EEG activity evoked in preparation for multi-talker listening by adults and children

Emma Holmes[a], Padraig T. Kitterick[b,c], and A. Quentin Summerfield[a,d]

[a]Department of Psychology, University of York, UK

[b]NIHR Nottingham Hearing Biomedical Research Unit, UK

[c]Division of Clinical Neuroscience, School of Medicine, University of Nottingham, UK

[d]Hull York Medical School, University of York, UK

Corresponding author: Emma Holmes[1], eholme5@uwo.ca

[1] Current postal address: Department of Psychology, The Brain and Mind Institute, University of Western Ontario, London N6A 5B7, ON, Canada

Abstract

Selective attention is critical for successful speech perception because speech is often encountered in the presence of other sounds, including the voices of competing talkers. Faced with the need to attend selectively, listeners perceive speech more accurately when they know characteristics of upcoming talkers before they begin to speak. However, the neural processes that underlie the preparation of selective attention for voices are not fully understood. The current experiments used electroencephalography (EEG) to investigate the time course of brain activity during preparation for an upcoming talker in young adults aged 18-27 years with normal hearing (Experiments 1 and 2) and in typically-developing children aged 7-13 years (Experiment 3). Participants reported key words spoken by a target talker when an opposite-gender distractor talker spoke simultaneously. The two talkers were presented from different spatial locations (± 30° azimuth). Before the talkers began to speak, a visual cue indicated either the location (left/right) or the gender (male/female) of the target talker. Adults evoked preparatory EEG activity that started shortly after (< 50 ms) the visual cue was presented and was sustained until the talkers began to speak. The location cue evoked similar preparatory activity in Experiments 1 and 2 with different samples of participants. The gender cue did not evoke preparatory activity when it predicted gender only (Experiment 1) but did evoke preparatory activity when it predicted the identity of a specific talker with greater certainty (Experiment 2). Location cues evoked significant preparatory EEG activity in children but gender cues did not. The results provide converging evidence that listeners evoke consistent preparatory brain activity for selecting a talker by their location (regardless of their gender or identity), but not by their gender alone.

Key words

Speech recognition; Multi-talker listening; Auditory; Attention; EEG

1. Introduction

Listeners often face the challenge of understanding speech against a background of competing voices (e.g. Darwin, 2008). In this situation, intelligibility is improved if listeners know characteristics of the target talker before he or she begins to speak. Experiments with adults have shown benefits of knowing the spatial location (Ericson *et al.* 2004; Kidd *et al.* 2005; Best *et al.*, 2007; Best *et al.*, 2009) and the identity (Kitterick, Bailey, & Summerfield, 2010) of the target talker. Experiments with children (Dhamani, Leung, Carlile, & Sharma, 2013) and adults (Kitterick, Bailey, & Summerfield, 2010) have shown benefits of knowing *when* the target talker will speak. However, although these behavioural advantages have been observed consistently, the time course of brain activity evoked when adults are cued to talker characteristics is unknown. In addition, it is unclear whether children evoke similar brain activity as adults when attending to a target talker during multi-talker listening.

There is substantial evidence that preparatory brain activity can be observed before a target stimulus is presented, in response to an instructive cue that directs attention to a particular stimulus attribute. In the visual modality, preparatory activity is observed in dorsal and ventral cortical regions that are specialised for processing the cued dimension (Giesbrecht *et al.*, 2003; Slagter *et al.*, 2007; Woldorff *et al.*, 2004). In these regions, the amplitude of pre-target BOLD activity correlates with behavioural performance (Giesbrecht, Weissman, Woldorff, & Mangun, 2006). In the auditory modality, Voisin, Bidet-Caulet, Bertrand, and Fonlupt (2006) showed modulation of activity by spatial attention in auditory cortex. An arrow cued attention to the left or to the right and participants had to detect the presence of a noise burst that emerged with increasing intensity. Contrasts between left and right trials revealed activity in the superior temporal sulcus (including Heschl's gyrus and surrounding areas) that occurred contralateral to the cued side. Taken together, the results of these previous experiments demonstrate that preparatory activity occurs following an instructive visual cue and that such activity is necessary for successful behavioural performance.

Only two previous experiments (Hill & Miller, 2010; Lee *et al.*, 2013) have measured brain activity when participants prepare their attention for an upcoming talker during multi-talker listening tasks. Together, these experiments showed high overlap in the brain regions active when participants were cued visually to either the spatial location or the fundamental frequency (F0) of a target talker (i.e. reflecting domain-general preparatory activity); although, the magnitude of activity within a subset of these regions differed when participants prepared for location compared with F0 (i.e. showing aspects of preparatory activity that are cue-specific). Hill and Miller (2010) measured brain activity using functional magnetic resonance imaging (fMRI). On each trial, three simultaneous talkers were presented, differing in simulated spatial location and average F0. Before the acoustic stimuli began, a visual cue indicated either the location (left/right/centre) or the F0 (high/low/middle) of the target talker. The visual cue evoked activity in a left-hemisphere fronto-parietal network. The detailed pattern of activity within the network depended on whether participants were preparing to select the upcoming target talker based on location or F0. The results therefore provide evidence for both domain-general and cue-specific brain activity when participants are cued to location and F0.

A similar experiment by Lee *et al.* (2013) measured preparatory brain activity using magneto-encephalography (MEG). On each trial, two spoken digits were presented simultaneously, differing in simulated spatial location (left/right) and F0 (high/low). Similar to the experiment of Hill and Miller (2010), the visual cue preceded the acoustical stimuli and indicated either the spatial location or the F0 of the target talker. Lee *et al.* found greater preparatory activity in the left dorsal precentral sulcus and gyrus during attend-location trials and in the left posterior superior temporal sulcus during attend-F0 trials. These results, like Hill and Miller's, demonstrate cue-specific brain activity during preparatory attention.

Neither Hill and Miller (2010) nor Lee at al. (2013) addressed the question of how soon attentional preparation is manifest in neural activity. Hill and Miller's experiment revealed brain activity only with the low temporal resolution of fMRI. Lee *et al.* did not analyse MEG data until 600 ms after the start of the visual cues. They displayed the visual cue together with a fixation dot for 300 ms; they then extinguished the cue, leaving only the dot for 700 ms, at which point the acoustical stimuli were presented. They analysed MEG data in 400-ms windows immediately before and after the onset of the acoustical stimuli. Thus, 600 ms elapsed between the onset of the visual cue and the start of the first analysis window. However, research investigating preparation for an upcoming *visual* stimulus has revealed preparatory brain activity less than 250 ms after the onset of the cue (Yamaguchi, Tsuchiya, & Kobayashi, 1994). Evidence of attentional preparation with a similar latency would not have been shown by the analyses of Lee *et al.* Thus, a key goal of the current experiments was to explore the time course of attentional preparation and selection during multi-talker listening.

In addition, a possible shortcoming of the experiments of Hill and Miller (2010) and Lee *et al.* (2013) is that differences in the feature to be used for selection (i.e. location or F0) were confounded with differences in the visual cues. The present experiments sought to measure brain activity during preparatory attention in children as well as in adults; as a result, we used cues that were less abstract, and hence more physically elaborate, than those used by Hill and Miller and by Lee *et al*. Thus, we also implemented a control condition, which aimed to check whether cue-specific effects could be explained by physical differences (e.g. in luminance or complexity) of the visual stimuli that we used to cue attention.

One motivation for testing children was to establish whether they evoke similar preparatory brain activity to adults. Dhamani *et al.* (2013) showed that children aged 10–15 years, like adults, achieve better speech intelligibility when they are cued to an informative feature of the target talker. However, the ability to extract speech from interfering sources of sound has a long developmental time-course (Cameron & Dillon, 2007; Cameron *et al.*, 2009; Vaillancourt, Laroche, Giguère, & Soli, 2008; Wightman, Kistler, & Brungart, 2006), meaning that children might evoke different, or less consistent, preparatory brain activity than adults. We sought to investigate this issue by establishing the similarity between children and adults in the timing of significant preparatory brain activity.

Against this background, the aim of the three experiments reported in this paper was to measure the temporal dynamics of brain activity in a two-talker listening task in adults and in children aged 7–13 years. We measured brain activity using electro-encephalography (EEG). Participants reported key words spoken by a target talker in the presence of a simultaneous competing talker. On each trial, a visual cue was presented before the talkers spoke to inform participants about either the spatial location of the target talker (left/right of fixation) or their gender (male/female).

We expected to observe both similarities and differences between the event-related potentials (ERPs) evoked when participants were instructed to select the target talker on the basis of location or gender. Similarities were expected to reflect domain-general processing of location and gender information, akin to the similarities in brain activity observed by Hill and Miller (2010) when listeners attended to talkers on the basis of location or F0. Differences in ERPs were expected to reflect cue-specific processing. Like Hill and Miller (2010) and Lee *et al.* (2013), we were interested in activity that arose in two phases of the task: (1) following the onset of the visual cue before the acoustic stimuli started, and (2) during the acoustic stimuli. We refer to the first phase as the "Preparatory Phase" and the second phase as the "Selective Phase". During the Preparatory Phase, we measured responses evoked by a visual cue. During the Selective Phase, the acoustical stimuli were natural spoken sentences from the Co-ordinate Response Measure corpus (CRM; Moore, 1981). We chose CRM stimuli because they have been shown to be engaging stimuli for children in tasks of selective attention (Rothpletz *et al.*, 2012; Wightman & Kistler, 2005; Wightman, Kistler, & O'Bryan, 2010).

2. Experiment 1

Experiment 1 aimed to investigate the timing of EEG activity during multi-talker listening in adults who were cued to the location or gender of a target talker. We aimed to identify robust attentional activity that did not reflect differences in physical aspect of the visual stimuli used to cue attention, so we also implemented a control condition that measured brain activity evoked by the visual cues when they did not have implications for auditory attention.

2.1. Methods

2.1.1. Participants

Participants were 16 young adults (8 male), aged 18–24 years (mean [M] = 20.4, standard deviation [SD] = 1.5). They were self-declared right-handed native English speakers with no history of hearing problems. They had 5-frequency average pure-tone hearing levels of 20 dB HL or better, tested in accordance with BS EN ISO 8253-1 (British Society of Audiology, 2004). The study was approved by the Research Ethics Committee of the Department of Psychology, University of York.

2.1.2. Apparatus

The experiment was conducted in a 5.3 m x 3.7 m single-walled test room (Industrial Acoustics Co., NY) located within a larger sound-treated room. Participants sat facing three loudspeakers (Plus XS.2, Canton, Germany) arranged in a circular arc at a height of 1 m at 0° azimuth (fixation) and at 30° to the left and right (Fig. 1). A 15-inch visual display unit (VDU; NEC AccuSync 52VM) was positioned directly below the central loudspeaker.

< INSERT FIG. 1 HERE >

2.1.3. Stimuli

Visual cues

Four visual cues, “left”, “right”, “male”, and “female”, were defined by white lines on a black background. Left and right cues were leftward- and rightward-pointing chevrons, respectively; male and female cues were stick figures (Fig. 2A–D). A composite visual stimulus was created by overlaying the four cues (Fig. 2E).

< INSERT FIG. 2 HERE >

Acoustical test stimuli

Acoustical test stimuli were sentences from the Co-ordinate Response Measure corpus (CRM; Moore, 1981) spoken by native British-English talkers (Kitterick, Bailey, and Summerfield, 2010). CRM sentences have the form ‘Ready <call sign>, go to <colour> <number> now’. In the sub-set used in the experiment, there were eight call-signs (‘Arrow’, ‘Baron’, ‘Charlie’, ‘Eagle’, ‘Hopper’, ‘Laker’, ‘Ringo’, ‘Tiger’), four colours (‘blue’, ‘red’, ‘green’, ‘white’), and four numbers (‘one’, ‘two’, ‘three’, ‘four’). An example is “Ready Charlie, go to green two now”. Sentences spoken by three male talkers and three female talkers were selected from the corpus. The sentences had an average duration of 2.5 seconds.

The levels of the digital recordings of the sentences were normalised to the same root mean square (RMS) power.

Acoustical control stimuli

Control stimuli were single-channel noise-vocoded representations of concurrent pairs of CRM sentences. Each control stimulus was created by summing a pair of sentences digitally with their onsets aligned, extracting the temporal envelope of the combination with the Hilbert Transform (Hilbert, 1912), and using the envelope to modulate the amplitude of random noise whose long-term spectrum matched the average spectrum of all of the pairs of sentences.

< INSERT FIG. 3 HERE >

2.1.4. Procedures

Test Condition

Fig. 3A illustrates the trial structure for the Test Condition. At the start of each trial, a fixation cross was presented for 1000 ms. Next, the visual composite stimulus was presented. After 800 ms, elements of the composite stimulus faded over 200 ms to reveal the visual cue. We used a decrease, rather than an increase, in luminance to reveal the cue in order to minimise any onset response to the visual cue in the EEG recording. After the cue had been fully revealed for 1000 ms, two different CRM sentences were presented concurrently, one from the left loudspeaker, the other from the right. The sentences started simultaneously, but contained different call signs and different colour-number combinations. The two talkers were selected quasi-randomly on each trial, with the restriction that one talker was male and the other was female. Over the course of the experiment, each of the six talkers was presented equally often from each location.

The visual cue directed attention to the target talker and varied from trial to trial. The cue remained on the screen throughout the duration of the acoustic stimuli so that participants did not have to retain the cue in memory. After both sentences had ended, participants were instructed to report the colour-number combination that was spoken by the target talker by pressing a coloured digit on a touch screen. The inter-trial interval varied randomly from 1000 to 1500 ms to desynchronise anticipatory activity for the next trial. Each participant completed 384 trials (96 in each cueing condition), with a break after every 48 trials.

The average presentation level of concurrent pairs of test sentences was set to 63 dB(A) SPL (range 61.6—66.2 dB) measured with a B&K (Brüel & Kjær, Nærum, Denmark) Sound Level Meter (Type 2260 Investigator) and 0.5-inch Free-field Microphone (Type 4189) placed in the centre of the arc at the height of the loudspeakers with the participant absent.

Control Condition

The trial structure of the Control Condition (Fig. 3B) was the same as the Test Condition with the exception that an acoustical control stimulus, presented from a single loudspeaker at 0° azimuth, replaced the pair of acoustical test stimuli. The task was to press the picture on the touch screen that corresponded to the visual cue that was presented. Each participant completed 216 trials (54 in each visual stimulus condition), with a break every 36 trials. The presentation level of the Control stimuli was set so that their average sound pressure level matched the average level of the pairs of Test stimuli when measured at the listening position. Participants undertook the Control Condition before the Test Condition. This task order was a crucial feature of the design, since we wanted to measure activity in response to the visual cues *before* participants had learnt the association between the visual cues and the acoustical stimuli presented in the Test Condition.

The logic behind the design of the Control Condition was that the stimuli lacked the spectral detail and temporal fine structure required for the perception of pitch (Moore, 2008). In addition, because the stimuli were presented from one loudspeaker, they did not provide the interaural differences in level and timing required for their constituent voices to be localised separately. In these ways, the acoustic cues required to segregate the sentences by gender and by location were neutralised, while the overall energy and gross fluctuations in amplitude of the test stimuli were preserved.

2.1.5. EEG recording and processing

Continuous EEG was recorded using the ANT WaveGuard-64 system (ANT, Netherlands; www.ant-neuro.com) with Ag/AgCl electrodes mounted on an elasticated cap. Electrode AFz was used as a ground site. The horizontal electro-oculogram (EOG) was measured with a bipolar lead attached to the outer canthi of the left and right eyes and the vertical EOG was measured with a bipolar lead above and below the right eye. The EEG was amplified and digitised with an ANT High-Speed Amplifier at a sampling rate of 1000 samples/s per channel.

The continuous EEG recordings were exported to MATLAB 7 (The MathWorks, Inc., Natick, MA, USA) and analysed using the EEGLAB toolbox (Version 9; http://sccn.ucsd.edu/eeglab/). Before statistical analysis, the data were band-pass filtered using a Butterworth filter between 0.25 and 30 Hz. The data were filtered, reversed, and filtered again to ensure zero phase shifting. We also conducted post-hoc analyses using different high pass filter values (0.1, 0.2, 0.3, 0.4, and 0.5 Hz) and found that the results did not change substantially between 0.2 and 0.4 Hz (Supplementary Table 2). The amplitude at each electrode was referenced to the average amplitude of the electrode array. Epochs were created with 4700 ms duration, including a baseline interval of 200 ms at the end of the fixation-cross period. Epochs were rejected for further analysis if they contained high-amplitude artifacts (absolute amplitude in any channel greater than ±200 μV) or if the behavioural response to the trial was incorrect. This method led to the rejection of approximately 12.5% of trials. Independent component analysis (ICA) was used to correct for any remaining eye-blink artifacts, which were identified by a stereotyped scalp topography and a product-moment correlation with the vertical EOG recording that exceeded 0.6 for >70% of trials containing high-amplitude peaks.

