Neural mechanisms for integrating consecutive and interleaved natural events

Juha M. Lahnakoski[1,2,3], Iiro P. Jääskeläinen[1], Mikko Sams[1], Lauri Nummenmaa[4,5]

1. Department of Neuroscience and Biomedical Engineering (NBE), School of Science, Aalto University, FI-00076, Espoo, Finland

2. Advanced Magnetic Imaging (AMI) Centre, Aalto NeuroImaging, School of Science, Aalto University, FI-00076, Espoo, Finland

3. Independent Max Planck Research Group for Social Neuroscience, Max Planck Institute of Psychiatry, DE-80804 Munich, Germany

4. Turku PET Centre, University of Turku, FI-20521, Turku, Finland

5. Department of Psychology, University of Turku, FI-20521, Turku, Finland

Corresponding author:

Juha M. Lahnakoski

Email: juha.lahnakoski@aalto.fi

Abstract

To understand temporally extended events, the human brain needs to accumulate information continuously across time. Interruptions that require switching of attention to other event sequences disrupt this process. To reveal neural mechanisms supporting integration of event information, we measured brain activity with functional magnetic resonance imaging (fMRI) from 18 participants during viewing of 6.5-minute excerpts from three movies i) consecutively and ii) as interleaved segments of approximately 50-s in duration. We measured inter-subject reliability of brain activity by calculating inter-subject correlations (ISC) of fMRI signals and analyzed activation timecourses with a general linear model (GLM). Interleaving decreased the ISC in posterior temporal lobes, medial prefrontal cortex, superior precuneus, medial occipital cortex and cerebellum. In the GLM analyses, posterior temporal lobes were activated more consistently by instances of speech when the movies were viewed consecutively than as interleaved segments. By contrast, low-level auditory and visual stimulus features and editing boundaries caused similar activity patterns in both conditions. In the medial occipital cortex, decreases in ISC were seen in short bursts throughout the movie clips. By contrast, the other areas showed longer-lasting differences in ISC during isolated scenes depicting socially-relevant and suspenseful content, such as deception or inter-subject conflict. The areas in the posterior temporal lobes also showed sustained activity during continuous actions and were deactivated when actions ended at scene boundaries. Our results suggest that the posterior temporal and dorsomedial prefrontal cortices, as well as the cerebellum and dorsal precuneus, support integration of events into coherent event sequences.

Keywords: Temporal integration, interleaved processing, movie, social perception, speech, conflict
Conflicts of interest: none

Introduction

Perception of dynamic events in changing environments requires constant accumulation and integration of temporally unfolding information. Interruptions can severely disturb such integration. For example, additional effort is needed to recall what happened in a television series following a commercial break. To integrate temporally unfolding information, the brain builds a predictive model (schema) of the current situation. This facilitates interpretation of future events, detection of event boundaries, and integration of the perceived events with information in memory into a coherent whole (for a review, see Kurby and Zacks 2008).

Attention modulates processing of both targets and semantically related objects in widespread areas of the cerebral cortex (Çukur et al., 2013; see also Lahnakoski et al., 2014). The temporoparietal junction and parietal cortical areas participate in detection of perceptual event boundaries (Zacks et al. 2001; 2010; 2011), possibly by monitoring the continuity of action sequences from one event (i.e. independent subset of actions) to another. In particular, temporoparietal regions, complemented by the precuneus and superior occipital lobe, are more sensitive to events where the same action continues across a cut-point (i.e., edit) between scenes in a video compared with cut-points where the location or action abruptly changes (Magliano and Zacks 2011). These regions may support the integration of ongoing actions across varying viewpoints into a coherent whole. Moreover, differences in gray matter volume in frontal and temporal cortices are associated with both abnormal perception of event boundaries and difficulties in performing action sequences after seeing them in videos (Bailey et al., 2011). Specifically, reduced gray matter volume in the medial temporal lobe was associated with omitting actions seen in videos whereas, against predictions, increased gray matter volume in the dorsolateral prefrontal cortex was correlated with the number of inaccurately performed actions (e.g. packing the correct items but putting them in an inappropriate container) (Bailey et al., 2011).

Temporally scrambled versus intact movies (Hasson et al., 2008a) and audio stories (Lerner et al., 2011) result in lower inter-subject correlation (ISC; Hasson et al., 2004) in frontal and posterior temporal regions, while ISC in sensory areas remain unaffected. This may reflect different temporal scales of information accumulation in sensory *vs.* higher-order areas (Hasson et al., 2015). Scrambling a movie reduces prior information that the brain can use about preceding events. This compromises the construction of a coherent predictive model of the movie that, in an unscrambled version, would facilitate understanding of events unfolding over long time windows (Hasson et al., 2015). However, experiments where the chronological order of movie events is simply scrambled cannot distinguish between integration of ongoing, temporally continuous information and integration of information between two parts of the same event sequence when separated in time by intermittent presentation of events from another sequence.

Another line of research has shown that when low-level visual continuity (rather than temporal order) of a continuous action sequence is disrupted with camera angle changes, ISC in the frontal cortex remains unaffected while ISCs in the occipital and posterior parietal cortices increase (Herbec et al., 2015). This may be due to the increased visual changes at the cut-points eliciting synchronized onset-responses or higher attentional synchrony introduced by, e.g., particular camera angles (shot sizes) in the videos (Herbec et al., 2015).

Posterior temporal lobes and midline structures such as precuneus and dorsomedial prefrontal cortex process multiple aspects of social information (see e.g. Lahnakoski et al., 2012b). Social perception and particularly speech comprehension relies strongly on contextual prediction of social information and cues (see e.g. Smirnov et al., 2014). Usually, such contextual information accumulates over time. Social perception is thus more likely to be affected by manipulation of the temporal structure of incoming stimuli than processing of simple stimulus features or perception of short and contextually isolated event sequences.

The current study

Here we aimed at resolving how the human brain integrates temporally extended event sequences by showing subjects excerpts from three movies i) consecutively and ii) as interleaved segments with the order of events within each movie retained. We calculated ISCs across subjects in both conditions, and tested whether the strength of the ISC differed across the conditions. We also annotated scene transitions, low-level visual and auditory features, socially relevant content, and suspense in the movies, to reveal those features whose processing is most sensitive to flipping between event sequences. As an additional control analysis, we tested whether continuity disruptions caused deactivations (and continuous cuts activations; Magliano and Zacks 2011) in the brain regions that presumably integrate event-sequence information as

suggested by the differences between the conditions in the ISC and general linear model (GLM) analyses. Finally, we used an explorative "reverse correlation" approach (extended from Hasson et al., 2004) for revealing the events in the movies that caused the highest ISC differences across conditions.

We hypothesized that viewing the movies as intact *vs.* interleaved segments would trigger stronger ISC in posterior temporal and frontal brain regions (Hasson et al., 2008a; Lerner et al., 2011). We specifically predicted that the processing of social events, rather than low-level physical features, would explain the ISC differences because understanding social events requires longer time periods. By contrast, interleaving is unlikely to affect the processing of low-level features as suggested by findings showing that the inter-subject reliability of early sensory-cortical responses is high even when a movie is played backwards (Hasson et al., 2008a).

Materials and methods

Subjects

Eighteen healthy subjects (21–34 years, mean 28; three female; five left-handed) participated in the study. Subjects reported no history of neurological or psychiatric disorders, and were not taking medications affecting the central nervous system at the time of the study. All subjects provided an informed consent as a part of the protocol approved by the ethics committee of the Hospital District of Helsinki and Uusimaa. The study was carried out in accordance with the guidelines of the declaration of Helsinki.

Stimuli and procedure

The subjects watched audiovisual excerpts of three feature films: *Star Wars: Episode IV - A New Hope* (SW; length 6 min 38 s), *Indiana Jones: Raiders of the Lost Ark* (IJ; length 6 min 22 s) and the James Bond-movie *Golden Eye* (GE; length 6 min 24 sec). We selected these movies since their storylines are easy to follow and each has a clearly different theme (space opera, classic adventure, and agent adventure, respectively) making it easy for the subjects to know which movie they are currently viewing. The excerpts were taken from the beginning of the films, excluding opening credits and other text. They were subsequently cut into eight short segments (mean length 49 seconds ± 14 seconds, 24 segments in total) following the original cut-points in the movies, so that the abrupt audiovisual changes at the cut-points were similar in the two presentation conditions (see below).

The clips were presented in two ways (**Figure 1**) In the *Consecutive*-condition, the excerpts from each movie were presented one after another without breaks. Thus, the participants viewed the 6-minute excerpts as if they would have been watching the original movies. In the *Interleaved*-condition, the clips were also shown without breaks but their order was interleaved in a pseudo-random manner so that the movie (and thus also the plotline) changed after each segment, while the temporal order of segments *within* each original movie was retained. In other words, all the storylines unfolded in their original logical order yet the participants were required to switch between the storylines approximately once every 50 seconds. The presentation order of the conditions was counterbalanced across subjects to remove order effects between the conditions. Subjects were instructed to follow the plot of each of the movies, and to prepare to answer questions about the movie contents after the experiment. However, only ratings of familiarity of the movies and evaluations of the clip transitions were collected after the experiment. To aid the following of the movie events, each movie excerpt depicted only a single sequence of events. Thus, the subjects had to integrate temporally fragmented information only in the interleaved condition.

