Wen et al. BMC Neuroscience (2024) 25:40
https://doi.org/10.1186/s12868-024-00889-6

BMC Neuroscience

RESEARCH

Open Access

Modulation of audiovisual integration in the left and right sides: effects of side and spatial coherency

XiaoHui Wen[1*], GuoQiang Li[1], XuHong Wang[1], XiaoLan Hu[1] and HongJun Yang[1]

Abstract

Background Using event-related potentials (ERPs), we aimed to investigate audiovisual integration neural mechanisms during a letter identification task in the left and right sides. Unimodal (A,V) and bimodal (AV) stimuli were presented on either side, with ERPs from unimodal (A,V) stimuli on the same side being compared to those from simultaneous bimodal stimuli (AV). Non-zero results of the AV-(A + V) difference waveforms indicated audiovisual integration on the left/right side.

Results When spatially coherent AV stimuli were presented on the right side, two significant ERP components in the integrated differential wave were noted. The N134 and N262, present in the first 300 ms of the AV-(A + V) integration difference wave, indicated significant audiovisual integration effects. However, when these stimuli were presented on the left side, there were no significant integration components. This audiovisual integration difference may stem from left/right asymmetry of cerebral hemisphere language processing.

Conclusions Audiovisual letter information presented on the right side was easier to integrate, process, and represent. Additionally, only one significant integrative component peaked at 140 ms in the parietal cortex for spatially non-coherent AV stimuli and provided audiovisual multisensory integration, which could be attributed to some integrative neural processes that depend on the spatial congruity of the auditory and visual stimuli.

Keywords Letter recognition, Audiovisual integration, Event-related potential, Spatial consistency, Neural mechanism

Background

The visual and auditory modalities are important sources of obtaining information from the outside world. When visual and auditory stimuli simultaneously appear and point to the same event, they are perceived as a coherent stimulus, with this perceptual process being referred to as audiovisual integration [1–10]. Studies examining the neural mechanisms underlying audiovisual speech integration suggest that the superior temporal sulcus is the main brain region crucially involved in the integration of visual and auditory speech information [11–17]. Additionally, the frontal cortex [18] and parietal region [19] seem to be involved. However, there have been inconsistent reports regarding the activation of these regions.

In modern society, written language has played an important role in peoples' lives. Reading and writing are products of human civilization, and the connection between visual letters and speech sounds have been artificially defined and acquired through long-term learning. It is rather impossible to have regions in the human brain that have naturally evolved to integrate this type of audiovisual stimulus [20]. Nevertheless, most people can easily

*Correspondence:
XiaoHui Wen
27308917@qq.com
[1] Hunan University of Humanities, Science and Technology, Loudi 417000, China

learn and grasp the association of visual letters with their corresponding speech sounds, and perceive both as a unified whole. There has been extensive interest in and research on the audiovisual integration of letters and sounds within the fields of basic and applied research, which emphasizes the important role of the parietotemporal area. Previous studies that directly examined the neural mechanisms underlying letter and speech integration revealed that the superior temporal sulcus is a multisensory area for speech integration [3, 6, 20–27].

Compared with other audio-visual stimuli integration, the integration of visual letters and speech sounds has shown similar commonalities in terms of brain mechanisms and influencing factors; however, it is unique in terms of its specific cross-modal connections. Notably, these studies all presented visual and auditory stimuli in the central side without distinguishing the left/right sides. For stimuli presented in the left/right sides, visual information was first projected to the contra-cerebral hemisphere for further processing. However, there was a clear left/right asymmetry in speech processing, with most people having a verbal advantage in the left hemisphere, as evidenced by behavioral data showing higher cognitive accuracy and greater speed for verbal stimuli presented on the right side (Campbell [7]; Dormal et al. [12]; Erickson et al. [2, 28, 29]). Previous neuroimaging studies on the processing mechanisms of visual letters [22, 23, 30, 31] revealed that brain regions that process such stimuli are mainly located in the left hemisphere. The superior temporal sulcus, which is involved in the auditory processing and visual imaging of visual letters in the left hemisphere, also contains critical areas for understanding spoken and written words.

Additionally, in the real world, signals from different sensory modalities of the same object or event are usually temporally and spatially coherent. From this perspective, multisensory integration should diminish or disappear when each single stimulus is far apart in time and space, since in this case, two objects or events are perceived rather than a single object or event. Therefore, the spatial coherence of a stimulus plays a critical role in multisensory integration [4, 14, 24, 25, 32–34]. Contrastingly, apart from visual flash and pure sound used in previous audiovisual integration studies, the link between visual letter and speech sound formed through previous learning (rather than during human evolution) was more readily perceived as a whole. The present study aimed to address whether spatial factors influenced the integration of such audiovisual stimuli and whether the degree of influence and the involved brain areas differed from those previously reported.

Considering the dominant processing effect of visual stimuli on the right visual side, we postulated that there may be differences in the audiovisual letter integration effects induced by stimuli presented in different visual sides. Based on this hypothesis, stimuli were presented on the left or right visual side, and participants were required to always look at the central fixation point during the experiment. Identified visual letters and speech sounds were presented unimodally (audio [A]; visual [V]) and bimodally (AV). Comparison of the ERP waveforms evoked by the stimuli on the left/right side allowed further elucidation of the underlying neural mechanisms involved in audiovisual integration during letter recognition in order to provide electrophysiological evidence for deeper insight into the brain mechanisms underlying letter and speech integration.

Methods

Participants

We used G*Power 3.1 software to calculate the sample size required to ensure sufficient statistical efficacy, which revealed that 24 participants met the minimum sample requirement for this experiment. Since there are many right-handed university undergraduates in our college, to minimize confusion, we randomly recruited 34 (16 male [mean age: 21.4 years] and 18 female participants [mean age: 20.9 years]) participants, all of whom were right-handed, had no history of mental disease or brain injury, and had normal or corrected vision and normal hearing. This study was approved by the Research Ethics Committee of Hunan University of Humanities, Science, and Technology (approval number: jy2021-072); further, all participants provided written informed consent. Additionally, this study was conducted in accordance with the tenets of the Declaration of Helsinki.

