Exp Brain Res (2008) 185:11–20
DOI 10.1007/s00221-007-1127-5

RESEARCH ARTICLE

Asynchrony from synchrony: long-range gamma-band neural synchrony accompanies perception of audiovisual speech asynchrony

Sam M. Doesburg · Lauren L. Emberson · Alan Rahi · David Cameron · Lawrence M. Ward

Received: 24 January 2007 / Accepted: 4 September 2007 / Published online: 6 October 2007

Abstract Real-world speech perception relies on both auditory and visual information that fall within the tolerated range of temporal coherence. Subjects were presented with audiovisual recordings of speech that were offset by either 30 or 300 ms, leading to perceptually coherent or incoherent audiovisual speech, respectively. We provide electroencephalographic evidence of a phase-synchronous gamma-oscillatory network that is transiently activated by the perception of audiovisual speech asynchrony, showing both topological and time–course correspondence to networks reported in previous neuroimaging research. This finding addresses a major theoretical hurdle regarding the mechanism by which distributed networks serving a common function achieve transient functional integration. Moreover, this evidence illustrates an important dissociation between phase-synchronization and stimulus coherence, highlighting the functional nature of network-based synchronization.

Electronic supplementary material The online version of this article (doi:10.1007/s00221-007-1127-5) contains supplementary material, which is available to authorized users.

S. M. Doesburg (✉) · L. L. Emberson · A. Rahi · D. Cameron · L. M. Ward
Psychophysics and Cognitive Neuroscience Laboratory, Department of Psychology, University of British Columbia, D.T. Kenny Bldg, 2136 West Mall, Vancouver, BC, Canada V6T 1Z4
e-mail: sam@psych.ubc.ca

L. M. Ward
The Brain Research Centre, University of British Columbia, Vancouver, BC, Canada

L. L. Emberson
Cognitive Neuroscience Lab, Department of Psychology, Cornell University, New York, NY, USA

Keywords EEG · Neural synchronization · Gamma · Multimodal speech · Integration

Introduction

Our conscious window on the world is a multisensory one, but the neural bases for multisensory integration have remained elusive. Mechanisms by which brain activity can match sensory features from different modalities to a single perceptual object and bind these features together to form a unified percept are necessary to support multisensory perception. Our robust perceptual system tolerates a moderate degree of temporal incoherence between sensory features from different modalities, but when the temporal incoherence goes beyond this range multisensory integration fails to occur (Dixon and Spitz 1980; Lewkowicz 1996). A related problem is that of the perception of simultaneity or repeated simultaneity, also referred to as perceptual synchrony (King 2005).

Audiovisual integration in speech is a promising arena for the investigation of multisensory integration as it provides complex, continuous stimuli for whose perception we typically rely on both auditory and visual information. The classic example of the profound effect of audiovisual integration on speech perception is the McGurk effect, where the presence of coordinated visual information alters the categorical perception of a phoneme from what the auditory information alone would produce (McGurk and MacDonald 1975).

In a comprehensive review of the neuroimaging literature examining audiovisual temporal integration, Calvert () concluded that crossmodal processing involves the activation of multiple brain regions likely working in concert to match and integrate multimodal inputs. A distributed

network of brain regions, composed of the superior temporal sulcus (STS), intraparietal sulcus, Heschl's gyrus and the inferior frontal gyrus, has been associated with perception of fused bimodal speech stimuli (Miller and D'Esposito 2005). However, other localization studies have related audiovisual speech fusion to restricted areas of association cortex located in the superior temporal sulcus (STS) and/or superior temporal gyrus, rather than to the activity of a distributed network (Calvert 2001; Wright et al. 2003). Complementary to the neural mechanisms for audiovisual speech fusion are those that identify mismatches between auditory and visual speech streams. This system functions to ascertain when an auditory speech stream is incongruent with a visual speech stream, and should accordingly belong to a different perceptual object. A robust group of imaging studies has identified a brain network composed of posterior parietal, insular, prefrontal and cerebellar areas that is associated with the perception of asynchronous bimodal speech stimuli (Jones and Callan 2003; Kaiser et al. 2005; Miller and D'Esposito 2005).

It remains a general problem in cognitive neuroscience how these, or any, distributed networks of neural groups achieve functional coupling for the performance of a particular task. Long-range phase synchronization, particularly within the gamma band, has been proposed to be a mechanism that enacts transient functional integration of task-relevant, neural populations and enables perceptual binding (Engel and Singer 2001; Varela et al. 2001). In this view, neural populations participating in an active task-relevant network or representing features belonging to the same perceptual object oscillate synchronously in the gamma-band (Tallon-Baudry and Bertrand 1999).

Previous research supports the hypothesis that long-range gamma synchrony mediates networks associated with the perception of audiovisual stimuli. First, gamma-band activation (GBA), reflective of *local* neural synchrony, has been associated with audiovisual perception in both EEG and magnetoencephalographic (MEG) recordings. Attention increases GBA more for bimodal versus unimodal audiovisual stimuli, underscoring that GBA is affiliated with integrative brain processes (Senkowski et al. 2005). Perceptual fusion of bimodal speech stimuli has also been related to enhanced GBA (Callan et al. 2001). Furthermore, information integration across modalities, based on semantic congruity between auditory and visual speech streams, is associated with enhanced GBA (Yuval-Greenberg and Deouell 2007). Second, transient *long-range* synchronization in the gamma band has been correlated with coherent perception in both unimodal visual and unimodal auditory perception (Doesburg et al. 2005; Eggermont 2000; Rodriguez et al. 1999). Perceptual integration across modalities, for example between tactile and visual stimuli, has also linked to long-distance gamma-band synchronization (Kanayama et al. 2007). Of more specific relevance to speech perception, long-range synchronization between brain areas has been found to relate to the matching of auditory speech stimuli to expectations about the speaker's own voice. Gamma-band coherence between frontal and temporal electrodes is greater when subjects hear their own undistorted speech, relative to when they hear their speech frequency-shifted down by one semi-tone (Ford et al. 2005). This finding indicates that long-range interactions may also be involved in the matching of auditory speech streams to their appropriate object, in this case to the speaker.

