Published in final edited form as:

Neuroimage. 2016 January 1; 124(0 0): 858–868. doi:10.1016/j.neuroimage.2015.09.044.

Interacting parallel pathways associate sounds with visual identity in auditory cortices

Jyrki Ahveninen¹, Samantha Huang¹, Seppo P. Ahlfors¹, Matti Hämäläinen^{1,2,3}, Stephanie Rossi¹, Mikko Sams³, and Iiro P. Jääskeläinen³

¹ Athinoula A. Martinos Center for Biomedical Imaging, Department of Radiology, Massachusetts General Hospital/Harvard Medical School, Charlestown, MA, USA ² Harvard-MIT Division of Health Sciences and Technology, Cambridge, MA, USA ³ Brain and Mind Laboratory, Department of Neuroscience and Biomedical Engineering, Aalto University School of Science, Espoo, Finland

Abstract

Spatial and non-spatial information of sound events is presumably processed in parallel auditory cortex (AC) “what” and “where” streams, which are modulated by inputs from the respective visual-cortex subsystems. How these parallel processes are integrated to perceptual objects that remain stable across time and the source agent's movements is unknown. We recorded magneto- and electroencephalography (MEG/EEG) data while subjects viewed animated video clips featuring two audiovisual objects, a black and a gray cat. Adaptor-probe events were either linked to the same object (the black cat meowed twice in a row in the same location) or included a visually conveyed identity change (the black and then the gray cat meowed with identical voices in the same location). In addition to effects in visual (including fusiform, middle temporal or MT areas) and frontoparietal association areas, the visually conveyed object-identity change was associated with a release from adaptation of early (50–150 ms) activity in posterior ACs, spreading to left anterior ACs at 250–450 ms in our combined MEG/EEG source estimates. Repetition of events belonging to the same object resulted in increased theta-band (4–8 Hz) synchronization within the “what” and “where” pathways (e.g., between anterior AC and fusiform areas). In contrast, the visually conveyed identity changes resulted in distributed synchronization at higher frequencies (alpha and beta bands, 8–32 Hz) across different auditory, visual, and association areas. The results suggest that sound events become initially linked to perceptual objects in posterior AC, followed by modulations of representations in anterior AC. Hierarchical what and where pathways seem to operate in parallel after repeating audiovisual associations, whereas the resetting of such associations engages a distributed network across auditory, visual, and multisensory areas.

Corresponding Author: Jyrki Ahveninen, Ph.D., MGH/MIT/HMS-Martinos Center, Bldg. 149 13th Street, Charlestown MA 02129, phone (617) 726-6584; fax (617) 726-7422, jyrki@nmr.mgh.harvard.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

1. Introduction

Perceptual objects refer to constructs that associate individual events with a specific source agent, and integrate them into entities that remain stable across time and the agent's movements in space (Bizley and Cohen, 2013). Indeed, one readily perceives dynamic information from different senses as belonging to a single entity, for example, when a cat moves about and meows as it goes along, rather than seeing and hearing a series of disparate events. In contrast to this everyday experience, a prominent organizational principle of sensory systems is parallel processing where information of different stimulus attributes is processed in separate sensory channels and feature pathways. The human visual pathway is believed to branch into two major streams that process spatial (“where”) and object-identity related (“what”) information (Ungerleider and Haxby, 1994). Many studies (Ahveninen et al., 2013; Ahveninen et al., 2006; Altmann et al., 2007; Clarke et al., 2002; Lomber and Malhotra, 2008; Rauschecker and Scott, 2009; Rauschecker and Tian, 2000) support the existence of an analogous division between parallel posterior “where” and anterior “what” pathways in auditory cortices (AC) as well (however, for alternative AC models, see also Bizley and Cohen, 2013; Griffiths and Warren, 2002; Recanzone and Cohen, 2010). What is still unknown is where and at which latencies spatial and identity-related feature information of auditory and visual events becomes linked to objects that remain perceptually constant despite their fluctuations across time and space.

The prevailing view has been that integration of information across feature pathways and sensory modalities occurs through hierarchical convergence of parallel pathways at higher-level association areas (Konorski, 1967). Previous studies suggest that the parallel “where” and “what” auditory pathways converge with their visual counterparts in dorsal/superior frontal (frontal eye fields, FEF; dorsolateral prefrontal cortex, DLPFC) vs. inferior frontal cortex (IFC) areas, respectively (Alain et al., 2001; Altmann et al., 2012; Arnott et al., 2005; Barrett and Hall, 2006; Clarke et al., 2002; Leung and Alain, 2011; Maeder et al., 2001; Romanski et al., 1999; Weeks et al., 1999). Furthermore, in humans there is abundant evidence of super-additive (or sub-additive) multisensory interactions in areas such as the posterior superior temporal sulcus (pSTS) and middle temporal gyrus (MTG) (Beauchamp et al., 2004; Bischoff et al., 2007; Calvert et al., 2000; Raji et al., 2000; Werner and Noppeney, 2010), which may also include neurons contributing to face and voice identity integration (von Kriegstein et al., 2005). The existence of a slow higher-order feature-attribute and crossmodal binding mechanism is also supported by psychophysical evidence of temporal judgments (Fujisaki and Nishida, 2010). The hierarchical convergence model is confronted by increasing evidence of interactions across sensory areas during multisensory stimulus processing (Senkowski et al., 2008). Indeed, even primary AC neurons can be crossmodally modulated (Bizley et al., 2007; Budinger et al., 2006; Pekkola et al., 2005; Raji et al., 2010), and these influences become progressively stronger in non-primary areas (Calvert et al., 1997; Kayser et al., 2005; Romanski et al., 1999; Smiley et al., 2007).

In contrast to multisensory areas such as pSTS and MTG that are activated by a variety of object-related events (Beauchamp et al., 2004), modulations of “unimodal” AC areas have been mostly studied with visual cues related to sound production (Calvert et al., 1997; Jääskeläinen et al., 2004b; Pekkola et al., 2005; van Wassenhove et al., 2005). These cues

have, thus, predicted spectrotemporal (McGurk and MacDonald, 1976) or spatial (Bonath et al., 2007) properties of isolated events, rather than tapping into perceptual objects that remain constant despite their fluctuations across time and space. Interestingly, indices of a specific role of AC areas in object processing were found in a single-unit study in rhesus monkeys, which documented a small number of neurons that responded to specific voice types coming from a particular direction in posterior non-primary ACs (Tian et al., 2001). Although most of the neurons sampled in this study were predominantly spatially selective, the linking of feature representations to more complex perceptual entities could be initiated in this area. Since response latencies were not examined in this pioneering monkey study, it remains to be determined whether such an effect occurs at early latencies driven by bottom-up processing, or whether it results from later feedback signals from higher-order cortical regions. Further, whether these neurons are sensitive to corresponding inputs from other sensory modalities was not tested.

The goal of the present study is to examine non-invasively in humans in which anatomical areas and at which latencies information from parallel auditory pathways becomes associated with the perceptual object's stable visual identity, beyond cues related to voice/sound production itself. Functional properties of AC neurons have been often studied by measuring neuronal adaptation, or repetition suppression, of neuronal responses to recurring stimulus attributes (Jääskeläinen et al., 2011). In populations sensitive to the feature of interest, a release from adaptation is expected when the consecutive “adaptor” and “probe” stimuli differ from one another with respect to this attribute (Ahveninen et al., 2006; Altmann et al., 2007). In ACs, such effects can be measured by using the MEG/EEG response N1, which shows adaptation effects with, for example, speech-related visual stimuli (Jääskeläinen et al., 2004b). Analogously, one might also test whether two consecutive stimuli are related to the same multisensory object or not: a release from adaptation is expected after two events related to different objects (e.g., a cat of different color but similar voice).

Accumulating evidence further suggests that crossmodal influences on AC function are mediated by inter-regional synchronization of neuronal oscillations (for a review, see Senkowski et al., 2008). Whereas local integration effects may occur at the high-frequency gamma band (~30–100 Hz) (for a review, see Jensen et al., 2007), longer-range coupling might be supported by lower frequencies at which neurons have more robust spike timing delays (Engel et al., 2001; Ermentrout and Kopell, 1998; Roelfsema et al., 1997). Non-invasive evidence of crossmodal oscillatory mechanisms has been obtained using both EEG (Doesburg et al., 2008; Hipp et al., 2011; van Driel et al., 2014; von Stein et al., 1999) and MEG (Alho et al., 2014) in humans. Long-range synchronization at the 2–4 Hz delta and 4–8 Hz theta bands could help provide low-level timing information of the speech rhythm from visual system to AC neurons (Arnal et al., 2011; Luo et al., 2010; Schroeder et al., 2008). Oscillatory phase synchronization at the alpha (8–16 Hz) (van Driel et al., 2014) and gamma (30–100 Hz) (Doesburg et al., 2008) frequencies may also play a role in temporal integration of audiovisual inputs during multisensory object formation. Therefore, it is conceivable that coordinated oscillatory interactions across ACs and visual areas also contribute to the formation of perceptual objects.