The beginning of the first video was synchronized to the beginning of fMRI sequence using Presentation software (Neurobehavioral Systems Inc., Albany, California, USA). The stimuli were back-projected on a semitransparent screen using a 3-micromirror data projector (Christie X3, Christie Digital Systems Ltd., Mönchengladbach, Germany). The subjects viewed the screen at 34 cm viewing distance via a mirror located above their eyes. Audio was played to the subjects with an UNIDES ADU2a audio system (Unides Design, Helsinki, Finland) via plastic tubes through EAR-tip (Etymotic Research, ER3, IL, USA) earplugs. The intensity of the auditory stimulation was selected to be loud enough to be heard over the scanner noise and was adjusted slightly (±2 dB) if the subject reported the sound to be too loud or too quiet before the experiment.

In a separate behavioral experiment after the scanning, subjects first saw the *last* 10 seconds of a movie segment, which was immediately followed by a one-second-duration arrow indicating the onset of the *first* 10 seconds of the subsequent segment. Then subjects used a 9-point Likert scale to rate i) how easy it was to follow the transition from one segment to another and ii) how natural the transition was. Ratings were given for all possible transitions in the Consecutive and Interleaved movies conditions. In

Figure 1: Illustrations of the stimuli used in the current study (top) and experimental design with Consecutive (bottom left) and Interleaved (bottom right) event sequence conditions. The background color of each condition indicates the movie and numbers indicate the index of the clip within the movie. In the Consecutive condition the movie clips were shown in the correct temporal order and movies were shown one after another. In the Interleaved condition the movie changed after each short clip, but the order of the clips within each movie was retained. The presentation order of the conditions was counterbalanced across subjects.

the Consecutive condition these constituted of 20-s excerpts of the natural flow of events in the movies (apart from the two transitions where the last two movies (IJ and GE) started, which were preceded by a different movie in both conditions). Transitions were presented in a random order. Finally, subjects indicated on a 9-point Likert scale how familiar they were with each of the movies (1 - not at all familiar, 9 - extremely familiar). Ratings of one subject were lost due to technical problems.

Functional magnetic resonance imaging

Functional brain imaging was carried out in the Advanced Magnetic Imaging (AMI) Centre of Aalto University with a 3.0 T GE Signa VH/i MRI scanner with HDxt upgrade (GE Healthcare Ltd., Chalfont St Giles, UK) using a 16-channel receiving head coil (MR Instruments Inc., MN, USA). The imaging area consisted of 36 functional gradient-echo planar oblique slices covering the whole brain (thickness 4 mm, no gap, in-plane resolution 3.75 mm x 3.75 mm, voxel matrix 64 x 64, TE 30 ms, TR 1800 ms, flip angle 75°, ASSET parallel imaging, R factor 2). ASSET calibration was performed prior to each imaging session. Total of 690 images were acquired continuously during both functional imaging sessions. In addition, a T_1-weighted spoiled gradient echo volume was acquired for anatomical alignment (SPGR pulse sequence, TE 1.9 ms, TR 10 ms, flip angle 15°, ASSET parallel imaging, R factor 2). The T_1 image consisted of 182 axial slices, in-plane resolution was 1 mm x 1 mm, matrix 256 x 256 and slice thickness 1 mm with no gap.

Preprocessing

Data were preprocessed with FSL (Smith et al., 2004; Woolrich et al., 2009). Motion correction was applied using MCFLIRT (Jenkinson et al., 2002), and non-brain matter was removed using BET (Smith 2002). The intensity distortions of the anatomical images were corrected prior to brain extraction, and values for intensity threshold and threshold gradient in BET were searched manually by changing the parameters and visually inspecting each brain-extracted volume. Using FLIRT, the datasets were registered to 2 mm MNI152 (Montreal Neurological Institute) standard-space template using the brain extracted T_1 weighted image of each individual subject as an intermediate step (Jenkinson et al., 2002). Registration both from functional to anatomical images and from anatomical images to standard template volumes was done using 9 degrees of freedom (DOF), allowing for translation, rotation, and scaling. Volume data were smoothed using a Gaussian kernel with full width at half maximum (FWHM) of 8.0 mm. High-pass temporal filtering was applied using Gaussian-weighted least-squares straight line fitting, with sigma 200 s, with the first 4 volumes of each dataset discarded to allow the magnetization to stabilize in the tissue (a fixation cross was presented during this time). The last 37 volumes of the data, which were recorded after the end of the stimulus, were removed leaving 649 volumes for the final analyses.

Analysis of behavioral data

Mean ratings for easiness to follow and naturalness of the transitions in the Consecutive *vs.* Interleaved conditions were compared across subjects with paired t-tests. The transitions to the first clips of each movie were excluded from the

analysis because they required reorienting to a new storyline in both conditions.

fMRI data analysis

All fMRI results are visualized on the "Conte-69" cerebral surface atlas and "Colin" cerebellar surface using Caret software (version 5.616; Van Essen et al., 2001).

Inter-subject correlation analysis

The ISCs for each voxel of the brain over both conditions were calculated using ISC toolbox (https://www.nitrc.org/projects/isc-toolbox/; Kauppi et al., 2014). To test for significance of ISC, we calculated the ISC between conditions (i.e. across pairs where the two participants were watching different (intact *vs.* segmented) versions of the stimuli) to build an empirical null distribution of ISC values. Because the first 50 seconds of stimulation in both conditions was the same, the ISC values in the null distribution were inflated due to the correlated activity during this time window. Thus, we used the 99th percentile of the observed values as a conservative threshold of significance without additional corrections[1]. Differences in within-condition ISC strength between Consecutive and Interleaved conditions were tested using Pearson-Filon statistics on Z transformed correlation coefficients, as implemented in the ISC toolbox (Kauppi et al., 2014). Specifically, this is accomplished by using a non-parametric permutation test that randomly re-labels the conditions (switching signs of pairwise differences) 25,000 times with the null hypothesis that relabeling does not change the ISC difference across conditions. The test concurrently controls the family-wise error rate (FWER) at p<0.05 by building the null distribution out of the maximal group differences observed in the whole brain. These differences across conditions were also essentially unchanged when the data was analyzed after excluding the first clip.

General linear model analysis

All GLM analyses were performed in SPM12 (http://www.fil.ion.ucl.ac.uk/spm). Canonical hemodynamic response function was used as the basis function and high-pass filter period was set to

___

[1] For the sake of comparison, we also calculated the ISCs after removing the first movie clip (which was identical in both conditions) and used the maximum observed ISC value between conditions as a threshold for the within-condition ISC maps. This procedure yielded essentially identical results to the ones calculated with the full timecourses, thus confirming that the chosen threshold effectively controls for multiple comparisons in the current data.

128 seconds. Event-related analysis of the movie transitions was performed by modeling the movie onsets as events in the design matrix, and a contrast was calculated between the transitions in the Consecutive *vs.* Interleaved conditions. The first clips of each movie were excluded from the event related analysis because an abrupt change in the visual stimulus and event context preceded these clips in both conditions.

To control for activity related to low-level sensory features, we extracted the visual contrast edges and sound power (root-mean-squared power; RMS) from the stimuli (for details, see Lahnakoski et al., 2012a). Additionally, optic flow was estimated for each frame of the movie using a Matlab toolbox (http://people.csail.mit.edu/celiu/OpticalFlow/, Liu 2009) and color intensity was estimated as the product of the mean color saturation and color value (brightness) of each frame (i.e. highly saturated bright colors received a high score and subdued and/or dark colors received a low score). Speech within the sound track of each movie segment was annotated using Audacity 1.3.12-beta (http://audacity.sourceforge.net/). Additionally, prominence of faces, biological motion and non-biological motion in the movies were annotated manually with 5-Hz sampling frequency by two independent raters using a web-based rating system (based on Nummenmaa et al., 2012). The rating was done by moving a scrollbar beside the video up or down with the mouse while watching the movies. The annotations were loaded to MATLAB, reordered to appropriate order for each condition, rounded to 10-ms resolution, and down-sampled to match the TR of the fMRI data prior to analysis.

