

The influence of visual information on auditory processing in individuals with congenital amusia: An ERP study



Xuejing Lu^{a,b,*}, Hao T. Ho^a, Yanan Sun^{b,c}, Blake W. Johnson^{b,c}, William F. Thompson^{a,b}

^a Department of Psychology, Macquarie University, Sydney, NSW, Australia

^b ARC Centre of Excellence in Cognition and its Disorders, NSW, Australia

^c Department of Cognitive Science, Macquarie University, Sydney, NSW, Australia

ARTICLE INFO

Article history:

Received 17 March 2016

Accepted 17 April 2016

Available online 27 April 2016

Keywords:

Congenital amusia

Audiovisual

ERP

Multisensory

Music perception

ABSTRACT

While most normal hearing individuals can readily use prosodic information in spoken language to interpret the moods and feelings of conversational partners, people with congenital amusia report that they often rely more on facial expressions and gestures, a strategy that may compensate for deficits in auditory processing. In this investigation, we used EEG to examine the extent to which individuals with congenital amusia draw upon visual information when making auditory or audio-visual judgments. Event-related potentials (ERP) were elicited by a change in pitch (up or down) between two sequential tones paired with a change in spatial position (up or down) between two visually presented dots. The change in dot position was either congruent or incongruent with the change in pitch. Participants were asked to judge (1) the direction of pitch change while ignoring the visual information (AV implicit task), and (2) whether the auditory and visual changes were congruent (AV explicit task). In the AV implicit task, amusic participants performed significantly worse in the incongruent condition than control participants. ERPs showed an enhanced N2–P3 response to incongruent AV pairings for control participants, but not for amusic participants. However when participants were explicitly directed to detect AV congruency, both groups exhibited enhanced N2–P3 responses to incongruent AV pairings. These findings indicate that amusics are capable of extracting information from both modalities in an AV task, but are biased to rely on visual information when it is available, presumably because they have learned that auditory information is unreliable. We conclude that amusic individuals implicitly draw upon visual information when judging auditory information, even though they have the capacity to explicitly recognize conflicts between these two sensory channels.

© 2016 Elsevier Inc. All rights reserved.

1. Introduction

Congenital amusia is a disorder of musical abilities including pitch perception (Ayotte et al. 2002; Foxtan et al. 2004; Hyde and Peretz 2004; Peretz 2013; Peretz et al. 2002; Peretz and Hyde 2003), melodic contour processing (Albouy et al. 2013), timbre perception (Marin et al. 2012) and pitch memory (Gosselin et al. 2009; Tillmann et al. 2009; Williamson and Stewart 2010). Experimental findings indicate that congenital amusia can also affect aspects of speech perception, such as the recognition of linguistic prosody. Interestingly, however, this deficit occurs mainly when speech is stripped of semantic information: affected individuals rarely show difficulties with natural speech and other complex sounds (e.g., environmental sounds) in everyday life (e.g., Ayotte et al. 2002; Liu et al. 2010; Patel et al. 2005). One possible explanation for this discrepancy is that amusics make use of other available cues, such as semantic or visual information,

to compensate for their auditory deficits. For example, Thompson et al. (2012) found that individuals with amusia show reduced sensitivity to prosody of spoken sentences conveying a happy, sad, tender or irritated emotional state. When queried, the same individuals reported that they often rely on facial expressions and gestures to interpret the moods and feelings of people with whom they interact. Thus, the reduced sensitivity to emotional prosody may not pose a problem to amusic individuals in real life because they can fall back on cues delivered through the visual modality.

A number of researchers have noted the tendency to compensate for unreliable sensory information by drawing upon information from intact sensory systems (e.g., Lessard et al. 1998; Massaro and Light 2004; Neville 1990; Rauschecker 1995). For instance, individuals with hearing loss benefit substantially from visible cues, and hence engage in lip-reading for optimal speech recognition (Grant et al. 1998; Middelweerd and Plomp 1987; Massaro and Cohen 1999). Visual information arising from the face and lips helps to compensate for hearing impairment, leading to more efficient auditory processing (i.e., faster and more accurate recognition of speech). Neurologically, the dependency on visual cues for individuals with a handicap in

* Corresponding author at: Department of Psychology, Macquarie University, NSW 2109, Australia.

E-mail address: xuejing.lu@mq.edu.au (X. Lu).

auditory processing, such as cochlear implant recipients, results in increased brain activation within the visual cortex in response to meaningful sounds (e.g., words and environmental sounds), even in the absence of visual stimulation (Giraud et al. 2000; Giraud et al. 2001; Giraud and Truy 2002; Zatorre 2001). Even for people with normal hearing, speech perception improves considerably when a speaker's lip movements are visible (Besle et al. 2004). As such, it is likely that amusic individuals also draw heavily from visual information during auditory tasks, in order to compensate for their deficits in pitch perception.

Albouy et al. (2015) provided indirect evidence that individuals with congenital amusia use a compensatory strategy. They showed that auditory perception in amusics benefits from the presence of informative visual information. Amusic and control participants were asked to detect a deviant pitch in a 5-tone sequence, which was simultaneously presented with or without visual stimuli – 5 dots appearing consecutively from the left to the right along a horizontal line but at the same vertical position. Even though the visual stimuli provided no task-relevant (pitch) information about the deviant tone (it did not vary in height, for example), the mere presence of the dots facilitated amusics' ability to detect the deviant. The authors argued that the presence of dots provided an additional temporal signal that “prepared” the processing of the concurrent auditory stimuli, thereby enhancing detection of deviant pitches (Jones 1976; Nickerson 1973). However, it is also possible that the visual stimuli improved performance by increasing the confidence levels of participants with amusia. If amusics tend to use visual information in daily life to compensate for their auditory difficulties, then they may have found it reassuring to have a visual accompaniment while judging the auditory signal. In other words, because amusics believe that visual information is potentially valuable, its presence may increase their confidence in decision making, which may benefit performance.

To explore this issue further, we required amusics and normal hearing controls to judge the direction (up or down) of pitch change in two consecutive tones. Pitch change stimuli were presented by themselves (auditory only, henceforth AO) or concurrently with visual cues on a computer screen (audio-visual, henceforth AV). In the AV condition, a sequence of two dots appeared on the screen, the second of which appeared either above or below the first one. Within the AV condition, the pitch change of the two tones was either congruent (second tone higher pitch and second dot higher location, or second tone lower pitch and second dot lower location) or incongruent (second tone higher pitch and second dot lower location, or vice versa) with the spatial change of the two dots. If the visual enhancement effect is entirely task independent, we predicted that amusics should show the same increment in performance for both AV conditions relative to AO condition, because the temporal cues and visual information in the AV-congruent and AV-incongruent displays are identical. However if there is an element of task-dependence in the enhancement effect, then amusics should show a larger increment in performance in the AV-congruent task relative to the AV-incongruent task. Notably, the intervals involved in Albouy et al.'s study (i.e., 12.5, 25 and 50 cents) were below or close to amusics' pitch threshold for pitch change detection (Foxton et al. 2004; Hyde and Peretz 2004; Tillmann et al. 2009). Therefore, we cannot tell whether facilitation by visual information is restricted to ambiguous contexts involving small intervals, or whether it can be observed even for pitch intervals that are well above the pitch threshold for amusic individuals. Typically, people combine and integrate multiple sources of sensory information by weighting them in proportion to their reliability (e.g., Alais and Burr 2004; Ernst and Banks 2002). Therefore, we also examined performance in detecting both near threshold and well-above threshold pitch changes.