Here, we specifically hypothesized that regions of AC will show release from adaptation when a visual cue changes the object identity, indicative of a presence of neuron groups within AC that are associated with perceptual objects. We further hypothesized that the activity in such AC regions will exhibit oscillatory functional connectivity with other brain areas processing auditory and visual object-feature and spatial information, and that the connectivity patterns will be different for repeating vs. changing audiovisual object associations. To test these hypotheses, we used an adaptation paradigm that taps into the formation of non-speech related audio-visual perceptual objects. Stimulus-related and interregional oscillatory processes during audiovisual processing were estimated using a multimodal cortically constrained MEG/EEG source modeling approach (Huang et al., 2014; Lin et al., 2006; Sharon et al., 2007). In this approach, the MEG and EEG source locations are restricted to the cortical mantle derived from anatomical MRI to reduce the potential solution space (Dale and Sereno, 1993). Additional improvements are achieved by combining the complementary information provided by simultaneously measured MEG and EEG, which helps provide better accuracy and smaller point spread of the source estimates than either modality alone (Ding and Yuan, 2013; Henson et al., 2009; Liu et al., 2002; Sharon et al., 2007).

2. Materials and Methods

2.1. Subjects, task and stimuli

Eleven subjects (age 21–50 years, 5 females) were studied. One subject of the initial $N=12$ was excluded due to excessive blink artifacts. The subjects had normal hearing and normal or corrected-to-normal vision. Human subjects' approval was obtained and voluntary consents were signed before each measurement, in accordance with the experimental protocol approved by the Massachusetts General Hospital Institutional Review Board.

During MEG/EEG acquisition, the subjects watched 3.1-s animated movie clips consisting of audiovisual Adaptor/Probe stimulus pairs (**Fig. 1a**). In each clip, cartoon faces of a gray and a black cat first appeared at the screen 5° to the right and left from a central fixation mark. Each cat subtended $\sim 1.7^\circ \times 1.7^\circ$. At $t_1 = 1$ s, the black cat meowed (Adaptor stimulus, meow sound and a visual presentation of the opening of the mouth, duration 450 ms). After that both cats moved to the center and returned to either their original places or switched sides. At $t_2 = 2.4$ s, the cat that had ended up in the black cat's initial position uttered a second meow (Probe stimulus, duration 450 ms), which sounded exactly the same as the Adaptor meow. Thus, 50% of the trials included a repetition of audiovisual events linked to the same object (AV_{Repeat}): the black cat meowed twice in a row in the same place. The other 50% of the trials included a *visual identity change* ($AV_{Repeat}V_{Change}$), that is, the black and gray cat meowed in a row with exactly the same voice in the same location. The meow sound stimulus (**Fig. 1b**), modified from a public domain recording, was presented at a clearly audible and comfortable level, being always simulated from the same direction as the corresponding visual stimulus using generic head-related transfer functions (Algazi et al., 2001). The position of the meowing cat was counterbalanced across the subjects ($N_{Right}=6$, $N_{Left}=5$) and controlled as a between-subjects nuisance factor in our statistical analyses. The

movie clips were separated by 3 s and presented in a random order, the interval between trial onsets was 6.1 s. A total of 288 trials were presented in two runs.

The subjects were asked to look at the fixation mark, and to press a button with one hand if the cats switched sides and another button with the other hand if the cats returned to their initial places. The subjects were asked to emphasize accuracy over speed, and to press the button only after the cats had disappeared from the screen. The order of responding hand was counterbalanced across subjects and controlled as a between-subject nuisance factor in statistical models. Cognitively, the purpose of this task was, on one hand, to confirm that the subject viewed the stimuli, but without knowing that the primary effect of interest was related to the emergence of an association between the visual and auditory stimulus. Although attention modulates audiovisual integration in complex settings (Fujisaki et al., 2006; van Ee et al., 2009), binding sounds to visual stimuli occurs normally with no conscious effort (De Meo et al., 2015). On the other hand, by keeping the subjects naïve about the experiment's goals we intended to avoid any circumstantial modulations of sound processing due to the subject attempting to find acoustical differences between Adaptors and Probes. In an additional control experiment conducted in a subpopulation of six subjects of the original sample, the movie clips were presented with no sound (144 trials in one run).

2.2. MEG/EEG and MRI data acquisition, processing, and source estimation

Simultaneous 306-channel MEG and 70-channel EEG data were recorded in a magnetically shielded room at 0.03–330 Hz, 1000 samples/s (Neuromag Vectorview system, Elekta-Neuromag, Helsinki, Finland). An MEG compatible EEG electrode cap was utilized (EasyCap, Herrsching, Germany). The head position relative to the MEG sensor array was monitored continuously using four head-position indicator (HPI) coils attached to the scalp. Electro-oculogram (EOG) was recorded to monitor eye artifacts. In off-line analysis, external MEG noise was suppressed and subject movements estimated at 200-ms intervals were compensated for using the signal-space separation method (Taulu et al., 2005). After downsampling (0.5–110 Hz, 333 samples/s), MEG/EEG epochs coinciding with over 150 μ V EOG, 100 μ V EEG, 2 pT/cm MEG gradiometer, or 5 pT MEG magnetometer peak-to-peak signals were excluded from further analyses. The average reference was used in all analyses of the EEG data. T1-weighted structural MRIs were obtained for combining anatomical and functional data using a multi-echo MPRAGE pulse sequence (TR=2510 ms; 4 echoes with TEs = 1.64 ms, 3.5 ms, 5.36 ms, 7.22 ms; 176 sagittal slices with $1 \times 1 \times 1$ mm³ voxels, 256×256 mm² matrix; flip angle = 7°) using 3T Siemens TimTrio scanner (Siemens, Erlangen, Germany). Multi-echo FLASH data were obtained with 5° and 30° flip angles (TR = 20 ms, TE = 1.89 + 2n ms [n = 0–7], 256×256 mm² matrix, 1.33-mm slice thickness) for reconstruction of head boundary-element models (BEM).

To calculate cortically-constrained ℓ_2 minimum-norm estimates (MNE) (Hämäläinen et al., 1993; Lin et al., 2006), the information from MRI structural segmentation and the MEG sensor and EEG electrode locations were used to compute the forward solutions for all putative source locations in the cortex using a three-compartment BEM (Hämäläinen et al., 1993). The Freesurfer software was used to determine the cortical surface as well as the boundaries separating the scalp, skull, and brain compartments from the anatomical MRI

data (<http://surfer.nmr.mgh.harvard.edu/>). The forward solution for individual subjects were computed for current dipoles placed on the cortical surfaces at ~5,000 locations per hemisphere. The noise covariance matrix was estimated from the raw MEG/EEG data using a 200-ms pre-stimulus baseline. For connectivity analyses, noise covariance matrices were obtained from a separate two-minute resting-state run, during which the subjects were asked to remain still and look at a fixation mark. The MEG/EEG data at each time point were multiplied by the MNE inverse operator and noise normalized to yield the estimated source activity, as a function of time, on the cortical surface (Lin et al., 2006). For group-level analysis, the cortical source estimates were normalized into a spherical standard brain representation (Fischl et al., 1999).

2.3. Analysis of stimulus-evoked activity

In the whole-cortex mapping, individual subjects' noise-normalized MNEs (dSPM) of the difference responses between the $A_{\text{Repeat}}V_{\text{Change}}$ and AV_{Repeat} conditions were averaged over time windows, to mitigate the inter-subject variability of regional time courses and to reduce the number of statistical comparisons (see, e.g., Ahveninen et al., 2011; Papanicolaou et al., 2006; Raji et al., 2000; Todorovic et al., 2011). The data from four consecutive 100-ms windows (50–150 ms encompassing the rising and falling phases of the N1 activity, 150–250, 250–350, and 350–450 ms after the Probe event onset) were analyzed using a random-effects model, which included the stimulus and response laterality as between-subjects nuisance factors and which was weighted by each subject's number of accepted trials. To control for multiple comparisons, the resulting statistical estimates were tested against an empirical null distribution of maximum cluster size across 10,000 iterations with a vertex-wise threshold of $P < 0.05$ and cluster-forming threshold of $P < 0.05$, yielding clusters corrected for multiple comparisons across the surface.

For region-of-interest (ROI) analyses, the superior temporal gyrus (STG) label of the Freesurfer Desikan-Killiany (Desikan et al., 2006) atlas, which encompasses non-primary ACs, was divided to anterior (aSTG) and posterior (pSTG) portions according to the corresponding divisions of the FSL Harvard-Oxford anatomical atlas (<http://fsl.fmrib.ox.ac.uk/fsl/fslwiki/Atlases>). The ROI time courses were obtained for each subject and ROI in the $A_{\text{Repeat}}V_{\text{Change}}$ and AV_{Repeat} conditions by averaging the dSPM waveforms for each dipole within the ROI, and by using a 15-ms sliding time integration. The time courses were analyzed using cluster-based randomization tests of the Fieldtrip Matlab toolbox (Maris and Oostenveld, 2007). We also tested our specific a priori hypothesis concerning N1 adaptation in the four ROIs at 50–150 ms (Ahveninen et al., 2011) using Bonferroni-corrected weighted ANOVA models that were computed separately for the aSTG and pSTG regions, with the stimulus laterality and responding hand as between-subject nuisance factors and the hemisphere as a within-subject factor.