Given the evidence relating GBA to perceptual binding within and across modalities, together with evidence that long-range gamma-band phase synchrony correlates with coherent perception, our hypothesis was that greater local and long-range gamma synchrony would be observed during coherent audiovisual speech percepts relative to incoherent percepts. Perception of speech asynchrony, however, does not simply reflect a failure of the neural systems responsible for binding speech stimuli. Rather the activity of a distributed network of brain areas has been linked to the detection of speech asynchrony. We accordingly hypothesized that increased long-range gamma-band synchronization would also be observed during the perception of incoherent audiovisual speech, relative to a prestimulus baseline, as these brain areas were activated and functionally integrated. This latter pattern of synchronization was predicted to be less pronounced than synchronization observed during perception of coherent audiovisual speech as it does not involve crossmodal binding. This pattern of synchrony should show a different topography, reflective of the distinct underlying network involved in detection of speech asynchrony, as opposed to that responsible for perceptual binding of audiovisual speech stimuli. To test these hypotheses we recorded 64 channel EEG while subjects viewed audiovisual speech stimuli in which the auditory and visual streams were offset by either a small or a large amount, leading to integrated or nonintegrated multimodal perception, respectively. Phase locking values were calculated between signals recorded at selected electrodes in order to measure long-distance synchronization in various frequency bands.

Methods

Audiovisual stimuli were recorded from four human speakers (two female). For each speaker, two segments were identified that contained a complete phrase uttered in approximately 2 s. The visual stream of each of these eight recordings was shifted forward by 30 ms relative to the auditory stream (Fig. 1) and then the 2-s video segments

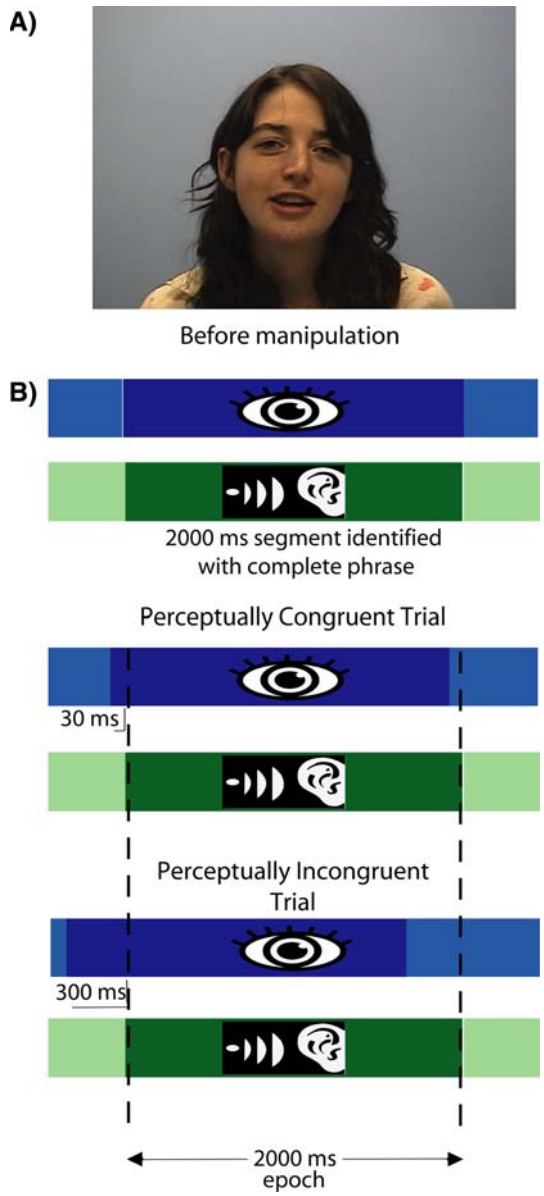

Fig. 1 **a** The stimulus display. **b** Construction of congruent and incongruent audiovisual speech stimuli. Two-second epochs of speech containing a complete phrase were identified. The visual stream was shifted forward relative to the auditory stream (30 ms for congruent trials, 300 ms for incongruent trials) and then 2,000 ms stimuli were extracted. There was no SOA between the visual and auditory and visual stimulus streams, as they both began and ended simultaneously

that contained the complete auditory phrases were extracted. Thus, there was no stimulus onset asynchrony (SOA) between visual and auditory stimulus streams but there was a temporal incongruence between the streams as the visual events were shifted in time relative to the corresponding audio events. This process was repeated with an offset of 300 ms, producing an additional eight 2-s stimuli for a total of 16 videos for experimental presentation. The offsets were chosen to be well within (30 ms) or outside (300 ms) the range of temporal integration for audiovisual speech stimuli producing perceptually congruent (synchronous) and incongruent (asynchronous) audiovisual streams, respectively (Dixon and Spitz 1980; Miller and D'Esposito 2005). For all stimuli, the audio and visual streams began and ended simultaneously, as the visual stream was shifted relative to the auditory stream before the 2-s stimulus segments were extracted.