Stimulus-feature dependent brain responses were analyzed similarly as those related to the movie clip onsets except that each stimulus feature was modeled with an event and a parametric modulator with one value for each functional volume. Correlation matrix and timecourses of the stimulus features after convolution with a canonical hemodynamic response function are shown in **Figure 2**. While most features were independent from each other, presence of faces and speech, non-biological motion and RMS, as well as optic flow and biological motion, were correlated. We performed the analyses both in independent GLMs for each feature and in a combined GLM with all features entered as separate parametric modulators without orthogonalization.
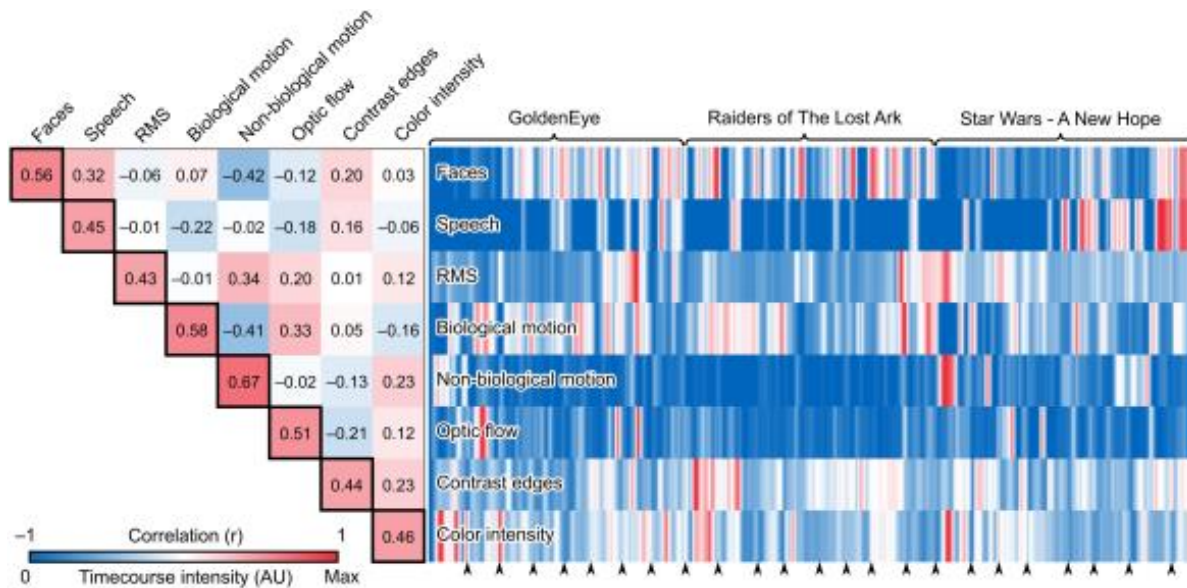
Figure 2: Correlation matrix and timecourses of the stimulus models. The pairwise correlation coefficients (*r*) of the stimulus features are presented in the upper triangle entries. Color coding indicates the correlation coefficient. The diagonal entries with black borders indicate how much of each stimulus feature model is explained by all other stimulus features together (*r*) after linear fitting. Normalized timecourses of the stimulus models convolved with a canonical HRF are shown next to the correlation matrix. Color coding indicates the amplitude of the stimulus features at each time point. The arrow heads indicate the starting points of the movie clips.

Results were similar in both approaches, suggesting that the collinearity does not confound the results.

As a validation analysis, we also replicated the analyses from a prior study (Magliano and Zacks 2011) on action-continuity processing in the brain during movie viewing. To this end, we performed a conjunction analysis of areas showing positive responses to all cut-points in the movies and negative responses to action discontinuities at a subset of the cut-points. Each first-level map was thresholded at FDR corrected q<0.05 and the binary volumes were multiplied together to produce the conjunction results. For this, all cut-points in all movies were semi-automatically extracted from the videos in Matlab. First, candidate cut-points were extracted based on root-mean-square change of pixel values between subsequent frames. These cut-points were then verified by viewing through the identified cut-intervals. Incorrectly identified cut points were then removed, and missed cut-points were added frame by frame by annotating each frame either as a cut (first frame after cut-point) or a non-cut. Then, each of the cut-points was inspected to reveal whether the action presented in the movie was continuous over the cut-point. The responses to (i) cut-points depicting discontinuous actions and (ii) all cut-points were then inserted as event onsets in GLM analyses for estimating their conjunction. The analyses were

repeated with both models in a single GLM and in two independent GLMs but the results were nearly identical due to low correlation between the models. Thus, we present only the results where the models were estimated independently.

Finally, we tested whether segmentation of the movies would break down their suspense, thus possibly leading to decreased ISC in brain regions involved in attention and emotion. To that end, we recruited 10 additional participants who rated continuously their experienced suspense while viewing the videos in either the Consecutive or Interleave condition (five observers per condition). Mean condition-wise suspense ratings were used to predict BOLD responses in a GLM as described earlier, with the exception that we used a lower high-pass filter cutoff frequency (512 s period, corresponding to the length of the full movie excerpts) due to the slowly changing suspense ratings. Given the slow and highly autocorrelated signals, we also performed the GLM analysis 20 times with temporally shifted models (shift >±5 samples; independent shift for each subject) to estimate the true null probability. We used these null analyses to build the null distribution of t-values and find the FDR corrected threshold of statistical significance (q>0.05).

Exploratory analysis of timepoints contributing to the ISC differences

To localize timepoints that were contributing to ISC differences across conditions, we reordered the timepoints in the Interleaved condition so that volumes recorded during the same moment in the videos were aligned in both conditions. We then standardized the subjects' voxelwise timecourses and computed a matrix of pairwise products of these values over subjects for each timepoint and each voxel:

$$Z_{x,y,z}(i,j) = (X_{i,x,y,z} - \mu_{i,x,y,z})/\sigma_{i,x,y,z} \circ (X_{j,x,y,z} - \mu_{j,x,y})/\sigma_{j,x,y,z}$$

where $X_{i,x,y,z}$ is the BOLD timecourse of subject i at voxel $[x,y,z]$, μ and σ are the means and standard deviations of the BOLD timecourses, and $\circ$ indicates the Hadamard (element-wise) product. Averaging these products over subject pairs and summing over time and dividing by the number of timepoints minus one directly gives the ISC value for each voxel:

$$\Sigma_{i,j,t} Z(i,j,t) / (n*(n-1)/2) / (T-1),$$

where n is the number of subjects, $n*(n-1)/2$ is the total number of subject pairs and T is the number of timepoints. Thus, the timepoints showing the largest differences are the ones that contribute most to the group difference over the entire dataset. For visualization, we performed paired t-tests over the pairs of subjects for each timepoint and voxel in the brain to find significant differences ($p<0.05$, Bonferroni corrected). Further, to highlight the instances that showed the most robust differences across groups in the regions showing group differences over the whole movie, we visualized at which time points the voxels in these regions on average showed significant differences ($p<0.05$, Bonferroni corrected by number of clusters) in the mean Z-score products over subject pairs. For a more stringent hypothesis testing without assumptions regarding the underlying distributions, we also performed a non-parametric permutation test (1000 permutations) of difference normalized by standard error of the mean (corresponding to t-values) across conditions by randomly relabeling the conditions for each subject and by building a null distribution from the observed differences. This test was performed both for individual voxels within the regions of interest (ROIs) and for mean ROI timecourses.

To simplify the exploratory visualization of the timecourses, the clusters showing significant ISC differences were first pruned by 2 mm on all sides to detach weakly connected parts from the main

clusters. Clusters smaller than 27 voxels (3x3x3 voxels) were removed to reduce the number of ROIs in the exploratory analysis. Edge-voxels that correlated poorly ($r<0.5$ in the across-subjects mean signal) with the mean timecourse of the clusters were pruned removing in total 71 voxels. This left 2586 voxels for the analysis. Finally, the voxels within the ROIs were grouped with hierarchical clustering (correlation distance, average distance linkage) to find temporally consistent regions that showed distinct synchrony patterns compared with each other. Clustering was repeated with 2–10 clusters to find the optimal number of clusters where within-cluster correlation distribution was maximally separable from across-clusters distribution (i.e. there was minimal overlap of correlation values between these two distributions). The optimal number of clusters was 4 with this measure. For visualization, the remaining clusters were inflated by 3 mm on all sides to make them better visible when plotted on the inflated brain surface.

To compare temporal distributions of peaks in the Z-score products, we calculated the kurtosis of the mean timecourse values of each cluster over both conditions, which describes the heaviness of the tails of the distribution. We also calculated the kurtosis of the across-conditions difference of the mean timecourses. To characterize how uniformly synchrony peaks were distributed over time we further calculated the burstiness (B) of the timecourses (Goh and Barabási 2008) that is defined as

$$B \equiv (\sigma_\tau - m_\tau)/(\sigma_\tau + m_\tau),$$

where m_τ and σ_τ are the mean and standard deviation of the number of events in time windows of a given size. A completely regular signal would thus have a burstiness value of -1 and a maximally bursty signal a value 1. For defining the threshold of an event (peak) we binarized the mean timecourses at Bonferroni corrected (over 4 clusters and 649 timepoints) $p<0.05$ from a surrogate distribution of a million values calculated assuming that the underlying signal values for each subject were independently drawn from a standard normal distribution. We repeated the calculation for windows of 20 to 200 samples in 20-sample intervals (20,40,...,200).

To find the most consistent peaks of synchrony across voxels for reverse correlation exploration, we performed a t-test over the voxels in the clusters at each timepoint to find occasions where differences across conditions were maximal, thus maximally contributing to the ISC difference. To

pinpoint the most robust differences, we also performed a permutation test with 1000 permutations similar to the one described above for difference in voxel and ROI timecourses over subjects.

Results

Following the movies after clip transitions was significantly more difficult in the Interleaved (M=8.56) *vs.* Consecutive (M=6.36) condition ($t(16) \approx 4.1$, $p \approx 0.001$; **Figure 3a**). Transitions within a movie (M=8.29) were also perceived as significantly more natural than those between movies *(M=1.67; $t(16) \approx 25.6$, $p \approx 2*10^{-14}$; **Figure 3b**).

Figure 4 depicts brain areas with significant ($p < 0.01$ in across-conditions ISC distribution; uncorrected) ISC during both conditions. These areas were very similar including wide bilateral areas in occipital, temporal and parietal cortex, frontal eye fields (Paus 1996) extending to ventral premotor cortex, and cerebellum.

