Neuroscience 343 (2017) 157–164

VISUAL FORM PREDICTIONS FACILITATE AUDITORY PROCESSING AT THE N1

TIM PARIS,* JEESUN KIM AND CHRIS DAVIS

The MARCS Institute, University of Western Sydney, Sydney, Australia

Abstract—Auditory-visual (AV) events often involve a leading visual cue (e.g. auditory-visual speech) that allows the perceiver to generate predictions about the upcoming auditory event. Electrophysiological evidence suggests that when an auditory event is predicted, processing is sped up, i.e., the N1 component of the ERP occurs earlier (N1 facilitation). However, it is not clear (1) whether N1 facilitation is based specifically on predictive rather than multisensory integration and (2) which particular properties of the visual cue it is based on. The current experiment used artificial AV stimuli in which visual cues predicted but did not co-occur with auditory cues. Visual form cues (high and low salience) and the auditory-visual pairing were manipulated so that auditory predictions could be based on form and timing or on timing only. The results showed that N1 facilitation occurred only for combined form and temporal predictions. These results suggest that faster auditory processing (as indicated by N1 facilitation) is based on predictive processing generated by a visual cue that clearly predicts both what and when the auditory stimulus will occur. © **2016 Published by Elsevier Ltd on behalf of IBRO.**

Key words: audiovisual, prediction, N1 latency, EEG.

INTRODUCTION

In many ecological settings, multisensory signals that indicate the presence and identity of an object/event occur at similar times and provide redundant and sometimes complementary information. Research investigating the sensory, perceptual and cognitive processing of these multisensory cues has largely focused on the combination of information, so what has often been overlooked is the role that the temporal order of these cues has on processing. That is, it is commonplace with multisensory signals for a cue from one modality to precede the other, for example, in the case of auditory-visual (AV) speech, the movements of the lips and jaw often begin before the auditory signal is produced. Recent research indicates that prior visual information can be used to generate a prediction about the upcoming sound such as what (Kim and Davis, in press; van Wassenhove et al., 2005), when (Vroomen and Stekelenburg, 2010) and where (Stekelenburg and Vroomen, 2012) it will occur, and this information results in changes to subsequent auditory processing.

In this regard, one particularly interesting suggestion has been that predictions derived from the visual modality can speed up auditory processing (van Wassenhove et al., 2005; Paris et al., 2013). This suggestion was based on the finding that the N100 (N1) ERP in response to auditory speech occurs earlier when preceded by visual speech (Arnal et al., 2009; van Wassenhove et al., 2005). Although it was known that auditory stimulus features (e.g., intensity, Jacobson et al., 1992) and auditory expectations (Budd and Michie, 1994) influence N1 latency, what was intriguing about this 'N1 facilitation effect' was that it demonstrated that predictions generated in one modality (visual) could facilitate processing in another (auditory). Further, it was suggested that the visually induced shift in auditory N1 latency is caused by AV interactions in sensory cortices (Besle et al., 2004; van Wassenhove et al., 2005; Arnal et al., 2009).

The AV interactions that give rise to the N1 facilitation effect have been suggested to reflect predictive processing and not that of multisensory integration per se. The argument for this is that the onset latency of the N1 is affected by the characteristics of the visual signal that begins prior to the acoustic event and not by the relationship between auditory and visual cues that is likely determined during and/or after the auditory event. For example, Arnal and colleagues (2009) have demonstrated that more salient visual speech (i.e., well-marked, distinct movements of the lips and jaw, associated with the articulation of a particular sound, e.g., 'pa') resulted in an earlier N1 latency compared to less salient visual speech (e.g., 'ga') and the validity of the prediction did not influence the amount of N1 facilitation. That is, an invalid prediction (such as in the lip movements of 'pa' paired with the sound of 'ga') resulted in the same latency facilitation as a valid prediction (visual 'pa' paired with auditory 'pa').

Although these results have been interpreted as indicating that N1 facilitation is due to visually based prediction, one feature of these experiments potentially undermines this interpretation. Typically, experiments have used AV signals that overlap in time, i.e.,

*Corresponding author. Address: The MARCS Institute, University of Western Sydney, Bullecourt Avenue, Milperra 2214, NSW, Australia.
E-mail address: t.paris@uws.edu.au (T. Paris).

Abbreviations: AV, auditory-visual; RT, response times; VO, visual only.

http://dx.doi.org/10.1016/j.neuroscience.2016.09.023
0306-4522/

participants saw visual speech that occurred both prior to and during auditory speech presentation. Since the AV signals overlapped, it cannot be unequivocally concluded that N1 facilitation occurs purely due to visually based predictions.

Further, even if N1 facilitation is due to prediction based on the leading signal, other interpretative issues remain. First, it is unclear whether prediction per se is sufficient to produce N1 facilitation or whether the pairing of the AV signals first needs to be over-learned. That is, N1 facilitation has almost exclusively been demonstrated with AV stimuli that have been extensively learnt, such as AV speech stimuli or with other ecological stimuli such as clapping hands (Stekelenburg and Vroomen, 2007). It therefore is not clear whether N1 facilitation requires pre-established AV pairings or whether it would occur whenever the presentation of one stimulus reliably predicts the occurrence of another. The former proposal is consistent with the suggestion that N1 facilitation occurs for speech stimuli because lip movements and speech sounds have a well-learned and tight mapping that allow predictions to be established (van Wassenhove et al., 2005).

