in attention could also be evaluated by standard neuropsychological testing and related to the level of multisensory integration.

Perceptual stability is another factor that could account for the within-subject correlation reported here. Including another perceptual illusion task that does not implicate sensory integration, such as Necker cube reversal, would control for this. As such, strong evidence for an underlying mechanism uniquely responsible for sensory integration can only be provided when specific factors such as attention, motivation and perceptual stability are teased out. Nevertheless, the preliminary findings reported here show that the physical nature of sensory stimuli interacting to produce a perceptual illusion cannot entirely explain individual patterns of illusory perception. Rather, specific factors directly related to multisensory integration or reflecting a general mechanism of sensory processing also have an impact on the degree to which integration occurs. Although the present study cannot determine which factors are involved in this process and to what degree they influence illusory perception, it shows the importance of probing individual-level multisensory integration to better understand the complex interaction between sensory modalities.

CONCLUSION

The data presented here show that the strength of two audio-visual illusions with distinct properties (speech *vs* non-speech, visual dominance *vs* auditory dominance) is correlated at the individual level, which suggest common individual properties in the integration of multisensory material. More studies are needed to establish what specific individual factors underlie the level to which audio-visual integration occurs.

ACKNOWLEDGEMENTS

This work was supported by grants from CIHR, CFI and FRSQ to HT. CT was supported by a scholarship from FRSQ and FC by a scholarship from CIHR. We thank Dr Yukiyasu Kamitani for help programming the illusory flash stimulus.

REFERENCES

[1] McGurk H, MacDonald J. Hearing lips and seeing voices. Nature 1976; 264: 746-48.

[2] Massaro DW, Cohen MM. Perception of synthesized audible and visible speech. Psychol Sci 1990; 1: 55-63.

[3] Rosenblum LD, Saldana HM. An audiovisual test of kinematic primitives for visual speech perception. J Exp Psychol Hum Percept Perform 1996; 22: 318-31.

[4] Shams L, Kamitani Y, Shimojo S. What you see is what you hear. Nature 2000; 408: 788.

[5] Giolas TG, Butterfield EC, Weaver SJ. Some motivational correlates of lipreading. J Speech Hear Res 1974; 17: 18-24.

[6] Sekiyama K, Tohkura Y. McGurk effect in non-English listeners: few visual effects for Japanese subjects hearing Japanese syllables of high auditory intelligibility. J Acoust Soc Am 1991; 90: 1797-805.

[7] Irwin JR, Whalen DH, Fowler CA. A sex difference in visual influence on heard speech. Percept Psychophys 2006; 68: 582-92.

[8] Alsius A, Navarra J, Campbell R, Soto-Faraco S. Audiovisual integration of speech falters under high attention demands. Curr Biol 2005; 15: 839-43.

[9] Rosenblum LD, Schmuckler MA, Johnson JA. The McGurk effect in infants. Psychophys 1997; 59: 347-57.

[10] Shams L, Kamitani Y, Thompson S, Shimojo S. Sound alters visual evoked potentials in humans. Neuroreport 2001; 12: 3849-52.

[11] Shams L, Kamitani Y, Shimojo S. Visual illusion induced by sound. Cogn Brain Res 2002; 14: 147-52.

[12] Tuomainen J, Andersen TS, Tiippana K, Sams M. Audio-visual speech perception is special. Cognition 2005; 96: 13-22.

[13] Calvert GA. Crossmodal processing in the human brain: insights from functional neuroimaging studies. Cereb Cortex 2001; 11: 1110-23.

[14] Macaluso E, George N, Dolan R, Spence C, Driver J. Spatial and temporal factors during processing of audiovisual speech: a PET study. Neuroimage 2004; 21: 725-32.

[15] Sestieri C, Di Matteo R, Ferretti A, *et al.* "What" versus "where" in the audiovisual domain: an fMRI study. Neuroimage 2006; 33: 672-80.

[16] Calvert GA, Campbell R, Brammer MJ. Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex. Curr Biol 2000; 10: 649-57.

[17] Olson IR, Gatenby JC, Gore JC. A comparison of bound and unbound audio-visual information processing in the human cerebral cortex. Brain Res Cogn Brain Res 2002; 14: 129-38.

[18] Jones JA, Callan DE. Brain activity during audiovisual speech perception: an fMRI study of the McGurk effect. Neuroreport 2003; 14: 1129-1133.

[19] Sekiyama K, Kanno I, Miura S, Sugita Y. Auditory-visual speech perception examined by fMRI and PET. Neurosci Res 2003; 47: 277-287.

[20] Kaiser J, Hertrich I, Ackermann H, Mathiak K, Lutzenberger W. Hearing lips: gamma-band activity during audiovisual speech perception. Cereb Cortex 2005; 15: 646-653.

[21] Mishra J, Martinez A, Sejnowski TJ, Hillyard SA. Early cross-modal interactions in auditory and visual cortex underlie a sound-induced visual illusion. J Neurosci 2007; 27: 4120-4131.

Received: September 3, 2007 Revised: September 19, 2007 Accepted: September 20, 2007