

Neural correlates of binding lyrics and melodies for the encoding of new songs



Irene Alonso^{a,b,c,e}, Lila Davachi^{e,f}, Romain Valabrègue^c, Virginie Lambrecq^{b,c,d},
Sophie Dupont^{b,c}, Séverine Samson^{a,b,*}

^a Laboratoire PSITEC (EA 4072), Université de Lille, France

^b Epilepsy Unit, Hôpital de la Pitié-Salpêtrière, Paris, France

^c UPMC UMR S 1127, Inserm U 1127, CNRS UMR 7225, CENIR, Institut du Cerveau et de la Moelle épinière, Paris, France

^d IHU-ICM, F-75013 Paris, France

^e Department of Psychology, New York University, New York, USA

^f Center for Neural Science, New York University, New York, USA

ARTICLE INFO

Article history:

Received 17 June 2015

Accepted 11 December 2015

Available online 17 December 2015

Keywords:

Memory

Binding

Lyrics

Melodies

Songs

Hippocampus

Basal ganglia

IFG

ABSTRACT

Songs naturally bind lyrics and melody into a unified representation. Using a subsequent memory paradigm, we examined the neural processes associated with binding lyrics and melodies during song encoding. Participants were presented with songs in two conditions: a unified condition (melodies sung with lyrics), and a separate condition (melodies sung with the syllable “la”). In both cases, written lyrics were displayed and participants were instructed to memorize them by repeating them covertly or by generating mental images of the songs. We expected the unified condition to recruit the posterior superior temporal gyrus, known to be involved in perceptual integration of songs, as well as the left inferior frontal gyrus (IFG). Conversely, we hypothesized that the separate condition would engage a larger network including the hippocampus to bind lyrics and melodies of songs, and the basal ganglia and the cerebellum to ensure the correct sequence coupling of verbal and musical information in time. Binding lyrics and melodies in the unified condition revealed activation of the left IFG, bilateral middle temporal gyrus (MTG), and left motor cortex, suggesting a strong linguistic processing for this condition. Binding in the separate compared to the unified condition revealed greater activity in the right hippocampus as well as other areas including the left caudate, left cerebellum, and right IFG. This study provides novel evidence for the role of the right hippocampus in binding lyrics and melodies in songs. Results are discussed in light of studies of binding in the visual domain and highlight the role of regions involved in timing and synchronization such as the basal ganglia and the cerebellum.

© 2015 Published by Elsevier Inc.

Introduction

Although enjoyable, widespread, and seemingly effortless, learning and remembering songs entail high-level cognitive processes that require the binding of verbal and musical information into a unified representation. Neuroimaging studies of song perception have led to enlightening debates on the independence or interaction of music and language networks (Schön et al., 2005). Two studies have suggested the involvement of the posterior areas of the superior temporal gyrus (STG) and superior temporal sulcus (STS) in the integration of lyrics and melodies within songs (Schön et al., 2010; Sammler et al., 2010). Despite these few studies that investigated the perceptual integration of verbal and musical information within songs, the neural mechanisms supporting the binding of lyrics and melodies in memory have received

relatively little attention. To further explore this issue, we examined the cerebral structures underlying the binding of lyrics and melodies during encoding of new songs.

One widely used approach to study successful memory formation has been the subsequent memory paradigm (Wagner et al., 1998; Kim, 2011; Paller and Wagner, 2002). This paradigm allows identifying brain activity during encoding that can predict later success in recognition by comparing subsequently recognized trials with subsequently forgotten trials, known as the subsequent memory effect (SME). Such a paradigm has been instrumental in revealing the distinct contributions of regions in the medial temporal lobe (MTL) to memory (Davachi et al., 2003; Staresina and Davachi, 2006; Staresina and Davachi, 2009; Staresina et al., 2011). One recent study used the subsequent memory paradigm to examine the binding of visual objects features (i.e. shape and colors) in memory (Staresina and Davachi, 2009). In one condition, objects images were presented with a color in a unified manner. In two other conditions, gray-scale objects and color were presented separately. The object in gray-scale was presented with a

* Corresponding author at: Laboratoire PSITEC (EA 4072), Université de Lille, Pont de Bois BP 60149 F-59 653 Villeneuve d'Ascq Cedex, France. Tel.: +33 320416443.

E-mail address: severine.samson@univ-lille3.fr (S. Samson).

color frame displayed either simultaneously with the object (spatial discontinuity condition) or with a short delay (spatial–temporal discontinuity condition). In these discontinuity conditions, participants were instructed to use imagery to bind these features into a unified representation. Interestingly, the magnitude of the SME in the hippocampus, unlike the neighboring perirhinal cortex, increased as a function of the spatiotemporal discontinuity of the presented object and color representations.

To further explore the role of the hippocampus in memory binding, we developed a new paradigm involving auditory information. In this case, participants were required to bind lyrics with their associated melodies during song encoding. Hence, the present study aims at exploring the underlying brain mechanisms that support the binding of lyrics and melodies within songs. We examined subsequent memory for songs under two encoding conditions inspired by [Staresina and Davachi's study \(2009\)](#): 1) following a unified presentation of song components (sung lyrics) and 2) following a separate presentation of song components (simultaneous presentation of sung melody on the syllable “la” and of written lyrics). In both conditions, participants were asked to covertly and repeatedly sing the resulting song until the next trial and to retain the song for a recognition test. Critically, these two conditions are inherently different in terms of the cognitive effort of the participant. The unified presentation is a perceptually richer condition that already provides an integrated signal while the separate condition requires the participant to actively create a mental representation of the song.

Taking in consideration memory models describing the MTL binding function ([Davachi, 2006](#); [Diana et al., 2007](#); [Ranganath, 2010](#)) as well as recent empirical findings in the visual domain ([Staresina and Davachi, 2009](#)), we suggest that the hippocampus may be required to integrate various elements of a song into a unified memory trace. Based on results reported by [Staresina and Davachi \(2009\)](#), we suggest that the involvement of the hippocampus during encoding of songs will enhance when melody and lyrics are presented separately rather than in a unified presentation (sung lyrics), the hippocampus being particularly important to actively integrate separate components of an event.

As previously discussed, binding is required to integrate information both across space and time ([Eichenbaum, 2013](#)). It is therefore worth noting that these dimensions might interact differently in the visual and the auditory domains. On one hand, vision is largely governed by space, requiring rapid binding of synchronous events ([Engel et al., 1997](#)). On the other hand, audition is constrained by time variations ([Sanders et al., 2011](#); [Kiebel et al., 2008](#); [Demany et al., 2010](#)) and consequently requires precise timing and binding of unfolding sequences. Thus, the evoked binding network cooperating with the hippocampus may vary according to spatial and temporal demands of the task. In particular, the cerebellum and the basal ganglia, which are highly interconnected ([Bostan et al., 2013](#)), might be critical for binding auditory information. A non-motor role of the basal ganglia on speech perception has been previously proposed (for a review see [Kotz et al., 2009](#); [Kotz and Schwartze, 2010](#)), and cerebellum activations have been frequently found during song perception and production (for a review see [Callan et al., 2007](#)). From a more general perspective both the cerebellum and basal ganglia have been shown to be crucial for sequence learning ([Graybiel et al., 1994](#); [Kotz et al., 2009](#); [Shin and Ivry, 2003](#)), and for generating temporal ([Cope et al., 2014](#); [Penhune et al., 1998](#); [Schwartz and Kotz, 2013](#); for a review see [Schwartz et al., 2012](#)) and harmonic expectations ([Seeger et al., 2013](#)). It is possible that activation of the cerebellum and basal ganglia may allow temporal coupling of syllables with musical notes during covert rehearsal, enhancing learning of the song sequence through auditory–motor binding. This idea is supported by studies reporting striatal contributions to auditory–motor binding in other not-singing tasks such as synchronized tapping to sounds ([Hove et al., 2013](#)). For instance, [Schmitz et al. \(2013\)](#) have shown functional connectivity between STG and basal ganglia in an auditory–motor binding task. By presenting sounds that were congruent or not with a human avatar breaststroke, the participants

were asked to judge small perceptual differences in velocity in the scanner. The STG and STS showed greater activation for sound–movement bound representations. In line with the suggested temporal coupling role, the basal ganglia and the cerebellum showed greater functional connectivity with the STG when the sounds were congruent to the movement with which they were integrated. This supports the idea that the basal ganglia and cerebellum may be crucial for sensory–motor binding to couple unfolding sound with other distinct elements. Consequently, we hypothesized that the basal ganglia and cerebellum as well as the hippocampus may play an important role in binding lyrics and melodies through covert rehearsal.

Interestingly, speech perception and memory for sentences may strongly rely on semantic and syntactic integration of words into coherent sequences ([Snijders et al., 2009](#)). These integrative processes are considered as binding and have been shown to be supported by the left IFG ([Hagoort, 2005](#); [Opitz and Friederici, 2003](#)). This idea is in line with the dual-stream model of speech proposed by [Hickok and Poeppel \(2007\)](#) and is in agreement with findings reporting an encoding role in the left IFG for words ([Kirchhoff et al., 2000](#)) and semantic associative processes ([Addis and McAndrews, 2006](#)). The binding view proposed by [Opitz \(2010\)](#), further predicts that, although the hippocampus may be involved in the integration of the word chain into a sentence early during encoding (as a form of relational binding already described), the hierarchical rule-based integration of words within their grammatical role and position depends on the left IFG. Based on this literature, it might be possible that the left IFG could be selectively involved in binding contiguous information, at least within the verbal domain. To this regard, melody is an inherent characteristic of the lyrics in the unified condition, hence a stronger involvement of the left IFG than the hippocampus could be expected. Although the integration of linguistic content embedded in a melody has not been studied yet, activation in the left IFG has been observed during musical imitation and vocal learning in singing ([Brown et al., 2004](#)) as well as during perceptual integration of lyrics and melodies of songs ([Schön et al., 2010](#)). This evidence suggests the potential involvement of the left IFG to support the mnemonic binding of sung lyrics, particularly following the covert rehearsal imitating a given song.

