

## ORIGINAL ARTICLE

# Encoding of Auditory Temporal Gestalt in the Human Brain

Michael P. Notter<sup>1,2,3</sup>, Michael Hanke<sup>4,5</sup>, Micah M. Murray<sup>1,2,3,6,7</sup>  
and Eveline Geiser<sup>1,2,8</sup>

<sup>1</sup>Department of Radiology, <sup>2</sup>Neuropsychology and Neurorehabilitation Service, <sup>3</sup>EEG Brain Mapping Core, Center for Biomedical Imaging (CIBM), Lausanne University Hospital and University of Lausanne, 1011 Lausanne, Switzerland, <sup>4</sup>Institute of Psychology, Otto-von-Guericke-University, <sup>5</sup>Center for Behavioral Brain Sciences, 39106 Magdeburg, Germany, <sup>6</sup>Ophthalmology Department, University of Lausanne and Fondation Asile des Aveugles, 1003 Lausanne, Switzerland, <sup>7</sup>Department of Hearing and Speech Sciences, Vanderbilt University, Nashville, TN 37203-5721, USA and <sup>8</sup>McGovern Institute, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

Address correspondence to Eveline Geiser, University Hospital Center and University of Lausanne, Department of Radiology and Department of Neuropsychology and Neurorehabilitation, Avenue Pierre Decker 5, NES0.098, CH-1011 Lausanne, Switzerland. eveline.geiser@unil.ch

## Abstract

The perception of an acoustic rhythm is invariant to the absolute temporal intervals constituting a sound sequence. It is unknown where in the brain temporal Gestalt, the percept emerging from the relative temporal proximity between acoustic events, is encoded. Two different relative temporal patterns, each induced by three experimental conditions with different absolute temporal patterns as sensory basis, were presented to participants. A linear support vector machine classifier was trained to differentiate activation patterns in functional magnetic resonance imaging data to the 2 different percepts. Across the sensory constituents the classifier decoded which percept was perceived. A searchlight analysis localized activation patterns specific to the temporal Gestalt bilaterally to the temporoparietal junction, including the planum temporale and supramarginal gyrus, and unilaterally to the right inferior frontal gyrus (pars opercularis). We show that auditory areas not only process absolute temporal intervals, but also integrate them into percepts of Gestalt and that encoding of these percepts persists in high-level associative areas. The findings complement existing knowledge regarding the processing of absolute temporal patterns to the processing of relative temporal patterns relevant to the sequential binding of perceptual elements into Gestalt.

In everyday hearing, humans are able to recognize a familiar rhythm, for example, the rhythm of a song, independently of the tempo at which it is played. This is exemplary of the fact that all auditory experiences unfold in time, comprising percepts of auditory Gestalt based on the rhythmic structure of sounds. The pattern resulting from the relative temporal relationship between acoustic events is the basis for the perceptual emergence of temporal Gestalt. This Gestalt percept is invariant to the absolute temporal relationship between constituent acoustic events, but depends on the relative temporal relationship

between them (Hulse et al. 1992). It is unknown, where in the brain percepts of temporal Gestalt are generated.

Temporal information influences audition across domains from the perception and identification of general environmental sounds, referred to as auditory objects (Griffiths and Warren 2004), their localization and movement detection in space (Carlile and Leung 2016) to speech (Poeppel 2014) and music (Grahn 2012) perception. Rhythmic patterns carry emotional information in speech (Trost et al. 2017), inform interpersonal communication (Keller et al. 2014), and are a key factor influencing auditory

stream segregation (Bendixen 2014). Specifically, rhythmic patterns define perception by influencing sensory processing in speech and nonspeech sounds (Bonte et al. 2009; Rimmele et al. 2011; Arnal and Giraud 2012; Geiser et al. 2012; Geiser and Gabrieli 2013; Bendixen et al. 2015; Kotz and Schmidt-Kassow 2015; Schwartz and Kotz 2016; Ghitza et al. 2012; Arnal et al. 2015) possibly through predictive mechanisms. Thus, rhythmic temporal patterns guide our listening and scaffold the auditory experience across domains of audition.

How the brain represents auditory temporal Gestalt is, to our knowledge, unknown. The processing of absolute temporal aspects in sounds, such as the discrimination of two rhythms at a given tempo, is associated with activity in the supplementary motor area, the dorsolateral prefrontal cortex and basal ganglia (Schubotz et al. 2000; Rao et al. 2001; Belin et al. 2002; Bengtsson et al. 2009; Geiser et al. 2012; Konoike et al. 2012; Thaut et al. 2014; Teki et al. 2011; Chen et al. 2008; Grahn and Rowe 2009). While these brain areas are involved when two sequences with two specific absolute temporal patterns are compared, the representation of the related temporal Gestalt might lay within these or other brain areas. In a recent paper, it was hypothesized how the brain could encode relative temporal patterns (Geiser et al. 2014). The time-locked neural activity to each acoustic event could first provide a passive representation of the global temporal pattern, potentially in sensory cortices. Then, higher-level, relative timing computations on this temporal representation binding them into percepts of Gestalt could be performed by distributed neuronal ensembles.

The challenge of investigating auditory temporal Gestalt, like many other Gestalt percepts, is the nonlinearity of its emergence linked to its perceptual invariance. That is, the emergence of the percept is not directly proportional to its constituent sensory input. Consequently, evidence on Gestalt processing is often derived from the two perceptual extremes, thus comparing the processing of a Gestalt with the processing of non-Gestalt (Keil et al. 1999; Altmann et al. 2003; Grassi et al. 2016). While this approach reveals brain areas involved in the processing of Gestalt, it cannot identify where in the brain the Gestalt percept is itself represented. The loci of representation can only be revealed by comparing two different Gestalt percepts. At the same time, the percepts must be investigated independently of their sensory basis, by matching for sensory input so that activation differences do not represent sensory input differences, and must also be induced equally by different sensory sequences. In the present study, we are able to investigate tempo invariant encoding of temporal Gestalt in the brain by applying a machine-learning pattern classification approach to fMRI data.

## Materials and Methods

### Participants

A total of 19 participants performed the experiment (9 females, aged  $26.0 \pm 5.23$  years; mean  $\pm$  SD, min/max = 21/41). Participants gave written informed consent in accordance with procedures approved by the local committee on the use of humans as experimental participants and according to the World Medical Association Helsinki Declaration as revised in October 2008. Participants were paid for their participation. All participants were right-handed, according to the Annett-Handedness-Questionnaire (Annett 1992), and reported no history of neurological, psychiatric, or hearing disorders.

### Stimuli and Design

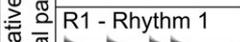
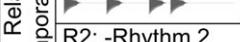
Participants listened to sound sequences and responded to questions presented after each sound sequence. The auditory

stimuli contained 4 consecutive tones with a duration of 200 ms and were constructed by means of Audacity (2.0.6). An isochronous baseline condition (BA) was constructed in which the temporal interval between all tones was identical. Two categories of temporally grouped experimental conditions (R1 and R2) represented each a specific rhythm, thus, a specific relative temporal pattern (Fig. 1). Both categories started with the same temporal “base” interval and comprised an interval that was shortened and another that was lengthened (each by 50% of the “base” interval). Both BA and experimental conditions R1 and R2 were presented with 3 different absolute temporal intervals as “base” interval ( $t_1 = 300$ ,  $t_2 = 400$ ,  $t_3 = 500$  ms) resulting in a tempo of 1.67 ( $t_3$ ), 2 ( $t_2$ ), and 2.5 ( $t_1$ ) beats-per-second, respectively. Thus, the stimulus material consisted of a total of  $3$  (BA, R1, R2)  $\times$   $3$  ( $t_1$ ,  $t_2$ ,  $t_3$ ) conditions. To induce perceptual variance,  $t_1$  was constructed with a drum sound (264 Hz base frequency) and  $t_2$  and  $t_3$  were constructed with a pluck sound (261 Hz base frequency). Each condition was presented 55 times. All presented stimuli were normalized based on the root-mean-square amplitude.

