Exp Brain Res (2018) 236:13–29
https://doi.org/10.1007/s00221-017-5104-3

RESEARCH ARTICLE

Touching lips and hearing fingers: effector-specific congruency between tactile and auditory stimulation modulates N1 amplitude and alpha desynchronization

Guannan Shen[1] · **Andrew N. Meltzoff**[1,2] · **Peter J. Marshall**[1]

Received: 30 May 2017 / Accepted: 5 October 2017 / Published online: 16 October 2017

Abstract Understanding the interactions between audition and sensorimotor processes is of theoretical importance, particularly in relation to speech processing. Although one current focus in this area is on interactions between auditory perception and the motor system, there has been less research on connections between the auditory and somatosensory modalities. The current study takes a novel approach to this omission by examining specific auditory–tactile interactions in the context of speech and non-speech sound production. Electroencephalography was used to examine brain responses when participants were presented with speech syllables (a bilabial sound /pa/ and a non-labial sound /ka/) or finger-snapping sounds that were simultaneously paired with tactile stimulation of either the lower lip or the right middle finger. Analyses focused on the sensory-evoked N1 in the event-related potential and the extent of alpha band desynchronization elicited by the stimuli. N1 amplitude over fronto-central sites was significantly enhanced when the bilabial /pa/ sound was paired with tactile lip stimulation and when the finger-snapping sound was paired with tactile stimulation of the finger. Post-stimulus alpha desynchronization at central sites was also enhanced when the /pa/ sound was accompanied by tactile stimulation of the lip. These novel findings indicate that neural aspects of somatosensory–auditory interactions are influenced by the congruency between the location of the bodily touch and the bodily origin of a perceived sound.

Keywords Touch · Speech · Audition · Multisensory integration · ERP · Alpha oscillations

Introduction

The integration of sensory information from different modalities is crucial for our everyday functioning. One focus of work on multisensory integration has been on the perception of speech, including the extent to which non-auditory information can influence speech perception. Research in this area has demonstrated that concurrent visual or somatosensory stimulation relevant to speech articulation exerts modulatory effects on auditory speech perception (Visual: D'Ausilio et al. 2014; Badin et al. 2010; Kuhl and Meltzoff 1982; McGurk and MacDonald 1976; Somatosensory: Bruderer et al. 2015; Ito et al. 2014; Gick and Derrick 2009). These and other related findings have highlighted linkages between speech perception and production, although the nature and function of these connections are debated (Schomers and Pulvermüller 2016; Hickok 2014; Hickok et al. 2011; Pulvermüller and Fadiga 2010).

One aspect of this debate concerns the interpretation of evidence showing activation of cortical sensorimotor systems during speech perception. Functional neuroimaging studies have reported that listening to speech activates sensorimotor areas involved in speech production (e.g., Correia et al. 2015; Pulvermüller et al. 2006; Wilson et al. 2004), although the specifics of this association are under discussion (Arsenault and Buchsbaum 2015; Behroozmand et al. 2015). Other evidence suggests that sensorimotor areas are more involved when listening to non-native speech sounds that are harder to decode, compared with listening to native languages (Kuhl et al. 2014; Callan et al. 2004).

✉ Guannan Shen
guannan.shen@temple.edu

[1] Department of Psychology, 1701 N 13th Street, Philadelphia, PA 19122, USA

[2] Institute for Learning and Brian Sciences, University of Washington, Seattle, USA

Much of the literature on the role of sensorimotor influences in speech perception has focused on the motor system. However, also relevant is a line of research demonstrating a different multimodal interaction—the modulation of somatosensory perception when perceiving speech. Illustrative behavioral studies reported that listening to voiced speech sounds modulates self-reported perceptual intensity of tactile stimulation in the lip and laryngeal regions (Champoux et al. 2011), and alters facial skin sensation (Ito and Ostry 2012), suggesting a linkage between auditory and somatosensory perception. There is a similar report linking visual perception of speech and somatosensation: viewing someone producing speech without sound (lip reading) increases the perceived intensity of tactile stimulation to one's own lip (Thomas et al. 2013). This perceptual enhancement is stronger when viewing the production of bilabial words than for control words, suggesting a degree of effector-based specificity. A related study showed that lip reading enhances the event-related potential (ERP) elicited by tactile lip stimulation, but does not affect the ERP response to median nerve stimulation (Möttönen et al. 2005).

The finding that viewing specific articulatory acts enhances cortical responses to tactile lip stimulation connects with a wider line of research concerning the processes involved when observing the actions of other people. Various studies have suggested that observing actions of others that are carried out with specific bodily effectors influences neural responses to somatosensory stimulation of the corresponding effector in the observer. For instance, activation of primary somatosensory cortex in response to hand stimulation is enhanced during the viewing of hand actions (Hasson et al. 2004; Avikainen et al. 2002; Rossi et al. 2002). Viewing actions carried out with a specific finger increases the amplitude of the ERP response to somatosensory stimulation to the corresponding finger, compared with the response elicited by stimulation of a different finger (Deschrijver et al. 2016). A number of functional neuroimaging studies have also found that action observation alone activates sensorimotor areas associated with the corresponding effectors, including feet as well as hands (Gazzola et al. 2006; Wheaton et al. 2004; Buccino et al. 2001). Related developmental evidence comes from work with infants. An electroencephalographic (EEG) study with 14-month-old infants showed greater desynchronization of the alpha-range mu rhythm over sensorimotor hand areas during observation of hand actions than during the observation of foot actions. Similarly, the observation of foot actions was associated with greater mu rhythm desynchronization at electrodes over the foot area than over the hand area (Saby et al. 2013).

There is also evidence that listening to sounds produced by human actions can induce particular sensorimotor neural activation linked to the effectors that normally produce these sounds. For example, neural activation in sensorimotor cortices is modulated by listening to a clapping sound or sound of someone knocking on a door (Pizzamiglio et al. 2005). A related effect was observed when novice piano players were listening to practiced notes (Lahav et al. 2007). One developmental study also found that infants exhibited stronger mu rhythm desynchronization over sensorimotor cortex when listening to the sound of shaking a rattle compared to non-action-related control sounds (Paulus et al. 2012).

The foregoing behavioral and neurophysiological evidence suggests that both listening to speech sounds and watching the corresponding articulations in the absence of sounds can modulate the observer's own somatosensory perception and, moreover, that activation of cortical sensorimotor networks occurs both during the observation of human actions and in response to hearing action-related sounds. The enhanced neurophysiological activity elicited by concurrent presentation of stimuli from different sensory modalities has been interpreted as an indicator of multisensory integration (Henschke et al. 2015; Stanford and Stein 2007; Foxe et al. 2000). However, such enhancement does not occur under all circumstances, and is influenced (increased) by temporal synchrony (Gick, Ikegami and Derrick 2010; Van Atteveldt et al. 2007; Senkowski et al. 2007; Calvert et al. 2000), spatial alignment (Macaluso and Driver 2005), and auditory–tactile frequency congruency (Wilson et al. 2010; Ro et al. 2009), as well as higher level cognitive factors, such as cross-modal semantic congruency (Doehrmann and Naumer 2008). One explanation for the observed facilitating effect of cross-modality congruency relates to the *unity assumption*, which refers to the degree to which sensory stimuli are inferred as arising from a single unitary object or event (Talsma et al. 2010; Welch 1999). The binding of congruent versus incongruent sensory stimuli and the enhanced processing of congruent multisensory stimuli reduces ambiguity and maximizes a coherent and unified perception of the external world (Tsilionis and Vatakis 2016).

One intriguing aspect of prior work that deserves further exploration concerns the effector-specific nature of perceptual enhancements or sensorimotor activation in response to seeing or hearing speech or action-related stimuli. In the current study, we aimed to extend prior findings on this body part specificity by recording brain responses to tactile stimulation of the lips or fingers that occurred simultaneously with either a bilabial speech sound or a finger-snapping sound. Our primary question concerned the extent to which any multisensory enhancement of neural responses depends on a congruence between the sound and the effector that naturally produces that sound. In the current study, two neurophysiological indices of cortical activation to the paired stimuli were extracted from the continuous EEG data: the sensory-evoked N1 in the ERP, and post-stimulus event-related desynchronization (ERD) in the alpha frequency range (8–14 Hz).

The N1 is a negative-going potential that peaks around 100 ms after the onset of sensory stimulation and is associated with early feature processing in primary and secondary sensory cortices (Molholm et al. 2004; Näätänen and Picton 1987). In terms of the ERD response, the suppression of alpha oscillations reflects aspects of cortical excitation (Foxe et al. 1998), with distinct patterns of desynchronization being observable over different scalp regions depending on the modality of stimulation (Sadaghiani and Kleinschmidt 2016). Following somatosensory stimulation, alpha ERD is often observed at central regions overlying somatosensory cortex in both EEG and MEG recordings (Lange et al. 2011; Schubert et al. 2008; Bauer et al. 2006; Cheyne et al. 2003; Della Penna et al. 2004). Similarly, several MEG studies have found that alpha ERD is induced in auditory cortex around 200–750 ms following auditory stimulation onset (Müller and Weisz 2012; Tiihonen et al. 1991; Fujioka and Ross 2008).

The extent of post-stimulus alpha desynchronization also appears to be related to multisensory integration. For example, alpha oscillations over auditory cortex were more suppressed for congruent audio–visual speech stimuli than for incongruent bimodal stimuli (Lange et al. 2013). A recent study on auditory–somatosensory interactions showed that post-stimulus alpha desynchronization was significantly stronger when the bimodal stimuli were spatially integrated than when they were perceived as coming from different directions (Plöchl et al. 2016). In the current study, both N1 amplitude and the extent of the alpha ERD response were used to index aspects of neural activation in response to the simultaneous presentation of auditory and tactile stimuli.

Although there is behavioral evidence that listening to speech modulates somatosensory perception (Thomas et al. 2013; Champoux et al. 2011), related evidence from neurophysiological studies remains limited. A previous study showed an enhancement of ERP amplitude to tactile stimulation of the lip during the visual observation of articulatory movements, but did not find a similar enhancement when participants were simply hearing speech (Möttönen et al. 2005). The speech stimuli in that particular study were generated by an experimenter who read a book aloud while participants received tactile stimulation. The reported absence of an auditory–tactile interaction in that particular study could be due to the lack of temporal alignment between auditory and tactile stimulation. In the current study, to maximize the potential effect of auditory–tactile sensory interactions, we employed a protocol that included simultaneous presentation of congruent auditory and tactile stimuli, specifically a bilabial speech sound (/pa/) paired with tactile stimulation of the lower lip, and a finger-snapping sound paired with tactile stimulation delivered to the tip of the right middle finger. We further examined responses to tactile stimulation of the lip and finger paired with a comparison speech sound that does not involve lips as the place of articulation (/ka/).