The use of congruent and incongruent AV pairings in the present study provided an interesting opportunity to obtain additional information about the neural mechanisms underlying visual enhancement of auditory information in amusics. A number of EEG studies have

reported that multisensory stimuli that are incongruent in spatial location or temporal synchrony elicit a N2 component, a negative polarity event-related potential (ERP) component with a latency of about 200 ms after the multisensory stimulus onset (Forster and Pavone 2008; Lindstrom et al. 2012). This component is strongly associated with conflict detection at both a response level and a stimulus-representation level (Yeung et al. 2004), and its amplitude is typically larger following incongruent than congruent stimuli (Nieuwenhuis et al. 2003). We reasoned that if amusic individuals tend to ignore AV conflicts because of an over reliance on visual information, then the N2 response to incongruent stimuli may be attenuated for amusic individuals but not control participants. However, an inability to detect the AV conflict in amusics (i.e., failure to discriminate incongruent from congruent pairings) would also account for an absence of N2 effect. To assess whether amusics can explicitly detect AV conflicts, we included a task in which participants were instructed to pay attention to both auditory and visual information, and indicate whether the auditory and visual changes were congruent with one another. In addition, the combination of auditory and visual stimuli allowed us to examine the N1 response to multisensory stimuli, which reflects a relatively earlier stage of perceptual processing compared to the N2 component (Luck, 2014). Amusic individuals tend to show normal N1 responses to pitch changes despite their impairments in fine-grained pitch perception (Moreau et al. 2009; Peretz et al. 2005; but see Albouy et al. 2013); but it is unknown whether amusics exhibit typical N1 responses to multisensory stimuli.

2. Materials and method

2.1. Participants

Sixteen individuals with congenital amusia and 16 control participants took part in the present study. To screen for individuals with deficits in pitch processing, three melodic subtests (Scale, Contour, and Interval) of the Montreal Battery of Evaluation of Amusia (MBEA; Peretz et al. 2003) were administered. Participants were diagnosed as congenital amusics when their composite scores (based on the melodic subtests) were equal or less than 65 out of 90 points, that is, 72% correct (Liu et al. 2010). All participants were right-handed and had normal hearing and normal or corrected-to-normal vision. None reported any auditory, neurological, or psychiatric disorder. The amusic and control groups were matched in terms of age, gender, and education (see Table 1). The study was approved by the Macquarie University Ethics Committee, and written informed consent for participation was obtained from all participants prior to testing.

2.2. Stimuli

Auditory and visual stimuli were delivered separately in two unimodal tasks and then combined with each other in two bimodal tasks (see *Tasks and conditions* below). The auditory stimuli comprised 800 ms tones, which had a flute timbre and were generated with the computer software Garageband (Version 6.0.4, Apple Inc., USA). The visual stimuli were white dots (60 × 60 pixels, screen resolution: 1980 × 1024 pixels) presented for 800 ms at the center of a black computer background. Each trial contained two stimuli. We refer to the first and second stimulus in a pair as the *standard* and *target*, respectively. The auditory standard could have one of the following tones in C major scale – C (261.63 Hz), D (293.66 Hz), E (329.63 Hz), F (349.23 Hz), and G (392 Hz). The pitch of the auditory target was shifted upward or downward by 3 and 4 semitones (small interval) or 8 and 9 semitones (large interval) with respect to the standard. These intervals were selected as amusic individuals show a relatively higher threshold for pitch direction discrimination (Liu et al. 2010). The visual standard always appeared at the center of the computer screen, while the visual target could be shifted 300 pixels upward or downward

Table 1
Participants' characteristics, mean \pm SD percent correct on the melodic subtests of MBEA, and independent-sample t-test results between the amusic and control group. Amusic participants performed significantly worse than control participants in the three melodic subtests of MBEA (all $p < 0.001$). However, the two groups were matched in other aspects (all $p > 0.40$). DF refers to the degrees of freedom and is corrected if the equal variances assumption is violated.

	Amusics	Controls	DF	T	p (2-tailed)
Age	21.28 \pm 3.89	20.61 \pm 2.82	30	0.56	0.58
Gender	8F/8M	8F/8M	–	–	–
Year of education	14.38 \pm 1.89	14.13 \pm 1.78	30	0.38	0.70
Years of musical training	0.53 \pm 1.02	0.72 \pm 1.22	30	–0.47	0.64
Hours of music listening daily	2.55 \pm 2.55	1.97 \pm 0.85	18.25	0.87	0.40
MBEA (percent correct)					
Scale	64.58 \pm 10.17	92.50 \pm 6.50	25.49	–9.25***	< 0.001
Contour	67.29 \pm 11.00	86.88 \pm 7.93	30	–5.79***	< 0.001
Interval	64.38 \pm 10.66	85.00 \pm 5.02	21.34	–7.00***	< 0.001
Total	65.42 \pm 4.85	88.13 \pm 4.60	30	–13.59***	< 0.001

*** $p < 0.001$.

along with the vertical line in relation to the standard. In the bimodal tasks, auditory and visual stimuli were presented simultaneously, and auditory and visual target could be congruent or incongruent in terms of the direction of their shift (e.g., *congruent*: both upward; *incongruent*: auditory upward, but visual downward).

2.3. Tasks and conditions

Participants performed four different tasks – two unimodal and two bimodal – in four blocks of trials (for more details, see *Procedure* below). In the AO task, participants were asked to judge whether the auditory target was shifted upward or downward with respect to the standard. In a similar manner, they had to indicate whether the visual target

was shifted upward or downward with respect to the standard in the visual only (henceforth, VO) task. In the AV implicit task, participants performed the same task as in the AO task but were instructed to ignore the visual information. Finally, in the AV explicit task, participants were required to determine whether the visual change in direction was congruent with the change in pitch. Participants performed the four tasks in the following fixed order: (1) AO, (2) AV implicit, (3) AV explicit, and (4) VO.

2.4. Procedure

As illustrated in Fig. 1, each trial began with a fixation-cross that appeared for 500–800 ms randomly in the centre of the screen. Subsequently, standard and target stimuli (each of 800 ms duration) were presented consecutively with a jittered inter-stimulus interval (ISI) of 300–500 ms. To minimize the possibility of interference in the response stage, participants made their responses by pressing one of two response keys with no time limit. The inter-trial interval (ITI) was also jittered between 300 and 500 ms. Participants underwent 240 trials in total in each AV task, that is, 60 trials per interval (small and large) and congruency (congruent and incongruent). The total number of trials was 120 (60 trials per interval) for the AO task, and 60 for the VO task. All trials were presented in randomized order. Participants were given a break after every 60 trials. They completed ten practice trials prior to each block. Feedback was provided in the practice but not in the actual experiment. Participants were seated approximately 50 cm from the computer screen in an electrically shielded and sound-attenuated room with dimmed light. All sounds were presented at a comfortable level via headphones (Sennheiser HD 280).

2.5. EEG Recording

The EEG was recorded with a sampling rate of 1000 Hz from 32 Ag-AgCl electrodes placed according to the extended International 10–20 electrode system and mounted to a head-cap (EASYCAP GmbH, Germany). The left mastoid electrode served as reference electrode during the recording. Four additional electrodes were employed to monitor horizontal and vertical eye movements. Electrode impedances were below 5 k Ω . Using a SynAmps 2.0 RT amplifier (Compumedics Neuroscan, USA), the EEG was filtered online with an analogue band-pass filter (0.05–200 Hz).

2.6. EEG processing

Offline processing of the EEG was performed in MATLAB (R2013b; Mathwork, USA) using the EEGLAB toolbox (Delorme and Makeig 2004). The raw data was first downsampled to 500 Hz and re-referenced to the average of the left and right mastoids. It was then segmented into one-second epochs extending from 200 ms

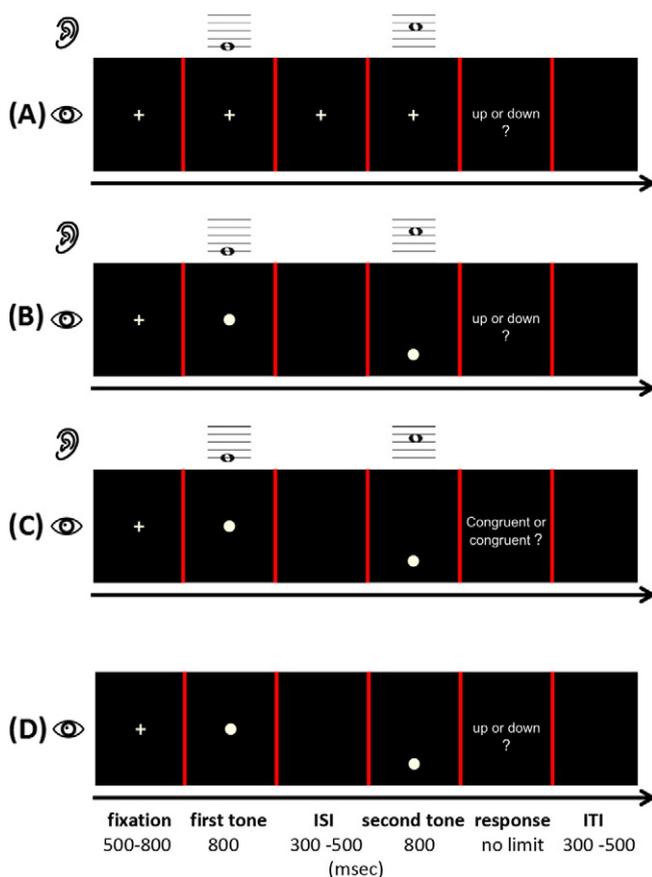


Fig. 1. Schematic illustration of the stimuli employed for four experimental tasks: (A) AO task, (B) AV implicit task, (C) AV explicit task, and (D) VO task.

