

Tilburg University

An event-related potential investigation of the time-course of temporal ventriloquism

Stekelenburg, J.J.; Vroomen, J.

Published in:
Neuroreport

Publication date:
2005

[Link to publication in Tilburg University Research Portal](#)

Citation for published version (APA):
Stekelenburg, J. J., & Vroomen, J. (2005). An event-related potential investigation of the time-course of temporal ventriloquism. *Neuroreport*, 16(6), 641-644.

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

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.

An ERP investigation of the time-course of temporal ventriloquism

Jeroen J. Stekelenburg^{CA} and Jean Vroomen

Psychonomics Laboratory, Tilburg University, P.O. Box 90153, 5000 LE, Tilburg, The
Netherlands

^{CA}Corresponding Author: J.J.Stekelenburg@uvt.nl

Running head: Time-course of temporal ventriloquism

ABSTRACT

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 (FLE). A sound presented before the flash reduced both the size of the FLE and the amplitude of visual N1 compared to when the sound lagged the flash. We attribute the modulation of the FLE to a modulation of facilitation of visual processing. The time-course (190 ms) and localization (occipito-parietal cortex) of this particular auditory-visual interaction confirms the sensory nature of temporal ventriloquism.

Key words: Temporal ventriloquism, Multisensory perception, Flash-lag effect, ERP

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 lead or lagged the flash at intervals ranging from -100 ms 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 if the sound attracted the temporal occurrence of the flash.

Available behavioural evidence suggest 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] as well as 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 [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, or 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 right or left from the moving bar. ERPs evoked by the flash in the asynchronous conditions (sound leading or lagging) were compared to those in the synchronous and visual-only conditions.

MATERIALS AND METHODS

Participants: Fourteen healthy participants (6 females, 8 males) 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 ($3.3^\circ \times 1.2^\circ$) with a luminance of 6 cd/m^2 moved from the left to the right over a distance of 12.5° at a constant velocity of $9.3^\circ/\text{s}$ on a grey background (10 cd/m^2 luminance). A solid white disk (120 cd/m^2 luminance) with a diameter matching the width of the bar (1.2°) was presented for one refresh cycle (16.7 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 ms, 33.4 ms, 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 either simultaneously with the flash, or it lead 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 0.8° from the bottom of the bar and at 2.4° from centre of the flash. The task was to decide whether the flash occurred left or right from the bar, using two designated buttons. The response could only be given after the bar had reached the end of its trajectory, that is, about 700 ms after the flash. The next trial followed at 1 s after the response. The experiment consisted of eight identical blocks. For each combination of condition (Synchronous, Sound Lead, Sound Lag, and Visual-only) and SOA (16.7 ms, 33.4 ms, or 50.1 ms) a total of 96 randomized trials were administered. Catch trials (6.25% of trials) were included to assess whether participants focused on the fixation. In catch trials, the fixation cross changed shape into an 'X' for 332 ms, starting when the bar was 10 refresh cycles from middle of the screen. (Pilot tests had shown that when the gaze was directed at the moving bar instead of the fixation cross, the change in shape remained unnoticed). Participants were instructed to refrain from responding during catch trials. An auditory-only condition was included to control for purely auditory contributions to the audiovisual ERP in the audiovisual conditions. The auditory-only condition consisted of the moving bar and the sound, but without a flash. Auditory-only trials were randomly interspersed between the

experimental conditions. Participants pushed any button after its completion to start the next trial. Prior to the start of the experiment a practice block of 27 trials was given.

[insert Fig. 1 about here]

ERP recording and analysis: EEG was recorded at a sample rate of 512 Hz from 49 locations using active Ag-AgCl electrodes (BioSemi Active 2) mounted in an elastic cap. Data were off-line referenced to an averaged reference and band-pass filtered (1–30 Hz, 24 dB/octave). Horizontal and vertical eye-movements were recorded using electrodes at the outer canthus of each eye as well as above and below the right eye, respectively. The raw data were segmented into epochs of 1000 ms, including a 200-ms prestimulus baseline. After EOG correction, epochs with an amplitude change exceeding $\pm 100 \mu\text{V}$ at any channel were automatically rejected. ERPs were averaged separately for the four (Synchronous, Lead, Lag, and Visual-only) conditions across the three SOAs, resulting in a maximum of 288 trials per condition. Collapsing the ERPs across SOAs was justified, as we found no significant differences between the ERPs of different SOAs. To investigate auditory effects on visual processing, we subtracted the ERP evoked in the auditory-only condition (288 trials) from the audiovisual ERPs (AV-A). In this procedure, the auditory ERP was first aligned in time to the auditory part of the audiovisual ERP. The AV-A difference wave therefore represents the EEG activity evoked by the flash plus the effect of the auditory-visual temporal asynchrony, but without the contribution of the auditory component as such.