As shown in **Figure 5,** Consecutive *vs.* Interleaved movies resulted in stronger ISC in temporal regions (left posterior STS, temporoparietal junction and middle temporal gyrus and right posterior STS), in bilateral superior and medial occipital cortices, and in the precuneus and cerebellum. The areas that showed different sensitivity to speech across conditions (white outlines in **Figure 5**; see also **Figure 7**) overlap extensively with the posterior temporal lobe areas showing significantly stronger ISC. Cluster centroid coordinates are listed in **Table I**. Interleaved *vs.* Consecutive movies caused higher ISCs in small clusters in bilateral extrastriate visual cortex and right superior temporal gyrus.

Due to the differences in difficulty of following movie *vs.* scene transitions, we also tested whether responses to the transitions were driving the ISC differences across conditions in our analyses. **Figure 6** shows brain regions responding significantly (in a GLM) to scene and movie transitions in the Consecutive and Interleaved conditions. Transitions significantly activated the precuneus, posterior cingulate cortex, lateral parietal cortex, and Crus II of the cerebellum. These areas did not overlap with the clusters showing ISC differences indicated with white outlines. We found no statistically significant differences between the two conditions ($q > 0.05$ FDR corrected) although, in the thresholded results, movie transitions in the Interleaved condition additionally correlated significantly with activity in anterior middle frontal gyrus (MFG) / frontal pole (FP) and small clusters in the anterior cingulate.
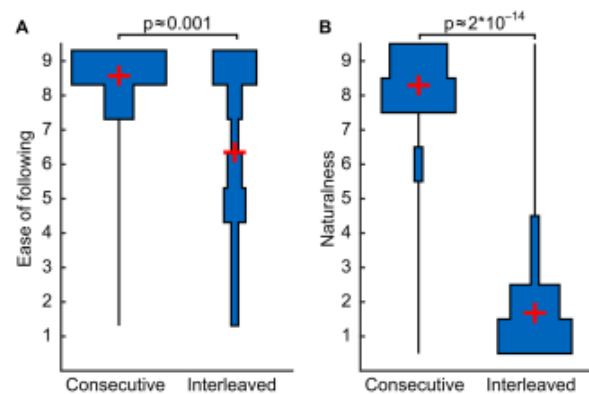

Figure 3: Distributions of self-reported ease of following events from one clip to another (A) and naturalness of movie scene transitions (B) in the Consecutive and Interleaved conditions. Distributions are depicted as symmetrical binned histogram violin plots to show the full distribution of ratings where zero is at the middle of the bars parallel to the y-axis. Empty bins are indicated by the line in the middle of the histogram (zero). *P*-values are from paired *t*-tests.
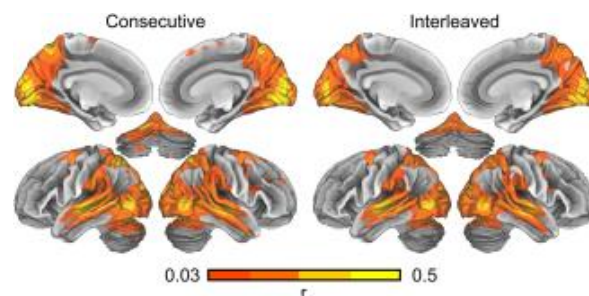

Figure 4: Brain areas showing significant inter-subject correlation during movie viewing in the Consecutive and Interleaved conditions. The data are thresholded at $P < 0.01$ (uncorrected *P*-value in the ISC distribution calculated across conditions).

Next we tested whether continuous *vs.* Interleaved presentation of the stimuli selectively influences the processing of a set of physical stimulus features (**Figure 7**). Hot colors indicate brain regions responding to different physical stimulus features over both conditions. Auditory and visual areas responded to low-level sensory features (RMS energy of soundtrack, high-frequency contrast edges, color intensity, and optic flow in the video). The loudness of the sound track also correlated with the activity of lateral and ventral visual regions, presumably due to relatively high correlations with visual features and sound RMS (see **Figure 2**). Optic flow and biological motion activated similar areas in or near the V5/MT+
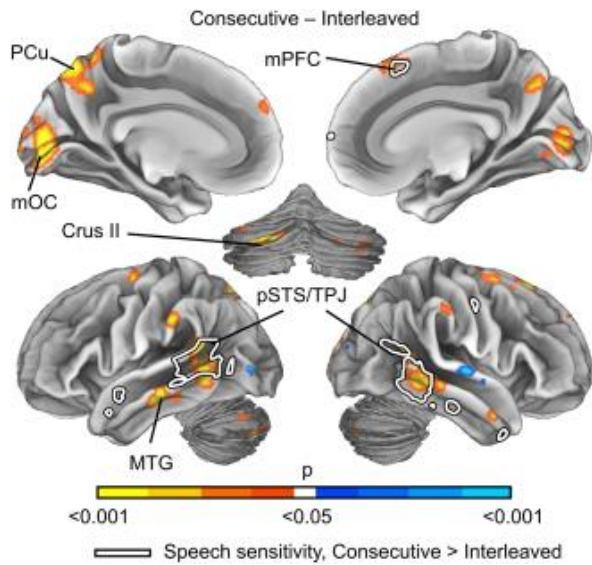
Figure 5: Brain regions showing different ISC between Consecutive and Interleaved conditions. The data are thresholded at $P < 0.05$, FWER corrected. Warm colors correspond to areas showing increased ISC in Consecutive and cold colors increased ISC in Interleaved condition. White line indicates areas showing different sensitivity to speech in Consecutive vs. Interleaved condition in Figure 7. Abbreviations: mOC, medial occipital cortex; mPFC, medial prefrontal cortex; MTG, middle temporal gyrus; PCu, precuneus; pSTS, posterior superior temporal sulcus; TPJ, temporoparietal junction.

Figure 6: Brain regions responding to scene transitions in the Consecutive (red) and Interleaved (blue) conditions ($q < 0.05$, FDR). Statistically significant effects in both conditions are shown in yellow. ISC differences from Figure 5 are indicated by white outlines. Abbreviations: FP, frontal pole; MFG, middle frontal gyrus; LPC, lateral parietal cortex; PCC, posterior cingulate cortex.

and superior occipital cortex. Speech and faces activated relatively similar areas in the temporal lobe, however, speech also activated wide spread areas of the superior temporal lobes around the primary auditory cortices whereas faces did not. Conversely, faces but not speech activated occipital regions. The brain responses to low-level visual and acoustic features were very similar in the Interleaved vs. Consecutive conditions. Significant differences were observed only for speech (filled green outline), for which especially posterior temporal responses were stronger in the Consecutive condition. These areas are highly similar to regions showing significant differences in ISC strength across conditions (see **Figure 5**).

The areas showing ISC differences were visually very similar to those reported to participate in action continuity processing in a prior study by Magliano and Zacks (2011). Therefore, we next examined whether the areas showing significant differences in ISC and speech-related responses would also be 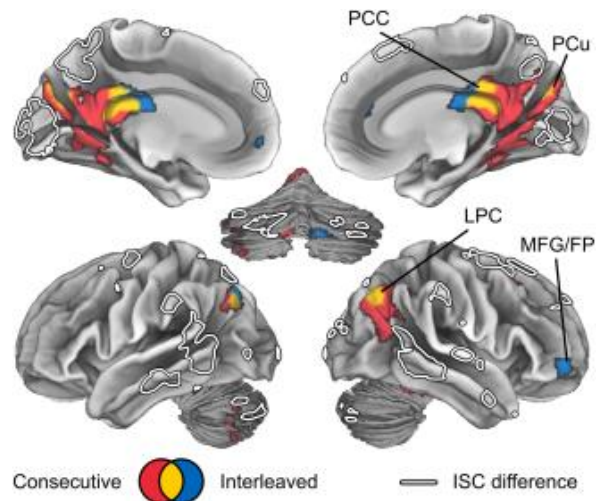selectively active during continuous actions vs. action discontinuities in our data. To this end, we replicated a conjunction analysis of Magliano and Zacks (2011) to reveal regions that were selectively activated by cut-points depicting continuous actions and deactivated by cut-points with action discontinuities. As shown in **Figure 8**, areas that were active during action continuity processing (hot colors) are partly overlapping with the areas of the pSTS and MTG that showed differences in ISC (white outlines) and speech responses (turquoise outlines) across conditions. This was, however, observed only in the Consecutive condition. In the Interleaved condition, only small bilateral clusters in the inferior bank of the STS was significantly activated by action continuity processing, yet the difference between Consecutive vs. Interleaved conditions was not statistically significant. Other areas showing ISC differences (Pcu, mPFC and Crus II) also showed negative, yet non-significant, correlations with action discontinuities. The mOC was an exception, as it was activated in all cut-points, including those where action was discontinuous (see Supplementary Figure S1).