before to 800 ms after the onset of the target. Trials with incorrect response¹ or in which the potential exceeded $\pm 150 \mu\text{V}$ were excluded from further data processing. The mean of each epoch was removed (see [Groppe et al. 2009](#) for details) before an independent component analysis (ICA) was performed using the *runica* algorithm. We used an ICA-based method to identify and reject trials with unusual activity (i.e., rejection by kurtosis; see [Delorme et al. 2001](#) for details). In addition, we employed the EEGLAB plugin *ADJUST* to automatically reject ocular ICs ([Mognon et al. 2011](#)). After IC rejection, the EEG data was low-pass filtered with a Windowed Sinc FIR Filter ([Widmann and Schröger 2012](#)) and a cut-off frequency at 30 Hz (Blackman window; filter order: 276). Epochs were subsequently averaged per participant, condition and task and baseline corrected by subtracting a pre-stimulus interval of 200 ms.

2.7. ERP Analysis

The main analysis focused on the comparison of the ERP components of interest involved in congruent and incongruent visual information processing between amusic and control groups in both implicit and explicit AV tasks. Based on visual inspection of the grand averages and previous studies, two pronounced ERP deflections within the following time windows were selected: 100–180 ms and 260–380 ms after the stimulus onset, encompassing the N1 component and N2–P3 complex respectively (e.g., [Folstein and van Petten 2008](#); [Talsma et al. 2007](#)). The N2 ERP overlapped with a larger and longer duration P3. Consequently, although the N2 is a negative-going ERP, peak amplitudes in both congruent and incongruent conditions were positive. Note that even in such circumstances, the experimental contrast (e.g., subtracting the ERP in the congruent condition from that in the incongruent condition) yields a negative difference N2 “component” ([Luck 2014](#)). Additionally, we identified two negative ERP deflections in the AO and VO tasks corresponding to the auditory and visual N1 ERPs, peaking within a post-stimulus time window of 90–150 ms ([Näätänen and Picton 1987](#)) and 120–180 ms ([Luck 2014](#); [Vogel and Luck 2000](#)), respectively. A positive ERP deflection within 200–380 ms after stimulus onset was found in both AO and VO tasks, although both groups showed a negative-going deflection in response to targets with small interval (see [Fig. 3](#)) within this time window in the AO task. Finally, for each participant, condition and task, we computed the mean amplitudes within these pre-defined time windows. Furthermore, given the topographical distributions of ERP components of interest, the analyses were conducted at the central sites (CP3, CPZ, CP4, C3, CZ, and C4) for the AO and two AV tasks, and at the centroparietal sites (CP3, CPZ, CP4, P3, PZ, and P4) for the VO task. Scalp topographies were assessed using all scalp electrodes.

2.8. Statistical analysis

For the statistical analysis, task performance on unimodal tasks and the AV implicit task was evaluated by mean percent correct (PC). However, we computed *d*'-prime (*d*') ([Macmillan and Creelman 2005](#)) for the AV explicit task to exclude the potential response biases reported in previous amusia studies (e.g., [Henry and McAuley 2013](#); [Pfeifer and Hamann 2015](#)). Repeated-measures ANOVA for the AV implicit task was conducted, with the between-subject factor Group (amusic and control) and the within-subject factors Interval (large and small) and Congruency (congruent and incongruent). The factor Congruency was excluded from the analysis of the AV explicit and AO tasks, and independent two-sample *t*-tests were conducted for the VO task. Similar analyses were conducted on the mean amplitudes of ERPs for each task and

time window. Below we report and discuss in detail only significant results of interest (see Supplementary Tables 1–6 for full results).

3. Results

3.1. Behavioural results

We first compared the task performance of the amusic and control group in the two unimodal tasks. For the AO task, the results revealed a significant interaction between Group and Interval, $F(1, 30) = 14.78$, $p = 0.001$, $\eta_p^2 = 0.33$. Further analyses showed that amusic participants performed as well as controls when the interval was large (amusics: $M = 0.95$, $SE = 0.01$; controls: $M = 0.96$, $SE = 0.01$), $F(1, 30) = 0.03$, $p = 0.87$, $\eta_p^2 < 0.01$. However, their performance significantly deteriorated compared to that of controls when the interval was small (amusics: $M = 0.80$, $SE = 0.02$; controls: $M = 0.89$, $SE = 0.02$), $F(1, 30) = 7.71$, $p < 0.01$, $\eta_p^2 = 0.21$. In addition, both amusic and control groups performed better when interval size was large, $F(1, 30) = 89.31$, $p < 0.001$, $\eta_p^2 = 0.75$, and $F(1, 30) = 16.11$, $p < 0.001$, $\eta_p^2 = 0.35$, respectively. For the VO task, both groups performed exceedingly well ($> 99\%$); the statistical results confirmed that there was no significant group difference, $t(30) = 1.75$, $p = 0.10$.

Next we examined how participants' performance was affected by the presence of visual information in the AV tasks. When participants were instructed to ignore visual information, as in the AV implicit task, we found all main effects and two-way interactions were significant (all $p < 0.01$). Although there was a trend, the three-way interaction of Group, Congruency and Interval did not reach significance, $F(1, 30) = 3.49$, $p = 0.07$, $\eta_p^2 = 0.10$. Further examination of the interaction between Group and Congruency revealed that the performance of amusics was significantly worse than controls in small (amusics: $M = 0.72$, $SE = 0.03$; controls: $M = 0.87$, $SE = 0.03$) and large interval conditions (amusics: $M = 0.92$, $SE = 0.01$; controls: $M = 0.96$, $SE = 0.02$), $F(1, 30) = 12.63$, $p = 0.001$, $\eta_p^2 = 0.30$, when the visual change in direction was incongruent with the pitch change. However, amusics (small: $M = 0.88$, $SE = 0.02$; large: $M = 0.97$, $SE = 0.01$) performed just as well as controls (small: $M = 0.93$, $SE = 0.02$; large: $M = 0.98$, $SE = 0.01$), $F(1, 30) = 3.64$, $p = 0.07$, $\eta_p^2 = 0.11$, when visual and auditory information was congruent. In addition, all participants were affected by unattended visual information to some degree, suggested by significant congruency effect in both amusic, $F(1, 30) = 58.39$, $p < 0.001$, $\eta_p^2 = 0.66$, and control groups, $F(1, 30) = 9.71$, $p < 0.01$, $\eta_p^2 = 0.24$.

To further explore the extent to which unattended visual information affects one's pitch change direction identification, additional comparisons (separate paired-sample *t*-tests) between AO and implicit AV tasks were conducted for each group (see [Fig. 2](#)). The mean difference is considered significant at a significance level of 0.0125, due to the adjustment for multiple comparisons. When congruent visual information was presented, task performance of both amusics and controls was significantly improved in small interval condition (all $p < 0.006$), but not in large interval condition due to a ceiling effect (all $p > 0.03$). More interestingly, amusics' performance was significantly worse in the presence of incongruent visual information, regardless of the interval size (all $p < 0.005$). By contrast, controls' performance was unaffected by incongruent visual information (all $p > 0.14$).

