Temporal ventriloquism refers to the phenomenon that a sound presented in close temporal proximity of a visual stimulus attracts its perceived temporal occurrence. Here, we investigate the time-course of the neuronal processes underlying temporal ventriloquism, using event-related brain potentials. To measure shifts in perceived temporal visual occurrence, we used a paradigm in which a sound modulates the magnitude of a visual illusion called the flash–lag effect. A sound presented before the flash reduced both the size of the flash–lag effect and the amplitude of visual N1 compared with when the sound lagged the flash. We attribute the modulation of the flash–lag effect to a modulation of facilitation of visual processing. The time-course (190 ms) and localization (occipitoparietal cortex) of this particular auditory–visual interaction confirms the sensory nature of temporal ventriloquism.

**Key words:** Event-related potentials; Flash–lag effect; Multisensory perception; Temporal ventriloquism

**INTRODUCTION**

It is generally acknowledged that signals from a specific modality can influence the perception of signals from another modality [1]. An example of such a crossmodal interaction is temporal ventriloquism [2–5]. It is the illusion that the perceived temporal occurrence of a visual event is temporally attracted toward a sound when both stimuli are presented with a small temporal discrepancy. Temporal ventriloquism has been demonstrated in various paradigms [2–5]. Vroomen and de Gelder [5], for example, used a phenomenon called the flash–lag effect (FLE) to investigate whether audition can capture temporal visual occurrence. The FLE refers to the phenomenon that when a static flash is projected on a moving object, it appears to lag behind [6]. Vroomen and de Gelder [5] found that the size of the FLE was modulated by sounds that either led or lagged the flash at intervals ranging from −100 to +100 ms (a negative sign refers to the sound before the flash, whereas a positive sign refers to the sound after the flash). A sound before the flash decreased the FLE, whereas a sound following the flash increased the FLE as though the sound attracted the temporal occurrence of the flash.

Available behavioural evidence suggests that temporal ventriloquism reflects a genuine perceptual effect and is not the result of a postperceptual response bias [3–5]. However, the neural mechanisms underlying temporal ventriloquism and its time-course are still unknown. The aim of the present study was to investigate the time-course of temporal ventriloquism using event-related potentials (ERPs). ERPs have already proven to be an appropriate tool for studying the temporal characteristics of auditory–visual interactions because of their excellent timing. Several studies have revealed early [7,8] and late [9,10] crossmodal modulations.

In the current study, we investigated whether the timing and amplitude of neural activity underlying typical visual processes are affected by temporal asynchrony between an auditory and visual stimulus. If temporal ventriloquism is a perceptual phenomenon rather than the result of a response bias, one would expect crossmodal interactions to occur at the early (<200 ms) brain potentials. We therefore examined whether a shift in the perceived occurrence of a visual event is reflected at the electrophysiological level as a shift in the latency of visually evoked potentials such as P1 and N1 or as a difference in the ERP amplitude.

The same FLE paradigm as in Vroomen and de Gelder’s work [5] was used. A centrally presented flash was projected on a horizontally moving bar just before they were physically aligned. A click sound was presented either synchronously with the flash, at 100 ms before the flash, or at 100 ms after the flash. Visual-only and auditory-only conditions were included as ERP baseline. Participants judged whether the flash appeared to the right or left of the moving bar. ERPs evoked by the flash in the asynchronous conditions (sound leading or lagging) were compared with those in the synchronous and visual-only conditions.

**MATERIALS AND METHODS**

**Participants:** Fourteen healthy participants (six women, eight men) with normal hearing and normal or corrected-to-normal vision volunteered to take part in the experiment and gave written informed consent. Their age ranged from 18 to 29 years with mean age of 20.6 years.