The other interpretive issue is about which features of the visual prediction are important in triggering N1 facilitation effects. Previous research has demonstrated how either form or temporal predictions can influence early neural responses using non-speech stimuli. Firstly in terms of form predictions, a number of studies used simple AV associations to investigate the effect of prior visual cues on auditory evoked responses (Widmann et al., 2004; Laine et al., 2007; Lindström et al., 2012). For example, in the study by Widmann and colleagues, participants were presented with horizontal bars of varying heights that were associated with the pitch of subsequent tones. The results of this study showed that valid visual form cues reduced the amplitude of the N1 relative to invalid ones, yet importantly there was no evidence of a latency shift in N1 (see also Laine et al., 2007; Lindström et al., 2012).

In terms of temporal prediction, a study by Vroomen and Stekelenburg (2010) presented participants with moving disks that collided with a central rectangle at which time a sound was produced. When these disks reliably preceded the sound (i.e., the disks provided a prediction as to when the sound would occur) a small N1 facilitation effect occurred. However, this effect was not robust as a second experiment using the same methods did not replicate this finding. Further, it is unclear whether the observed N1 facilitation effect in the study was due to temporal prediction alone as only a single tone was used in this experiment, thus also rendering the form predictable. Given this, whether temporal predictions alone can induce N1 facilitation is yet to be confirmed. Taken together, the above studies suggest that N1 facilitation does not occur for form-only predictions and it is unclear whether facilitation occurs for temporal-only predictions. In light of this it should be noted that ecological stimuli that do show facilitation effects contain both form and temporal cues (Paris et al., 2013).

The current study will address the above issues using non-ecological (artificial) AV stimuli (expanding shapes and tones, see below) that allow complete control of stimulus parameters. We designed visual stimuli that contained key features of AV speech, i.e., the form of the stimuli evolved over time to enable the prediction of the onset and type of the upcoming sound, but unlike speech the visual and auditory components did not overlap. In addition, we also controlled other features such as the time-course and salience of the visual prediction. In this way we could test three main questions: First, whether dynamic visual form predictions that occur prior to the sound would facilitate N1 latency; second, whether this effect would be moderated by stimulus form salience; and third, whether timing information only would facilitate responses.

EXPERIMENTAL PROCEDURES

Participants

Seventeen female participants from the University of Western Sydney took part in the experiment. Their age ranged from 17 to 40 with a mean age of 25 years. All participants reported having normal or corrected-to-normal vision and hearing and were right-handed. The study was conducted with approval of the ethics committee of the University of Western Sydney.

Experimental design and stimuli

The functional properties of the audiovisual stimuli were designed to be similar to those of ecological stimuli (i.e., a brief dynamic visual stimulus that changed to signal the onset of a specific sound). To satisfy these requirements, three types of videos were created that each consisted of a different expanding visual shape that preceded an auditory event. Each type began with a small fixation circle (shown centered in a square), which lasted for a variable window of 300 to 1100 ms. The shape then expanded into one of three possible shapes (sharp, round and rounded diamond, see Fig. 1B) occurred for 500 ms, whereupon the shape made contact with the edge of the square and disappeared as a high (1000 Hz) or low tone (333 Hz) played (for a duration of 100 ms and a rise/fall-time of 10 ms). An example of the time-course of an AV trial is shown in Fig. 1.

In the study there were two unimodal (AO and VO) and three AV conditions. The AO condition consisted of a display of a static fixation circle followed by one of the two tones and the VO condition consisted of a video with no sound. The AO condition acted as the ‘unpredicted’ stimuli as no visual cue preceded the sound. The AV condition consisted of two types of form and temporal prediction cues: (AVvalid) in which visual form and timing cues validly predicted a tone or (AVinvalid) where the form and timing cues provided an invalid prediction. The third AV condition consisted of a visual cue that only provided reliable temporal information (AVtemp). In order to create predictions