Continuous EEG was recorded while participants were presented with each of the various possible combinations of audio–tactile stimulation. If auditory–somatosensory integration is modulated by effector congruency, we expected to find differential responses to congruent and incongruent stimulus pairs in N1 amplitude and in the extent of alpha band desynchronization. We hypothesized that neural processing of sensory input would be enhanced—as indicated by greater N1 amplitude and increased alpha desynchronization—when auditory stimuli are simultaneously accompanied by somatosensory stimulation of the bodily effectors associated with production of the sounds.

Methods

Participants

A total of 23 participants (16 females, mean age = 22.09 years, SD = 3.91) were recruited for the study. All were right-handed and had no history of neurological abnormality. Informed consent was obtained from all individual participants included in the study. Data from 21 participants were used in the final EEG analyses. One participant's dataset was excluded because of hardware problems, and another was excluded because the participant was too fatigued to complete the task. For the behavioral data, because of a temporary foot pedal malfunction, the responses of two participants were not registered. A total of 19 participants were, therefore, included in the analyses of reaction time. All participants gave their informed consent to participate in this study, which was approved by the Temple University Institutional Review Board.

Tactile stimulation

Tactile stimuli were delivered to the distal tip of the right middle finger and the right side of the lower lip using an inflatable membrane (10 mm diameter) mounted in a plastic casing. The membrane was inflated by a short burst of compressed air delivered via flexible polyurethane tubing (3 m length, 3.2 mm outer diameter). The compressed air delivery was controlled by STIM stimulus presentation software in combination with a pneumatic stimulator unit (both from James Long Company) and an adjustable regulator that restricted the airflow to 60 psi. To generate each tactile stimulus, the STIM software delivered a trigger (10 ms duration) that served to open and close a solenoid in the pneumatic stimulator. Expansion of the membrane started

15 ms after trigger onset and peaked 40 ms later. As shown in Fig. 1, the total duration of membrane movement was around 100 ms, with a peak force of 2.2 N as measured using a custom calibration unit (James Long Company). This stimulation method has been successfully used in a number of prior EEG and MEG studies (Pihko et al. 2009; Saby et al. 2013; Shen et al. 2017).

Auditory stimulation

The speech sounds /pa/, /ka/, and a finger-snapping sound were recorded and were then scaled to a total duration of 200 ms with 40 ms voice onset time (VOT) for speech sounds. The peak intensity of all three sounds was adjusted to 75 dB (Fig. 1). The speech sounds /pa/ and /ka/ were pronounced by an adult female speaker, while the finger-snapping sounds were produced by multiple volunteers. All manipulation of the speech sound recordings was done in Praat (Boersma and Weenink 2009) and Audacity (www.audacityteam.org). The auditory stimuli were presented binaurally via Etymotic ER-3A insert earphones (www.etymotic.com).

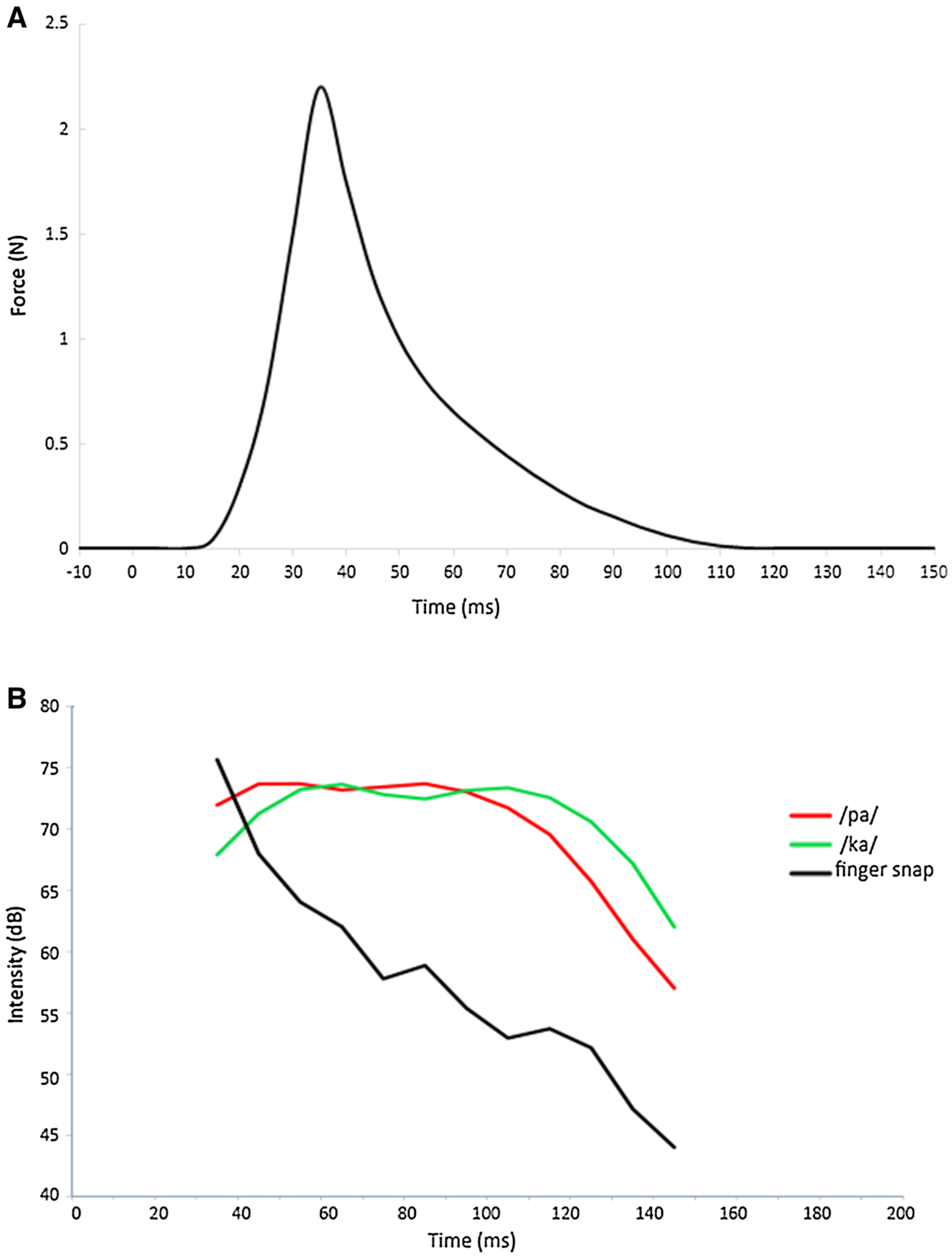

Fig. 1 **a** Response curve showing the force exerted over time by the tactile stimulator. 0 ms corresponds to trigger onset, and 0 ms in ERP epochs. **b** Intensity contours of the auditory stimuli /pa/ (red), /ka/ (green), and the finger-snap sound (black). 0 ms corresponds to auditory stimuli onset and time zero (0 ms) in the ERP computation

Procedure

Participants were fitted with an EEG cap and were seated comfortably in front of a computer screen. After tactile stimulators were attached to their right middle finger and the right side of their lower lip, participants put the insert earphones into both ears. Participants were told that they should focus on the fixation cross on the screen and respond to each auditory or tactile stimulus by pressing a foot pedal with their right foot. Reaction time was recorded for each trial.

The experiment was divided into two blocks. The first block consisted of auditory unimodal stimuli (80 trials of each of the 3 sounds); the second block consisted of bimodal simultaneous auditory–tactile stimuli (80 trials of each of the 6 audio–tactile combinations). Within each block, the stimuli were presented in a random order with a jittered stimulus onset asynchrony of between 1500 and 1900 ms (in 100 ms intervals).

EEG recording

EEG signals were recorded from 32 sites (Fp1, FPz, Fp2, F3, F4, Fz, F7, F8, C3, C4, Cz, T7, T8, P3, P4, Pz, P7, P8, POz, O1, Oz, O2, M1, M2, Fc1, Fc2, Fc5, Fc6, Cp1, Cp2, Cp5, Cp6) using a cap (ANT Neuro, Germany) with electrodes placed according to the international 10–20 system. Vertical electrooculogram (EOG) activity was collected from electrodes placed above and below the left eye. Scalp electrode impedances were kept under 25 kΩ although most cases impedance values were below 15 kΩ. All EEG and EOG channels were continuously recorded with a sampling frequency of 512 Hz. Hardware bandpass filter settings were 0.1 Hz (high pass) and 100 Hz (low pass) and the gain was 4000 for EEG channels and 1000 for the EOG channel. The EEG signals were collected referenced to Cz with an AFz ground, and were re-referenced offline to the average of the left and right mastoids.

Data analysis

Pre-processing of EEG data

Processing and initial analysis of the EEG signals were performed using the EEGLAB 13.5.4b toolbox (Delorme and Makeig 2004) implemented in MATLAB. Epochs of 1500 ms duration were extracted from the continuous EEG data, with each epoch extending from − 500 to 1000 ms relative to stimulus onset. Independent component analysis (ICA) was used to identify and remove eye movement artifacts (Hoffmann and Falkenstein 2008). Visual inspection of the EEG signal was used to reject epochs containing other movement artifacts. The mean number of artifact-free trials per condition was 73 (SD = 6.8). There was no significant difference in the number of usable trials across all unimodal and bimodal conditions ($p = 0.826$).

ERP analysis

To prepare the data for ERP analysis, artifact-free epochs were low-pass filtered at 30 Hz before being averaged and baseline corrected relative to a 100 ms pre-stimulus baseline. Amplitude of the N1 was calculated with an adaptive mean technique, which allows for individual differences in peak latency (see, e.g., Clayson et al. 2013). For each participant, the most negative peak between 90 and 160 ms following stimulus onset was identified. N1 amplitude for that participant was computed as the mean amplitude in a window extending ± 10 ms around the identified negative peak.

Previous ERP studies of auditory–somatosensory interaction have typically focused on central (e.g., Foxe et al. 2000; Brett-Green et al. 2008; Murray et al. 2005) and fronto-central (Russo et al. 2010) electrode sites. Working from these prior studies, and from prior work on the somatosensory N1 showing a distribution of this component over frontal and central sites (Töllner et al. 2009; Nishimura et al. 1986), the region of interest (ROI) for the ERP analyses in the current study comprised frontal (F3, F4), frontal–central (FC1, FC2) and central (C3, C4) recording sites.

Repeated-measures ANOVAs on mean N1 amplitude were conducted with the factors Body Part (lip, finger), Sound (/pa/, /ka/, finger-snap), Region (frontal, fronto-central, central), and Hemisphere (left, right). Greenhouse–Geisser adjustments to the degrees of freedom were performed. Post hoc analyses were conducted using pairwise *t* tests with FDR correction.

Analyses were also carried out for N1 amplitude in the unimodal auditory conditions, with a two-way repeated-measures ANOVA being conducted with the factors Electrode (C3/C4/Cz) and Sound (/pa/, /ka/, finger-snap). Since the auditory N1 is known to occur at central sites around the vertex (Virtanen et al. 1998), the analysis of responses to unimodal stimuli focused on the electrodes Cz, C3, and C4. The aim of these analyses of the unimodal stimuli was to ensure that any effects observed in the multimodal paradigm were not due to the physical differences between the sounds.