Auditory stimulus sequences were presented through fMRI-compatible insert earphones by Sensimetrics (<http://www.sens.com/>). Participants' ears were covered with foam ear-defenders for comfort. Stimulus sequences were presented at a comfortable listening level and participants' responses were recorded using Presentation software (Neurobehavioral Systems, <http://www.neurobs.com/>). For each participant 5 runs of stimuli were acquired, each comprising 99 experimental stimuli and 22 empty trials. Each trial was preceded by a fixation cross presented for 300 ms followed by a blank screen for 100 ms. The order of the stimuli was pseudorandomized and counterbalanced such that each of the conditions was equally often preceded by all stimulus conditions (Aguirre et al. 2011). Each run lasted approximately 12.17 min, resulting in a total of approximately 61 min of functional data. The intertrial interval (ITI) was jittered resulting in an ITI of either 6 or 6.5 s (50%) balanced over experimental conditions. The questions were presented jittered in steps of 200 ms between 1.395 and 2.395 s after the offset of the experimental stimulus and the jitter was balanced over experimental conditions.

### Procedure and Apparatus

The task applied in this paradigm had to ensure that listeners paid attention to the relative temporal pattern of the sound sequences, while the types of answers had to be orthogonal to the sound categories so the classifier does not encode the response type. After each stimulus participants had to answer

9 conditions		Tempo [BPS]		
		t1: 2.50	t2: 2.00	t3: 1.67
Relative temporal pattern	BA - Baseline 	pluck	pluck	drum
	R1 - Rhythm 1 	pluck	pluck	drum
	R2: -Rhythm 2 	pluck	pluck	drum

**Figure 1.** Left: Oscillograms visually representing the relative temporal patterns of the 2 temporally grouped experimental conditions (R1 and R2) and the isochronous baseline condition (BA). Right: 9 experimental conditions comprising 3 different tempi ( $t_1$ ,  $t_2$ ,  $t_3$ ). Per relative temporal pattern, each tempo results in a different absolute temporal pattern.

as quickly as possible a previously unknown question regarding the sound sequence. For example: “Were pauses 2 and 3 different?” or “Was it a guitar sound?” Importantly, participants had to keep a representation of the pattern in mind in order to be able to answer some of the questions, as categorizing them into R1 and R2 was not enough to answer the questions regarding the perceived temporal pattern. A total of 20 different questions was constructed and pseudorandomly distributed over experimental conditions in such a way that the number of answers per type (e.g., “yes”–“no”) was balanced over experimental conditions. Two potential answers were given side by side on the screen, and participants indicated their answer in the form of a two-alternative forced choice on each presented sequence by pressing 1 of 2 buttons. Participants were motivated to indicate their answer as quickly and as correctly as possible as there was a limited time window to answer (2.5 s) and as they received financial reimbursement for each correctly answered question in addition to their hourly payment. Participants performed a series of practice trials outside of the scanner until they were familiar with the task.

### fMRI Acquisition

Structural and functional data were collected on a 3 T Siemens Prisma scanner equipped with a 64-channel head coil at the Center for Biomedical Imaging (CIBM) at the University Hospital Lausanne (CHUV). A 3-dimensional high-resolution isotropic T1-weighted sequence (TR/TE/flip angle = 5.0 s/2.88 ms/0°) provided 176 contiguous slices with 240 × 256 mm in-plane resolution and a slice thickness of 1.2 mm (voxel size = 1 × 1 × 1.2 mm) (Marques et al. 2010). Functional MRI images were continuously acquired using a standard gradient echo sequence (TR/TE/flip angle = 2 s continuous/30 ms/90°) that acquired 33 axial functional images with 192 × 192 mm in-plane resolution, 2.5 mm slice thickness and a 0.25 mm interslice gap (voxel size = 1.96 × 1.96 × 2.75 mm) in ascending order covering the majority of voxels in the temporal, frontal and occipital lobes (see Supplementary Material S1).

### fMRI Analysis

Imaging data processing was carried out by Nipype v0.11.0 (Gorgolewski et al. 2011), integrating algorithms of the software packages FSL v5.0 (Smith et al. 2004), AFNI v7.18.1710 (Cox 1996), FreeSurfer v5.3.0 (Dale et al. 1999), ANTS v2.1.0 (Avants et al. 2011), ART (as implemented in Nipype) and SPM12. Structural data underwent cortical reconstruction and parcellation of the anatomical images using the default processing stream in FreeSurfer, the accuracy of which was verified manually via visual inspection.

#### Preprocessing

The anatomical image was used for spatial normalization to the MNI152-template. An anatomical mask was used to constrain all analyses. A mask including voxels that were covered in at least 5% of the volumes in each session was used for the first-level analysis. For the group analyses, a conjunction of all ANTS-normalized subject specific masks was further constrained by including only gray-matter voxels based on the Harvard-Oxford cortical and subcortical structural atlases (<http://fsl.fmrib.ox.ac.uk/fsl/fslwiki/Atlases>). Voxels were considered as being gray-matter if the probability was above 50%. EPI-data were despiked by means of AFNI on a voxel-by-voxel basis using interpolation to replace this voxel's value with a value based on neighboring

voxels. Volumes were corrected for slice-time acquisition relative to the 17th (i.e., middle) slice. Correction for motion was done by realigning every functional volume to the mean functional volume of the current run. Slice-time correction, motion correction and spatial smoothing (FWHM: 4 mm). To correct for session-related mean and scaling effects, we applied second-order detrending as implemented in the TSNR function of Nipype. The functional mean image resulting from this correction was then used to flag motion and intensity artifacts in the functional volumes. This was done with “artifact detection,” an algorithm implemented in Nipype that excluded volumes with a global intensity differing from the time series mean by more than 3 standard deviations or those in which a participant's composite head motion (the Euclidian combination of head translations and rotations) exceeded 1 mm between adjacent volumes. The mean functional image resulting from the TSNR function was used to calculate a coregistration matrix between the functional and the anatomical space using FreeSurfer's `bbregister`. Flagged outliers were excluded from the analysis by regressing them out in the individual statistics.

#### Univariate Analysis

Individual statistics were based on a least-square estimation using the GLM for serially autocorrelated observations and was performed separately on each voxel in the individual participant's space (Friston et al. 1995) with SPM12. Nine covariates of interest were calculated, representing the stimulus onset of the 9 experimental conditions. Covariates were convolved with the canonical hemodynamic response function. The serial autocorrelation of the BOLD time series was modeled using a first-order autoregressive mode. The 5 experimental runs were treated as separate sessions. No global intensity normalization was applied. Realignment-parameters were included as regressors of no interest in the statistical model. A temporal high-pass filter of 128 s was applied to remove low-frequency drifts over a timescale longer than this cutoff. The first-level model was masked with the mean functional image over all 5 sessions resulting from the TSNR function. On average, 3.8 out of 365 volumes per session and subject were identified as outliers and regressed out from the individual analysis. For the random-effects analysis contrasts of interest on the individual level were calculated and normalized to an MNI152-template using ANTS with a spatial resolution of 2 × 2 × 2 mm<sup>3</sup>. One-sample *t*-tests and *F*-tests were performed on the second level. *T*- and *F*-maps were thresholded on the voxel level at *P* < 0.001, and topological False Discovery Rate (FDR)-correction for multiple comparisons, as implemented in Nipype, was applied on the whole-brain activation cluster-extent level at *P* < 0.05 (Table 1).