Eleven subjects (average age 20.09; SD 1.76, five male) participated in the study after giving informed consent. Subjects were seated 35 cm in front of a computer monitor and received the auditory speech stimuli through headphones. Visual speech stimuli were presented centrally and subtended a visual angle of 31.92° horizontally and 20.14° vertically (the average size of the speakers' heads within this display was $11.87 \times 11.05°$). Between trials subjects viewed a grey screen with a centrally-located fixation point ($0.82 \times 0.82°$) for 2 s. Six blocks of 100 stimuli, in which randomly selected congruent and incongruent stimuli occurred in equal numbers, were presented to each subject. Subjects were told that half of the stimuli they would see were congruent and to press the X button on a computer keyboard when they perceived a congruent speech stimulus and the Y button when they perceived an incongruent speech stimulus. Subjects were told to respond only after the end of the 2-s-duration speech stimulus so as to minimize contamination of the EEG recordings by motor artifacts.

EEG data were recorded continuously from 59 electrodes positioned at standard 10–10 locations as well as three at non-standard sites below the inion using an SA isolated bioelectric amplifier, and were referenced to the right mastoid electrode. The electro-oculogram (EOG) was recorded bipolarly using electrodes beside and below the right eye. Electrode impedances were kept below 15 kΩ (sufficient since amplifier input impedance was >2 GΩ). EEG and EOG were amplified using a gain of 20,000, bandpass filtered from 0.1 to 100 Hz, digitally sampled at 500 Hz, and recorded for offline analysis. The protocol used in this research was approved by the Behavioural Research Ethics Board of the University of British Columbia and was conducted in accordance with the principles of the Helsinki Declaration.

Epochs were extracted for trials in which the subjects reported congruent perception when presented with video segments offset by 30 ms, and when subjects reported incongruent perception when presented with video segments offset by 300 ms. Only trials in which subjects responded after the offset of the stimulus were included in the analysis. Epochs containing ocular and nonocular artifacts were removed. Epochs containing artifacts were defined as those in which the EOG contained values outide of the ±75 μV range, or on which the voltage difference

between the beginning and the end of the epoch exceeded 70 μV for any electrode other than the EOG (as identified by the automatic rejection algorithm in EEGLAB; Delorme and Makeig 2004). Each subject's data were manually inspected to ensure that the artifact rejection algorithm was working as intended. For some subjects the criteria were slightly adjusted to more accurately identify artifacts. As the surrogate method for PLV analysis used here requires data to be combined across subjects (Lachaux et al. 1999), only the first 200 acceptable epochs from each condition were taken from each subject. This was done to prevent any observed effects in the group data being driven by responses present only in a small number of subjects. This procedure resulted in seven subjects who had more than, or nearly, 200 artifact-free epochs for each condition being included in the main analysis (average age 21.43; SD 1.99, four male, see Table 1 for numbers of epochs for each subject/condition). This yielded a total of 1,374 epochs for congruent percepts and 1,354 epochs for incongruent percepts.

Epochs were extracted from 400 ms before stimulus onset until 1200 ms after stimulus onset, but we report analyses only from –200 ms until 1,000 ms to avoid distorting edge effects from the phase locking value (PLV) analysis (see Freeman 2004). To attenuate volume conduction and to remove spurious synchronization resulting from the use of a right mastoid reference electrode, source current density (SCD) was derived from scalp potentials recorded at 62 electrode locations (excluding the EOG and reference channels). This was performed prior to the calculation of PLVs. To compute SCD we used a MATLAB script supplied by Carsten Allefeld (http://www.agnld.uni-potsdam.de/~allefeld/index.html) that implements the algorithm of Perrin et al. (1987, see their Eqs. 3 and 5). The SCD is a reference-free measure and largely reflects the activity of superficial neural generators near the recording electrodes (Lachaux et al. 1999). For this reason we refer to electrodes and cortical regions beneath those electrodes interchangeably, as all reported results pertain to the PLV_z values based on SCD. SCD data were analyzed for a 19-electrode montage, selected from the full 62-electrode set, positioned evenly about the scalp (minimum distance between electrodes ≈4 cm). Data were bandpass filtered digitally at 1-Hz intervals (passband = $f \pm 0.05f$, where f represents the filter frequency) between 6 and 60 Hz. The analytic signal

$$\varsigma(t) = f(t) + i\tilde{f}(t) = A(t)e^{i\phi(t)}$$

of each filtered waveform, where $\tilde{f}(t)$ is the Hilbert transform of the EEG recording, $f(t)$, and $i = \sqrt{-1}$, was calculated to determine the instantaneous amplitude, $A(t)$, and the instantaneous phase, $\phi(t)$, for each sample point (Pikovski et al. 2001). Instantaneous phases between various pairs of electrodes were compared to calculate PLVs between electrodes j and k for each time point t across epochs N (Lachaux et al. 1999):

$$PLV_{j,k,t} = N^{-1}\left|\sum_{N} e^{i\left[\phi_j(t)-\phi_k(t)\right]}\right|$$

PLV is a real value between 0 (random phase difference) and 1 (constant phase difference). To reduce the influence of volume conduction from constant sources on PLVs by removing the record of ongoing synchrony, we standardized them relative to a 200 ms pre-stimulus baseline. Standardization was accomplished by subtracting the mean baseline PLV from the PLV for each data point and dividing the difference by the standard deviation of the baseline PLV. The resulting PLV_z scores indicate changes from the average baseline PLV expressed in units of standard deviation; PLV_z scores usually range from –3 to +3, although they can be more negative or more positive. PLV between each electrode pair was standardized relative to the 200 ms baseline period for that electrode pair for trials within a particular condition. The baselines for the two conditions are essentially equivalent, however, as trials in each condition occurred randomly and with equal probability in each block.