In the AV explicit task, participants judged the congruency of the visual and auditory information. Unsurprisingly, participants' sensitivity to AV incongruence increased with the increase of interval size between standard and target stimuli, revealed by a significant interval effect, $F(1, 30) = 70.28$, $p < 0.001$, $\eta_p^2 = 0.70$, suggesting. Moreover, a significant group difference, $F(1, 30) = 5.07$, $p = 0.03$, $\eta_p^2 = 0.15$, and Interval \times Group interaction were found, $F(1, 30) = 4.61$, $p = 0.04$, $\eta_p^2 = 0.13$. Post-hoc analysis revealed that amusics performed significantly worse in comparison to controls in small interval condition (amusics: $M = 1.77$, $SE = 0.26$; controls: $M = 2.75$; $SE = 0.26$),

¹ The analyses on all trials (including the trials with incorrect response) yielded the same results, except that the interaction between Group and Congruency was non-significant within the N1 time window for the AV explicit task ($p = 0.15$). Please see full results in the Supplementary Table 5.

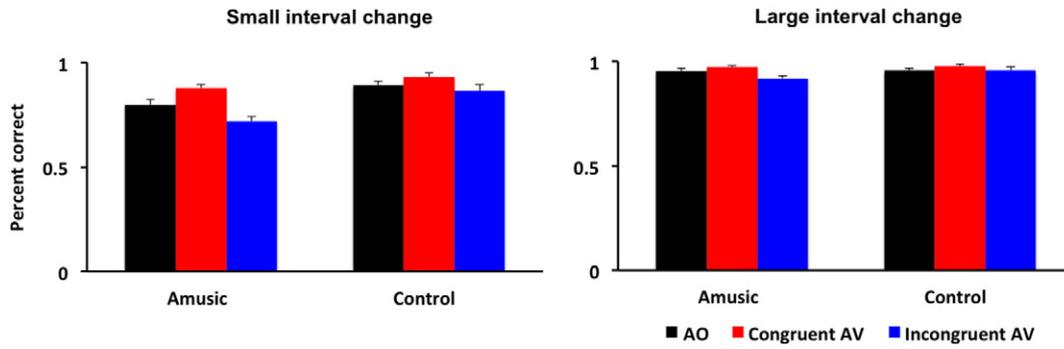


Fig. 2. Behavioural performance on the AO (black bar) and AV implicit (congruent AV pairings: red bar; incongruent AV pairings: blue bar) tasks evaluated by percent correct. Error bars represent 1 SE.

$F(1, 30) = 7.33, p < 0.05, \eta_p^2 = 0.20$, but not in large interval condition (amusics: $M = 3.23, SE = 0.21$; controls: $M = 3.61; SE = 0.21$), $F(1, 30) = 1.16, p = 0.22, \eta_p^2 = 0.05$.

3.2. ERP results

Amusic and control participants showed no significant differences in the AO and VO task within any of the selected time windows (all $p > 0.05$, see Fig. 3 and 4).

In line with behavioural results of the AO task, the amplitude of N1 component was larger for the large than small pitch change, $F(1, 30) = 8.14, p < 0.01, \eta_p^2 = 0.21$. No other significant effect was observed in the unimodal tasks. Group differences were only found in the two bimodal tasks.

For the implicit task (see Fig. 5), both groups exhibited larger amplitude response to AV pairings with large interval change within the N1 and N2–P3 time windows, as revealed by the main effect of

Interval with no other interactions, $F(1, 30) = 13.91, p = 0.001, \eta_p^2 = 0.32$, and $F(1, 30) = 48.36, p < 0.001, \eta_p^2 = 0.62$, respectively. More importantly, the statistical results yielded a significant interaction between Congruency and Group in the N2–P3 time window, $F(1, 30) = 4.75, p < 0.05, \eta_p^2 = 0.14$. For the control group, the negativity of the N2–P3 complex was greater for incongruent trials ($M = 2.15 \mu V, SE = 0.83 \mu V$) than for congruent trials ($M = 3.77 \mu V, SE = 0.70 \mu V$), $F(1, 30) = 12.02, p < 0.01, \eta_p^2 = 0.29$. For the amusic group, the enhanced negativity of the N2–P3 complex for incongruent trials was not found (incongruent: $M = 1.51 \mu V, SE = 0.83 \mu V$; congruent: $M = 1.69 \mu V, SE = 0.70 \mu V$), $F(1, 30) = 0.15, p = 0.70, \eta_p^2 < 0.01$. No other significant effect was observed in the AV implicit task.

By contrast, in the AV explicit task, both control and amusic groups showed enhanced N2–P3 complex response to incongruent AV pairings (controls: $M = 2.16 \mu V, SE = 0.84 \mu V$; amusics: $M = 1.00 \mu V, SE = 0.84 \mu V$) relative to congruent AV pairings (controls: $M = 3.61 \mu V, SE = 0.79 \mu V$; amusics: $M = 2.50 \mu V, SE = 0.79 \mu V$) AV pairings,

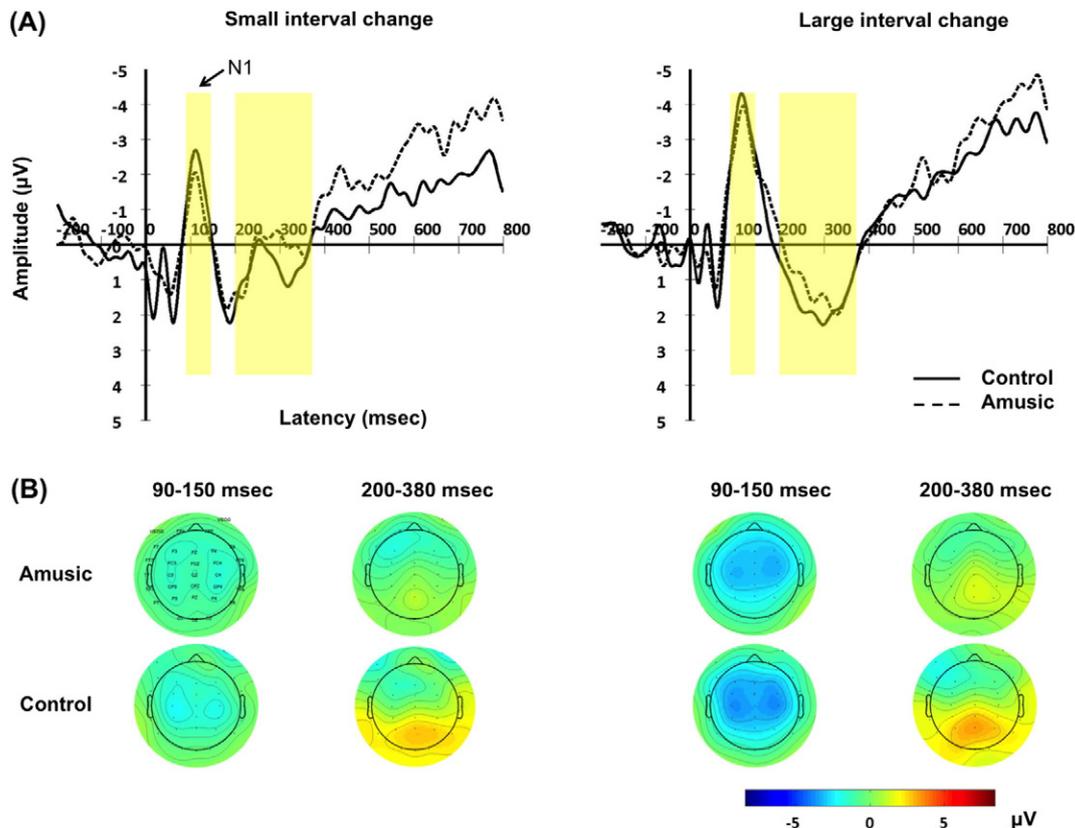


Fig. 3. ERP results of the AO task. (A) Grand-averaged ERPs of amusics (dash line) and controls (solid line) at electrode CZ. No group difference was observed in two time windows of interest (highlighted by yellow). (B) Group average topographic maps for each time window.