#### Multivariate Pattern Analysis

In order to identify if activation patterns in the brain differ between rhythmic conditions, a classifier was iteratively trained on 80% of the data to discriminate between experimental conditions, and the resulting prediction accuracy on the remaining 20% of the data was averaged (Haxby et al. 2001; Haxby 2012). The classification of the data was done with PyMVPA v2.4.1 (Hanke et al. 2009). A three-way classification “R1 versus R2 versus BA” was performed. To prepare for the analysis, the  $\beta$ -maps from the GLM were normalized by voxel-wise *z*-scoring per run across conditions and conditions necessary for each classification in subject's native space. Thus, each cross-validation was based on 30 (6 × 5)  $\beta$ -maps. Whole head *z*-statistic maps were then used for the classification analysis.

**Table 1** Brain areas resulting from the univariate analyses

MNI coordinates (mm)			Mean t-values	Extent (voxels of $2 \times 2 \times 2 \text{ mm}^3$ )	Peak location according to Harvard-Oxford probabilistic atlas
x	y	z			
<b>Temporally grouped rhythms versus isochronous baseline rhythms (R &gt; BA)</b>					
38	-48	52	3.99	1360	47% Superior_parietal_lobule, 22% Angular_gyrus, 6% Supramarginal_gyrus_posterior_division
42	-36	40	3.87	208	25% Supramarginal_gyrus_posterior_division, 14% supramarginal_gyrus_anterior_division, 8% postcentral_gyrus, 5% superior_parietal_lobule
50	-34	50	3.85	96	29% Supramarginal_gyrus_posterior_division, 29% supramarginal_gyrus_anterior_division, 13% postcentral_gyrus, 1% angular_gyrus
<b>Isochronous baseline rhythms versus temporally grouped rhythms (BA &gt; R)</b>					
18	68	4	4.06	2624	73% Frontal_pole
54	-24	20	4.18	2392	58% Parietal_operculum_cortex, 10% planum_temporale, 4% central_opercular_cortex, 4% supramarginal_gyrus_anterior_division
-6	58	6	4.11	1232	65% Frontal_pole, 15% paracingulate_gyrus, 4% frontal_medial_cortex
-40	-4	-6	4.10	1192	98% Insular_cortex
-38	6	-10	4.23	120	95% Insular_cortex
-44	0	6	3.99	72	64% Central_opercular_cortex, 20% insular_cortex
-36	-22	14	3.91	56	49% Insular_cortex, 14% Heschl's_gyrus_(includes_H1_and_H2), 9% central_opercular_cortex, 3% parietal_operculum_cortex
-36	-26	10	3.61	8	60% Heschl's_gyrus_(includes_H1_and_H2), 3% planum_polare, 3% insular_cortex, 2% planum_temporale
-34	-22	18	3.75	8	53% Insular_cortex, 11% central_opercular_cortex, 8% parietal_operculum_cortex

For the classification analysis a linear support vector machine (SVM) from the LIBSVM package (<https://www.csie.ntu.edu.tw/~cjlin/libsvm>) as implemented in PyMVPA was used. As the dataset was presented in 5 sessions, a 5-fold leave-one-run-out cross-validation (i.e., training on 80% and testing on 20% of the data) was performed on the whole-brain (Pereira et al. 2009). The SVM C hyperparameter was automatically scaled according to the data norm for each training data fold. Voxels representing the top 5% of all F-values from a one-way ANOVA on all experimental contrasts in the training set entered the whole-brain cross-validation. The group-level accuracy and confusion matrix was obtained by averaging the confusion matrix or accuracy of all participants. Significance testing in the cross-validation was performed on accuracy values by a one-sampled t-test with the test-value 33.33% for the 3-fold discrimination. Bonferroni-Holmes correction was applied for post hoc testing.

### Searchlight Analysis

As an approach to localize brain regions with high discriminative information, we performed a searchlight analysis that is more sensitive to functional brain organization (Kriegeskorte et al., 2006). The searchlight analysis was performed across the brain in subject space and for each subject and pair of conditions, separately. Every second voxel was taken as a center point for a sphere with a radius of 3 voxels (i.e., 7 mm radius, up to 123 voxels). For each of those spheres, a 5-fold leave-one-out cross-validation was performed, resulting in 1 accuracy value per sphere. The classification accuracy of a given voxel was the average accuracy of all spheres that included this voxel. The resulting subject specific accuracy maps per classification pair were normalized with ANTS into the MNI152-template space and averaged over all participants to obtain the group classification accuracy map. As a three-way classification in the searchlight approach impairs interpretability by hiding which particular category discriminations are driving the overall accuracy, 2 two-way classifications were done including "R1 versus R2" and "R (R1 and R2) versus BA." For the second classification, an

unbalanced number of elements per condition was available ( $R = 30$ ,  $BA = 15$ ). Therefore, we report an average of 30 classifications, each of which was based on a random set of 15 elements from R.

Significance testing in the searchlight analysis was performed by using permutation and bootstrap sampling methods, followed by cluster thresholding with correction for multiple comparisons as suggested by Stelzer et al. (2013) and implemented by PyMVPA on images spatially normalized to the MNI space with a voxel resolution of  $2 \times 2 \times 2 \text{ mm}^3$ . The permutation was achieved by conducting the previously mentioned searchlight for each subject an additional 99 times, each time with randomly permuted data labels resulting in one condition relevant accuracy map and 99 "noise" accuracy maps per subject. For the bootstrapping process, new group-level "noise" accuracies maps were generated by selecting randomly 1 of the 100 accuracy maps per subject and averaging them into a new group-level chance accuracy map. This process was repeated 100 000 times to create a voxelwise null distribution of the classification accuracy map on the group level. All volumes were thresholded with a voxelwise threshold of  $P < 0.001$ . Afterwards, the distribution of maximal cluster size was determined from all thresholded chance accuracy maps as a family-wise error (FWE) measure. Probability of clusters in the thresholded empirical accuracy map was evaluated using this distribution (threshold  $P < 0.05$ ).

## Results

### Behavioral Data

Participants were proficient in performing the task indicating that the Gestalt percepts were, in general, successfully induced. Mean accuracy of performance was on average  $82.3 \pm 1.1\%$  (standard error), ranging from 70.6% to 90.4%. Correct responses were significantly more frequent in the isochronous baseline ( $95.8 \pm 1.1\%$ ) compared with the grouped condition, R2 ( $78.5 \pm 1.5\%$ ;  $t_{(18)} = 247.11$ ,  $P < 0.001$ ) and significantly more frequent in the grouped condition, R2, compared with the grouped condition, R1 ( $72.7 \pm 1.1\%$ ,  $t_{(18)} = 40.2$ ,  $P < 0.001$ ). As not all questions

asked in the task related to the percepts of Gestalt, the few error trials cannot be directly attributed to a failure in representing the Gestalt percept. All trials were subsequently used for fMRI analysis.