In sum, the present study investigated the cerebral network involved in memory binding of songs in two different conditions, where lyrics and melodies were presented either in a unified, or in a separate way. Given the involvement of the posterior areas of the STG/STS and the left IFG in the processing of integrated song representations and the left IFG in binding verbal information and vocal imitation, we predicted that these regions would show a SME in the unified condition. Based on memory models of binding, we hypothesized that the hippocampus would be more involved in binding resulting in greater SME in the separate than in the unified condition in this region. Furthermore, the separate condition may require higher temporal and mental imagery processing demands to allow the coupling of lyrics and melodies in time. Hence, we predicted a SME in the basal ganglia and the cerebellum in the separate condition.

Methods

Participants

Twenty-two healthy native French speakers participated in this study. They were all right-handed as assessed by the Edinburgh Handedness Inventory ([Oldfield, 1971](#)), and reported to have normal hearing. All participants were non-musicians and musical experience was assessed by the Music Expertise Questionnaire ([Ehrlé, 1998](#)), indicating that participants were not music experts (mean score \pm SD: 5.45 ± 2.79 ; max score: 27 points).

Participants were included after a screening phase to verify that they were all able to sing and imagine songs. General auditory imagery was measured with the Bucknell Auditory Imagery Scale ([Zatorre et al.,](#)

2010), confirming that participants had normal auditory imagery abilities (Vividness scale mean: 5.34 ± 0.68 ; Control scale mean: 5.62 ± 0.66). To test the specific imagery of songs during the screening phase, participants listened to six songs and had to repeat them aloud and provide a subjective response on their ability to covertly rehearse each of them. Similarly, participants were also presented six melodies sung with the syllable “la” and accompanied by written lyrics. They were asked to sing aloud the correspondent songs and provide a subjective rating on their ability to imagine each song. To further ensure that participants were able to imagine a song by combining a melody with some written lyrics correctly, we developed a song plausibility test. In this test, participants were presented with 24 melodies and 24 written lyrics, where half of the resulting songs had a one-to-one mapping between the number of syllables in the lyric and the number of sounds in the melody, and half did not. Participants had to imagine the song and respond if the combination of melody and lyrics was plausible or not. After the screening of 37 volunteers, 13 participants were excluded because they did not achieve an accuracy score over 66% in this test. Additionally, two females randomly selected among those with lower plausibility scores (67%) were excluded for gender counterbalance purposes, resulting in the final inclusion of 22 participants (11 female; mean age \pm SD: 24.09 ± 5.45 ; mean year of education \pm SD: 16 ± 1.69 ; mean song plausibility accuracy \pm SD: $74\% \pm 0.05$) in the MRI study.

The local ethics committee has approved this study and all participants provided signed written informed consent beforehand.

Stimuli

Two sets of 54 pairs of songs with interchangeable lyrics and melodies were specifically created based on a collection of 19th Century French folk songs (Robine, 1994). Syllables were matched to musical notes in a one-to-one basis. Each lyric consisted of 6 to 9 pronounced syllables and each melody consisted of 6 to 9 notes. The resulting songs had a variable duration between 2.5 and 3 s. One set of 54 pairs of songs was composed of *original* songs. The other set of 54 pairs of songs was created by recombining the lyrics from one song of the *original* pair with the melody from the other song of the *original* pair and vice versa (*rearranged* songs). By following this procedure, the level of familiarity for lyrics and melodies remained equivalent for the *original* and *rearranged* songs. Melodies of each *original* song were either sung with lyrics or sung using the syllable “la”. Melodies of each *rearranged* song used for recognition were sung with lyrics. The same female professional pop singer recorded all singing stimuli. Additionally, written lyrics were used during encoding.

Procedure

Before entering the scanner, participants received a short song imagery warm-up session based on the plausibility test previously described. Scanning was divided into six sessions to allow pauses every 10 min (without leaving the scanner or moving). Each session included 3 blocks, for a total acquisition of 18 blocks (Fig. 1). Each block included an encoding phase followed by an immediate recognition phase (Fig. 2).

The encoding phase included 6 encoding trials consisting of *original* songs and 2 baseline trials presented in random order. An encoding trial consisted in the double presentation of each *original* song, followed by a fixation-cross for a period of 6 s. During that period, participants were asked to rehearse the song by singing covertly and repeatedly until the next trial. In baseline trials, participants were prompted with a written number presented with white noise for 6 s, to control for visual and auditory stimulation relative to encoding trials. Following the stimulus presentation, a fixation-cross appeared in the screen and remained for a period of 6 s, and participants were asked to count down covertly from the displayed number until the next trial. Thus, the inter-stimulus interval was 12 s varying with a jitter of ± 1 s that was

naturally provided by the difference in length of songs to allow for fitting of HRF. The total duration of an encoding phase was 1 min 36 s. At the end of the encoding phase, a slide was displayed for 7 s to make the transition to the recognition phase.

The recognition phase consisted of 6 *original* songs and 6 *rearranged* songs, all of which were presented in random order. No visual cue was provided for the lyrics during recognition. After the presentation of each song, participants were asked whether the lyric had been previously sung with the presented melody or not, using a confidence rating scale. Participants had 3 s to press one of four buttons: “No, sure”, “No, not sure”, “Yes, not sure”, “Yes, sure”. Each recognition phase took 1 min 24 s.

Blocks were randomly assigned to two different encoding conditions. In the unified condition, *original* songs were presented in a unified manner (sung lyrics). In the separate condition, the melody of the *original* songs was sung using the syllable “la” in place of lyrics and the lyrics were presented visually. In each encoding trial, auditory stimuli were presented twice in a row to facilitate encoding, while written lyrics were simultaneously displayed in the screen. The assignment of songs to either the unified or separate condition was counterbalanced across participants so that all songs were used in both conditions. Blocks from each condition were presented in an interleaved fashion and the order of the blocks was counterbalanced across participants.

The total duration of the functional acquisition was approximately 54 min \pm 5 min depending on the length of the brief pauses between sessions.

fMRI parameters

Functional magnetic resonance imaging (fMRI) was performed on a 3 T Siemens TRIO scanner (Siemens) at the Centre de Neuroimagerie de Recherche at the Pitié-Salpêtrière Hospital in Paris. Before the functional scans, high-resolution T1-weighted images ($1 \times 1 \times 1$ mm voxel size) were acquired for anatomical coregistration using a magnetization-prepared rapid acquisition gradient echo (MPRAGE) sequence (TR = 2300 ms, TE = 4.18 ms).

Six series of 264 volumes of oxygenation level-dependent (BOLD) images were obtained using a single-shot echo-planar gradient-echo (EPI) pulse sequence (TR = 2100 ms, TE = 29 ms, the first four volumes were later discarded to allow for T1 saturation). Forty slices were acquired perpendicular to the hippocampal plane (voxel size 3 mm isotropic, with a slice gap of 0.3 mm) with a field of view of 204×204 mm, and a flip angle of 90° . Additionally seven series of 8 volumes (the first four volumes were later discarded to allow for T1 saturation) were acquired with the same parameters except the phase encoding, which was set in the opposite direction. Those extra series were only used for the deformation field correction purposes as described in Anderson et al. (2003) and Smith et al. (2004). Auditory stimulation was delivered through MR Confon headphones and participants wore earplugs to minimize the scanner noise during stimulus presentation and imagery.

Behavioral analyses

Behavioral data from the recognition phases were analyzed in PASW statistics v18.0 and used for two different purposes: 1) provide an overall measure of recognition success, and 2) provide an individual success score for each encoded item which was later related to fMRI data during encoding. First, recognition responses were analyzed using receiver operating curves, by calculating the Areas Under the Curve (AUC) for the unified and the separate encoding conditions, which took into account both accuracy and confidence ratings.

Individual memory scores for each encoding trial were calculated using recognition accuracy and confidence responses for the related *original* and *rearranged* songs in a combined score. A score of 0.5 was given to each *original* or *rearranged* song that was correctly recognized or correctly rejected with high confidence, and 0.25 for low-confident correct scores. Furthermore, for each missed *original* song or false alarm,

Encoding (fMRI analysis)

Task: “Imagine the correspondent song and sing it covertly and repeatedly until the next trial.”

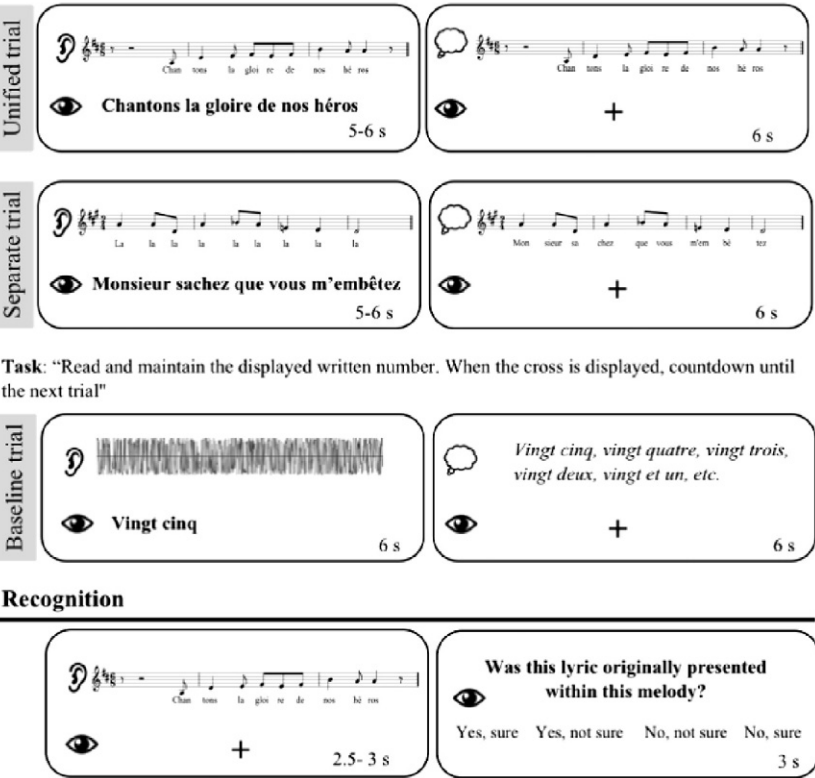


Fig. 1. Experimental trials.

the item was penalized with a score of -0.5 for high-confident responses, or -0.25 for low-confident responses. Thus, the maximum memory score possible was 1, the sum of 0.5 from a high-confident hit of the *original* song and 0.5 from a high-confident rejection of the *rearranged* song (see Table 1a). Trials with a memory score of zero or below were considered unsuccessful or at chance and were transformed to zero, since negative memory scores were not expected to modulate brain activity.