2.4. Analysis of functional connectivity

To examine modulations of functional connectivity, seed-based phase locking estimates (Huang et al., 2014) were calculated between all cortical locations and the aSTG and pSTG seed regions using MNE-Python tools (Gramfort et al., 2013). A fast Fourier transform with a 700 ms Hanning window was applied on epochs centered at 350 ms after the Probe

stimulus onset. A three-octave frequency band of 4–32 Hz was selected based on previous studies suggesting that lower-frequency oscillations might specifically contribute to the longer-range coupling that supports distributed networks (Kopell et al., 2000). At the smallest actual frequency bin of 4.3 Hz within this range of interest the 700-ms analysis window encompassed three full cycles. Phase locking was estimated using the weighted phase lag index (wPLI), which helps avoid spurious inflation of synchronization indices due to EEG and MEG point spread and crosstalk (Vinck et al., 2011). Detailed comparisons between task conditions were conducted within three consecutive one-octave bands, which corresponded to the theta (4–8 Hz), alpha (8–16 Hz), and beta (16–32 Hz) frequency ranges. Similarly to the statistical analysis of the whole-cortex evoked activity described above, significant differences in the connectivity values were determined using a random-effects model and cluster-based randomization tests.

3. Results

During simultaneous MEG/EEG measurements, subjects viewed animated video clips featuring two audiovisual perceptual objects, a black and a gray cat (**Fig. 1**). Paired Adaptor-Probe events were either linked to the same object (the black cat meowed twice in a row in the same location; AV_{Repeat}) or included a visual identity change (the black and then the gray cat meowed with identical voices in the same location; $A_{Repeat}V_{Change}$). Behaviorally, all subjects were able to follow the instructions of discriminating whether or not the cats switched positions (group-average hit rate 99%).

3.1. Stimulus-evoked activity patterns

Consistent with our hypotheses, MEG/EEG source estimates for the cortical activity evoked by the Probe stimuli suggested a significant modulation of N1 adaptation in posterior STG (pSTG) areas, followed by a modulation in the left anterior STG (aSTG), due to visually conveyed identity change of the audiovisual object (**Fig. 2**).

In whole-cortex mapping analyses (**Fig. 2a**), significant activation differences between the $A_{Repeat}V_{Change}$ and AV_{Repeat} conditions were observed in left superior temporal auditory areas at 250–350 and 350–450 ms after the Probe stimulus onset ($p < 0.05$, cluster-based simulation tests). The significant clusters were centered at the left aSTG and planum polare (PP), with the maxima at the Montreal Neurological Institute (MNI) coordinate locations $MNI_{x,y,z} = [-50, 6, -15]$ at 250–350 ms and at $[-42, 0, -24]$ at 350–450 ms. In addition, significant effects were observed near right ACs at 50–150 and 150–250 ms, with the clusters centered at the right supramarginal cortex (~the temporo-parietal junction, TPJ) and insula, respectively. Evidence of modulations of visual areas was found in the right fusiform area (50–150, 150–250 ms), as well as in a right occipitotemporal / inferior parietal cluster overlapping the likely location of the human area MT (150–250 ms). Significant differences were also seen in the left paracentral (50–150 ms) and right superior frontal cortices (250–350 ms).

The signal polarity for the contrast between the $A_{Repeat}V_{Change}$ and AV_{Repeat} conditions suggested enhanced cortical surface negativity in the left AC, which extended from Heschl's gyrus (HG; 250–350 ms) to aSTG and PP (250–350 and 350–450 ms) (**Fig. 2b**). A ROI time

course analysis showed prominent negative deflections relative to the pre-stimulus baseline in both conditions (**Fig. 2c**). This confirmed that the significant negative contrast in left ACs reflects enhanced inward currents, consistent with previously described feedback effects in AC (Ahveninen et al., 2011), instead of decreases of surface positivities after visual identity changes. In the ROI analyses, significant increases of negativity after the $A_{\text{Repeat}}V_{\text{Change}}$ vs. AV_{Repeat} conditions were found in the right pSTG during the ascending phase of N1 activity ($p < 0.05$, cluster-based randomization test), and subsequently in the left aSTG (two significant time clusters, $p < 0.05$ and $p < 0.01$). The aSTG effects coincided with the 250–350 and 350–450 ms time windows of the whole-cortex mapping results (**Figs. 2a, 2b**). An analysis of N1 amplitude from an average over 50–150 ms window, finally, showed a significantly larger amplitude during $A_{\text{Repeat}}V_{\text{Change}}$ than in the AV_{Repeat} condition in the pSTG ROIs ($F_{1,7} = 19.5$, $p < 0.01$, Bonferroni corrected factorial ANOVA) (**Fig. 2d**).

An ROI analysis was also conducted on the control data obtained using silent movie clips (**Fig. 3**). In aSTGs, the time courses were similar across the two conditions (**Fig. 3b**). In pSTG ROIs, a slow positive trend was observed during the AV_{Repeat} condition, and the difference was significant at 250–310 ms for the left pSTG and 160–210 ms for the right pSTG ($p < 0.05$, cluster-based randomization test). Notably, these effects did not coincide with the significant effects observed in the main experiment. Furthermore, the comparisons of ROI amplitudes within the 50–150 ms time window showed no significant N1 differences. The Estimated Marginal Means ($\pm$ standard errors) of dSPM amplitudes were 0.17 ± 0.10 for $A_{\text{Repeat}}V_{\text{Change}}$ and 0.36 ± 0.09 for AV_{Repeat} in pSTGs, and 0.03 ± 0.03 for $A_{\text{Repeat}}V_{\text{Change}}$ and 0.09 ± 0.05 for AV_{Repeat} in aSTGs. An additional control analysis that compared responses to the audiovisual and silent Adaptor stimuli is shown in **Figure 3c**. In the Adaptor responses, traces of N1-like responses that have slightly longer peak latencies than for the comparable audiovisual events were observable even in the silent condition.

3.2. Functional connectivity

Cortico-cortical phase locking between the four AC ROIs and the rest of the cerebral cortex was compared between the two conditions 0–700 ms after the Probe onset at the theta (4–8 Hz), alpha (8–16 Hz), and beta (16–32 Hz) frequency ranges (**Fig. 4**). Repetition of events linked to the same object (*i.e.*, $AV_{\text{Repeat}} > A_{\text{Repeat}}V_{\text{Change}}$) increased synchronization primarily at the theta range (6 out of 8 significant clusters, see **Fig. 4b** and **Table 1**; cluster-based simulation tests, $p < 0.05$). The majority of the significant connectivity clusters were limited within the respective anterior/ventral “what” and posterior/dorsal “where” pathways. Specifically, the left and right aSTG seeds (auditory “what” stream) showed increased theta-band weighted phase-lag index (wPLI) (Vinck et al., 2011) with right fusiform/inferior temporal cortices, which belong to the ventral visual “what” stream (Ungerleider and Haxby, 1994) and are also reportedly activated by animal faces (Tong et al., 2000). Theta-band wPLI also increased between the left aSTG seed and a cluster centered at the right orbitofrontal cortex, an area believed to provide feedback to inferior temporal areas during visual object recognition (Bar et al., 2006). For the left pSTG seed (auditory “where” stream) increased beta-band synchronization was seen with a left dorsal/superior frontal cluster overlapping with the left FEF, which is a known visual and auditory spatial processing site (for a review, see Rauschecker and Tian, 2000). For the right pSTG seed,

increased theta connectivity was observed with left superior parietal cortex areas, which are involved with auditory and visual spatial cognition. In the $AV_{Repeat} > A_{Repeat}V_{Change}$ comparison, the only connections across the putative “what” and “where” domains occurred between the left pSTG seed and left ventral occipital and inferior temporal cortices at the alpha band. There was also a significant theta-band cluster that overlapped the left pSTG seed region itself, probably explainable by the average phase of the seed as a whole being slightly different from the phase at this specific location (*i.e.*, a non-zero phase lag).

In contrast, visually conveyed identity changes ($A_{Repeat}V_{Change} > AV_{Repeat}$) increased synchronization in distributed networks between auditory, visual, and multisensory areas. The majority of significant clusters occurred at the beta and alpha ranges (8 and 6, respectively, out of 15 significant clusters, see **Fig. 4c, Table 2**; cluster-based simulation tests, $p < 0.05$), and the most common pattern was a connectivity increase between the “what” and “where” pathways (6 out of 15 clusters). Specifically, the left aSTG (~auditory “what”) seed showed increased synchrony with right frontal “where” areas including the right FEF (beta range) and right dorsolateral prefrontal cortex (DLPFC, alpha range); The right aSTG increased synchronization with right posterior ACs (beta range) and right posterior parietal “where” areas including the intraparietal sulcus (IPS, alpha range); The left pSTG (~auditory “where”) seed showed increased connectivity with a cluster that extended to left inferior / orbital frontal areas that presumably belong to the “what” stream; The right pSTG showed increased connectivity with inferior frontal “what” regions of both hemispheres (alpha range).

Feature-domain specific connectivity increases were observed in four out of the 15 significant clusters (**Fig. 4c, Table 2**): The left aSTG seed was increasingly connected with the corresponding right-hemispheric anterior AC area (beta range), with visual “what” areas of the right lingual gyrus (alpha range), and also with a right-hemispheric alpha cluster that also encompassed parts of the inferior frontal cortex (auditory and visual “what” streams); The right aSTG showed increase connectivity with a cluster that extended from the left inferior temporal areas (visual “what”) to left anterior ACs (beta range). Increased connectivity patterns were also observed in clusters involving the known multisensory areas: The right aSTG seed increased connectivity with two clusters overlapping with the right MTG and pSTS (alpha, beta ranges) and also with the left anterior MTG/STS.