To reveal the specific moments that triggered across-conditions differences in ISC strength, we explored the time-courses of the clusters showing significantly higher ISC in Consecutive vs. Interleaved condition. **Figure 9** shows the time-courses of products of Z-scores
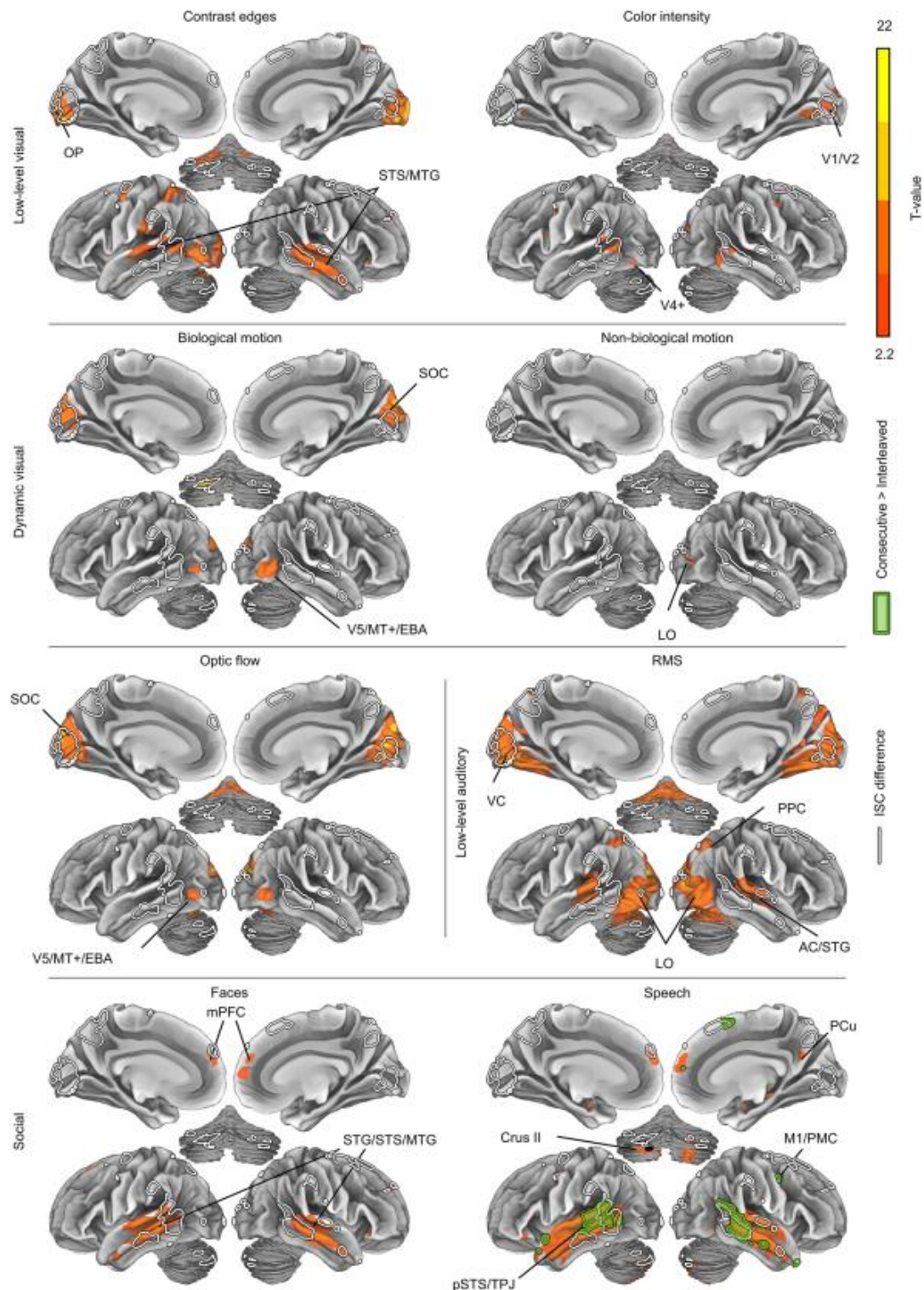
Figure 7: Brain regions responding to physical stimulus features across both conditions. Filled green outline indicates regions showing stronger stimulus-dependent responses to speech in the Consecutive versus Interleaved condition (no differences were observed to other features). White outlines indicate the areas showing significant differences in ISC across conditions. Abbreviations as in Figure 5, additionally: AC, auditory cortex; EBA, extrastriate body area; LO, lateral occipital cortex; M1, primary motor cortex; OP, occipital pole; PMC, premotor cortex; PPC, posterior parietal cortex; SOC, superior occipital cortex; STG, superior temporal gyrus; STS, superior temporal sulcus; V5/MT+, middle temporal visual area; VC, visual cortex. The data are thresholded at $q < 0.05$, FDR
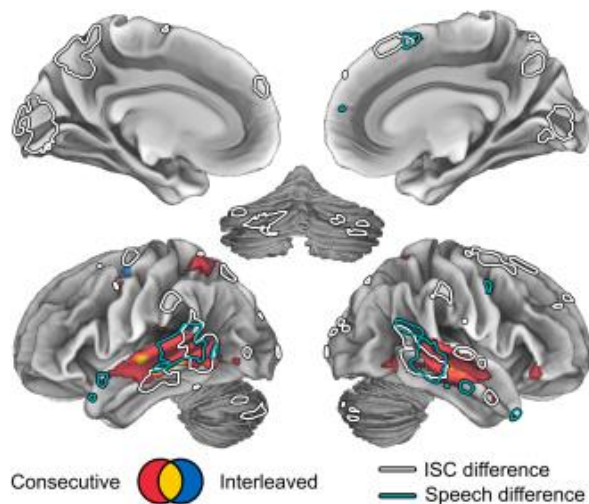
Figure 8: Areas activated while following continuous action in Consecutive (red) and Interleaved (blue) conditions. Colored areas indicate the conjunction of the responses to (positive) cut-points and (negative) action discontinuities. Overlapping areas in both conditions are indicated by yellow. Both cut-points and action discontinuities have been thresholded at $q < 0.05$ (FDR corrected). The area of conjunction corresponds to the areas where both effects overlap. Outlines indicate areas showing differences in ISC (white) and speech sensitivity (turquoise) across conditions (see Figures 5 and 7, respectively).

(instantaneous similarity) of subject pairs. The timecourses have been reorganized so that responses to the same timepoints in the stimuli are aligned in both conditions. Panel A shows the hierarchical clusters of voxels, displayed with color-coding in the correlation matrix of the timecourses. Different clusters are indicated by the same colors on the right and bottom of the matrix, and the same color-coding indicates the location of the clusters on the brain surface.

Generally, the voxels form two main groups: medial occipital cortex (yellow), and the other brain areas. The latter are further subdivided into ventral temporal cluster (green), PCu-mPFC cluster (purple) and a cluster consisting of areas of bilateral pSTS/MTG, right mPFC and left Crus II (orange). Panel B shows mean plots of Z-score products for both conditions and timepoints of significant group difference for the clusters. Overall, the yellow mOC cluster shows several short synchrony peaks and across-condition differences that are relatively uniformly distributed over time. By contrast, time courses of other clusters show longer-duration peaks

that are specific to isolated parts of the movies. This is corroborated by lower kurtosis and consistently lower burstiness of timecourses in the mOC compared with the other clusters (kurtosis of timecourses in mOC was 9.6 *vs.* 13.0 to 21.8 in other areas, mean burstiness over time window sizes -0.6 *vs.* -0.3 to -0.1; kurtosis for difference across conditions 6.7 *vs.* 7.4 to 22.0, burstiness –0.6 *vs.* –0.4 to –0.2). Thus, the relatively long periods of significant group differences in the pSTS/MTG, Crus II, PCu, mPFC and vTC presumably corresponded to events that are particularly differently processed when task switching is required. Panel C shows the consistent between-condition differences (in t-tests) in the orange, purple and green clusters that assumedly show moments of content-specific synchrony. These moments are described in **Table II** and depicted in **Video 1**. They contain mainly instances of conflict, deception and escape, and also include the opening scenes of the second and third movies when subjects start following a new long narrative. Timepoints that survived a more conservative permutation test are marked with an asterisk. Instances of higher ISC in Consecutive *vs.* Interleaved conditions that were caused by increased deactivations in Consecutive condition are marked with minus signs in panel C and **Table II** (see red and blue dashed lines in **Figure 9 C**).