In analyzing the ERP data we did not compute or compare N1 latencies across different conditions. Previous studies have found that speech and non-speech sounds can evoke N1 responses with different latencies (differing by ~ 10 ms; Tiitinen et al. 1999; Eulitz et al. 1995). Since our focus was on congruency effects in multisensory processing, and not the auditory processing of speech/non-speech sounds, we did not include an analysis of latency differences across conditions.

Time–frequency analyses

Event-related changes in alpha oscillations were compared across the various stimulus conditions. Time–frequency decompositions of the EEG data were conducted using event-related spectral perturbation (ERSP) analysis (Makeig 1993) for a 1500 ms window that extended from − 500 to 1000 ms relative to stimulus onset. ERSP was computed using a Morlet wavelet decomposition over a frequency range of 5–25 Hz, with 100 overlapping windows starting with a 3-cycle wavelet at the lowest frequency with a Hanning-tapered window applied. The baseline was defined as the 500-ms period immediately before stimulus onset. Event-related desynchronization (ERD) is indicated by negative ERSP values, which signify a decrease in power relative to the baseline. To assess whether differential alpha band desynchronization was induced by the bimodal stimuli, mean ERSP in the alpha band (8–14 Hz) for a window of 250–600 ms post-stimulus onset was computed for each participant and condition. Because alpha desynchronization has been reported around primary sensory cortices following sensory input (Schubert et al. 2008; Müller and Weisz 2012; Fujioka and Ross 2008; Lange et al. 2011), and spatial effects on auditory–tactile integration have been observed in auditory cortex (Plöchl et al. 2016), the ROI for alpha band analysis comprised central (C3, C4) and temporal (T7, T8) electrodes that overlying somatosensory and auditory cortices, respectively. A four-way repeated measures ANOVAs on mean alpha ERSP were carried out involving the factors Body Part (lip, finger), Region (central vs. temporal), Hemisphere (left vs. right) and Sound (/pa/, /ka/, finger-snap sound). The Greenhouse–Geisser correction factor was employed, and post hoc analyses were conducted using pairwise *t* tests with FDR correction. Additional analyses involved a comparison of ERSP responses to the three unisensory auditory stimuli.

Results

Behavioral results

Participants were instructed to respond to each auditory or tactile stimulus by pressing a foot pedal with their right foot, which constituted the behavioral responses. These responses were required chiefly to ensure that participants remained alert throughout the experiment. Although the reaction time responses were analyzed, they remain secondary to the primary measures of the study.

A significant difference in reaction time was found between the unimodal and bimodal conditions ($F(1,18) = 10.637$, $p = 0.004$, $\eta^2 = 0.051$). Participants responded significantly faster to bimodal stimuli ($M = 403$ ms) than to unimodal stimuli ($M = 449$ ms). For unimodal stimuli, there was no significant difference in reaction time to the three sounds $(2, 36) = 2.191$, $p = 0.152$). For the six combinations of bimodal audio–tactile stimuli, a one-way ANOVA revealed a significant main effect of condition ($F(5, 90) = 4.056$, $p = 0.005$, $\eta^2 = 0.004$). Post hoc pairwise comparisons with FDR correction showed that the coupling of tactile lip stimulation with the finger-snapping sound yielded significantly longer reaction times than the pairing of tactile lip stimuli with the speech sounds /pa/ ($p = 0.038$) and /ka/ ($p = 0.038$). No other comparisons were statistically significant (Fig. 2).

N1 to bimodal stimuli

The N1 elicited by the simultaneous auditory and tactile stimuli had a primarily central and frontal–central scalp distribution (see Fig. 3). As noted above, analysis of the ERP waveforms focused on left and right central electrodes (C3, C4), left and right fronto-central electrodes (FC1, FC2), and left and right frontal electrodes (F3, F4). The ERP waveforms at these electrodes are shown in Figs. 4 and 5. A four-way repeated-measure ANOVA was then conducted on N1 amplitude with the following factors: Body Part (lip, finger), Sound (/pa/, /ka/, finger-snap), Region (frontal, fronto-central, central), and Hemisphere (left, right). The analysis revealed a significant main effect of Hemisphere ($F(1, 20) = 10.153$, $p = 0.005$, $\eta^2 = 0.005$), with greater amplitude in the left hemisphere, contralateral to the tactile stimulation. There was also a significant interaction between Body Part and Sound ($F(2, 40) = 10.356$, $p < 0.001$, $\eta^2 = 0.025$). To explore this interaction, two three-way ANOVAs were then conducted separately for lip and finger stimulation.

For tactile stimulation of the lip, the ANOVA for N1 amplitude showed a significant main effect of Sound ($F(2,40) = 5.236$, $p = 0.016$, $\eta^2 = 0.019$). Post hoc pairwise *t* tests with FDR correction revealed that simultaneous pairing of the tactile lip stimulation and /pa/ elicited a significantly larger N1 than did the simultaneous pairing of the lip stimulation and /ka/ ($p < 0.001$), or the pairing of the lip stimulation and the finger-snap sound ($p < 0.001$). There was no significant difference between the pairing of lip stimulation with /ka/ versus pairing of lip stimulation with the finger-snap sound ($p = 0.368$). There were no significant main effects or interactions involving Hemisphere or Region.

For tactile stimulation of the finger, there was a significant main effect of Sound ($F(2, 40) = 6.757$, $p = 0.008$, $\eta^2 = 0.053$). Post hoc pairwise *t* tests with FDR correction showed that N1 amplitude was significantly larger for finger stimulation paired with the finger-snap sound than for finger stimulation paired with either of the speech sounds (both comparisons $p < 0.001$). There was no significant

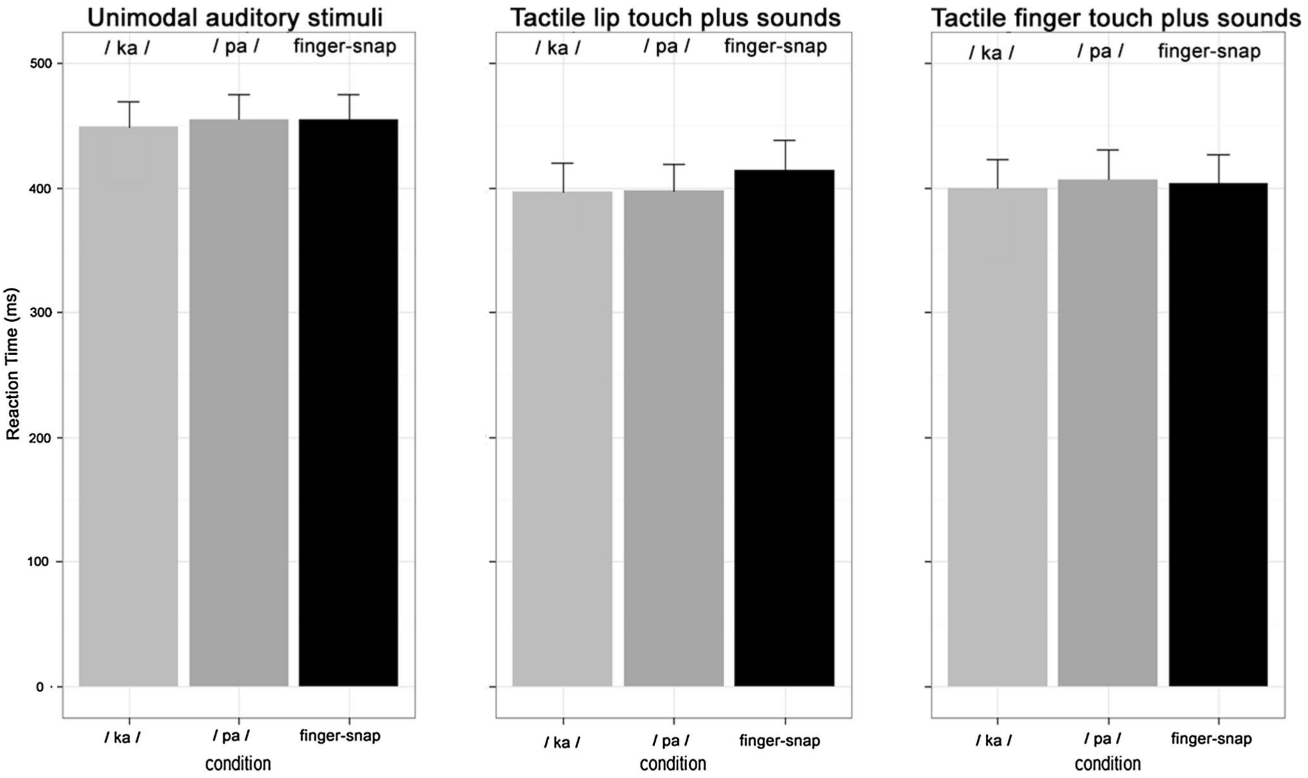

Fig. 2 Mean reaction time across the experimental conditions. Error bars indicate one standard error

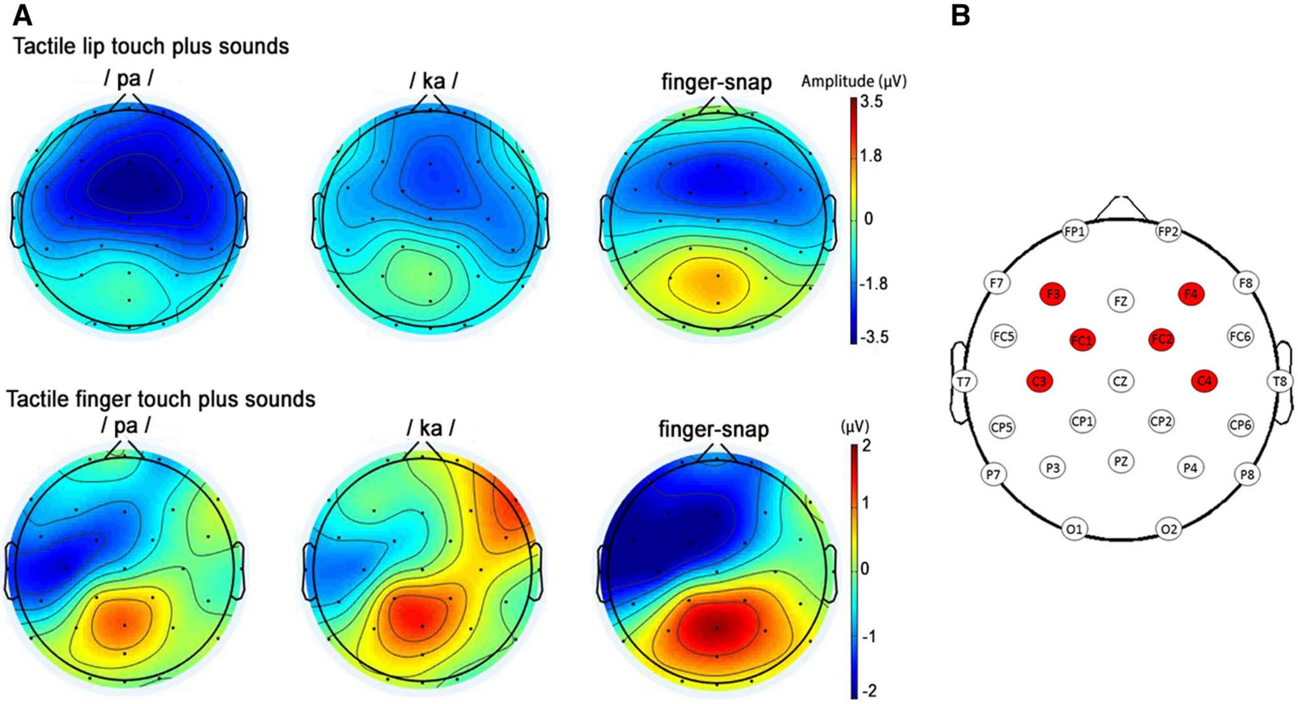

Fig. 3 **a** Mean N1 amplitude (90–150 ms) for tactile stimulation of the lip (upper panel) and of the right finger (lower panel) paired with /ka/ (left), /pa/ (middle) and finger-snap sounds (right). **b** Electrode montage with analyzed sites highlighted in red