4. Discussion

Here, we studied in which cortical areas and at which latencies auditory and visual information of an event's identity and location become associated with integrated perceptual object representations. We addressed this question by using cortically constrained MEG/EEG source estimates to compare activation and connectivity patterns during an audiovisual object-tracking task, consisting of trials with the same sound stimulus being either repeatedly associated with the same object or linked to two different visually conveyed object identities. The visual identity changes increased activations in visual cortices (fusiform cortex, MT), as well as in association areas potentially involved in audiovisual integration (paracentral region/SMA, supramarginal, inferior parietal / STS). The shortest-latency effects on sound processing triggered by visual changes in object

identity were observed in posterior AC areas 50–150 ms from the event onset (N1 repetition suppression effect), followed by robust modulations of left anterior AC areas at 250–450 ms (enhanced cortical inward source currents). Importantly, these effects took place despite the auditory stimulus being exactly the same, thus disclosing AC processes that could be associated with representations of perceptual objects. Connectivity analysis revealed distinct hierarchical/modular vs. distributed network patterns during object processing, depending on the previous stimulation history and situational requirements. Repetition of audiovisual events related to the same object increased synchrony at the lower-frequency theta band in a network of modular connectivity patterns (e.g., from anterior ACs to ventral visual pathways), whereas visually induced changes to the object identity engaged a broadly distributed network across auditory, visual, and multisensory cortices predominantly in the alpha and beta bands.

AC neuron populations are prone to stimulus-specific adaptation that suppresses their responsiveness to repetitive sounds (May et al., 1999), whereas a release from adaptation is observed when a probe stimulus includes a change relative to the preceding adaptor stimulus. Adaptation designs may thus more accurately reveal the specificity of population responses than recording responses to two classes separately (Jääskeläinen et al., 2011), which has been the case in many previous multisensory integration studies. Here, the visually conveyed identity changes after the Probe stimuli otherwise identical to the Adaptor stimuli resulted in significant N1 enhancements, as evidenced in the ascending N1 phase in the right pSTG time course and in bilateral pSTGs in the 50–150 ms time averages. These effects could be viewed as a change response (Jääskeläinen et al., 2011), signaling that the underlying AC neuron populations associate the Adaptor and Probe events with distinct objects although the sound content remained the same. Tentatively, this early posterior AC effect could also reflect initial fast and coarse processing of both “what” and “where” aspects of the perceptual object in the posterior “where” processing pathway. Indeed, a previous single-unit study in rhesus monkeys documented a small number of posterior AC neurons that were sensitive to the type of monkey call coming from a specific spatial location, in contrast to the majority of sampled units that were sensitive to spatial changes only (Tian et al., 2001).

The most prominent AC modulations were observed in left anterior non-primary areas at 250–450 ms after Probe onsets, demonstrated as a sustained enhancement of cortical surface negativity. This pattern is consistent with previous studies associating processing of object-related sounds with anterior superior temporal areas (Ahveninen et al., 2013; Rauschecker and Tian, 2000), which may also contain neurons sensitive to voice identity (Belin and Zatorre, 2003; Petkov et al., 2008) and its crossmodal congruency with concurrent visual information (Perrodin et al., 2014). The present results are also consistent with a recent EEG/fMRI study that suggested a ventriloquism-effect related event-related potential (ERP) component at 260 ms (Bonath et al., 2007). This component was estimated to originate in PT, close to the longer-latency activity in left HG seen in our cortically constrained combined MEG/EEG estimates, which may provide more accuracy than unimodal EEG source localization (Sharon et al., 2007). Our study also utilized more “naturalistic” audiovisual design, which targeted perceptual objects that span across time and the source agent's movements in space, instead of isolated feature events.

MEG/EEG signals result primarily from post-synaptic currents in the dendrites of cortical pyramidal cells (Hämäläinen et al., 1993). The direction of the source current presumably depends on the type and the dendritic location of the synaptic input (Allison et al., 2002). Feedback inputs connect mainly to the top laminae (Rockland and Pandya, 1979), resulting in negative MEG/EEG dipolar currents (Ahlfors et al., 2015; Cauller, 1995; Jones et al., 2007). Neurophysiological studies in non-human primates suggest that the majority of visual inputs to non-primary ACs have a feedback-type laminar profile (Schroeder and Foxe, 2002). According to previous non-human primate studies (Ghazanfar et al., 2005), ACs also receive directed modulations from association areas including STS during multisensory integration. Consistent with these findings, our network analyses suggested increased alpha/beta phase-locking between anterior ACs and STS/MTG in the $A_{\text{Repeat}} V_{\text{Change}} > A V_{\text{Repeat}}$ comparison. Therefore, in addition to local release from adaptation, the enhanced AC surface negativity effects after visually conveyed identity changes might involve increased cortico-cortical feedback. This may specifically apply to the later (and stronger) left aSTG pattern that was broadly distributed over time (van Atteveldt et al., 2014), whereas the earlier N1 effect might be predominantly related to release from adaptation that increases local recurrent excitation of top-layer synapses (May et al., 1999).

A topic of an intensive debate has been whether perceptual objects are formed via hierarchical convergence of sensory pathways at higher-order areas (Konorski, 1967), interactions across parallel pathways already at sensory cortices (Senkowski et al., 2008), or distributed processing across dynamic networks (Bizley and Cohen, 2013). The present functional connectivity results offer a novel perspective by suggesting that the prevailing processing scenario might vary depending on the previous stimulation history and the requirements of the particular setting. Repetition of events related to the same object increased synchronization within a network that was primarily feature-domain specific. That is, the anterior AC seed regions (auditory “what” areas) were connected with ventral visual cortex areas involved with the processing of animal faces (Tong et al., 2000) and right orbitofrontal “what” areas that may provide feedback to inferior temporal areas during visual object recognition (Bar et al., 2006). The right pSTG seed (auditory “where” area) showed, correspondingly, increased synchrony with dorsal visual pathway areas of medial/superior parietal cortex, and the left pSTG with frontal areas overlapping the left FEF, a known spatial processing area that has abundant bilateral fiber connections with posterior ACs (Rauschecker and Tian, 2000; Romanski et al., 1999). The only exception was the connectivity of left pSTG seed with two clusters encompassing the left ventral visual pathway. Overall, although multisensory object formation clearly involves extensive interactions across multiple brain areas (Senkowski et al., 2008), processing of object identity may follow a more predetermined, hierarchical route, converging within the higher-order “what” pathway when the object identity rules have been more firmly established through repeated events.

In contrast to the feature-domain specific pattern of activation observed after repetition, the visually identity changes resulted in distributed functional connectivity patterns across the putative “what” and “where” processing streams of auditory and visual modalities. These enhancements were observed within the auditory system between the right aSTG seed and clusters extending to right pSTG and the supramodal spatial processing area IPS. The pSTG

seeds, in turn, exhibited enhanced connectivity patterns with clusters that extended to the right and left IFC, which is generally associated with higher-level processing of auditory object information (Rauschecker and Tian, 2000). In addition to the connections across feature domains, visual identity changes also increased alpha and beta phase locking between the sound-object/voice-identity sensitive aSTG seeds and visual-cortex clusters centered at the right lingual gyrus that might overlap with the color-sensitive area V4 (Ungerleider and Haxby, 1994) and the left fusiform gyrus, which also strongly responds to cat faces in humans (Tong et al., 2000). These effects, together with the enhanced currents in aSTG at longer latencies, may be related to the reestablishment of object “binding” followed by breaking down of the binding when the cats had switched positions and the meow sound was produced by the other cat. The top-down control for this resetting process could be provided by the putative multisensory integration areas pSTS/MTG (Beauchamp et al., 2004; Calvert et al., 2000; Raij et al., 2000; Werner and Noppeney, 2010), which showed increased alpha/beta phase locking with aSTG areas after visual identity changes.

The control data collected using the visual stimulus without the meow sounds suggested small modulations of AC even without sound stimulation, consistent with previous studies (Calvert et al., 1997; Pekkola et al., 2005; Raij et al., 2010). However, none of the significant modulations coincided with the differences caused by visual identity changes in the main experiment. The effects elicited by the silent Probe stimuli were temporally quite smooth, and observable only in posterior ACs, whereas the longer-latency AC effect in the main experiment was observed in the left anterior AC. For the Probe stimuli, the modulations were not as clear as, for example, in our recent MEG/EEG study, which used basic stimuli with a sharp temporal onset (Raij et al., 2010). However, the supporting analysis of the silent Adaptors did indeed show small N1-like responses that, consistent with our previous study (Raij et al., 2010), had longer latencies than the responses elicited by the audiovisual Adaptors. Overall, these control analyses support the view that our main results were due to the combination of auditory and visual inputs, instead of visual changes alone.