Finally, because the participants might have experienced higher suspense during the continuous movies we investigated the role of suspense in modulating ISC. **Figure 10 A** shows the areas where brain activity correlated with the suspense ratings across both conditions (positive: parametric in green, non-parametric in hot colors; negative: parametric in purple, non-parametric in cold colors). Statistically significant differences across conditions inside the areas showing significant suspense-related activity are indicated by red (Consecutive>Interleaved) and blue (Interleaved>Consecutive) outlines. The areas showing ISC differences are outlined in white. Suspense ratings correlated positively with nearly all regions showing significant ISCs in the standard parametric t-test. However, only regions in the vicinity of V5/MT+, FEF, posterior parietal cortex and opercular parietal cortical regions survived the non-parametric test. Negative responses were seen in lateral parietal and anterior cingulate cortices (ACC), but only a small cluster in the right ACC was significant in the non-parametric test. Figure 10 B shows the timecourses of the suspense ratings in Consecutive (red) and Interleaved (blue) conditions. The scenes showing highest ISC differences in **Figure 9 C** are indicated by gray bars at the bottom of the plot.
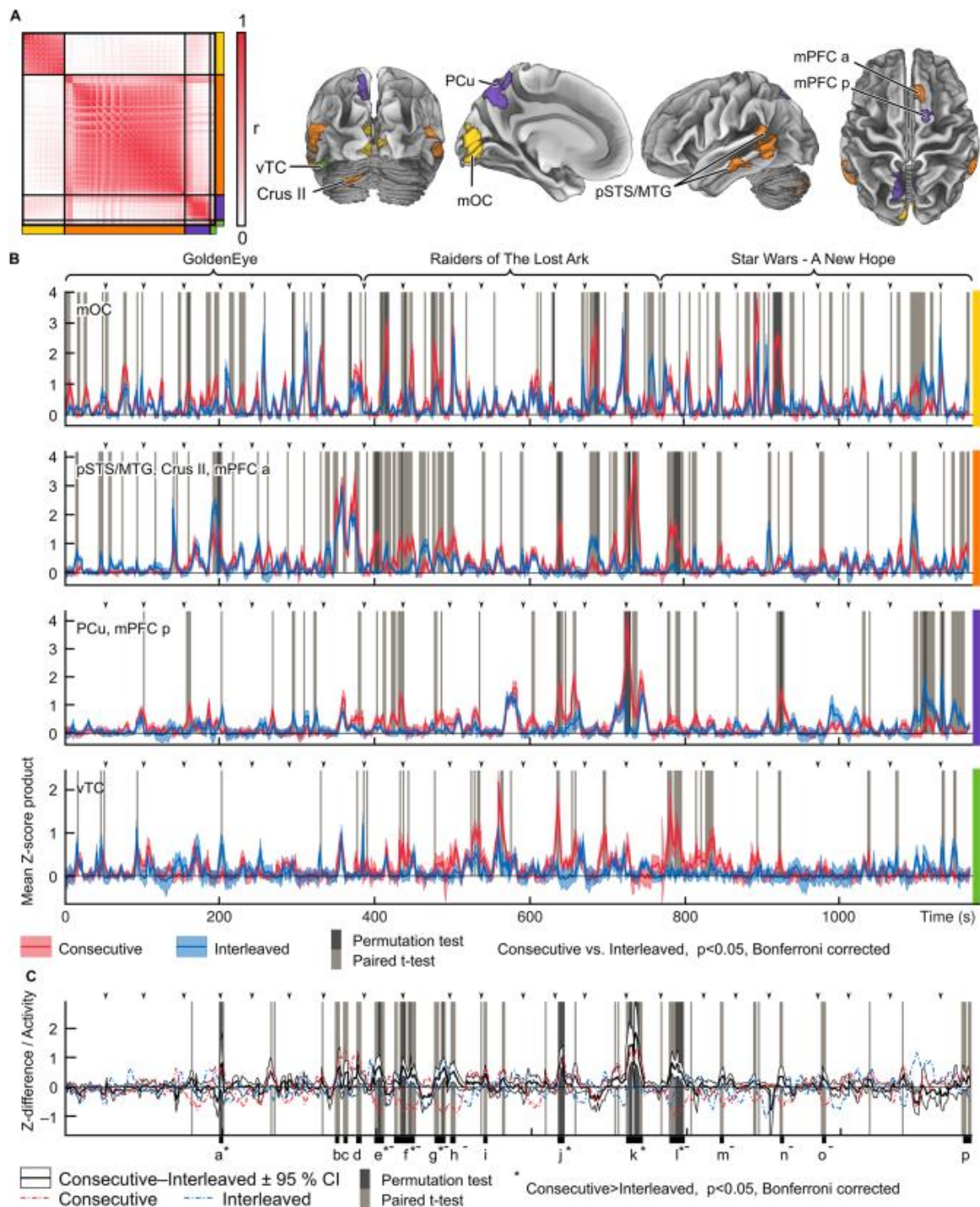
10

Figure 9: Timecourses of products of Z-scores over subject pairs (instantaneous similarity) contributing to the mean ISCs and ISC group differences. A: (Left) Correlation matrix and clustering of the voxels showing significantly higher ISC in Consecutive versus Interleaved condition. The different clusters are separated by black lines and indicated by the colors on the right and bottom of the matrix. (Right) Locations of the clusters are indicated by colors on the brain visualizations on the right. Abbreviations as in Figure 5, in addition: vTC, ventral temporal cortex, dmPFC a/p, dorsomedial prefrontal cortex, anterior/posterior. B: Mean timecourses of Z-score products across subject pairs for the clusters. Consecutive condition is show in red and Interleaved condition in blue. Timepoints where the mean timecourse of each cluster showed significant ($P < 0.05$, Bonferroni corrected; t-test light gray, permutation test dark gray) between-condition difference are indicated with gray vertical bars. Arrow heads indicate the onset times of each short movie segment and the full duration of each movie excerpt is indicated by the braces on top of the plots. The clusters are indicated by the colored bars on

(Figure 9 continued) the right side of the plots. C: Mean difference of Z-score products between conditions across all voxels in the orange, purple and green clusters. Significant differences ($P < 0.05$, Bonferroni corrected) are indicated by gray bars (t-test result with lighter and permutation test results in darker gray). Red and blue dashed lines indicate the mean standardized activity over subjects and voxels in the orange, purple and green clusters. Time windows marked with a–p indicate instances where the mean timecourses showed most consistent differences across all voxels in the three clusters. Windows shorter than 3 TRs have been omitted and windows that are separated by less than 3 TRs have been collapsed into one longer window. The error bars on all plots indicate 95% confidence intervals. Time windows containing timepoints that were significant in the permutation test are indicated by the asterisks (*). Timepoints of significant difference across conditions caused by deactivations in the Consecutive condition are marked by minus (−) signs.

Discussion

Our main finding was that when the natural flow of action sequences was interrupted and frequent switching between parallel event-sequences was required, the inter-subject reliability of brain responses was decreased in the bilateral posterior STS region and left MTG, mOC and smaller clusters in the mPFC and cerebellum (**Figure 5**). Particularly, responses associated with speech in the posterior temporal lobe were significantly weaker during interleaved movies (**Figure 7**). By contrast, low-level sensory responses were similar in both conditions. Explorative timepoint-by-timepoint reverse correlation comparison revealed that events depicting social conflict and deception contributed strongly to the ISC group difference in temporal lobes, medial prefrontal cortex and Crus II. By contrast, differences in the occipital visual areas were diffusely distributed in time throughout the movie stimuli (**Figure 9**). Our results suggest that i) temporal interleaving specifically modulates processing of social aspects of stimuli, particularly speech, in the STS/MTG, Crus II, vTC and mPFC, and that ii) early sensory processes are either not affected or, in the case of early visual cortex, show general reduction of ISC that is not specific to particular events.

Temporal integration of events

The ISC maps in both Consecutive and Interleaved conditions were visually very similar (**Figure 4**) and accord with those reported in prior studies (see e.g. Hasson et al., 2004; Hasson et al., 2008a; Lerner et al., 2011). However, statistically testing revealed that shifting back and forth between the different movies markedly reduced ISC in occipitotemporal cortices (pSTS/TPJ) occipital cortex and precuneus (**Figure 5**). Prior research has established that these regions maintain a temporally consistent model of action sequences depicted in movies across "continuity edits" that depict the same continuous action from different camera angles; such edits are continuous over time, space and action. This is in contrast with edits with a discontinuity, such as between two consecutive shots that show different locations or people. Discontinuous edits where place,

time or action changed during the transition showed little consistent activity (Magliano and Zacks 2011). In line with this, shuffling a movie using discontinuity edits markedly lowers ISC in higher-order associative and frontal cortices across observers (Hasson et al., 2008a; Lerner et al., 2011). In contrast, adding continuity edits to depiction of continuous action actually increases ISC in the occipital, ventral temporal and parietal cortices (Herbec et al., 2015) leaving ISCs in the posterior temporal regions unaffected. Indeed, we did not observe significant differences in the event-related responses to scene onsets when the clips were preceded by the previous clip of the same movie (Consecutive condition) *vs.* a clip from a different movie (Interleaved condition; **Figure 6**) despite the reported decrease of naturalness and ease of following the movie *vs.* scene transitions (**Figure 3**). Moreover, most subjects found following the events in the Interleaved condition relatively easy and the event related activity due to movie/scene transitions was markedly different from the foci showing differences across conditions in ISC and speech-related GLM analysis. This suggests that the circuitry activated by scene transitions between movies is also activated similarly at scene transitions during the natural flow of a movie, which also requires reorientation and updating the contents of working memory (Kurby and Zacks 2008). Together with the present results (see **Figure 5**), these findings suggests that the pSTS region, precuneus, ventral temporal and medial prefrontal cortices and part of Crus II are involved in building and maintaining a temporally extended model of observed events.

