Auditory processing and audiovisual integration revealed by combining psychophysical and fMRI experiments
Tomaskovic, Sonja

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2006

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):

Copyright
Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

Take-down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): http://www.rug.nl/research/portal. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.
Chapter 4

Hearing with the eyes:
the McGurk effect
Chapter 4

Abstract

When the speech sound /aba/ is presented simultaneously with a moving mouth pronouncing the sound /aga/, subjects tend to report hearing /ada/. This perceptual illusion is known as a McGurk effect. This effect is a useful tool for studying audiovisual (AV) integration, since, in the McGurk effect, the auditory perception is changed due to AV integration. In the current experiment the strength of the McGurk effect was studied as a function of AV synchronicity. AV time shifts ranging from -620 ms to +420 ms were used. The results show that the strength of AV integration depends on AV synchronicity.

4.1 Introduction

In speech perception both auditory and visual information, if available, is important. Especially in a noisy environment, seeing the speaker’s face can significantly improve the intelligibility of speech (Sumby and Pollack, 1954). To investigate how auditory and visual information is combined we studied the McGurk effect (McGurk and MacDonald, 1976). McGurk and McDonald showed that when the speech sound /aba/ is presented while a moving mouth pronouncing the sound /aga/ is shown, one perceives a different syllable /ada/. Another combination of syllables has shown comparable effect: audio /apa/ with visual /aka/ leads to the perception of /ata/.

Summerfield (1987) explained this perceptual effect with the hypothesis that visual and acoustical information are complementary in speech perception. This hypothesis is called the VPAM (Visual: Place, Auditory: Manner) hypothesis. The visual cue gives information about the place of articulation (e.g. bilabial, alveolar or velar), and the acoustical cue, about the manner of articulation (e.g. voiceless, voiced or nasal).

At the phonemic level, some cues to the places of articulation may be visible, while cues to manner are distributed between the two modalities
Chapter 4

(e.g. Binnie et al., 1974). Previous studies indicated that in the absence of auditory input, lip movements for /aga/ are frequently perceived as /ada/, and those for /aka/ are perceived as /ata/ (McGurk and MacDonald, 1976; Binnie et al., 1974). Thus, there is visual information for /aga/ and /ada/ and audio information with features common to /ada/ and /aba/. By responding to the common information for both cues, a subject would arrive at the unify percept /ada/ (McGurk and MacDonald, 1976). The perception of the McGurk effect has been shown for a variety of languages, combination of stimuli, and ages of subjects. It is a surprisingly robust effect and does not habituate over time, despite objective knowledge of the illusion involved (McGurk and MacDonald, 1976). Dubbing the male voice onto the female face did not have an influence on the perception of the McGurk effect (Green et al., 2001). The mechanism for integrating speech information from the auditory and the visual modalities is not disrupted by a gender incompatibility even when it is perceptually evident. McGurk and McDonald (1976) reported that where responses were dominated by a single modality, for adults this tends to be visual, and auditory for children. Although, the effect is observed across a wide age range of subjects, there are clearly age related dependencies in susceptibility, which can indicate that the McGurk effect, although not experience based, is influenced by experience. The main interest of the present study was to investigate how the strength of the McGurk effect depends on audiovisual synchronicity. This psychophysical experiment was carried out as an introduction to an extensive neuroimaging study of the audiovisual integration and the McGurk effect (chapter 5).