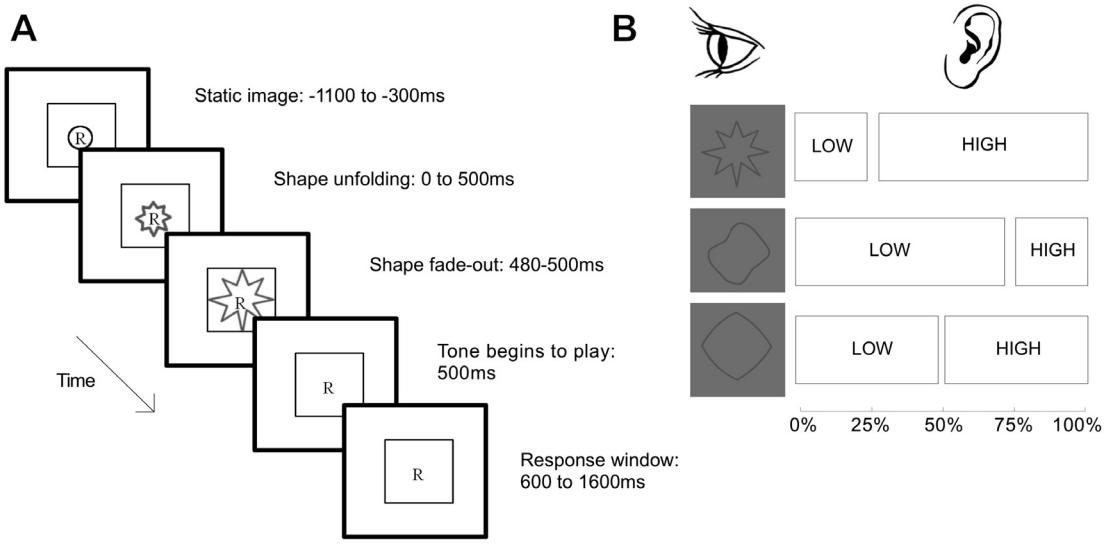

Fig. 1. Experimental design. (A) Time-course of a single AV trial with the sharp shape. In this example, 'R' denotes that a response is required (note, the stimulus shapes were displayed on a gray background, see B). (B) AV stimulus pairings (sharp shape: top; round shape: center and rounded-diamond shape). Rectangle width represents the proportion of times the high or low tone occurred after the corresponding video, e.g., the high tone followed the sharp shape in 75% of AV trials.

from the visual stimuli, we manipulated the proportion of times specific shapes preceded specific sounds (Fig. 1B). In the AV form prediction conditions, the sharp and round shapes were used. In trials with these shapes, 75% were followed by a matching tone (AVvalid; i.e., a sharp shape followed by high tone and round shape followed by low tone) and 25% were followed by a mismatched tone (AVinvalid; i.e., sharp shape followed by low tone and round shape followed by high tone). In the temporal prediction (AVtemp) condition, the diamond shape was presented and both high and low tones were equally likely to follow. In summary, the diamond shape consistently predicted only the temporal onset of the sound, whereas the sharp and round shapes also predicted which sound was likely to occur.

Two aspects of the method are worth emphasizing. First, to increase the likelihood that participants learnt the mapping between specific shapes and sounds, we chose to use items that had a known 'cross modal correspondence' (Spence, 2011). That is, the tendency to associate features of cross-modal stimuli such as sharper shapes with higher pitch (Marks, 1987; Walker et al., 2010). Although this general association may have been present prior to the experiment, the specific characteristics of this association, such as the dynamic form (expanding shape) and timing (colliding with the square) of the stimuli were novel and had to be acquired during the experiment.

Second, the type of shape determined level of salience: The sharp shape had more corners and because both shapes developed from an initial circle, the sharp shape was more apparent earlier than the round or the diamond shape. Furthermore, the round and the diamond shapes were more similar to each other than to the sharp one. For these reasons, the sharp shape was used as the high salience condition. The round shape was used as the low salience condition and the diamond shape was for the temporal only condition.

An initial behavioral study using a speeded identification task was conducted to verify that the round and sharp shapes differed in salience. The experiment consisted of a visual identification task using the three shapes used in the main study. Stimulus videos were cut to different lengths (gates) each 40 ms between 40 and 400 ms (i.e., 10 gating conditions). To determine at what point reliable visual information was available for each shape, 13 participants (6 female, mean age = 26.3) were asked to indicate which shape they saw with a button press based on three response options (sharp, round or diamond). Each shape was presented 14 times in each gating condition (420 trials).

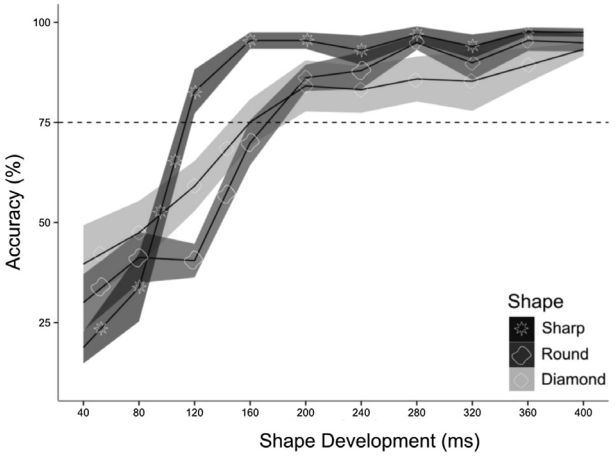

Fig. 2. Mean response accuracy in each gating condition (from 40 to 400 ms) for the sharp, round and diamond shape videos. Error ribbons represent standard error of the mean. The 75% threshold is indicated by the dotted line.

To estimate an exposure for which discrimination could clearly be made, the time point at which identification accuracy passed 75% was used (see Fig. 2). The mean exposure time at this level of accuracy for the sharp shape (114 ms) was significantly less than both the diamond shape (161 ms), $t(12) = 2.50$, $p < 0.05$ and the round shape (173 ms), $t(12) = 3.74$, $p < 0.05$. Responses to the round shape did not differ significantly from the diamond shape, $t(12) = 0.23$, $p > 0.05$. The similar levels of performance for the round and diamond shapes during shape development indicate that these shapes were often confused with each other.