However, simply following continuity of actions between movie cut-points cannot explain the observed differences in brain responses between conditions, although areas revealed by the conjunction analysis partially overlapped those showing ISC differences in the pSTS/MTG (no differences across conditions; **Figure 8**). Moreover, the regions of the temporal lobe showing the greatest deactivations to action discontinuities extend anteriorly from the regions showing ISC/speech-sensitivity differences and those reported
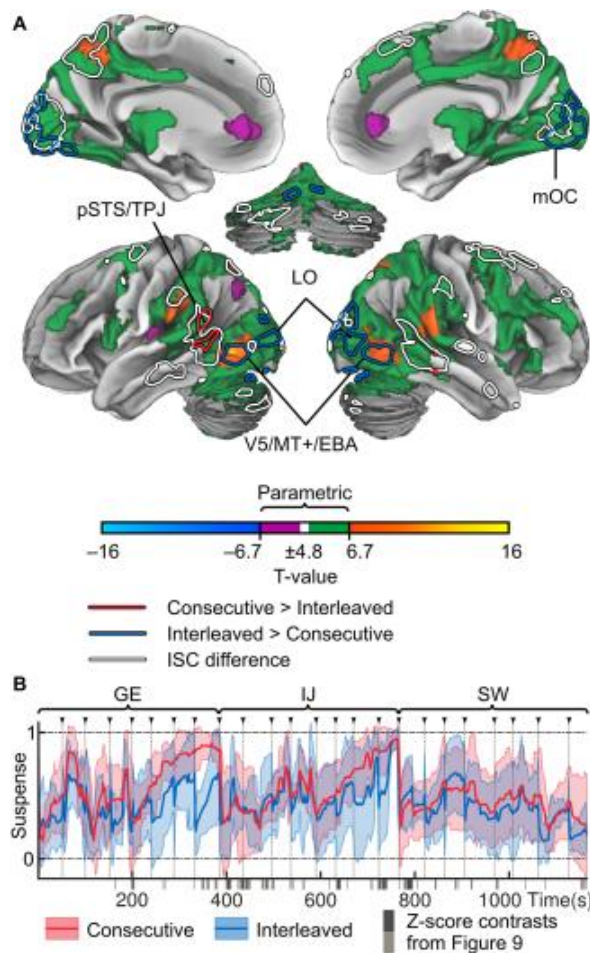
Figure 10: Brain regions responding to experienced suspense. A: Main effects of suspense in a GLM analysis across both conditions. Due to the autocorrelation structure of the slow signals in the model, the results are thresholded ($P < 0.05$, FDR corrected) based on both standard parametric t-test and null GLMs where the models were circularly shifted. Parametric effects are shown in green (positive) and purple (negative) tones. Differences across conditions in the GLM results are shown in red (Consecutive > Interleaved) and blue (Interleaved > Consecutive) outlines. White outlines indicate areas showing ISC differences in Figure 5. B: Mean ratings (±95% CI) of suspense in the Consecutive (red) and reordered Interleaved (blue) condition. Movie clip onsets are indicated by arrows and gray dashed lines. Gray bars at the bottom of the plot show the differences across conditions in Figure 9C. Abbreviations as in Figures 5–7.

previously (Magliano and Zacks 2011). Furthermore, no significant effects were observed in the precuneus and posterior parietal cortex, which have previously been associated with continuity processing (Magliano and Zacks 2011). This may depend on the definition of discontinuity or the complexity of the stimuli. For example, the Magliano and Zacks (2011) study used a children's movie with relatively simple actions and practically no speech, while we found that speech consistently activated the anterior parts of STS/MTG that also showed continuity effects in our data. Altogether these data suggest that not all aspects of action continuity processing are similarly affected by action sequence interleaving in the timescales studied here.

The capacity for prediction and temporal integration may be one of the key properties that allow the pSTS region to serve as a 'hub' for social processing (Lahnakoski et al., 2012b). The pSTS/TPJ region has been proposed to participate in several functions such as integration of "what" and "where" information (Allison et al., 2000), inference of intentionality (Nummenmaa and Calder 2009) and perception of social awkwardness in naturalistic scenes (Pantelis et al., 2015). These functions could be subserved by a more general function of updating a model of the concurrent situational context used to predict the upcoming events and detect prediction violations. Similar predictive schemes have been proposed to underlie comparison of "efference copies" from the motor/premotor system to current percepts in perception-action cycles, for example, during speech production (for a review, see e.g. Rauschecker and Scott 2009). Updating predictions of the current situation is a continuous process, which may be interrupted by a sudden change in stimulus events. This, in turn, may lead to reduction in the response reliability across observers.

Social and sensory feature driven brain responses

We found that speech-sensitive areas in the pSTS/TPJ showed greatest interleaving effects, and presence of human speech correlated more with the activity in these areas in the Consecutive *vs.* Interleaved condition (**Figure 7**). By contrast, the low-level stimulus features (visual motion, contrast, color, and sound loudness) were associated with activation of the corresponding sensory regions in both conditions, with no differences across conditions in the GLM analysis. This suggests that the observed differences in brain activity may relate specifically to processing of language and possibly other socially relevant movie content. However, our stimulus model is not exhaustive and we cannot rule out that processing of non-modeled low-level features could also be affected due to top-down influences, such as attention (see e.g. Çukur et al., 2013), although the stimulus material itself was the same in both conditions.

13

Presence of speech also predicted activity in the Crus II of the cerebellum whose synchrony timecourse was similar to that in pSTS. This is in accordance with the emerging view of cerebellar contribution to social processing (Van Overwalle et al., 2014). In speech perception, cerebellum appears to participate in temporal, or sequential, prediction (for a review, see Moberget and Ivry 2016). This finding supports its role in prediction and following of ongoing social events, suggested here.

Reverse correlation analysis

In the pSTS/MTG, mPFC, vTC and Crus II, reduced ISC in Interleaved *vs.* Consecutive condition was mainly explained by a few scenes and not by either constantly higher ISC or differences at the transition moments between movie clips. Reverse correlation analysis revealed that the scenes showing most consistent between-conditions differences in the temporal and frontal regions contained most often deception or conflict (e.g. a man sneaking up and pulling a gun behind Indiana Jones before being disarmed by him; discussion about princess Leia hiding plans in an escape pod after she was questioned).

Half of the scenes showing significantly higher ISCs in Consecutive *vs.* Interleaved condition were driven by stronger deactivations rather than activations in the Consecutive *vs.* Interleaved condition. ISC differences in all of the scenes containing speech were driven by high activation peaks, as suggested by the stronger speech related responses. In contrast, none of the scenes where ISC difference was driven by deactivation contained speech. While some of the peaks of synchrony in the precuneus coincide with those in the speech-sensitive areas, we did not see speech-related responses in those parts of precuneus that show ISC differences. The synchrony peaks in precuneus were infrequent compared with those in the speech sensitive regions and thus difficult to interpret alone (see **Figure 9 C** for timepoints, **Table II** for scene descriptions and **Video 1** for excerpts from the marked timepoints and the associated synchrony of brain activity). Importantly, since the scenes with the greatest ISC differences do not consistently coincide with movie transitions in the Interleaved condition, there is no obvious need for integrating information differently during these scenes in the two conditions. This suggests that there is a more general difference in the mode of information processing of these brain regions in the two conditions that warrants further study in the future.

In contrast to the areas discussed above, the mOC showed ISC differences throughout the experiment in activity peaks that were relatively uniformly distributed in time. While most of these peaks were of short duration, some peaks lasted for several seconds. Most prominent peaks correspond to scenes with significant change in lighting (lit torch, golden glow of an idol and ending the shooting of lasers). However, these are not the only scenes with prominent lighting changes. Additionally, in contrast to the temporal lobes, suspense modulated the activity in the visual cortex more strongly in the Interleaved *vs.* Consecutive condition. This result was in a direction opposite to the ISC results, and the difference in suspense-related brain activity was slightly more posterior than the areas showing differences in ISC.

Brain responses to suspense during movie viewing

Scenes showing highest differences in ISC timecourses also contained moments of high suspense. In the GLM analysis, suspense was associated with activity in corticolimbic areas whose ISC and BOLD-GLM responses is modulated by affective arousal (Nummenmaa et al., 2012; Nummenmaa et al., 2014), which is an important component of suspense experience (Lehne and Koelsch 2015). However, due to the slowness of the signals, the parametric test was very liberal. With a stricter FDR corrected non-parametric threshold, the significant effects were restricted to the dorsal attention network (posterior parietal cortex and areas surrounding the bilateral middle temporal visual areas and frontal eye fields, respectively), according with the well-established influence of affective significance on attention (Vuilleumier 2005).

Suspense ratings also differed across the conditions, but of the areas showing ISC differences, only small parts of the left posterior temporal lobe showed higher suspense-related responses in the Consecutive *vs.* Interleaved condition. Moreover, the differences in ISC did not seem to coincide exclusively with moments of high suspense or with moments where the ratings of suspense differed across conditions. Therefore, we conclude that suspense may have contributed to the differences seen in the ISCs and speech responses between conditions in the temporal lobes, but it is not sufficient to explain all of the differences seen in other analyses.

Time scales of integration and memory

The time windows studied here were relatively long in comparison to time-scales studied for lower-level integration effects such as change detection in the auditory cortex at a millisecond scale (Mustovic et al., 2003) or behavioral accounts of event integration in videos in the order of 2–3

seconds (Fairhall et al., 2014). Prior fMRI studies employing scrambling the temporal order of video clips or audio stories (Hasson et al., 2008a; Lerner et al., 2011) have found that the correlation of activity in areas of the posterior temporal lobes, parts of mPFC and precuneus is decreased – either across or within subjects – when the stimuli are scrambled in short *vs.* long time windows. Temporal integration windows in fMRI data seems to increase up to timescales of 30–40 seconds, where the areas of significant ISC appear indistinguishable from those elicited by intact videos. However, by testing the magnitude of ISC statistically, we demonstrated that even at scales of 50 seconds, brain processing of stimuli is affected even if the regions in the thresholded ISC maps are visually very similar. Moreover, in contrast to the studies employing temporal scrambling, the subjects of the current study could still follow the storylines in their entirety, as the order of movie parts remained unchanged for each stimulus. Additionally, earlier studies did not test which aspects of the stimuli were integrated over long time periods in these brain areas. Our results demonstrate that even though similar levels of ISC differences may be seen in multiple brain regions, different areas may have different profiles that account for the increased ISC. Thus, in future studies, it is important to consider what particular aspects of stimuli are integrated in each area.