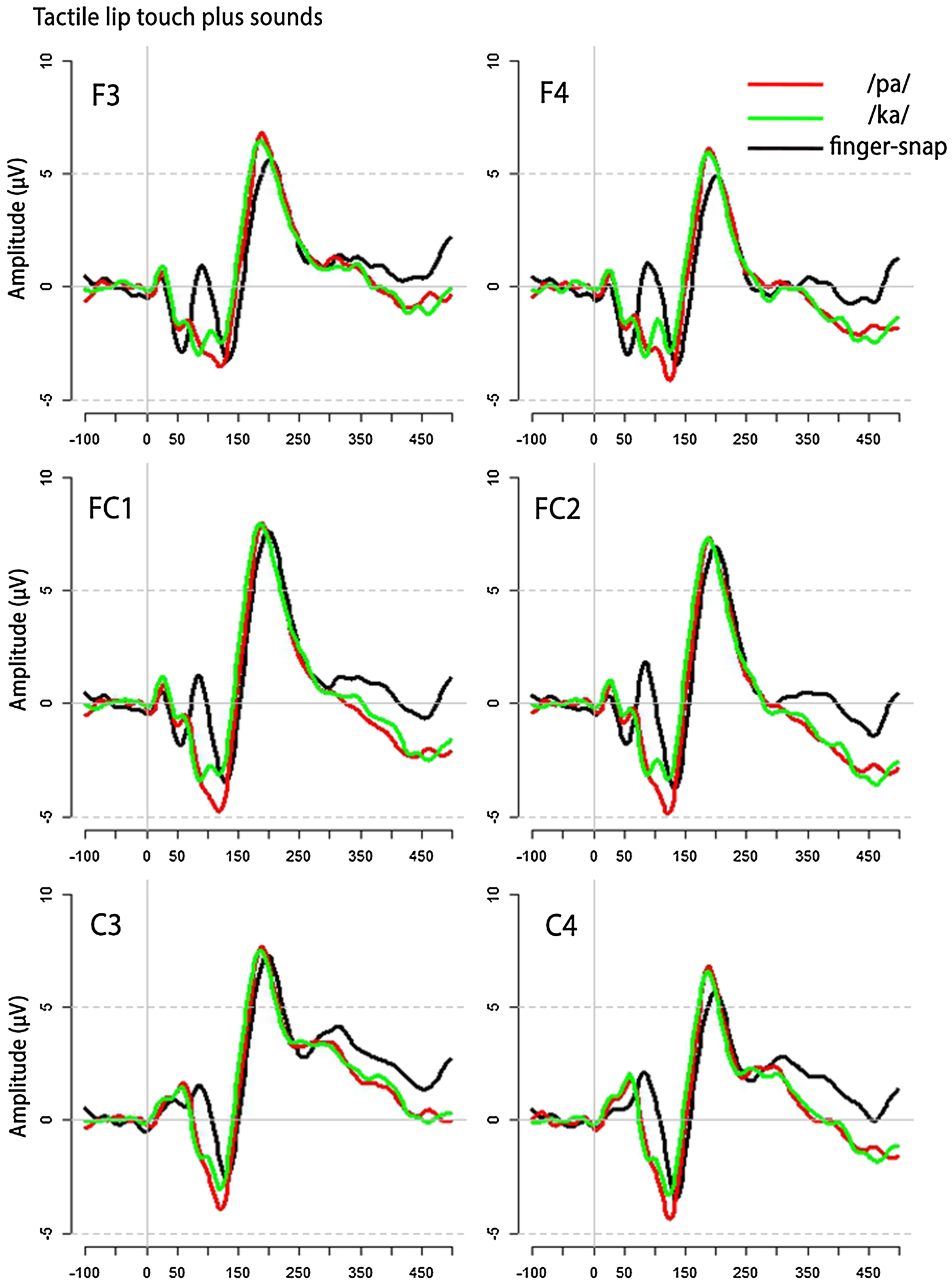

Fig. 4 Grand average ERP waveforms for tactile stimulation of the lip paired with /ka/ (green), /pa/ (red), and the finger-snap sound (black)

difference in N1 amplitude between the two speech sounds paired with finger stimulation ($p = 0.82$). The analysis also revealed a significant main effect of Hemisphere ($F(1, 20) = 28.816$, $p < 0.001$, $\eta^2 = 0.031$), with significantly larger N1 amplitude in the left hemisphere (contralateral to the side of touch) than in the right hemisphere,

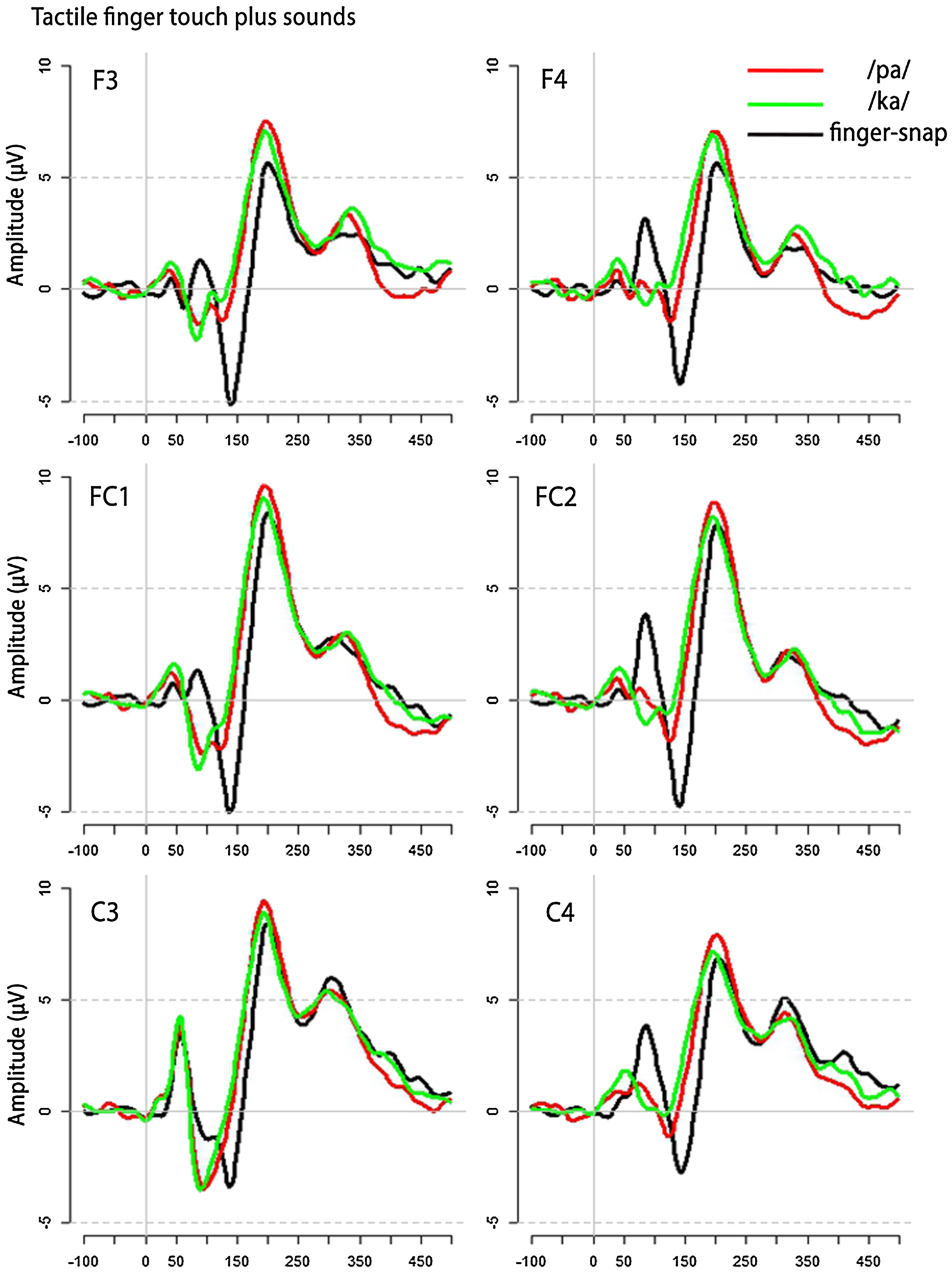

Fig. 5 Grand average ERP waveforms for tactile stimulation of the finger paired with /ka/ (green), /pa/ (red) and the finger-snap sound (black)

and a significant interaction between Region and Hemisphere ($F(2, 40) = 5.054$, $p = 0.028$, $\eta^2 = 0.005$). Post hoc pairwise t test with FDR correction showed that N1 was significantly larger at the fronto-central electrode than frontal and central electrodes on the right hemisphere ($p < 0.001$ for both comparisons), but the difference between regions was not significant in the left hemisphere.

N1 to unimodal auditory stimuli

The ERP waveforms and topographic plots of N1 responses elicited by the unimodal auditory stimuli are shown in Fig. 6. A two-way repeated-measures ANOVA on N1 amplitude revealed a main effect of Electrode ($F(2, 40) = 4.456$, $p = 0.022$, $\eta^2 = 0.019$). Post hoc pairwise t tests with FDR correction showed that the N1 response was significantly larger at Cz than at C3 ($p < 0.001$) or C4 ($p < 0.001$). There was no significant main effect of Sound, and no significant interaction between sound and electrode.