Procedure

The experiment took place in an electrically shielded room. Visual stimuli were presented on a 1024 × 768 resolution 51 cm CRT monitor positioned 1 m from the participant so that the videos subtended a visual angle of 3.15 degrees. Sounds were presented through transducer earphones.

The experiment consisted of a single EEG session interspersed with blocks of behavioral trials. Behavioral response time and error data were collected in addition to the EEG data in order to gauge whether participants had learnt the visual predictions, i.e., faster and more accurate responses would reflect learning of the cross-modal predictions. There were a total number of 1000 trials in the experiment (700 EEG trials and 300 behavioral trials). EEG trials consisted of 100 AO, 100 VO, 100 AVtemp 100 AVinvalid and 300 AVvalid. Behavioral trials consisted of 50 AO, 50 AVtemp, 50 AVinvalid and 150 AVvalid.

Within the EEG session, blocks consisting of behavioral trials occurred at random times and ranged from 15 to 25 trials each in length (rectangular distribution). This design was chosen to keep participants engaged throughout the experiment, as participants did not know whether the next trial would require a response. Participants were made aware of whether a trial was an EEG or a behavioral trial by a letter displayed at fixation. For behavioral (Response trials) an 'R' was shown and for EEG trials (No Response) an 'N' was shown. In behavioral trials participants were instructed to complete a speeded classification task by pressing one of two buttons (left or right) corresponding to the high or low tone as quickly and as accurately as possible. Response mapping was counter-balanced across blocks. In EEG trials participants were instructed to attend to the stimuli, however no response was required. EEG responses to behavioral trials were not analyzed due to artifacts that arise from movement and motor preparation (Luck, 2005).

Before the start of the experiment 50 practice trials familiarized participants with the task and stimuli. Following this participants were asked to explicitly state the pairing between the shapes and sound, which all participants were able to do.

EEG recording & analysis

The EEG was recorded using 64 active electrodes (Biosemi, Amsterdam, The Netherlands) positioned according to the extended standard 10–10 (Oostenveld and Praamstra, 2001) system at a sampling rate of 512 Hz. An additional eight electrodes were used: Two electrodes on the mastoids, four ocular electrodes to detect eye blinks (horizontal and vertical EOG) and two electrodes served as reference (CMS/DRL). Before cap placement, participants brushed their hair to improve the conductance between scalp and electrodes (Mahajan and McArthur, 2010). The EEG was referenced off-line to the mastoids and band-pass filtered (0.5–30 Hz, 12 dB octave). Large artifacts were removed from the data prior to ICA decomposition and eye-blink stereotyped components were removed. Pre-processing and data analysis was performed using EEGLAB version 10 (Delorme and Makeig, 2004) and custom Matlab scripts (The Mathworks, Natick, MA, USA).

The data were segmented into epochs of 1000 ms (700 ms before and 300 ms after the auditory onset), including a baseline period −100 to 0 ms. Any epochs with amplitude exceeding +/− 100 uV at any channel were rejected. Across all conditions, over 90% of trials were included after this rejection process. AV-VO conditions were created so that comparisons between AO and AV could be made. ERP peak latency and mean amplitude was identified for each participant and each condition between the window 70–150 ms (N1) and 120–250 ms (P2). This window was chosen as it has shown the largest and most robust peaks and has been used previously for latency and amplitude measurements using AV predictions (Stekelenburg and Vroomen, 2007; Paris et al., 2016). 'N1 suppression' was defined as the difference in N1 mean amplitude between two conditions and 'N1 facilitation' was defined as the difference in N1 peak latency values between two conditions'.

RESULTS

Behavioral analysis

For behavioral trials, an analysis of response times (RT) and error rates was conducted. For the RT analysis, incorrect responses were removed. Mean response times for the different conditions are plotted in Fig. 3.

Overall, participants responded within half a second after the auditory event. RTs in the AO condition were significantly slower than in the average of all AV conditions, $t(16) = 10.51$, $p < 0.05$. Responses to the AVvalid items (332 ms) were significantly faster than AVtemp items (355 ms), $t(16) = 3.62$, $p < 0.05$ and AVinvalid items (382 ms), $t(16) = 4.57$, $p < 0.05$). AVinvalid responses were also significantly slower than AVtemp items, $t(16) = 3.73$, $p < 0.05$.