Finally, it is possible that differences in ISC may be related to subsequent memory of the events. Earlier results suggest that improved memory encoding during movie viewing is linked with increased ISC (Hasson et al., 2008b) in STS and TPJ that partly overlap with the areas showing ISC differences in the current study. It is thus possible that some of the differences reported here are related to the encoding to or retrieval from long-term memory.

Conclusions

Following multiple interleaved events forces the brain to constantly reorient to a new scene, thus reducing the predictability of the events. We conclude that posterior temporal, medial parietal and lateral frontal cortical areas are the key regions participating in the integration of ongoing events. We propose that these regions build and maintain a temporally extended predictive model of ongoing sequence of events.

Acknowledgements

Supplementary information

Supplementary video:

http://onlinelibrary.wiley.com/store/10.1002/hbm.23591/asset/supinfo/hbm23591-sup-0001-suppinfo1.mpg?v=1&s=f57395acd3d4f6ef458aa288be526b745e974364

Suplementary figure 1:

http://onlinelibrary.wiley.com/store/10.1002/hbm.23591/asset/supinfo/hbm23591-sup-0001-suppinfo2.tiff?v=1&s=926761c6c2bf135d1bfabcd28774466ff73999a1

References

Allison T, Puce A, and McCarthy G (2000): Social perception from visual cues: role of the STS region. Trends Cogn Sci 4:267–278.

Bailey HR, Kurby CA, Giovannetti T, and Zacks JM (2011): Action perception predicts action performance. Neuropsychologia 51:2294–2304.

Çukur T, Nishimoto S, Huth AG, and Gallant JL (2013): Attention during natural vision warps semantic representation across the human brain. Nat Neurosci 16:763–770.

Fairhall SL, Albi A, and Melcher D (2014): Temporal integration windows for naturalistic visual sequences. PLoS One 9:e102248.

Goh K-I and Barabási A-L (2008): Burstiness and memory in complex systems. Europhys Lett 81:48002.

Hasson U, Chen J, and Honey CJ (2015): Hierarchical process memory: memory as an integral component of information processing. Trends Cogn Sci 19:304–313.

Hasson U, Furman O, Clark D, Dudai Y, and Davachi L (2008b): Enhanced intersubject correlations during movie viewing correlate with successful episodic encoding. Neuron 57:452–462.

Hasson U, Nir Y, Levy I, Furhmann G, and Malach R (2004): Intersubject synchronization of

cortical activity during natural vision. Science 303:1634-1640.

Hasson U, Yang E, Vallines I, Heeger DJ, and Rubin N (2008a): A hierarchy of temporal receptive windows in human cortex. J Neurosci 28:2539–2550.

Herbec A, Kauppi J-P, Jola C, Tohka J, and Pollick FE (2015): Differences in fMRI intersubject correlation while viewing unedited and edited videos of dance performance. Cortex In press:doi:10.1016/j.cortex.2015.06.026.

Jenkinson M, Bannister PR, Brady JM, and Smith SM (2002): Improved optimisation for the robust and accurate linear registration and motion correction of brain images. NeuroImage 17:825–841.

Kauppi J-P, Pajula J, and Tohka J (2014): A versatile software package for inter-subject correlation based analyses of fMRI. Front Neuroinform 8:doi: 10.3389/fninf.2014.00002.

Kurby CA and Zacks JM (2008): Segmentation in the perception and memory of events. Trends Cogn Sci 12:72–79.

Lahnakoski JM, Glerean E, Jääskeläinen IP, Hyönä J, Hari R, Sams M, and Nummenmaa L (2014): Synchronous brain activity across individuals underlies shared psychological perspectives. NeuroImage 100:316–324.

Lahnakoski JM, Glerean E, Salmi J, Jääskeläinen IP, Sams M, Hari R, and Nummenmaa L (2012b): Naturalistic fMRI Mapping Reveals Superior Temporal Sulcus as the Hub for the Distributed Brain Network for Social Perception. Front Hum Neurosci 6:233.

Lahnakoski JM, Salmi J, Jääskeläinen IP, Lampinen J, Glerean E, Tikka P, and Sams M (2012a): Stimulus-Related Independent Component and Voxel-Wise Analysis of Human Brain Activity during Free Viewing of a Feature Film. PLoS One 7:e35215.

Lehne M and Koelsch S (2015): Toward a general psychological model of tension and suspense. Front Psychol 11:doi:10.3389/fpsyg.2015.00079.

Lerner Y, Honey CJ, Silbert LJ, and Hasson U (2011): Topographic mapping of a hierarchy of temporal receptive windows using a narrated story. J Neurosci 31:2906–2915.

Liu C (2009). Beyond Pixels: Exploring New Representations and Applications for Motion Analysis. Massachusetts Institute of Technology.

Magliano JP and Zacks JM (2011): The impact of continuity editing in narrative film on event segmentation. Cognitive Sci 35:1–29.

Moberget T and Ivry RB (2016): Cerebellar contributions to motor control and language comprehension: searching for common computation principles. Ann NY Acad Sci 1369:154–171.

Mustovic H, Scheffler K, Di Salle F, Esposito F, Neuhoff JG, Hennig J, and Seifritz E (2003): Temporal integration of sequential auditory events: silent period in sound pattern activates human planum temporale. NeuroImage 20:429–434.

Nummenmaa L and Calder AJ (2009): Neural mechanisms of social attention. Trends Cogn Sci 13:135–143.

Nummenmaa L, Glerean E, Viinikainen M, Jääskeläinen IP, Hari R, and Sams M (2012): Emotions promote social interaction by synchronizing brain activity across individuals. Proc Natl Acad Sci USA 109:9599–9604.

Nummenmaa L, Saarimäki H, Glerean E, Gotsopoulos A, Jääskeläinen IP, Hari R, and Sams M (2014): Emotional speech synchronizes brains across listeners and engages large-scale dynamic brain networks. NeuroImage 102:498–509.

Pantelis PC, Byrge L, Tyszka JM, Adolphs R, and Kennedy DP (2015): A specific hypoactivation of right temporo-parietal junction/posterior superior temporal sulcus in response to socially awkward situations in autism. Soc Cogn Affect Neurosci doi: 10.1093/scan/nsv021.

Paus T (1996): Location and function of the human frontal eye-field: A selective review. Neuropsychologia 34:475–483.

Rauschecker JP and Scott SK (2009): Maps and streams in the auditory cortex: nonhuman primates illuminate human speech processing. Nat Neurosci 12:718–724.

Smirnov D, Glerean E, Lahnakoski JM, Salmi J, Jääskeläinen IP, Sams M, and Nummenmaa L (2014): Fronto-parietal network supports context-dependent speech comprehension. Neuropsychologia 63:293–299.

Smith SM (2002): Fast robust automated brain extraction. Hum Brain Mapp 17:143–155.

Smith SM, Jenkinson M, Woolrich MW, Beckmann CF, Behrens TEJ, Johansen-Berg H, Bannister PR, De Luca M, Drobnjak I, Flitney DE, Niazy

RK, Saunders J, Vickers J, Zhang Y, De Stefano N, Brady JM, and Matthews PM (2004): Advances in functional and structural MR image analysis and implementation as FSL. NeuroImage 23:S208–S219.

Van Essen DC, Dickson J, Harwell J, Hanlon D, Anderson CH, and Drury HA (2001): An Integrated Software System for Surface-based Analyses of Cerebral Cortex. J Am Med Inform Assoc 8:443–459.

Van Overwalle F, Baetens K, Mariën P, and Vandekerckhove M (2014): Social cognition and the cerebellum: a meta-analysis of over 350 fMRI studies. NeuroImage 86:554–572.

Vuilleumier P (2005): How brains beware: neural mechanisms of emotional attention. Trends Cogn Sci 9:585–594.

Woolrich MW, Jbabdi S, Patenaude B, Chappell M, Makni S, Behrens T, Beckmann C, Jenkinson M, and Smith SM (2009): Bayesian analysis of neuroimaging data in FSL. NeuroImage 45:S173–S186.

Zacks JM, Braver TS, Sheridan MA, Donaldson DI, Snyder AZ, Ollinger JM, Buckner RL, and Raichle ME (2001): Human brain activity time-locked to perceptual event boundaries. Nat Neurosci 4:651–655.

Zacks JM, Kurby CA, Eisenberg ML, and N Haroutunian (2011): Prediction error associated with the perceptual segmentation of naturalistic events. J Cogn Neurosci 23:4057–4066.

Zacks JM, Speer NK, Swallow KM, and Maley CJ (2010): The brain's cutting-room floor: segmentation of narrative cinema. Front Hum Neurosci 4:doi: 10.3389/fnhum.2010.00168.