SUSCEPTIBILITY TO AN AUDIO-VISUAL SPEECH ILLUSION IN OLDER AND YOUNGER ADULTS

Annalisa Setti[12], Kate E. Burke[12], RoseAnne Kenny[13] and Fiona N. Newell[12]
[1]Trinity College Institute of Neuroscience [2] School of Psychology, [3]Department of Medical Gerontology, Trinity College Dublin.
asetti@tcd.ie

Abstract

In the present work we investigated older and younger control participants' susceptibility to an audio-visual speech illusion known as the McGurk effect that occurs when an incongruent phoneme [|ba|] and viseme [ga] are perceived as a new fused percept ['da']. We hypothesized that if older persons integrate information from different senses more than younger, their susceptibility to the illusion should be higher. The results confirmed this hypothesis. We suggest that difficulty in focusing on one channel (audition) while simultaneously perceiving inputs from other channels (vision) is the reason of this enhanced integration.

Multiple sources of information are available in the environment. The senses capture these inputs and the brain is responsible for combining congruent information appropriately, while separating information that belongs to different objects or events (Stein & Meredith, 1993). During the ageing process the quality of sensory inputs diminish due to degradation of the sensory organs (Fozard & Gordon-Salant, 2001; Gordon-Salant, 2005; Schieber, 2006). The ageing brain is different to the younger brain in that it adapts to these changes by relying on the combination of different (sometimes declining) sensory inputs (Laurienti, Burdette, Maldjian, & Wallace, 2006), thus taking advantage of redundant information in the environment (Calvert, Spence, & Stein, 2004). Older participants show greater benefit from audio-visual information compared to younger adults, i.e. greater probability to combine auditory and visual inputs (Laurienti et al., 2006; Peiffer, Mozolic, Hugenschmidt, & Laurienti, 2007). They appear to be able to compensate for their worsening eyesight or poor hearing by combining the inputs from these senses.

However when older adults are presented with irrelevant information, they find it more difficult to ignore (Hugenschmidt, Mozolic, & Laurienti, 2009; Hugenschmidt, Mozolic, Tan, Kraft, & Laurienti, 2009).

A prime example of audio-visual speech integration is the McGurk illusion (McGurk & MacDonald, 1976). In the original experiment the auditory syllable /ba/ was combined with a face articulating [ga]. The participants reported to perceive 'da', a new percept obtained by the fusion of the sound and the visual input.

In the present work we used the susceptibility to the McGurk illusion to assess audio-visual integration abilities in younger and older persons. To our knowledge only one study investigated the McGurk illusion in older adults with normal hearing abilities and no overall difference was found in the percentage of fused responses reported by the younger and older participants (Cienkowski & Carney, 2002). However measures of integration at syllable, word and sentence level do not always correlate (Sommers, Tye-Murray, & Spehar, 2005).

We hypothesised that if performance at syllable level does not reflect performance at word level multisensory integration could be enhanced in older adults if we used words rather than syllables to test the McGurk effect (e.g. Hugenschmidt et al., 2009).

Method

Participants

Twenty-six participants volunteered for this study, 13 were younger adults (6 male) (mean age = 22, SD = 4) and 13 were community dwelling older adults (6 male) (mean age= 65.5, SD = 4). The experiment was approved by the St. James Hospital Ethics Committee and by the School of Psychology Research Ethics Committee, Trinity College Dublin and they conform to the Declaration of Helsinki. All participants provided informed, written consent to take part in the study.

All older participants obtained a MMSE score higher than 27 (M = 29, SD = 1), young participants were not tested for their MMSE score as we assumed that it was higher than 27. Older participants were also screened for vision (logMAR) and hearing (Hughson Westlake test with Kamplex BA 25 Screening Audiometer). Their vision was normal to corrected (M = 0.05, SD = 0.05), and their hearing was normal for their age range. Younger participants had normal hearing and vision by self-report. The two groups of participants were matched for level of education (1 junior certificate; 7 diploma; 5 college).