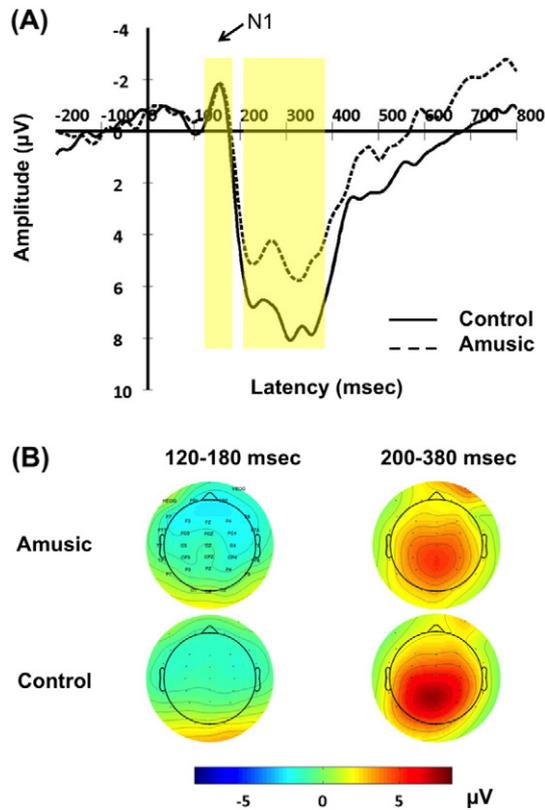


Fig. 4. ERP results of the VO task. (A) Grand-averaged ERPs of amusics and controls at electrode CPZ. No significant group difference was observed in either time window of interest (highlighted by yellow). (B) Group average topographic maps for each time window.

as the main effect of Congruency indicated, $F(1, 30) = 29.08$, $p < 0.001$, $\eta^2 = 0.49$. Furthermore, the interaction between Congruency and Group was found to be non-significant, $F(1, 30) = 0.01$, $p = 0.93$, $\eta^2 < 0.001$. Within the N1 time window, however, there was a significant interaction between Group and Congruency in the AV explicit task, $F(1, 30) = 5.84$, $p < 0.05$, $\eta^2 = 0.16$, with no three-way interaction involving Interval, $F(1, 30) = 0.19$, $p = 0.67$, $\eta^2 < 0.01$. As can be seen in Fig. 6, the mean N1 amplitude elicited by incongruent AV pairs is larger than that elicited by congruent AV pairings in the amusic group (congruent: $M = -2.29 \mu\text{V}$, $SE = 0.76 \mu\text{V}$; incongruent: $M = -3.20 \mu\text{V}$, $SE = 0.72 \mu\text{V}$), $F(1, 30) = 8.87$, $p < 0.01$, $\eta^2 = 0.23$. No such difference was observed for the control group (congruent: $M = -2.56 \mu\text{V}$, $SE = 0.76 \mu\text{V}$; incongruent: $M = -2.43 \mu\text{V}$, $SE = 0.72 \mu\text{V}$), $F(1, 30) = 0.19$, $p = 0.67$, $\eta^2 < 0.01$. It should be noted that the analyses on all trials (including trials with incorrect response) showed a different picture, as the interaction between Group and Congruency within the N1 time window was no longer significant, $F(1, 30) = 2.14$, $p = 0.15$, $\eta^2 = 0.07$. In addition, an enhanced negativity of the N1 and N2–P3 complex was elicited by large pitch changes across participants, suggested by the main effect of Interval, $F(1, 30) = 11.19$, $p < 0.01$, $\eta^2 = 0.27$, and $F(1, 30) = 30.40$, $p < 0.001$, $\eta^2 = 0.50$, respectively, which was in line with the results of the AV implicit task.

4. Discussion

Congenital amusia is characterised by a variety of deficits in musical perception. The present study tested the hypothesis that these auditory processing deficits shown by amusic individuals can be compensated for by visuospatial information presented simultaneously. There were several salient observations from the present results:

- (1) amusics performed significantly worse than controls in the AO task, when pitch changes were small;

- (2) the two groups performed equally well in the VO task;
- (3) in the presence of unattended visual information, the performance of amusics improved significantly when it was congruent with auditory information involving small interval change, but deteriorated significantly when they were incongruent;
- (4) the performance of controls was only affected by unattended visual change in direction, when it was congruent with the direction of the small pitch change;
- (5) in controls, a robust N2–P3 complex was elicited by incongruent AV pairings, whether implicit or explicit;
- (6) in amusics, the N2–P3 complex was elicited by explicit AV pairings but not by implicit AV pairings.

4.1. Behavioural results

Amusic participants displayed difficulties in the AO task when the interval was small (i.e., 3 or 4 semitones), whereas they could identify the direction of the interval just as well as controls when the interval was large (i.e., 8 or 9 semitones). This finding is consistent with previous claims that amusics can discriminate large pitch differences because such differences exceed their threshold for discriminating pitch change direction (Liu et al. 2010; Williamson and Stewart 2010).

Bimodal stimuli are typically detected and recognized more rapidly and accurately than unimodal stimuli by both non-amusic (Besle et al. 2004; Giard and Peronnet 1999; Laurienti et al. 2004; Teder-Sälejärvi et al. 2005; Teder-Sälejärvi et al. 2002) and amusic participants (Albouy et al. 2015). Indeed, when congruent visual information was supplied in the AV implicit task, amusics' ability to identify the contour of small intervals improved significantly than when auditory stimuli presented alone. However, our results also showed that amusics' performance deteriorated significantly when the visual information was incongruent with the auditory information. More interestingly, the deterioration in performance shown by amusic participants, especially in large interval condition, did not result from an inability to detect the AV incongruence, as suggested by the results of the AV explicit task. Instead, the finding in the AV implicit task is most likely to reveal visual influences on auditory processing for amusics. The interference may occur at either the perceptual stage, where amusics perceived the changes in pitch direction, or the response stage, where they have to respond. Given that the task required participants to pay attention to the auditory modality while ignoring the visual modality, and allowed a non-speeded response, it is reasonable to conclude that the amusic individuals are readily affected by the unattended visual stimuli at the perceptual stage of auditory processing.

4.2. Neurophysiological results

For the AO task, amusics displayed comparable ERPs in comparison with control participants (see Fig. 3) despite their behaviourally impaired performance for small changes in pitch. This discrepancy between behavioral and neural results for the small interval condition has been observed in other investigations (Peretz et al. 2009; Peretz et al. 2005), and may indicate that early stages of auditory processing (as reflected in ERPs) are intact in amusics, when the task is not demanding (e.g., Moreau et al. 2009; see Albouy et al. 2013 for further discussion). With regard to the VO task, both amusic and control participants performed equally well and showed no difference in terms of ERPs.

For the AV implicit task, control participants exhibited an increase in the negativity of N2–P3 complex to incongruent relative to congruent AV pairings, suggesting a processing of conflict detection even when the visual information was task-irrelevant (Forster and Pavone 2008; Lindstrom et al. 2012; Nieuwenhuis et al. 2003; Yeung et al. 2004). By contrast, amusics failed to show a conflict response in this task,

