Available online at www.sciencedirect.com
SCIENCE DIRECT
Cognitive Brain Research 25 (2005) 499 – 507
COGNITIVE BRAIN RESEARCH
www.elsevier.com/locate/cogbrainres

Research Report

Exposure to asynchronous audiovisual speech extends the temporal window for audiovisual integration

Jordi Navarra[a,b,*], Argiro Vatakis[b], Massimiliano Zampini[b], Salvador Soto-Faraco[a], William Humphreys[b], Charles Spence[b]

[a]Grup de Recerca Neurociència Cognitiva, Parc Científic de Barcelona, Spain
[b]Department of Experimental Psychology, University of Oxford, UK

Accepted 31 July 2005
Available online 31 August 2005

Abstract

We examined whether monitoring asynchronous audiovisual speech induces a general temporal recalibration of auditory and visual sensory processing. Participants monitored a videotape featuring a speaker pronouncing a list of words (Experiments 1 and 3) or a hand playing a musical pattern on a piano (Experiment 2). The auditory and visual channels were either presented in synchrony, or else asynchronously (with the visual signal leading the auditory signal by 300 ms; Experiments 1 and 2). While performing the monitoring task, participants were asked to judge the temporal order of pairs of auditory (white noise bursts) and visual stimuli (flashes) that were presented at varying stimulus onset asynchronies (SOAs) during the session. The results showed that, while monitoring desynchronized speech or music, participants required a longer interval between the auditory and visual stimuli in order to perceive their temporal order correctly, suggesting a widening of the temporal window for audiovisual integration. The fact that no such recalibration occurred when we used a longer asynchrony (1000 ms) that exceeded the temporal window for audiovisual integration (Experiment 3) supports this conclusion.

Theme: Neural basis of behavior
Topic: Cognition

Keywords: Multisensory integration; Asynchrony; Temporal recalibration; Speech; Music; Temporal order judgment

1. Introduction

One of the best examples of human multisensory integration is provided by audiovisual speech perception (see [6]). The integration of audiovisual speech information occurs automatically whenever both acoustic and visual (lip movement) information are available simultaneously, even if this results in an illusory percept. This is perhaps best illustrated by the so-called McGurk effect, whereby the observer experiences hearing /da/ when presented with the sound /ba/ while viewing the lip movements associated with /ga/ (e.g., [22]; see also [33]).

Temporal coincidence has been identified as one of the most important factors determining whether or not multisensory integration takes place (see [6,8,31], for reviews). For instance, the audiovisual integration of speech breaks down if the asynchrony between the visual lip movements and the auditory speech sounds becomes too great (e.g., [9,13,19,20,24]). However, strict temporal overlap is not necessary, as the perceptual system can accommodate some degree of asynchrony, especially between correlated multisensory inputs (i.e., the lips match the sounds). This supports the idea that there is a temporal window within which multisensory integration can take place (e.g., [34,40]). The McGurk illusion, for

* Corresponding author. Departament de Psicologia Bàsica, Universitat de Barcelona, Pg. Vall d'Hebron, 171, 08035 Barcelona, Spain. Fax: +34 93 402 13 63.
E-mail address: jordi.navarra@ub.edu (J. Navarra).

0926-6410/$ - see front matter © 2005 Elsevier B.V. All rights reserved.
doi:10.1016/j.cogbrainres.2005.07.009

example, persists even when the visual information leads (by up to 240 ms), or lags (by up to 60 ms) the auditory input [9,24]. Similarly, modern technology can also lead to asynchronous audiovisual stimulus presentation as, for example, with satellite TV broadcasts, in which there is often a lag between the auditory and visual signals (cf. [18,29,30]). The ability of the human perceptual system to reconcile small temporal asynchronies suggests a certain flexibility in the underlying mechanisms of multisensory integration.

Here, we investigated the nature of this temporal flexibility by addressing whether exposure to a continuous stream of complex audiovisual stimuli (such as speech or a recording of a hand playing a piano) presented asynchronously can induce a general temporal recalibration between audition and vision. Across three experiments, we analyzed the effects of monitoring asynchronous speech or music on performance in a temporal order judgment (TOJ) task where participants had to judge which of two events, a light flash and a burst of white noise, had been presented first (see [16]).

