Brain Research 1661 (2017) 79–87

Contents lists available at ScienceDirect

Brain Research

journal homepage: www.elsevier.com/locate/bres

Research report

Temporal and identity prediction in visual-auditory events: Electrophysiological evidence from stimulus omissions

Thijs van Laarhoven *, Jeroen J. Stekelenburg, Jean Vroomen

Department of Cognitive Neuropsychology, Tilburg University, P.O. Box 90153, 5000 LE Tilburg, The Netherlands

ARTICLE INFO

Article history:
Received 6 October 2016
Received in revised form 13 January 2017
Accepted 13 February 2017
Available online 16 February 2017

Keywords:
Stimulus omission
Predictive coding
Event-related potentials
Visual-auditory

ABSTRACT

A rare omission of a sound that is predictable by anticipatory visual information induces an early negative omission response (oN1) in the EEG during the period of silence where the sound was expected. It was previously suggested that the oN1 was primarily driven by the *identity* of the anticipated sound. Here, we examined the role of *temporal* prediction in conjunction with identity prediction of the anticipated sound in the evocation of the auditory oN1. With incongruent audiovisual stimuli (a video of a handclap that is consistently combined with the sound of a car horn) we demonstrate in Experiment 1 that a natural match in identity between the visual and auditory stimulus is *not* required for inducing the oN1, and that the perceptual system can adapt predictions to unnatural stimulus events. In Experiment 2 we varied either the auditory onset (relative to the visual onset) or the identity of the sound across trials in order to hamper temporal and identity predictions. Relative to the natural stimulus with correct auditory timing and matching audiovisual identity, the oN1 was abolished when either the timing or the identity of the sound could not be predicted reliably from the video. Our study demonstrates the flexibility of the perceptual system in predictive processing (Experiment 1) and also shows that precise predictions of timing and content are both essential elements for inducing an oN1 (Experiment 2).

1. Introduction

One of the main and arguably most basal functions of the human brain is to ‘make sense’ of our environment. Understanding which events in the outside world caused activation of specific sensory systems is what is generally considered to be the essence of perception (Lochmann and Deneve, 2011). This notion is central to the predictive coding theory, in which perceiving is considered a process of inferring the most probable causes explaining sensory signals (Friston, 2005). A key element of predictive coding is the assumption that the brain generates internal templates of the world in higher cortical areas (Mumford, 1992). These templates supposedly contain specific activation patterns of sensory systems that an occurring stimulus would normally elicit. The generated templates are presumed to be sent from higher to lower cortical processing areas (top-down), where they induce a predicted pattern of activation (Friston, 2005). If the bottom-up activation pattern induced by a stimulus matches the prediction, recognition of the stimulus occurs. Any violation of the predicted patterns by the sensory input is sent from lower sensory levels to higher cortical processing areas, reflecting the prediction error (Arnal and Giraud, 2012; Wacongne et al., 2012).

An approach that has been applied recently to explore the neurophysiological mechanisms of sensory prediction relies on the electrophysiological responses to infrequent unexpected stimulus *omissions*. According to the predictive coding framework, early sensory responses reflect the difference between the prediction and sensory input (Friston, 2005; Wacongne et al., 2012). During stimulus omissions there is no sensory input and the neural response to stimulus omissions is thus hypothesized to represent the neural code of top-down prediction devoid of stimulus-evoked sensory processing (Arnal and Giraud, 2012; SanMiguel et al., 2013b). An auditory event can be made predictable either by a motor act or anticipatory visual information regarding the onset and identity of the sound (SanMiguel et al., 2013b; Stekelenburg and Vroomen, 2015). An occasional unexpected omission of the sound evokes an early negative omission response (oN1), likely originating in the auditory cortex, suggesting that both motor and visual predictions are able to activate a sensory template of an expected auditory stimulus in the auditory cortex.

Abbreviations: VA, visual-auditory; MA, motor-auditory; oN1, omission N1; oN2, omission N2; oP3, omission P3.

* Corresponding author.
E-mail address: T.J.T.M.vanLaarhoven@TilburgUniversity.edu (T. van Laarhoven).

http://dx.doi.org/10.1016/j.brainres.2017.02.014

While the available data agree that the oN1 response is an electrophysiological indicator of automatic predictive processing, it is not yet fully understood whether auditory prediction is primarily driven by *temporal* information (timing) or by the *identity* of the anticipated sound. In the motor-auditory (MA) domain, a study of SanMiguel et al. (2013a) suggests that auditory omission responses are primarily driven by *identity* prediction, with only a modulatory effect of temporal prediction. In their study either a single sound or a random sound was presented after a self-paced button press. Prediction-related auditory omission responses were only observed in the single sound condition, suggesting that the sensory system, even with exact foreknowledge of the stimulus onset, does not formulate predictions if the identity of the predicted stimulus cannot be anticipated (SanMiguel et al., 2013a). However, the timing of the sound was not specifically manipulated in their study, which calls upon further investigation of the role of temporal prediction using a stimulus omission paradigm.