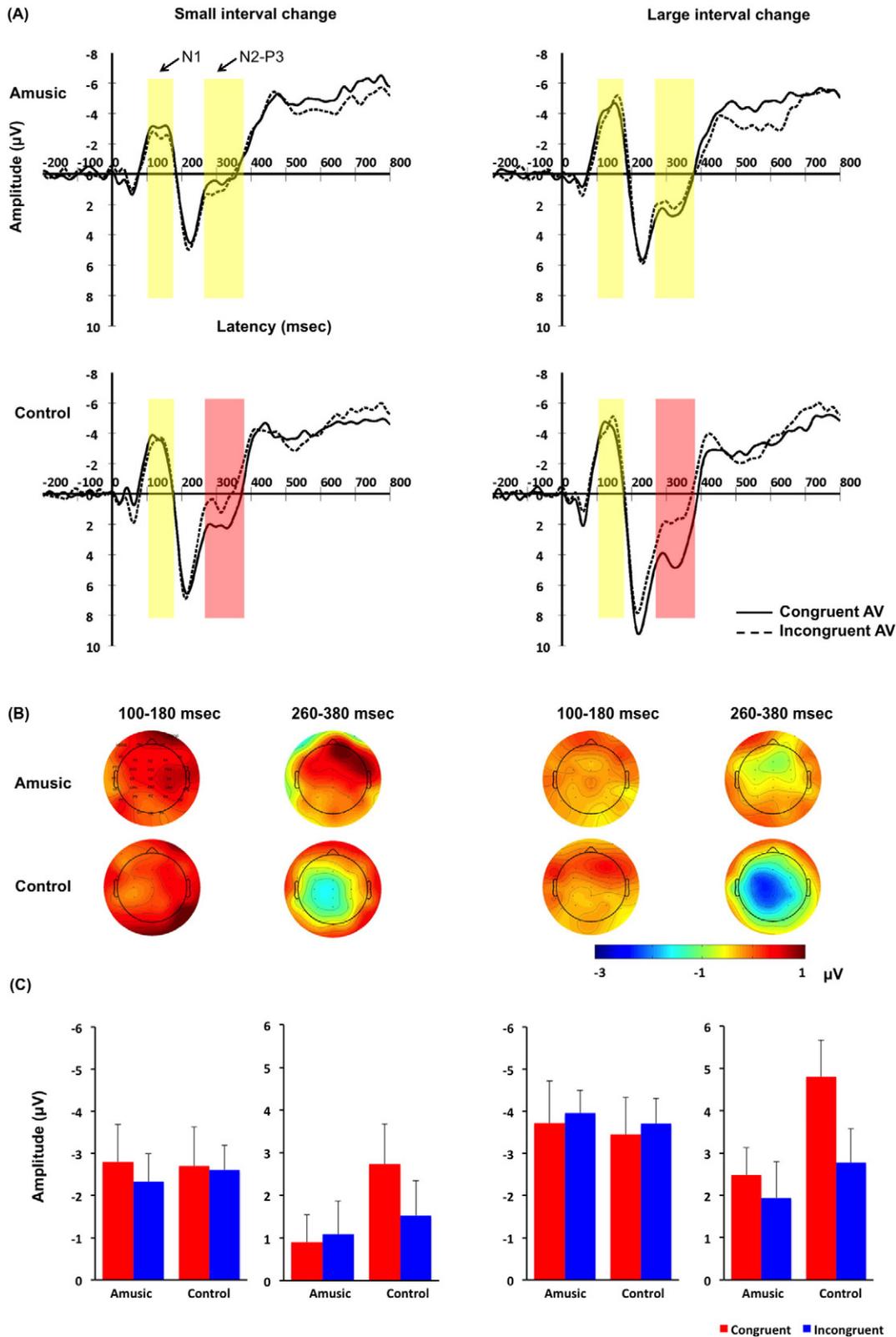


Fig. 5. Results of the AV implicit task. (A) Grand-averaged ERPs of amusic and control groups at electrode CZ in response to auditory stimuli with congruent (solid line) and incongruent (dash line) visual stimuli. For each time window, a significant congruency effect is highlighted by red, and a non-significant congruency effect is highlighted by yellow. (B) Topographic maps of difference amplitude by subtracting congruent trials from incongruent trials for each group within each time window of interest. (C) Mean amplitude in response to congruent trials (red bar) and incongruent trials (blue bar) over ROI for each time window. Error bars represent 1 SE.

as reflected by an absence of N2–P3 effect. However, the absence of N2–P3 effect cannot be explained by impaired neural circuitry, as amusic participants exhibited a similar increase in the negativity of N2–P3

complex to incongruent pairings when compared with congruent AV pairings in the AV explicit task, regardless of whether the interval was small or large. Therefore, the absence of N2–P3 effect in amusic

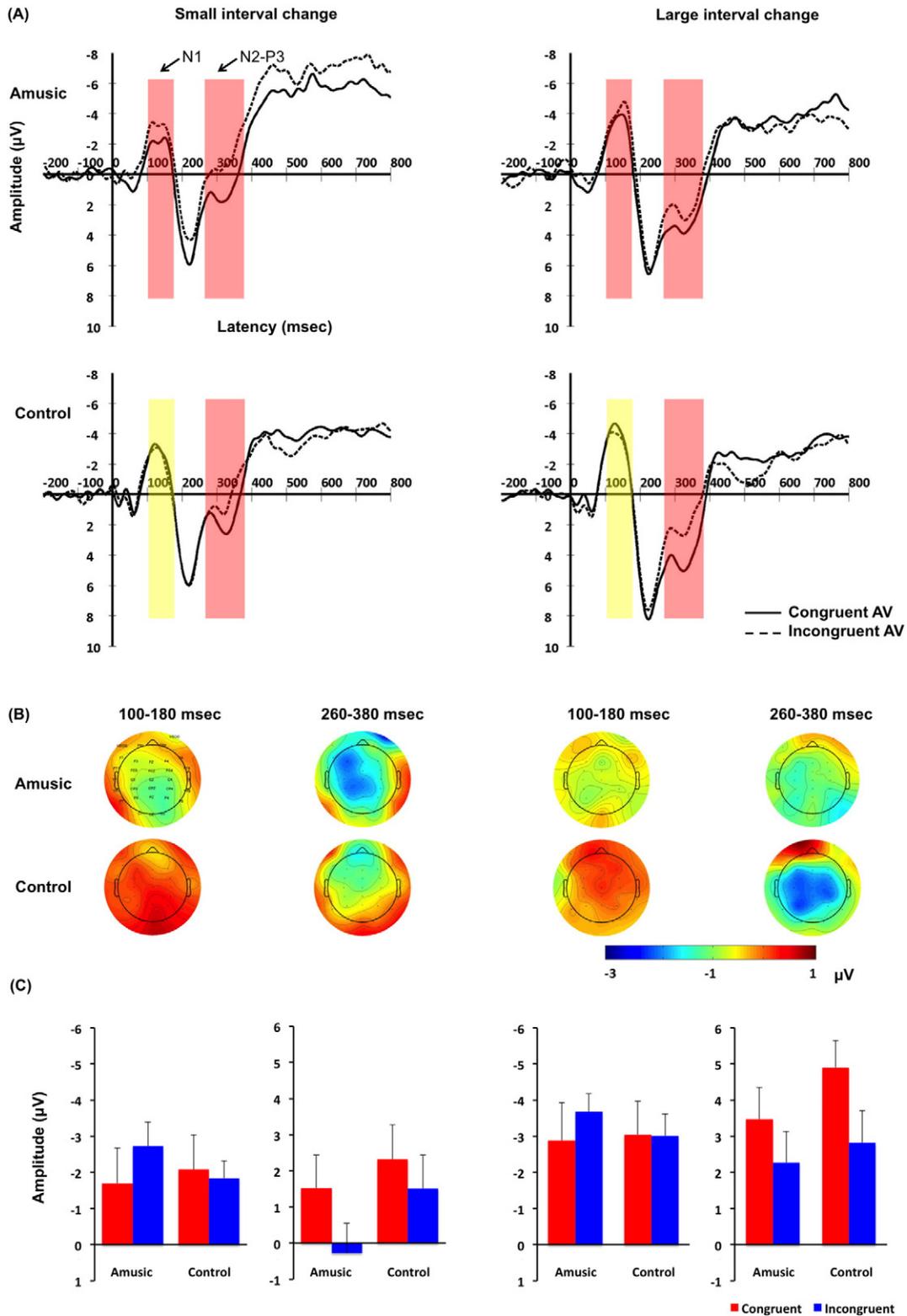


Fig. 6. Results of the AV explicit task. (A) Grand-averaged ERPs of amusic and control groups at electrode CZ in response to auditory stimuli with congruent (solid line) and incongruent (dash line) visual stimuli. For each time window, a significant congruency effect is highlighted by red, and a non-significant congruency effect is highlighted by yellow. (B) Topographic maps of difference amplitude by subtracting congruent trials from incongruent trials for each group within each time window of interest. (C) Mean amplitude in response to congruent trials (red bar) and incongruent trials (blue bar) over ROI for each time window. Error bars represent 1 SE.

individuals observed in the AV implicit task is more likely due to an unconscious neglect of AV conflict. This explanation is further supported by the discrepancy between behavioural and ERP results in the AV implicit task, in which amusics depended on unattended visual information to complete the auditory task, as behavioural results suggested,

thereby ignoring AV conflicts, as reflected by the absence of N2–P3 complex.