4.2 Materials and methods

4.2.1 Stimuli

Different VCV (vowel-consonant-vowel) stimuli were pronounced by a woman of Dutch origin and were recorded with a Panasonic 3 CCD 24x dig. zoom
video camera. The frame rate was 25 Hz; the audio signal had a sampling rate of 48 kHz and a bit depth of 16 bits per sample. Stimuli that were used in the experiment: visual /aga/ with audio /aba/, and visual /aka/ with audio /apa/, with different audiovisual (AV) delays (from -620 to +420 ms). Pink noise was added to the audio stimuli to mask environmental noise. Auditory stimuli were presented via headphones (Telephonics TDH-49P) and visual stimuli were presented at a computer screen at a distance of approximately 60 cm. The audio signal was always in the middle of the stimulus. McGurk stimuli with AV delays were made by shifting the video signal. The duration of the VCV was approximately 730 ms. In order to have a silent period before and after stimuli, all stimuli were made 2 s long. In Adobe Premiere time shifts were made with minimum delay steps of 40 ms. After a pilot study, the following set of audio delays were selected: -620, -500, -380, -260, -180, -100, -60, -20, +20, +60, +100, +140, +220, +300, +420 ms. The A-weighted sound pressure levels (SPL(A)) of the stimuli /aba/ and /apa/ were 75 dB(A), and the pink noise 54 dB(A), so the stimuli were presented at an S/N ratio of 21 dB.

4.2.2 Subjects

Twenty subjects participated in the experiment (age 24 to 54; 14 male). They were all of Dutch origin\(^1\) and had normal or corrected to normal vision.

4.2.3 Experimental setup

In one experimental session 30 different stimuli, were presented in random order, and repeated seven times. In total three experimental sessions were carried out. The subjects were sitting in front of a computer monitor and they were asked to look at the screen constantly and to listen carefully. After each

\(^1\)We tested six additional subjects of Italian origin, and three of them were susceptible to the McGurk effect. We excluded Italian subjects from the further analysis to have uniform group of subjects.
stimulus the subjects were asked to report what they heard by pressing one of four labeled keys on the keyboard. The four possible answers the subjects could give were the two real audio stimuli (/aba/ and /apa/) and the two McGurk stimuli (/ada/ and /ata/). They were presented in alphabetical order, /aba/, /ada/, /apa/, and /ata/ (figure 4.1).

4.2.4 Data analysis

The programs to present the stimuli and to analyzed the data were made in Borland Delphi 6. Kyplot 2.0 (https://www.kyenslab.com, 1997-2000) was used to present results.

4.3 Results

The overall results of twenty subjects is shown on figure 4.2. On the horizontal axes time delays between audio and visual information are presented (negative numbers on the abscissa indicate that audio stimuli preceded the
4.4 Discussion

The data presented in the current study is in agreement with previous studies of the McGurk effect. The perfect synchronization of visual and auditory speech information is not the major determinant of audiovisual integration in speech perception. The McGurk effect is the strongest when visual information lead for 60 ms, and the effect is still present for the extreme delays (-620 ms and +420 ms) in our experiment. McGrath and Summerfield (1985) demonstrated that the time delays, between audio and visual information, of 40 ms should not affect audiovisual speech perception. Furthermore, subjects’ perception was influenced by the visual stimuli even when the auditory information lagged the visual information by 180 ms (Munhall et al., 1996). Similarly, increasing the angle between the auditory and visual stimu-
lus sources has little effect on the McGurk effect (Jones and Munhall, 1997).

It is easier to identify syllables on audio information alone than on visual information alone (Girin et al., 2001). Just watching the mouth pronouncing /aga/ or /aka/, without an audio input is much harder to resolve. Furthermore, for the extreme audio delay -620 ms, when there is no overlap between audio and visual information, 100% perception of the real audio signal was expected. At the other extreme of +420 ms, the visual stimulus has already finished when the audio stimulus begins and should therefore have no effect on what is heard. Shortly, two blocks of information are independent, in both cases, but there is never 100% score, on a real audio stimuli, accomplished. The subjects indeed heard the real audio signal but the memory of the visual stimulus is still present, and it might change subjects’ perception. Thus, their responses altered between real audio stimuli, /aba/ and /apa/ and integrated, ‘McGurk’ stimuli, /ada/ and /ata/.

4.5 Conclusions

This paper confirms previous findings that the McGurk effect is the strongest, not for synchronized audio and visual stimuli, but when the visual stimulus leads the audio. Furthermore, for long time delays between audio and visual stimuli (up to 620 ms) it seems that there is still audiovisual interaction present, although to a lower extent.

Acknowledgements

Under the supervision of Esther Wiersinga - Post these measurements were preformed and the credit should be given to her and Femke de Smit for making programs for stimulus presentation, and data analysis.
References