To test the statistical significance of phase synchronization and desynchronization between pairs of electrodes, epochs were shuffled to create surrogate distributions of PLV_z values. Our surrogate distributions contained 200 random shufflings for each electrode pair at each data point. We considered a measured PLV_z above the 97.5th percentile of the surrogate distribution to be a significant increase in synchronization, and one below the 2.5th percentile a significant decrease in synchronization. Only long-range synchronization/desynchronization effects that reached statistical significance by this measure are discussed.

Table 1 Total number of epochs for each subject for the congruent and incongruent conditions

Subject	No. of epochs congruent	No. of epochs incongruent
1	200	200
2	180	199
3	200	198
4	200	200
5	200	200
6	200	157
7	194	200
Total	1,374	1,354

Results

Since the degree of temporal offset was chosen according to known windows for perceptual integration of audiovisual speech, subjects responded with high accuracy. However, trials with a 30 ms offset were successfully identified as congruent (3.9% errors) more accurately than the 300 ms offset trials were identified as incongruent (10.2% errors).

Long-range gamma-band phase synchronization was observed 170–250 ms after the onset of incongruent speech stimuli in a band of frequencies ranging from 37 to 44 Hz (Fig. 2a). A pattern of global phase synchronization in a somewhat lower frequency band, from 23 to 28 Hz, was observed 40–70 ms after the onset of congruent speech stimuli (Fig. 2b). These effects are also clearly visible in a direct comparison of the PLVs from the two stimulus conditions (Fig. 2c). No other reliable differences in long-distance synchrony were present during the analyzed epoch. A sustained increase in synchronization from 6 to 9 Hz beginning at the onset of the audiovisual speech stimuli, and a sustained decrease in synchronization from 10 to 19 Hz beginning at 170 ms, however, occurred for both congruent and incongruent speech trials (Fig. 2a, b). Long-range synchrony during the perception of audio-visually incongruent speech was anchored primarily at electrodes over frontal areas and left posterior cortex, accompanied by substantial synchronization between left temporal cortex and other electrodes (Fig. 3a). Although far less long-range gamma-band synchronization was seen 170–250 ms after the onset of audio-visually congruent speech, the topology of observed long-distance synchronization was similar to that seen during incongruent speech perception (Fig. 3a). These patterns of synchronization were robust as they were also observed at neighboring frequencies (see supplementary material).

It should also be noted that throughout the period of increased long-range synchronization increase there was also a distinct pattern of decrease in long-range gamma-band synchronization (long-range *desynchronization*) that also was considerably more pronounced during incongruent percepts, relative to congruent percepts (Fig. 3b). This long-range desynchronization appears between left frontal and posterior locations, and has a more medial topography than does the aforementioned long-range synchronization. Perception of coherent speech, conversely, was associated with long-range desynchronization in the gamma-band between right temporal and parietal electrodes and other, widespread locations (Fig. 3b). Since this long-range desynchronization occurred within a time–frequency window that was dominated by long-range *synchronization*, as indicated by the appearance of only synchronization increases in that time–frequency window in Fig. 2, they should be considered of secondary importance to the increased phase synchrony displayed in Fig. 3a.

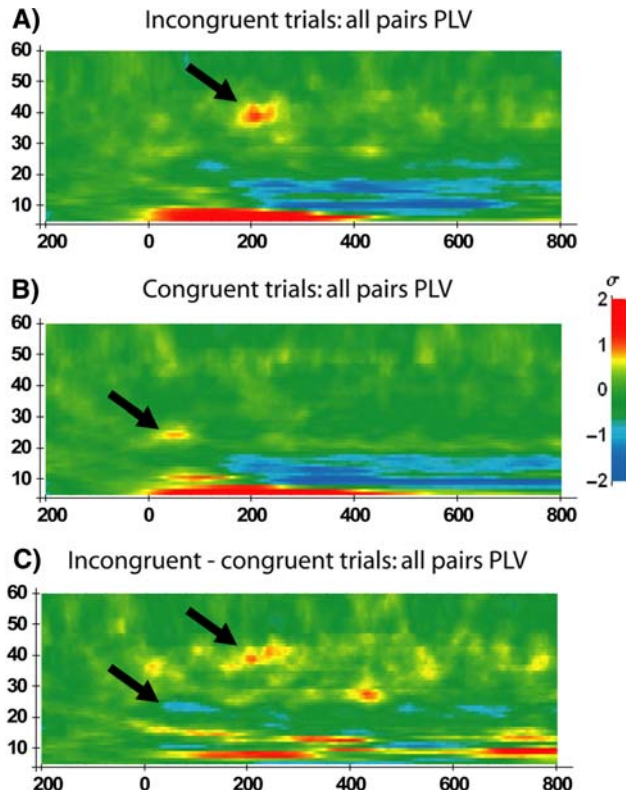

Fig. 2 **a** Global long-range synchrony increase accompanying incongruent perception, time-locked to the onset of audiovisual stimuli. Shown is PLV_z averaged across all analyzed electrode pairs (19 electrodes; 171 pairs). *Arrow* indicates the 37 to 44 Hz burst of increased phase-locking occurring 170 to 250 ms after the onset of incongruent speech stimuli. Changes from the pre-stimulus baseline, both positive (synchronization increase) and negative (synchronization decrease), are measured in standard deviation units (σ in *legend on right*). **b** Global long-range phase synchrony increase following the onset of congruent stimuli. Note the burst of increased phase-locking from 23 to 28 Hz occurring from 40 to 70 ms after stimulus onset. **c** Difference map displaying a direct comparison of the two stimulus conditions (incongruent–congruent). This demonstrates that effects shown in Fig. 2a, b are also visible in a direct comparison, as denoted by the *arrows*. The difference map was calculated by subtracting PLV_z, averaged across all 171 electrode pairs, for each data point for the congruent dataset from the corresponding data points in the incongruent dataset. Since the baseline mean and standard deviations are roughly the same for all points, this procedure reveals approximate average differences in PLV in units of standard deviation of the baseline PLV