Experimental methods

Stimuli materials

Chinese letter processing involves relatively stronger connection between the form and meaning, while English letter processing involves relatively stronger connection between form and sound. In this study, the participants were college students who had been learning English for >10 years; moreover, the present study applied simple letters that all participants were familiar with as stimuli. Therefore, to better investigate the characteristics of audio-visual integration, we borrowed from previous studies and used letters rather than words as stimulus materials. The stimuli were presented unimodally or bimodally (single visual [V], single auditory [A], and simultaneous audiovisual [AV]). The visual stimuli included the English alphabet (A, D, E, I, J, K, O, R, and T; font style, Arial; font size, 36) projected on a screen at a distance of 2.83 m from the participants using a projector, and presented at a position of 9.6° to the left (VL) or right (VR) of the gaze point for a duration of 60 ms. The auditory stimulus comprised the recording of the same nine letters of the alphabet (75 dB; 5 ms rise/fall; duration, 60 ms), presented through speakers on the left and right (AL and AR), which were 19.2° apart. Neurophysiological and behavioral studies [33] have shown that the relative position between single morphological stimuli plays a key role in multisensory integration and that between-stimuli positional differences will affect their integration effects. Therefore, auditory stimuli were presented using speakers (rather than head-phones) placed in the same position as that of the letter presentation [14, 33]. In the condition of the simultaneous presentation of audiovisual stimuli, both spatially congruent (ALVL and ARVR) and incongruent (ALVR and ARVL) locations of visual and auditory stimuli were applied. Accordingly, the entire experiment included eight stimulus presentation conditions: AL, AR, VL, VR, ALVL, ARVR, ALVR, ARVL (Abbreviations: VL, visual stimuli presented on the screen at a position of 9.6° to the left of the gaze point(+); VR, visual stimuli presented on the screen at a position of 9.6° to the right of the gaze point(+); AL, auditory stimulus presented through speakers on the left located (19.2° apart) behind the screen corresponding to the location of the VL; AR, auditory stimulus presented through speakers on the right located (19.2° apart) behind the screen corresponding to the location of the VR; ALVL, auditory and visual stimuli presented simultaneously to the left of the gaze point(+); ARVR, auditory and visual stimuli presented simultaneously to the right of the gaze point(+); ARVL, auditory stimulus presented to the right of the gaze point(+) and visual stimuli presented simultaneously to the left of the gaze point(+); and ALVR, auditory stimulus presented to the left of the gaze point(+) and visual stimuli presented simultaneously to the right of the gaze point(+)). All the conditions were presented randomly. Figure 1 shows a schematic of the experimental design.

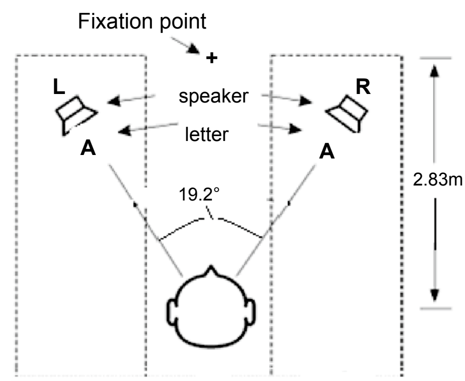

Fig. 1 Schematic of the experimental design. Visual stimuli (VL, VR) presented to the left (L) or right (R) of the fixation point; auditory stimulus (AL, AR) presented through speakers on the left and right (19.2° apart) located behind the screen corresponding to the location of the visual stimulus presentation

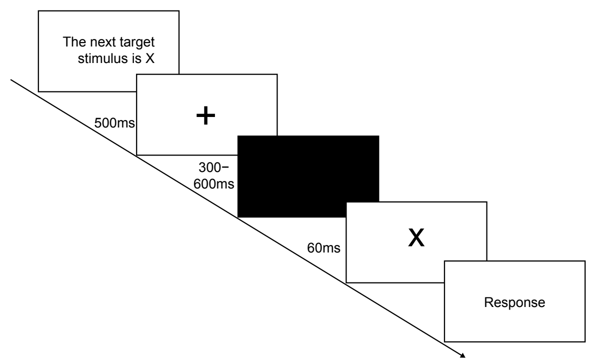

Fig. 2 The experimental procedure of the letter identification task

Procedure and task

The experiment was conducted in a dark, soundproof laboratory with the participant sitting in a comfortable chair and gazing at the center of the screen with both eyes. The experiment started with displaying the + (gaze point) for 500 ms. Next, a black screen was displayed for 300–600 ms, followed by a random presentation of stimuli to the left or right of the gaze point for 60 ms. To exclude the confounding effect of brain region activity induced by other factors (decision making, response preparation, etc.) on brain region activity induced by cross-morphological interaction and perceptual processing, we adopted the odd-ball experimental paradigm, which requires participants to respond to low-probability target stimuli. The participant's task was to press the "1" key as quickly as possible after the target stimulus was shown in any channel. During the experiment, the target stimulus was not fixed; instead, it was changed once per group (72 trials). Before each target stimulus changed, a message—"The next target stimulus is X (a letter presented as a picture)"—was displayed at the center of the screen. Throughout the experiment, participants were asked to look at the central gaze point on the screen with both eyes, to try not to move or blink when the stimulus was presented, and to respond as quickly and accurately as possible. Figure 2 shows the experimental procedure of the task.