### fMRI Data

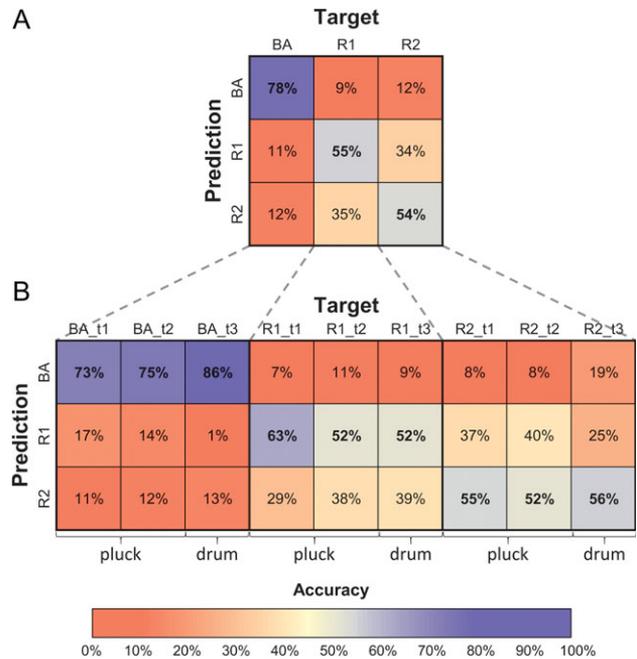
Based on a univariate analysis approach, no statistically significant group activation difference was observed when comparing the two temporally grouped experimental conditions (R1 </> R2) at a voxelwise threshold of  $P < 0.05$  and a cluster-wise threshold of  $P < 0.05$  corrected. All temporally grouped experimental conditions (R1 and R2) compared with the isochronous baseline condition resulted in activity in the right superior parietal lobe including a cluster in the angular gyrus (AG) and two clusters in the supramarginal gyrus (SMG). The reverse contrast revealed activity in the frontal pole (FP), the left and right insula, the left Heschl's gyrus including the planum temporale (PT), and the left opercular cortex (Table 1, Supplementary Material S2).

An multivariate pattern analysis (MVPA) analysis revealed with what precision brain activation patterns can predict which percept was induced in the brain and identified the loci of brain activation patterns between percepts. In a multivariate cross-validation approach a classifier on voxel-based brain activation patterns trained across the tempi of the sequences (t1, t2, t3) was able to predict not only whether a temporal grouped experimental stimulus (R) or an isochronous baseline stimulus (BA) had been presented but, more importantly, which of the two temporally grouped experimental conditions (R1 or R2) was perceived. The discrimination accuracy for the three-way classification of BA, R1, and R2 was 62.46% ( $t_{(18)} = 11.11$ ,  $P < 0.001$ ). Each of the three experimental conditions was successfully classified (BA = 78%,  $t_{(18)} = 16.81$ ,  $P < 0.001$  corr.; R1 = 55%,  $t_{(18)} = 6.39$ ,  $P < 0.001$  corr.; and R2 = 54%,  $t_{(18)} = 5.75$ ,  $P < 0.001$  corr.). Figure 2 illustrates the performance of the three-way classifier by separately indicating hit-rates and misidentifications for each experimental condition thus illustrating the contribution of each experimental condition to the average discrimination accuracy.

To identify the brain regions decoding the different rhythms we performed a searchlight analysis on the two critical comparisons; the comparison between two temporally grouped experimental conditions, R1 and R2, and the comparison between both temporally grouped experimental conditions, R, and the isochronous baseline condition, BA. We identified brain locations in which patterns discriminated between the two temporally grouped experimental conditions in the right inferior frontal gyrus (IFG) and bilaterally in the temporo-parietal junction (TPJ) including the PT and SMG (Fig. 3, top, Table 2). Classification of the difference between temporally grouped experimental conditions (R) and the isochronous baseline condition (BA) yielded large parts of the brain, including the regions that discriminated between the 2 temporally grouped experimental conditions. Differentiation at a classification accuracy higher than 68.47%, corresponding to values above the 95th percentile of all voxels, yielded 8 clusters (Fig. 3, bottom, Table 2) including the left and right lateral occipital lobe (LOL), the right middle frontal gyrus (MFG), the left precuneus (PCUN), the left supplementary motor area (SMA), the right premotor cortex, the right insula (IN), the left and right supramarginal gyrus (SMG), the superior parietal lobe and angular gyrus, and right the frontal pole (FP).

### Discussion

The perceptual binding of acoustic events across time is inherent to hearing and allows perceiving acoustic scenes rather than

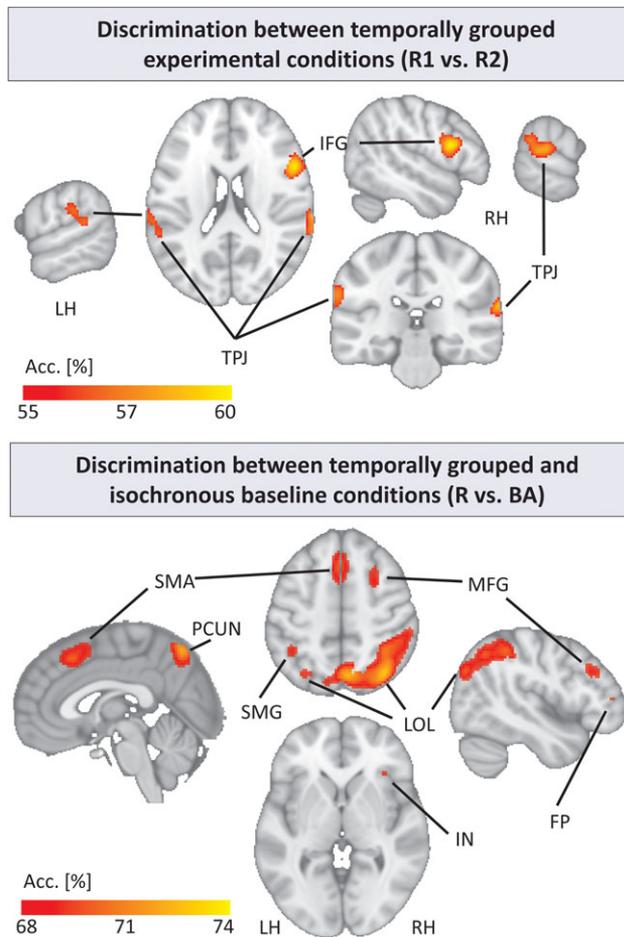


**Figure 2.** Results of the three-way multivariate cross-validation in which the classifier was trained on a sample set of stimuli from the category BA, R1, and R2 across all 3 tempi. (A) depicts predictions in % per condition (chance level identification = 33%). (B) visualizes the same classification indicating predictions separately for each subcondition. Identifications above chance level are highlighted in bold.

sequences of independent sounds. Based on brain activation patterns, a linear SVM classifier identified which relative temporal pattern had been presented. Importantly, the classifier was trained across absolute temporal patterns (tempi) of sequences providing a strong indication that the classifier had identified the neural activation pattern associated with the Gestalt percept, thus the relative temporal pattern. A searchlight analysis identified the regions involved in the representation of a specific temporal Gestalt. Activation patterns in the TPJ of both hemispheres (including the PT and parts of the SMG) and the IFG in the right hemisphere discriminated between the two different temporally grouped percepts of temporal Gestalt. Our findings reveal for the first time secondary and higher cortical areas which contribute to the representation of temporal Gestalt.