Stimuli and Material

The stimuli were digital recordings of a female face pronouncing words (e.g. |bale| / [cale]). The recordings were made in a quiet room with natural light illumination with a JVC high band digital video camera and were edited with Adobe Premiere® software.

Thirty-six auditory and visual words (17 of which taken from Bargary et al., 2009) were re-combined to originate the McGurk illusion condition, one of these combinations was used as training (2 combinations were excluded due to technical problems from further analyses). In the Auditory only condition the auditory stimuli were presented with a pixelated version of the original visual stimulus, the pixelation comprised on average 6 pixels in the horizontal axis (from ear to ear) and 12 in the vertical axis (from chin to forehead). In the Visual only condition the visual words in the Auditory + Visual condition were presented with white noise as auditory stimulus.

A separate list with the expected fused words (Congruent condition) was also presented to the participants, in order to check whether there was any bias due to the McGurk words themselves, i.e. they were not familiar to the participants.

Procedure

Participants were seated in front of the computer screen with their chin comfortably positioned on a chin-rest at a distance of 57cm from the screen. They were informed that they would hear and see a person pronouncing words and that their task was to report what the person said. The order of the Audio only, Visual only and Auditory+Visual conditions was counterbalanced between participants. The Congruent block was always presented last. The responses were collected by the experimenter and typed into a Microsoft Excel® file.

At the beginning of each trial a fixation cross appeared at the centre of the computer screen for 700ms followed by the word (A, V or AV condition). The presentation of the trials was self paced.

Results and Discussion

The percentage of responses provided was calculated for each participant in each condition (Auditory Only, Visual Only and Auditory+Visual). In the Auditory+Visual condition the percentage of responses provided was very high both for older and younger adults, 92.5% and 98.6% respectively. In the Auditory Only condition the percentage of responses was also high for both groups (older adults 95.3%; younger adults 97%), while it decreased substantially in the Visual Only condition with 56.6% of responses (either correct

or incorrect) for older participants and 69.7% for younger participants. In the Visual Only the number of responses provided also varied dramatically between participants, with some participants attempting to respond to the high majority of trials and other participants responding to none or very few (Non responses: Older adults mean = 8.8, SD = 9; Younger adults mean = 10.8, SD = 11).

The proportion of correct responses in the Visual Only (older, mean = 0.083; younger, mean = 0.046) ($t(1,24) = 1.47$, $p = 0.15$) and in the Auditory Only ($t(1,24) = 1.12$, $p = 0.27$) condition did not differ between age groups (older, mean = 0.51; younger, mean = 0.48).

Responses in the Auditory+Visual condition have been categorised as McGurk fusion when the participant's response corresponded to the expected fused response either embedded in the expected word or in a similar word. An auditory response was when the participant reported the word presented as auditory stimulus, a visual response was when the participant reported the word presented as visual stimulus and an Other response was when the participant reported a word that did not correspond to any of the previous categories. The other category included words similar to the auditory word that could not be considered as fused responses as the place of articulation was the same for the Auditory and the reported word, suggesting that they could have been simple misperceptions of the auditory stimulus [e.g. |peek| – [tea] – expected fused response 'key', production classified as Other 'pea']. It also could include a completely unrelated word (e.g. |pin| – [tin] – expected fused response 'kin', production classified as Other 'elf').

We then analysed the proportion of McGurk fusion responses in the two groups of participants. Older participants produced significantly more fused responses than younger participants [$t(1,24) = 3.04$, $p < 0.01$] (Figure 1).

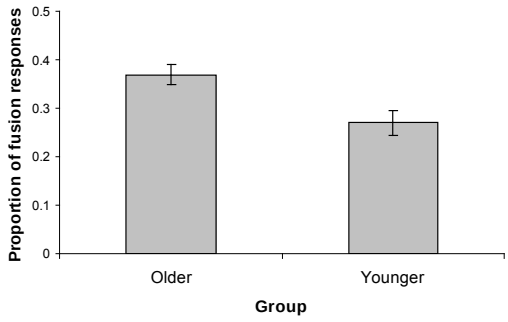

Fig. 1. Proportion of McGurk illusions (fused percept) in older and younger control participants.