Before initiating the formal experiment, each participant was given 20 practice trials to ensure familiarity with the experimental procedure and task requirements. In the formal experiment, each stimulus presentation condition was repeated 90 times, for a period of 1/9th of the target stimuli. The rest time at the end of each group of trials was freely decided by the participants in order to eliminate fatigue.

ERP recording and analysis

The ERP recording and analysis system (Brain Products GmbH, Gilching, Germany) was used to record the EEG signals according to the international 10–20 system extended by 64 electrode caps. The reference electrodes were placed on the bilateral mastoid line with the forehead grounded. Horizontal electrooculography (HEOG) was recorded through the electrodes at the lateral corners of the left and right eyes, while vertical electrooculography (VEOG) was recorded through the electrodes located in the upper and lower orbits of the left eye. The signal was amplified with an amplifier, filtered with a band pass of 0.05–80 Hz, and sampled at 500 Hz. The impedance between the scalp and electrodes was < 5 kΩ, with off-line super position processing after the completion of recording. VEOG and HEOG were automatically corrected, while eye-movement and other artifacts (voltages exceeding 100 μV) were automatically rejected. After superimposition, the participants whose eye gaze point deviated from the central gaze point by > 0.2 and those whose eye gaze's wave amplitude was > 3 μV in the − 300 ms interval were first excluded [35]. The results showed that all participants had gaze point deviations of < 0.2°.

For the purpose of this study, ERPs that were evoked only when participants responded correctly to non-target stimuli were included in the analysis; additionally, EEGs were averaged for each of the eight conditions and filtered at a low-pass cutoff frequency of 30 Hz. Furthermore, to exclude confounding of early cross-morphological interactions by pre- and post-stimulus slow potentials, the mean EEG waveforms were filtered using a high-pass cutoff filter at 2 Hz [33]. ERP was analyzed at 1000 ms after stimulus presentation and 100 ms before stimulus onset, as a baseline. Multisensory interaction effects were examined by subtracting the sum of the ERPs of the two single-channel stimuli from the ERP of the two-channel stimuli. Specifically, the interaction effect was equal to the AV-(A + V) integration difference wave. Integration difference waves for spatially position-consistent and non-consistent *AV* stimulus pairs were calculated using the following equations.

(i) Difference waves of spatially coherent AV stimuli:

$$\mathrm{L(left)\!:\!ALVL} - (\mathrm{AL} + \mathrm{VL}) = \mathrm{SL},$$

$$\mathrm{R\left(right\right)\!:\!ARVR} - (\mathrm{AR} + \mathrm{VR}) = \mathrm{SR}.$$

(ii) Difference waves of spatially non-coherent AV stimuli:

$$\mathrm{L(left)\!:\!ARVL} - (\mathrm{AR} + \mathrm{VL}) = \mathrm{DLR},$$

$$\mathrm{R\left(right\right)\!:\!ALVR} - (\mathrm{AL} + \mathrm{VR}) = \mathrm{DRL}.$$

The aforementioned paradigm has been used in typical multisensory integration ERP research. However, this differential wave equation may present an issue. If there was a common activation I in three ERP waveforms (visual [V], auditory [A], and audiovisual [AV]), such as the expected slow wave caused by target expectations, the comparison would have been inappropriate. Since the common activation I was reduced twice (i. e., $AVC-(AC+VC)=-C$ [36]) by dualform ERP, the resulting item may have reflected multisensory interaction and the co-activation component. To examine the impact of spatial consistency on audiovisual integration more accurately, we used the sum of positional consistent difference waves to subtract the sum of positional inconsistent difference waves. This "double difference" wave was mathematically similar to the ERP of a positional consistent stimulus pair minus the ERP of a positional inconsistent stimulus pair (i.e., $(SL+SR)-(DLR+DRL)=(ALVL+ARVR)-(ALVR+ARVL)$). This equation contained two subtracted numbers (ALVL and ARVR, where L denoted left and R denoted right) and two negative numbers (ALVR and ARVL), which allowed exclusion of contamination by other artifacts. The period of deviation from zero in the double difference wave was determined to be the processing stage of the spatial representation of the integration of the two sensory information types.

To rule out the confounding effect of brain region activation by other factors (decision making and preparation for response, etc.) on cross-modality interaction and sensory processing-induced brain region activity [9], we only analyzed the ERP components in the first 300 ms of the differential wave. According to the total average map and topographic map of ERP, fifteen electrode positions (FP1, F 3, F 5, c 5, P 5, O 1, FZ, CZ, POZ, FP2, F 4, F 6, C 6, P 6, O 2) in the central-parietal region and bilateral (left-sided) prefrontal-temporal regions were selected for analysis of variance with two-factor (stimulus presentation conditions and recording points) repeated measurements. The significance of the S_L and S_R (positional consistent stimulus pair) integration difference waves was tested separately using Student's *t*-tests (i.e., SL, SR, and the double difference waves were compared with zero). Moreover, *p*-values were reported with Bonferroni correction. The level of significance was set at $p < 0.001$.

Results

Spatially coherent AV stimuli

Figure 3 shows the ERP components of the original waveforms evoked by spatially coherent visual (VL, VR), auditory (AL, AR), and audiovisual (ALVL, ARVR) stimuli. The auditory component included significant P1, N1, and P2 components, which were mainly in the forehead and central regions. The visual ERP component included significant N1 and P2 components, which were mainly in the posterior part of the scalp. The components evoked by auditory and visual stimuli could also be distinguished in the ERP waveforms of bimodal stimuli.

AV-(A + V) difference waves were analyzed to reveal the cross-modal integration effect of audiovisual stimuli located simultaneously to the right and left of the gaze point (see Methods for the specific calculation of the interaction effect). Student's t-tests results demonstrated that the SL integration difference wave had no significant components ($p > 0.05$; Fig. 4a), while the SR integration difference wave had two significant components between 72–145 ms and 250–280 ms (Fig. 4b).