In the main analyses (**Fig. 2**), the pSTG time courses showed a clear event-related response pattern, with a N1-like negative (i.e., inward current) response followed by a positive deflection that resembles the P2 component. Whereas analogous peaks were also identifiable for the right aSTG, these deflections are more difficult to identify in the left aSTG time courses. These differences could be related to both neurophysiological and technical factors. For one thing, previous studies generally suggest that the temporal integration windows lengthen and the temporal precision decreases as the activations spread along the posterior-anterior axis of superior temporal plane (DeWitt and Rauschecker, 2012; Jääskeläinen et al., 2004a; Santoro et al., 2014). Weaker phase locking of aSTG neurons to the auditory events could partially explain the lack of as clear event-related peaks as in pSTGs. At the same time, the observation that the left aSTG showed less clear peaks than the right aSTG might be in part explainable by differential cancellation patterns of electromagnetic fields due to the local sulcal and gyral patterns (Shaw et al., 2013). The hemispheric differences could also reflect greater susceptibility of left vs. right auditory areas to suppression of vocalization responses due to simultaneous visual stimuli, a phenomenon that has been previously evidenced in audiovisual speech studies (Jääskeläinen et al., 2008). Differential

suppression of responses by concurrent visual stimuli could also contribute to the differences between anterior and posterior areas. Unfortunately, the present study did not include a control condition with auditory stimuli alone to examine this possibility. In the same vein, cat communication sounds (i.e., “meow”) exhibit complex pitch- and timbre shifts (**Fig. 1b**). Several previous studies suggest that sounds containing complex frequency modulations and, for example, melodies are processed more anteriorly than sounds consisting of fixed pitch patterns (Barrett and Hall, 2006; Patterson et al., 2002; Puschmann et al., 2010). The sensitivity of aSTGs to transient pitch and timbre changes in the meow stimulus could have resulted in a more complex time course that then masks some of the initial onset-responses that dominate the pSTG time courses.

In addition to ACs, the visual identity changes activated a broad range of other areas, most of which were also involved in the oscillatory networks revealed in the connectivity analyses. The early activations in the fusiform cortex and in the area MT could be related to the processing of visual stimulus changes. There were also modulations in paracentral regions, near SMA, which according to previous studies is activated by synchronous auditory and visual stimuli (Marchant et al., 2012). Change-related event-related modulations were also observed in the right supramarginal, insular, and inferior parietal cortices, areas that in previous studies have been shown to be involved in audiovisual interactions (Beauchamp et al., 2004; Benoit et al., 2010; Bischoff et al., 2007; Calvert et al., 2000), as well as in the right superior frontal cortex.

Many studies suggest that multisensory interaction effects, including the McGurk and ventriloquism illusion (McGurk and MacDonald, 1976; Sams et al., 1991), are independent of attention (De Meo et al., 2015). Therefore, we presumed that using naturalistic content-congruent stimuli, such as meowing cats, feature integration across modalities and feature pathways occurs without active effort. The behavioral task during scanning was intended to control the level of subjects’ attention and alertness. This differs from certain previous studies in which feature binding within modalities as well as multisensory integration were studied using abstract feature combinations, during attentionally demanding tasks whose instructions were more directly related to the integration process itself (Fujisaki et al., 2006; Fujisaki and Nishida, 2010; Talsma et al., 2010; van Ee et al., 2009). Indeed, whereas attention-independent mechanisms may suffice when the level of competition is low and/or the crossmodal associations are well-learned, the role of attention is more critical when multiple objects compete for processing resources (Alsus et al., 2005; Fujisaki et al., 2006; Fujisaki and Nishida, 2010; Munhall et al., 2009; Talsma et al., 2010). Further, the capacity of audiovisual integration has been shown to be limited to only one item at the time (Van der Burg et al., 2013), and psychophysical studies suggest that feature binding across auditory and visual systems and between feature pathways within these modalities is considerably slower than within a modality or feature domain (Fujisaki and Nishida, 2010). This evidence emphasizes the importance of slow attentionally controlled feature binding mechanisms under challenging stimulation conditions. In future studies, it would be highly interesting to test how the AC effects and interregional modulations shown in the present study might be replicated in such settings.

Further studies are needed to verify to which degree the effects observed in this study depend on the congruence between the visual and auditory content. At the neuronal level, object-related associations could occur to even simple feature combinations such as black vs. gray squares making a beep sound. Here, we chose the cat stimuli because we anticipated that this would provide stronger effects and a better signal-to-noise ratio. Another topic needing further experimental corroboration is the role of spatial congruence in the emergence of audiovisual object-related integration effects. We simulated the sounds from a direction congruent with the visual location of the cat to boost the congruency effect. However, given the poorer spatial resolution of auditory vs. visual spatial perception, evidenced as the proneness of auditory localization to be dominated by a coinciding visual stimulus (Bischoff et al., 2007; Busse et al., 2005; Lewald, 2002; Recanzone, 1998), the effect would likely have been quite similar if the sound had originated from straight ahead. Finally, the present task design was deliberately made as simple as possible, as it was assumed the modulations in sound processing regions by the visual identity change, with no sound-associated dynamic modulations comparable to, e.g., McGurk experiments, would be quite small in amplitude. Further studies with more dynamic settings, where both auditory and visual aspects of stimuli are spatially dynamic, are needed to further elaborate to which degree the present effects relate to processing of audiovisual object-identity associations vs. perceptual object formation.

5. Conclusions

In conclusion, our results revealed potential cortical bases of multisensory perceptual object representations. Visually conveyed change in the identity of a perceptual object resulted in early 50–150-ms latency release from adaptation in posterior AC areas, followed by enhancement of inward currents in aSTG 250–450 ms from sound onset. These results suggest that ACs play a role in the formation of multisensory object representations involving sound stimuli. Our cortico-cortical phase locking estimates, further, suggested that hierarchical feature-domain specific networks might emerge to support the processing of repeated and familiar objects, whereas a broadly distributed network across auditory, visual, and multisensory areas is activated when the association between a sound and a visual stimulus changes.

Acknowledgements

In memory of our friend and mentor Dr. John W. Belliveau. We thank Nao Suzuki and Dr. Wei-Tang Chang for their advice and support. This work was supported by National Institutes of Health (NIH) Awards R01MH083744, R21DC014134, R01HD040712, R01NS037462, and 5R01EB009048, and by the Academy of Finland (authors IPJ and MS). The research environment was supported by the NIH awards P41EB015896, S10RR014978, S10RR021110, S10RR019307, S10RR014798, and S10RR023401.

References

- Ahlfors SP, Jones SR, Ahveninen J, Hämäläinen MS, Belliveau JW, Bar M. Direction of magnetoencephalography sources associated with feedback and feedforward contributions in a visual object recognition task. *Neurosci Lett*. 2015; 585:149–154. [PubMed: 25445356]
- Ahveninen J, Hämäläinen M, Jääskeläinen IP, Ahlfors SP, Huang S, Lin FH, Raij T, Sams M, Vasios CE, Belliveau JW. Attention-driven auditory cortex short-term plasticity helps segregate relevant sounds from noise. *Proc Natl Acad Sci U S A*. 2011; 108:4182–4187. [PubMed: 21368107]

- Ahveninen J, Huang S, Nummenmaa A, Belliveau JW, Hung AY, Jääskeläinen IP, Rauschecker JP, Rossi S, Tiitinen H, Raij T. Evidence for distinct human auditory cortex regions for sound location versus identity processing. *Nat Commun.* 2013; 4:2585. [PubMed: 24121634]
- Ahveninen J, Jääskeläinen IP, Raij T, Bonmassar G, Devore S, Hämäläinen M, Levänen S, Lin FH, Sams M, Shinn-Cunningham BG, Witzel T, Belliveau JW. Task-modulated “what” and “where” pathways in human auditory cortex. *Proc Natl Acad Sci U S A.* 2006; 103:14608–14613. [PubMed: 16983092]
- Alain C, Arnott SR, Hevenor S, Graham S, Grady CL. “What” and “where” in the human auditory system. *Proc Natl Acad Sci U S A.* 2001; 98:12301–12306. [PubMed: 11572938]
- Algazi, VR.; Duda, RO.; Thompson, M.; Avendano, C. The CIPIC HRTF database. IEEE Workshop on Applications of Signal Processing to Audio and Electroacoustics. Mohonk Mountain House; New Paltz, NY: 2001. p. 99-102.
- Alho J, Lin FH, Sato M, Tiitinen H, Sams M, Jääskeläinen IP. Enhanced neural synchrony between left auditory and premotor cortex is associated with successful phonetic categorization. *Front Psychol.* 2014; 5:394. [PubMed: 24834062]
- Allison T, Puce A, McCarthy G. Category-sensitive excitatory and inhibitory processes in human extrastriate cortex. *J Neurophysiol.* 2002; 88:2864–2868. [PubMed: 12424319]
- Alsius A, Navarra J, Campbell R, Soto-Faraco S. Audiovisual integration of speech falters under high attention demands. *Curr Biol.* 2005; 15:839–843. [PubMed: 15886102]
- Altmann CF, Bledowski C, Wibral M, Kaiser J. Processing of location and pattern changes of natural sounds in the human auditory cortex. *Neuroimage.* 2007; 35:1192–1200. [PubMed: 17320413]
- Altmann CF, Getzmann S, Lewald J. Allocentric or craniocentric representation of acoustic space: an electrotomography study using mismatch negativity. *PLoS ONE.* 2012; 7:e41872. [PubMed: 22848643]
- Arnal LH, Wyart V, Giraud AL. Transitions in neural oscillations reflect prediction errors generated in audiovisual speech. *Nat Neurosci.* 2011; 14:797–801. [PubMed: 21552273]
- Arnott SR, Grady CL, Hevenor SJ, Graham S, Alain C. The functional organization of auditory working memory as revealed by fMRI. *J Cogn Neurosci.* 2005; 17:819–831. [PubMed: 15904548]
- Bar M, Kassam KS, Ghuman AS, Boshyan J, Schmidt AM, Dale AM, Hämäläinen MS, Marinkovic K, Schacter DL, Rosen BR, Halgren E. Top-down facilitation of visual recognition. *Proc Natl Acad Sci U S A.* 2006; 103:449–454. [PubMed: 16407167]
- Barrett DJ, Hall DA. Response preferences for “what” and “where” in human non-primary auditory cortex. *Neuroimage.* 2006; 32:968–977. [PubMed: 16733092]
- Beauchamp MS, Lee KE, Argall BD, Martin A. Integration of auditory and visual information about objects in superior temporal sulcus. *Neuron.* 2004; 41:809–823. [PubMed: 15003179]
- Belin P, Zatorre RJ. Adaptation to speaker's voice in right anterior temporal lobe. *Neuroreport.* 2003; 14:2105–2109. [PubMed: 14600506]
- Benoit MM, Raij T, Lin FH, Jääskeläinen IP, Stufflebeam S. Primary and multisensory cortical activity is correlated with audiovisual percepts. *Hum Brain Mapp.* 2010; 31:526–538. [PubMed: 19780040]
- Bischoff M, Walter B, Blecker CR, Morgen K, Vaitl D, Sammer G. Utilizing the ventriloquism-effect to investigate audio-visual binding. *Neuropsychologia.* 2007; 45:578–586. [PubMed: 16620884]
- Bizley JK, Cohen YE. The what, where and how of auditory-object perception. *Nat Rev Neurosci.* 2013; 14:693–707. [PubMed: 24052177]
- Bizley JK, Nodal FR, Bajo VM, Nelken I, King AJ. Physiological and anatomical evidence for multisensory interactions in auditory cortex. *Cereb Cortex.* 2007; 17:2172–2189. [PubMed: 17135481]
- Bonath B, Noesselt T, Martinez A, Mishra J, Schwiecker K, Heinze HJ, Hillyard SA. Neural basis of the ventriloquist illusion. *Curr Biol.* 2007; 17:1697–1703. [PubMed: 17884498]
- Budinger E, Heil P, Hess A, Scheich H. Multisensory processing via early cortical stages: Connections of the primary auditory cortical field with other sensory systems. *Neuroscience.* 2006; 143:1065–1083. [PubMed: 17027173]