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 condition ($t(13)=2.77, p<0.05$) and between the Lead and Visual-only condition ($t(13)=3.41, p<0.01$).

[insert Fig. 2 about here]

ERPs: Fig. 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. Amplitude and latency of P1 and N1 were scored in the window of 100-200 ms and 150-250 ms relative to prestimulus baseline. P1 peaked at 140 ms and had a central occipito-parietal maximum. Peak N1 had a latency of approximately 190 ms and a bilateral occipito-parietal 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 the experimental manipulations. N1 amplitude and latency were analyzed 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 ($F(3,11)=5.13, p<0.05$). Fig. 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 each other (all F values > 5.37), except for Lead vs. Synchronous and Lag vs. Visual-only. There were no other significant effects of experimental manipulation on amplitude or latency of ERP components after the N1 at any electrode position. We additionally tested the correspondence between the size of the FLE and the N1 amplitude modulation across conditions. 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.

[insert Fig. 3 about here]

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 compared to when the sound appeared after the flash. Here, we showed that the largest FLE (in the Lag and Visual-only condition) 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 there is a functional link 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

Gelder [5], a sound accompanying the flash (either 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 with 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 Lindsley [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 event-related potentials. 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 evoked potentials [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 as well as 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 compared to 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 simultaneously, 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 lead 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

1. Calvert G, Spence C, Stein BE. *The handbook of multisensory processes*. MIT Press: Cambridge; 2004.
2. Bertelson P, Aschersleben G. Temporal ventriloquism: crossmodal interaction on the time dimension. 1. Evidence from auditory-visual temporal order judgment. *Int J Psychophysiol* 2003; **50**:147-155.
3. Fendrich R, Corballis PM. The temporal cross-capture of audition and vision. *Percept Psychophys* 2001; **63**:719-725.
4. Morein-Zamir S, Soto-Faraco S, Kingstone A. Auditory capture of vision: examining temporal ventriloquism. *Brain Res Cogn Brain Res* 2003; **17**:154-163.
5. Vroomen J, de Gelder B. Temporal ventriloquism: sound modulates the flash-lag effect. *J Exp Psychol Hum Percept Perform* 2004; **30**:513-518.
6. Nijhawan R. Neural delays, visual motion and the flash-lag effect. *Trends Cogn Sci* 2002; **6**:387.
7. Giard MH, Peronnet F. Auditory-visual integration during multimodal object recognition in humans: a behavioral and electrophysiological study. *J Cogn Neurosci* 1999; **11**:473-490.
8. Molholm S, Ritter W, Murray MM, Javitt DC, Schroeder CE, Foxe JJ. Multisensory auditory-visual interactions during early sensory processing in humans: a high-density electrical mapping study. *Brain Res Cogn Brain Res* 2002; **14**:115-128.
9. Schröger E, Widmann A. Speeded responses to audiovisual signal changes result from bimodal integration. *Psychophysiology* 1998; **35**:755-759.
10. Stekelenburg JJ, Vroomen J, de Gelder B. Illusory sound shifts induced by the ventriloquist illusion evoke the mismatch negativity. *Neurosci Lett* 2004; **357**:163-166.

11. Miller J. Divided attention: evidence for coactivation with redundant signals. *Cognit Psychol* 1982; **14**:247-279.
12. Miller J. Timecourse of coactivation in bimodal divided attention. *Percept Psychophys* 1986; **40**:331-343.
13. Donchin E, Lindsley DB. Average evoked potentials and reaction times to visual stimuli. *Electroencephalogr Clin Neurophysiol* 1966; **20**:217-223.
14. Morrell LK, Morrell F. Evoked potentials and reaction times: a study of intra-individual variability. *Electroencephalogr Clin Neurophysiol* 1966; **20**:567-575.
15. Wilkinson RT, Morlock HC. Auditory evoked response and reaction time. *Electroencephalogr Clin Neurophysiol* 1967; **23**:50-56.
16. Besle J, Fort A, Delpuech C, Giard MH. Bimodal speech: early suppressive visual effects in human auditory cortex. *Eur J Neurosci* 2004; **20**:2225-2234.
17. van Wassenhove V, Grant KW, Poeppel D. Visual speech speeds up the neural processing of auditory speech. *Proc Natl Acad Sci U S A* 2005.
18. Regan D, Spekrijse H. Auditory-visual interactions and the correspondence between perceived auditory space and perceived visual space. *Perception* 1977; **6**:133-138.
19. Shipley T. Auditory Flutter-Driving of Visual Flicker. *Science* 1964; **145**:1328-1330.