These scores were used for subsequent memory analysis in two ways. For the GLM model, scores were used to tease apart successful (hits) and unsuccessful (misses) trials building separate regressors for SPM analysis. Only trials with both successful recognition of the *original* songs and

rejection of the *rearranged* songs were considered successful. Trials with a score of 0.25, corresponding to a high-confident correct answer ($+0.5$) and a low-confident incorrect answer (-0.25) were left out of the analysis, since they may reflect familiarity processes that are difficult to interpret within the limits of the current design (see Table 1b).

Memory scores (ranging from 0 to 1) were also entered as a parametric modulator of encoding, representing a putative sign of binding success. The parametric analysis of fMRI data assumes a gradation in activation of brain areas involved in binding which may be modulated not only by accuracy, but also by the degree of confidence in these two responses, which might reflect memory strength of the encoded item. In this way, when responses for both *original* and *rearranged* songs are both correct and confident, the studied song may have been better encoded, with greater detail, than when the participant was not sure (see Table 1b).

fMRI GLM analysis

All fMRI analyses were computed using SPM8 (Wellcome Trust Centre for Neuroimaging). Preprocessing of functional images included spatial realignment and coregistration of the functional and anatomical

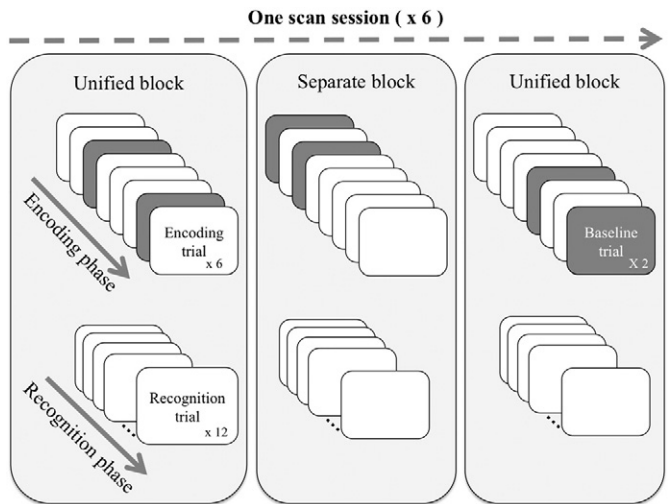


Fig. 2. Scanning session.

Table 1a
Binding success scores according to accuracy and confidence.

		Correct response	
		Correct “yes” (for original)	Correct “no” (for rearranged)
Subject response	Yes	Sure	0.5
		Not sure	0.25
	No	Sure	-0.5
		Not sure	-0.25

Table 1b
Binding scores for fMRI analysis.

	Memory scores	GLM
Both correct and sure	1	Hit
Both correct one sure	0.75	Hit
Both correct non-sure	0.5	Hit
One correct sure & one incorrect not sure	0.25	–
One correct not sure & one incorrect not sure	0	Miss
One correct sure & one incorrect sure	0	Miss
One correct not sure & one incorrect sure	(–0.25) 0	Miss
Both incorrect non-sure	(–0.5) 0	Miss
Both incorrect one sure	(–0.75) 0	Miss
Both incorrect and sure	(–1) 0	Miss

data. Then, images were distortion corrected and normalized to the MNI space and smoothed using a 3D Gaussian kernel of 8 mm full-width at half maximum (FWHM). In the first level, the analysis was carried within a normalized gray matter mask based on the normalized mean anatomical image.

Data was modeled and estimated using the GLM to compare successful versus unsuccessful encoding trials. For the GLM, trials were binned by condition as well as encoding success (success sorting described in the previous section) to create four different regressors (unified-successful, unified-unsuccessful, separate-successful, separate-unsuccessful). Additionally, two regressors were also created for the baseline encoding trials and all recognition trials. Initially, stimulus presentation and imagery periods were modeled separately at the first level (within-subjects level) and then combined at the second level (between-subjects level) to form encoding trials.

First, we looked at task effects using a factorial design with one factor (encoding) and looking at the positive effect of encoding reflected by a *t* contrast. Additionally, we used one-sample *t*-tests to compare activity during successful trials only to the baseline condition, together with one-sample *t*-tests of successful trials in each condition against baseline. General activation on the separate and unified conditions, irrespective of memory performance was compared using one-sample *t*-tests. Second, we looked at SMEs comparing successful against unsuccessful trials using one-sample *t*-tests in each condition.

fMRI parametric analysis

High confidence during recognition of *original* and *rearranged* is an emergent property of memory strength, such that a greater degree of confidence in both responses reflects stronger memory (Shrager et al., 2008; for a review see Wais, 2008). The hippocampus has been shown to be particularly sensitive to strong item and context associations (Davachi, 2006), consistently with the binding role of the hippocampus previously described. Therefore, to look at the combined effect of confidence and accuracy on subsequent memory, we carried out a parametric modulation analysis. We examined the extent to which BOLD activation was modulated by memory performance, using the memory scores for each encoding trial previously described (see Table 1b). By looking at modulations of activity with memory performance at each encoding trial, the parametric approach offers the additional advantage that it is not contaminated by possible differences in difficulty between conditions, as opposed to the GLM. One sample *t*-tests were used at the group level to evaluate SME for each encoding condition (unified and separate) and, most importantly, for the contrast between conditions (separate > unified, and unified > separate).

Correlation between BOLD activation and memory performance

Additionally, to explore the relationship between neural changes and behavior within our *a priori* regions of interest (ROIs), brain activation results from the parametric analysis from each condition were correlated with memory performance across participants. The ROI analysis allows

a more sensitive exploration of relevant areas controlling for Type I error (Poldrack, 2007). We used the Automated Anatomical Labeling (AAL) atlas (Tzourio-Mazoyer et al., 2002) to define neuroanatomical ROIs including subregions within the medial temporal lobe (hippocampus, parahippocampal gyrus), basal ganglia (caudate, pallidum, and putamen) and IFG (orbital part, triangular part, and opercular part).

Beta values from the parametric analysis for the separate and the unified conditions respectively were extracted within each ROI using the MarsBar toolbox (Brett et al., 2002) and were used as a neural measure of encoding. Beta values were then correlated to recognition performance, reflected by each subject's AUC for each condition. Note that both neural and behavioral measures here are sensitive to accuracy and confidence in a similar manner, being therefore suitable correlates. Of note, unlike statistical parametric maps, this analysis illustrates individual differences across participants, with the prediction that participants with greater activation on the ROIs would also show more accurate recognition, incurring in greater successful binding overall. Moreover, a comparison of ROI activation between above and below chance performers was conducted for the separate condition.

Results

a) Behavior.

Recognition memory performance was measured using the AUC for each encoding condition, thus taking accuracy and confidence into account. All participants recognized songs above chance in the unified condition ($p < 0.05$). This suggests that participants could successfully bind lyrics and melodies and that they followed the instructions. In the separate condition, 7 participants did not surpass this threshold and were thus excluded from the SPM analysis. The statistic results of the AUC for each participant in both conditions are reported in Table 2.

b) fMRI GLM analysis.

All SPMs were threshold at $p < 0.001$ (uncorrected) with a minimum cluster extent of 5 voxels (see Table 3). Anatomical location of significant clusters was carried out based on the Automated Anatomical Labeling (AAL) atlas (Tzourio-Mazoyer et al., 2002).

Task effects

A comparison of all encoding trials against the baseline trials revealed significant activation in the bilateral STG, middle temporal gyrus (MTG), left IFG, right caudate, left putamen and several regions of the cerebellum (see Fig. 3A). The full list of regions activated by the task is reported in Table 3. A comparison of unified against separate condition revealed significantly greater activation in the bilateral STG, bilateral MTG as well as orbital part of the left IFG and right anterior cingulum (see Fig. 3B).

Restricting this analysis to successful trials only against the baseline trials, additional activation was found in the right and the left hippocampus, as well as extended clusters in the right IFG and right caudate. When looking at successful trials against baseline in the unified condition only, we found activation in the bilateral caudate, the left putamen and the left hippocampus. Conversely, the separate condition revealed activation in both the right and the left hippocampus, together with clusters in the right and the left caudate.

Subsequent memory effects

Activation related to the subsequent memory effect (SME) for the binding of lyrics and melodies was assessed comparing successful with unsuccessful trials in each condition. SME in the unified condition revealed a cluster in the left IFG together with activation in the bilateral MTG and the left postcentral gyrus (Fig. 4A). Of note, we should specify that a supplementary analysis on the unified condition including all participants did not replicate these results. Conversely, SME during the separate condition revealed significant activation in the right hippocampus,

Table 2
Summary of behavioral results.

Subject	Separate condition			Unified condition			Demographic data				
	AUC	SD	<i>p</i>	AUC	SD	<i>p</i>	Plausibility	Age	Sex	Education	Musical expertise
C01	0.589	0.055	0.110*	0.655	0.053	0.005	0.71	20	f	13	11
C02	0.658	0.053	0.005	0.822	0.042	0.000	0.79	21	m	15	7
C03	0.796	0.044	0.000	0.889	0.033	0.000	0.79	22	f	17	7
C04	0.503	0.056	0.963*	0.714	0.050	0.000	0.79	19	m	12	3
C05	0.547	0.056	0.395*	0.836	0.039	0.000	0.71	20	f	14	2
C06	0.551	0.056	0.363*	0.735	0.048	0.000	0.67	23	f	16	3
C07	0.837	0.040	0.000	0.785	0.044	0.000	0.67	23	f	18	4
C08	0.788	0.043	0.000	0.849	0.036	0.000	0.79	33	f	20	10
C09	0.733	0.049	0.000	0.811	0.042	0.000	0.67	21	m	16	3
C10	0.579	0.056	0.157*	0.664	0.052	0.003	0.75	31	f	17	6
C11	0.706	0.050	0.000	0.843	0.037	0.000	0.71	24	f	17	5
C12	0.639	0.053	0.013	0.685	0.050	0.001	0.71	25	m	17	4
C13	0.723	0.050	0.000	0.717	0.050	0.000	0.83	25	f	14	6
C14	0.718	0.049	0.000	0.867	0.037	0.000	0.67	25	m	16	4
C15	0.863	0.037	0.000	0.865	0.037	0.000	0.75	20	f	16	4
C16	0.687	0.051	0.001	0.701	0.051	0.000	0.71	21	m	16	10
C17	0.628	0.055	0.022	0.644	0.054	0.010	0.83	21	m	16	9
C18	0.763	0.046	0.000	0.800	0.042	0.000	0.79	21	m	16	2
C19	0.560	0.055	0.282*	0.645	0.053	0.009	0.75	43	m	17	4
C20	0.756	0.047	0.000	0.780	0.045	0.000	0.79	26	m	16	3
C21	0.504	0.056	0.941*	0.713	0.050	0.000	0.75	22	m	17	9
C22	0.683	0.051	0.001	0.801	0.044	0.000	0.75	24	f	16	4
Mean	0.67	0.05		0.76	0.04		0.74	24.09	11 f	16.00	5.45
SD	0.11	0.01		0.08	0.01		0.05	5.46	11 m	1.69	2.79

* Performance at chance level. SD: standard deviation; f: female; m: male.

left caudate, the orbital part of the right IFG, right superior frontal gyrus, left middle frontal gyrus, MTG and bilateral cerebellum (Fig. 4B). A full list of regions and mean *t* values within each region is reported in Table 4.

c) fMRI parametric analysis.