Gamma-band amplitude, reflective of local gamma synchrony as opposed to synchronization between distant brain regions, showed a pattern of results different from that observed for long-range synchronization. An initial burst of gamma-band activation centred at ~40 Hz and peaking around 200 ms after the onset of audiovisual speech stimuli was present in both conditions (Fig. 4a). Similar GBAs appear at regular intervals throughout the

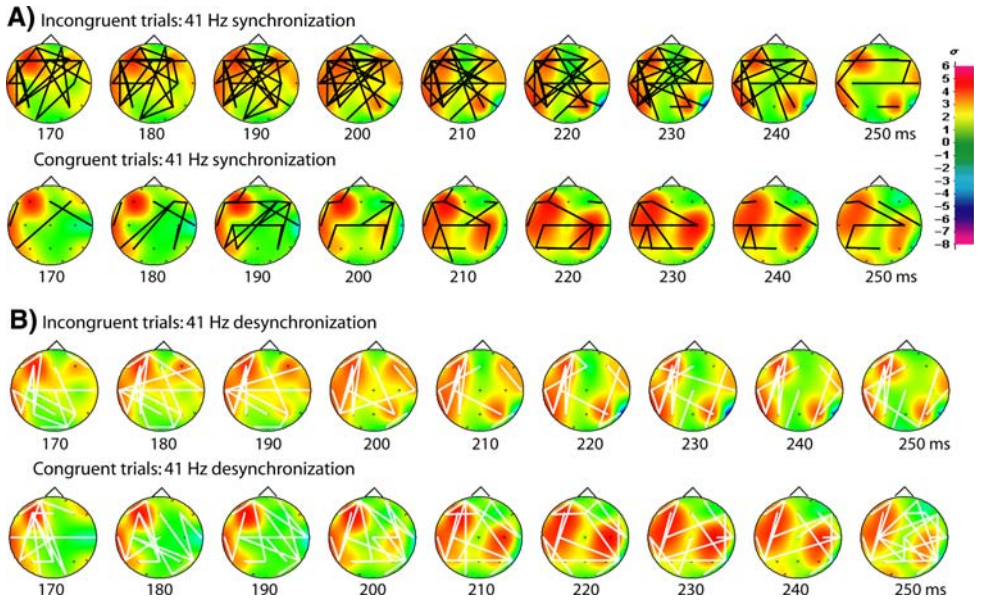

Fig. 3 **a** Topography of long-range gamma-band phase synchronization accompanying incongruent perception. *Black lines* indicate statistically significant increases in phase-locking between electrode pairs. Note the anchoring of the network at frontal and left posterior sites. *Colours* represent spline-interpolated instantaneous SCD amplitudes, expressed in standard deviations from the pre-stimulus baseline (*legend at right*). **b** Topography of long-range *desynchonization* during the burst of long-range gamma-band *synchronization*. This effect is anchored at left frontal and left posterior electrodes, and is more medial than the long-range synchronization. As this effect occurs during a time–frequency window in which synchronization is dominant (see Fig. 2a), it should be considered to be subsidiary to the synchronization shown in Fig. 3a

analyzed epoch, possibly indicating that gamma activity was periodically "refreshed" at a lower frequency, consistent with previous studies indicating a low carrier frequency mediating coordination of gamma rhythms across distant cortical areas during cognitive processing (Canolty et al. 2006; Doesburg et al. 2005; von Stein and Sarnthein 2000; Ward 2003). Congruent perception culminated in a period of enhanced GBA, relative to incongruent, from about 700 to 900 ms after stimulus onset (Fig. 4a). The scalp topography of gamma-band activity during the zenith of the long-range synchronization (and early GBA enhancement) was roughly similar across congruent and incongruent percepts, showing widespread activations over left frontal and temporal–parietal locations and more restricted activations over similar areas in the right cortex (Fig. 4b). Some differences, however, are evident between these topographies for congruent and incongruent percepts including a stronger left frontal activation during congruent percepts. Also, the topography of the right-hemispheric GBA appears to be somewhat more distributed for incongruent percepts, with distinct frontal and parietal components, whereas during congruent perception there is a single parietocentral component.

Several other notable amplitude changes were observed during the course of audiovisual speech perception, although these effects did not appear to vary with audiovisual congruence (Fig. 4a). Most notably, a broadband amplitude decrease occurred from about 8 to 30 Hz beginning about 150 ms after stimulus onset and continuing until the end of the analyzed epoch. The frequency of this effect expanded into lower frequencies at around 500 ms after stimulus onset. This amplitude decrease accompanied the decrease in synchronization in this frequency range that was noted earlier (Fig. 2a, b). Increases in instantaneous amplitude, centred at 14 and 30 Hz, also occurred from about 70 to 140 ms post stimulus onset. The increased 30 Hz activity occurred in the same time range as the long-range synchronization increase that occurred on the congruent trials, but the 14 Hz activation did not appear to co-occur with any long-distance synchronization changes.