Event-related alpha band responses to bimodal stimulation

Figure 7 shows the scalp distribution of post-stimulus alpha ERSP for each of the six audio–tactile conditions. Based on previous literature on sensory-evoked alpha ERD (Müller and Weisz 2012; Fujioka and Ross 2008; Lange et al. 2011), statistical analyses focused on the central electrodes C3 and C4, and the temporal electrodes T7 and T8. Plots of the time–frequency decomposition at C3 and T7 (contralateral to the sites of tactile stimulation) are shown in Fig. 8. Inspection of these plots showed that the post-stimulus alpha desynchronization started around 250 ms. This timeframe for the response is consistent with the time windows used in previous studies of alpha ERD to somatosensory stimulation (Dockstader et al. 2008; Fujioka and Ross 2008; Schubert et al. 2008). For lip stimulation, we found significant alpha ERSP differences across three different sounds starting around 200 ms ($p < 0.05$, with FDR correction), whereas no difference across sounds was found for finger stimulation (Fig. 7).

Mean ERSP in the 8–14 Hz frequency band for the 250–600 ms window was compared across the conditions using a four-way repeated measures ANOVA with the following factors: Body Part (lip, finger), Sound (/pa/, /ka/, finger-snap), Hemisphere (left, right) and Region (central, temporal). Results showed a significant main effect of Sound ($F(2, 40) = 4.089$, $p = 0.038$, $\eta^2 = 0.25$). Pairwise t tests with FDR correction revealed that the /pa/ and /ka/ sounds elicited stronger alpha ERD than the finger-snap sound (/pa/ > finger-snap, $p < 0.001$; /ka/ > finger-snap, $p < 0.001$; /pa/ > /ka/, $p = 0.38$). There were also a significant main effect of Hemisphere ($F(1,20) = 15.623$, $p < 0.001$, $\eta^2 = 0.034$), with greater alpha ERD over the left hemisphere than the right hemisphere, and a main effect of Region ($F(1, 20) = 7.898$, $p = 0.011$, $\eta^2 = 0.023$), with greater alpha ERD in central regions than temporal regions. The four-way ANOVA also revealed a significant interaction between Body Part and Sound ($F(2, 40) = 7.481$, $p = 0.002$, $\eta^2 = 0.032$), and a significant interaction between Body Part and Region (F(2,

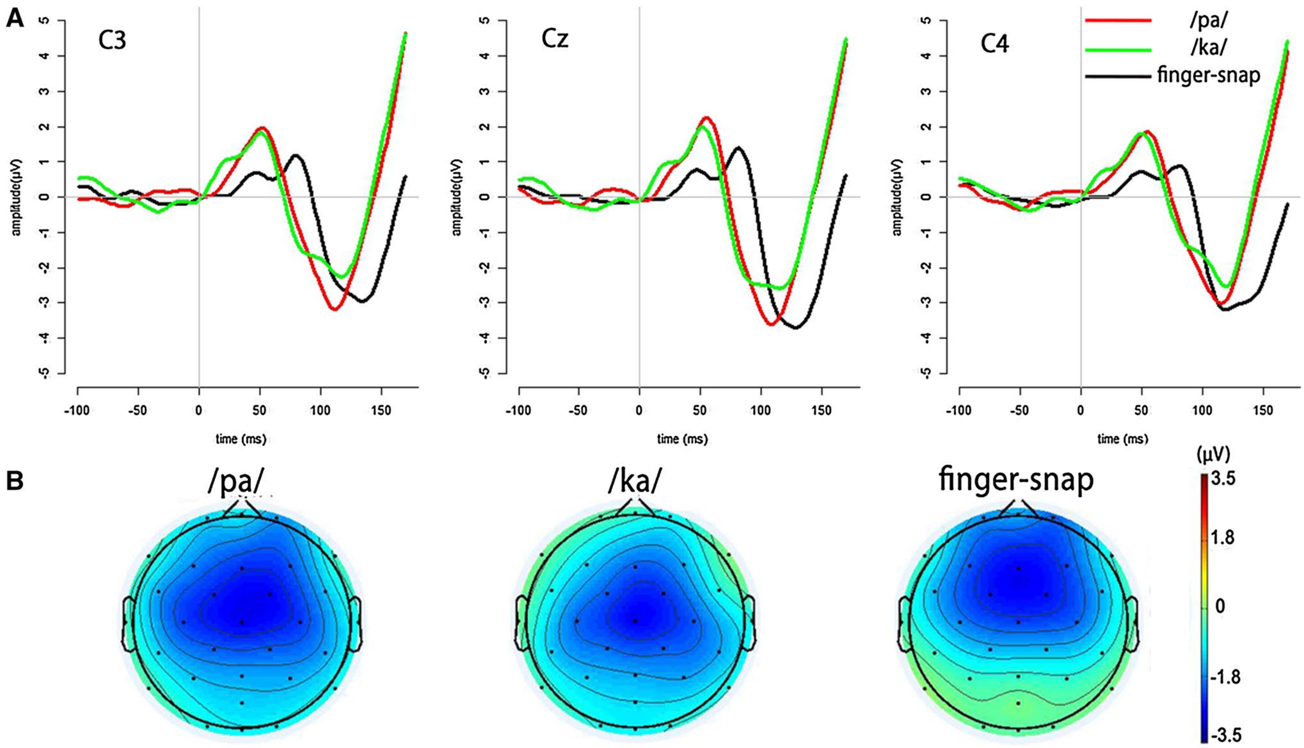

Fig. 6 **a** Grand average ERPs elicited to /ka/ (green), /pa/ (red) and finger-snap (black) sounds at C3 (left), Cz (center) and C4 (left) electrodes. **b** Mean auditory N1 amplitude (90–150 ms) for /pa/ (left), /ka/ (center), and the finger-snap sound (right)

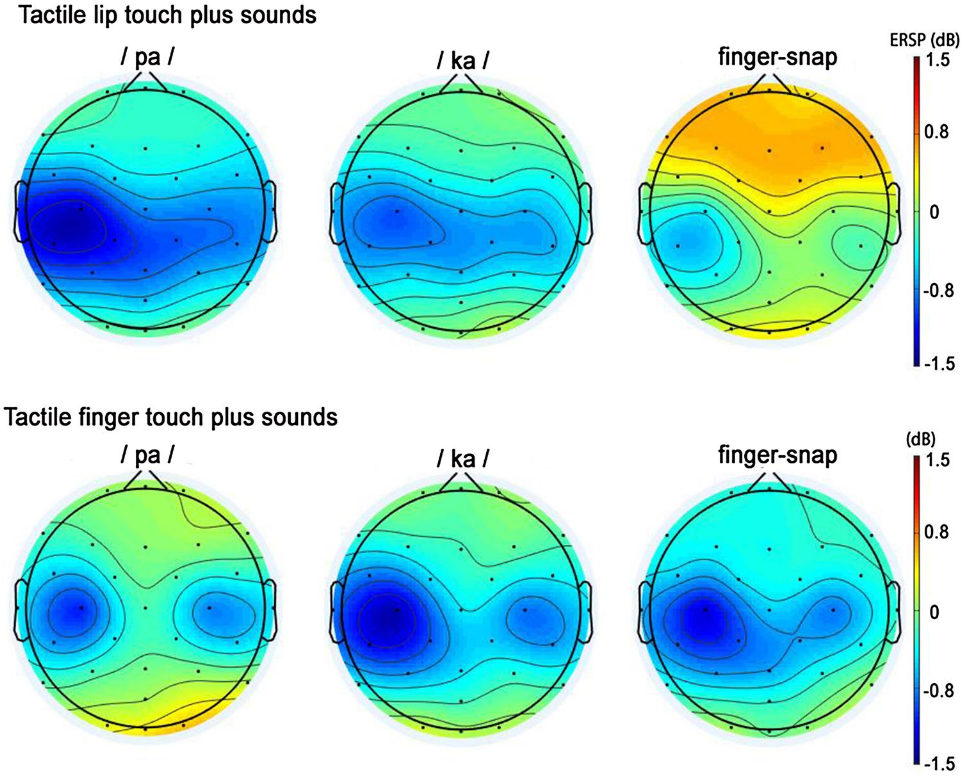

Fig. 7 Mean ERSP in the alpha band (8–14 Hz) for simultaneous audio–tactile stimuli. Mean values are computed for a window of 250–600 ms post-stimulus

40) = 6.674, $p = 0.018$, $\eta^2 = 0.004$). To follow up on this significant interaction, two separate three-way ANOVAs were conducted for each tactile stimulation site (lip, finger).

For tactile stimulation of the lip, there were significant main effects of Sound ($F(2, 40) = 10.214$, $p = 0.001$, $\eta^2 = 0.106$) and Hemisphere ($F(1, 20) = 9.495$, $p = 0.006$, $\eta^2 = 0.047$), with stronger alpha ERD in the left hemisphere, contralateral to the site of tactile stimulation. There was no significant interaction between Sound and Electrode. Pairwise *t* tests with FDR correction revealed that alpha ERD following presentation of the lip stimulation paired with /pa/ was significantly larger than when lip stimulation was paired with the /ka/ or finger-snap sound (/pa/ > /ka/, $p < 0.001$; /pa/ > finger-snap, $p < 0.001$; /ka/ > finger-snap, $p < 0.001$).

For tactile stimulation of the finger paired with auditory stimuli, there was a significant main effect of Hemisphere ($F(1, 20) = 5.416$ $p = 0.031$, $\eta^2 = 0.023$), with greater alpha ERD in the left hemisphere, as well as a significant main effect of Region ($F(1, 20) = 16.651$, $p < 0.001$, $\eta^2 = 0.041$), with greater alpha ERD in the central than temporal regions. There was no significant main effect of Sound, and no significant interaction between the factors.

Event-related alpha band responses to unimodal auditory stimuli

For alpha ERSP elicited to the unimodal auditory stimuli, there was a main effect of Hemisphere ($F(1, 20) = 9.917$, $p = 0.005$, $\eta^2 = 0.021$), with significantly greater alpha ERD in the left than in the right hemisphere. There was also a main effect of region ($F(1, 20) = 15.111$, $p < 0.001$, $\eta^2 = 0.059$), with greater alpha ERD in the central region than in the temporal region. Most importantly, there was no significant main effect of Sound ($F(2, 40) = 2.637$, $p = 0.094$) and no significant interaction between any of the factors. Figure 9 shows topographic plots of alpha ERD for the unisensory auditory conditions.

Discussion

The neural processes involved in multisensory interactions have been studied using both single neuron recordings in animals (e.g., Henschke et al. 2015; Bizley et al. 2007; Lakatos et al. 2007; Kayser et al. 2005) and functional neuroimaging methods with humans (e.g., Ro et al. 2012; Beauchamp 2005; Foxe et al. 2000). These investigations have contributed to our understanding of basic functions of multisensory processes, such as stimulus detection and spatial localization, which mainly involve multisensory neurons in the superior colliculus, thalamus, and association cortex (Wu et al. 2015; Stein and Stanford 2008). However, the role of multisensory interactions in higher level functions, such as object recognition, action perception, speech perception and even social perception, remains poorly understood (Alais et al. 2010; Campanella and Belin 2007).