Further analysis of the SR difference wave showed that the component within the 72–145 ms interval peaked at 134 ms (N134) and was significantly distributed in the central-parietal region ($p < 0.001$); the component within the 250–280-ms interval peaked at 262 ms (N262) and was significant in both (left-skewed) prefrontal-temporal regions ($p < 0.05$) (see Fig. 5).

Spatially non-coherent AV stimuli

Figure 6 shows the ERP components in the original waveforms induced by non-coherent AV stimuli (ARVL, ALVR). The ERP waveforms for the bimodal stimuli included significant N1 (peaked at ≈130 ms) and P2 (peaked at ≈250 ms) components.

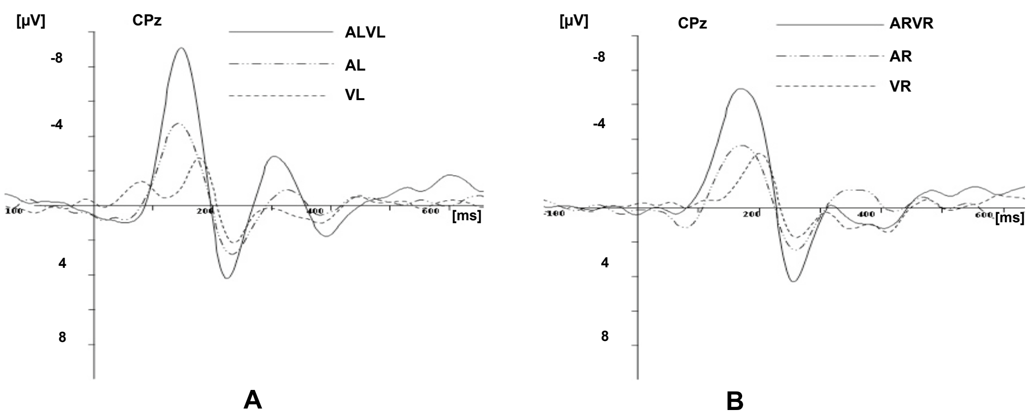

Fig. 3 **a** Plot of the total mean ERP evoked by AL, VL and ALVL at the CPZ recording point; **b** plot of the total mean ERP induced by AR, VR, and ARVR at the CPZ recording point. ERP: event related potentials; VL and VR: visual stimuli presented to the left (VL) and right (VR) of the gaze point; AL and AR: auditory stimulus presented through speakers on the left (AL) and right (AR) located (19.2° apart) behind the screen corresponding to the location of the visual stimulus presentation

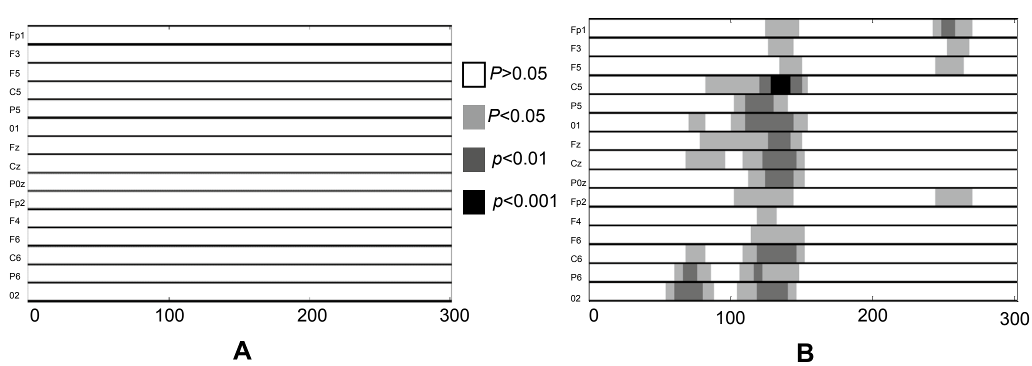

Fig. 4 **a** Student's t-test plotted to demonstrate the p-values of the ALVL- (AL + VL) difference wave (SL) in the 0–300-ms interval for some electrodes (compared with the zero value). **b** Student's t-test plotted to demonstrate the p-values of the ARVR- (AR + VR) difference wave (SR) in the 0–300-ms interval for some electrodes (compared with the zero value)