2.1.6. Behavioural analyses

Trials were separated into Location (average left/right cues) and Gender (average male/female cues) groups, separately for the Test and Control Conditions. Responses were scored as correct if both the colour and number key words were reported correctly in the Test Condition, and if the visual cue was reported correctly in the Control Condition.

2.1.7. Analyses of ERPs

Our primary aims were to determine the time course of attentional preparation in relation to the onset of the visual cue (Preparatory Phase) and the time course of attentional selection in relation to the onset of the acoustical stimuli (Selective Phase). While we expected to find significant differences between the Test and Control Conditions in both phases, we had no prior expectations about the timing of significant differences within each phase. Accordingly, in seeking significant differences, a Spatio-temporal Cluster-based Permutation Analysis was conducted (Maris & Oostenveld, 2007), in which the cluster statistic was calculated as the sum of the *t*-values within the cluster (at each space-by-time point). A strength of the method is that it does not require prior assumptions about the spatial or temporal location of significant effects while overcoming the problem of making multiple comparisons across electrodes and temporal samples. The analysis was used to make two types of comparison. Type-I analyses compared the amplitudes of ERPs between the Test and Control Conditions, separately for Location and Gender trials. Clusters found during Type-I analyses in the Preparatory Phase could not arise from sensory or perceptual processes because the stimuli did not differ between the conditions in this phase. Rather, such differences were interpreted as arising from contrasting attentional activity between the Test and Control Conditions. Type-I clusters found in the Selective Phase, in contrast, could arise *either* from differences in attentional activity *or* from differences between the acoustical structure of the Test and Control stimuli. An alternative interpretation, that Type I clusters could reflect differences in generalised arousal rather than focussed attention, is considered in discussing the results.

The use of the average reference was likely to produce complementary clusters with opposite polarities at different scalp locations, which could reflect underlying activity at a single source. We, thus, implemented a cluster naming system to respect this complementarity. Clusters are numbered respective to their onset latencies, with the suffix P or N, indicating positive (Test > Control) or negative

(Control > Test) polarities, respectively. An additional suffix (E: early; L: late) was added if two clusters of one polarity overlapped with one cluster of the opposite polarity.

Type-II analyses compared Location with Gender trials in the Test Condition only. These analyses identified clusters where ERPs differed significantly depending on whether participants were receiving cues for, and directing attention towards, location or gender. We implemented a similar cluster naming system as for Type-I analyses, except that the suffixes P and N refer to greater amplitude in Location than Gender trials (Location > Gender) or greater amplitude in Gender than Location trials (Gender > Location), respectively. Such differences could be evoked *either* by different attentional processes *or* by physical differences between the visual cues. Accordingly, we compared the amplitudes of ERPs on Location and Gender trials—averaged over the space-by-time-points in the cluster—between the Test and Control Conditions in a 2 x 2 ANOVA. The rationale was that differences in the visual cues between Location and Gender trials were also present in the Control Condition, but the attentional activity evoked by the cues should be present in the Test but not the Control Condition. Thus, a significant two-way interaction meant that the cluster could not be fully explained by the influence of physical differences in the visual cues between conditions.

2.2. Results

2.2.1. Behavioural results

Conjoint accuracy in identifying the colour and number key words in the Test and Control Conditions was high and, therefore, the data were converted to rationalized arcsine units (RAU; Studebaker, 1985) before paired-sampled *t*-tests were conducted. Accuracy in the Test Condition did not differ between Location (M = 95.3%, SD = 0.05) and Gender (M = 94.8%, SD = 0.05) trials, $t(15) = 1.1$, $p = 0.29$. There was also no significant difference in the accuracy with which the visual cue was identified in the Control Condition between Location (M = 99.4%, SD = 0.01) and Gender (M = 99.1%, SD = 0.02) trials, $t(15) = 0.3$, $p = 0.75$.

2.2.2. Event-related potentials

Type-I analyses: Differences between Test and Control Conditions

Location trials

Fig. 4 illustrates the results of the Type-I analyses on trials in which a Location cue (left/right) was presented. During the 1000-ms Preparatory Phase, one significant cluster of activity (Cluster 1N) was identified (Fig. 4B). The existence of Cluster 1N demonstrates that differences in brain activity arise between a condition in which a visual cue has no implications for auditory attention and a condition in which the same cue directs listeners to prepare to select an upcoming talker on the basis of their location. Cluster 1N began 227 ms after the visual cue began to appear and 27 ms after the visual cue was fully revealed. The polarity, location, onset time, and duration of Cluster 1N are tabulated in Table 1.

During the Selective Phase, four significant clusters of activity were identified (Clusters 2–3). Clusters 2N and 2P were complementary, since they showed opposite polarity at overlapping time points. Cluster 2N spanned the interval from 69 to 1029 ms (Fig. 4C), relative to the start of the phase, and Cluster 2P spanned the interval from 81 to 671 ms (Fig. 4D). Cluster 3N (1072 to 2200 ms; Fig. 4E) started shortly after Cluster 2N had finished. Cluster 3P (1696 to 2200 ms; Fig. 4F) started towards the end of the Selective Phase. Overall, significant Type-I differences occurred throughout the majority of the Selective Phase of Location trials (Fig. 4A).

Gender trials

The second of the Type-I analyses compared ERPs between the Test and Control Conditions on trials in which a Gender cue (male/female) was presented. Panels G–J of Fig. 4 illustrate these results. No significant clusters were identified during the Preparatory Phase. During the Selective Phase, three significant clusters were identified (Clusters 4–5; Fig. 4G). Clusters 4N (108 to 1030 ms; Fig. 4H) and 4P (495 to 1038 ms; Fig. 4I) were complementary. Cluster 5P (1717 to 2200 ms; Fig. 4J) occurred later during the Selective Phase. Many of the electrodes in Cluster 5P overlapped with the electrodes that contributed to Cluster 4P.

< INSERT FIG. 4 HERE >

Type-II analyses: Differences between Location and Gender trials

Fig. 5 illustrates the results of Type-II analyses that compared ERPs between Location and Gender trials in the Test Condition. The analysis identified three significant clusters during the Preparatory Phase (Clusters 6P, $6N_E$ and $6N_L$; Fig. 5B–D). The polarity, location, onset time, and duration of these clusters are listed in the third column of Table 2. Two of the clusters were complementary and occurred towards the beginning of the Preparatory Phase (Cluster 6P: 29 ms to 628 ms; Cluster $6N_E$: 40 to 429 ms). The third cluster arose later during the Preparatory Phase (Cluster $6N_L$: 484 to 948 ms).

The clusters identified towards the beginning of the Preparatory Phase (Clusters 6P and $6N_E$) showed the same patterns of activity in the Control Condition ($p \leq 0.011$; Fig. 6) as in the Test Condition. For these clusters, the interaction between cue type (Location/Gender) and condition (Test/Control) was not significant. Therefore, it is not possible to rule out the explanation that Type-II clusters that occurred towards the beginning of the Preparatory Phase arose from differences in the visual cues, rather than from differences in attentional processes triggered by the cues.

Table 1. (Continued on next page). Summary of results for the Gender and Location Condition comparisons (Type-I analysis) between the Test and Control Conditions in Experiments 1–3. The rows headed 'Cluster *p*-value' show the results of the Spatio-temporal Cluster-based Permutation Analysis.

Phase		**Experiment 1 Location**	**Experiment 2 Location**	**Experiment 3 Location**	**Experiment 1 Gender**	**Experiment 2 Gender**	**Experiment 3 Gender**
Preparatory	**Cluster Number**	1N	8N	15N	-	10N	-
	Cluster *p*-value	0.001	0.014	0.039	-	0.03	-
	Polarity	Control > Test	Control > Test	Control > Test	-	Control > Test	-
	Electrode Locations	Central	Central	Posterior	-	Central	-
	Onset of cluster (ms)	27	43	210	-	84	-
	Duration of cluster (ms)	664	638	245	-	300	-
Selective	**Cluster Number**	2N	9N	16N	4N	11N	17N
	Cluster *p*-value	0.002	0.001	< 0.001	0.002	< 0.001	0.010
	Polarity	Control > Test	Control > Test	Control > Test	Control > Test	Control > Test	Control > Test
	Electrode Locations	Posterior + Central	Central	Posterior + Central	Central + Anterior	Central	Central
	Onset of cluster (ms)	69	286	71	108	112	573
	Duration of cluster (ms)	960	1176	1432	922	1107	494
Selective *(continued on next page)*	**Cluster Number**	2P	9P	16P	4P	11P	17P
	Cluster *p*-value	0.041	0.014	0.034	0.010	0.007	0.021
	Polarity	Test > Control	Test > Control	Test > Control	Test > Control	Test > Control	Test > Control

Table 1. (Continued from previous page)

Phase		**Experiment 1 Location**	**Experiment 2 Location**	**Experiment 3 Location**	**Experiment 1 Gender**	**Experiment 2 Gender**	**Experiment 3 Gender**
Selective *(continued from previous page)*	**Electrode Locations**	Non-Central	Posterior	Non-Central	Posterior	Posterior	Posterior
	Onset of cluster (ms)	81	298	77	495	502	483
	Duration of cluster (ms)	590	704	864	543	529	556
Selective	**Cluster Number**	3N	-	-	-	12N	-
	Cluster *p*-value	0.002	-	-	-	0.003	-
	Polarity	Control > Test	-	-	-	Control > Test	-
	Electrode Locations	Central	-	-	-	Central	-
	Onset of cluster (ms)	1072	-	-	-	1261	-
	Duration of cluster (ms)	1128	-	-	-	939	-
Selective	**Cluster Number**	3P	-	-	5P	12P	-
	Cluster *p*-value	0.027	-	-	0.033	0.047	-
	Polarity	Test > Control	-	-	Test > Control	Test > Control	-
	Electrode Locations	Posterior	-	-	Posterior	Posterior	-
	Onset of cluster (ms)	1696	-	-	1717	1844	-
	Duration of cluster (ms)	504	-	-	483	356	-

The cluster that occurred later during the Preparatory Phase (Cluster $6N_L$) showed a different pattern of results to Clusters 6P and $6N_E$. Although the interaction between cue type and condition was not significant [$F(1,15) = 2.32$, $p = 0.15$; Fig. 6C], the difference between Location and Gender trials was significant in the Test condition but not the Control condition ($p = 0.90$). The finding that ERPs did not differ between Location and Gender trials in the Control Condition implies that activity within this cluster might reflect differences in the attentional processes triggered by the cues in the Test Condition. However, the finding of no significant interaction means that it was not possible to fully rule out the explanation that the cluster arose from differences in the visual cues.

There were two significant clusters during the Selective Phase (Fig. 5E–F), which overlapped in time (Cluster 7N: 371 to 1206 ms; Cluster 7P: 590 to 869 ms). These clusters did not show the same pattern in the Control Condition ($p \geq 0.26$). For Cluster 7P, there was a significant interaction between cue type (Location/Gender) and condition (Test/Control) [$F(1,15) = 11.07$, $p = 0.005$], although the interaction was not significant for Cluster 7N [$F(1,15) = 3.46$, $p = 0.08$]. Overall, the finding of a significant interaction for Cluster 7P provides strong evidence for differences in the processes for attending selectively to a talker between Location and Gender trials. Whereas, Cluster 7N provides weaker evidence since it was not possible to fully rule out the explanation that the cluster arose from differences in the visual cues.

< INSERT FIG. 5 & FIG. 6 HERE >

Table 2. (Continued on next page). Summary of results for the Test Condition comparison between Location and Gender trials (Type-II analysis) across Experiments 1–3. A tick in the row headed ‘Significant in Control Condition?’ indicates that the difference in the amplitude of ERPs between Location and Gender trials was significant in the Control Condition across the spatio-temporal points of the cluster (p-values displayed underneath). A tick in the row headed ‘Significant Test/Control Interaction?’ indicates that an ANOVA with the factors cue type (Location/Gender) and condition (Test/Control) revealed a significant two-way interaction (p-values displayed underneath).

Phase	Properties	Experiment 1	Experiment 2	Experiment 3
Preparatory	**Cluster Number**	6P	13P	18P
	Cluster *p*-value	< 0.001	0.004	0.014
	Polarity	Loc > Gen	Loc > Gen	Loc > Gen
	Electrode Locations	Posterior	Posterior	Posterior
	Onset of cluster (ms)	29	53	72
	Duration of cluster (ms)	599	342	372
	Significant in Control Condition?	✓ $p = 0.011$	✓ $p = 0.017$	✓ $p < 0.001$
	Significant Test/Control Interaction?	✗ $p = 0.82$	✗ $p = 0.85$	✓ $p = 0.003$

Table 2. (Continued from the previous page)

Preparatory	**Cluster Number**	$6N_E$	13N	-
	Cluster *p*-value	0.003	0.005	-
	Polarity	Gen > Loc	Gen > Loc	-
	Electrode Locations	Anterior + Central	Anterior + Central	-
	Onset of cluster (ms)	40	103	-
	Duration of cluster (ms)	389	288	-
	Significant in Control Condition?	✓ *p* = 0.002	✓ *p* = 0.014	-
	Significant Test/Control Interaction?	✗ *p* = 0.80	✗ *p* = 0.80	-
Preparatory	**Cluster Number**	$6N_L$	-	-
	Cluster *p*-value	0.010	-	-
	Polarity	Gen > Loc	-	-
	Electrode Locations	Central	-	-
	Onset of cluster (ms)	484	-	-
	Duration of cluster (ms)	464	-	-
	Significant in Control Condition?	✗ *p* = 0.15	-	-
	Significant Test/Control Interaction?	✗ *p* = 0.90	-	-
Selective	**Cluster Number**	7N	14N	19N
	Cluster *p*-value	< 0.001	0.018	0.022
	Polarity	Gen > Loc	Gen > Loc	Gen > Loc
	Electrode Locations	Posterior + Central	Central	Central
	Onset of cluster (ms)	371	807	1069
	Duration of cluster (ms)	835	396	455
	Significant in Control Condition?	✗ *p* = 0.56	✗ *p* = 0.31	✗ *p* = 0.07
	Significant Test/Control Interaction?	✗ *p* = 0.08	✓ *p* = 0.044	✓ *p* = 0.001
Selective	**Cluster Number**	7P	-	-
	Cluster *p*-value	0.049	-	-
	Polarity	Loc > Gen	-	-
	Electrode Locations	Anterior	-	-
	Onset of cluster (ms)	590	-	-
	Duration of cluster (ms)	279	-	-
	Significant in Control Condition?	✗ *p* = 0.26	-	-
	Significant Test/Control Interaction?	✓ *p* = 0. 005	-	-

2.3. Discussion

During the Preparatory Phase, significantly different ERPs were evoked in the Test Condition compared with the Control Condition (Type-I analyses), but only on Location trials (Figs. 4A–B) and not on Gender trials (Fig. 4G). During this phase, no acoustical stimuli were presented and the visual stimuli did not differ between the Test and Control Conditions. The result is compatible with the interpretation that a visual cue for spatial location can trigger preparatory attentional activity. Moreover, it does so with a short latency (< 50 ms) after the full reveal of the visual cue.

There were significant differences between Location and Gender trials during the Preparatory Phase of the Test Condition (Type-II analyses). These differences had a similar latency to Type-I differences during Location trials (Figs. 5B–C). However, a difference between Location and Gender trials also occurred in the Control Condition at the same electrodes and time points (Figs. 6A–B). Thus, it is not possible to rule out the explanation that these early clusters were evoked largely by physical differences between the visual cues for location compared with gender, rather than by differences in preparatory attentional processes triggered by the different cue types. The physical differences encompassed both their luminance and structural complexity. A further contribution may have come from differences between the cognitive processes evoked by the representation of an inanimate object (a chevron) compared with a human being (Caramazza & Shelton, 1998; Downing, Chan, Peelen, Dodds, & Kanwisher, 2006).

During the Selective Phase, differences in ERPs arose between Location and Gender trials that could not be attributed to differences in the visual cues (Cluster 7P, Fig. 6E). Given that the acoustical stimuli were identical in Location and Gender trials, the different ERPs presumably reflect differences in the analysis and grouping of acoustic cues for selecting a voice by location or by gender. This result is discussed further in the General Discussion (Section 5).