Discussion

Our results demonstrate increased long-range gamma-band phase synchrony during the perception of temporally incongruent audiovisual speech signals 170–250 ms after stimulus onset. This pattern of synchronization is also visible during the perception of congruent audiovisual speech stimuli, but it is markedly less pronounced there than it is during incongruent perception. Decreased long-distance synchronization was also observed during the same time–frequency window. This desynchronization, however, showed a different topography, involved fewer

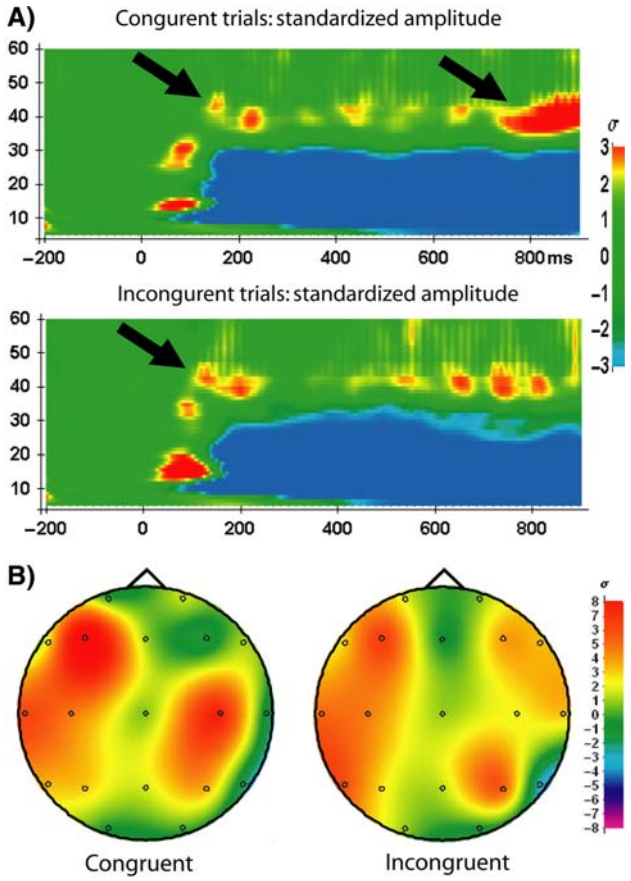

Fig. 4 **a** Global instantaneous amplitude of SCD time-locked to stimulus onset for both incongruent and congruent trials, calculated by averaging the instantaneous amplitude across all electrodes (expressed as number of standard deviations, σ, above or below the pre-stimulus baseline, *legend at right*). *Arrows* indicate the increase in gamma-band activity, apparent for both congruent and incongruent percepts, during the period where increased *long-range* synchrony was seen for incongruent perception. Also denoted by an *arrow* is a late increase in gamma-band activity that was more pronounced during congruent perception. **b** Topography of 41 Hz instantaneous amplitude at the zenith of long-range gamma-band synchronization for incongruent perception (210 ms after stimulus onset)

electrode pairs, and was much less pronounced than the observed synchronization. Previous neuroimaging research has identified a distributed network of brain regions displaying increased activation during the perception of audiovisual speech asynchrony, comprised principally of insular, prefrontal and posterior parietal regions, with a left hemisphere bias (Jones and Callan 2003; Kaiser et al. 2005; Miller and D'Esposito 2005). These regions are also activated during the detection of audiovisual asynchrony in nonspeech stimuli, indicating that they play a more general role in perceptual segregation (Bushara et al. 2001). When the network is activated by nonspeech stimuli, however, the lateralization tendencies observed during the detection of audiovisual speech asynchrony are not present. This suggests that the although neural generators associated audiovisual speech segregation are an outgrowth of more general mechanisms, it is still a distinct process. The topographical pattern of increased gamma-band synchronization observed in the present study displays both a left hemisphere bias and a prefrontal-parietal orientation (Fig. 2). This implies an anatomical similarity between the imaging activations and the increased long-range gamma-band synchrony beginning at 170 ms after onset of a perceptually incongruent stimulus. We interpret the observed increase of large-scale gamma-synchronization during incongruent percepts as indicative of the functional integration of this network of cortical areas associated with the detection of audiovisual speech synchrony. Recent evidence also gives credence to the notion that the operation of this network is associated with GBA. MEG recordings during perception of audiovisual speech incongruence have revealed significant increase of GBA beginning at about 160 ms (Kaiser et al. 2005), which is strikingly similar to the onset of gamma-synchronization reported here at 170 ms. Moreover, induced ~42 Hz activity is positively correlated with the veridical detection of auditory deviations from visual speech stimuli (Kaiser et al. 2006). Our results, together with those from the more anatomically precise measures of fMRI and PET, give a convergent account of a functionally integrated yet distributed network of brain regions that is associated with the detection of asynchronous audiovisual speech, and that achieves transient functional integration through gamma-band phase synchronization.

It is not clear what functional role is played by the long-range desynchronization also observed for incoherent percepts during the same 170–250 ms time period. Previous research has attributed such phase scattering to the functional decoupling of neural populations (Rodriguez et al. 1999). In the context of perceptual binding by synchronization (see Engel and Singer 2001), this effect could reflect neural signals indicating that (or effecting that) the two stimulus streams are not integrated into a coherent percept.