**Stimuli and procedure:** The experiment took place in a dark, sound-attenuated and electrically shielded chamber.
Stimuli were presented on a 17-inch monitor positioned at eye-level at a distance of 70 cm from the participant's head. A vertical black bar (3.3° × 1.2°) with a luminance of 6 cd/m² moved from the left to the right over a distance of 12.5° at a constant velocity of 9.3°/s on a grey background (10 cd/m² luminance). A solid white disk (120 cd/m² luminance) with a diameter matching the width of the bar (1.2°) was presented for one refresh cycle (16 ms) at the horizontal centre of the screen, at the level of the vertical middle of the bar (Fig. 1). The disk was always flashed before the bar reached the horizontal centre of the screen, at three stimulus onset asynchronies (SOAs), namely 16.7, 33.4, or 50.1 ms. These three SOAs were chosen on the basis of a pilot study to approximately equate the number of left/right responses. The auditory stimulus was a 70-dB white noise of 16.7-ms duration coming from a loudspeaker located below and in front of the monitor. Three audiovisual asynchronies were used: the sound was presented simultaneously with the flash, or it led or lagged the flash by 100 ms. In the fourth (Visual-only) condition, the flash was not accompanied by a sound. Participants were required to focus on a red fixation cross (+) located at the horizontal middle of the screen, at an interval of −100 (Lead), 0 (Sync), or 100 ms (Lag). Participants judged the position of the flash relative to the bar.

RESULTS

Performance: In 99.6% of the catch trials, participants correctly refrained from giving a response, indicating that they kept their gaze on the fixation cross. To estimate, at the behavioural level, the effect of the sound on the FLE, psychometric functions were computed by fitting a straight line through the data points of the three visual SOAs, separately for each condition (Fig. 2a). The point of subjective equality (PSE; i.e. the position where the flash appears to be on the bar) was derived from the psychometric functions and subjected to a multivariate analysis of variance for repeated measures. The PSE serves as a measure of the magnitude of the FLE. As expected, the PSEs differed significantly between the audiovisual conditions [F(3,11)=6.16, p<0.05] (Fig. 2b). The size of the FLE was lowest in the Lead condition (38.3 ms) and highest in the Lag (41.3 ms) and Visual-only (41.4 ms) conditions. Post-hoc tests revealed a significant difference in PSE between the Lead and the Lag conditions [t(13)=2.77, p<0.05] and between the Lead and Visual-only conditions [t(13)=3.41, p<0.01].

Event-related potentials: Figure 3 depicts the averaged occipital ERPs, timed relative to the onset of the flash. The main question was whether early visual ERP components (P1 and N1) were affected by the auditory–visual temporal asynchrony, but without the contribution of the auditory component as such.
asynchrony. Amplitude and latency of P1 and N1 were scored in the windows of 100–200 and 150–250 ms relative to the prestimulus baseline. P1 peaked at 140 ms and had a central occipitoparietal maximum. Peak N1 had a latency of approximately 190 ms and a bilateral occipitoparietal scalp distribution. Using a multivariate analysis of variance for repeated measures, P1 latency and amplitude were tested with the factors Condition (Lead, Synchronous, Lag, and Visual-only) and Electrode (PO3, POz, PO4, O2, Oz, and O1). P1 latency and amplitude were not significantly affected by experimental manipulations. N1 amplitude and latency were analysed using the factors Condition, Hemisphere (left, right), and Electrode (P5/6, PO3/4, PO7/8, P7/8, O1/2). N1 latency did not significantly differ between conditions, but a main effect of Condition was found for N1 amplitude \(F(3,11) = 5.13, p < 0.05\). Figure 3 shows that N1 amplitude was largest in the Lag and Visual-only conditions and smallest in the Lead condition. No other main effects or interactions were significant. Post-hoc analysis showed that each condition differed significantly from the other (all F values > 5.37), except for Lead versus Synchronous and Lag versus Visual-only. No other significant effects of experimental manipulation on amplitude or latency of ERP components after the N1 at any electrode position were observed. We additionally tested the correspondence between the size of the FLE and the amplitude of the visual N1 component. N1 amplitudes and PSEs were first transformed into z-scores to make the scales comparable. As is clear from Fig. 2b, there was no hint of an interaction \(p > 0.15\), suggesting that the amplitude of the visual N1 component and the size of the FLE were similarly affected by the auditory–visual temporal asynchrony.