The absence of evidence of preparatory activity on Gender trials in Fig. 4 is informative. The result argues against the interpretation that Cluster 1N, found on Location trials, arose from greater arousal in the Test Condition than the Control Condition. Greater arousal would have been expected to occur irrespective of the trial type and thus be shown on Gender trials as well as Location trials. Instead, it is possible that the result arose from a feature of the design. Whereas there were only two possible locations, there were three possible male and three possible female talkers. As a result, there was more variation in the evidence of gender (e.g. in average values of the F0 and formant frequencies) than in the evidence of location. Thus, the cues for location were more specific than the cues for gender. Even though the difference in specificity was not reflected in differences in behavioural accuracy, it might have influenced the patterns of brain activity that were observed during the Preparatory Phase.

3. Experiment 2

To avoid differences in the specificity of the visual cues for attributes of the target talker between Location and Gender trials, the same male and female talker were presented for the entire experiment, rather than employing three instances of each gender as in Experiment 1. Thus, gender cues indicated the gender but also the *identity* of the target talker. Participants were also familiarised with the locations and genders before the Test Condition was administered.

Experiment 2 tested two hypotheses: first, that gender cues can evoke preparatory brain activity (similar to that observed on Location trials) provided that variation in the evidence of gender is minimised, and second, that differential activity emerges between Location and Gender trials when both types of cue are similarly specific. An additional aim was to determine whether the overall pattern of results of Experiment 1 could be replicated with a different set of participants.

3.1. Methods

3.1.1. Participants

Participants were 16 young adults (8 male), aged 18–27 years (M = 21.3, SD = 2.1), none of whom had taken part in Experiment 1. All participants were self-declared right-handed native English speakers with no history of hearing problems. Participants all had 5-frequency average pure-tone hearing levels of 20 dB HL or better, tested in accordance with BS EN ISO 8253-1 (British Society of Audiology, 2004). The study was approved by the Research Ethics Committee of the Department of Psychology, University of York.

3.1.2. Stimuli and procedure

Stimuli and procedures were the same as those in Experiment 1 except that only one of the male and one of the female talkers were used. After participants had completed the Control Condition, but before they undertook the Test Condition, a block of trials was presented, which aimed to familiarise participants with the two locations and the two talkers. Familiarisation involved 52 trials in which only one or other of the two talkers, but not both, was presented during the Selective Phase. The trial structure was the same as the Test Condition except that there was no competing talker and EEG was not recorded.

3.1.3. EEG recording, processing, and analyses

The EEG recording, processing, and analysis procedures were identical to those used in Experiment 1.

3.2. Results

3.2.1. Behavioural results

Conjoint accuracy in identifying the colour and number key words in the Test and Control Conditions was high and, therefore, the data were converted to rationalized arcsine units (RAU; Studebaker, 1985) before paired-samples *t*-tests were conducted. Accuracy in the Test Condition did not differ between Location (M = 96.5%, SD = 0.02) and Gender (M = 95.9%, SD = 0.02) trials, $t(15) = 1.5$, $p = 0.17$. There were also no significant differences in the accuracy with which the visual cue was identified in the Control Condition between Location (M = 99.6%, SD = 0.01) and Gender (M = 99.6%, SD = 0.01) trials, $t(15) = 0.3$, $p = 0.75$.

3.2.2. Event-related potentials

Type-I analyses: Differences between Test and Control Conditions

Location trials

Panels A-D of Fig. 7 illustrate the results of the Type-I analyses that compared ERPs between the Test and Control Conditions on trials in which a Location cue was presented. One significant cluster of activity was identified during the Preparatory Phase (Fig. 7B) and two significant clusters were identified during the Selective Phase (Fig. 7C–D). The polarity, location, onset time, and duration of these clusters are listed in Table 1.

< INSERT FIG. 7 HERE >

Gender trials

Panels E–J of Fig. 7 illustrate the results of the Type-I analysis that compared ERPs between the Test and Control Conditions on trials in which a Gender cue was presented. One significant cluster was identified during the Preparatory Phase (Fig. 7F) and four significant clusters were identified during the Selective Phase (Fig. 7G–J; Table 1).

Type-II analyses: Differences between Location and Gender Conditions

Fig. 8 illustrates the results of Type-II analyses that compared ERPs between Location and Gender trials in the Test Condition. The analysis identified two significant clusters during the Preparatory Phase (Fig. 8B–C) and one significant cluster during the Selective Phase (Fig. 8D). The polarity, location, onset time, and duration of these clusters are listed in Table 2.

< INSERT FIG. 8 & FIG. 9 HERE >

The clusters identified during the Preparatory Phase (Clusters 13P and 13N) showed the same patterns of activity in the Control Condition ($p \leq 0.017$; Fig. 9). For these clusters, the interaction between cue type (Location/Gender) and condition (Test/Control) was not significant. Therefore, it is not possible to rule out the explanation that Type-II clusters during the Preparatory Phase arose from differences in the visual cues, rather than from differences in attentional processes triggered by the cues.

The cluster identified during the Selective Phase (Cluster 14N) did not show the same pattern in the Control Condition ($p = 0.31$; Fig. 9C). The interaction between cue type (Location/Gender) and condition (Test/Control) was significant [$F(1,15) = 4.82$, $p = 0.044$]. This finding is compatible with the idea that the cluster during the Selective Phase arose from differences in the processes for attending selectively to a talker between Location and Gender trials.

3.3. Discussion

Experiment 2 partially replicated the results of Experiment 1. Both experiments provided evidence of activity during the Preparatory Phase of Location trials that began earlier than 50 ms after the visual cue was fully revealed, lasted longer than 600 ms, and was characterised by more negative amplitudes for the Test than Control Condition at central electrodes (Figs. 4B and 7B). Additionally, both experiments revealed Type-II differences between Location and Gender trials during the Preparatory Phase that were present both in the Test and in the Control Conditions. The findings during the Selective Phases are also similar: Type-I differences in Location and Gender trials occurred throughout the Selective Phase of Experiments 1 and 2, characterised by more negative amplitudes during the Test than Control Condition at central electrodes and more positive amplitudes at non-central (typically posterior) electrodes. Type-II results in both experiments also showed more negative amplitudes on Location than Gender trials at central electrodes during the Selective Phase, which were not present in the Control Condition.

A difference between the results of the two experiments is the finding of significant activity during the Preparatory Phase of Gender trials in Experiment 2 that was not present in Experiment 1. The result is compatible with the idea that adult listeners evoke consistent brain activity in preparing auditory attention to select a particular voice in response to a cue for gender, but that activity is less consistent in space and time when only the gender of the target talker is known and there is uncertainty about the talker's identity and vocal characteristics.

4. Experiment 3

Experiment 3 examined whether typically-developing children aged 7–13 years display evidence of preparatory and selective attention similar to that shown by adults in Experiments 1 and 2. There is extensive evidence that that the ability to extract speech from interfering sounds develops during childhood and is not adult-like until late in the teenage years (e.g. Cameron & Dillon, 2007; Cameron *et al.*, 2009; Vaillancourt *et al.*, 2008; Wightman *et al.*, 2006). It is less clear whether the ability to benefit from a cue which guides attentional selection also develops slowly. We had identified one experiment which demonstrates that children are able to use advance knowledge about the time at which a talker will start speaking to improve speech intelligibility in background noise (Dhamani *et al.*, 2013). However, to our knowledge, there has been no direct comparison between adults and children of the extent to which an attentional cue improves speech intelligibility. In addition, no experiments to our knowledge have demonstrated that children can use cues for location or gender to improve speech intelligibility, akin to the cues employed in the current experiments. Therefore, we aimed to contrast two hypotheses. One was that children do not prepare attention for location or gender in advance of the onset of an acoustical stimulus; in which case, we would find no evidence of preparatory brain activity. The alternative was that children do prepare attention, in which case it would be possible to measure preparatory brain activity in children, but it might be of lower amplitude and longer latency than the corresponding activity measured in adults.

Concerning brain activity during the Selective Phase, even if children at a particular developmental stage are able to extract the speech of a target talker with similar behavioural accuracy as adults, the underlying brain processes may differ between children and adults. While we expected to find differences between adults and children in ERPs during the selective phase, we had no a priori expectations about when in time the differences would occur.

4.1. Methods

4.1.1. Participants

Participants were 26 children (12 male), aged 7–13 years (M = 10.5, SD = 1.7). All participants were declared by their parents to be right-handed native English speakers with no history of hearing problems. All participants had 5-frequency pure-tone average hearing threshold levels of 35 dB or better, tested in accordance with BS EN ISO 8253-1 (British Society of Audiology, 2004). Two participants were excluded from the analysis—one due to a technical problem during data collection and another due to poor behavioural performance in Location trials during the Test Condition (20.8%). It was evident that the child had forgotten the association between the location cues and the target talker. Thus, analyses are based on data from 24 participants. The study was approved by the Research Ethics Committee of the Department of Psychology, University of York.

4.1.2. Stimuli and procedure

Stimuli and procedures were the same as those in Experiment 2, except that children completed only 96 trials in the Control Condition and between 96 and 144 trials in the Test Condition (depending on their level of fatigue). Participants received a short break every 16 trials and a longer break every 48 trials. Before undertaking the Test Condition, children completed 16 familiarisation trials (4 in each cue type condition).

4.1.3. <u>EEG recording, processing, and analyses</u>

EEG recording, processing, and analyses procedures were the same as those in Experiment 2, with one exception. Due to the higher rate of artefacts in EEG data from children than adults, the artefact rejection criteria were relaxed to maintain a similar proportion of rejected trials as in the adult EEG data (< 12.5%). For participants that showed evidence of artefacts following artefact rejection, ICA was applied to correct for remaining eye-blink artefacts.

4.2. Results

4.2.1. <u>Behavioural results</u>

The data were converted to rationalized arcsine units (RAU; Studebaker, 1985) before paired-sampled *t*-tests were conducted. Conjoint accuracy in identifying the colour and number key words in the Test Condition was moderately high and did not differ between Location (M = 89.4%, SD = 7.46) and Gender (M = 88.6%, SD = 7.98) trials, $t(23) = 0.8$, $p = 0.42$. There were no significant differences in the accuracy with which the visual cue was identified in the Control Condition between Location (M = 97.5%, SD = 3.25) and Gender (M = 98.0%, SD = 2.08) trials, $t(23) = 0.1$, $p = 0.91$.

1.1.1. <u>Event-related potentials</u>

Type-I analyses: Differences between Test and Control Conditions

Location trials

Fig. 10 illustrates the results of the Type-I analyses. Panels A–D report the analysis that compared ERPs between the Test and Control Conditions on trials in which a Location cue was presented. One significant cluster of activity was identified during the Preparatory Phase (Fig. 10B) and two significant clusters were identified during the Selective Phase (Fig. 10C–D; Table 1).

< INSERT FIG. 10 HERE >

Gender trials

Panels E–G of Fig. 10 illustrate the results of the Type-I analysis that compared ERPs between the Test and Control Conditions on trials in which a Gender cue was presented. No significant clusters of activity were identified during the Preparatory Phase, but two significant clusters were identified during the Selective Phase (Fig. 10F–G; Table 1).

Type-II analyses: Differences between Location and Gender trials

Fig. 11 illustrates the results of Type-II analyses that compared ERPs between Location and Gender trials in the Test Condition. The analysis identified one significant cluster during the Preparatory Phase (Fig. 11B) and one significant cluster during the Selective Phase (Fig. 11C; Table 2).

< INSERT FIG. 11 & FIG. 12 HERE >

The cluster identified during the Preparatory Phase (Cluster 18P) showed a greater difference between Location and Gender trials in the Control Condition ($p < 0.001$; Fig. 12A), which was demonstrated by a significant interaction between cue type (Location/Gender) and condition (Test/Control) [$F(1,23) = 10.74$, $p = 0.003$; Fig. 12A]. Therefore, it is not possible to rule out the explanation that the cluster arose from differences in the visual cues, rather than from differences in attentional processes triggered by the cues.

The cluster identified during the Selective Phase (Cluster 19N), however, did not show the same pattern in the Control Condition ($p = 0.07$; Fig. 12B). Furthermore, the interaction between cue type (Location/Gender) and condition (Test/Control) was significant [$F(1,23) = 13.19$, $p = 0.001$; Fig. 12B]. This finding demonstrates that the cluster during the Selective Phase arose from differences in the processes for attending selectively to a talker between Location and Gender trials.

4.3. Discussion

The children tested in Experiment 3 produced strikingly similar patterns of ERPs to those produced by adults in Experiments 1 and 2. The children showed significant differences in activity between the Test and Control conditions during the Preparatory Phase of Location trials (Type-I analyses, Fig. 10A–B) and during the Selective Phase of both Location and Gender Trials (Type-I analyses, Fig. 10A,C–E). They also showed significant differences between Location and Gender trials during both the Preparatory and Selective Phases of the Test Condition (Type-II analyses, Fig. 11A–C).

The difference between Location and Gender trials during the Preparatory Phase was present in both the Test and Control Conditions (Fig. 12A), suggesting that both differences could be attributed to the physical differences between the visual cues. However, the difference during the Selective Phase of the Test Condition differed significantly from the corresponding difference in the Control Condition. The same result was shown by adults in Experiments 1 and 2 and indicates that all three groups evoke significantly different activity depending on whether they are selecting a voice by location or gender.

Unlike the adults in Experiment 2, the children did not display significant activity during the Preparatory Phase of Gender trials (Fig. 10E). Behavioural accuracy was high, showing that the children understood what the cues meant and were able to select the correct talker based on the gender information provided. One explanation would be that children can perform the computation to determine a talker's gender from acoustic evidence including their F0 and formant frequencies in order to select that talker, but they cannot prepare in advance to select those values. That problem may have been exacerbated by the fact that the children completed only 16 familiarisation trials, due to time constraints, whereas the adults completed 52 familiarisation trials. It is possible that 16 trials were not sufficient for the children to learn the distinguishing characteristics of the two talkers. An alternative explanation would be that children can prepare, but that they differ from one another in the way that they prepare, with the result that they do not display consistent patterns of EEG activity as a group.

5. General Discussion

All three experiments revealed significant preparatory EEG activity in a multi-talker listening task when participants were cued in advance to the *location* of the target talker. This result was shown by significant differences in ERPs between the Test and Control Conditions (Figs. 4, 7, and 10), despite the fact that the stimuli were identical between the conditions during the Preparatory Phase of each trial. Adult listeners (Experiments 1 and 2) displayed preparatory activity for location which started within 50 ms of the full reveal of the visual cue and was sustained for longer than 600 ms during the 1000-ms Preparatory Phase (Table 1). This finding suggests that adults begin to prepare their attention early after a location cue is revealed and utilise preparatory brain activity for a large portion of the available time. The alternative explanation that the activity arises from a difference in arousal between Test and Control Conditions is less likely because similar activity was not evoked consistently on trials in which the gender of the target talker was cued.

Together, the results of Experiments 1 and 2 suggest that adults are able to attend to a talker based on a cue for gender, but prepare to a greater extent if the specific talker is known in advance. When the specific talker was known in advance (Experiment 2), preparation for gender started within 100 ms of the full reveal of the visual cue, was sustained for 300 ms, and was observed at similar scalp locations as preparation for the location of the target talker (Fig. 7).

Children aged 7–13 years (Experiment 3) displayed similar patterns of brain activity during the Preparatory and Selective Phases as adults (Experiments 1 and 2). Overall, the groups showed similarities both in the timing of significant differences and in the scalp locations at which significant differences occurred (Table 1). Although, as a group, the children did not show consistent evidence of preparation for a target talker's gender, unlike the adults in Experiment 2.

5.1. Preparation by location or gender

Previous experiments have demonstrated improved speech intelligibility for a target talker who speaks in a mixture of competing talkers when participants know in advance the spatial location (Ericson *et al.* 2004; Kidd *et al.* 2005; Best *et al.*, 2007; Best *et al.*, 2009) or the identity (Kitterick, Bailey,

& Summerfield, 2010) of the target talker, compared to when these characteristics are not known in advance. The current results are consistent with the idea that listeners take advantage of an advance cue and may utilise preparatory brain activity to improve speech intelligibility. This brain activity might reflect focused attention to the cued location or gender before the talkers begin to speak, which might help participants ignore talkers at unattended locations or talkers who possess different voice characteristics to the target talker. In some everyday situations, a talker's location and gender can be identified from visual information and, therefore, this information may be utilised regularly to improve speech intelligibility in adverse listening conditions.