Another discrepancy between behavioural and ERP results was found in the AV explicit task, in which amusic showed reduced sensitivity to AV incongruence in small interval condition, while

exhibiting a comparable N2–P3 effect (i.e., enhanced N2–P3 complex in response to incongruent than congruent AV pairings), to controls. The failure of amusics to detect AV incongruence with a small interval change may reflect a limited access to pitch information consciously, even though their neural response to incongruent AV pairings differed from that to congruent pairings. This observation is in good agreement with the “pitch awareness hypothesis”, which emphasises that pitch information can be registered but not accessed consciously by individuals who suffer from congenital amusia (e.g., Hyde et al. 2011; Loui et al. 2009; Loui et al. 2008; Loui et al. 2011; Lu et al. 2015; Moreau et al. 2013; 2009; Peretz et al. 2009). One predication of the “pitch awareness hypothesis” is that amusics would show low confidence level in tasks that require explicit pitch processing owing to a lack of pitch awareness (Lu et al., under review). Therefore, it is possible that the discrepancy between behavioural and ERP results observed in the AV explicit task results from a low confidence level in cases with small interval changes.

In line with previous study examining the ERPs to neutral AV stimuli (Stekelenburg and Vroomen 2007), the N1 component shown by controls in both AV tasks did not depend on the informational congruency between audio and visual stimuli, revealed by comparable N1 amplitude for incongruent and congruent AV pairings. However, we observed significantly larger N1 amplitudes for incongruent than congruent AV pairings in the amusic group but not the control group in the explicit AV task, when the trials with correct response included only (i.e., when participants detected the AV incongruence successfully). Interestingly, the effect shown by amusics was no longer significant when trials with incorrect response were also included. Since there were few trials with incorrect response, we can only speculate that the enhanced N1 for incongruent AV pairings is necessary for detecting AV conflict for amusics but not controls, resulting from a (strategic) mechanism that they use to compensate for deficient processing of pitch.

4.3. Implications

Based on the observations of the present study, we argue that amusics tend to make use of available visual cues in order to compensate for their difficulties in pitch perception. The tendency by amusics to use visual cues is compatible with the “optimal-integration hypothesis,” which suggests that perceivers are more likely to rely on one modality over the other depending on how reliable the information is (Ernst and Banks 2002; Ernst and Bulthoff 2004). For instance, when presented with flashed visual stimuli accompanied by auditory beeps, the perceived number of flashes is influenced by the number of beeps (Shams et al. 2000). In this case, audition dominates over vision, as the auditory modality is more reliable and more precise at determining temporal information.

For amusic individuals, their pitch impairment means that auditory information is not reliable. Therefore, amusic individuals tend to use contextual or facial cues to boost their auditory perception (Albouy et al. 2015; Thompson et al. 2012). Our results show that amusics rely on unattended visual information even when pitch differences are well above their threshold. This finding extends the “optimal-integration hypothesis” to circumstances where there is a long-term bias in weighing visual over auditory information established on the basis of daily experience, generalizing the phenomenon from a particular source of reliability in a sensory modality to the entire modality.

5. Conclusion

In summary, the present study is the first ERP study showing that individuals with congenital amusia rely heavily on unattended visual information when doing auditory task due to their deficits in auditory processing, providing the theoretical basis for using visual information

to improve amusics' auditory perception. Furthermore, the presence of visual information may boost auditory encoding at the perceptual stage in congenital amusia, thereby leading to more accurate neural representations of input stimuli, which yields new insights into the mechanisms of the combination of auditory and visual information in this special population.

Conflict of interest

The authors declare no conflict of interest.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <http://dx.doi.org/10.1016/j.neuroimage.2016.04.043>.

References

- Alais, D., Burr, D., 2004. The ventriloquist effect results from near-optimal bimodal integration. *Curr. Biol.* 14, 257–262.
- Albouy, P., Mattout, J., Bouet, R., Maby, E., Sanchez, G., Aguera, P.E., Daligault, S., Delpuech, C., Bertrand, O., Caclin, A., Tillmann, B., 2013. Impaired pitch perception and memory in congenital amusia: the deficit starts in the auditory cortex. *Brain* 136, 1639–1661.
- Albouy, P., Leveque, Y., Hyde, K.L., Bouchet, P., Tillmann, B., Caclin, A., 2015. Boosting pitch encoding with audiovisual interactions in congenital amusia. *Neuropsychologia* 67, 111–120.
- Ayotte, J., Peretz, I., Hyde, K., 2002. Congenital amusia: a group study of adults afflicted with a music-specific disorder. *Brain* 125, 238–251.
- Besle, J., Fort, A., Delpuech, C., Giard, M.H., 2004. Bimodal speech: early suppressive visual effects in human auditory cortex. *Eur. J. Neurosci.* 20, 2225–2234.
- Delorme, A., Makeig, S., 2004. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J. Neurosci. Methods* 134, 9–21.
- Delorme, A., Makeig, S., Sejnowski, T., 2001. Automatic artifact rejection for EEG data using high-order statistics and independent component analysis. *Proceedings of the Third International ICA Conference*, pp. 9–12.
- Ernst, M.O., Banks, M.S., 2002. Humans integrate visual and haptic information in a statistically optimal fashion. *Nature* 415, 429–433.
- Ernst, M.O., Bulthoff, H.H., 2004. Merging the senses into a robust percept. *Trends Cogn. Sci.* 8, 162–169.
- Folstein, J.R., van Petten, C., 2008. Influence of cognitive control and mismatch on the N2 component of the ERP: a review. *Psychophysiology* 45, 152–170.
- Forster, B., Pavone, E.F., 2008. Electrophysiological correlates of crossmodal visual distractor congruency effects: evidence for response conflict. *Cogn. Affect. Behav. Neurosci.* 8, 65–73.
- Foxton, J.M., Dean, J.L., Gee, R., Peretz, I., Griffiths, T.D., 2004. Characterization of deficits in pitch perception underlying “tone deafness”. *Brain* 127, 801–810.
- Giard, M.H., Peronnet, F., 1999. Auditory–visual integration during multimodal object recognition in humans: a behavioral and electrophysiological study. *J. Cogn. Neurosci.* 11, 473–490.
- Giraud, A.L., Truy, E., 2002. The contribution of visual areas to speech comprehension: a PET study in cochlear implants patients and normal-hearing subjects. *Neuropsychologia* 40, 1562–1569.
- Giraud, A.L., Lorenzi, C., Ashburner, J., Wable, J., Johnsrude, I., Frackowiak, R., Kleinschmidt, A., 2000. Representation of the temporal envelope of sounds in the human brain. *J. Neurophysiol.* 84, 1588–1598.
- Giraud, A.L., Price, C.J., Graham, J.M., Truy, E., Frackowiak, R.S., 2001. Cross-modal plasticity underpins language recovery after cochlear implantation. *Neuron* 30, 657–664.
- Gosselin, N., Jolicoeur, P., Peretz, I., 2009. Impaired memory for pitch in congenital amusia. *Ann. N. Y. Acad. Sci.* 1169, 270–272.
- Grant, K.W., Walden, B.E., Seitz, P.F., 1998. Auditory–visual speech recognition by hearing-impaired subjects: consonant recognition, sentence recognition, and auditory–visual integration. *J. Acoust. Soc. Am.* 103, 2677–2690.
- Groppe, D.M., Makeig, S., Kutas, M., 2009. Identifying reliable independent components via split-half comparisons. *NeuroImage* 45, 1199–1211.
- Henry, M.J., McAuley, J.D., 2013. Failure to apply signal detection theory to the Montreal Battery of Evaluation of Amusia may misdiagnose amusia. *Mus. Percept.* 30, 480–496.
- Hyde, K.L., Peretz, I., 2004. Brains that are out of tune but in time. *Psychol. Sci.* 15, 356–360.
- Hyde, K.L., Zatorre, R.J., Peretz, I., 2011. Functional MRI evidence of an abnormal neural network for pitch processing in congenital amusia. *Cereb. Cortex* 21, 292–299.
- Jones, M.R., 1976. Levels of structure in the reconstruction of temporal and spatial serial patterns. *J. Exp. Psychol.* 2, 475–488.
- Laurienti, P.J., Kraft, R.A., Maldjian, J.A., Burdette, J.H., Wallace, M.T., 2004. Semantic congruence is a critical factor in multisensory behavioral performance. *Exp. Brain Res.* 158, 405–414.
- Lessard, N., Pare, M., Lepore, F., Lassonde, M., 1998. Early-blind human subjects localize sound sources better than sighted subjects. *Nature* 395, 278–280.