- Busse L, Roberts KC, Crist RE, Weissman DH, Woldorff MG. The spread of attention across modalities and space in a multisensory object. *Proc Natl Acad Sci U S A*. 2005; 102:18751–18756. [PubMed: 16339900]
- Calvert G, Campbell R, Brammer M. Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex. *Current Biology*. 2000; 10:649–657. [PubMed: 10837246]
- Calvert GA, Bullmore ET, Brammer MJ, Campbell R, Williams SC, McGuire PK, Woodruff PW, Iversen SD, David AS. Activation of auditory cortex during silent lipreading. *Science*. 1997; 276:593–596. [PubMed: 9110978]
- Caulier L. Layer I of primary sensory neocortex: where top-down converges upon bottom-up. *Behav Brain Res*. 1995; 71:163–170. [PubMed: 8747184]
- Clarke S, Bellmann Thiran A, Maeder P, Adriani M, Vernet O, Regli L, Cuisenaire O, Thiran JP. What and where in human audition: selective deficits following focal hemispheric lesions. *Exp Brain Res*. 2002; 147:8–15. [PubMed: 12373363]
- Dale A, Sereno M. Improved localization of cortical activity by combining EEG and MEG with MRI cortical surface reconstruction: A linear approach. *J Cog Neurosci*. 1993; 5:162–176.
- De Meo R, Murray MM, Clarke S, Matusz P. Top-down control and early multisensory processes: Chicken vs. egg. *Front Integr Neurosci* In press. 2015
- Desikan R, Segonne F, Fischl B, Quinn B, Dickerson B, Blacker D, Buckner R, Dale A, Maguire R, Hyman B, Albert M, Killiany R. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *Neuroimage*. 2006; 31:968–980. [PubMed: 16530430]
- DeWitt I, Rauschecker JP. Phoneme and word recognition in the auditory ventral stream. *Proc Natl Acad Sci U S A*. 2012; 109:E505–514. [PubMed: 22308358]
- Ding L, Yuan H. Simultaneous EEG and MEG source reconstruction in sparse electromagnetic source imaging. *Hum Brain Mapp*. 2013; 34:775–795. [PubMed: 22102512]
- Doesburg SM, Emberson LL, Rahi A, Cameron D, Ward LM. Asynchrony from synchrony: long-range gamma-band neural synchrony accompanies perception of audiovisual speech asynchrony. *Exp Brain Res*. 2008; 185:11–20. [PubMed: 17922119]
- Engel AK, Fries P, Singer W. Dynamic predictions: oscillations and synchrony in top-down processing. *Nat Rev Neurosci*. 2001; 2:704–716. [PubMed: 11584308]
- Ermentrout GB, Kopell N. Fine structure of neural spiking and synchronization in the presence of conduction delays. *Proc Natl Acad Sci U S A*. 1998; 95:1259–1264. [PubMed: 9448319]
- Fischl B, Sereno MI, Tootell RB, Dale AM. High-resolution intersubject averaging and a coordinate system for the cortical surface. *Hum Brain Mapp*. 1999; 8:272–284. [PubMed: 10619420]
- Fujisaki W, Koene A, Arnold D, Johnston A, Nishida S. Visual search for a target changing in synchrony with an auditory signal. *Proc Biol Sci*. 2006; 273:865–874. [PubMed: 16618681]
- Fujisaki W, Nishida S. A common perceptual temporal limit of binding synchronous inputs across different sensory attributes and modalities. *Proc Biol Sci*. 2010; 277:2281–2290. [PubMed: 20335212]
- Ghazanfar AA, Maier JX, Hoffman KL, Logothetis NK. Multisensory integration of dynamic faces and voices in rhesus monkey auditory cortex. *J Neurosci*. 2005; 25:5004–5012. [PubMed: 15901781]
- Gramfort A, Luessi M, Larson E, Engemann DA, Strohmeier D, Brodbeck C, Parkkonen L, Hämäläinen MS. MNE software for processing MEG and EEG data. *Neuroimage*. 2013
- Griffiths TD, Warren JD. The planum temporale as a computational hub. *Trends in Neurosciences*. 2002; 25:348–353. [PubMed: 12079762]
- Hämäläinen M, Hari R, Ilmoniemi R, Knuutila J, Lounasmaa O. Magnetoencephalography - theory, instrumentation, and applications to noninvasive studies of the working human brain. *Rev. Mod. Phys*. 1993; 65:413–497.
- Henson RN, Mouchlianitis E, Friston KJ. MEG and EEG data fusion: simultaneous localisation of face-evoked responses. *Neuroimage*. 2009; 47:581–589. [PubMed: 19398023]
- Hipp JF, Engel AK, Siegel M. Oscillatory synchronization in large-scale cortical networks predicts perception. *Neuron*. 2011; 69:387–396. [PubMed: 21262474]

- Huang S, Rossi S, Hämäläinen M, Ahveninen J. Auditory conflict resolution correlates with medial-lateral frontal theta/alpha phase synchrony. *PLoS ONE*. 2014; 9:e110989. [PubMed: 25343503]
- Jääskeläinen IP, Ahveninen J, Andermann ML, Belliveau JW, Raij T, Sams M. Short-term plasticity as a neural mechanism supporting memory and attentional functions. *Brain Res*. 2011; 1422:66–81. [PubMed: 21985958]
- Jääskeläinen IP, Ahveninen J, Bonmassar G, Dale AM, Ilmoniemi RJ, Levänen S, Lin FH, May P, Melcher J, Stufflebeam S, Tiitinen H, Belliveau JW. Human posterior auditory cortex gates novel sounds to consciousness. *Proc Natl Acad Sci U S A*. 2004a; 101:6809–6814. [PubMed: 15096618]
- Jääskeläinen IP, Kauramäki J, Tujunen J, Sams M. Formant transition-specific adaptation by lipreading of left auditory cortex N1m. *Neuroreport*. 2008; 19:93–97. [PubMed: 18281900]
- Jääskeläinen IP, Ojanen V, Ahveninen J, Auranen T, Levänen S, Möttönen R, Tarnanen I, Sams M. Adaptation of neuromagnetic N1 responses to phonetic stimuli by visual speech in humans. *Neuroreport*. 2004b; 15:2741–2744. [PubMed: 15597045]
- Jensen O, Kaiser J, Lachaux JP. Human gamma-frequency oscillations associated with attention and memory. *Trends Neurosci*. 2007; 30:317–324. [PubMed: 17499860]
- Jones SR, Pritchett DL, Stufflebeam SM, Hämäläinen M, Moore CI. Neural correlates of tactile detection: a combined magnetoencephalography and biophysically based computational modeling study. *J Neurosci*. 2007; 27:10751–10764. [PubMed: 17913909]
- Kayser C, Petkov CI, Augath M, Logothetis NK. Integration of touch and sound in auditory cortex. *Neuron*. 2005; 48:373–384. [PubMed: 16242415]
- Konorski J. Integrative activity of the brain; an interdisciplinary approach. University of Chicago Press; Chicago: 1967.
- Kopell N, Ermentrout GB, Whittington MA, Traub RD. Gamma rhythms and beta rhythms have different synchronization properties. *Proc Natl Acad Sci U S A*. 2000; 97:1867–1872. [PubMed: 10677548]
- Leung AW, Alain C. Working memory load modulates the auditory “What” and “Where” neural networks. *Neuroimage*. 2011; 55:1260–1269. [PubMed: 21195187]
- Lewald J. Rapid adaptation to auditory-visual spatial disparity. *Learn Mem*. 2002; 9:268–278. [PubMed: 12359836]
- Lin FH, Belliveau JW, Dale AM, Hämäläinen MS. Distributed current estimates using cortical orientation constraints. *Hum Brain Mapp*. 2006; 27:1–13. [PubMed: 16082624]
- Liu A, Dale A, Belliveau J. Monte Carlo simulation studies of EEG and MEG localization accuracy. *Human Brain Mapping*. 2002; 16:47–62. [PubMed: 11870926]
- Lomber SG, Malhotra S. Double dissociation of 'what' and 'where' processing in auditory cortex. *Nat Neurosci*. 2008; 11:609–616. [PubMed: 18408717]
- Luo H, Liu Z, Poeppel D. Auditory cortex tracks both auditory and visual stimulus dynamics using low-frequency neuronal phase modulation. *PLoS Biol*. 2010; 8:e1000445. [PubMed: 20711473]
- Maeder PP, Meuli RA, Adriani M, Bellmann A, Fornari E, Thiran JP, Pittet A, Clarke S. Distinct pathways involved in sound recognition and localization: a human fMRI study. *Neuroimage*. 2001; 14:802–816. [PubMed: 11554799]
- Marchant JL, Ruff CC, Driver J. Audiovisual synchrony enhances BOLD responses in a brain network including multisensory STS while also enhancing target-detection performance for both modalities. *Hum Brain Mapp*. 2012; 33:1212–1224. [PubMed: 21953980]
- Maris E, Oostenveld R. Nonparametric statistical testing of EEG- and MEG-data. *J Neurosci Methods*. 2007; 164:177–190. [PubMed: 17517438]
- May P, Tiitinen H, Ilmoniemi RJ, Nyman G, Taylor JG, Näätänen R. Frequency change detection in human auditory cortex. *J Comput Neurosci*. 1999; 6:99–120. [PubMed: 10333158]
- McGurk H, MacDonald J. Hearing lips and seeing voices. *Nature*. 1976; 264:746–748. [PubMed: 1012311]
- Munhall KG, ten Hove MW, Brammer M, Pare M. Audiovisual integration of speech in a bistable illusion. *Curr Biol*. 2009; 19:735–739. [PubMed: 19345097]