Preparatory activity was found consistently when participants were cued to the location of the target talker, but not when participants were cued to the talker's gender. In Experiment 1, where there were several possible male and female talkers to which the gender cue could refer, participants did not show temporally specific preparation that was evoked in location trials (Fig. 4). In Experiment 2, on the other hand, where there was only one male and one female talker, the cue for gender was also a cue for the identity of the talker and so might have provided participants with the opportunity to prepare their attention for the particular F0 range and vocal tract length of the target talker. Consistent with this idea, the results from Experiment 2 showed preparatory EEG activity on Gender trials that was similar to that evoked on Location trials (Fig. 7). The result is compatible with the finding of better intelligibility when participants know the identity of an upcoming target talker than when they do not (Kitterick *et al.*, 2010). The idea that differences in ERPs between Experiments 1 and 2 result from differences in the specificity of evidence for gender is also consistent with previous experiments that show that the amplitudes of ERPs are affected by the predictability of an attribute for which a stimulus cues (e.g. Horvath, Sussman, Winkler, & Schröger, 2011; Sussman, Winkler, & Schröger, 2003).

5.2. Time-course of preparation

A previous MEG experiment (Lee *et al.*, 2013) using a similar design found evidence of preparatory attention between 600 and 1000 ms after the onset of their visual cues, which is consistent with the timing of the later activity that we observed during the Preparatory Phase. However, Lee *et al.* did not analyse activity before 600 ms. The current experiments demonstrate that participants also evoke preparatory brain activity within 50 ms of the reveal of the visual cue.

Early effects of attention, with latencies less than 50 ms after stimulus onset, have been observed in previous experiments when participants selectively attended to acoustical stimuli (e.g. Woldorff & Hillyard, 1991). The time course of preparatory activity observed in response to the visual cue started with a similar latency as the P20-50 component reported by Woldorff and Hillyard, but also overlapped with the time range of the N1 and P3 components (typically occurring at approximately 100 and 300 ms, respectively), which are known to be sensitive to attention (e.g. Hillyard *et al.*, 1973; Hink & Hillyard, 1976). This overlap can be observed in the ERP diagrams displayed in Figs. 4, 7, and 10. The use of a Spatio-Temporal Cluster-Based Permutation analysis in the current study not only provided information on differences between conditions at such well-established peaks in the ERP waveform, but also provided estimates of how long these differences lasted. The results suggest that participants prepare early for location and gender attributes and sustain preparatory attention for a large portion of the available time.

5.3. Selection by location or gender

Consistent differences between the Test and Control Conditions occurred during the Selective Phases of Location and Gender trials (Table 1), although may have resulted from differences in the acoustical stimuli that were presented in the Test and Control conditions. The acoustical stimuli in the Control Condition were designed to have the same overall energy and gross fluctuations in amplitude as the pairs of sentences in the Test Condition. However, eliminating cues for location and gender meant that the acoustical stimuli differed in spectral detail, temporal fine structure, and inter-aural differences in level and timing.

Of greater interest, consistent differences were found *between* Location and Gender trials during the Selective Phase of the Test Condition in all three experiments, even though the acoustical stimuli were identical for Location and Gender trials. It was also possible to rule out the explanation that physical aspects of the *visual* stimuli were responsible for these differences (Figs. 6, 9, and 12). Rather, the activity is likely to reflect differences in the mechanisms that participants use to select a talker based on their location or gender. Differential activity between Location and Gender trials began more than 350 ms after the onset of acoustical stimuli and lasted up to 1500 ms (Figs. 5, 8, and 11). In these experiments, the first portion of the sentence did not contain key words that participants were required to report. Rather, the key words occurred towards the end of each sentence. Thus, the long latency of ERPs is consistent with the interpretation that participants focussed attention on the target talker to the greatest extent at the time in the sentence at which the key words were spoken. Another possible explanation for the long latency of ERPs is that the ability to separate auditory streams is thought to build up over time (Deike, Heil, Böckmann-Barthel, & Brechmann, 2012; Moore & Gockel, 2012). Thus, participants might not have separated the mixture of talkers successfully at the beginning of the sentence and were, instead, only able to direct selective attention to the target talker towards the end of the sentence.

5.4. Domain-general and cue-specific effects

5.4.1. Preparatory Phase

The finding that ERPs were similar during the Preparatory Phases of Location and Gender trials in Experiment 2 (Table 1) provides evidence for domain-general preparatory attention, which is consistent with the fMRI results reported by Hill and Miller (2010). They reported overlapping activity in a left-dominant fronto-parietal network in response to a visual cue for location or F0 before three talkers started speaking. Given similarity between the design of the current experiments and that of Hill and Miller, it is likely that the ERPs that occurred during location and gender trials in the current experiments arose due to activity in a similar fronto-parietal network as was identified by Hill and Miller.

Domain-general preparatory activity may be underpinned by greater cortical excitability in cortical networks that are relevant for attending to both location and gender (He & Raichle, 2009; O'Connell *et al.*, 2009). Since visual inspection of the data showed low frequency activity throughout the trial, we reanalysed the data with a lower high-pass filter (0.1 Hz). This analysis revealed sustained differences in ERPs across most of the Preparatory Phase (Supplementary Tables 2–4), which might reflect slow cortical potentials underpinned by fluctuations in cortical excitability (Birbaumer, 1999; Bosch, Mecklinger, & Friederici, 2001; Elbert, 1993).

The comparison between Location and Gender trials aimed to reveal whether there was evidence for cue-specific processing, as reported by Hill and Miller (2010) and Lee *et al.* (2013). A consistent difference between Location and Gender trials was observed early during the Preparatory Phase of the Test Condition (Figs. 5, 8, and 11), although similar differences occurred in the Control Condition (Figs. 6, 9, and 12). Therefore, this result likely reflects differences in physical attributes of the visual cues between Location and Gender trials, such as luminance, structural complexity, or differences in the cognitive processes evoked by animate (human stick figures) and inanimate (chevron) cues, rather than differences in attentional processing of the cues (Table 2). Although the visual cues presented by Hill and Miller and Lee *et al.* had higher similarity than the cues presented in the current experiments, the experiments of Hill and Miller and Lee *et al.* did not rule out the explanation that differences in the orientation of the visual cues contributed to differences in brain activity. It is, therefore, possible that activity reported in their experiments reflect a combination of activity evoked by physical aspects of the visual cues and attentional activity evoked by those cues.

Alternatively, the absence of cue-specific preparatory activity in the current experiments might be because detection of significant cue-specific activity relied on a second-order comparison (in which differences between Location and Gender trials in the Test Condition were required to be significantly larger than in the Control Condition). This aspect of the design would reduce the statistical power for detecting cue-specific activity.

Another possible explanation for the absence of cue-specific preparatory activity is that our task may have been too easy for normally-hearing adults. Consequently, participants would have gained no benefit from preparing their attention differently for location and gender before the talkers started speaking. Additional activity may be evoked when preparation is necessary for accurate speech intelligibility and, possibly, the magnitude of preparatory brain activity may depend on the difficulty of the task, since previous experiments relate the magnitude of ERPs during selective attention to task difficulty (e.g. Hillyard, Hink, Schwent, & Picton, 1973; Sabri, Liebenthal, Waldron, Medler, & Binder, 2006).

5.4.2. Selective Phase

During the Selective Phase of the present experiments, there was evidence for consistent cue-specific activity that could not be explained by differences in the visual cues (Table 2). This finding is consistent with the results of Hill and Miller (2010) and Lee *et al.* (2013), who both found significant differences in brain activity when participants selectively attended to a talker, depending on whether participants received information about the talker's spatial location or their F0. However, it was necessary for Hill and Miller to select a high-performing sub-set of their participants in order to detect cue-specific activity. In the current experiments, both children and adults achieved high (> 85%) accuracy, which might have contributed to consistent observations of cue-specific activity during the Selective Phase of the three experiments.

Cue-specific activity is likely to be mediated by activity in different neural generators. There is consistent evidence that "what" and "where" processing occurs in dorsal and ventral pathways, respectively (Adriani *et al.*, 2003; Ahveninen *et al.*, 2006; Alain, Arnott, Hevenor, Graham, & Grady, 2001; Clarke & Thiran, 2004; Leavitt, Molholm, Gomez-Ramirez, & Foxe, 2011; Warren & Griffiths, 2003). For example, Ahveninen *et al.* (2006) presented Finnish vowel sounds from two possible locations: 0 degrees azimuth (straight ahead) or 45 degrees to the right. They presented two sequential vowels, which were either identical or differed in either spatial location or phonetic identity. They measured brain activity using fMRI and MEG when participants attended to spatial or phonetic attributes of the vowels. Regions specialised for spatial processing, such as posterior temporal cortex and posterior parietal regions, displayed significantly greater activity when attending to location; whereas attending to phoneme identity increased activity in anterior and superior temporal cortex. Although it is not possible from the current results to localise the neural generators of activity with high spatial precision, the results are consistent with the idea that attending to different attributes of speech produces activity in different areas of the brain.

5.5. Differences between adults and children

The similarities between ERPs evoked by children and adults were more striking than the differences. The groups showed similarities both in the timing of significant differences and in the scalp locations at which significant differences occurred (Table 1). These results are compatible with reports that children, like adults, achieve higher accuracy of speech intelligibility in noisy listening environments when they are cued to a talker in advance of them speaking compared to when they receive no advance cue (Dhamani *et al.*, 2013). The results of Experiment 3 extend those results by showing that children aged 7–13 years display similar patterns of brain activity as adults during multi-talker listening.

The main differences between children and adults were that the children displayed fewer significant clusters and their clusters generally had shorter durations (Table 1). There are at least three possible explanations for this finding that cannot be distinguished here: (1) Children display consistently weaker preparatory and selective attention than adults; (2) Preparatory and selective attention is variable amongst children, such that some children engage preparatory and selective attention but others do not; or (3) Given that the children contributed fewer trials than the adults, poorer signal-to-noise ratio may make it more difficult to detect preparatory and selective attention in children than in adults.

5.6. Considerations for future research

Overall, these experiments demonstrate the consistency of the spatio-temporal cluster-based permutation method for analysing ERPs in adults and children. Given that the type of clusters to which the analysis is most sensitive depends on how the data are filtered, we considered how the EEG filter settings affected the timing of the clusters observed in these experiments. In a series of post-hoc analyses, we found largely consistent results using high-pass filters between 0.2 and 0.4 Hz (Supplementary Tables 2–4). Thus, the results of the cluster-based permutation analysis were relatively stable across a range of high-pass filter settings.

It is possible that comparisons between the Test and Control conditions were confounded by order effects, since the Control Condition was always presented before the Test Condition. The Control Condition was presented first to measure EEG responses evoked by the visual stimuli *before* participants had learnt the association between the visual cues and the acoustical stimuli; it was, therefore, a necessary feature of the current within-subjects design. Given that the current experiments indicate that preparatory attention could potentially be indexed by comparisons between the Control and Test Conditions used in these experiments, future studies seeking to examine these effects may wish to control for order effects by using a between-subjects design in which each group is either assigned the Test or Control Condition.

Experiments that have analysed cortical rhythms when participants are asked to detect acoustical target stimuli at a cued spatial location show increased alpha power over auditory and parieto-occipital cortex, which occurs ipsilateral to the cued location (Banerjee, Snyder, Molholm, & Foxe, 2011; Müller & Weisz, 2012) and may reflect the suppression of distracting information (for reviews, see Foxe & Snyder, 2011, and Strauß, Wöstmann, & Obleser, 2014). Therefore, as a next step, it would be interesting to examine the role of oscillatory activity on advance cues for location and gender in the current experiments. Effects similar to those reported by Banjeree *et al.* and Müller & Weisz are unlikely to be visible in the current results, since conditions in which participants attended to left and right locations were averaged when conducting the analyses. Nevertheless, we would expect to observe similar effects if we compared attend-left with attend-right trials. It would also be interesting to examine whether oscillatory activity differs between location and gender trials and whether, on gender trials, oscillatory activity differs when participants attend to the male or female talker.

5.7. Conclusions

Young adults (aged 18-27 years with normal hearing) and typically-developing children (aged 7-13 years) show consistent evidence of preparatory brain activity when they are cued visually to the location of an upcoming target talker in a mixture of two talkers. Preparatory EEG activity in adults starts less than 50 ms after the cue is fully revealed. Activity is then sustained for more than 600 ms. Preparatory activity in children starts later and lasts for a shorter time. Adults, but not children, also display preparatory brain activity when they know the gender of an upcoming talker, but only when the cue for gender predicts the specific identity of the target talker. Once the talkers have started to speak, both groups display significant differences in brain activity depending on whether they are selecting the target talker by location or gender. Considered overall, young adults and typically-developing children display evidence of striking similarities in brain activity both in preparation for, and in the execution of, multi-talker listening.

The experiments achieved the goal of validating a technique that can be used both with adults and children to study the timing of the deployment of attention in selecting one talker from a mixture of talkers. In the future, the technique might be applied to populations where attention is suspected of being atypical. The fact that the technique reveals evidence of preparatory attention—measured before acoustic stimuli are presented—opens up the possibility of detecting abnormalities in preparatory auditory attention independently of effects of impairments in peripheral auditory processing.

6. Acknowledgements

EH was supported by a studentship from the Goodricke Appeal Fund. PTK was supported by a Value-in-People Award to the University of York from the Wellcome Trust.

7. References

Adriani, M., Maeder, P. P., Meuli, R., Thiran, A. B., Frischknecht, R., Villemure, J.-G., … Clarke, S. (2003). Sound recognition and localization in man: Specialized cortical networks and effects of acute circumscribed lesions. *Experimental Brain Research*, *153*(4), 591–604. http://doi.org/10.1007/s00221-003-1616-0

Ahveninen, J., Jääskeläinen, I. P., Raij, T., Bonmassar, G., Devore, S., Hämäläinen, M. S., … Belliveau, J. W. (2006). Task-modulated "what" and "where" pathways in human auditory cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(39), 14608–13. http://doi.org/10.1073/pnas.0510480103

Alain, C., Arnott, S. R., Hevenor, S., Graham, S., & Grady, C. L. (2001). "What" and "where" in the human auditory system. *Proceedings of the National Academy of Sciences of the United States of America*, *98*(21), 12301–6. http://doi.org/10.1073/pnas.211209098

Best, V., Marrone, N., Mason, C. R., Kidd, G., & Shinn-Cunningham, B. G. (2009). Effects of sensorineural hearing loss on visually guided attention in a multitalker environment. *Journal of the Association for Research in Otolaryngology*, *10*(1), 142–9. http://doi.org/10.1007/s10162-008-0146-7

Best, V., Ozmeral, E. J., & Shinn-Cunningham, B. G. (2007). Visually-guided attention enhances target identification in a complex auditory scene. *Journal of the Association for Research in Otolaryngology, 8*(2), 294–304. doi:10.1007/s10162-007-0073-z

Birbaumer, N. (1999). Slow cortical potentials: Plasticity, operant control, and behavioral effects. *The Neuroscientist*, *5*(2), 74–78. http://doi.org/10.1177/107385849900500211

Bosch, V., Mecklinger, A., & Friederici, A. D. (2001). Slow cortical potentials during retention of object, spatial, and verbal information. *Cognitive Brain Research*, *10*(3), 219–37. http://doi.org/10.1016/S0926-6410(00)00040-9

Cameron, S., Brown, D., Keith, R., Martin, J., Watson, C., & Dillon, H. (2009). Development of the North American Listening in Spatialized Noise-Sentences test (NA LiSN-S): Sentence equivalence, normative data, and test-retest reliability studies. *Journal of the American Academy of Audiology*, *20*(2), 128–46. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/19927676

Cameron, S., & Dillon, H. (2007). Development of the Listening in Spatialized Noise-Sentences test (LISN-S). *Ear and Hearing*, *28*(2), 196–211. http://doi.org/10.1097/AUD.0b013e318031267f

Caramazza, A., & Shelton, J. R. (1998). Domain-specific knowledge systems in the brain the animate-inanimate distinction. *Journal of Cognitive Neuroscience*, *10*(1), 1–34. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9526080

Clarke, S., & Thiran, A. B. (2004). Auditory neglect: what and where in auditory space. *Cortex*, *40*(2), 291–300. http://doi.org/10.1016/S0010-9452(08)70124-2

Deike, S., Heil, P., Böckmann-Barthel, M., & Brechmann, A. (2012). The build-up of auditory stream segregation: A different perspective. *Frontiers in Psychology*, *3*, 1–7. http://doi.org/10.3389/fpsyg.2012.00461

Dhamani, I., Leung, J., Carlile, S., & Sharma, M. (2013). Switch attention to listen. *Scientific Reports*, *3*, 1297. http://doi.org/10.1038/srep01297

Downing, P. E., Chan, A. W.-Y., Peelen, M. V, Dodds, C. M., & Kanwisher, N. (2006). Domain specificity in visual cortex. *Cerebral Cortex*, *16*(10), 1453–61. http://doi.org/10.1093/cercor/bhj086

Elbert, T. (1993). *Slow cortical potentials reflect the regulation of cortical excitability*. Springer US. Retrieved from http://physrev.physiology.org/content/70/1/1.short

Foxe, J. J., & Snyder, A. C. (2011). The role of alpha-band brain oscillations as a sensory suppression mechanism during selective attention. *Frontiers in Psychology*, *2*, 1–13. http://doi.org/10.3389/fpsyg.2011.00154

He, B. J., & Raichle, M. E. (2009). The fMRI signal, slow cortical potential and consciousness. *Trends in Cognitive Sciences*, *13*(7), 302–9. http://doi.org/10.1016/j.tics.2009.04.004.The

Hilbert, D. (1912). *Grundzüge einer Allgemeinen Theorie der linearen Integralgleichungen (Basics of a general theory of linear integral equations)*. Leizpig: Teubner.