Long-range phase synchronization between neural groups provides a mechanism for increased information transfer and provides an account of how distributed networks of brain regions are able to exhibit transient enhancement of functional connectivity (Fries 2005). One criticism of this hypothesis is that synchronous activity between brain regions could, among other causes, merely reflect the temporally cohesive nature of the stimuli themselves (see Shadlen and Movshon 1999, for a review of such criticisms). The present paper uniquely reports an increase in long-range gamma-band synchrony indicating function global synchrony associated with a functionally identified network of brain regions in the absence of both temporal synchrony in the stimulus and temporal coherence of the resulting percept.

It is not clear why no induced increases in long-range gamma-band synchrony were seen during the perception of congruent speech. A distributed network of brain areas has also been identified for the perceptual fusion of audiovisual stimuli, and transient functional integration must be achieved within this network in order to perceptually fuse congruent speech stimuli (Miller and D'Esposito 2005). Previous research has linked operational synchrony in lower frequency bands, indicative of temporal coherence between processing onsets across different cortical regions, to a large-scale network underlying perceptual fusion of audiovisual speech stimuli (Fingelkurts et al. 2003). Our results show an early increase in long-distance synchronization centered at 27 Hz was seen from 40 to 70 ms after the onset of congruent stimuli. This synchronization, however, occurs much too early to account for the perceptual fusion of complex stimuli such as speech. It is consistent, however, with evoked synchronization related to multisensory perception (Callan et al. 2001; Senkowski et al. 2007). This finding, moreover, is consistent with a burgeoning body of evidence indicating that early multimodal processing is mediated by fast interactions between primary sensory regions, although the anatomical origin of this effect cannot be confirmed in our data (Schroeder and Foxe 2005).

One possible explanation for why our study produced no evidence for a distributed network responsible for the fusion of audiovisual speech is that no such network exists. Although some fMRI studies have identified multiple locations associated with bimodal speech integration (see Miller and D'Esposito 2005), others have not. Wright and colleagues (2003), for example, conducted an fMRI study to localize bimodal speech integration areas, and failed to identify any of the brain regions implicated for speech fusion by Miller and D'Esposito (2005) except for the STS/STG. Calvert et al. (2000) conducted a localization study using stricter criteria for a brain region to qualify as a speech fusion area, namely, (1) responsiveness to stimulation in each modality in isolation, (2) a superadditive response to congruent audiovisual speech (AV > A + V), and (3) a subadditive response to incongruent stimuli (AV < A + V). Only the STS met all of these criteria. Moreover, a single sweep EEG study lends further credence to this notion, as two components were found: STS activation peaking from about 150 to 300 ms after stimulus onset, and a distributed component spanning parietal, occipital, sensorimotor, temporal and prefrontal areas (Callan et al. 2001). This second component was attributed to a task-induced effect unrelated to the binding of audiovisual speech. These results are in accordance with the conclusion that the activity of the STS alone, not an anatomically distributed network of brain areas, is responsible for audiovisual speech integration (Calvert et al. 2000; Calvert 2001). If binding auditory and visual speech streams is accomplished largely within a single cortical region the absence of a long-range synchronization effect is unsurprising, as this measure indexes connectivity between multiple task-relevant regions.

An alternative interpretation of our results is that both congruent and incongruent audiovisual speech perception rely upon a common distributed network of brain areas. The topography and time–course of both long-range gamma-band synchronization effects, as well as the distribution of gamma-band current on the scalp, are similar (Figs. 3a, 4b) in the time range of 170 to 250 ms after stimulus onset. Consider, for example, that activity from a single network governing the integration of auditory and visual speech streams produces the effects reported here. Our hypothesis was that stimuli that were perceived as congruent would be accompanied by increased gamma-band synchronization as a result of the binding of stimulus features across modalities, and because of detection and functional integration within the distributed network of brain areas responsible for audiovisual speech integration. It is possible, however, that it is not whether the stimulus streams are effectively integrated, but rather the amount of processing needed to determine if the streams should be integrated, that determines the level of network activation in this task. Synchronization between neural populations relevant for the performance of a particular task is supposed to enable the effective transfer of information (Fries 2005). Adaptive resonance theory (e.g., Grossberg 1995) would predict that the process of constructing a multisensory perception involves reentrant processing wherein feedforward and feedback neural signals are exchanged between distributed brain regions until these areas have found a "match." In this view, reentrant information exchanges, mediated by neural synchrony, would continue until ambiguity about how the perception should be organized would be resolved. Accordingly, the amount of information exchange, and thus neural synchrony, should be higher in situations where there is greater uncertainty about whether or not stimulus features from temporally-shifted audiovisual streams should be grouped. This predicts that audiovisual streams that clearly correspond, or clearly do not correspond, should produce less synchronization then those for which this is a more difficult determination. A parabolic relationship would thus exist between the amount of temporal mismatch between the auditory and visual streams and the amount of synchronization between relevant brain areas. In light of such a theory, it may be that the incongruent stimuli presented in this study produced increased synchronization between brain areas because more information exchange between relevant brain areas was required to determine that they were a nonmatch than was required for the determination

that congruent stimuli should be grouped. This interpretation is given some credence by our behavioural data, as incongruent speech stimuli (89.8% correct) were correctly identified less frequently than were congruent stimuli (96.1% correct). Further studies, however, with consistent results over more conditions of audiovisual temporal incongruence, would be needed before it could be concluded that this explanation is to be preferred to the idea that the STS alone mediates audiovisual speech fusion.