- Papanicolaou AC, Pazo-Alvarez P, Castillo EM, Billingsley-Marshall RL, Breier JI, Swank PR, Buchanan S, McManis M, Clear T, Passaro AD. Functional neuroimaging with MEG: normative language profiles. *Neuroimage*. 2006; 33:326–342. [PubMed: 16887368]
- Patterson RD, Uppenkamp S, Johnsrude IS, Griffiths TD. The processing of temporal pitch and melody information in auditory cortex. *Neuron*. 2002; 36:767–776. [PubMed: 12441063]
- Pekkola J, Ojanen V, Autti T, Jääskeläinen IP, Mottonen R, Tarkiainen A, Sams M. Primary auditory cortex activation by visual speech: an fMRI study at 3 T. *Neuroreport*. 2005; 16:125–128. [PubMed: 15671860]
- Perrodin C, Kayser C, Logothetis NK, Petkov CI. Auditory and visual modulation of temporal lobe neurons in voice-sensitive and association cortices. *J Neurosci*. 2014; 34:2524–2537. [PubMed: 24523543]
- Petkov CI, Kayser C, Steudel T, Whittingstall K, Augath M, Logothetis NK. A voice region in the monkey brain. *Nat Neurosci*. 2008; 11:367–374. [PubMed: 18264095]
- Puschmann S, Uppenkamp S, Kollmeier B, Thiel CM. Dichotic pitch activates pitch processing centre in Heschl's gyrus. *Neuroimage*. 2010; 49:1641–1649. [PubMed: 19782757]
- Raij T, Ahveninen J, Lin FH, Witzel T, Jääskeläinen IP, Letham B, Israeli E, Sahyoun C, Vasios C, Stufflebeam S, Hämäläinen M, Belliveau JW. Onset timing of cross-sensory activations and multisensory interactions in auditory and visual sensory cortices. *Eur J Neurosci*. 2010; 31:1772–1782. [PubMed: 20584181]
- Raij T, Uutela K, Hari R. Audiovisual integration of letters in the human brain. *Neuron*. 2000; 28:617–625. [PubMed: 11144369]
- Rauschecker JP, Scott SK. Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. *Nat Neurosci*. 2009; 12:718–724. [PubMed: 19471271]
- Rauschecker JP, Tian B. Mechanisms and streams for processing of “what” and “where” in auditory cortex. *Proc Natl Acad Sci U S A*. 2000; 97:11800–11806. [PubMed: 11050212]
- Recanzone GH. Rapidly induced auditory plasticity: the ventriloquism aftereffect. *Proc Natl Acad Sci U S A*. 1998; 95:869–875. [PubMed: 9448253]
- Recanzone GH, Cohen YE. Serial and parallel processing in the primate auditory cortex revisited. *Behav Brain Res*. 2010; 206:1–7. [PubMed: 19686779]
- Rockland KS, Pandya DN. Laminar origins and terminations of cortical connections of the occipital lobe in the rhesus monkey. *Brain Res*. 1979; 179:3–20. [PubMed: 116716]
- Roelfsema PR, Engel AK, König P, Singer W. Visuomotor integration is associated with zero time-lag synchronization among cortical areas. *Nature*. 1997; 385:157–161. [PubMed: 8990118]
- Romanski LM, Tian B, Fritz J, Mishkin M, Goldman-Rakic PS, Rauschecker JP. Dual streams of auditory afferents target multiple domains in the primate prefrontal cortex. *Nat Neurosci*. 1999; 2:1131–1136. [PubMed: 10570492]
- Sams M, Aulanko R, Hämäläinen M, Hari R, Lounasmaa O, Lu S, Simola J. Seeing speech: visual information from lip movements modifies activity in the human auditory cortex. *Neuroscience Letters*. 1991; 127:141–145. [PubMed: 1881611]
- Santoro R, Moerel M, De Martino F, Goebel R, Ugurbil K, Yacoub E, Formisano E. Encoding of natural sounds at multiple spectral and temporal resolutions in the human auditory cortex. *PLoS Comput Biol*. 2014; 10:e1003412. [PubMed: 24391486]
- Schroeder CE, Foxe JJ. The timing and laminar profile of converging inputs to multisensory areas of the macaque neocortex. *Brain Res Cogn Brain Res*. 2002; 14:187–198. [PubMed: 12063142]
- Schroeder CE, Lakatos P, Kajikawa Y, Partan S, Puce A. Neuronal oscillations and visual amplification of speech. *Trends Cogn Sci*. 2008; 12:106–113. [PubMed: 18280772]
- Senkowski D, Schneider TR, Foxe JJ, Engel AK. Crossmodal binding through neural coherence: implications for multisensory processing. *Trends Neurosci*. 2008; 31:401–409. [PubMed: 18602171]
- Sharon D, Hämäläinen MS, Tootell RB, Halgren E, Belliveau JW. The advantage of combining MEG and EEG: comparison to fMRI in focally stimulated visual cortex. *Neuroimage*. 2007; 36:1225–1235. [PubMed: 17532230]

- Shaw ME, Hämäläinen MS, Gutschalk A. How anatomical asymmetry of human auditory cortex can lead to a rightward bias in auditory evoked fields. *Neuroimage*. 2013; 74:22–29. [PubMed: 23415949]
- Smiley JF, Hackett TA, Ulbert I, Karmas G, Lakatos P, Javitt DC, Schroeder CE. Multisensory convergence in auditory cortex, I. Cortical connections of the caudal superior temporal plane in macaque monkeys. *J Comp Neurol*. 2007; 502:894–923. [PubMed: 17447261]
- Talsma D, Senkowski D, Soto-Faraco S, Woldorff MG. The multifaceted interplay between attention and multisensory integration. *Trends Cogn Sci*. 2010; 14:400–410. [PubMed: 20675182]
- Taulu S, Simola J, Kajola M. Applications of the Signal Space Separation Method. *IEEE Trans. Signal Proc*. 2005; 53:3359–3372.
- Tian B, Reser D, Durham A, Kustov A, Rauschecker JP. Functional specialization in rhesus monkey auditory cortex. *Science*. 2001; 292:290–293. [PubMed: 11303104]
- Todorovic A, van Ede F, Maris E, de Lange FP. Prior expectation mediates neural adaptation to repeated sounds in the auditory cortex: an MEG study. *J Neurosci*. 2011; 31:9118–9123. [PubMed: 21697363]
- Tong F, Nakayama K, Moscovitch M, Weinrib O, Kanwisher N. Response properties of the human fusiform face area. *Cogn Neuropsychol*. 2000; 17:257–280. [PubMed: 20945183]
- Ungerleider LG, Haxby JV. 'What' and 'where' in the human brain. *Curr Opin Neurobiol*. 1994; 4:157–165. [PubMed: 8038571]
- van Atteveldt NM, Peterson BS, Schroeder CE. Contextual control of audiovisual integration in low-level sensory cortices. *Hum Brain Mapp*. 2014; 35:2394–2411. [PubMed: 23982946]
- Van der Burg E, Awh E, Olivers CN. The capacity of audiovisual integration is limited to one item. *Psychol Sci*. 2013; 24:345–351. [PubMed: 23389426]
- van Driel J, Knapen T, van Es DM, Cohen MX. Interregional alpha-band synchrony supports temporal cross-modal integration. *Neuroimage*. 2014; 101:404–415. [PubMed: 25042447]
- van Ee R, van Boxtel JJ, Parker AL, Alais D. Multisensory congruency as a mechanism for attentional control over perceptual selection. *J Neurosci*. 2009; 29:11641–11649. [PubMed: 19759311]
- van Wassenhove V, Grant KW, Poeppel D. Visual speech speeds up the neural processing of auditory speech. *Proc Natl Acad Sci U S A*. 2005; 102:1181–1186. [PubMed: 15647358]
- Vinck M, Oostenveld R, van Wingerden M, Battaglia F, Pennartz CM. An improved index of phase-synchronization for electrophysiological data in the presence of volume-conduction, noise and sample-size bias. *Neuroimage*. 2011; 55:1548–1565. [PubMed: 21276857]
- von Kriegstein K, Kleinschmidt A, Sterzer P, Giraud AL. Interaction of face and voice areas during speaker recognition. *J Cogn Neurosci*. 2005; 17:367–376. [PubMed: 15813998]
- von Stein A, Rappelsberger P, Sarnthein J, Petsche H. Synchronization between temporal and parietal cortex during multimodal object processing in man. *Cereb Cortex*. 1999; 9:137–150. [PubMed: 10220226]
- Weeks RA, Aziz-Sultan A, Bushara KO, Tian B, Wessinger CM, Dang N, Rauschecker JP, Hallett M. A PET study of human auditory spatial processing. *Neurosci Lett*. 1999; 262:155–158. [PubMed: 10218879]
- Werner S, Noppeney U. Superadditive responses in superior temporal sulcus predict audiovisual benefits in object categorization. *Cereb Cortex*. 2010; 20:1829–1842. [PubMed: 19923200]