**DISCUSSION**

The goal of our study was to investigate the time-course of electrophysiological correlates of temporal ventriloquism. Consistent with the study of Vroomen and de Gelder [5], a flash was perceived as occurring earlier (i.e. a smaller FLE) when a sound was presented before the flash than when the sound appeared after the flash. Here, we showed that the largest FLE (in the Lag and Visual-only conditions) was associated with the highest N1 amplitude, and the smallest FLE (in the Lead condition) was associated with the lowest N1 amplitude (Fig. 2b). To establish whether a functional link exists between the N1 amplitude effects and the modulation of the FLE, we first need to consider the behavioural data. In the current study and in that of Vroomen and de Gelder [5], a sound accompanying the flash (presented simultaneously, leading, or lagging) induced a reduction of the size of the FLE relative to the Visual-only condition. Vroomen and de Gelder argued that a sound combined with the flash speeds up processing of the visual...
stimulus, thereby reducing the magnitude of the FLE. This explanation fits the observation that detection of stimuli containing redundant bimodal information is faster than that of its unimodal inputs (the so-called redundant target effect) [7,8,11,12]. Similarly, here we propose that the FLE is reduced when the flash is presented together with a sound because of enhanced visual processing. Modulation of the size of the reduced FLE is attributed to a difference in the extent to which processing of the flash is enhanced by the sound. Visual facilitation is maximal when a sound, within certain limits, precedes the flash and minimal when a sound is lagging the flash.

At the electrophysiological level, behavioural facilitation is associated with modulation of N1 amplitude. Donchin and Linsley [13] found in a simple reaction time task that N1 was largest for the fastest reaction times, suggesting that enhanced N1 reflects neural facilitation (see also [14,15]). However, when a target stimulus is accompanied by other redundant information, (e.g., a visual target presented with an irrelevant sound), there is not only behavioural facilitation (the redundant target effect) but also a reduction of the ERPs. Faster reaction times in the bimodal condition relative to the unimodal visual condition are thus associated with a decreased N1 amplitude [7,8]. Similar effects are found in the speech domain where seeing lip movements improves speech intelligibility and decreases auditory ERP components [16,17]. The reduced N1 response in the redundant target condition is interpreted as reflecting a lesser energetic demand (neural facilitation) from the visual system for detecting visual stimuli, made more salient by the addition of an auditory accessory stimulus [7]. The fact that the depression of N1 amplitude was strongest when the FLE was most reduced supports the view that the crossmodal effect on the FLE was indeed induced by enhanced visual processing. Thus, both behavioural and electrophysiological data suggest that the FLE is mediated by the extent to which visual processing is facilitated by a task-irrelevant sound.

Whereas the amplitude of visual N1 varied as a function of the auditory–visual temporal asynchrony, the latency of the visual ERP components (P1 and N1) was unaffected. This finding corresponds with data of Regan and Spekreijse [18] who also showed that the timing of the visual occipital ERP was not influenced by auditory–visual asynchrony in a phenomenon where auditory flutter drives visual flicker [19]. So, available data suggest that auditory–visual temporal discrepancies are not resolved by a temporal shift of visual processing. An alternative explanation for the absence of a temporal shift of the ERP components may lie in the small size of the temporal ventriloquist effect. Here, a leading sound made the flash appear earlier by only about 3 ms than a lagging sound. This temporal difference may not be reliably reflected in the ERPs because this order of magnitude reaches the lower limit of the temporal resolution of the sampled EEG.

**CONCLUSION**

Manipulation of the temporal asynchrony between a visual target and a task-irrelevant sound (presented simulta-

ously, leading, or lagging) in the FLE paradigm systematically affected (i.e., decreased) the amplitude of visual N1. Depression of the sensory-specific N1 to bimodal stimulation was explained as reflecting facilitation of visual processing. We therefore interpreted the modulation of N1 decrement as an expression of the extent to which the sound facilitated the processing of the flash. It was maximal when the sound led the flash and minimal when the sound lagged the flash. The latency of this effect (less than 200 ms) and the fact that crossmodal interactions were found in visual cortical areas support the notion that temporal ventriloquism can be regarded as a sensory phenomenon.

**REFERENCES**