We also analysed the proportion of correct responses in the Congruent condition, i.e. when participants were presented with the auditory and the visual version of the expected McGurk fusion stimuli instead of with an incongruent audio-visual stimulus. The average percentage of correct responses for older participants was 82% and 80% for younger participants [$t(1,24) = 0.53$, $p = 0.6$].

In sum, in this Experiment we compared the susceptibility of younger and older participants to the McGurk illusion. We found no differences in their unisensory perception, i.e. both groups were equally able to report the audio and visual word in the Auditory Only and Visual Only conditions respectively, although for both groups the Visual Only task was very difficult.

Importantly older participants were more susceptible to the illusion.. This finding is in line with recent studies showing that older adults are more susceptible to audio-visual interactions than their younger counterparts (Laurienti et al., 2006; Hugenschmidt et al., 2009). In a recent study (Setti, Burke, Kenny, & Newell, 2009) we have also shown that older persons are more susceptible to a sound induced visual illusion (Shams, Kamitani, & Shimojo, 2000). Older participants reported perceiving 2 flashes when 1 flash was presented with 2 auditory beeps more often than younger adults. We suggest that the higher susceptibility to the illusions in older adults is due to their increased difficulty in ignoring sensory inputs that are perceptually salient but not task relevant.

Acknowledgements

This research was completed as part of a wider programme of research within the TRIL Centre, (Technology Research for Independent Living). The TRIL Centre is a multi-disciplinary research centre, bringing together researchers from UCD, TCD, NUIG & Intel, funded by Intel and IDA Ireland. www.trilcentre.org.

References

Calvert, G. A., Spence, C., & Stein, B. E. (Eds.). (2004). *The Handbook of Multisensory Processes*. Boston: MIT.

Cienkowski, K. M., & Carney, A. E. (2002). Auditory-visual speech perception and aging. *Ear Hearing, 23*(5), 439-449.

Fozard, J. L., & Gordon-Salant, S. (2001). Changes in vision and hearing with aging. In J. E. Birren & K. Schaie, W. (Eds.), *Handbook of the psychology of aging* (pp. 241-266). San Diego: Academic Press.

Gordon-Salant, S. (2005). Hearing loss and aging: New research findings and clinical implications. *42*(4, Suppl. 2), 9-24.

Hugenschmidt, C., Mozolic, J., & Laurienti, P. J. (2009). Suppression of multisensory integration by modality-specific attention in aging. *Neuroreport: For Rapid Communication of Neuroscience Research, 20*, 349-353.

Hugenschmidt, C., Mozolic, J., Tan, H., Kraft, R., & Laurienti, P. (2009). Age-Related Increase in Cross-Sensory Noise in Resting and Steady-State Cerebral Perfusion. *Brain Topography*.

Laurienti, P. J., Burdette, J. H., Maldjian, J. A., & Wallace, M. T. (2006). Enhanced multisensory integration in older adults. *Neurobiology of Aging, 27*(8), 1155-1163.

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

Peiffer, A. M., Mozolic, J. L., Hugenschmidt, C. E., & Laurienti, P. J. (2007). Age-related multisensory enhancement in a simple audiovisual detection task. *Neuroreport: For Rapid Communication of Neuroscience Research, 18*(10), 1077-1081.

Schieber, F. (2006). Vision and aging. In *Handbook of Psychology and Aging, VI Ed.* Burlington, MA: Elsevier.

Setti, A., Burke, K. E., Kenny, R. A., & Newell, F. N. (2009). *Audio-visual integration in fall-prone olde adults*. Paper presented at the International Multisensory Research Forum.

Shams, L., Kamitani, Y., & Shimojo, S. (2000). What you see is what you hear. *Nature, 408*(6814), 788-788.

Sommers, M. S., Tye-Murray, N., & Spehar, B. (2005). Auditory-visual speech perception and auditory-visual enhancement in normal-hearing younger and older adults. *Ear Hearing, 26*(3), 263-275.

Stein, B. E., & Meredith, A. (1993). *The merging of the senses.* : Cambridge, MA: MIT Press.