Hill, K. T., & Miller, L. M. (2010). Auditory attentional control and selection during cocktail party listening. *Cerebral Cortex*, *20*(3), 583–90. http://doi.org/10.1093/cercor/bhp124

Hillyard, S. A., Hink, R. F., Schwent, V. L., & Picton, T. W. (1973). Electrical signs of selective attention in the human brain. *Science*, *182*(4108), 177–80. http://doi.org/10.1126/science.182.4108.177

Horvath, J., Sussman, E., Winkler, I., & Schröger, E. (2011). Preventing distraction: Assessing stimulus-specific and general effects of the predictive cueing of deviant auditory events. *Biological Psychology*, *87*(1), 35–48. http://doi.org/10.1016/j.biopsycho.2011.01.011

Kitterick, P. T., Bailey, P. J., & Summerfield, A. Q. (2010). Benefits of knowing who, where, and when in multi-talker listening. *The Journal of the Acoustical Society of America*, *127*(4), 2498–508.

http://doi.org/10.1121/1.3327507

Leavitt, V. M., Molholm, S., Gomez-Ramirez, M., & Foxe, J. J. (2011). “What” and “where” in auditory sensory processing: A high-density electrical mapping study of distinct neural processes underlying sound object recognition and sound localization. *Frontiers in Integrative Neuroscience*, *5*, 1–15. http://doi.org/10.3389/fnint.2011.00023

Lee, A. K. C., Rajaram, S., Xia, J., Bharadwaj, H., Larson, E., Hämäläinen, M. S., & Shinn-Cunningham, B. G. (2013). Auditory selective attention reveals preparatory activity in different cortical regions for selection based on source location and source pitch. *Frontiers in Neuroscience*, *6*, 1–9. http://doi.org/10.3389/fnins.2012.00190

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods*, *164*(1), 177–90. http://doi.org/10.1016/j.jneumeth.2007.03.024

Moore, B. C. J. (2008). The role of temporal fine structure processing in pitch perception, masking, and speech perception for normal-hearing and hearing-impaired people. *Journal of the Association for Research in Otolaryngology*, *9*(4), 399–406. http://doi.org/10.1007/s10162-008-0143-x

Moore, B. C. J., & Gockel, H. E. (2012). Properties of auditory stream formation. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1591), 919–31. http://doi.org/10.1098/rstb.2011.0355

O’Connell, R. G., Dockree, P. M., Robertson, I. H., Bellgrove, M. A, Foxe, J. J., & Kelly, S. P. (2009). Uncovering the neural signature of lapsing attention: electrophysiological signals predict errors up to 20 s before they occur. *The Journal of Neuroscience*, *29*(26), 8604–11. http://doi.org/10.1523/JNEUROSCI.5967-08.2009

Sabri, M., Liebenthal, E., Waldron, E. J., Medler, D. A., & Binder, J. R. (2006). Attentional modulation in the detection of irrelevant deviance: A simultaneous ERP/fMRI study. *Journal of Cognitive Neuroscience*, *18*(5), 689–700.

Strauß, A., Wöstmann, M., & Obleser, J. (2014). Cortical alpha oscillations as a tool for auditory selective inhibition. *Frontiers in Human Neuroscience*, *8*, 350. http://doi.org/10.3389/fnhum.2014.00350

Studebaker, G. A. (1985). A “Rationalized” Arcsine Transform. *Journal of Speech Language and Hearing Research*, *28*(3), 455. http://doi.org/10.1044/jshr.2803.455

Sussman, E., Winkler, I., & Schröger, E. (2003). Top-down control over involuntary attention switching in the auditory modality. *Psychonomic Bulletin & Review*, *10*(3), 630–7.

http://doi.org/10.3758/BF03196525

Vaillancourt, V., Laroche, C., Giguère, C., & Soli, S. D. (2008). Establishment of age-specific normative data for the canadian French version of the hearing in noise test for children. *Ear and Hearing*, *29*(3), 453–66. http://doi.org/10.1097/01.aud.0000310792.55221.0c

Warren, J. D., & Griffiths, T. D. (2003). Distinct mechanisms for processing spatial sequences and pitch sequences in the human auditory brain. *The Journal of Neuroscience*, *23*(13), 5799–804. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/12843284

Wightman, F. L., Kistler, D. J., & Brungart, D. S. (2006). Informational masking of speech in children: Auditory-visual integration. *The Journal of the Acoustical Society of America*, *119*(6), 3940–9.

Wightman, F. L., Kistler, D. J., & O'Bryan, A. (2010). Individual differences and age effects in a dichotic informational masking paradigm. *The Journal of the Acoustical Society of America,* *128*(1), 270–279. doi:10.1121/1.3436536

Yamaguchi, S., Tsuchiya, H., & Kobayashi, S. (1994). Electroencephalographic activity associated with shifts of visuospatial attention. *Brain*, *117*, 553–62. http://doi.org/10.1093/brain/117.3.553

Figure 1. [one column, B&W] Layout of loudspeakers (blue squares) and visual display unit (grey rectangle) relative to a participant's head.

Figure 2. [one column, B&W] Visual cues for location (A,B) and gender (C,D). A visual composite stimulus (E) was created by overlaying the four visual cues.

Figure 3. [two columns, colour] Schematic showing the trials structure in (**A**) the Test Condition and (**B**) the Control Condition. Stimuli for an example trial are displayed below, with an example of the visual stimuli (left), acoustical stimuli (centre) and response buttons (right).

Figure 4. [two columns, colour] (Continued on next page). Experiment 1: Results from Type-I Spatio-temporal Cluster-based Permutation Analyses for Location (**A** to **F**) and Gender (**G** to **J**) trials. (**A** and **G**) Coloured rectangles indicate the time-span of significant (p < 0.05) clusters of activity. For clarity in plotting, time on the x-axis is relative to the onset of the acoustical stimuli (i.e. relative to the start of the Selective Phase), rather than relative to the start of the phase in which the cluster occurred (which is how the cluster latencies are described in the main text). Rows on the y-axis show separate significant clusters. For clusters plotted as red rectangles, the average amplitude, over all space-by-time points in the cluster, was more positive in the Test Condition than the Control Condition. For clusters plotted as blue rectangles, the average amplitude was more negative in the Test Condition than the Control Condition. Further information about each cluster is displayed in (**B** to **F** and **H** to **J**) where, for each cluster, the topographical map shows the electrodes that contributed to the cluster, the graph shows the ERPs averaged across those electrodes over the time course of the trial, and the time-span of the cluster is indicated by a dashed rectangle.

Figure 5. [two columns, colour] Experiment 1: Results of the Type-II Spatio-temporal Cluster-based Permutation Analyses. This analysis contrasted Location and Gender trials in the Test Condition. (**A**) Coloured rectangles indicate the time-span of significant (p < 0.05) clusters of activity. Time on the x-axis is relative to the onset of the acoustical stimuli. Rows on the y-axis show separate significant clusters. For clusters plotted as red rectangles, the average amplitude, over all space-by-time points in the cluster, was more positive on Location trials than Gender trials. For clusters plotted as blue rectangles, the average amplitude was more negative on Location trials than Gender trials. Further information about each cluster is displayed in (**B**)-(**F**) where, for each cluster, the topographical map shows the electrodes that contributed to the cluster, the graph shows the ERPs averaged across those electrodes over the time course of the trial, and the time-span of the cluster is indicated by a dashed rectangle.

Figure 6. [two columns, colour] Experiment 1: Comparison of differences in the amplitude of ERPs between Location and Gender trials in the Test and Control Conditions for each significant Type-II cluster in Experiment 1. This analysis investigated whether differences between Location and Gender trials in the Test Condition were also present in the Control Condition (i.e. investigating whether the clusters could be explained by physical aspects of the visual cues). Clusters are labelled on the left of the figure, with their corresponding electrode topographies. Graphs (**A**)-(**E**) plot the mean amplitude for Location and Gender trials in the Test and Control Conditions, averaged across participants and space-time points. Error bars show 95% within-subjects confidence intervals. Narrow brackets display the significance level of the comparison between Location and Gender trials in the Test and Control Conditions. Wider brackets display the significance level of the two-way interaction (* p < 0.050; ** p < 0.010; *** p < 0.001). Graphs (**F**)-(**J**) display the difference of the differences in Gender and Location trials between the Test and Control conditions in 50-ms time windows repeated every 10 ms within the cluster (right axis) and the uncorrected p-values resulting from a paired-samples t-test

comparing the differences (left axis). The mid-point of each time window relative to the onset of acoustic stimuli is displayed on the x-axis.

Figure 7. [two columns, colour] (Continued on next page). Experiment 2: Results from Type-I Spatio-temporal Cluster-based Permutation Analyses for the Location Condition (**A** to **F**) and the Gender Condition (**E** to **J**). (**A** and **E**) Coloured rectangles indicate the time-span of significant ($p < 0.05$) clusters of activity. Time on the x-axis is relative to the onset of the acoustical stimuli. Rows on the y-axis show separate significant clusters. For clusters plotted as red rectangles, the average amplitude, over all space-by-time points in the cluster, was more positive in the Test Condition than the Control Condition. For clusters plotted as blue rectangles, the average amplitude was more negative in the Test Condition than the Control Condition. Further information about each cluster is displayed in (**B** to **D** and **F** to **J**) where, for each cluster, the topographical map shows the electrodes that contributed to the cluster, the graph shows the ERPs averaged across those electrodes over the time course of the trial, and the time-span of the cluster is indicated by a dashed rectangle.

Figure 8. [two columns, colour] Experiment 2: Results from the Type-II Spatio-temporal Cluster-based Permutation Analysis. This analysis contrasted Location and Gender trials in the Test Condition. (**A**) Coloured rectangles indicate the time-span of significant ($p < 0.05$) clusters of activity. Time on the x-axis is relative to the onset of the acoustical stimuli. Rows on the y-axis show separate significant clusters. For clusters plotted as red rectangles, the average amplitude, over all space-by-time points in the cluster, was more positive on Location trials than Gender trials. For clusters plotted as blue rectangles, the average amplitude was more negative on Location trials than Gender trials. Further information about each cluster is displayed in (**B**)-(**D**). For each cluster, the topographical map shows the electrodes that contribute to the cluster, the graph shows the ERPs averaged across those electrodes over the time course of the trial, and the time-span of the cluster is indicated by a dashed rectangle.

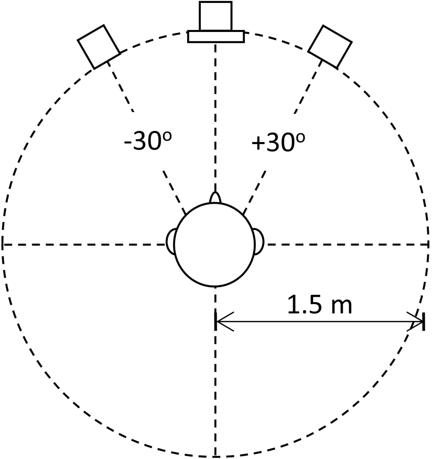
Figure 9. [two columns, colour] Experiment 2: Comparison of differences in the amplitude of ERPs between Location and Gender trials in the Test and Control Conditions for each significant Type-II cluster in Experiment 2. This analysis investigated whether differences between Location and Gender trials in the Test Condition were also present in the Control Condition (i.e. investigating whether the clusters could be explained by physical aspects of the visual cues). Clusters are labelled on the left of the figure, with their corresponding electrode topographies. Graphs (**A**)-(**C**) plot the mean amplitude for Location and Gender trials in the Test and Control Conditions, averaged across participants and space-time points. Error bars show 95% within-subjects confidence intervals. Narrow brackets display the significance level of the comparison between Location and Gender trials in the Test and Control Conditions. Wider brackets display the significance level of the two-way interaction (* $p < 0.050$; ** $p < 0.010$; *** $p < 0.001$). Graphs (**D**)-(**F**) display the difference of the differences in Gender and Location trials between the Test and Control conditions in 50-ms time windows repeated every 10 ms within the cluster (right axis) and the uncorrected *p*-values resulting from a paired-samples *t*-test comparing the differences (left axis). The mid-point of each time window relative to the onset of acoustic stimuli is displayed on the x-axis.

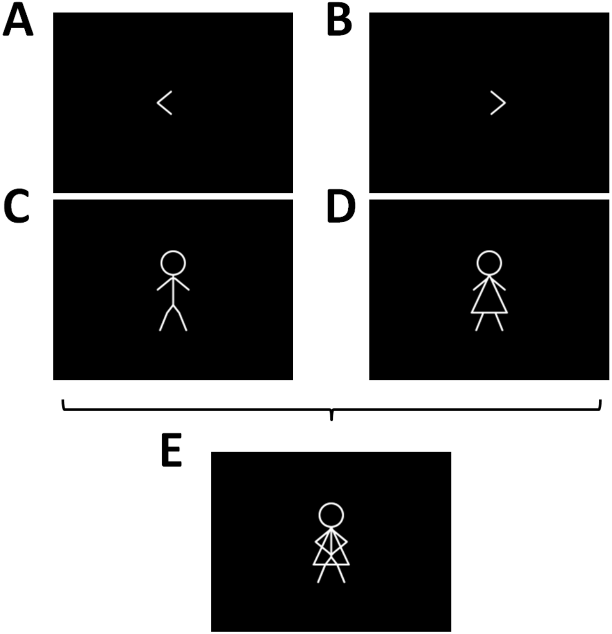
Figure 10. [two columns, colour] (Continued on next page). Experiment 3: Results from Type-I Spatio-temporal Cluster-based Permutation Analyses for Location (**A** to **D**) and Gender (**E** to **G**) trials. (**A** and **E**) Coloured rectangles indicate the time-span of significant ($p < 0.05$) clusters of activity. Time on the x-axis is relative to the onset of the acoustical stimuli. Rows on the y-axis show separate significant clusters. For clusters plotted as red rectangles, the average amplitude, over all space-by-time points in the cluster, was more positive in the Test Condition than the Control Condition. For clusters plotted as blue rectangles, the average amplitude was more negative in the Test Condition