Temporally grouped (Gestalt) and isochronous baseline rhythms and, moreover, two different forms of temporally grouped experimental rhythms were successfully classified based on cerebral activation patterns across the whole brain. This classification was achieved across tempi and timbres. While the isochronous baseline was identified at the highest extent, importantly, the two temporally grouped conditions were identified to a similar extent by the classifier (Fig. 2B). The classification is therefore based on brain activity that is specific to the temporal Gestalt percept induced by the stimulus and independent of the absolute temporal relationship between the constituent sounds. These results indicate that a specific brain activation pattern encodes the temporal Gestalt of an acoustic sequence, mirroring the perceptual characteristic of temporal Gestalt, that is, perceptual invariance across tempi (Hulst et al. 1992).

The representation of the specific temporally grouped Gestalt was localized bilaterally in the TPJ and in the right IFG. To our knowledge, this is the first study that successfully isolates brain activity specific to temporal Gestalt. Previous studies had focused on the processing of rhythmic pattern differences



**Figure 3.** Searchlight result: Regions that show a significant classification above chance level ( $P < 0.05$ , cluster-level corrected) on a standard MNI152-brain for (top) 2 temporally grouped experimental conditions (R1 vs. R2) and (bottom) temporally grouped experimental and isochronous baseline conditions (R vs. BA, >95th percentile of classification accuracy). The group maps were created by averaging the individual discriminative voxel maps of all participants. Color coding indicates discrimination accuracy. RH = right hemisphere; LH = left hemisphere.

between beat and non-beat based rhythms and between isochronous and non-isochronous rhythms (Grahn and Brett 2007; Geiser et al. 2008, 2012; Bengtsson et al. 2009) or task related activity in listening, discrimination, and reproduction tasks (Chen et al. 2008; Konoike et al. 2012; Thaut et al. 2014) and were not designed to identify the encoding of Gestalt. Here, we show for the first time that the relative temporal distance between acoustic events in a sequence of sounds and their resulting Gestalt percept is encoded in two specific cortical regions, namely the TPJ and the IFG.

Activation patterns in the TPJ that decoded temporal Gestalt extended bilaterally from the PT to lateral-parietal regions. The PT is known as a computational hub in auditory processing. It is suggested to underlie the functional specialization of hemispheres due to its sensitivity to specific temporal intervals (Griffiths and Warren 2002; Josse et al. 2003; Poeppel 2003). Our results point to an additional function of the PT in the integration of temporal Gestalt. We did not simply test the distinction between two temporal patterns, but between two relative temporal patterns. Thus, in this brain area a specific voxel activation pattern contributes to the representation of a temporal Gestalt

that can emerge from a slower or a faster sequence of tones, thus from different absolute temporal intervals. Our findings indicate that the PT encodes the integration of temporal intervals into a temporal Gestalt and could, thus, underlie perceptual invariance of temporal Gestalt across temporal inputs. Discriminating voxel activation patterns in TPJ also included multisensory areas in the lateral parietal cortex that have earlier been associated with Gestalt representation in the visual domain based on both simultanagnosic behavior after lesions in TPJ (Valenza et al. 2004; Huberle and Karnath 2006, 2010; Ritzinger et al. 2012) and imaging studies in healthy participants (Huberle and Karnath 2012; Rennig et al. 2015). In the auditory domain, parietal aspects of TPJ are known to process sounds in space (Romanski et al. 1999; Maeder et al. 2001; Tian et al. 2001; Altmann et al. 2008; Ducommun et al. 2002; Tata and Ward 2005; De Santis et al. 2007). Our results not only confirm the role of the TPJ in Gestalt processing by expanding it to auditory temporal Gestalt, but also suggest that already secondary auditory cortices, namely the posterior superior temporal gyrus (pSTG) are sensitive to temporal Gestalt and that this process relies on both the left and the right hemispheres. This is an additional function of the pSTG that complements its known role in temporal processing.

Activation patterns in the frontal cortex that decoded temporal Gestalt were localized to the right IFG, namely, the pars opercularis and extended into the precentral gyrus. As part of the ventral pathway in human auditory perception (Clarke et al. 2000; Thiran and Clarke 2003; Arnott et al. 2004; Zundorf et al. 2016; Alain et al. 2001), this region likely plays a role in auditory object recognition (De Lucia et al. 2009). This is the first account of an involvement of these brain regions in temporal Gestalt encoding. One limitation of the decoding approach as used in this study is, that it does not provide information about why specific brain areas display increased classification accuracy (Serences and Saproo 2012). It is possible that two distinct, spatially separated populations of neurons encode different experimental conditions or, alternatively, that neurons encoding two conditions are spatially intertwined in both populations of neurons. In the case of temporal Gestalt it is likely that the TPJ encodes a representation that is closer to the timing characteristic inducing the percept, while the IFG represents the higher or more abstract contextual binding of a temporal Gestalt.

The classification between temporally grouped experimental conditions (R1 and R2) and the isochronous baseline condition (BA) in the searchlight analysis localized discriminating patterns in a wide network of medial and lateral cortical and subcortical areas, including the PCUN, SMA, MFG, right premotor cortex, and superior parietal lobe. These results conform to a large literature associating these areas with temporal tasks. Areas that are frequently reported in temporal discrimination tasks or rhythm reproduction tasks include SMA (Schubotz et al. 2000; Rao et al. 2001; Belin et al. 2002; Lewis et al. 2004; Grahn and Brett 2007; Geiser et al. 2008, 2012; Bengtsson et al. 2009; Teki et al. 2011; Konoike et al. 2012; Thaut et al. 2014; Riecker et al. 2002; Chen et al. 2008; Grahn and McAuley 2009; Schwartz et al. 2011) and dorsolateral prefrontal cortex (Chen et al. 2008; Grahn and Rowe 2009). We restricted the results of this searchlight analysis to the most significant voxels, as the standard significance threshold had revealed discriminating patterns in even larger parts of the cortex. Explanation for this large network could be two-fold. First, in the baseline condition the identification of the category was often sufficient to perform the task while the two experimental conditions required the representation of the temporal Gestalt. Thus, as suggested by the behavioral results, performing

**Table 2** Brain areas discriminating between experimental conditions according to a searchlight analysis

MNI coordinates (mm)			Mean prediction/ classification accuracy (%)	Extent (voxels of $2 \times 2 \times 2 \text{ mm}^3$ )	Cluster location according to Harvard-Oxford probabilistic atlas
x	y	z			
<b>Discrimination between temporally grouped rhythms and isochronous baseline rhythms (R and BA) (clusters = voxels &gt; 95% discrimination accuracy)</b>					
32	-70	46	70.52	25 728	Lateral occipital cortex, superior division: 32.79%, precuneus cortex: 11.81%, angular gyrus: 7.91%, superior parietal lobule: 5.68%, supramarginal gyrus, posterior division: 5.36%
-2	20	48	69.58	3680	Paracingulate gyrus: 38.97%, superior frontal gyrus: 23.98% (supplementary motor area (SMA) according to AAL)
28	4	58	69.83	3072	Middle frontal gyrus: 25.93%, superior frontal gyrus: 22.52%
-32	-72	44	68.96	1296	Lateral occipital cortex, superior division: 61.42%
-38	-46	44	69.22	1288	Superior parietal lobule: 26.19%, supramarginal gyrus, posterior division: 24.43%, angular gyrus: 9.84%
46	28	30	69.02	824	Middle frontal gyrus: 47.74%, frontal pole: 8.12%
34	26	0	68.67	56	Frontal orbital cortex: 37.57%, insular cortex: 15.86%, frontal operculum cortex: 14.71%
46	44	8	68.56	16	Frontal pole: 89.5%
<b>Discrimination between temporally grouped rhythms (R1 and R2)</b>					
54	14	18	57.24	4216	Inferior frontal gyrus, pars opercularis: 31.24%, precentral gyrus: 26.48%
68	-28	18	56.93	2016	Supramarginal gyrus, posterior division: 24.50%, Superior temporal gyrus, posterior division: 23.02%, planum temporale: 8.75%
-64	-30	26	56.75	1832	Supramarginal gyrus, anterior division: 32.75%, parietal operculum cortex: 12.25%, planum temporale: 12.15%, postcentral gyrus: 8.68%, superior temporal gyrus, posterior division: 5.57%, supramarginal gyrus, posterior division: 5.42%