Figure captions

Fig. 1. Experimental paradigm. A black bar moved from left to right at a constant speed against a grey background. The flash (a white disk) was projected for one refresh cycle (16 ms) at various timings (16, 33, 50 ms) before the flash and the bar were physically aligned. The flash was either presented alone or could be accompanied by a tone at an interval of -100 (Lead), 0 (Sync), or 100 ms (Lag). Participants judged the position of the flash relative to the bar.

Fig. 2. (a) The proportion of the lag responses as function of the timing of the flash relative to the bar for the Lead, Synchronous (Sync), Lag, and Visual-only (V) conditions. (b) The size of the FLE and the amplitude of the visual occipito-parietal N1 (collapsed over electrodes P5/6, PO3/4, PO7/8, P7/8, O1/2) as function of condition. The left vertical axis shows the size of the FLE estimated by the time (in milliseconds) when the disk had to be flashed relative to the bar so that both were perceived to be at the same location. The right vertical axis shows the amplitude of the N1 in microvolts.

Fig. 3. Grand averaged visual-only ERP (V) and the difference waves ($AV - A$) of the Lead (Ld), Synchronous (S), and Lag (Lg) audiovisual conditions, at a representative electrode (O1).

Fig. 1

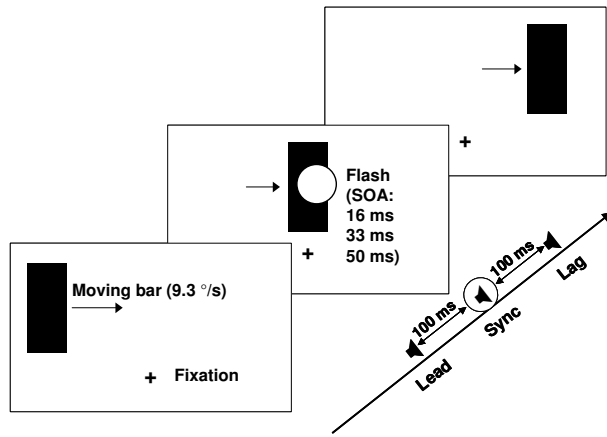


Fig. 2

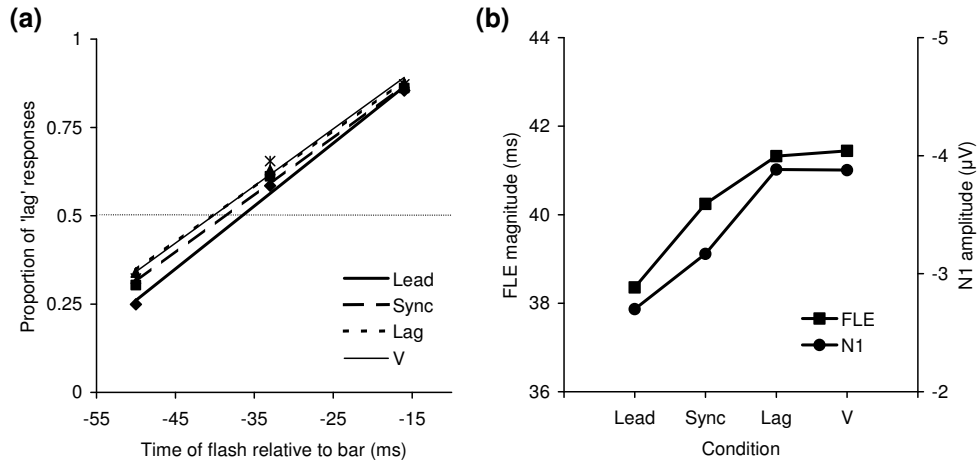


Fig. 3