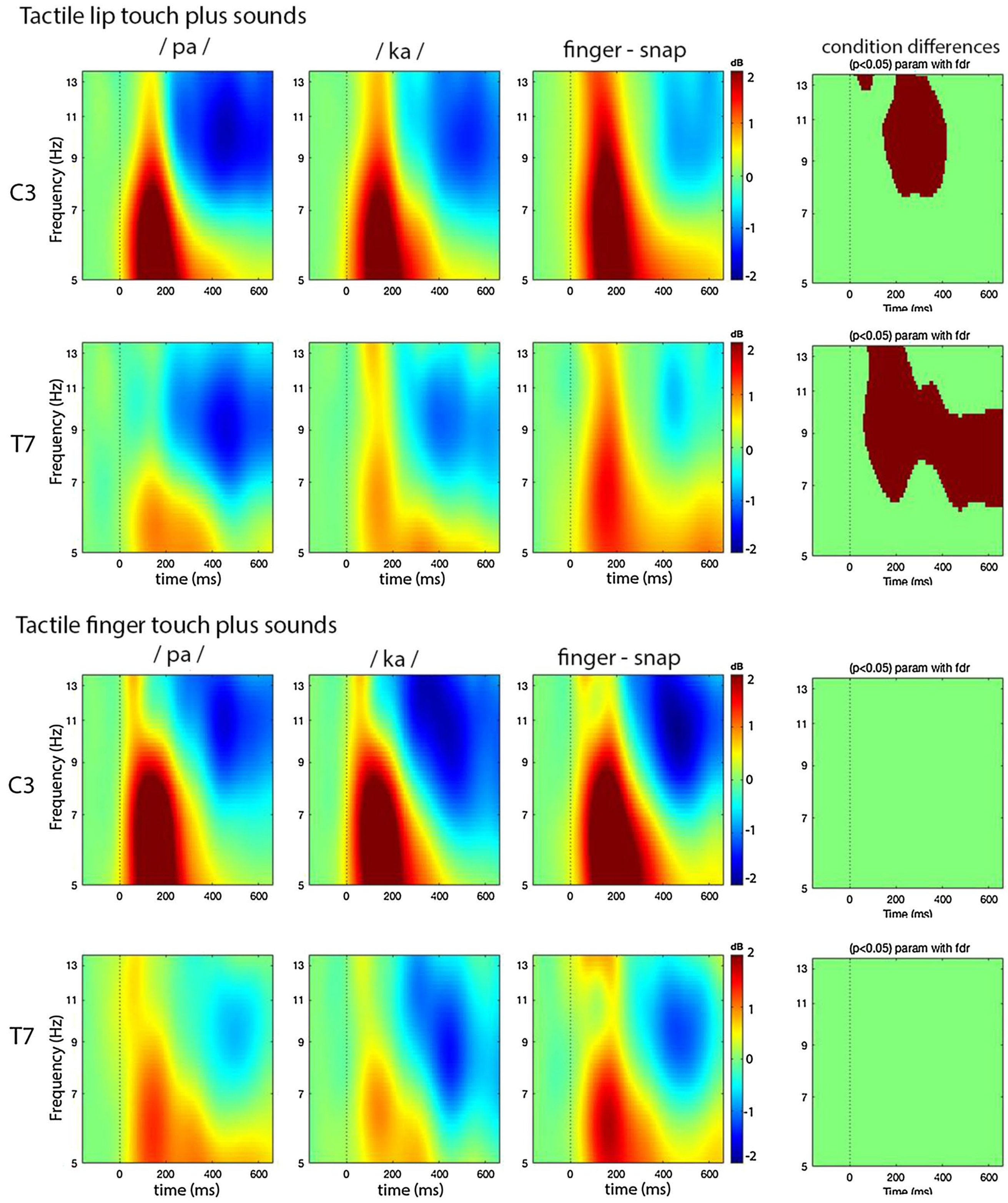

Fig. 8 Time–frequency plots at C3 and T7 following tactile stimulation of the lip paired with /pa/, /ka/, and the finger-snap sound (upper panel) and after tactile stimulation of the finger paired with each sound (lower panel). The right column shows the time and frequency range where the ERSP differs significantly between the sounds ($p < 0.05$, with FDR correction)

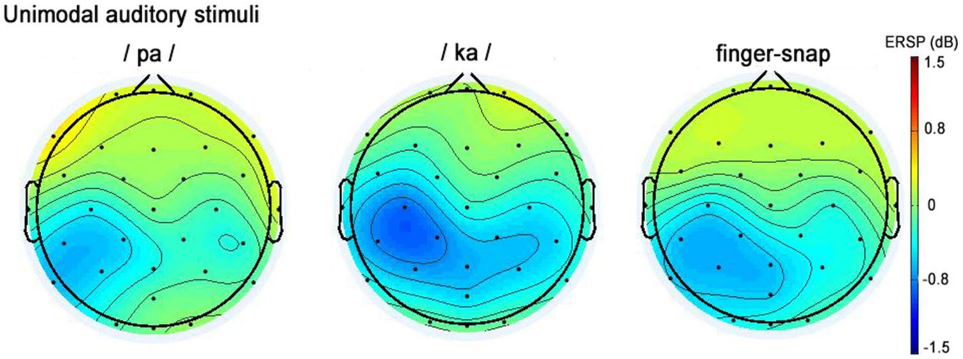

Fig. 9 Mean ERSP in the alpha (8–14 Hz) band for unisensory auditory stimuli, averaged across the window of 250–600 ms post-stimulus

Sensory inputs that are perceptually coherent or semantically congruent usually elicit enhanced neural responses in cortical areas associated with multisensory processing, such as superior temporal sulcus and gyrus, as well as primary auditory cortex (Barraclough et al. 2005; Wright et al. 2003; Calvert et al. 1999). The current findings are consistent with prior work showing that neural processing of multisensory input is modulated by cross-modal congruency (Alais, et al. 2010; Driver and Noesselt 2008). Our results extend the understanding of these effects by demonstrating the importance of effector-specific congruency in the context of stimuli related to human actions. Although the low-density EEG arrays used in the current study limit the ability to infer the specific cortical origins of the enhanced neural processing of bimodal input, the overall enhanced neural processing of matching auditory–tactile stimuli is consistent with previously proposed functional roles for congruency multisensory integration.

The current study further adds to the extant literature by providing novel neural evidence concerning the connections between auditory perception and somatosensation, which to this point have mainly been investigated using behavioral measures (Bruderer et al. 2015; Thomas et al. 2013; Ito and Ostry 2012; Gick and Derrick 2009). To our knowledge, the current study is the first to establish neurophysiological evidence for an effector-specific correspondence or congruency effect on auditory–somatosensory integration.

We found that enhancement of sensory-evoked neural activation (as indexed by N1 amplitude) depended on the congruence between the body part that was stimulated and the bodily site involved in the production of the sounds, for both speech and non-speech domains. Increased N1 amplitudes were observed in response to simultaneous auditory and tactile stimuli that were congruent in terms of the origin of sound production and the site of the tactile stimulation, relative to stimulus pairs that did not match. Specifically, amplitude of the N1 response was significantly larger when tactile lip stimulation was concurrently presented with the bilabial syllable /pa/ than when tactile lip stimulation was paired with either the non-labial syllable /ka/ or the finger-snap sound. Similarly, the coupling of tactile finger stimulation with the finger-snap sound elicited a larger N1 response than the pairing of tactile finger stimulation with the speech sounds. Analyses of the response elicited by the unimodal auditory stimuli revealed no difference in N1 amplitude to the three sounds, suggesting that the congruency effect on the auditory–tactile N1 response was not simply reducible to different acoustical features of the auditory stimuli.

The effect of bimodal congruency on N1 amplitude in the ERP signal fits with previous findings for auditory and visual stimuli (Molholm et al. 2004). Importantly, our findings further this understanding by showing congruency effects for tactile stimulation of the effectors involved with the production of the specific sounds. Prior work has found that ERP response to tactile lip stimulation is augmented by viewing speech movements (Möttönen et al. 2005). Those authors did not find similar effects for simply listening to speech, but their study did not involve precise temporal coordination of auditory and tactile stimulation. By employing a protocol involving simultaneous bimodal stimulation, we were able to show enhancement of N1 amplitude when tactile lip stimulation was paired with hearing a bilabial speech sound.

Our findings for N1 amplitude also provide novel evidence for an effector-specific auditory–tactile congruency effect outside of the speech domain. Specifically, enhanced N1 amplitude was also found when the finger-snapping sound was paired with tactile stimulation of the finger. This finding provides a novel connection between the current study and a broader literature on the bimodal linkages between action production and action perception. Much of the prior work in this area has relied on paradigms involving visual observation of others' actions, and has tended to emphasize the role of the motor system. However, the role of somatosensory factors in mediating self-other correspondences is an area of growing interest in experimental studies involving both adults (Keysers et al. 2010) and infants (Marshall and Meltzoff 2015; Meltzoff and Moore 1997). The findings from the current study add to this literature by showing the potential for studying interactions between auditory perception and tactile stimulation in the context of sounds that are produced by bodily actions.

As with the N1 response, the extent of post-stimulus alpha desynchronization showed an effector-specific response to the pairing of tactile lip stimulation with speech sounds. The concurrent presentation of tactile lip stimulation with the bilabial syllable /pa/ elicited significantly greater alpha ERD than did tactile lip stimulation accompanied by either the non-labial syllable /ka/ or the finger-snap sound. Alpha oscillations over sensory cortex are thought to reflect local sensory cortex excitability (Foxe et al. 1998), and the extent of alpha desynchronization is thought to be sensitive to multisensory integration (Plöchl et al. 2016; Lange et al. 2013). The current results extend work in this area by showing that cortical excitation is enhanced when tactile stimulation to the lip is accompanied by a speech sound that involves the contact between both lips as the primary place of articulation (bilabial production).

One point of note here is that the alpha ERD was less sensitive than N1 amplitude as a measure of the multimodal congruency effects, perhaps because of the nature of the task employed. It is worth noting that participants responded to each stimulus by pressing a foot petal with their right foot (around 400 ms after stimulus onset); therefore, the alpha ERD observed in the left hemisphere starting around 200 ms may have been potentially affected by the foot movement. The action of pressing the foot pedal may have resulted in background alpha desynchronization at central sites across all conditions, which may account for the absence of a congruency effect for alpha ERD in response to the pairing of tactile finger stimulation with the finger-snap sound. Additionally, because of differences in experience with the bimodal stimulus combinations and differential expertise with the motor acts, auditory–somatosensory integration for speech may be more robust than for finger-snapping. We experience concurrent lip sensation with bilabial sounds during our everyday speech productions, whereas finger-snapping is less frequent. In addition, previous studies have suggested a privileged speech multimodal network starting in infancy (e.g., Kuhl et al. 2014; Kuhl and Meltzoff 1984, 1996), with neural connections in the adult brain between auditory cortex, sensorimotor areas, and multisensory regions in the superior temporal sulcus (Doehrmann and Naumer 2008; Kayser et al. 2008; Calvert and Thesen 2004; van Atteveldt et al. 2004; Calvert et al. 2000). Auditory–tactile interactions connected with speech articulation may be more robust and easier to detect across multiple neural measures than for non-speech actions. Future studies are needed to further test whether effector-specific congruency for auditory–somatosensory interactions outside of the speech domain can be documented beyond early stimulus feature processing (as indexed by the amplitude of the N1).