The present study investigated the neural mechanisms of temporal and identity auditory predictions in the visual-auditory (VA) domain by using infrequent auditory stimulus omissions. We conducted two separate experiments. In both experiments, we used a video of an actor performing a single handclap (Fig. 1) as a visual stimulus containing anticipatory information about sound identity and sound onset (Stekelenburg and Vroomen, 2007, 2015).

In the first experiment, we examined whether visual-to-auditory predictions (reflected in the omission response) are flexible and adapt, in short-term, to unnatural VA incongruences, or rather depend on long-term established associations. Compared to auditory prediction by a self-generated motor act, prediction of a sound by vision might be more affected by the informational association between the visual and auditory stimulus. While strict informational associations are not necessarily involved in the act of a button press – as a button press can elicit various different sounds in daily practice – a video of a natural visual event may induce relatively strong auditory associations based on lifelong experience. Furthermore, although previous studies have shown that unnatural VA pairings may lead to enhancements in auditory processing (Fort et al., 2002; Giard and Peronnet, 1999; Thorne and Debener, 2008), it is unclear whether auditory omission responses are affected by VA congruency of identity or not. Hence, the first experiment was conducted to examine the influence of VA congruency of identity on prediction-related auditory omission responses. VA congruency was manipulated block-wise in two separate conditions. The video of the handclap was presented synchronously with either the sound of the actual handclap (natural condition) or the sound of a car horn (incongruent condition). The timing of the incongruent sound matched the timing of the natural sound. The sound of a car horn was specifically chosen to obtain a high level of VA incongruence with respect to real-world situations. VA trials were interspersed with unpredictable omissions of the sound in 12% of the trials in both conditions, c.f. SanMiguel et al. (2013a) and Stekelenburg and Vroomen (2015). Based on previous findings (SanMiguel et al., 2013b; Stekelenburg and Vroomen, 2015), three distinct omission ERP components – elicited by rare omissions of the expected sound – were expected for the natural condition: an initial negative deflection at around 50–100 ms after the expected sound onset (oN1), reflecting prediction error, followed by a second negative response at around 200 ms (oN2), and finally a more broadly distributed positive response at 300 ms (oP3), presumably reflecting higher-order error evaluation, attention orienting and subsequent updating of the forward model (Baldi and Itti, 2010; Polich, 2007). A statistically significant difference between the omission responses of the natural and incongruent conditions would suggest that the omission response depends on long-term learned VA associations.

In the second experiment, we examined the separate contributions of temporal and identity information on VA omission responses by randomizing (on a trial-to-trial basis) either auditory onset relative to visual onset or sound identity. Three experimental conditions were included: a *natural* condition, a *random-timing* condition and a *random-identity* condition (Table 1). The natural condition was identical to the natural condition of Experiment 1. In the other two conditions, either the onset (random-timing condition) or the identity (random-identity condition) of the sound was unpredictable. Temporal prediction was disrupted in the random-timing condition by presenting VA stimuli (88% of total number of trials) for which sound and vision were always asynchronous. The magnitude of asynchrony varied on a trial-to-trial basis in order to prevent adaptation to temporal asynchrony (Vroomen et al., 2004). In the random-identity condition the identity of the sound was different for each trial (c.f. the random-sound condition in SanMiguel et al. (2013a)). Based on previous findings in the MA domain, prediction-related neural activity induced by auditory omissions was expected to be most evident in the natural condition (SanMiguel et al., 2013a; Stekelenburg and Vroomen, 2015), and to be diminished in the random-identity condition (SanMiguel et al., 2013a). Assuming that timing of the sound is also of importance in the VA domain (Vroomen and Stekelenburg, 2010), we expected that the omission responses would also be diminished in the random-timing condition.

2. Results

2.1. Experiment 1

Three distinct deflections in the omission ERP were observed for both the natural and incongruent condition (Fig. 2). The first negative component peaked in a time-window of 45–80 ms and is denoted as oN1. A second negative component reached its maximum at 120–240 ms (oN2). The two negative components were followed by a broadly distributed positive deflection in a window of 240–500 ms (oP3). The oN1 deflection showed a bilateral scalp distribution with a right preponderance in both conditions, while the oN2 and oP3 components had a bilateral scalp distribution with no clear preponderance towards either hemisphere (Fig. 3). Based on these scalp distributions, a left fronto-temporal (F7, F5, FT7, FC5) and right temporal (FC6, FT8, C6, T8) ROI were selected for the oN1 time-window. A frontal (F1, Fz, F2) and frontal-central (FC1, FCz, FC2) ROI was selected for the oN2 and oP3 time-window respectively. Mean amplitudes were calculated for

Fig. 1. Time-course of the video used in all experimental conditions administered in Experiment 1 and Experiment 2.