Conclusion

We present evidence that a large-scale network of brain areas responsible for detection of temporal mismatch between audio and visual streams of speech stimuli achieves transient functional integration through gamma-band phase-synchronization. Our results provide a striking anatomical and time–course corroboration of previous findings. This work addresses a major theoretical hurdle regarding the mechanism for transient functional integration and provides an important dissociation between phase-synchronization and stimulus coherence, highlighting the functional nature of network-based synchronization.

References

Bushara KO, Grafman J, Hallet M (2001) Neural correlates of auditory-visual stimulus onset asynchrony. J Neurosci 21(1):300–304

Callan DE, Callan AM, Kroos C, Vatikiotis-Bateson E (2001) Multimodal contribution to speech perception revealed by independent component analysis: a single-sweep EEG case study. Cog Brain Res 10:349–353

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

Calvert GA (2001) Crossmodal processing in the human brain: insights from functional neuroimaging studies. Cer Cortex 11:1110–1123

Canolty RT, Edwards E, Dalal SS, Soltani M, Nagarajan SS, Kirsh HE, Berger MS, Barbaro NM, Knight RT (2006) High gamma power is phase-locked to theta oscillations in the human neocortex. Science 313:1626–1628

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

Dixon NF, Spitz L (1980) The detection of auditory visual desynchrony. Perception 9:719–721

Doesburg SM, Kitajo K, Ward LM (2005) Increased gamma-band synchrony precedes switching of conscious perceptual objects in binocular rivalry. Neuroreport 2:229–239

Eggermont JJ (2000) Sound-induced synchronization of neural activity between and within three cortical areas. J Neurophysiol 83:2708–2722

Engel AK, Singer W (2001) Temporal binging and the neural correlates of sensory awareness. Trends Cogn Sci 5:16–25

Fingelkurts AA, Fingelkurts AA, Krause CM, Möttönen R, Sams M (2003) Cortical operational synchrony during audio–visual speech integration. Brain Lang 85:297–312

Ford JM, Gray M, Faustman WO, Heinks, Mathalon DH (2005) Reduced gamma-band coherence to distorted feedback during speech when what you say is not what you hear. Int J Psychophysiol 57:143–150

Freeman WJ (2004) Origin, structure, and role of background EEG activity. Part 1. Analytic amplitude. Clin Neurophysiol 115:2077–2088

Fries P (2005) A mechanism for cognitive dynamics: neuronal communication through neuronal coherence. Trends Cogn Sci 9:474–479

Grossberg S (1995) The attentive brain. Am Sci 83:483–449

Jones JA, Callan DE (2003) Brain activity during audiovisual speech perception: an fMRI study of the McGurk effect. Neuroreport 14(8):1129–1133

Kaiser J, Hertrich I, Ackermann H, Lutzenberger W (2006) Gamma-band activity over early sensory areas predicts detection of changes in audiovisual speech stimuli. Neuroimage 15:646–653

Kaiser J, Hertrich I, Ackermann H, Mathiak K, Lutzenberger W (2005) Hearing lips: gamma-band activity during audiovisual speech perception. Cer Cortex 15:646–653

Kanayama N, Sato A, Ohira H (2007) Crossmodal effect with rubber hand illusion and gamma-band activity. Psychophysiol 44: 392–402

King AJ (2005) Multisensory integration: strategies for synchronization. Curr Biol 15(9):R339–R341

Lachaux JP, Rodriguez E, Martinerie J, Varela FJ (1999) Measuring phase synchrony in brain signals. Hum Brain Mapp 8(4):94–208

Lewkowicz DJ (1996) Perception of auditory–visual temporal synchrony in human infants. J Exp Psychol: Hum Percept Perform 22(5):1194–1106

McGurk H, MacDonald J (1975) Hearing lips and seeing voices. Nature 264:746–748

Miller LM, D'Esposito M (2005) Perceptual fusion and stimulus coincidence in the cross-modal integration of speech. J Neurosci 25(25):5884–5893

Perrin F, Bertrand O, Pernier J (1987) Scalp current density mapping: value and estimation from potential data. IEEE Trans Biomed Eng BME 34:283–288

Pikovski A, Rosenblum M, Kurths J (2001) Synchronization: a universal concept in nonlinear sciences. Cambridge University Press, Cambridge 432 p

Rodriguez E, George N, Lachaux JP, Martinirie J, Renault B, Varela FJ (1999) Perception's shadow: long-distance synchronization of human brain activity. Nature 397:430–433

Schroeder E, Foxe J (2005) Multisensory contributions to low-level "unisensory" processing. Curr Opin Neurobiol 6(4):285–296

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

Senkowski D, Talsma D, Herrmann CS, Woldorff M (2005) Multisensory processing and oscillatory gamma responses: effects of spatial selective attention. Exp Brain Res 166:411–426

Shadlen MN, Movshon JA (1999) Synchrony unbound: a critical evaluation of the temporal binding hypothesis. Neuron 24:67–77

Tallon-Baudry C, Bertrand O (1999) Oscillatory gamma activity in human and its role in object representation. Trends Cogn Sci 3(4):151–162

Varela F, Lachaux JP, Rodriguez E, Martinerie J (2001) The brainweb: phase synchronization and large-scale integration. Nat Rev Neurosci 2(4):229–239

von Stein A, Sarnthein J (2000) Different frequencies for different scales of cortical integration: from local gamma to long range alpha/theta. Int J Psychophysiol 38(3):301–313

Ward LM (2003) Synchronous neural oscillations and cognitive processes. Trends Cogn Sci 17:553–559

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

Yuval-Greenberg S, Deouell LY (2007) What you see is not (always) what you hear: induced gamma band responses reflect cross-modal interactions in familiar object recognition. J Neurosci 27(5):1090–1096