To look at SME as a function of both accuracy and confidence and to compare SME effects between our two encoding conditions, we carried out a parametric modulation analysis. Parametric modulation results were also threshold at $p < 0.001$ (uncorrected) with a minimum cluster extent of 5 voxels. Results for the separate condition, unified condition, the comparison between separate and unified conditions and vice versa are fully reported in Table 5.

In the unified condition, SMEs were shown in areas within the left hemisphere, including the left MTG, the left IFG and the left postcentral gyrus (see Fig. 5A). This was not replicated in the supplementary analysis with all participants. In the separate condition, SMEs revealed the engagement a greater number of regions as described in Table 5 (see Fig. 5B).

No differences were observed when comparing the unified against the separate condition. However, the comparison of the separate against the unified condition revealed significant modulation of activity with memory scores in various regions. In the left hemisphere, few clusters of activations were found in the MTG, the insula and the anterior cingulum. In the right hemisphere, activations were found in the hippocampus and the IFG, together with a cluster in the fusiform gyrus, and few cerebellum activations in the left IV–V, left vermis IV–V and right vermis III (see Fig. 5C).

d) Correlation between BOLD activation and memory performance.

Beta values from the parametric analysis were averaged within each anatomically defined ROI and correlated with memory performance for each condition. Similar to previous analyses, only subjects that performed above chance were included in this analysis. No significant correlation was found for the unified condition. For the separate condition, a significant Spearman correlation was shown in the right pallidum ($\rho = 0.44$; $p < 0.05$, one-tailed). Additionally, *t*-tests between above and below chance performers on average beta values within ROIs were conducted for the separate condition, revealing greater activation of the right hippocampus ($t = 2.70$; $df = 20$; $p < 0.05$, two-tailed) (see Fig. 6), left caudate ($t = 3.09$; $df = 20$; $p < 0.01$, two-tailed) and right putamen ($t = 2.31$; $df = 20$; $p < 0.05$, two-tailed) associated with above chance memory performance.

Discussion

The present study aimed to examine the neural substrates underlying the binding of lyrics and melodies during song encoding. To this end, we asked if binding differed as a function of binding demands during song encoding. In the unified condition, participants listened to melodies sung with lyrics while also viewing the written lyrics. By contrast, in the separate condition, participants listened to melodies sung with the syllable “la” while viewing written lyrics and were required to mentally combine these two separate features into a song. Following each stimulus presentation in both conditions, participants were asked to mentally rehearse and maintain each song. We measured mnemonic binding using a memory test in which the correct combination of lyrics and melody corresponding to the *original* songs had to be recognized among other *rearranged* songs.

Behavioral results

Behavioral analysis using receiving operating curves revealed that all 22 participants recognized songs in the unified condition, whereas only 15 participants were able to recognize songs above chance in the separate condition. Despite all participants were able to imagine a song by combining a melody with some written lyrics, as shown by the song plausibility test taken prior to the experiment, behavioral results indicate that the additional cognitive effort required for binding under the separate encoding condition trades off with memory performance when compared to the unified condition. It could be argued that creating a song representation from separate lyrics and melodies poses an additional process for the encoding of a song. Moreover the higher perceptual similarity between encoding and retrieval on the unified condition may have facilitated the recognition under this condition. Of note, we used the syllable “la” for the aural presentation of the melody on the separate condition. This methodological choice, made to enhance the timbre similarity between conditions and facilitate encoding (as opposed to an instrumental timbre, see Weiss et al., 2012), introduces certain verbal interference during the presentation of the stimuli. Thus, it is complicated to dissociate the effects of difficulty from possible differences on the nature of binding itself. Due to this overall difference on performance between conditions, our experimental design cannot

Table 3
GLM task effects.

Cluster size (k)	MNI coordinates	Region	Side	Region size (k)	Mean t
265	56, 0, 46	Precentral	R	216	4.152
		Middle frontal gyrus	R	49	3.997
9	48, 38, 28	Middle frontal gyrus	R	7	3.736
15	56, 32, 2	IFG triangular	R	15	3.652
6778	−66, −22, 8	IFG triangular	L	1664	4.589
		IFG orbital	L	440	4.658
		IFG operculum	L	742	4.453
		Middle frontal gyrus	L	552	4.180
		Middle frontal gyrus orbital	L	98	4.904
		Heschl gyrus	L	35	5.468
		Insula	L	213	3.875
		Postcentral	L	82	3.864
		Precentral	L	999	4.267
		Rolandic operculum	L	97	4.280
		MTG	L	666	5.249
		Superior temporal pole	L	196	5.603
		STG	L	861	5.647
1623	64, −4, −2	STG	R	1008	5.494
		Superior temporal pole	R	297	4.829
		MTG	R	189	4.610
		Rolandic operculum	R	34	4.225
		Heschl gyrus	R	26	5.049
		Middle temporal pole	R	9	3.970
1259	6, 2, 66	Supp motor area	L	569	4.299
		Superior frontal gyrus	L	245	4.021
		Supp motor area	R	205	4.496
		Middle frontal gyrus	L	140	3.867
		Superior frontal gyrus (medial)	L	95	3.789
667	−48, −44, 52	Inferior parietal lobe	L	382	3.862
		Supramarginal	L	161	4.017
		STG	L	96	3.910
13	−20, −72, 6	Calcarine	L	13	3.489
2295	28, −64, 6	Calcarine	R	704	4.009
		Cerebellum VI	R	573	4.332
		Calcarine	L	312	3.839
		Cerebellum Crus1	R	308	3.973
		Lingual	R	175	4.010
		Cuneus	R	56	3.825
		Cuneus	L	44	3.932
		Superior occipital gyrus	R	35	3.872
		Vermis VI	R	33	3.979
		Middle occipital gyrus	R	19	3.778
		Fusiform	R	15	3.799
63	−8, −74, −14	Cerebellum VI	L	57	3.700
		Lingual	L	5	3.560
195	−42, −66, −26	Cerebellum VI	L	110	3.848
		Cerebellum Crus1	L	85	4.001
250	−22, −70, −58	Cerebellum VIII	L	149	4.628
		Cerebellum VIIb	L	8	4.075
918	30, −64, −52	Cerebellum VIII	R	536	4.731
		Cerebellum Crus2	R	228	3.762
		Cerebellum VIIb	R	128	3.836
		Cerebellum Crus1	R	14	3.659
17	−20, 2, 8	Putamen	L	16	3.646
10	20, −6, 24	Caudate	R	10	3.529

exclude the interference of difficulty on comparison between conditions (particularly on the GLM analysis). Consequently, only participants with above chance performance in both conditions were included for neuroimaging analysis, and thus, the SME resulting from the comparison of trials within condition should not be contaminated by difficulty factors.

Task general activation

Given our encoding instructions, general task activations irrespective of memory performance were expected to reflect a network

associated with processes supporting covert singing. Brain activity during the task was shown in the bilateral cortical regions in the IFG, medial frontal gyrus (MFG), superior frontal gyrus, STG, MTG, the temporal pole, the SMA, the left IPL, left supramarginal gyrus areas VI, VIIb and VIII within the cerebellum bilaterally, and subcortical activations in the right caudate and left putamen. These activations are consistent with an extended covert singing network, in line with previous studies on covert singing (Langheim et al., 2002; Callan et al., 2006). Interestingly, the unified condition elicited greater activation in the STG and MTG, together with the orbital part of the left IFG and the left anterior cingulum relative to the separate condition. In the unified condition the auditory input already contained an integrated representation of the song, as opposed to the separate condition. Hence, it is possible that the reported greater general task activations for the unified condition may be reflecting such greater perceptual richness. This interpretation is in line with studies showing the implication of the posterior STG (Sammler et al., 2010; Schön et al., 2010) and the IFG (Schön et al., 2010) on the integration of lyrics and melody during song perception. Since these contrasts merged activation from subsequently recognized and forgotten trials, it is not surprising that no MTL activation was shown.

To have a more detailed idea of activations related to our encoding task, we focused only on successful trials (in which encoding was certainly happening) as compared to baseline trials. This contrast revealed additional activations in bilateral hippocampi, as well as the right IFG and right caudate, supporting the involvement of those structures during the present song encoding task. When we limited this comparison to the separate condition, bilateral activation was found in the hippocampus and the caudate. However, in the unified condition the comparison of subsequent successful encoding trials against baseline only revealed significant activation in the left hippocampus, and not the right. Although not being memory specific, the lateralization of this activation profile is in line with previous results (Alonso et al., 2014), which have shown the detrimental effect of a left hippocampal lesion on the integration of lyrics and melodies during listening to songs presented in a unified manner.