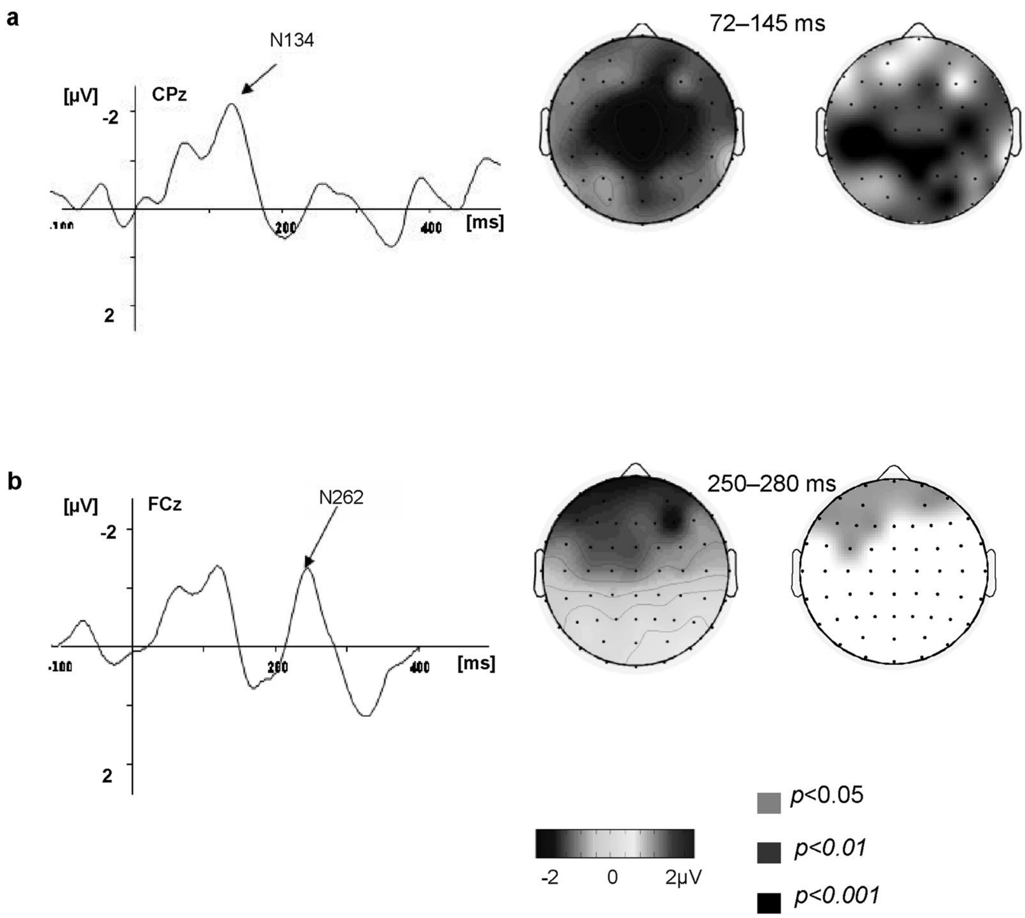

Fig. 5 Total event-related potential average map of ARVR- (AR + VR) difference wave (SR), with its topographic map and t-test topographic map within 72–145 ms (**a**) and 250–280 ms (**b**)

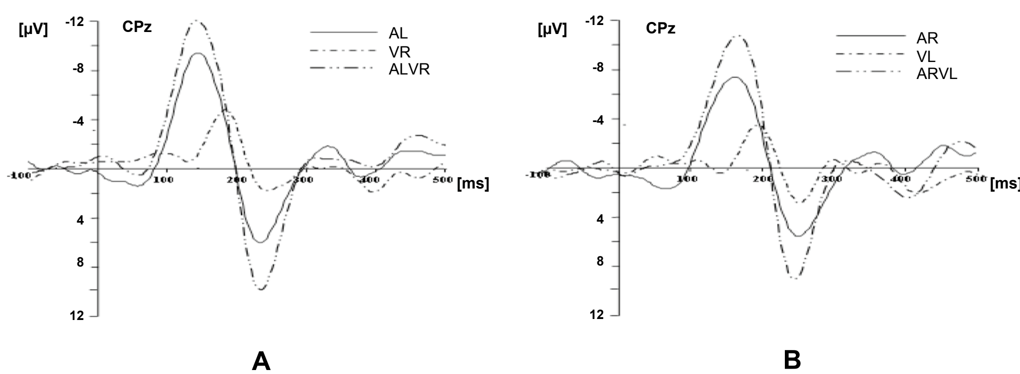

Fig. 6 **a** Plot of the total mean event-related potential (ERP) evoked by AL, VR, and ALVR at the CPZ recording point; **b** plot of the total mean ERP induced by AR, VL and ARVL at the CPZ recording point

We investigated the influence of spatial coherence on multisensory integration based on double difference waves, which could more thoroughly eliminate the confusion caused by common components in ERP waveforms (see the Methods section for details). As shown in Fig. 7, there was a marked integration component in the double difference wave; the component peaked at 140 ms (P140) and was significantly distributed in the parietal region ($p < 0.05$; Fig. 7).

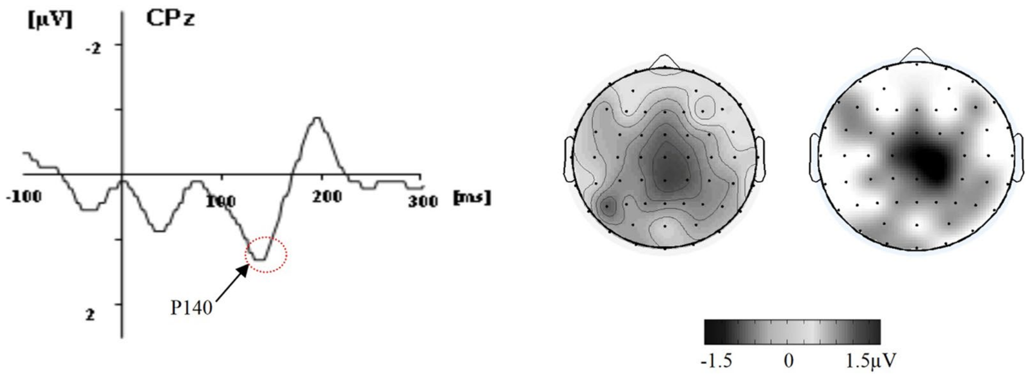

Fig. 7 Total event-related potential average map of (ALVL + ARVR) – (ALVR + ARVL) double difference waves on CPz recording sites, with its topographic map and t-test topographic map

Discussion

Spatially coherent AV stimuli

When people look precisely at the center, information presented to the right (i.e., right visual side) and left (i.e., left visual side) of the gaze point is projected in the left and right cerebral hemispheres, respectively. For visual-verbal stimuli, the processing brain regions are mainly located in the left hemisphere [22, 23]. In contrast to language processing [6, 7], auditory processing is more susceptible to visual influences at both the neural and perceptual levels. Given all of the above, we hypothesized that there may be asymmetric audiovisual integration of the left and right visual sides during letter recognition. In this study, the ERP waveforms evoked by two-channel stimuli (AV) in the left and right visual sides were subtracted from the algebraic sum of ERP waveforms evoked by two single-channel stimuli presented separately. This was to explore whether the integration effects between visual and auditory stimuli presented in different visual sides were similar as well as to further investigate the neural mechanisms underlying audiovisual integration. The results showed that when stimuli were presented in the right visual side, there were two major ERP components in the ARVR-(AR + VR) difference waveform. The first was a large negative deviation within 72–145 ms present in the central-parietal region (N134), while the second was a negative deviation within 250–280 ms present in the bilateral (left-skewed) prefrontal-temporal regions (N262). Contrastingly, for the condition where the stimuli were presented in the left visual side, no significant ERP component was present in the ALVL-(AL + VL) difference wave. The present experiment verifies the previously proposed hypothesis, i.e., there is a significant difference between the integration effects evoked by audiovisual stimuli presented in different sides, with only audiovisual stimuli presented in the right side showing integration effects.