For the behavioral measures taken in this study, all combinations of audio–tactile stimulation yielded faster response times to multisensory stimuli than to unimodal stimuli. This finding is consistent with previous findings of response facilitation to multisensory stimulation (Senkowski et al. 2007; Murray et al. 2005). For the bimodal stimuli, shorter reaction times were observed when tactile lip stimulation was paired with speech sounds (/pa/ and /ka/) than when paired with a non-speech (finger-snap) sound. The pattern of findings suggests that congruency effects in the behavioral responses were limited to a broad correspondence between tactile lip stimulation and speech sounds and not to the correspondence between finger sounds and finger stimulation. However, it should also be noted as a limitation of the behavioral measure that participants were only instructed to "respond" when they heard a sound or felt the tactile stimulation. They were not instructed to respond as fast as possible, since the response demand was mainly acting as a way to maintain participants' alertness throughout the experiment.

Our interpretation of the observed congruency effects in the ERP signal posits that the larger N1 and ERSP responses we observed are indicative of multisensory integration that occurs in a relatively automatic or pre-attentional fashion. However, it should also be noted that there has been intense debate in the literature concerning the role of attention in multisensory integration. Although one position in this debate is that multisensory integration precedes attentional selection, it has become clearer that the interaction between multisensory integration and attention is more nuanced than allowed for by this simple sequential model (Talsma et al. 2010). Although our task did not involve explicit instructions to attend to one modality over another, it is possible that repeated stimulation of the lip or finger implicitly and automatically drew participants' attention to those body parts, which could influence the processing of the accompanying tactile stimuli. Such an explanation does not negate our results, but instead points to the difficulty of resolving the debate about the interaction between attention and multisensory integration (Hartcher-O'Brien et al. 2017; Macaluso et al. 2016).

In summary, the current study demonstrated that audio–tactile stimuli that were either congruent or incongruent in terms of bodily effectors evoked different patterns of cortical activation, as characterized by N1 amplitude and alpha-range desynchronization in the EEG signal. Critically, neural processing of bimodal input was significantly enhanced when auditory and tactile stimuli were congruent—specifically when the auditory stimuli were paired with tactile stimulation of the particular effectors associated with production of the sounds. The audio–tactile enhancement reported here is, therefore, not a general effect, but is modulated by cross-modal congruency.

The different patterns of cortical activation to audio–tactile stimuli involving articulatory and hand actions may indicate privileged connections between speech sounds and the somatosensory cortices involved in articulatory acts, or they may reflect more associative familiarity effects. Exploring the balance of these possibilities should be the target of future studies. In particular, studies of infant participants may help to determine the role of speech and non-speech action experience in the development of the effector-specific enhancement effects seen here with adults. Studies using imaging methods that allow higher spatial resolution can further delineate neural responses in auditory and somatosensory cortices to congruent versus incongruent auditory–tactile pairings, and will elucidate the role of functional connections between primary sensory cortices and association cortex in modulating effector-specific multisensory effects.

Acknowledgements The authors thank Ashley Drew, Staci Weiss, Rebecca Laconi, Jeb Taylor and Rebecca Feldman for their help with data collection. The writing of this article was supported in part by awards from the National Institutes of Health (1R21HD083756) and the National Science Foundation (BCS-1460889 and SMA-1540619).

References

Alais D, Newell FN, Mamassian P (2010) Multisensory processing in review: from physiology to behaviour. See Perceiv 23(1):3–38

Arsenault JS, Buchsbaum BR (2015) No evidence of somatotopic place of articulation feature mapping in motor cortex during passive speech perception. Psychon Bull Rev 23(4):1231–1240

Avikainen S, Forss N, Hari R (2002) Modulated activation of the human SI and SII cortices during observation of hand actions. NeuroImage 15(3):640–646

Badin P, Tarabalka Y, Elisei F (2010) Can you “read” tongue movements? Evaluation of the contribution of tongue display to speech understanding. Speech Commun 52(6):493–503

Barraclough NE, Xiao DK, Baker CI, Oram MW, Perrett DI (2005) Integration of visual and auditory information by superior temporal sulcus neurons responsive to the sight of actions. J Cogn Neurosci 17(3):377–391

Bauer M, Oostenveld R, Peeters M, Fries P (2006) Tactile spatial attention enhances gamma-band activity in somatosensory cortex and reduces low-frequency activity in parieto-occipital areas. J Neurosci 11 26(2):490–501

Beauchamp MS (2005) See me, hear me, touch me: multisensory integration in lateral occipital-temporal cortex. Curr Opin Neurobiol 15(2):145–153

Behroozmand R, Shebek R, Hansen DR, Oya H, Robin DA, Howard MA, Greenlee JDW (2015) Sensory-motor networks involved in speech production and motor control: an fMRI study. NeuroImage 109:418–428

Bizley JK, Nodal FR, Bajo VM, Nelken I, King AJ (2007) Physiological and anatomical evidence for multisensory interactions in auditory cortex. Cereb Cortex 17(9):2172–2189

Boersma P, Weenink D (2009) Praat: doing phonetics by computer (Version 5.1. 05). Retrieved 1 May 2009

Brett-Green BA, Miller LJ, Gavin WJ, Davies PL (2008) Multisensory integration in children: a preliminary ERP study. Brain Res 1242:283–290

Bruderer AG, Danielson DK, Kandhadai P, Werker JF (2015) Sensorimotor influences on speech perception in infancy. Proc Natl Acad Sci 112(14):13531–13536

Buccino G, Binkofski F, Fink GR, Fadiga L, Fogassi L, Gallese V, Freund H-J (2001) Action observation activates premotor and parietal areas in a somatotopic manner: an fMRI study. Eur J Neurosci 13(2):400–404

Callan DE, Jones JA, Callan AM, Akahane-Yamada R (2004) Phonetic perceptual identification by native- and second-language speakers differentially activates brain regions involved with acoustic phonetic processing and those involved with articulatory-auditory/orosensory internal models. NeuroImage 22(3):1182–1194

Calvert GA, Thesen T (2004) Multisensory integration: methodological approaches and emerging principles in the human brain. J Physiol Paris 98(1–3):191–205

Calvert GA, Brammer MJ, Bullmore ET, Campbell R, Iversen SD, David AS (1999) Response amplification in sensory-specific cortices during crossmodal binding. NeuroReport 10(2):2619–2623

Calvert GA, Campbell R, Brammer MJ (2000) Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex. Curr Biol 10(12):649–657

Campanella S, Belin P (2007) Integrating face and voice in person perception. Trends Cogn Sci 11(12):535–543

Campbell T, Winkler I, Kujala T (2007) N1 and the mismatch negativity are spatiotemporally distinct ERP components: disruption of immediate memory by auditory distraction can be related to N1. Psychophysiology 44:530–540

Champoux F, Shiller DM, Zatorre RJ (2011) Feel what you say: an auditory effect on somatosensory perception. PLoS One 6(8):e22829

Cheyne D, Gaetz W, Garnero L, Lachaux JP, Ducorps A, Schwartz D, Varela FJ (2003) Neuromagnetic imaging of cortical oscillations accompanying tactile stimulation. Cogn Brain Res 17(3):599–611

Clayson PE, Baldwin SA, Larson MJ (2013) How does noise affect amplitude and latency measurement of event-related potentials (ERPs)? A methodological critique and simulation study. Psychophysiology 50(2):174–186

Correia JM, Jansma BMB, Bonte M (2015) Decoding articulatory features from fMRI responses in dorsal speech regions. J Neurosci 35(45):15015–15025

D’Ausilio A, Bartoli E, Maffongelli L, Berry JJ, Fadiga L (2014) Vision of tongue movements bias auditory speech perception. Neuropsychologia 63:85–91

Della Penna S, Torquati K, Pizzella V, Babiloni C, Franciotti R, Rossini PM, Romani GL (2004) Temporal dynamics of alpha and beta rhythms in human SI and SII after galvanic median nerve stimulation. A MEG study. NeuroImage 22(4):1438–1446

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

Deschrijver E, Wiersema JR, Brass M (2016) The interaction between felt touch and tactile consequences of observed actions: an action-based somatosensory congruency paradigm. Soc Cogn Affect Neurosci 11(7):1162–1172

Dockstader C, Gaetz W, Cheyne D, Wang F, Castellanos FX, Tannock R (2008) MEG eventrelated desynchronization and synchronization deficits during basic somatosensory processing in individuals with ADHD. Behav Brain Funct 4(1):8

Doehrmann O, Naumer MJ (2008) Semantics and the multisensory brain: how meaning modulates processes of audio-visual integration. Brain Res 1242:136–150

Driver J, Noesselt T (2008) Multisensory interplay reveals cross-modal influences on 'sensory-specific' brain regions, neural responses, and judgments. Neuron 57(1):11–23

Eulitz C, Diesch E, Pantev C, Hampson S, Elbert T (1995) Magnetic and electric brain activity evoked by the processing of tone and vowel stimuli. J Neurosci 15(4):2748–2755

Foxe JJ, Simpson GV, Ahlfors SP (1998) Parieto-occipital approximately 10 Hz activity reflects anticipatory state of visual attention mechanisms. NeuroReport 9(17):3929–3933

Foxe JJ, Morocz IA, Murray MM, Higgins BA, Javitt DC, Schroeder CE (2000) Multisensory auditory-somatosensory interactions in early cortical processing revealed by high-density electrical mapping. Brain Res Cogn Brain Res 10(1–2):77–83

Fujioka T, Ross B (2008) Auditory processing indexed by stimulus-induced alpha desynchronization in children. Int J Psychophysiol 68(2):130–140

Gazzola V, Aziz-Zadeh L, Keysers C (2006) Empathy and the somatotopic auditory mirror system in humans. Curr Biol 16(18):1824–1829

Gick B, Derrick D (2009) Aero-tactile integration in speech perception. Nature 462(7272):502–504

Gick B, Ikegami Y, Derrick D (2010) The temporal window of audio-tactile integration in speech perception. J Acoust Soc Am 128(5):342–346

Hartcher-O'Brien J, Soto-faraco S, Adam R (2017) Editorial: a matter of bottom-up or top-down processes: the role of attention in multisensory integration. Front Integr Neurosci 11:5. doi:10.3389/fnint.2017.00005