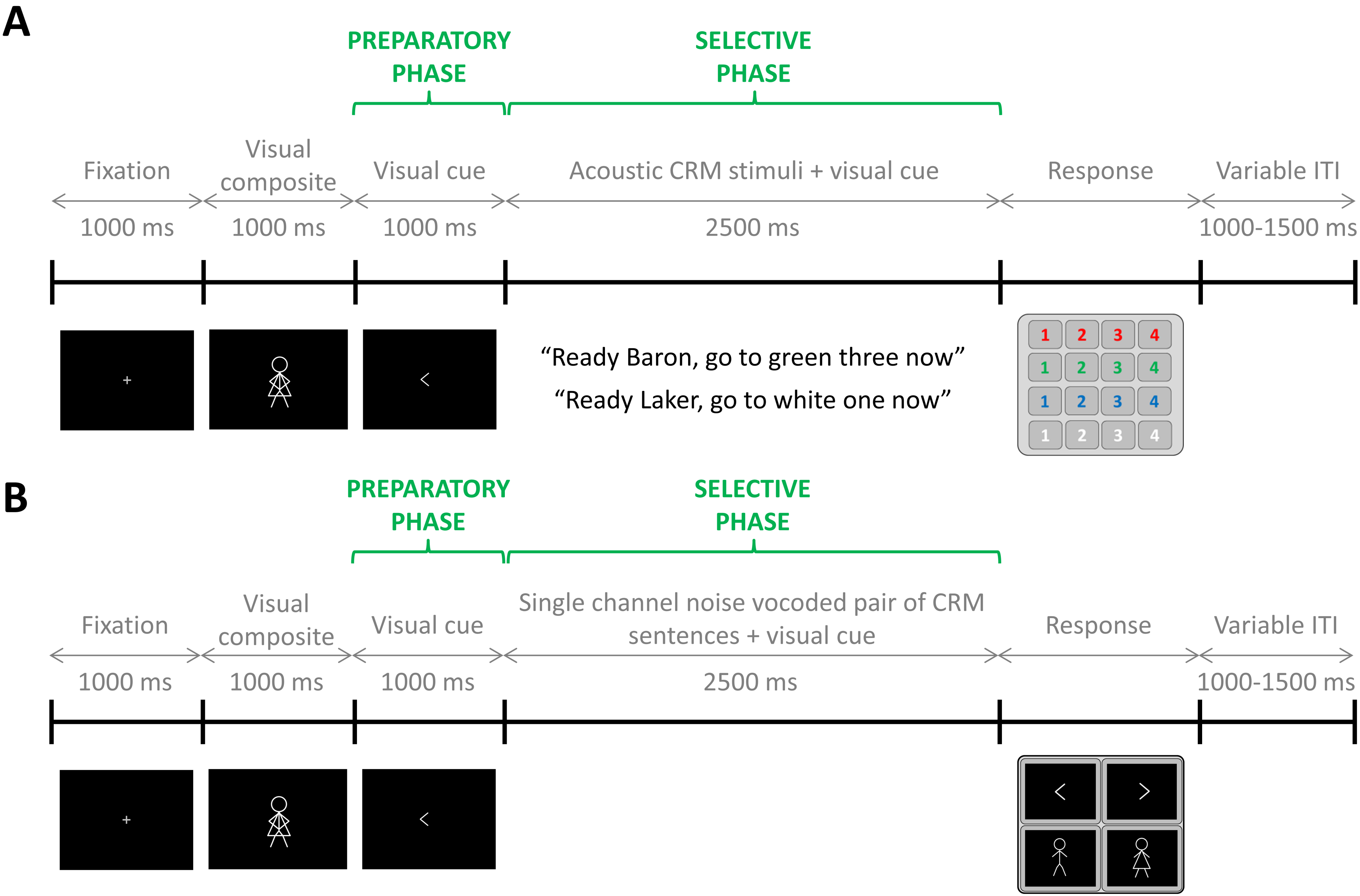
than the Control Condition. Further information about each cluster is displayed in (**B** to **D** and **F** to **G**) where, for each cluster, the topographical map shows the electrodes that contributed to the cluster, the graph shows the ERPs averaged across those electrodes over the time course of the trial, and the time-span of the cluster is indicated by a dashed rectangle.

Figure 11. [two columns, colour] Experiment 3: Results from the Type-II Spatio-temporal Cluster-based Permutation Analysis. This analysis contrasted Location and Gender trials in the Test Condition. (**A**) Coloured rectangles indicate the time-span of significant ($p < 0.05$) clusters of activity. Time on the x-axis is relative to the onset of the acoustical stimuli. Rows on the y-axis show separate significant clusters. For clusters plotted as red rectangles, the average amplitude, over all space-by-time points in the cluster, was more positive on Location trials than Gender trials. For clusters plotted as blue rectangles, the average amplitude was more negative on Location trials than Gender trials. Further information about each cluster is displayed in (**B**)-(**C**). For each cluster, the topographical map shows the electrodes that contribute to the cluster, the graph shows the ERPs averaged across those electrodes over the time course of the trial, and the time-span of the cluster is indicated by a dashed rectangle.

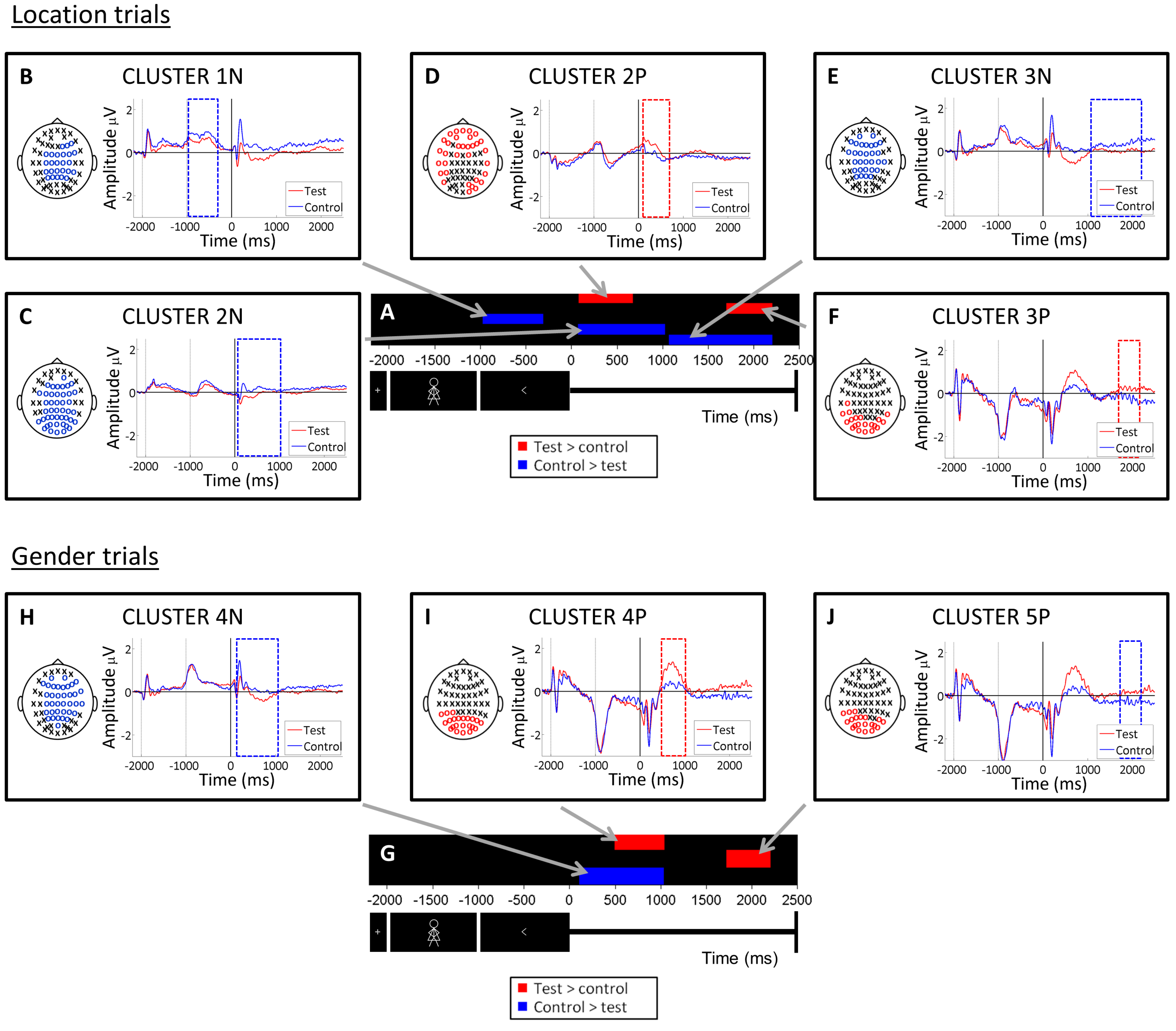
Figure 12. [two columns, colour] Experiment 3: Comparison of differences in the amplitude of ERPs between Location and Gender trials in the Test and Control Conditions for each significant Type-II cluster in Experiment 3. This analysis investigated whether differences between Location and Gender trials in the Test Condition were also present in the Control Condition (i.e. investigating whether the clusters could be explained by physical aspects of the visual cues). Clusters are labelled on the left of the figure, with their corresponding electrode topographies. Graphs (**A**)-(**B**) plot the mean amplitude for Location and Gender trials in the Test and Control Conditions, averaged across participants and space-time points. Error bars show 95% within-subjects confidence intervals. Narrow brackets display the significance level of the comparison between Location and Gender trials in the Test and Control Conditions. Wider brackets display the significance level of the two-way interaction (* $p < 0.050$; ** $p < 0.010$; *** $p < 0.001$). Graphs (**C**)-(**D**) display the difference of the differences in Gender and Location trials between the Test and Control conditions in 50-ms time windows repeated every 10 ms within the cluster (right axis) and the uncorrected *p*-values resulting from a paired-samples *t*-test comparing the differences (left axis). The mid-point of each time window relative to the onset of acoustic stimuli is displayed on the x-axis.

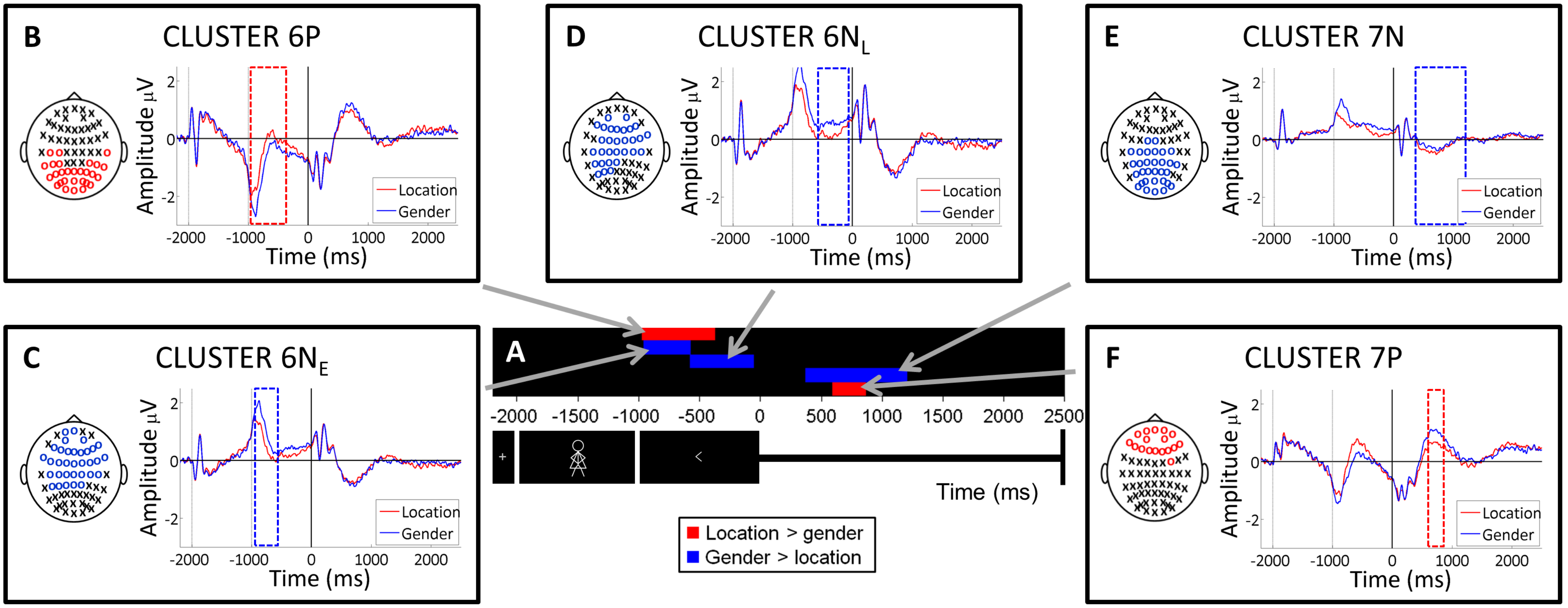

-30º
+30º
1.5 m

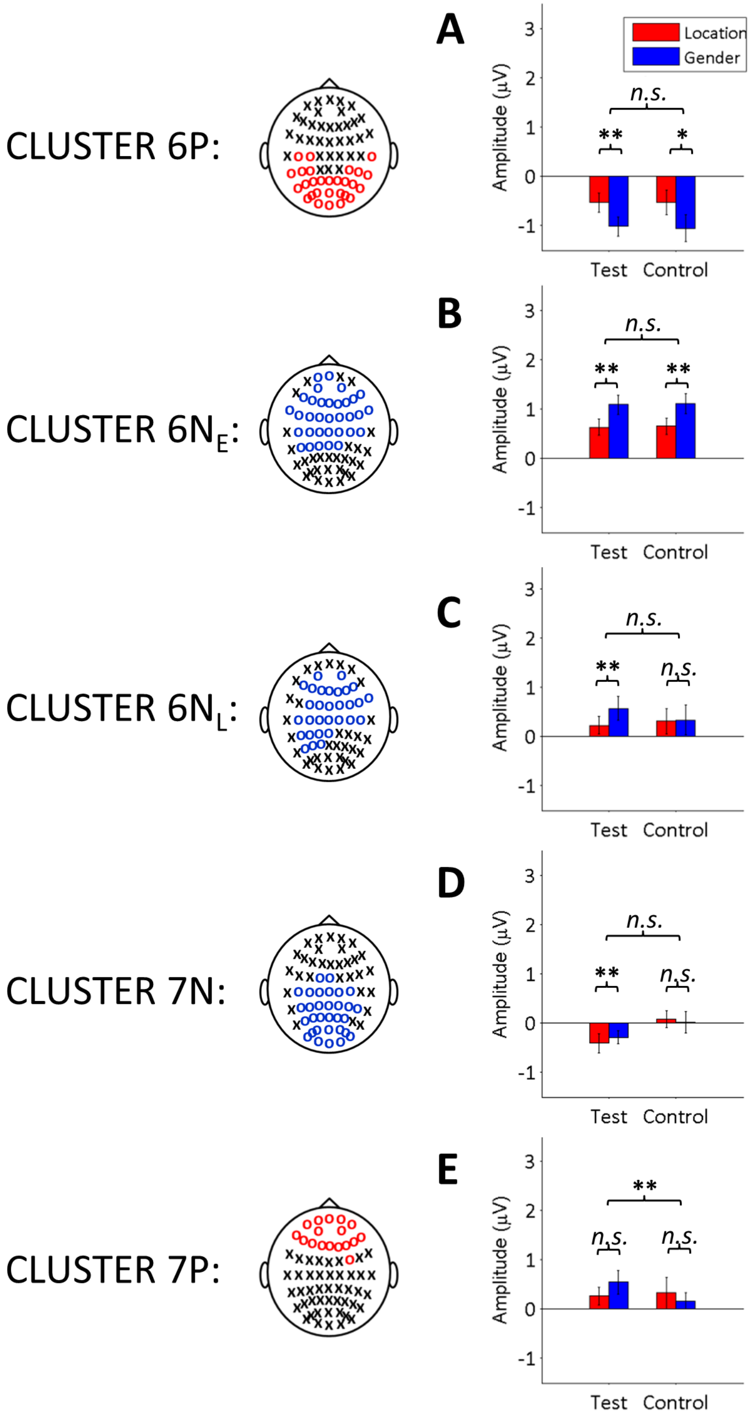

A
B
C
D
E



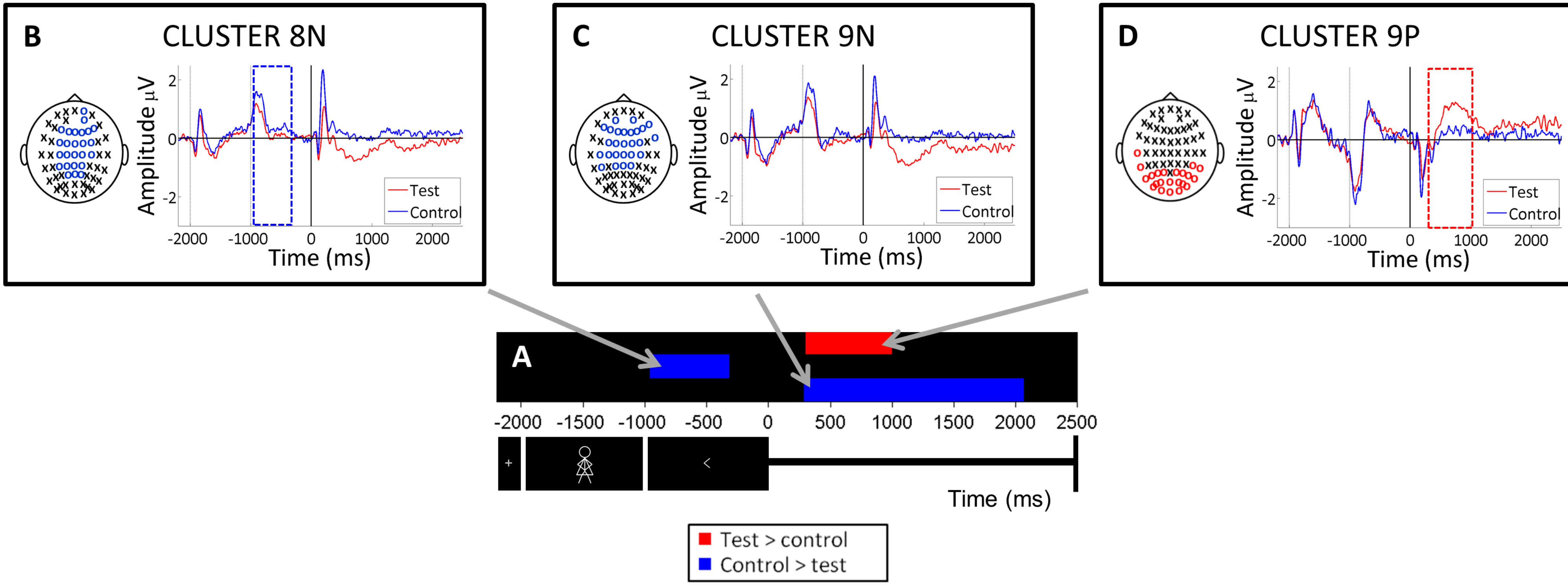
A
PREPARATORY PHASE
SELECTIVE PHASE
Fixation
Visual composite
Visual cue
Acoustic CRM stimuli + visual cue
Response
Variable ITI
1000 ms
1000 ms
1000 ms
2500 ms
1000-1500 ms
“Ready Baron, go to green three now”
“Ready Laker, go to white one now”
B
PREPARATORY PHASE
SELECTIVE PHASE
Fixation
Visual composite
Visual cue
Single channel noise vocoded pair of CRM sentences + visual cue
Response
Variable ITI
1000 ms
1000 ms
1000 ms
2500 ms
1000-1500 ms