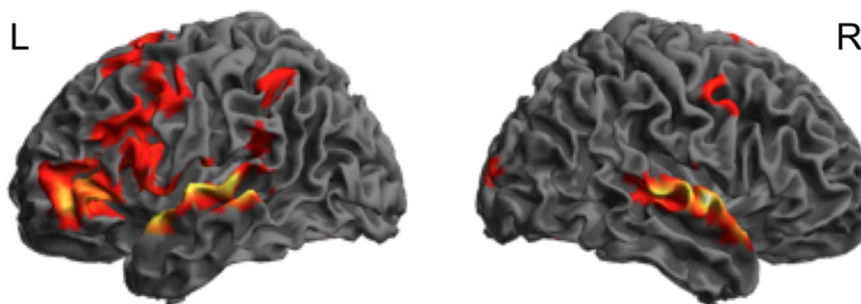
Binding of lyrics and melodies in unified condition

Subsequent memory effects in the unified condition revealed activations in the bilateral MTG, the left postcentral gyrus and, in line with our hypothesis, in the left IFG, suggesting the involvement of an auditory–articulatory network. The parametric analysis, more sensitive to memory strength, showed modulations of activity in relation to SME in similar regions to the GLM in the left hemisphere. Given that songs under the unified condition already provide an integrated percept to be covertly repeated, it is possible that this condition may have strongly taxed the phonological loop (Buchsbbaum et al., 2005; Buchsbbaum and d'Esposito, 2008; Buchsbbaum, 2013). This interpretation is coherent with previous data showing left IFG activation in relation to articulatory learning (Rauschecker et al., 2008). Although we expected the bilateral activation of the posterior STG/STS, cortical activation was found instead in the middle portion of the temporal lobes. The posterior STS and the MTG have been previously shown to be functionally related during the integration of audiovisual information (Beauchamp et al., 2004). It is possible that the visually provided lyrics may have elicited a more multimodal integration than the previous studies in song perception, which highlighted mainly the posterior STG/STS. It is interesting to note that both STG and MTG regions were more active in the unified than the separate condition, already showing a perceptual integration role. Nevertheless, it is difficult to determine why only MTG and not STG showed activation during encoding in the unified condition.

Both the GLM and the parametric analysis showed activation in the left IFG, in agreement with the binding role of Broca's area for language suggested by previous studies (Hagoort, 2005). Our results further extend the role of this area in the particular case of lyrics and melody integration, although these results were not replicated when performing

General task effects

A. Unified & Separate > Baseline



B. Unified > Separate

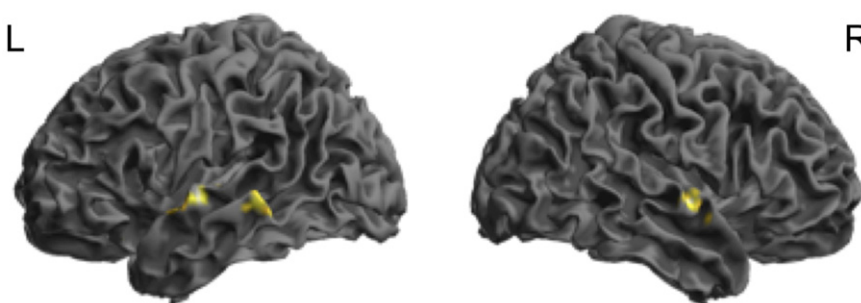


Fig. 3. General task effects.

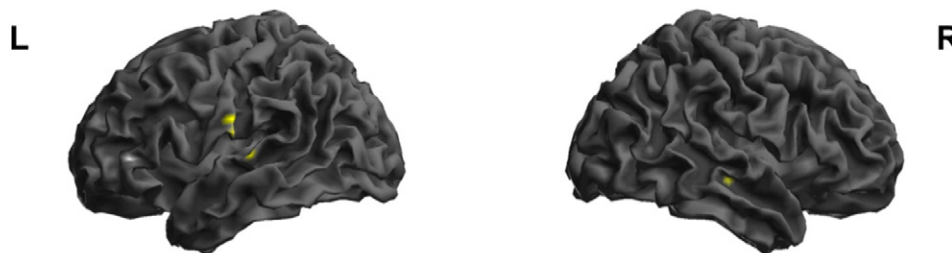
the analysis with all the participants (by including the 7 excluded participants). Taking into account the strong evolutionary impact of language, it may be the case that binding involving verbal (syntactic and articulatory) information, presented in a unified way, may be processed in a specialized region, such as Broca's area. This interpretation is in agreement with previous results showing a key role of the left inferior prefrontal cortex on word assembling and encoding (Clark and Wagner, 2003), notably driven by articulatory components of the

phonological system. In light of the music research literature, our results extend the role of IFG and MTG previously suggested for the recognition and mental imagery of familiar songs (Herholz et al., 2012) to the encoding and mental imagery of new songs. However, it remains open if other non-verbal auditory binding examples, such as for instance timbre and melody, may depend on similar mechanisms, or whether the involvement of the left IFG is restricted to binding verbal information.

GLM subsequent memory effects

A. Unified: Successful > Unsuccessful

$n = 15$; $p < 0.001$ uncorr.



B. Separate: Successful > Unsuccessful

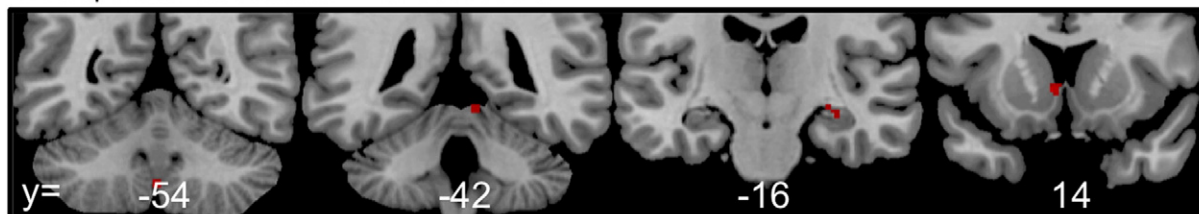


Fig. 4. GLM subsequent memory effects.

Table 4
GLM analysis of encoding.

Unified: successful > unsuccessful					Separate: successful > unsuccessful				
Region	Side	Size (k)	MNI coordinates	Mean t	Region	Side	Size (k)	MNI coordinates	Mean t
IFG-triangular	L	18	−46 34 4	4.35	MFG	L	7	−28 32 30	4.08
Postcentral gyrus	L	20	−64 −18 20	4.11	MTG	L	6	−48 −4 −16	3.91
MTG	L	9	−66 −28 2	4.11	Cerebellum IX	L	11	−2 −52 −44	4.05
MTG	R	6	46 −22 −12	4.21	Caudate	L	10	−6 14 −2	4.03
					SFG	R	8	22 40 34	4.42
					IFG — orbital	R	5	44 24 −10	4.21
					Cerebellum III	R	6	6 −42 −10	3.88
					Hippocampus	R	6	30 −18 −12	3.97

Binding of lyrics and melodies in separate condition

A different network of regions was obtained when lyrics and melodies were presented separately during encoding. In line with our hypothesis, results from the GLM revealed active clusters in the right hippocampus, left caudate, and the cerebellum bilaterally, as well as in other regions including the right IFG, right superior frontal gyrus (SFG), left middle frontal gyrus (MFG) and left MTG. The parametric analysis coincided with the GLM on the implication of the left MTG, left caudate and vermis III. However, the parametric analysis showed a slightly different network, suggesting that confidence nuances on accuracy that could be related to recollection processes might be particularly relevant to understand binding on the separate condition. Modulations of brain activity related to subsequent memory were shown as well in

the right caudate, SFG bilaterally, left MFG, a small region in the left IPL and some additional cerebellum areas.

It may be worth mentioning that, although we did not have an a priori hypothesis for the MTG and other reported SME modulation of activity in areas such as the IPL, the SFG and the anterior cingulum, these structures have been previously related to the inhibition of overt responses in language, essential for covert production of inner speech (Perrone-Bertolotti et al., 2014). Thus, we interpret the present findings in the context of the covert singing imagery demands. However, it should be noted that the MTG also holds a relevant function for processing complex linguistic lexical-meaning relations required for rapid reading (Fujimaki et al., 2004; Simos et al., 2002; Yarkoni et al., 2008). Although the experiment was constructed to depend on an aural representation of a song (recognition was only based on auditory

Table 5
Parametric analysis of encoding.

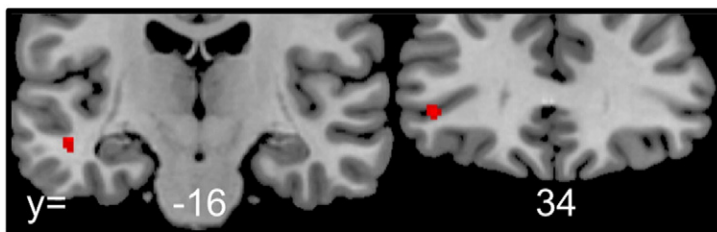
Unified					Separate				
Region (aal)	Cluster size (k)	Region size (k)	MNI coordinates (x, y, z)	Z	Region (aal)	Cluster size (k)	Region size (k)	MNI coordinates (x, y, z)	Z
<i>Left hemisphere</i>					<i>Left hemisphere</i>				
Middle temporal gyrus	7	7	−48 −16 −12	3.42	MTG	23	23	−60 −8 −22	3.78
					SFG (medial)	8	6	−12 26 60	3.77
					SFG (medial)	24	23	−2 52 16	3.36
					IPL	6	5	−62 −48 36	3.57
					Caudate	29	27	−6 12 −2	3.89
					Anterior cingulum	30	25	−4 38 4	3.56
					Cerebellum IV–V	8	4	−14 −32 −16	3.49
IFG — triangular	9	8	−44 34 −2	3.39	<i>Right hemisphere</i>				
Postcentral gyrus	5	4	−52 −4 16	3.30	SFG (medial)	10	8	2 62 14	3.33
<i>Right hemisphere</i>					MFG (p. orbitalis left)	43	30	0 42 −10	3.53
No significant activations					Fusiform gyrus	5	1	34 −70 −8	3.40
					Caudate	6	6	10 10 −10	3.18
					Vermis III	24	9	8 −46 −12	3.82
					Cerebellum II	71	54	24 −84 −38	4.20
					Cerebellum X	5	3	18 −38 −42	3.74
Unified > separate					Separate > unified				
Region (aal)	Cluster size (k)	Region size (k)	MNI coordinates (x, y, z)	Z	Region (aal)	Cluster size (k)	Region size (k)	MNI coordinates (x, y, z)	Z
<i>No significant activations for this condition</i>					<i>Left hemisphere</i>				
					MTG	30	30	−56 −4 −22	4.27
					Insula	16	11	−38 −12 −4	4.58
					Anterior cingulum	9	9	−6 40 6	3.45
					Anterior cingulum	7	5	−4 30 14	3.60
					Cerebellum IV–V	45	20	−14 −30 −14	3.73
					Vermis IV–V	18	7	−4 −64 −10	3.49
					<i>Right hemisphere</i>				
					IFG (p. orbitalis)	8	6	32 18 −22	3.33
					Fusiform gyrus	43	32	24 −74 −14	3.79
					Fusiform gyrus	6	5	26 −56 −12	3.27
					Right hippocampus	6	6	32 −20 −12	3.36
					Vermis III	43	23	6 −42 −10	3.85

Significant clusters ($p < 0.001$; $k \geq 5$), and the regions within them from the “automated anatomical atlas”.