the task on the two experimental conditions was more difficult than performing the task on the baseline condition. Task difficulty is often associated with brain activity that is spread across several cortical areas. Second, a searchlight analysis can potentially display over-sensitivity to activation pattern differences when discriminating voxels that are highly dispersed in the brain, because a minority of voxels suffices to identify a searchlight as informative (Etzel et al. 2013). Thus, it is likely that a few highly informative voxels associated with task difficulty account for the large network involved in discriminating between experimental and baseline stimuli.

The here described machine-learning experiment showed how neural mechanisms underlying invariance in perceptual categories can be investigated. Our results extend previous research on perceptual categories that are nonlinearly related to their sensory basis such as machine-learning research on speech perception (Lee et al. 2012; Borghesani et al. 2016; Markiewicz and Bohland 2016) and general perceptual object and concept processing (Linden et al. 2012; Ley et al. 2014; Klein and Zatorre 2015; Schindler et al. 2013; Zhang et al. 2015; Wurm and Lingnau 2015), but also research using alternative experimental paradigms such as bistable stimuli or illusory percepts (de Wit et al. 2012; Ishizu and Zeki 2014; Kok and de Lange 2014; Grassi et al. 2017; Murray and Herrmann 2013; Kubilius et al. 2015). Future experiments should additionally test whether classifiers resulting from MVPA can in fact generalize across physically different stimuli (Correia et al. 2015; Arsenault and Buchsbaum 2016; Derrfuss et al. 2017; de Borst et al. 2016) by testing the classifier on items of a perceptual category with a sensory basis unrelated to the one used for classifier training.

In conclusion, we investigated for the first time where in the brain temporal Gestalt percepts emerge and localized this process to the TPJ (including the PT and the SMG) and the right inferior frontal cortex. The results of this study are of relevance as Gestalt processing is a core capacity of perception. Rhythmic

aspects of auditory perception are considered a fundamental mechanism for correct language perception and suggested to be impaired in children with speech and language deficits (Cumming et al. 2015). Our results provide a keystone towards a model of rhythm perception by indicating that both sensory and higher cortical areas encode temporal Gestalt. Finally, the study provides an example of how we can tackle the neural basis of perceptual invariance using neuroimaging methods.

## Supplementary Material

Supplementary data is available at *Cerebral Cortex* online.

## Funding

Swiss National Science Foundation (PZOOPI\_148184 to E.G., 320030-149982, 320030-169206 to M.M.M., as well as project 51AU40-125759, "SYNAPSY, The Synaptic Bases of Mental Disease" to M.M.M.), the German federal state of Saxony-Anhalt and the European Regional Development Fund (ERDF) to M.H. (Project: Center for Behavioral Brain Sciences), a grantor advised by Carigest S.A. awarded to M.M.M., the Center d'Imagerie BioMédicale (CIBM) of the University of Lausanne (UNIL) and the Swiss Federal Institute of Technology Lausanne (EPFL). The neuropsychology and the radiology services at the University Hospital in Lausanne (CHUV) as well as the Leenaards and Jeantet Foundations.

## Authors' Contributions (CRediT)

Conceptualization, E.G. and MP.N.; Methodology, E.G. and MP.N.; Formal analysis, MP.N. and M.H.; Software, M.H.; Writing—original draft, E.G. and MP.N.; Writing—review and editing, E.G., MP.N., M.M.M., and M.H.; Visualization, MP.N.; Funding acquisition, E.G. and M.M.M.; Supervision, E.G.

## Notes

We thank Dr Eleonora Fornari from the Center d'Imagerie BioMédicale (CIBM) in Lausanne for her helpful assistance in data acquisition and 3 anonymous reviewers for their helpful comments. *Conflict of Interest*: None declared.