The N134 component induced by the right-side audiovisual integration was presented within 72–145 ms, verifying the proposed hypothesis that there is a significant difference between those of the auditory N1 component. This result is consistent with previous reports [3, 10, 37, 38]. We interpret this to mean that when audiovisual stimuli are presented on the right side simultaneously, not only does the presentation of auditory stimuli activate auditory-related cortical areas, but the simultaneous presentation of visual stimuli also contributes to the activation of auditory cortical areas.N262 in the difference wave was present at latencies of 250–280 ms, and it characterized the integration of visual letters with auditory speech in multisensory areas. Although that component was distributed in the prefrontal-temporal regions, similar components in aforementioned studies [26, 33] were mostly distributed in the superior temporal cortex. We attribute this to the fact that the stimuli used in the aforementioned studies were visual flashes and auditory noises, whereas the stimuli used in the present experiment were letters and speech sounds that were specifically related to each other, and additionally, the stimulus presentation time was significantly shorter (only 60 ms). During audiovisual speech processing, both temporal and prefrontal cortices are areas where sensory integration occurs, and activation of the temporal cortex can be transmitted to the prefrontal cortex. The interpretation of N262 in this study is supported by the results of the ERP study on audiovisual Chinese character recognition by Liu et al. [39], who found two components—N210 (right prefrontal-temporal cortex) and P270 (left prefrontal-temporal cortex). Based on the latency and location of these two components in the brain, they proposed that these two components characterize the activation of brain regions that occurs during audiovisual integration. The involvement of the prefrontal cortex in audiovisual integration was also verified by Raij et al. [3]. Van

Atteveldt et al. [20] also suggested that the prefrontal cortex is associated with speech processing and attention. Furthermore, studies conducted by Alsius et al. [5], Fleming et al. [2], and Gonzalo and Büchel [30] revealed that the left prefrontal cortex is the main area for learning audiovisual connections, and that the prefrontal area is strongly activated during the learning of novel audiovisual connections. However, the activation of this area is diminished if such connections are familiar and acquired over time.

In our experiments, visual, auditory, and audiovisual stimuli were presented on the left and right sides, and the integration effects evoked by stimulus presentation on the different sides and the brain area mechanisms involved were examined separately. The results showed that when the stimuli were presented in the right side, two major ERP components (N134 and N262) were present in the first 300 ms of the AV-(A + V) integration difference wave, indicating a significant audiovisual integration effect. However, when the stimuli were presented in the left side, there were no significant components in the integration difference wave, which indicates the absence of a significant audiovisual integration effect. We suggest that this difference in audiovisual integration between the left and right sides results from the functional variability of the left and right cerebral hemispheres. When letters are presented in the right visual side, the left brain, which primarily performs verbal processing, is more likely to integrate audiovisual letter information, process it, and represent it. Therefore, the participants are more likely to recognize the letters presented on that visual side. This is consistent with previous reports, which proposed that the left hemisphere was the main brain area involved in speech processing and recognition [22–25]. However, when audiovisual speech stimuli are presented in the left side, visual letter information projected to the right cerebral hemisphere may not be processed until it is delivered to the corresponding area in the left hemisphere, which delays the arrival of visual information in the superior temporal area. Therefore, in the condition where the stimuli were presented in the left side, the audiovisual stimuli were not well integrated due to the temporal factor. Accordingly, no significant integration effect could be observed. Our findings are consistent with previous reports regarding letter-speech integration, where Raij et al. [3] found significant audiovisual integration effects in the left prefrontal parietal area and left posterior superior temporal sulcus. Although an integration effect was also present in the right superior temporal sulcus, the effect appeared 70 ms later than in the left superior temporal sulcus, which reflects the possible transmission of the corpus callosum association through bilateral symmetric activation.

However, most previous studies have found latency and brain area distribution in visual cortical areas similar to that of the ERP component associated with audiovisual integration in visual N1 [33, 38], all of which were interpreted as a modulation of activity in visual cortical areas through the presentation of auditory stimuli. However, we did not observe any similar audiovisual interaction component in this area. This could be attributed to the fact that the visual stimuli in our experiment were statically presented with clear letters, whereas speech is continuous and constantly changing over time, which makes the recognition of auditory stimuli more difficult than that of visual letter symbols. Moreover, it further demonstrates that visual information has a greater impact on auditory processing. In language processing, auditory processing is more susceptible to visual influences at both the neural and perceptual levels, and auditory language areas (e.g., the temporal transverse gyrus) can be activated through visual language. This is consistent with previous reports regarding letter-speech integration [3, 7, 20, 21], where audiovisual integration modulates activity in the auditory-related cortex rather than in the visual-related cortex. A similar tuning effect was not observed in the visual-related cortex. This could be attributed to the fact that visual information has a relatively higher degree of reliability; therefore, its presentation has a greater effect on auditory cortical activation, whereas the presentation of auditory information has little effect on the visual cortex.