An ANOVA was conducted with Salience (low and high) and Validity (valid and invalid) conditions. There was a main effect of Salience (faster responses to the high salience items) and Validity (faster responses to valid items), $F(1,16) = 6.18$, $p < 0.05$ and $F(1,16) = 21.77$, $p < 0.05$). There was also a significant interaction, $F(1,16) = 22.75$, $p < 0.05$ such that the difference in RT's between AVvalid and AVinvalid was greater in the high salience condition. *T*-tests were

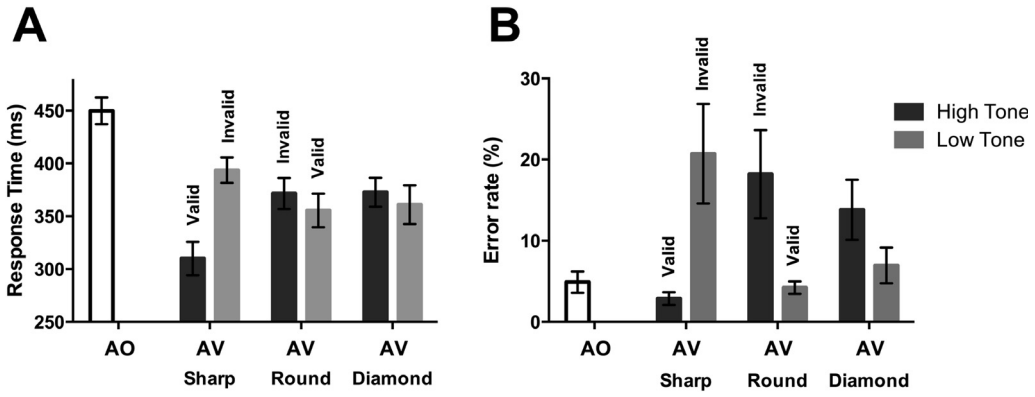

Fig. 3. Graph of mean tone classification for (A). Response time (ms) and (B). Error rates (percent) for AO, AVvalid/AVinvalid (Sharp, Round shape) and AVtemp (Diamond shape) conditions. Error bars represent the standard error of the mean.

conducted to compare the two validity conditions for high and low salience items. Responses to high salience items were significantly faster in the valid than in the invalid condition, $t(16) = 6.69$, $p < 0.05$, however there was no such difference for low salience items, $t(16) = 1.22$, $p > 0.05$ or the AVtemp items, $t(16) = 0.72$, $p > 0.05$.

The accuracy data showed that participants were on average 90.4% correct in classifying the correct (high or low) tone. There were significantly more correct responses to AVvalid items (96.5%) than AVinvalid items (80.6%), $t(16) = 2.94$, $p < 0.05$. This validity effect occurred for both high ($t(16) = 3.06$, $p < 0.05$) and low salience items, $t(16) = 2.83$, $p < 0.05$. Responses to AVvalid items were also more accurate than to the AVtemp ones (90.6%), $t(16) = 2.52$, $p < 0.05$. Responses to AVvalid items did not significantly differ from no prediction (AO) items (98.1%), $t(16) = 1.07$, $p > 0.05$

ERP analysis

For each participant, visual only (VO) scores were subtracted from AV scores to remove the contribution of the visual cue from the ERP. The following factors were used in the analysis; Electrode (Fz, Cz, Pz, C3, C4), Modality (AV, AO), Validity (AVvalid and AVinvalid) and Shape (Sharp, Round, Diamond).

To determine whether N1 latency responses did not significantly differ between valid and invalid trials (as in previous studies), a Modality × Validity × Electrode ANOVA was conducted. There was no main effect of Validity (mean = 96.5 ms; 97.2 ms), $F(1,16) = 0.15$, $p > 0.05$. Based on this result and previous studies (Arnal et al., 2009), levels of Validity were collapsed for subsequent analyses.

N1 latency was submitted to a Modality × Shape × Electrode ANOVA. There was a significant main effect of Modality, such that N1 latency occurred significantly earlier for AV relative to AO trials, $F(1,16) = 5.167$, $p < 0.05$. In addition there was a significant Modality × Shape interaction, $F(2,32) = 5.60$, $p < 0.05$. Neither of these effects were moderated by Electrode ($F(4,64) = 0.96$, $p > 0.05$, $F(8,128) = 1.21$, $p > 0.05$).

Due to the known differences in N1 latency for tones of different frequency (Jacobson et al., 1992), a Modality × Tone × Electrode ANOVA was conducted to test whether the effect of Modality was moderated by Tone frequency. The results again showed a significant effect of Modality, $F(1,16) = 5.96$, $p < 0.05$, as well as a significant effect of Tone, such that N1 latency occurred significantly earlier for high relative to low tones (mean = 97.7 ms; 104.4 ms, $F(1,16) = 45.30$, $p < 0.05$. In addition, there was no significant interaction between these effects, $F(1,16) = 0.64$, $p > 0.05$.

To further investigate the Shapes associated with facilitation, Modality × Electrode ANOVAs were conducted for each of the three Shape conditions (Sharp, Round and Diamond). N1 latency was shorter for the sharp shape relative to AO, $F(1,16) = 4.12$, $p < 0.05$ but this was not the case for either of the other shapes ($F(1,16) = 0.25$, $p > 0.05$, $F(1,16) = 0.89$, $p > 0.05$, corrected for multiple comparisons). The largest reduction in N1 latency occurred at electrode Cz (10.68 ms), with the average N1 latency in the AO condition occurring at 104.1 ms after sound onset. The size of the N1 facilitation effect for each AV condition and for the sharp and round shapes is shown in Fig. 5.