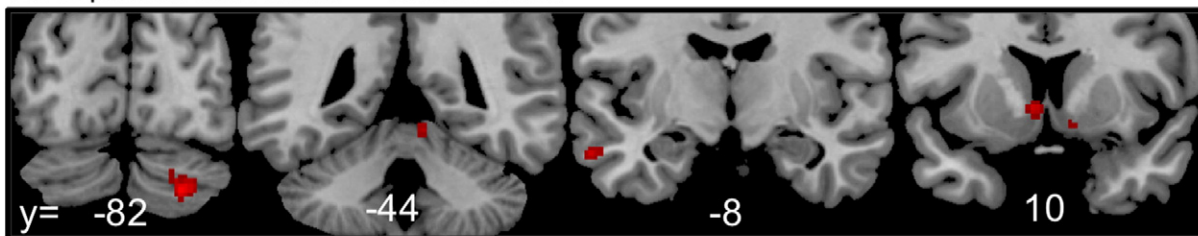
Parametric analysis subsequent memory effects

A. Unified: Successful > Unsuccessful

$n = 15$; $p < 0.001$ uncorr.



B. Separate: Successful > Unsuccessful



C. Successful: Separate > Unified

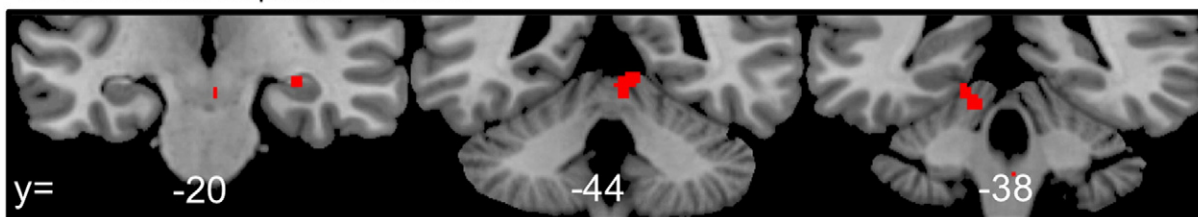


Fig. 5. Parametric analysis subsequent memory effects.

stimuli and the imagery period during encoding prompted the covert rehearsal of an auditory representation), it is difficult to quantify the possible contribution of the written lyric support that could have enhanced the activation of the left MTG during both unified and separate conditions.

Taken together, the GLM and the parametric analysis highlight the implication of the basal ganglia, specially the left caudate, and the cerebellum bilaterally, in line with our predictions. Moreover, a correlation was found between memory performance and right pallidum activation, together with greater activation of the left caudate and right putamen for above chance performers as compared to participants that performed below chance. Thus, the implication of the basal ganglia during encoding on the separate condition is consistent both within and across subjects. This result suggests the involvement of sequencing and timing networks (Penhune et al., 1998; Nenadic et al., 2003; Ivry and Spencer, 2004; Tillmann et al., 2008; Callan et al., 2007; Schwartze et al., 2012) for the binding of auditory information. Moreover, the separate condition showed greater SME modulation of activity as compared to the unified condition in various areas that seem to belong to an auditory-motor integration network (Zatorre et al., 2007; Brown et al., 2004), including the cerebellum, the left insula, left MTG, left anterior cingulate, right fusiform gyrus, right orbital IFG and, critically, the right hippocampus. These areas are relevant for the coupling of perception and action and have been already reported in studies involving speech and singing (Ackermann and Riecker, 2004; Callan et al., 2007; Herholz et al., 2012). In particular, we argue that the cerebellum seems to be crucial to provide a precise timing context (Janata and Grafton, 2003; Ivry and Spencer, 2004; Callan et al., 2007; Schwartze et al., 2012) in which lyrics and melodies could be bound when presented separately.

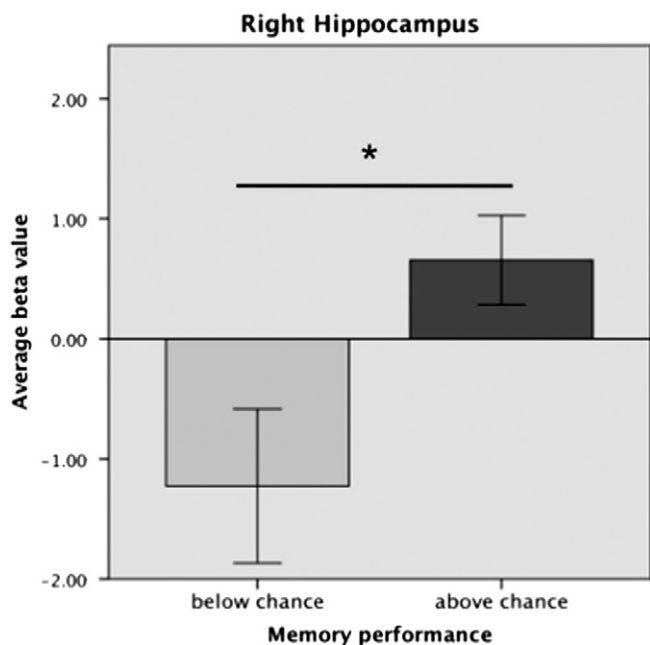


Fig. 6. Difference in right hippocampus activation during the separate condition between below and above chance performers (* $p < 0.05$). Error bars represent standard error of the mean.

Although the activation of the right hippocampus found in the GLM was not significant in the parametric analysis at a robust threshold, greater activation was found in the right hippocampus during the separate condition as compared with the unified condition, in line with our predictions. The correlation analysis of brain activity and memory performance across participants that performed above chance did not show any significant correlation. However, a comparison between above and below chance performers revealed greater activation of the right hippocampus for above chance performers. This pattern of results could suggest a threshold memory process for the involvement of the right hippocampus during encoding on the separate condition. The greater implication of the hippocampus on the separate condition is in agreement with similar studies in the visual domain suggesting an increasing involvement of the hippocampus with greater discontinuity of the elements to bind (Staresina and Davachi, 2009). However, a crucial novelty in our design is that subjects were asked to bind trial-unique information, as melodies from every song were different. Thus, the cognitive demands may differ from those in visual studies presenting a limited number of contexts (i.e. four colors) for possible binding (Staresina and Davachi, 2009).

This joint evidence suggests the cooperation of the right hippocampus with basal ganglia to achieve the binding of lyrics and melodies that are presented separately. The striatum and hippocampus have previously been found to cooperate via the prefrontal cortex in a variety of tasks (Seger, 2006), including encoding of words (Sadeh et al., 2011; for a review see Battaglia et al., 2011), spatial navigation (Brown et al., 2012; Hartley and Burgess, 2005), sequence learning and consolidation (Albouy et al., 2008; Kotz et al., 2009; Tubridy and Davachi, 2011; Schwartz et al., 2012; Albouy et al., 2013; Chan et al., 2013). This large body of research has challenged the competitive view of hippocampal and basal ganglia function, and led to a growing number of studies to also assess mnemonic functions of the striatum for both procedural and declarative memory (Scimeca and Badre, 2012; Badre et al., 2014; Censor et al., 2014). The present findings are in agreement with this new approach, and suggest that, when auditory information that unfolds over time is involved, the striatum may cooperate with the hippocampus to achieve encoding.

Parametric analyses further highlight the combined activation of frontal areas with the cerebello-striatal circuit. This is consistent with previous studies implicating such a network in the integration of temporal structure and memory content, including processing of durations (Hälbjg et al., 2002; Teki et al., 2012) and subsequent long-term storage of temporal structure (Schwartz et al., 2012). One possible interpretation of the greater right IFG activation found during separate as compared to unified condition is that it could reflect inhibition (for a review see Aron et al., 2004) related to the verbal suppression (Robinson et al., 2015) of the syllable “la”, which was required to combine the written lyrics with the melody. However, some studies suggest that the role of the right IFG may not be limited to inhibition (see Hampshire, 2015), supporting its involvement for the detection of important cues (Hampshire et al., 2010) and for encoding (Addis and McAndrews, 2006; Bor et al., 2004). Interestingly, a recent case study (Herbet et al., 2015) suggests that the direct stimulation of the pars opercularis of the right IFG may be related to the melodically intoned articulation of words such as during singing. Although a particular cognitive role of the right IFG remains unclear, laterality differences in the IFG have been previously reported in a study comparing the perception of songs to speech sounds (Merrill et al., 2012), which reported right and left lateralized activations respectively. Despite the fact that in our study songs were presented in both conditions, one plausible interpretation is that encoding under unified and separate conditions may have respectively rely more on the verbal (Opitz and Friederici, 2003) or on the musical structure (Koelsch and Friederici, 2003), resulting in the laterality differences found in the IFG.

Interestingly, although previous studies have described a particular role of the posterior STG during song integration (Sammler et al.,

2010; Alonso et al., 2014), our results did not show the involvement of this area in the separate condition. It is possible that integration effects found previously in this region may be dependent on the physical properties of songs that already integrate both components into an auditory signal, and thus are specific to encoding of sung lyrics. Conversely the profile of activity shown in response to the separate condition may correspond to a more independent perceptual processing of lyrics and melodies. Thus, we argue that the binding of song elements presented separately may have enhanced the involvement of other structures such as the right hippocampus, left caudate and cerebellum to bridge the gap between lyrics and melodies and encode an integrated song representation.