Spatially non-coherent AV stimuli

The comparison of ERP waveforms evoked by spatially congruent and non-congruent stimuli revealed a significant integration component within 120–160 ms in the double difference wave, which peaked at 140 ms (P140) and was mainly distributed in the parietal cortex. The latency and distribution of the brain regions of this component are similar to those of previous studies, where Fort et al. [40] observed a positive component extensively distributed in the parietal-occipital cortex at ≈170 ms. Teder-Sälejärvi et al. [33] also observed a central positive component around 175 ms. Calvert and Campbell [6] mentioned that the parietal cortex is involved in cross-morphological localization and in tuning spatial attention processing. Accordingly, the P140 in our results indicates a spatial location effect in the perceptual phase, while AV multisensory convergence is based on certain types of integrated neural processing that depend on spatial coherence between the auditory and visual stimuli (i.e., the spatial location between visual and auditory stimuli can affect audiovisual integration effects).

Conclusions

In this study, regarding spatially coherent audiovisual stimuli presented on the right side, there were two major ERP components; namely, the N134, which represents the audiovisual integration occurring in auditory areas, and the N262, which represents the audiovisual integration occurring in polymorphic sensory areas. However, no significant component of the integration difference wave emerged when the audiovisual stimuli were presented on the left side. This difference in audiovisual integration between the left and right sides may be attributed to the asymmetry of language processing in the left and right cerebral hemispheres. The audiovisual letter information presented simultaneously in the right visual side is more likely to be perceived as a unified integrated quality and more likely to be processed and characterized. Regarding audiovisual stimuli presented in incongruent locations, there was a significant integration component in the difference wave that peaked at 140 ms (P140), which was extensively distributed in parietal areas and represented the effect of spatial location on audiovisual integration in multisensory areas. Our results suggest that audiovisual integration is based on certain different types of integrated neural processing, some of which depend on the spatial coherence between visual and auditory stimuli, whereas others are not limited by spatial coherence.

Further research

In this study, letters were used as stimuli when examining the audiovisual integration in the left or right visual sides. There is need to verify whether the results will be consistent if the stimuli are replaced with Chinese or other characters. Furthermore, the influencing factors of multisensory integration include spatial and temporal factors. Future research can regulate the temporal order relationship between stimuli to examine the impact of temporal factors on audiovisual integration. Additionally, since only radially oriented sources would cause topographical maxima over these areas, especially with the folding of the auditory cortex, ERPs can appear over frontal areas, originating in auditory areas, rather than prefrontal cortical areas. Accordingly, subsequent studies are warranted to use the fMRI technique for accurate localization.

Abbreviations

ERP Event-related potential
STG Superior temporal gyrus

Acknowledgements

We sincerely thank all those who have helped us during the writing of this thesis. We gratefully acknowledge the help of Prof. ChaoLiang Liu and Prof. JunJun Cao, who have offered us valuable suggestions in academic studies. We thank Editage (www.editage.cn) for English language editing.

Author contributions

Conceptualization, XH. Wen, GQ.L., XH. Wang, XL.H., HJ.Y.; Methodology, XH. Wen, GQ.L., XH. Wang, XL.H., HJ.Y.; Formal analysis, XH. Wen, GQ.L., XH.Wang; Resources, XH. Wen, GQ.L., XH. Wang; Data curation, XH. Wen, GQ.L., XH. Wang; Writing—original draft preparation, XH. Wen, XL.H., HJ.Y.; Writing—review and editing, XH. Wen, GQ.L., XH.Wang, XL.H., HJ.Y.

Funding

2023 Annual Research on Teaching Reform in General Higher Education Institutions of Hunan Province (Grant ID: HNJG-20231180).

Availability of data and materials

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Declarations

Ethics approval and consent to participate

This study was approved by the Research Ethics Committee of Hunan University of Humanities, Science, and Technology (approval number: jy2021-072), and all participants provided written informed consent. Furthermore, this study was conducted in accordance with the tenets of the Declaration of Helsinki.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 21 March 2024 Accepted: 16 August 2024
Published online: 27 August 2024

References

1. Erickson LC, Zielinski BA, Zielinski JE, Liu G, Turkeltaub PE, Leaver AM, et al. Distinct cortical locations for integration of audiovisual speech and the McGurk effect. Front Psychol. 2014;5:534.
2. Fleming JT, Noyce AL, Shinn-Cunningham BG. Audio-visual spatial alignment improves integration in the presence of a competing audio-visual stimulus. Neuropsychologia. 2020;146: 107530.
3. Raij T, Uutela K, Hari R. Audiovisual integration of letters in the human brain. Neuron. 2000;28:617–25.
4. Tang XY, Wu YN, Peng X, Wang AJ, Li Q. The influence of endogenous spatial cue validity on audiovisual integration. Acta Psychol Sin. 2020;52:835–46.
5. Alsius A, Paré M, Munhall KG. Forty years after hearing lips and seeing voices: the McGurk effect revisited. Multisens Res. 2018;31:111–44.
6. Calvert GA, Campbell R. Reading speech from still and moving faces: the neural substrates of visible speech. J CognNeurosci. 2003;15:57–70.
7. Campbell R. The processing of audio-visual speech: empirical and neural bases. Philos Trans R Soc Lond B Biol Sci. 2008;363:1001–10.
8. McGurk H, MacDonald J. Hearing lips and seeing voices. Nature. 1976;264:746–8.
9. Mishra J, Martinez A, Sejnowski TJ, Hillyard SA. Early cross-modal interactions in auditory and visual cortex underlie a sound-induced visual illusion. J Neurosci. 2007;27:4120–31.
10. Molholm S, Ritter W, Murray MM, Javitt DC, Schroeder CE, Foxe JJ. Multisensory auditory-visual interactions during early sensory processing in humans: a high-density electrical mapping study. Brain Res Cogn Brain Res. 2002;14:115–28.
11. Beauchamp MS, Nath AR, Pasalar S. fMRI-Guided transcranial magnetic stimulation reveals that the superior temporal sulcus is a cortical locus of the McGurk effect. J Neurosci. 2010;30:2414–7.
12. Dormal G, Pelland M, Rezk M, Yakobov E, Lepore F, Collignon O. Functional preference for object sounds and voices in the brain of early blind and sighted individuals. J CognNeurosci. 2018;30:86–106.