For N1 amplitude, the Modality × Validity × Electrode ANOVA showed no main effect of Validity, $F(1,16) = 0.34$, $p > 0.05$. As in the latency analysis, levels of Validity were collapsed for subsequent analyses. The Modality × Shape × Electrode ANOVA showed a significant main effect of Modality, such that mean N1 amplitude was significantly lower for AV relative to the AO responses, $F(1,16) = 10.791$, $p < 0.05$, but this did not significantly differ between Shapes, $F(2,32) = 0.113$, $p > 0.05$ or between Electrodes, $F(4,64) = 1.41$, $p > 0.05$ (see Fig. 5).

Each latency and amplitude analysis was also conducted on the P2 time window, however no significant difference in latency or amplitude was observed for any of the previous comparisons (Fig. 4).

DISCUSSION

This study used newly associated AV stimuli to examine whether auditory N1 facilitation would occur with AV

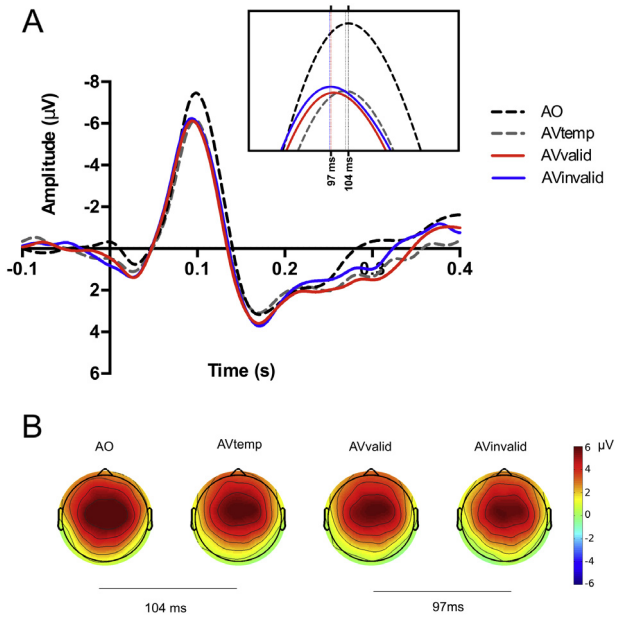

Fig. 4. (A) ERPs locked to auditory onset at electrode Cz for Auditory Only (no visual prediction), AVtemp (temporal only prediction), and AVvalid and AVinvalid (form and temporal predictions). The zoomed in plot highlights the extent of N1 facilitation and the vertical lines mark the average peak N1 latency in each condition. (B) Topographic maps at indicated peak N1 latencies for each condition.

presentations when there was no overlap between the visual and auditory signals (i.e., the visual signal predicted auditory one), and for different types of prediction (form-and-temporal and temporal-only). In addition, the study manipulated the salience of the visual form to test whether this factor would modulate N1 facilitation as reported in previous studies using AV speech (van Wassenhove et al., 2005; Arnal et al., 2009).

The behavioral data were collected to ascertain whether the experimentally defined association between auditory and visual signals had been established (i.e., that the different visual stimuli were seen to predict the high and low tones) showed that response times were significantly faster and more accurate to correctly predicted tones (AVvalid) compared to incorrectly predicted ones (AVinvalid) or to when only the onset of the tone was reliably predicted (AVtemp). In addition, response times in the validly cued condition were faster for the salient sharp shape than for the low salient round one. This shows that participants had formed the expected AV associations, and further used the salient cue to predict the upcoming tone.

For the ERP data, results showed that for the high salient visual cue (the sharp shape) N1 facilitation occurred for both valid and invalid prediction conditions (Arnal et al., 2009). No N1 facilitation was found for the low salient prediction cue (the round shape) even though it was associated with the low tone to the same extent as was the salient one with the high tone (75% valid prediction in both cases). No facilitation effect was found for the AVtemp condition.

Before discussing what the details of these findings reveal about N1 facilitation, two broad results are noteworthy. First, an N1 facilitation effect occurred even though the visual and auditory signals did not overlap. Second, this effect was produced from a newly learned association between a visual stimulus (the expanding sharp shape) and an auditory one (the high tone). These findings are consistent with the proposal that N1 facilitation occurs due to a prediction triggered by a preceding stimulus (Arnal et al., 2009). Further, the