Conclusion

The present study examined the neural substrate underlying the binding of lyrics and melodies during song encoding. Overall, we observe involvement of the right hippocampus along with an auditory-motor network engaged in timing and sequencing processing for binding lyrics and melodies when they are presented separately. Furthermore, we provide novel evidence that the left IFG is involved in the binding of verbal and melodic information that is presented in a unified song. Although the present study represents a first step to address the implication of the hippocampus in binding songs, future studies isolating factors such as number of contexts, salience, semantic content or rhythmical structure of the songs may contribute to better understand song encoding. Furthermore, in light of these results, we suggest that other structures such as the IFG, basal ganglia and the cerebellum are relevant for binding auditory information. We encourage future studies to explore the particular role of these brain structures in more detail to complete our understanding of binding auditory information within and between modalities.

Acknowledgments

The authors are grateful to Alexa Tomparry for her helpful insights on this study. The research leading to these results has received funding from an ERS Marie-Curie fellowship to Irene Alonso by the European Community's Seventh Framework Programme under the EBRAMUS project – grant agreement no. 238157, a personal fellowship to Irene Alonso by “LaCaixa Foundation”, from the program “Investissements d’avenir” ANR-10-IAIHU-06 and from Institut Universitaire de France to Séverine Samson.

References

- Ackermann, H., Riecker, A., 2004. The contribution of the insula to motor aspects of speech production: a review and a hypothesis. *Brain Lang.* 89 (2), 320–328.
- Addis, D.R., McAndrews, M.P., 2006. Prefrontal and hippocampal contributions to the generation and binding of semantic associations during successful encoding. *NeuroImage* 33 (4), 1194–1206.
- Aron, A.R., Robbins, T.W., Poldrack, R.A., 2004. Inhibition and the right inferior frontal cortex. *Trends Cogn. Sci.* 8 (4), 170–177.
- Albouy, G., King, B.R., Maquet, P., Doyon, J., 2013. Hippocampus and striatum: dynamics and interaction during acquisition and sleep-related motor sequence memory consolidation: hippocampus and striatum and procedural memory consolidation. *Hippocampus* 23 (11), 985–1004.
- Albouy, G., Sterpenich, V., Baeteau, E., Vandewalle, G., Desseilles, M., Dang-Vu, T., Darsaud, A., Ruby, P., Luppi, P.-H., Degueldre, C., Peigneux, P., Luxen, A., Maquet, P., 2008. Both the hippocampus and striatum are involved in consolidation of motor sequence memory. *Neuron* 58 (2), 261–272.
- Alonso, I., Sammler, D., Valabrègue, R., Dinkelacker, V., Dupont, S., Belin, P., Samson, S., 2014. Hippocampal sclerosis affects fMRI-adaptation of lyrics and melodies in songs. *Front. Hum. Neurosci.* 8, 111.
- Andersson, J.L., Skare, S., Ashburner, J., 2003. How to correct susceptibility distortions in spin-echo echo-planar images: application to diffusion tensor imaging. *NeuroImage* 20 (2), 870–888.
- Badre, D., Lebrecht, S., Pagliaccio, D., Long, N.M., Scimeca, J.M., 2014. Ventral striatum and the evaluation of memory retrieval strategies. *J. Cogn. Neurosci.* 26 (9), 1928–1948.
- Battaglia, F.P., Benchenane, K., Sirota, A., Pennartz, C.M., Wiener, S.I., 2011. The hippocampus: hub of brain network communication for memory. *Trends Cogn. Sci.* 15 (7), 310–318.