- Lindstrom, R., Paavilainen, P., Kujala, T., Tervaniemi, M., 2012. Processing of audiovisual associations in the human brain: dependency on expectations and rule complexity. *Front. Psychol.* 3, 159.
- Liu, F., Patel, A.D., Fourcin, A., Stewart, L., 2010. Intonation processing in congenital amusia: discrimination, identification and imitation. *Brain* 133, 1682–1693.
- Loui, P., Guenther, F.H., Mathys, C., Schlaug, G., 2008. Action–perception mismatch in tone-deafness. *Curr. Biol.* 18, R331–R332.
- Loui, P., Alsop, D., Schlaug, G., 2009. Tone deafness: a new disconnection syndrome? *J. Neurosci.* 29, 10215–10220.
- Loui, P., Kroog, K., Zuk, J., Winner, E., Schlaug, G., 2011. Relating pitch awareness to phonemic awareness in children: implications for tone-deafness and dyslexia. *Front. Psychol.* 2, 111.
- Lu, X., Ho, H.T., Liu, F., Wu, D., Thompson, W.F., 2015. Intonation processing deficits of emotional words among Mandarin Chinese speakers with congenital amusia: an ERP study. *Front. Psychol.* 6, 385.
- Luck, S., 2014. *An Introduction to the Event-related Potential Technique*. MIT Press, Cambridge.
- Macmillan, N.A., Creelman, C.D., 2005. *Detection Theory: A User's Guide*. Erlbaum, Mahwah, NJ.
- Marin, M.M., Gingras, B., Stewart, L., 2012. Perception of musical timbre in congenital amusia: categorization, discrimination and short-term memory. *Neuropsychologia* 50, 367–378.
- Massaro, D.W., Cohen, M.M., 1999. Speech perception in perceivers with hearing loss. *J. Speech Lang. Hear. Res.* 42, 21–41.
- Massaro, D.W., Light, J., 2004. Using visible speech to train perception and production of speech for individuals with hearing loss. *J. Speech Lang. Hear. Res.* 47, 304–320.
- Middelweerd, M.J., Plomp, R., 1987. The effect of speechreading on the speech-reception threshold of sentences in noise. *J. Acoust. Soc. Am.* 82, 2145–2147.
- Mognon, A., Jovicich, J., Bruzzone, L., Buiatti, M., 2011. ADJUST: an automatic EEG artifact detector based on the joint use of spatial and temporal features. *Psychophysiology* 48, 229–240.
- Moreau, P., Jolicœur, P., Peretz, I., 2009. Automatic brain responses to pitch changes in congenital amusia. *Ann. N. Y. Acad. Sci.* 1169, 191–194.
- Moreau, P., Jolicœur, P., Peretz, I., 2013. Pitch discrimination without awareness in congenital amusia: evidence from event-related potentials. *Brain Cogn.* 81, 337–344.
- Näätänen, R., Picton, T., 1987. The N1 wave of the human electric and magnetic response to sound: a review and an analysis of the component structure. *Psychophysiology* 24, 375–425.
- Neville, H.J., 1990. Intermodal competition and compensation in development. Evidence from studies of the visual system in congenitally deaf adults. *Ann. N. Y. Acad. Sci.* 608, 71–91.
- Nickerson, R.S., 1973. Intersensory facilitation of reaction time: energy summation or preparation enhancement? *Psychol. Rev.* 80, 489–509.
- Nieuwenhuis, S., Yeung, N., van den Wildenberg, W., Ridderinkhof, K.R., 2003. Electrophysiological correlates of anterior cingulate function in a go/no-go task: effects of response conflict and trial type frequency. *Cogn. Affect. Behav. Neurosci.* 3, 17–26.
- Patel, A.D., Foxton, J.M., Griffiths, T.D., 2005. Musically tone-deaf individuals have difficulty discriminating intonation contours extracted from speech. *Brain Cogn.* 59, 310–313.
- Peretz, I., 2013. The biological foundations of music: insights from congenital amusia. In: Deustch, D. (Ed.), *The Psychology of Music*. Academic Press, San Diego, pp. 551–564.
- Peretz, I., Hyde, K.L., 2003. What is specific to music processing? Insights from congenital amusia. *Trends Cogn. Sci.* 7, 362–367.
- Peretz, I., Ayotte, J., Zatorre, R.J., Mehler, J., Ahad, P., Penhune, V.B., Jutras, B., 2002. Congenital amusia: a disorder of fine-grained pitch discrimination. *Neuron* 33, 185–191.
- Peretz, I., Champod, A.S., Hyde, K., 2003. Varieties of musical disorders: the Montreal Battery of Evaluation of Amusia. *Ann. N. Y. Acad. Sci.* 999, 58–75.
- Peretz, I., Brattico, E., Tervaniemi, M., 2005. Abnormal electrical brain responses to pitch in congenital amusia. *Ann. Neurol.* 58, 478–482.
- Peretz, I., Brattico, E., Jarvenpää, M., Tervaniemi, M., 2009. The amusic brain: in tune, out of key, and unaware. *Brain* 132, 1277–1286.
- Pfeifer, J., Hamann, S., 2015. Revising the diagnosis of congenital amusia with the Montreal Battery of Evaluation of Amusia. *Front. Hum. Neurosci.* 9, 161.
- Rauschecker, J.P., 1995. Compensatory plasticity and sensory substitution in the cerebral cortex. *Trends Neurosci.* 18, 36–43.
- Shams, L., Kamitani, Y., Shimojo, S., 2000. Illusions: what you see is what you hear. *Nature* 408, 788.
- Stekelenburg, J.J., Vroomen, J., 2007. Neural correlates of multisensory integration of ecologically valid audiovisual events. *J. Cogn. Neurosci.* 19, 1964–1973.
- Talsma, D., Doty, T.J., Woldorff, M.G., 2007. Selective attention and audiovisual integration: is attending to both modalities a prerequisite for early integration? *Cereb. Cortex* 17, 679–690.
- Teder-Sälejärvi, W.A., McDonald, J.J., Di Russo, F., Hillyard, S.A., 2002. An analysis of audiovisual crossmodal integration by means of event-related potential (ERP) recordings. *Cogn. Brain Res.* 14, 106–114.
- Teder-Sälejärvi, W.A., Di Russo, F., McDonald, J.J., Hillyard, S.A., 2005. Effects of spatial congruity on audio-visual multimodal integration. *J. Cogn. Neurosci.* 17, 1396–1409.
- Thompson, W.F., Marin, M.M., Stewart, L., 2012. Reduced sensitivity to emotional prosody in congenital amusia rekindles the musical protolanguage hypothesis. *Proc. Natl. Acad. Sci. U. S. A.* 109, 19027–19032.
- Tillmann, B., Schulze, K., Foxton, J., 2009. Congenital amusia: a short-term memory deficit for non-verbal, but not verbal sounds. *Brain Cogn.* 71, 259–264.
- Vogel, E.K., Luck, S.J., 2000. The visual N1 component as an index of a discrimination process. *Psychophysiology* 37, 190–203.
- Widmann, A., Schröger, E., 2012. Filter effects and filter artifacts in the analysis of electrophysiological data. *Front. Psychol.* 3, 233.
- Williamson, V.J., Stewart, L., 2010. Memory for pitch in congenital amusia: beyond a fine-grained pitch discrimination problem. *Memory* 18, 657–669.
- Yeung, N., Botvinick, M.M., Cohen, J.D., 2004. The neural basis of error detection: conflict monitoring and the error-related negativity. *Psychol. Rev.* 111, 931–959.
- Zatorre, R.J., 2001. Do you see what I'm saying? Interactions between auditory and visual cortices in cochlear implant users. *Neuron* 31, 13–14.