13. Hwang J, Romanski LM. Prefrontal neuronal responses during audiovisual mnemonic processing. J Neurosci. 2015;35:960–71.
14. Macaluso E, George N, Dolan R, Spence C, Driver J. Spatial and temporal factors during processing of audiovisual speech: a PET study. Neuroimage. 2004;21:725–32.
15. Sekiyama K, Kanno I, Miura S, Sugita Y. Auditory-visual speech perception examined by fMRI and PET. Neurosci Res. 2003;47:277–87.
16. Tijms J, Fraga-González G, Karipidis II, Brem S. The role of letter-speech sound integration in normal and abnormal reading development. Front Psychol. 2020;11:1441.
17. Ozker M, Schepers IM, Magnotti JF, Yoshor D, Beauchamp MS. A double dissociation between anterior and posterior superior temporal gyrus for processing audiovisual speech demonstrated by electrocorticography. J CognNeurosci. 2017;29:1044–60.
18. Szycik GR, Tausche P, Münte TF. A novel approach to study audiovisual integration in speech perception: localizer fMRI and sparse sampling. Brain Res. 2008;1220:142–9.
19. Diaconescu AO, Alain C, McIntosh AR. The co-occurrence of multisensory facilitation and cross-modal conflict in the human brain. J Neurophysiol. 2011;106:2896–909.
20. Pekkola J, Ojanen V, Autti T, Jääskeläinen IP, Möttönen R, Tarkiainen A, et al. Primary auditory cortex activation by visual speech: an fMRI study at 3 T. NeuroReport. 2005;16:125–8.
21. van Atteveldt NM, Formisano E, Blomert L, Goebel R. The effect of temporal asynchrony on the multisensory integration of letters and speech sounds. Cereb Cortex. 2007;17:962–74.
22. James KH, James TW, Jobard G, Wong AC, Gauthier I. Letter processing in the visual system: different activation patterns for single letters and strings. Cogn Affect BehavNeurosci. 2005;5:452–66.
23. Hashimoto R, Sakai KL. Learning letters in adulthood: direct visualization of cortical plasticity for forming a new link between orthography and phonology. Neuron. 2004;42:311–22.
24. Li Q, Yang H, Sun F, Wu J. Spatiotemporal relationships among audiovisual stimuli modulate auditory facilitation of visual target discrimination. Perception. 2015;44:232–42.
25. Liu CL, Hue CW, Chen CC, Chuang KH, Liang KC, Wang YH, et al. Dissociated roles of the middle frontal gyri in the processing of Chinese characters. NeuroReport. 2006;17:1397–401.
26. Mishra S, Lunner T, Stenfelt S, Rönnberg J, Rudner M. Seeing the talker's face supports executive processing of speech in steady state noise. Front Syst Neurosci. 2013;7:96.
27. Karipidis II, Pleisch G, Di Pietro SV, Fraga-González G, Brem S. Developmental trajectories of letter and speech sound integration during reading acquisition. Front Psychol. 2021;12: 750491.
28. Tripathi KK, Mukundan CR. Differential processing of visual information from right and left central hemisides during binocular viewing. Indian J Aerosp Med. 2004;48:22–7.
29. Tzeng OJ, Hung DL, Cotton B. Visual interalisation effect in reading Chinese characters. Nature. 1979;282:499–501.
30. Gonzalo D, Büchel C. Audio-visual associative learning enhances responses to auditory stimuli in visual cortex. In: Functional NeuroImaging of visual cognition-attention and performance. Oxford: Oxford University Press; 2004, 225–39.
31. Gaffan D, Harrison S. Auditory-visual associations, hemispheric specialization and temporal-frontal interaction in the rhesus monkey. Brain. 1991;114:2133–44.
32. Gondan M, Niederhaus B, Rösler F, Röder B. Multisensory processing in the redundant-target effect: a behavioral and event-related potential study. Percept Psychophys. 2005;67:713–26.
33. Teder-Sälejärvi WA, Di Russo F, McDonald JJ, Hillyard SA. Effects of spatial congruity on audio-visual multimodal integration. J CognNeurosci. 2005;17:1396–409.
34. Vroomen J, de Gelder B. Sound enhances visual perception: cross-modal effects of auditory organization on vision. J Exp Psychol Hum Percept Perform. 2000;26:1583–90.
35. Woodman GF, Luck SJ. Serial deployment of attention during visual search. J Exp Psychol Hum Percept Perform. 2003;29:121–38.
36. Teder-Sälejärvi WA, McDonald JJ, Di Russo F, Hillyard SA. An analysis of audio-visual crossmodal integration by means of event-related potential (ERP) recordings. Brain Res Cogn Brain Res. 2002;14:106–14.
37. Föcker J, Hölig C, Best A, Röder B. Neural plasticity of voice processing: evidence from event-related potentials in late-onset blind and sighted individuals. Restor Neurol Neurosci. 2015;33:15–30.
38. Giard MH, Peronnet F. Auditory-visual integration during multimodal object recognition in humans: a behavioral and electrophysiological study. J CognNeurosci. 1999;11:473–90.
39. Liu Q, Qiu J, Chen A, Yang J, Zhang Q, Sun HJ. The effect of visual reliability on auditory-visual integration: an event-related potential study. NeuroReport. 2007;18:1861–5.
40. Fort A, Delpuech C, Pernier J, Giard MH. Early auditory-visual interactions in human cortex during nonredundant target identification. Brain Res Cogn Brain Res. 2002;14:20–30.

Publisher's Note