## References

- Aguirre GK, Mattar MG, Magis-Weinberg L. 2011. de Bruijn cycles for neural decoding. *Neuroimage*. 56:1293–1300.
- Alain C, Arnott SR, Hevenor S, Graham S, Grady CL. 2001. “What” and “where” in the human auditory system. *Proc Natl Acad Sci USA*. 98:12301–12306.
- Altmann CF, Bulthoff HH, Kourtzi Z. 2003. Perceptual organization of local elements into global shapes in the human visual cortex. *Curr Biol*. 13:342–349.
- Altmann CF, Henning M, Doering MK, Kaiser J. 2008. Effects of feature-selective attention on auditory pattern and location processing. *Neuroimage*. 41:69–79.
- Annett M. 1992. 5 Tests of hand skill. *Cortex*. 28:583–600.
- Arnal LH, Doelling KB, Poeppel D. 2015. Delta-beta coupled oscillations underlie temporal prediction accuracy. *Cereb Cortex*. 25:3077–3085.
- Arnal LH, Giraud A-L. 2012. Cortical oscillations and sensory predictions. *Trends Cogn Sci*. 16:390–398.
- Arnott SR, Binns MA, Grady CL, Alain C. 2004. Assessing the auditory dual-pathway model in humans. *Neuroimage*. 22:401–408.
- Arsenault JS, Buchsbaum BR. 2016. No evidence of somatotopic place of articulation feature mapping in motor cortex during passive speech perception. *Psychon Bull Rev*. 23:1231–1240.
- Avants BB, Tustison NJ, Song G, Cook PA, Klein A, Gee JC. 2011. A reproducible evaluation of ANTs similarity metric performance in brain image registration. *Neuroimage*. 54:2033–2044.
- Belin P, McAdams S, Thivard L, Smith B, Savel S, Zilbovicius M, Samson S, Samson Y. 2002. The neuroanatomical substrate of sound duration discrimination. *Neuropsychologia*. 40:1956–1964.
- Bendixen A. 2014. Predictability effects in auditory scene analysis: a review. *Front Neurosci*. 8:60.
- Bendixen A, Schwartze M, Kotz SA. 2015. Temporal dynamics of contingency extraction from tonal and verbal auditory sequences. *Brain Lang*. 148:64–73.
- Bengtsson SL, Ullen F, Ehrsson HH, Hashimoto T, Kito T, Naito E, Forssberg H, Sadato N. 2009. Listening to rhythms activates motor and premotor cortices. *Cortex*. 45:62–71.
- Bonte M, Valente G, Formisano E. 2009. Dynamic and task-dependent encoding of speech and voice by phase reorganization of cortical oscillations. *J Neurosci*. 29:1699–1706.
- Borghesani V, Pedregosa F, Buiatti M, Amadon A, Eger E, Piazza M. 2016. Word meaning in the ventral visual path: a perceptual to conceptual gradient of semantic coding. *Neuroimage*. 143:128–140.
- Carlile S, Leung J. 2016. The perception of auditory motion. *Trends Hear*. 20:1–19.
- Chen JL, Penhune VB, Zatorre RJ. 2008. Listening to musical rhythms recruits motor regions of the brain. *Cereb Cortex*. 18:2844–2854.
- Clarke S, Bellmann A, Meuli RA, Assal G, Steck AJ. 2000. Auditory agnosia and auditory spatial deficits following left hemispheric lesions: evidence for distinct processing pathways. *Neuropsychologia*. 38:797–807.
- Correia JM, Jansma BMB, Bonte M. 2015. Decoding articulatory features from fmri responses in dorsal speech regions. *J Neurosci*. 35:15015–15025.
- Cox RW. 1996. AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. *Comput Biomed Res*. 29:162–173.
- Cumming R, Wilson A, Leong V, Colling LJ, Goswami U. 2015. Awareness of rhythm patterns in speech and music in children with specific language impairments. *Front Hum Neurosci*. 9:672.
- Dale AM, Fischl B, Sereno MI. 1999. Cortical surface-based analysis—I. Segmentation and surface reconstruction. *Neuroimage*. 9:179–194.
- de Borst AW, Valente G, Jaaskelainen IP, Tikka P. 2016. Brain-based decoding of mentally imagined film clips and sounds reveals experience-based information patterns in film professionals. *Neuroimage*. 129:428–438.
- De Lucia M, Camen C, Clarke S, Murray MM. 2009. The role of actions in auditory object discrimination. *Neuroimage*. 48:475–485.
- De Santis L, Clarke S, Murray MM. 2007. Automatic and intrinsic auditory “what” and “where” processing in humans revealed by electrical neuroimaging. *Cereb Cortex*. 17:9–17.
- de-Wit LH, Kubilius J, Wagemans J, Op de Beeck HP. 2012. Bistable Gestalts reduce activity in the whole of V1, not just the retinotopically predicted parts. *J Vis*. 12.
- Derrfuss J, Ekman M, Hanke M, Tittgemeyer M, Fiebach CJ. 2017. Distractor-resistant STM is supported by transient changes in neural stimulus representations. *J Cogn Neurosci*. 29:1547–1565.
- Ducommun CY, Murray MM, Thut G, Bellmann A, Viaud-Delmon I, Clarke S, Michel CM. 2002. Segregated processing of auditory motion and auditory location: an ERP mapping study. *Neuroimage*. 16:76–88.
- Etzel JA, Zacks JM, Braver TS. 2013. Searchlight analysis: promise, pitfalls, and potential. *Neuroimage*. 78:261–269.
- Friston K, Holmes A, Poline J, Grasby P, Williams S, Frackowiak R, Turner R. 1995. Analysis of fMRI time-series revisited. *Neuroimage*. 2:45–53.
- Geiser E, Gabrieli JDE. 2013. Influence of rhythmic grouping on duration perception: a novel auditory illusion. *PLoS One*. 8:e54273.
- Geiser E, Notter M, Gabrieli JDE. 2012. A corticostriatal neural system enhances auditory perception through temporal context processing. *J Neurosci*. 32:6177–6182.
- Geiser E, Walker KMM, Bendor D. 2014. Global timing: a conceptual framework to investigate the neural basis of rhythm perception in humans and non-human species. *Front Psychol*. 5:159.
- Geiser E, Zaehle T, Jancke L, Meyer M. 2008. The neural correlate of speech rhythm as evidenced by metrical speech processing. *J Cogn Neurosci*. 20:541–552.
- Ghitza O, Giraud A-L, Poeppel D. 2012. Neuronal oscillations and speech perception: critical-band temporal envelopes are the essence. *Front Hum Neurosci*. 6:340.
- Gorgolewski K, Burns CD, Madison C, Clark D, Halchenko YO, Waskom ML, Ghosh SS. 2011. Nipype: a flexible, lightweight and extensible neuroimaging data processing framework in python. *Front Neuroinform*. 5:13–13.
- Grahn JA. 2012. Neural mechanisms of rhythm perception: current findings and future perspectives. *Top Cogn Sci*. 4:585–606.
- Grahn JA, Brett M. 2007. Rhythm and beat perception in motor areas of the brain. *J Cogn Neurosci*. 19:893–906.