Several recent studies provide evidence for a temporal equivalent of the well-known spatial ventriloquism effect. In its spatial version, the location of a sound source is illusorily misplaced toward the position of a concurrent visual stimulus (e.g., [17]; see [3] for a recent review). In a recent demonstration of the existence of the temporal analogue of the ventriloquism effect, Morein-Zamir, Soto-Faraco, and Kingstone [23] reported that the perceived onset time of a light can be attracted temporally toward the onset time of a sound that is presented slightly later (see also [1,2,10,28,41]). Moreover, recent studies have also shown that it is possible to induce temporal recalibration after-effects by exposing the observer to a continuous stream of desynchronized audiovisual stimuli (e.g., tones and lights) [12,42]. In the present study, we investigated whether it is possible to demonstrate temporal recalibration using an online adaptation method during exposure to more complex and ecologically valid stimuli, such as a face talking or a hand playing notes on a piano. We measured the transfer of any temporal recalibration effect caused by exposure to desynchronized complex stimuli (speech or music) to the perception of a different kind of stimuli, consisting of a simple flash of light and a burst of white noise (cf. [12]).

We used a videotaped recording of a speaker pronouncing a list of words or a hand playing a piano. In half of the experimental blocks, the auditory signal was delayed relative to the visual signal, whereas in the remainder of the blocks, the auditory and visual stimuli were presented in synchrony (see Fig. 1). While monitoring the speech (or music) stream for targets (male first names or a break in the musical pattern, respectively), participants were asked to judge the order in which a pair of stimuli (a burst of white noise and a briefly flashing LED) was presented (i.e., they performed an audiovisual TOJ task).[1] We predicted that if any adaptation to the asynchronous complex audiovisual stream were to take place, then it might be possible to find a general temporal recalibration of audiovisual processing (cf. [12]), thus leading to an influence in TOJ performance for light flashes and noise bursts.

In particular, the occurrence of adaptation might affect the just noticeable difference (JND), and/or the point of subjective simultaneity (PSS) in the TOJ task. The JND refers the smallest temporal interval between two stimuli needed for participants to be able to judge correctly which one was presented first on 75% of trials. Our prediction was that monitoring the complex asynchronous stimuli (either audiovisual speech or a hand playing a piano) might modify the perceiver's 'online' temporal resolution (i.e., a widening of the audiovisual temporal window for integration), thus participants would require a longer interval in the TOJ task to decide whether the light or the sound came first (i.e., the JND should be larger). The PSS indicates how much time one stimulus has to lead the other in order for the two to be judged as occurring simultaneously (i.e., the average SOA at which participants make each response equally often), and is sensitive to differences in neural processing latencies between auditory and visual stimuli [34]. In the present study, any shift in the PSS would presumably reflect a realignment in the temporal processing of one sensory modality relative to the other, consequent on the brain's ability to adapt to audiovisual asynchrony.

2. Experiment 1

2.1. Materials and methods

2.1.1. Participants

Twelve participants took part in this experiment. All were naive as to the purpose of the experiment and all reported normal hearing and normal or corrected-to-normal vision. All of the participants gave their informed consent prior taking part in the study, and the majority received a £5 (UK Sterling) gift voucher in return for their participation. All of the experiments reported in this study were non-invasive, were conducted in accordance with the Declaration of Helsinki, and had ethical approval from the Department of Experimental Psychology, University of Oxford, UK.

2.1.2. Apparatus and materials

We used an 18-min videotaped recording of a male speaker (consisting of a close-up of the mouth area, from the

[1] Given that temporal recalibration effects are smaller when the adaptor stimuli and test stimuli (used to measure temporal after-effects) are different (see [12]), we decided to present the speech monitoring and TOJ tasks simultaneously (i.e., rather than sequentially as in many studies of post-exposure after-effects; cf. [12,42]).