Location trials
B
CLUSTER 1N
Amplitude μV
Time (ms)
Test
Control
D
CLUSTER 2P
Amplitude μV
Time (ms)
Test
Control
E
CLUSTER 3N
Amplitude μV
Time (ms)
Test
Control
C
CLUSTER 2N
Amplitude μV
Time (ms)
Test
Control
A
-2000 -1500 -1000 -500 0 500 1000 1500 2000 2500
Time (ms)
Test > control
Control > test
F
CLUSTER 3P
Amplitude μV
Time (ms)
Test
Control
Gender trials
H
CLUSTER 4N
Amplitude μV
Time (ms)
Test
Control
I
CLUSTER 4P
Amplitude μV
Time (ms)
Test
Control
J
CLUSTER 5P
Amplitude μV
Time (ms)
Test
Control
G
-2000 -1500 -1000 -500 0 500 1000 1500 2000 2500
Time (ms)
Test > control
Control > test

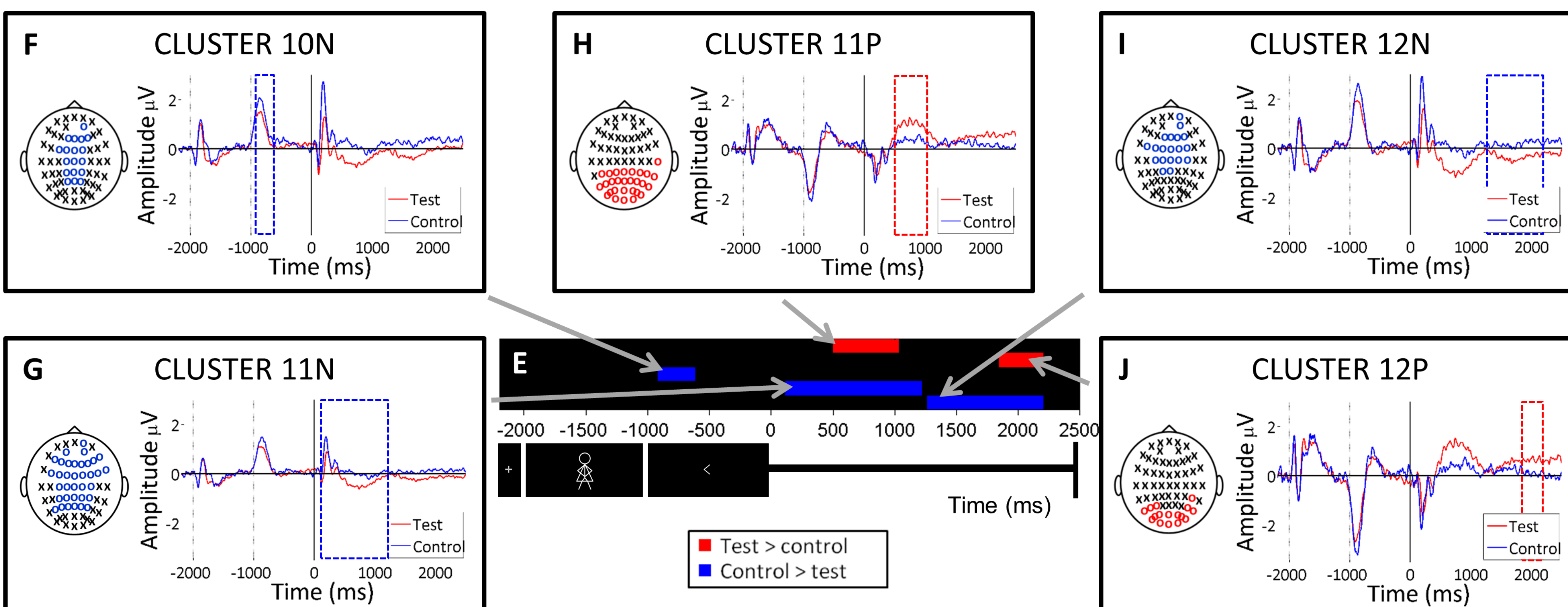

B
CLUSTER 6P
Amplitude µV
Time (ms)
Location
Gender
D
CLUSTER 6N$_L$
Amplitude µV
Time (ms)
Location
Gender
E
CLUSTER 7N
Amplitude µV
Time (ms)
Location
Gender
C
CLUSTER 6N$_E$
Amplitude µV
Time (ms)
Location
Gender
A
-2000 -1500 -1000 -500 0 500 1000 1500 2000 2500
Time (ms)
Location > gender
Gender > location
F
CLUSTER 7P
Amplitude µV
Time (ms)
Location
Gender

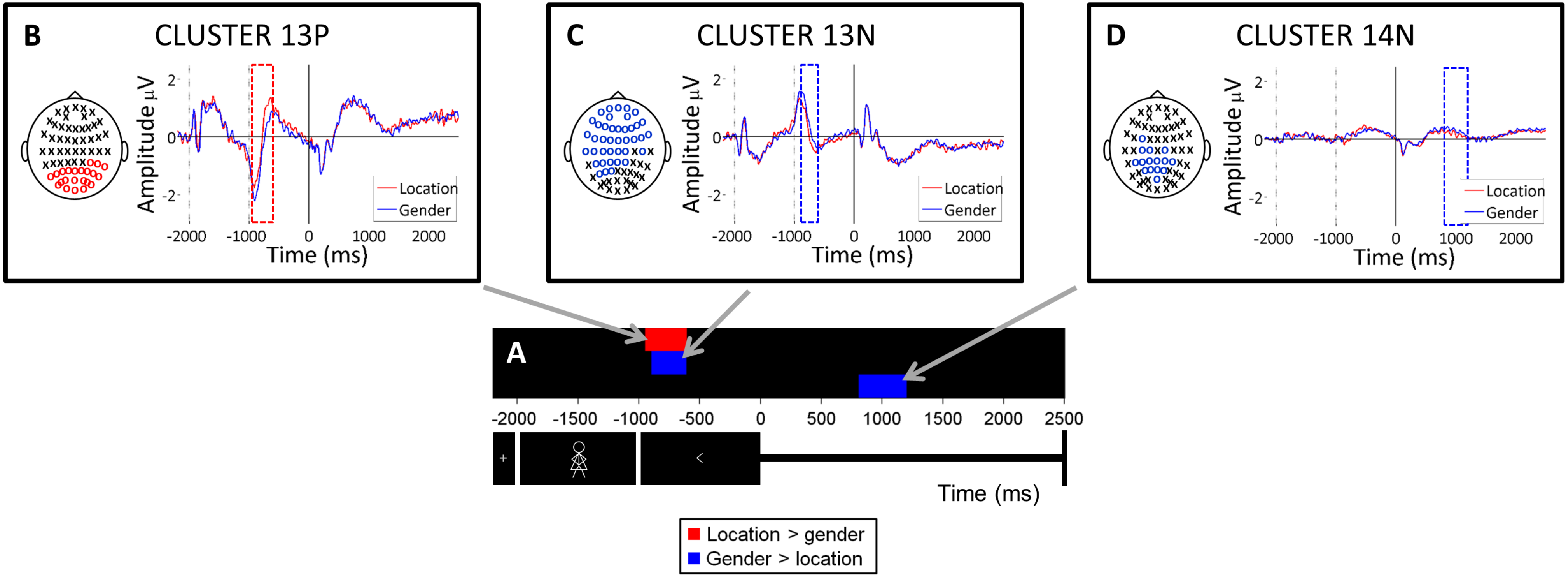

CLUSTER 6P:
CLUSTER 6N$_E$:
CLUSTER 6N$_L$:
CLUSTER 7N:
CLUSTER 7P:
A
B
C
D
E
Location
Gender
Amplitude (μV)
3
2
1
0
-1
n.s.
**
*
Test
Control

Location trials

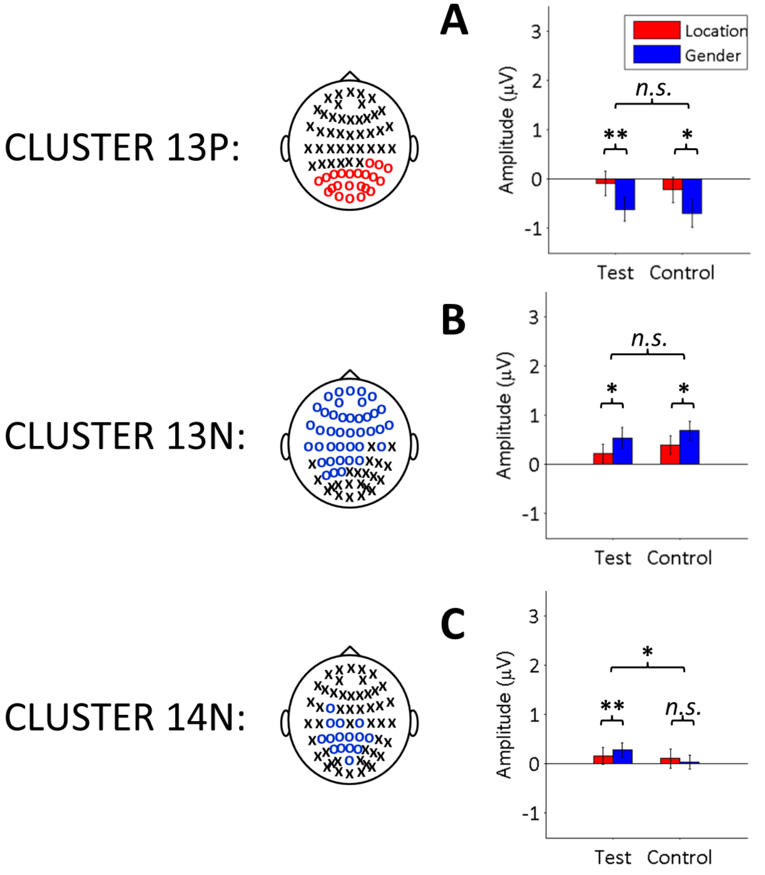


Gender trials

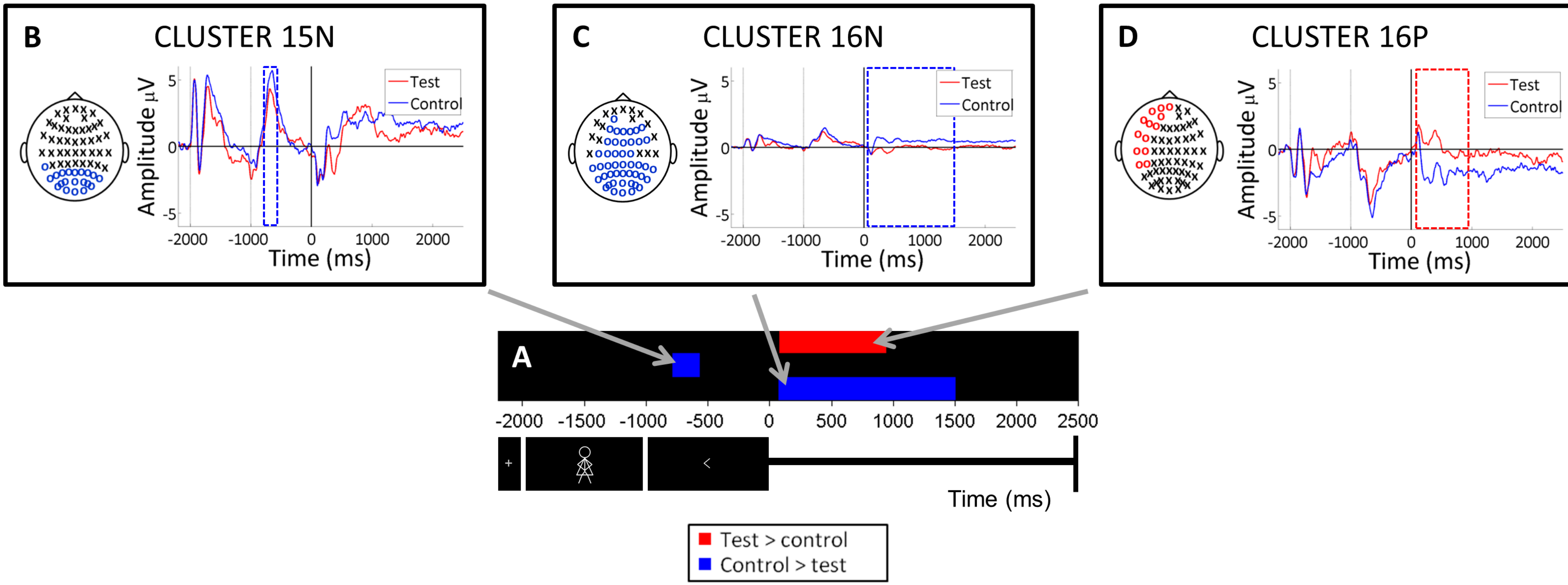




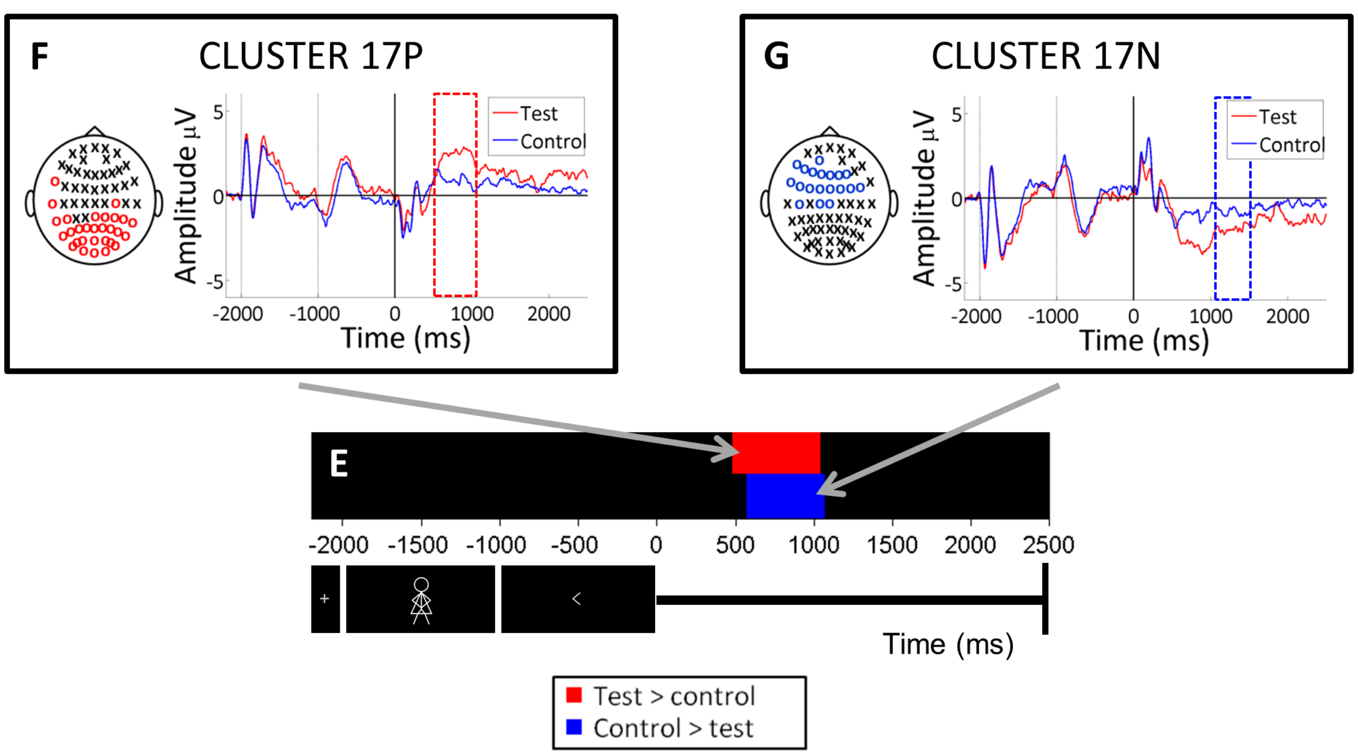
B
CLUSTER 13P
C
CLUSTER 13N
D
CLUSTER 14N
Amplitude μV
Time (ms)
Location
Gender
A
-2000 -1500 -1000 -500 0 500 1000 1500 2000 2500
Time (ms)
Location > gender
Gender > location