Highlights

- Sound representations are linked to perceptual objects already in auditory cortex.
- Posterior ACs initiate and anterior ACs consolidate the object associations.
- Repeating associations are processed in parallel in “what” and “where” pathways.
- New associations engage distributed auditory–visual–multisensory networks.

Figure 1.

Task and stimuli. **(a)** Audiovisual object tracking task. Subjects viewed movie clips containing an Adaptor-Probe audiovisual stimulus-event pair. A gray and a black cat were presented on the screen. The Adaptor stimulus consisted of the black cat meowing, after which both cats moved to the center and then either returned to the original positions or switched sides. The Probe stimulus was the meowing of the cat that was on the same position as the initial Adaptor stimulus. Thus, the sound stimulus (“meow”) was always the same, but the visually conveyed identity of the meowing cat changed in 50% in of the trials. In other words, the Adaptor and Probe events either belonged repeatedly to the same object (AV_{Repeat}) or different objects ($A_{Repeat}V_{Change}$). Subjects were instructed to respond with one hand if the cat returned in the same position and with another if they switched positions. Note that the motion pattern between Probe and Adaptor has been simplified for clarity. **(b)** Auditory stimulus. The time-domain representation (*top*) and a narrowband spectrogram (30-ms Hanning window, *bottom*) of the sound stimulus are shown.

Figure 2.

Cortical activations after visual identity changes ($A_{Repeat} V_{Change}$), as compared with the AV_{Repeat} condition. **(a)** Significant activity modulations were observed in visual, auditory, and association cortex areas. The views with significant AC modulations encircled. **(b)** Signal polarity within the significant left superior temporal clusters. Significant increases in cortical surface negativity, corresponding to enhanced inward source currents, extended from Heschl's gyrus (HG) (250–350 ms) to aSTG and the planum polare (PP) at 250–450 ms. **(c)** ROI time courses of noise-normalized MNE (dSPM), normalized for display by the within-ROI standard deviation, show significant increases of surface negativity after visual identity changes in the left aSTG around 250–450 ms. The error bars indicate the standard error of mean across subjects. **(d)** Comparisons of averaged dSPM amplitudes at 50–150 ms suggest significantly increased release from adaptation of N1 activity in pSTG areas after visual identity changes. (PT: planum temporale).

Figure 3.

AC ROI time course analysis during silent movie clips. **(a)** Locations of ROIs. **(b)** ROI time course analysis of silent Probe events. The visual stimulus alone resulted in a slow positive drift for the AV_{Repeat} condition in pSTGs. However, the observed effects were spatially and temporally different from the effects of the main experiment, as the time courses in left aSTG, where the most significant modulations were observed in the main experiment, were virtually similar across the respective conditions of the control experiment. No significant differences were observed in the average N1 amplitude during the 50–150 ms time window. The time courses are normalized for display to the same scale than the corresponding time courses in **Fig 2c**. **(c)** AC ROI time course analysis of Adaptor responses during audiovisual and silent movie clips. Consistent with our previous studies (Raij et al., 2010), weak onset responses that share resemblance to the N1 response were observed even to silent Adaptors (marked with an arrow). The time courses are normalized for display by the standard deviation of responses within each ROI during the audiovisual condition. In the panels **b** and **c**, the error bars indicate the standard error of mean across subjects at each time point.

Figure 4.

Phase locking of oscillatory activity between the AC seed regions and the rest of the cortex. (a) A diagram of auditory and visual dual pathways, and description of seed region and target cluster color labeling. (b) Repetition of events related to the same object ($AV_{Repeat} > A_{Repeat} V_{Change}$) increased theta synchronization in a hierarchical/modular fashion (for details, see **Table 1**). Both aSTG seeds were synchronized with areas associated with visual object recognition ("what"), including ventral visual cortex and inferior/orbitofrontal areas (see (Bar et al., 2006)). The pSTGs were in turn synchronized with areas traditionally associated with higher-order "where" stream, with the only exception being the strengthened connection between the left pSTG and left ventral temporal/occipital areas. (c) In trials with visually conveyed identity change ($A_{Repeat} V_{Change} > AV_{Repeat}$), synchronization increased in more distributed fashion, and predominantly at higher frequencies than in the opposite contrast. The largest number of clusters occurred at the beta band, and the most common connectivity pattern was between "what" and "where" streams (see, e.g., the right aSTG seed; for details, see **Table 2**).

Table 1

Significant clusters, showing enhanced wPLI with AC seed ROIs after repetition of events linked to the same object ($AV_{\text{Repeat}} > A_{\text{Repeat}}V_{\text{Change}}$). The maximum values of initial GLM reflect $-\log_{10}(p) \times \text{sign}(t)$.

Seed	Target hemisphere	Freq. range	Max value	Size (mm ²)	x_{MNI}	y_{MNI}	z_{MNI}	Cluster P	Anatomical site of maximum
Left pSTG									
	Left								
		Theta	−3.5	2170	−35	5	−38	0.0156	Inferior temporal
		Theta	−3.0	3510	−48	−24	−9	0.0006	STG
		Alpha	−2.7	3153	−15	−92	−8	0.0035	Lateral occipital
		Beta	−3.5	2939	−40	14	50	0.0017	Caudal middle frontal
Right pSTG									
	Left								
		Theta	−2.9	2755	−12	−82	32	0.0052	Superior parietal
Left aSTG									
	Right								
		Theta	−3.2	2516	19	25	−20	0.0055	Lateral orbitofrontal
		Theta	−2.8	3510	29	−74	−10	0.0004	Fusiform
Right aSTG									
	Right								
		Theta	−3.2	3418	53	−52	−19	0.0003	Inferior temporal

Table 2

Significant clusters showing enhanced wPLI with AC seed ROIs after visually conveyed identity change ($A_{\text{Repeat}}V_{\text{Change}} > AV_{\text{Repeat}}$). The maximum values of initial GLM reflect $-\log_{10}(p) \times \text{sign}(t)$.

Seed	Target hemisphere	Freq. range	Max value	Size (mm ²)	<i>x</i> _{MNI}	<i>y</i> _{MNI}	<i>z</i> _{MNI}	Cluster <i>P</i>	Anatomical site of maximum
Left pSTG									
	Left	Beta	2.7	2054	−25	16	−18	0.0162	Lateral orbitofrontal
		Right pSTG							
	Left	Alpha	2.9	4987	−41	29	20	0.0001	DLPFC
		Right							
	Alpha	2.6	2748	37	−26	3	0.0045	Insula	
	Beta	4.2	3726	48	−15	33	0.0004	Postcentral	
	Beta	2.9	2250	61	−17	3	0.0218	STG	
	Beta	2.3	2952	19	22	−19	0.0039	Lateral orbitofrontal	
	Left aSTG								
	Right	Theta	4.4	2026	16	−42	73	0.0241	Superior parietal
		Alpha	4.2	3863	14	−51	−3	0.0001	Lingual
	Alpha	2.2	3913	43	26	21	0.0001	DLPFC	
	Beta	3.4	2717	54	10	−14	0.0102	STG	
	Beta	2.8	3564	29	−34	65	0.0010	Postcentral	
	Right aSTG								
		Left	Beta	5.3	4426	−43	−28	−21	0.0001
Right									
Alpha		3.5	2277	34	−45	48	0.0194	Superior parietal / IPS	
Alpha		2.3	2454	43	−59	9	0.0118	Inferior parietal	
Beta		3.0	3622	57	−24	−14	0.0005	MTG	