- Beauchamp, M.S., Lee, K.E., Argall, B.D., Martin, A., 2004. Integration of auditory and visual information about objects in superior temporal sulcus. *Neuron* 41 (5), 809–823.
- Bor, D., Cumming, N., Scott, C.E., Owen, A.M., 2004. Prefrontal cortical involvement in verbal encoding strategies. *Eur. J. Neurosci.* 19 (12), 3365–3370.
- Bostan, A.C., Dum, R.P., Strick, P.L., 2013. Cerebellar networks with the cerebral cortex and basal ganglia. *Trends Cogn. Sci.* 17 (5), 241–254.
- Brett, M., Anton, J.-L., Valabregue, R., Poline, J.-B., 2002. Region of interest analysis using an SPM toolbox [abstract]. Presented at the 8th International Conference on Functional Mapping of the Human Brain, June 2–6, 2002, Sendai, Japan (Available on CD-ROM in) *NeuroImage* 16 (2) (abstract 497).
- Brown, S., Martinez, M.J., Hodges, D.A., Fox, P.T., Parsons, L.M., 2004. The song system of the human brain. *Cogn. Brain Res.* 20 (3), 363–375.
- Brown, T.I., Ross, R.S., Tobyn, S.M., Stern, C.E., 2012. Cooperative interactions between hippocampal and striatal systems support flexible navigation. *NeuroImage* 60 (2), 1316–1330.
- Buchsbaum, B.R., 2013. The role of consciousness in the phonological loop: hidden in plain sight. *Front. Psychol.* 4, 496.
- Buchsbaum, B.R., D'Esposito, M., 2008. The search for the phonological store: from loop to convolution. *J. Cogn. Neurosci.* 20 (5), 762–778.
- Buchsbaum, B.R., Olsen, R.K., Koch, P., Berman, K.F., 2005. Human dorsal and ventral auditory streams subservise rehearsal-based and echoic processes during verbal working memory. *Neuron* 48 (4), 687–697.
- Callan, D., Kawato, M., Parsons, L., Turner, R., 2007. Speech and song: the role of the cerebellum. *Cerebellum* 6 (4), 321–327.
- Callan, D.E., Tsytsarev, V., Hanakawa, T., Callan, A.M., Katsuhara, M., Fukuyama, H., Turner, R., 2006. Song and speech: brain regions involved with perception and covert production. *Neuroimage* 31 (3), 1327–1342.
- Censor, N., Horowitz, S., Cohen, L., 2014. Interference with existing memories alters offline intrinsic functional brain connectivity. *Neuron* 81 (1), 69–76.
- Chan, S.-H., Ryan, L., Bever, T.G., 2013. Role of the striatum in language: syntactic and conceptual sequencing. *Brain Lang.* 125 (3), 283–294.
- Clark, D., Wagner, A.D., 2003. Assembling and encoding word representations: fMRI subsequent memory effects implicate a role for phonological control. *Neuropsychologia* 41 (3), 304–317.
- Cope, T.E., Grube, M., Singh, B., Burn, D.J., Griffiths, T.D., 2014. The basal ganglia in perceptual timing: timing performance in multiple system atrophy and Huntington's disease. *Neuropsychologia* 52, 73–81.
- Davachi, L., 2006. Item, context and relational episodic encoding in humans. *Curr. Opin. Neurobiol.* 16 (6), 693–700.
- Davachi, L., Mitchell, J.P., Wagner, A.D., 2003. Multiple routes to memory: distinct medial temporal lobe processes build item and source memories. *Proc. Natl. Acad. Sci.* 100 (4), 2157–2162.
- Demany, L., Semal, C., Cazalets, J.-R., Pressnitzer, D., 2010. Fundamental differences in change detection between vision and audition. *Exp. Brain Res.* 203 (2), 261–270.
- Diana, R.A., Yonelinas, A.P., Ranganath, C., 2007. Imaging recollection and familiarity in the medial temporal lobe: a three-component model. *Trends Cogn. Sci.* 11 (9), 379–386.
- Eichenbaum, H., 2013. Memory on time. *Trends Cogn. Sci.* 17 (2), 81–88.
- Engel, A.K., Roelfsema, P.R., Fries, P., Brecht, M., Singer, W., 1997. Role of the temporal domain for response selection and perceptual binding. *Cereb. Cortex* 7, 571–582.
- Ehrlé, N., 1998. Traitement temporel de l'information auditive et lobe temporal Unpublished doctoral dissertation Université de Reims, Reims.
- Fujimaki, N., Hayakawa, T., Munetsuna, S., Sasaki, T., 2004. Neural activation dependent on reading speed during covert reading of novels. *Neuroreport* 15 (2), 239–243.
- Graybiel, A.M., Aosaki, T., Flaherty, A.W., Kimura, M., 1994. The basal ganglia and adaptive motor control. *Science* 265, 1826–1831.
- Hagoort, P., 2005. On broca, brain, and binding: a new framework. *Trends Cogn. Sci.* 9 (9), 416–423.
- Hälbjg, T. F., von Cramon, D. Y., Schmid, U. D., Gall, C., & Friederici, A. D. (2002). Processing of temporal duration information in working memory after frontodorsal tumor excisions. *Brain Cogn.* 50, 282–303.
- Hampshire, A., 2015. Putting the brakes on inhibitory models of frontal lobe function. *NeuroImage* 113, 340–355.
- Hampshire, A., Chamberlain, S.R., Monti, M.M., Duncan, J., Owen, A.M., 2010. The role of the right inferior frontal gyrus: inhibition and attentional control. *NeuroImage* 50 (3), 1313–1319.
- Hartley, T., Burgess, N., 2005. Complementary memory systems: competition, cooperation and compensation. *Trends Neurosci.* 28 (4), 169–170.
- Herbet, G., Lafargue, G., Almairac, F., Moritz-Gasser, S., Bonnetblanc, F., Duffau, H., 2015. Disrupting the right pars opercularis with electrical stimulation frees the song: case report. *J. Neurosurg.* 1–4.
- Herholz, S.C., Halpern, A.R., Zatorre, R.J., 2012. Neuronal correlates of perception, imagery, and memory for familiar tunes. *J. Cogn. Neurosci.* 24 (6), 1382–1397.
- Hickok, G., Poeppel, D., 2007. The cortical organization of speech processing. *Nat. Rev. Neurosci.* 8 (5), 393–402.
- Hove, M.J., Fairhurst, M.T., Kotz, S.A., Keller, P.E., 2013. Synchronizing with auditory and visual rhythms: an fMRI assessment of modality differences and modality appropriateness. *Neuroimage* 67, 313–321.
- Ivry, R., Spencer, R.M.C., 2004. The neural representation of time. *Curr. Opin. Neurobiol.* 14 (2), 225–232.
- Janata, P., Grafton, S.T., 2003. Swinging in the brain: shared neural substrates for behaviors related to sequencing and music. *Nat. Neurosci.* 6 (7), 682–687.
- Kiebel, S.J., Daunizeau, J., Friston, K.J., 2008. A hierarchy of time-scales and the brain. *PLoS Comput. Biol.* 4 (11), e1000209.
- Kirchhoff, B.A., Wagner, A.D., Maril, A., Stern, C.E., 2000. Prefrontal-temporal circuitry for episodic encoding and subsequent memory. *J. Neurosci.* 20 (16), 6173–6180.
- Kim, H., 2011. Neural activity that predicts subsequent memory and forgetting: a meta-analysis of 74 fMRI studies. *NeuroImage* 54 (3), 2446–2461.
- Koelsch, S., Friederici, A.D., 2003. Toward the neural basis of processing structure in music. *Ann. N. Y. Acad. Sci.* 999 (1), 15–28.
- Kotz, S.A., Schwartz, M., 2010. Cortical speech processing unplugged: a timely subcortical-cortical framework. *Trends Cogn. Sci.* 14 (9), 392–399.
- Kotz, S.A., Schwartz, M., Schmidt-Kassow, M., 2009. Non-motor basal ganglia functions: a review and proposal for a model of sensory predictability in auditory language perception. *Cortex* 45 (8), 982–990.
- Langheim, F.J., Callicott, J.H., Mattay, V.S., Duyn, J.H., Weinberger, D.R., 2002. Cortical systems associated with covert music rehearsal. *NeuroImage* 16 (4), 901–908.
- Merrill, J., Sammler, D., Bangert, M., Goldhahn, D., Lohmann, G., Turner, R., Friederici, A.D., 2012. Perception of words and pitch patterns in song and speech. *Front. Psychol.* 3, 76.
- Nenadic, I., Gaser, C., Volz, H., Rammsayer, T., Häger, F., Sauer, H., 2003. Processing of temporal information and the basal ganglia: new evidence from fMRI. *Exp. Brain Res.* 148, 238–246.
- Oldfield, R.C., 1971. The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia* 9, 97–113.
- Opitz, B., 2010. Neural binding mechanisms in learning and memory. *Neurosci. Biobehav. Rev.* 34 (7), 1036–1046.
- Opitz, B., Friederici, A.D., 2003. Interactions of the hippocampal system and the prefrontal cortex in learning language-like rules. *NeuroImage* 19 (4), 1730–1737.
- Paller, K.A., Wagner, A.D., 2002. Observing the transformation of experience into memory. *Trends Cogn. Sci.* 6 (2), 93–102.
- Penhune, V.B., Zatorre, R.J., Evans, A.C., 1998. A pet study of auditory and visual rhythm reproduction. *J. Cogn. Neurosci.* 10 (6), 752–765.
- Perrone-Bertolotti, M., Rapin, L., Lachaux, J.-P., Baciou, M., Lœvenbruck, H., 2014. What is that little voice inside my head? Inner speech phenomenology, its role in cognitive performance, and its relation to self-monitoring. *Behav. Brain Res.* 261, 220–239.
- Poldrack, R.A., 2007. Region of interest analysis for fMRI. *Soc. Cogn. Affect. Neurosci.* 2 (1), 67–70.
- Robine, M., 1994. In: Michel, A. (Ed.), *Anthologie de la chanson française –des trouvères aux grands auteurs du XIXème siècle*. Albin Michel, Paris.
- Robinson, G.A., Cipolletti, L., Walker, D.G., Biggs, V., Bozzali, M., Shallice, T., 2015. Verbal suppression and strategy use: a role for the right lateral prefrontal cortex? *Brain* 138 (4), 1084–1096.
- Ranganath, C., 2010. Binding items and contexts the cognitive neuroscience of episodic memory. *Curr. Dir. Psychol. Sci.* 19 (3), 131–137.
- Rauschecker, A.M., Pringle, A., Watkins, K.E., 2008. Changes in neural activity associated with learning to articulate novel auditory pseudowords by covert repetition. *Hum. Brain Mapp.* 29 (11), 1231–1242.
- Sadeh, T., Shohamy, D., Levy, D.R., Reggev, N., Maril, A., 2011. Cooperation between the hippocampus and the striatum during episodic encoding. *J. Cogn. Neurosci.* 23 (7), 1597–1608.
- Sammler, D., Baird, A., Valabregue, R., Clement, S., Dupont, S., Belin, P., Samson, S., 2010. The relationship of lyrics and tunes in the processing of unfamiliar songs: a functional magnetic resonance adaptation study. *J. Neurosci.* 30 (10), 3572–3578.
- Sanders, M.C., Chang, N.-Y.N., Hiss, M.M., Uchanski, R.M., Hullar, T.E., 2011. Temporal binding of auditory and rotational stimuli. *Exp. Brain Res.* 210 (3–4), 539–547.
- Schmitz, G., Mohammadi, B., Hammer, A., Heldmann, M., Samii, A., Münte, T.F., Effenberg, A.O., 2013. Observation of sonified movements engages a basal ganglia frontocortical network. *BMC Neurosci.* 14 (1), 32.
- Schön, D., Gordon, R., Campagne, A., Magne, C., Astésano, C., Anton, J., Besson, M., 2010. Similar cerebral networks in language, music and song perception. *NeuroImage* 51 (1), 450–461.
- Schön, D., Gordon, R.L., Besson, M., 2005. Musical and linguistic processing in song perception. *Ann. N. Y. Acad. Sci.* 1060 (1), 71–81.
- Schwartz, M., Tavano, A., Schröger, E., Kotz, S.A., 2012. Temporal aspects of prediction in audition: cortical and subcortical neural mechanisms. *Int. J. Psychophysiol.* 83 (2), 200–207.
- Schwartz, M., Kotz, S.A., 2013. A dual-pathway neural architecture for specific temporal prediction. *Neurosci. Biobehav. Rev.* 37 (10), 2587–2596.
- Scimeca, J., Badre, D., 2012. Striatal contributions to declarative memory retrieval. *Neuron* 75 (3), 380–392.
- Seger, C.A., 2006. Dynamics of frontal, striatal, and hippocampal systems during rule learning. *Cereb. Cortex* 16 (11), 1546–1555.
- Seger, C.A., Spiering, B.J., Sares, A.G., Quraini, S.I., Alpeter, C., David, J., et al., 2013. Corticostriatal contributions to musical expectancy perception. *J. Cogn. Neurosci.* 25 (7), 1062–1077.
- Shin, J.C., Ivry, R.B., 2003. Spatial and temporal sequence learning in patients with Parkinson's disease or cerebellar lesions. *J. Cogn. Neurosci.* 15 (8), 1232–1243.
- Shrager, Y., Kirwan, C.B., Squire, L.R., 2008. Activity in both hippocampus and perirhinal cortex predicts the memory strength of subsequently remembered information. *Neuron* 59 (4), 547–553.
- Simos, P.G., Breier, J.L., Fletcher, J.M., Foorman, B.R., Castillo, E.M., Papanicolaou, A.C., 2002. Brain mechanisms for reading words and pseudowords: an integrated approach. *Cereb. Cortex* 12 (3), 297–305.
- Snijders, T.M., Vosse, T., Kempen, G., Van Berkum, J.J., Petersson, K.M., Hagoort, P., 2009. Retrieval and unification of syntactic structure in sentence comprehension: an fMRI study using word-category ambiguity. *Cereb. Cortex* 19 (7), 1493–1503.
- Smith, S.M., Jenkinson, M., Woolrich, M.W., Beckmann, C.F., Behrens, T.E.J., Johansen-Berg, H., Bannister, P.R., De Luca, M., Drobnjak, I., Flitney, D.E., Niazy, R., Saunders, J., Vickers, J., Zhang, Y., De Stefano, N., Brady, J.M., Matthews, P.M., 2004. Advances in functional and structural MR image analysis and implementation as FSL. *NeuroImage* 23 (S1), 208–219.

- Staresina, B.P., Davachi, L., 2006. Differential encoding mechanisms for subsequent associative recognition and free recall. *J. Neurosci.* 26 (36), 9162–9172.
- Staresina, B.P., Davachi, L., 2009. Mind the gap: binding experiences across space and time in the human hippocampus. *Neuron* 63 (2), 267–276.
- Staresina, B.P., Duncan, K.D., Davachi, L., 2011. Perirhinal and parahippocampal cortices differentially contribute to later recollection of object- and scene-related event details. *J. Neurosci.* 31 (24), 8739–8747.
- Teki, S., Grube, M., Griffiths, T.D., 2011. A unified model of time perception accounts for duration-based and beat-based timing mechanisms. *Front. Integr. Neurosci.* 5, 90.
- Tillmann, B., Justus, T., Bigand, E., 2008. Cerebellar patients demonstrate preserved implicit knowledge of association strengths in musical sequences. *Brain Cogn.* 66 (2), 161–167.
- Tubridy, S., Davachi, L., 2011. Medial temporal lobe contributions to episodic sequence encoding. *Cereb. Cortex* 21 (2), 272–280.
- Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., Mazoyer, B., Joliot, M., 2002. Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *NeuroImage* 15 (1), 273–289.
- Wais, P.E., 2008. fMRI signals associated with memory strength in the medial temporal lobes: a meta-analysis. *Neuropsychologia* 46 (14), 3185–3196.
- Weiss, M.W., Trehub, S.E., Schellenberg, E.G., 2012. Something in the way she sings enhanced memory for vocal melodies. *Psychol. Sci.* 23 (10), 1074–1078.
- Wagner, A.D., Schacter, D.L., Rotte, M., Koutstaal, W., Maril, A., Dale, A.M., Rosen, B.R., 1998. Building memories: remembering and forgetting of verbal experiences as predicted by brain activity. *Science* 281 (5380), 1188–1191.
- Yarkoni, T., Speer, N.K., Balota, D.A., McAvoy, M.P., Zacks, J.M., 2008. Pictures of a thousand words: investigating the neural mechanisms of reading with extremely rapid event-related fMRI. *NeuroImage* 42 (2), 973–987.
- Zatorre, R.J., Chen, J.L., Penhune, V.B., 2007. When the brain plays music: auditory–motor interactions in music perception and production. *Nat. Rev. Neurosci.* 8 (7), 547–558.
- Zatorre, R.J., Halpern, A.R., Bouffard, M., 2010. Mental reversal of imagined melodies: a role for the posterior parietal cortex. *J. Cogn. Neurosci.* 22 (4), 775–789.