- Grahn JA, McAuley JD. 2009. Neural bases of individual differences in beat perception. *Neuroimage*. 47:1894–1903.
- Grahn JA, Rowe JB. 2009. Feeling the beat: premotor and striatal interactions in musicians and nonmusicians during beat perception. *J Neurosci*. 29:7540–7548.
- Grassi PR, Zaretskaya N, Bartels A. 2016. Parietal cortex mediates perceptual Gestalt grouping independent of stimulus size. *Neuroimage*. 133:367–377.
- Grassi PR, Zaretskaya N, Bartels A. 2017. Scene segmentation in early visual cortex during suppression of ventral stream regions. *Neuroimage*. 146:71–80.
- Griffiths TD, Warren JD. 2002. The planum temporale as a computational hub. *Trends Neurosci*. 25:348–353.
- Griffiths TD, Warren JD. 2004. What is an auditory object? *Nat Rev Neurosci*. 5:887–892.
- Hanke M, Halchenko YO, Sederberg PB, Hanson SJ, Haxby JV, Pollmann S. 2009. PyMVPA: a Python toolbox for multivariate pattern analysis of fMRI data. *Neuroinformatics*. 7:37–53.
- Haxby JV. 2012. Multivariate pattern analysis of fMRI: the early beginnings. *Neuroimage*. 62:852–855.
- Haxby JV, Gobbini MI, Furey ML, Ishai A, Schouten JL, Pietrini P. 2001. Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science*. 293:2425–2430.
- Huberle E, Karnath HO. 2006. Global shape recognition is modulated by the spatial distance of local elements—evidence from simultanagnosia. *Neuropsychologia*. 44:905–911.
- Huberle E, Karnath H-O. 2010. Saliency modulates global perception in simultanagnosia. *Exp Brain Res*. 204:595–603.
- Huberle E, Karnath H-O. 2012. The role of temporo-parietal junction (TPJ) in global Gestalt perception. *Brain Struct Funct*. 217:735–746.
- Hulse S, Takeuchi A, Braaten R. 1992. Perceptual invariances in the comparative psychology of music. *Music Percept*. 10:151–184.
- Ishizu T, Zeki S. 2014. Varieties of perceptual instability and their neural correlates. *Neuroimage*. 91:203–209.
- JTrost W, Labbe C, Grandjean D. 2017. Rhythmic entrainment as a musical affect induction mechanism. *Neuropsychologia*. 96:96–110.
- Josse G, Mazoyer B, Crivello F, Tzourio-Mazoyer N. 2003. Left planum temporale: an anatomical marker of left hemispheric specialization for language comprehension. *Brain Res Cogn Brain Res*. 18:1–14.
- Keil A, Muller MM, Ray WJ, Gruber T, Elbert T. 1999. Human gamma band activity and perception of a gestalt. *J Neurosci*. 19:7152–7161.
- Keller PE, Novembre G, Hove MJ. 2014. Rhythm in joint action: psychological and neurophysiological mechanisms for real-time interpersonal coordination. *Philos Trans R Soc Lond B Biol Sci*. 369:20130394.
- Klein ME, Zatorre RJ. 2015. Representations of invariant musical categories are decodable by pattern analysis of locally distributed BOLD responses in superior temporal and intraparietal sulci. *Cereb Cortex*. 25:1947–1957.
- Kok P, de Lange FP. 2014. Shape perception simultaneously up- and downregulates neural activity in the primary visual cortex. *Curr Biol*. 24:1531–1535.
- Konoike N, Kotozaki Y, Miyachi S, Miyauchi CM, Yomogida Y, Akimoto Y, Kuraoka K, Sugiura M, Kawashima R, Nakamura K. 2012. Rhythm information represented in the fronto-parieto-cerebellar motor system. *Neuroimage*. 63:328–338.
- Kotz SA, Schmidt-Kassow M. 2015. Basal ganglia contribution to rule expectancy and temporal predictability in speech. *Cortex*. 68:48–60.
- Kriegeskorte N, Goebel R, Bandettini P. 2006. Information-based functional brain mapping. *Proc Natl Acad Sci USA*. 103:3863–3868.
- Kubilius J, Baeck A, Wagemans J, Op de Beeck HP. 2015. Brain-decoding fMRI reveals how wholes relate to the sum of parts. *Cortex*. 72:5–14.
- Lee Y-S, Turkeltaub P, Granger R, Raizada RDS. 2012. Categorical speech processing in Broca's area: an fMRI study using multivariate pattern-based analysis. *J Neurosci*. 32:3942–3948.
- Lewis PA, Wing AM, Pope PA, Praamstra P, Miall RC. 2004. Brain activity correlates differentially with increasing temporal complexity of rhythms during initialisation, synchronisation, and continuation phases of paced finger tapping. *Neuropsychologia*. 42:1301–1312.
- Ley A, Vroomen J, Formisano E. 2014. How learning to abstract shapes neural sound representations. *Front Neurosci*. 8:132.
- Linden DEJ, Oosterhof NN, Klein C, Downing PE. 2012. Mapping brain activation and information during category-specific visual working memory. *J Neurophysiol*. 107:628–639.
- Maeder PP, Meuli RA, Adriani M, Bellmann A, Fornari E, Thiran JP, Pittet A, Clarke S. 2001. Distinct pathways involved in sound recognition and localization: a human fMRI study. *Neuroimage*. 14:802–816.
- Markiewicz CJ, Bohland JW. 2016. Mapping the cortical representation of speech sounds in a syllable repetition task. *Neuroimage*. 141:174–190.
- Marques JP, Kober T, Krueger G, van der Zwaag W, Van de Moortele P-F, Gruetter R. 2010. MP2RAGE, a self bias-field corrected sequence for improved segmentation and. *Neuroimage*. 49:1271–1281.
- Murray MM, Herrmann CS. 2013. Illusory contours: a window onto the neurophysiology of constructing perception. *Trends Cogn Sci*. 17:471–481.
- Pereira F, Mitchell T, Botvinick M. 2009. Machine learning classifiers and fMRI: a tutorial overview. *Neuroimage*. 45:S199–S209.
- Poeppel D. 2003. The analysis of speech in different temporal integration windows: cerebral lateralization as “asymmetric sampling in time”. *Speech Commun*. 41:245–255.
- Poeppel D. 2014. The neuroanatomic and neurophysiological infrastructure for speech and language. *Curr Opin Neurobiol*. 28:142–149.
- Rao SM, Mayer AR, Harrington DL. 2001. The evolution of brain activation during temporal processing. *Nature Neurosci*. 4:317–323.
- Rennig J, Himmelbach M, Huberle E, Karnath H-O. 2015. Involvement of the TPJ area in processing of novel global forms. *J Cogn Neurosci*. 27:1587–1600.
- Riecker A, Wildgruber D, Dogil G, Grodd W, Ackermann H. 2002. Hemispheric lateralization effects of rhythm implementation during syllable repetitions: an fMRI study. *Neuroimage*. 16:169–176.
- Rimmele J, Jolsvai H, Sussman E. 2011. Auditory target detection is affected by implicit temporal and spatial expectations. *J Cogn Neurosci*. 23:1136–1147.
- Ritzinger B, Huberle E, Karnath H-O. 2012. Bilateral theta-burst TMS to influence global gestalt perception. *PLoS One*. 7:e47820.
- Romanski LM, Tian B, Fritz J, Mishkin M, Goldman-Rakic PS, Rauschecker JP. 1999. Dual streams of auditory afferents target multiple domains in the primate prefrontal cortex. *Nature Neurosci*. 2:1131–1136.
- Schindler A, Herdener M, Bartels A. 2013. Coding of melodic gestalt in human auditory cortex. *Cereb Cortex*. 23:2987–2993.

- Schubotz RI, Friederici AD, von Cramon DY. 2000. Time perception and motor timing: a common cortical and subcortical basis revealed by fMRI. *Neuroimage*. 11:1–12.
- Schwartz M, Keller PE, Patel AD, Kotz SA. 2011. The impact of basal ganglia lesions on sensorimotor synchronization, spontaneous motor tempo, and the detection of tempo changes. *Behav Brain Res*. 216:685–691.
- Schwartz M, Kotz SA. 2016. Contributions of cerebellar event-based temporal processing and preparatory function to speech perception. *Brain Lang*. 161:28–32.
- Serences JT, Saproo S. 2012. Computational advances towards linking BOLD and behavior. *Neuropsychologia*. 50:435–446.
- Smith SM, Jenkinson M, Woolrich MW, Beckmann CF, Behrens TEJ, Johansen-Berg H, Bannister PR, De Luca M, Drobnjak I, Flitney DE, et al. 2004. Advances in functional and structural MR image analysis and implementation as FSL. *Neuroimage*. 23:S208–S219.
- Stelzer J, Chen Y, Turner R. 2013. Statistical inference and multiple testing correction in classification-based multi-voxel pattern analysis (MVPA): random permutations and cluster size control. *Neuroimage*. 65:69–82.
- Tata MS, Ward LM. 2005. Spatial attention modulates activity in a posterior “where” auditory pathway. *Neuropsychologia*. 43:509–516.
- Teki S, Grube M, Kumar S, Griffiths TD. 2011. Distinct neural substrates of duration-based and beat-based auditory timing. *J Neurosci*. 31:3805–3812.
- Thaut MH, Trimarchi PD, Parsons LM. 2014. Human brain basis of musical rhythm perception: common and distinct neural substrates for meter, tempo, and pattern. *Brain Sci*. 4:428–452.
- Thiran AB, Clarke S. 2003. Preserved use of spatial cues for sound segregation in a case of spatial deafness. *Neuropsychologia*. 41:1254–1261.
- Tian B, Reser D, Durham A, Kustov A, Rauschecker JP. 2001. Functional specialization in rhesus monkey auditory cortex. *Science*. 292:290–293.
- Valenza N, Murray MM, Ptak R, Vuilleumier P. 2004. The space of senses: impaired crossmodal interactions in a patient with Balint syndrome after bilateral parietal damage. *Neuropsychologia*. 42:1737–1748.
- Wurm MF, Lingnau A. 2015. Decoding actions at different levels of abstraction. *J Neurosci*. 35:7727–7735.
- Zhang F, Wang J-P, Kim J, Parrish T, Wong PCM. 2015. Decoding multiple sound categories in the human temporal cortex using high resolution fMRI. *PLoS One*. 10:e0117303.
- Zundorf IC, Lewald J, Karnath H-O. 2016. Testing the dual-pathway model for auditory processing in human cortex. *Neuroimage*. 124:672–681.