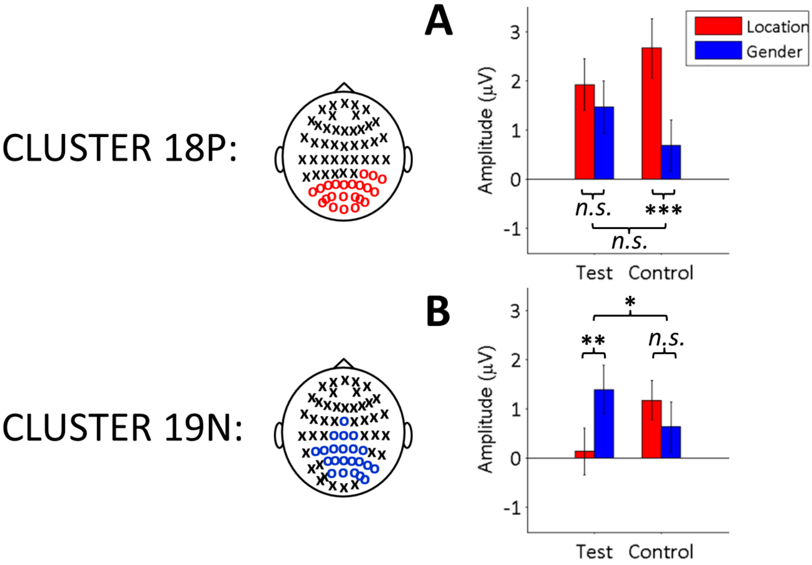

CLUSTER 13P:
CLUSTER 13N:
CLUSTER 14N:
A
B
C
Location
Gender
Amplitude (μV)
n.s.
**
*
Test
Control
3
2
1
0
-1

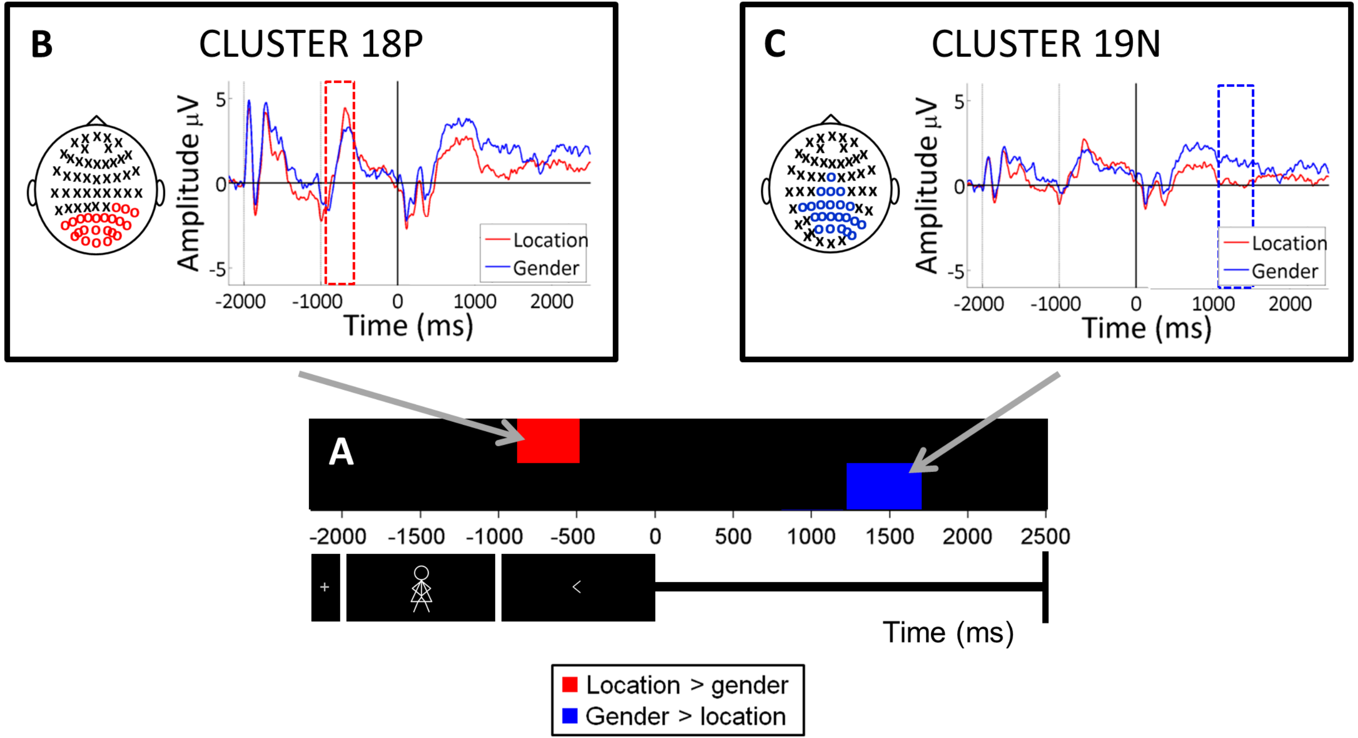
Location trials



Gender trials




CLUSTER 18P:
CLUSTER 19N:
A
B
Location
Gender
Amplitude (μV)
3
2
1
0
-1
n.s.
***
n.s.
*
**
n.s.
Test
Control


B
CLUSTER 18P
Amplitude µV
5
0
-5
-2000
-1000
0
1000
2000
Time (ms)
Location
Gender
C
CLUSTER 19N
Amplitude µV
5
0
-5
-2000
-1000
0
1000
2000
Time (ms)
Location
Gender
A
-2000
-1500
-1000
-500
0
500
1000
1500
2000
2500
+
<
Time (ms)
Location > gender
Gender > location

Supplementary Material

Supplementary Table 1. Summary of key terms and analyses.

Term	Definition
Preparatory Phase	The time interval between the start of the reveal of the visual cue and the start of the acoustic stimuli.
Selective Phase	The time interval between the start and end of the acoustic stimuli.
Type-I Analyses	Type-I analyses aimed to identify processes related to attentional preparation and selection by cues for location *or* gender. These analyses compared ERPs between the Test and Control Conditions, separately for Location and Gender trials. Type-I clusters found in the Preparatory Phase could not arise from sensory or perceptual processes, because the stimuli did not differ between the conditions in this phase. Type-I clusters found in the Selective Phase could arise *either* from differences in attentional activity *or* from differences between the acoustical structure of the Test and Control stimuli.
Type-II Analyses	Type-II analyses aimed to identify differences in ERPs between trials in which participants received cues for location and gender. These analyses compared ERPs between Location and Gender trials in the Test Condition only. Such differences could be evoked *either* by different attentional processes *or* by physical differences between the visual cues.
2 x 2 ANOVAs	To help interpret whether Type-II clusters arose from different attentional processes, these analyses compared ERPs between Location and Gender trials—averaged over the space-by-time-points in the Type-II cluster—between the Test and Control Conditions. A significant two-way interaction was interpreted as indicating that the cluster could not be fully explained by the influence of physical differences in the visual cues between conditions.

Supplementary Table 2. Experiment 1: Clusters ($p < 0.05$) by high-pass filter value (Hz). Values reported in paper (0.25 Hz) are displayed in bold font. Timing of clusters is relative to the onset of acoustical stimuli.

		0.1	0.2	**0.25**	0.3	0.4	0.5
Location	Preparatory	-1000 to -167 ms (central)	-1000 to -719 ms (central) -698 to -314 ms (central)	**-973 to -309 ms (central)**	-1000 to -722 (central) -700 to -313 (central)	-962 to -724 ms (central) -688 to -314 ms (central)	-478 to -313 ms, p = 0.068
		-1000 to 0 ms (posterior)	-534 to -184 ms (posterior)			-148 to 0 ms, p = 0.060	-152 to 0 ms (central)
	Selective	0 to 2200 ms (central)	73 to 987 ms (central) 1247 to 2200 ms (central)	**69 to 1029 ms (posterior + central)** **1072 to 2200 ms (central)**	63 to 907 ms (most of array) 1239 to 2200 ms (central)	64 to 871 ms (posterior + central) 1341 to 1798 ms (central) 1797 to 2200 ms (central)	20 to 867 ms (posterior + central) 1252 to 1798 ms (central) 2026 to 2200 ms (central + anterior)
		280 to 2200 ms (posterior)	449 to 1041 ms (posterior) 1513 to 2200 ms (posterior)	**81 to 671 ms (non-central)** **1696 to 2200 ms (posterior)**	122 to 943 ms (non-central) 1795 to 2200 ms (posterior)	0 to 314 ms (non-posterior) 385 to 941 ms (posterior) 1794 to 2200 ms (posterior)	0 to 316 ms (non-posterior) 389 to 883 ms (posterior)
Gender	Preparatory	-966 to 0 ms (central)					
		-862 to -255 ms (posterior)					-209 to 0 ms (central)
	Selective	0 to 2200 ms (central)	109 to 1040 ms (central) 1056 to 2200 ms (central)	**108 to 1030 ms (central)**	122 to 372 ms (central) 377 to 922 ms (central)	115 to 354 ms (central) 385 to 922 ms (central)	114 to 270 ms (central) 383 to 916 ms (central)
		290 to 2200 ms (posterior)	459 to 2200 ms (posterior)	**495 to 1038 ms (posterior)** **1717 to 2200 ms (posterior)**	422 to 927 ms (posterior) 1738 to 2162 ms (posterior)	502 to 865 ms (posterior) 1759 to 2166 ms (posterior)	112 to 272 ms (non-central) 492 to 906 ms (posterior)

Red = Test > Control

Blue = Control > Test

Grey = difference not significant (direction is the same as others in that row)

Supplementary Table 3. Experiment 2: Clusters ($p < 0.05$) by high-pass filter value (Hz). Values reported in paper (0.25 Hz) are displayed in bold font. Timing of clusters is relative to the onset of acoustical stimuli.

		0.1	0.2	**0.25**	0.3	0.4	0.5
Location	Preparatory	-996 to -146 ms (central)	-958 to -109 ms (central)	**-957 to -319 ms (central)**	-963 to -541 ms (central)	-958 to -603 ms (central)	-958 to -605 ms (central) -595 to -281 ms (central)
		-486 to -119 ms, p = 0.067	-574 to -213 ms (posterior)				-944 to -706 ms (posterior)
	Selective	0 to 2200 ms (central)	106 to 2200 ms (central)	**286 to 1462 ms (central)**	105 to 250 ms (central) 373 to 1081 ms (central) 1372 to 2064 ms (central)	101 to 245 ms (central) 361 to 1042 ms (central) 1347 to 1922 ms (central)	104 to 246 ms (central) 361 to 1037 ms (central) 1303 to 1916 ms (central)
		0 to 2200 ms (non-central)	276 to 1094 ms (posterior)	**298 to 1002 ms (posterior)**	296 to 991 ms (posterior)	463 to 1009 ms (posterior)	0 to 235 ms (non-central) 459 to 970 ms (posterior)
Gender	Preparatory	-985 to 0 ms (central)	-947 to -617 ms (central)	**-916 to -616 ms (central)**	-916 to -707 ms (anterior)	-928 to -736 ms (anterior)	-920 to -725 ms (anterior)
		-1000 to -731 ms (posterior) -591 to 0 ms (posterior)				-920 to -745 ms (posterior)	-921 to -747 ms (posterior)
	Selective	0 to 2200 ms (central)	107 to 2200 ms (central)	**112 to 1219 ms (central)** **1261 to 2200 ms (central)**	115 to 1067 ms (central) 1305 to 2200 ms (central)	121 to 297 ms (central) 376 to 1214 ms (central) 1320 to 2053 ms (central)	122 to 355 ms (central) 356 to 1218 ms (central) 1320 to 2080 ms (central)
		0 to 2200 ms (non-central)	274 to 1259 ms (posterior) 1600 to 2200 ms (posterior)	**502 to 1031 ms (posterior)** **1844 to 2200 ms (posterior)**	504 to 1033 ms (posterior) 1920 to 2189 ms, p = 0.068	417 to 1137 ms (posterior) 1757 to 2172 ms (posterior)	122 to 251 ms (non-central) 508 to 1136 ms (posterior) 1945 to 2188 ms (posterior)

Red = Test > Control

Blue = Control > Test

Grey = difference not significant (direction is the same as others in that row)

Supplementary Table 4. Experiment 3: Clusters (*p* < 0.05) by high-pass filter value (Hz). Values reported in paper (0.25 Hz) are displayed in bold font. Timing of clusters is relative to the onset of acoustical stimuli.

		0.1	0.2	**0.25**	0.3	0.4	0.5
Location	Preparatory	-1000 to -477 ms (posterior)	-899 to -492 ms (posterior)	**-790 to -545 ms (posterior)**	-829 to -445 ms (posterior)	-808 to -433 ms (posterior)	-817 to -413 ms (posterior)
			-827 to -428 ms (central + anterior)		-806 to -297 ms (central + anterior)	-790 to -354 ms (anterior)	-798 to -348 ms (anterior)
	Selective	633 to 1055 ms (central) 1520 to 2200 ms (central)		**71 to 1503 ms (posterior + central)**	75 to 586 ms (posterior + central) 1071 to 1558 ms (posterior)	126 to 585 ms (posterior + central) 1066 to 1562 ms (posterior)	117 to 583 ms (posterior + central) 730 to 1010 ms (posterior) 1017 to 1501 ms (central)
		675 to 962 ms (posterior)		**77 to 941 ms (non-central)**	46 to 627 ms (anterior)	62 to 621 ms (central + anterior) 1038 to 1517 ms (anterior)	54 to 621 ms (central + anterior) 1045 to 1542 ms (anterior)
Gender	Preparatory						
	Selective		1072 to 1427 ms (central)	**573 to 1067 ms (central)**	564 to 1464 ms (anterior + central)	562 to 1067 ms (anterior + central)	549 to 1061 ms (anterior)
		714 to 1672 ms (posterior)	782 to 1356 ms (posterior) 1390 to 1685 ms (posterior)	**483 to 1039 ms (posterior)**	583 to 1323 ms (posterior)	580 to 1071 ms (posterior)	632 to 858 ms (posterior) 863 to 1066 ms (posterior)

Red = Test > Control

Blue = Control > Test

Grey = difference not significant (direction is the same as others in that row)