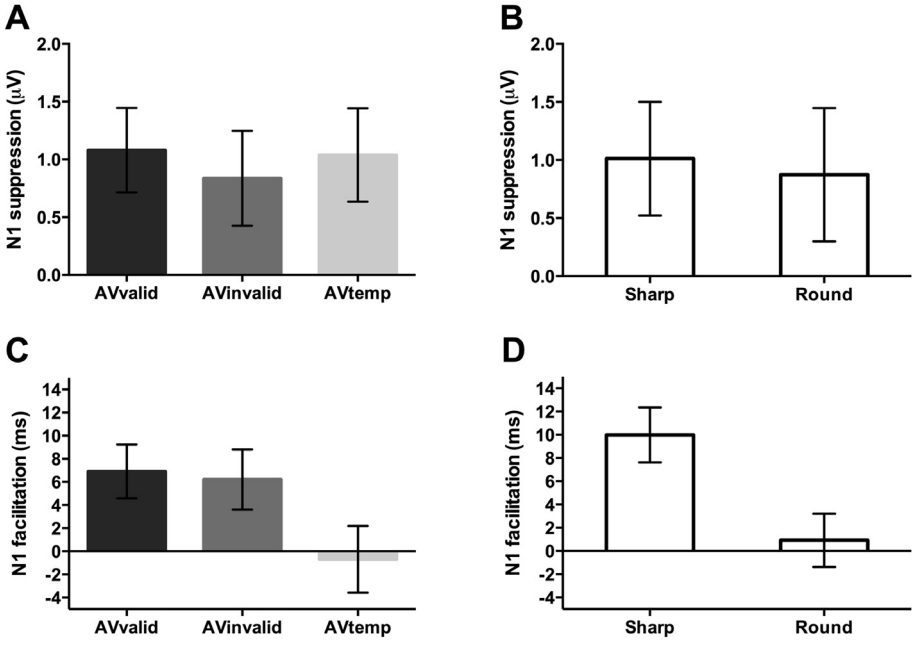

Fig. 5. Mean N1 suppression and peak latency (AV-AO) at electrode Cz for: AVvalid, AVinvalid and AVtemp conditions (A, C) and for different levels of salience (B, D). Positive values reflect reduced amplitudes and shorter latencies. Error bars represent the standard error of the mean.

results show that N1 facilitation is not specific to well learnt ecological stimuli.

In terms of the pattern of N1 facilitation, an important finding was that it occurred only with the salient visual cue. This is an important finding as it provides an insight into what specific characteristics of the predictive stimuli give rise to facilitation. As mentioned above, since the specific AV associations were experimentally defined, the N1 facilitation effect would likely have developed over the course of the experiment and since both the salient and less salient visual cues were paired with their respective tones the same number of times, N1 facilitation cannot simply be due to the frequency that AV stimulus types were presented together. Rather, the likely difference between these cues was the distinctiveness of the shape's visual features (such as number of points and sharpness) that differentiated it from the other cue types. The results of the initial behavioral study indicated that these visual features were perceived early in the unfolding of the salient visual cue. This could explain why the salient cue facilitated classification of its associated tone more than the low salient round shape.

Based on these results, we propose that the distinctiveness of the salient cue allowed participants to make (mostly) correct predictions about the upcoming auditory tone and to do so well before its presentation, and it was this that produced N1 facilitation. On the other hand, although participants would also learn that the round shape predicts the low tone (in 75% of trials), no facilitation effect was observed. This is likely because this shape was confused with the diamond shape, which did not reliably predict this tone, and therefore the round shape would only provide a clear predictive cue once participants could unambiguously identify it. Thus, we suggest that being able to make an effective prediction at an early time point (i.e., well before the presentation of the auditory stimulus) is important for N1 facilitation since this prediction must have time to affect auditory processing.

In regard to the validity of visual predictions, previous research has shown that the matching of the visual prediction with the subsequent sound does not influence the degree of N1 facilitation (Arnal et al., 2009). The current results are consistent with this finding as both valid and invalid predictions facilitated N1 responses. That is, when it was learned that a reasonably reliable prediction could be made using the sharp shape, N1 facilitation occurred when this cue was presented regardless of whether the prediction turned out to be correct. This indicates that N1 facilitation occurred at a level of acoustic processing prior to tone identification.

The current results also showed that a temporal cue alone was not sufficient to facilitate N1 responses. In the temporal-only condition the diamond shape provided a clear cue about when a tone would be presented but an unreliable one about which tone it would be (there was a 50% chance that either the high or low tone would follow). This result appears at odds with that of Vroomen and Stekelenburg (2010) where a 'temporal' cue to auditory onset produced a small N1 facilitation effect in one of two experiments. A possible explanation for the different outcomes is that in the Vroomen and Stekelenburg study only a single sound was used, thus the temporal cue was 100% reliable in predicting the sound type. In other words, the 'temporal' cue also specified the form of the upcoming sound, so it was not purely a temporal cue.

In addition to providing evidence concerning the basis of N1 facilitation, the current study also sheds light on how the salience of a cue might best be characterized. In a previous study by Arnal and colleagues (2009), the visual speech cue 'ga' that was regarded as having low salience still produced a small N1 facilitation effect. This contrasts with the null effect found for the 'low salient' round shape in the current study. One reason for this apparent discrepancy is that salience may be a relative rather than an absolute measure. That is, in previous studies salience has been defined as the amount of clear, unambiguous movement (e.g., van Wassenhove et al., 2005), yet the degree of ambiguity of a particular stimulus also depends on the nature of the stimulus items that it is presented with (i.e., its context, see Paris et al., 2013). In other words, prediction salience should not be defined by properties of the predictive cue in isolation but also with respect to the characteristics of the other presented stimuli and the confusion between stimuli (e.g., in the current study, the round stimulus had similar features to the diamond one, especially early in the presentation time-course).