Hasson U, Nir Y, Levy I, Fuhrmann G, Malach R (2004) Intersubject synchronization of cortical activity during natural vision. Science 303(5664):1634–1640

Henschke JU, Noesselt T, Scheich H, Budinger E (2015) Possible anatomical pathways for short-latency multisensory integration processes in primary sensory cortices. Brain Struct Funct 22(2):955–977

Hickok G (2014) The myth of mirror neurons: the real neuroscience of communication and cognition. WW Norton & Company, New York

Hickok G, Houde J, Rong F (2011) Sensorimotor integration in speech processing: computational basis and neural organization. Neuron 69(3):407–422

Hoffmann S, Falkenstein M (2008) The correction of eye blink artefacts in the EEG: a comparison of two prominent methods. PLoS One 3(8):e3004

Ito T, Ostry DJ (2012) Speech sounds alter facial skin sensation. J Neurophysiol 107(1):442–447

Ito T, Gracco VL, Ostry DJ (2014) Temporal factors affecting somatosensory-auditory interactions in speech processing. Front Psychol 5(10):1–10

Kayser C, Petkov CI, Augath M, Logothetis NK (2005) Integration of touch and sound in auditory cortex. Neuron 48(2):373–384

Kayser C, Petkov CI, Logothetis NK (2008) Visual modulation of neurons in auditory cortex. Cereb Cortex 18(7):1560–1574

Keysers C, Kaas JH, Gazzola V (2010) Somatosensation in social perception. Nat Rev Neurosci 11(6):417–428

Kuhl PK, Meltzoff AN (1982) The bimodal perception of speech in infancy. Science 218(4577):1138–1141

Kuhl PK, Meltzoff AN (1984) The intermodal representation of speech in infants. Infant Behav Dev 7(3):361–381

Kuhl PK, Meltzoff AN (1996) Infant vocalizations in response to speech: vocal imitation and developmental change. J Acoust Soc Am 100(4):2425–2438

Kuhl PK, Ramirez RR, Bosseler A, Lin JF, Imada T (2014) Infants' brain responses to speech suggest analysis by synthesis. Proc Natl Acad Sci 111(31):11238–11245

Lahav A, Saltzman E, Schlaug G (2007) Action representation of sound: audiomotor recognition network while listening to newly acquired actions. J Neurosci 27(2):308–314

Lakatos P, Chen CM, O'Connell MN, Mills A, Schroeder CE (2007) Neuronal oscillations and multisensory interaction in primary auditory cortex. Neuron 53(2):279–292

Lange J, Oostenveld R, Fries P (2011) Perception of the touch-induced visual double-flash illusion correlates with changes of rhythmic neuronal activity in human visual and somatosensory areas. Neuroimage 15 54(2):1395–1405

Lange J, Christian N, Schnitzler A (2013) Audio-visual congruency alters power and coherence of oscillatory activity within and between cortical areas. NeuroImage 79:111–120

Macaluso E, Driver J (2005) Multisensory spatial interactions: a window onto functional integration in the human brain. Trends Neurosci 28(5):264–271

Macaluso E, Noppeney U, Talsma D, Vercillo T, Hartcher-O'Brien J, Adam R (2016) The curious incident of attention in multisensory integration: bottom-up vs top-down. Multisens Res 29(6–7):557–583

Makeig S (1993) Auditory event-related dynamics of the EEG spectrum and effects of exposure to tones. Electroencephalogr Clin Neurophysiol 86(4):283–293

Marshall PJ, Meltzoff AN (2015) Body maps in the infant brain. Trends Cogn Sci 19(9):499–505

McGurk H, MacDonald J (1976) Hearing lips and seeing voices. Nature 264(5588):746–748

Meltzoff AN, Moore MK (1997) Explaining facial imitation: a theoretical model. Early Dev Parent 6(3–4):179

Molholm S, Ritter W, Javitt DC, Foxe JJ (2004) Multisensory visual-auditory object recognition in humans: a high-density electrical mapping study. Cereb Cortex 14(4):452–465

Möttönen R, Järveläinen J, Sams M, Hari R (2005) Viewing speech modulates activity in the left SI mouth cortex. NeuroImage 24(3):731–737

Müller N, Weisz N (2012) Lateralized auditory cortical alpha band activity and interregional connectivity pattern reflect anticipation of target sounds. Cereb Cortex 22(7):1604–1613

Murray MM, Molholm S, Michel CM, Heslenfeld DJ, Ritter W, Javitt DC, Foxe JJ (2005) Grabbing your ear: rapid auditory-somatosensory multisensory interactions in low-level sensory cortices are not constrained by stimulus alignment. Cereb Cortex 15(7):963–974

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

Nishimura S, Tomita Y, Tanaka T, Kasagi S, Takashima S, Takeshita K (1986) Developmental change of short latency somatosensory evoked potential waves between P3 and N1 components in children. Brain Dev 8(1):6–9

Paulus M, Hunnius S, Van Elk M, Bekkering H (2012) How learning to shake a rattle affects 8-month-old infants' perception of the rattle's sound: electrophysiological evidence for action-effect binding in infancy. Dev Cogn Neurosci 2(1):90–96

Pihko E, Nevalainen P, Stephen J, Okada Y, Lauronen L (2009) Maturation of somatosensory cortical processing from birth to adulthood revealed by magnetoencephalography. Clin Neurophysiol 120(8):1552–1561

Pizzamiglio L, Aprile T, Spitoni G, Pitzalis S, Bates E, D'Amico S, Di Russo F (2005) Separate neural systems for processing action- or non-action-related sounds. NeuroImage 24(3):852–861

Plöchl M, Gaston J, Mermagen T, König P, Hairston WD (2016) Oscillatory activity in auditory cortex reflects the perceptual level of audio-tactile integration. Sci Rep 6:33693

Pulvermüller F, Fadiga L (2010) Active perception: sensorimotor circuits as a cortical basis for language. Nat Rev Neurosci 11(5):351–360

Pulvermüller F, Huss M, Kherif F, del Prado M, Martin F, Hauk O, Shtyrov Y (2006) Motor cortex maps articulatory features of speech sounds. Proc Natl Acad Sci USA 103(2):7865–7870

Ro T, Hsu J, Yasar NE, Elmore LC, Beauchamp MS (2009) Sound enhances touch perception. Exp Brain Res 195(1):135–143

Ro T, Ellmore TM, Beauchamp MS (2012) A neural link between feeling and hearing. Cereb Cortex 23(7):1724–1730

Rossi S, Tecchio F, Pasqualetti P, Ulivelli M, Pizzella V, Romani G, Rossini P (2002) Somatosensory processing during movement observation in humans. Clin Neurophysiol 113(1):16–24

Russo N, Foxe JJ, Brandwein AB, Altschuler T, Gomes H, Molholm S (2010) Multisensory processing in children with autism: high-density electrical mapping of auditory–somatosensory integration. Autism Res 3(5):253–267

Saby JN, Meltzoff AN, Marshall PJ (2013) Infants' somatotopic neural responses to seeing human actions: i've got you under my dkin. PLoS One 8(10):e77905

Sadaghiani S, Kleinschmidt A (2016) Brain networks and α- oscillations: structural and functional foundations of cognitive control. Trends Cogn Sci 20(11):805–817

Schomers MR, Pulvermüller F (2016) Is the sensorimotor cortex relevant for speech perception and understanding? An integrative review. Front Hum Neurosci 10(9):435

Schubert R, Haufe S, Blankenburg F, Villringer A, Curio G (2008) Now you'll feel It-now You won't: EEG rhythms predict the effectiveness of perceptual masking. J Cogn Neurosci 21:2407–2419

Senkowski D, Talsma D, Grigutsch M, Herrmann CS, Woldorff MG (2007) Good times for multisensory integration: effects of the precision of temporal synchrony as revealed by gamma-band oscillations. Neuropsychologia 45(3):561–571

Shen G, Saby JN, Drew AR, Marshall PJ (2017) Exploring potential social influences on brain potentials during anticipation of tactile dtimulation. Brain Res 1659:8–18

Stanford TR, Stein BE (2007) Superadditivity in multisensory integration: putting the computation in context. NeuroReport 18(8):787–792

Stein BE, Stanford TR (2008) Multisensory integration: current issues from the perspective of the single neuron. Nature Rev Neurosci 9(4):255

Talsma D, Senkowski D, Soto-Faraco S, Woldorff MG (2010) The multifaceted interplay between attention and multisensory integration. Trends Cogn Sci 14(9):400–410

Thomas R, Sink J, Haggard P (2013) Sensory effects of action observation. Exp Psycol 60(5):335–346

Tiihonen J, Hari R, Kajola M, Karhu J, Ahlfors S, Tissari S (1991) Magnetoencephalographic 10-Hz rhythm from the human auditory cortex. Neurosci Lett 129(2):303–305

Tiitinen H, Sivonen P, Alku P, Virtanen J, Näätänen R (1999) Electromagnetic recordings reveal latency differences in speech and tone processing in humans. Cogn Brain Res 8(3):355–363

Töllner T, Gramann K, Müller HJ, Eimer M (2009) The anterior N1 component as an index of modality shifting. J Cogn Neurosci 21(9):1653–1669

Tsilionis E, Vatakis A (2016) Multisensory binding: is the contribution of synchrony and semantic congruency obligatory? Current Opinion Behav Sci 8:7–13

van Atteveldt N, Formisano E, Goebel R, Blomert L (2004) Integration of letters and speech sounds in the human brain. Neuron 43(2):271–282

Van Atteveldt NM, Formisano E, Blomert L, Goebel R (2007) The effect of temporal asynchrony on the multisensory integration of letters and speech sounds. Cereb Cortex 17(4):962–974

Virtanen J, Ahveninen J, Ilmoniemi RJ, Näätänen R, Pekkonen E (1998) Replicability of MEG and EEG measures of the auditory N1/N1m-response. Electroencephalogr Clin Neurophysiol 108(3):291–298

Welch RB (1999) Meaning, attention, and the "unity assumption" in the intersensory bias of spatial and temporal perceptions. Adv Psychol 129:371–387

Wheaton KJ, Thompson JC, Syngeniotis A, Abbott DF, Puce A (2004) Viewing the motion of human body parts activates different regions of premotor, temporal, and parietal cortex. NeuroImage 22(1):277–288

Wilson SM, Saygin AP, Sereno MI, Iacoboni M (2004) Listening to speech activates motor areas involved in speech production. Nat Neurosci 7(7):701–702

Wilson EC, Braida LD, Reed CM (2010) Perceptual interactions in the loudness of combined auditory and vibrotactile stimuli. J Acoust Soc Am 127(5):3038–3043

Wright TM, Pelphrey KA, Allison T, McKeown MJ, McCarthy G (2003) Polysensory interactions along lateral temporal regions evoked by audiovisual speech. Cereb Cortex 13(10):1034–1043

Wu C, Stefanescu RA, Martel DT, Shore SE (2015) Listening to another sense: somatosensory integration in the auditory system. Cell Tissue Res 361(1):233–250