The similarity between the current results (e.g., N1 facilitation depends on salience but not on validity) and those that have used ecological stimuli, suggests that, given the right stimulus properties and experimental manipulation, N1 facilitation could be produced with any predictive stimulus. Given the limited experience participants had with the current AV associations, it follows that such facilitation relies on the learning of visual to auditory predictions. This suggests that ecological stimuli used in previous studies represent a special case of such prediction whereby learning has occurred over the lifespan. In regard to the current stimuli configurations, what should be noted is that although the specific characteristics of the stimuli had to be acquired within the experiment (e.g., learning the timing between the auditory and visual cues as well as the form of the dynamically expanding shape), the general association between the sharp shape and high tone may have had some degree of prior association (Marks, 1987; Walker et al., 2010). This 'preferred' association may have had the effect of reducing the length of time it took for the N1 facilitation effect to manifest. Nevertheless the use of non-ecological visual and auditory cues used in this experiment indicates that N1 facilitation is not specific to well learnt ecological associations.

In sum, we found that visual predictions influenced N1 latency with auditory and visual cues that did not overlap. We argue that N1 facilitation depends on the visual cue providing an early, reliable and salient prediction to the identity of the upcoming sound. This study extends previous findings by demonstrating that N1 facilitation effects are not specific to well-learnt ecological

predictions but depend on the specific configuration of any cross modal stimuli.

COMPLIANCE WITH ETHICAL STANDARDS

T Paris, J Kim and C Davis declare that they have no conflict of interest.

All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Informed consent was obtained from all individual participants included in the study.

REFERENCES

Arnal LH, Morillon B, Kell CA, Giraud AL (2009) Dual neural routing of visual facilitation in speech processing. J Neurosci 29 (43):13445–13453.

Besle J, Fort A, Delpuech C, Giard M-H (2004) Bimodal speech: early suppressive visual effects in human auditory cortex. Eur J Neurosci 20:2225–2234.

Budd TW, Michie PT (1994) Facilitation of the N1 peak of the auditory ERP at short stimulus intervals. Neuroreport 5(18):2513–2516.

Kim, J., Davis, C. (in press). How visual timing and form information affect speech and non-speech processing. Brain and Language.

Delorme A, Makeig S (2004) EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. J Neurosci Methods 134(1):9–21.

Jacobson GP, Lombardi DM, Gibbens ND, Ahmad BK, Newman CW (1992) The effects of stimulus frequency and recording site on the amplitude and latency of multichannel cortical auditory evoked potential (CAEP) component N1. Ear Hear 13(5):300–306.

Laine M, Kwon MS, Hämäläinen H (2007) Automatic auditory change detection in humans is influenced by visual-auditory associative learning. Neuroreport 18(16):1697–1701.

Lindström R, Paavilainen P, Kujala T, Tervaniemi M (2012) Processing of audiovisual associations in the human brain: dependency on expectations and rule complexity. Front Psychol 3.

Luck S (2005) An introduction to the event-related potential technique. Cambridge, Mass: MIT Press.

Mahajan Y, McArthur G (2010) Does combing the scalp reduce scalp electrode impedances? J Neurosci Methods 188(2):287–289.

Marks LE (1987) On cross-modal similarity: Auditory–visual interactions in speeded discrimination. J Exp Psychol Hum Percept Perform 13(3):384.

Oostenveld R, Praamstra P (2001) The five percent electrode system for high-resolution EEG and ERP measurements. Clin Neurophysiol 112(4):713–719.

Paris T, Kim J, Davis C (2013) Visual speech form influences the speed of auditory speech processing. Brain Lang 126 (3):350–356.

Paris T, Kim J, Davis C (2016) Using EEG and stimulus context to probe the modelling of auditory-visual speech. Cortex.

Spence C (2011) Crossmodal correspondences: A tutorial review. Attention Percept Psychophysics 73(4):971–995.

Stekelenburg JJ, Vroomen J (2007) Neural correlates of multisensory integration of ecologically valid audiovisual events. J Cognit Neurosci 19:1964–1973.

Stekelenburg JJ, Vroomen J (2012) Electrophysiological correlates of predictive coding of auditory location in the perception of natural audiovisual events. Front Integr Neurosci 6.

Van Wassenhove V, Grant KW, Poeppel D (2005) Visual speech speeds up the neural processing of auditory speech. Proc Natl Acad Sci USA 102:1181–1186.

Vroomen J, Stekelenburg JJ (2010) Visual anticipatory information modulates multisensory interactions of newly established audiovisual stimuli. J Cognit Neurosci 22(7):1583–1596.

Walker P, Bremner JG, Mason U, Spring J, Mattock K, Slater A, Johnson SP (2010) Preverbal infants' sensitivity to synaesthetic cross-modality correspondences. Psychol Sci 21(1): 21–25.

Widmann A, Kujala T, Tervaniemi M, Kujala A, Schröger E (2004) From symbols to sounds: visual symbolic information activates sound representations. Psychophysiology 41(5):709–715.

(Accepted 12 September 2016)
(Available online 17 September 2016)