Behavioral/Systems/Cognitive

The Relationship of Lyrics and Tunes in the Processing of Unfamiliar Songs: A Functional Magnetic Resonance Adaptation Study

Daniela Sammler,[1,2,3] **Amee Baird,**[1] **Romain Valabrègue,**[3,4] **Sylvain Clément,**[1] **Sophie Dupont,**[2,4] **Pascal Belin,**[5,6] **and Séverine Samson**[1,2]

[1]Neuropsychologie et Cognition Auditive-JE2497, Université de Lille-Nord de France, 59653 Villeneuve d'Ascq, France, [2]Hôpital de la Pitié-Salpêtrière, 75651 Paris, France, [3]Centre de Neuroimagerie de Recherche, Paris, France, [4]Centre de Recherche de l'Institut du Cerveau et de la Moëlle Épinière, Universite Pierre et Marie Curie–Unité Mixte de Recherche 7225 Centre National de la Recherche Scientifique–UMRS 975 INSERM, Paris, France, [5]Centre for Cognitive Neuroimaging, Department of Psychology, University of Glasgow, Glasgow G12 8QB, United Kingdom, and [6]Laboratories for Brain, Music, and Sound, Université de Montréal, Montreal, Quebec H3C 3J7 and McGill University, Montréal, Quebec H3A 2T5, Canada

The cognitive relationship between lyrics and tunes in song is currently under debate, with some researchers arguing that lyrics and tunes are represented as separate components, while others suggest that they are processed in integration. The present study addressed this issue by means of a functional magnetic resonance adaptation paradigm during passive listening to unfamiliar songs. The repetition and variation of lyrics and/or tunes in blocks of six songs was crossed in a 2 × 2 factorial design to induce selective adaptation for each component. Reductions of the hemodynamic response were observed along the superior temporal sulcus and gyrus (STS/STG) bilaterally. Within these regions, the left mid-STS showed an interaction of the adaptation effects for lyrics and tunes, suggesting an integrated processing of the two components at prelexical, phonemic processing levels. The degree of integration decayed toward more anterior regions of the left STS, where the lack of such an interaction and the stronger adaptation for lyrics than for tunes was suggestive of an independent processing of lyrics, perhaps resulting from the processing of meaning. Finally, evidence for an integrated representation of lyrics and tunes was found in the left dorsal precentral gyrus (PrCG), possibly relating to the build-up of a vocal code for singing in which musical and linguistic features of song are fused. Overall, these results demonstrate that lyrics and tunes are processed at varying degrees of integration (and separation) through the consecutive processing levels allocated along the posterior–anterior axis of the left STS and the left PrCG.

Introduction

Song is one of the richest formats of human communication, as it tightly binds verbal and musical information. A contemporary debate in music cognition research concerns the relationship between lyrics and tunes in the processing of song. Several lines of evidence suggest a separate processing of both components, as demonstrated by the better performance of nonfluent aphasics in producing the melody than the lyrics of songs (Hébert et al., 2003; Racette et al., 2006), the dissociation of lyrics and tunes in song memory after temporal lobe damage [Samson and Zatorre

(1991), their Experiment 2; Peretz (1996); Hébert and Peretz (2001)], or the differential brain signatures in healthy participants during listening to melodic or semantic errors in familiar songs (Besson et al., 1998). In contrast, other studies suggest an integrated processing of lyrics and tunes, as shown by the interaction between the perception of single pitches and vowels (Lidji et al., 2009), harmonic and phonemic information (Bigand et al., 2001), or lexical and semantic information (Poulin-Charronnat et al., 2005; see also Schön et al., 2005), as well as the failure of listeners to ignore the lyrics when required to recognize the tunes of songs, and vice versa [Serafine et al. (1984, 1986); Crowder et al. (1990); Samson and Zatorre (1991), their Experiment 1].

These divergent accounts are, however, not necessarily mutually exclusive. Rather, they may represent the extremes of a continuum with a more or less accentuated integration/dissociation at different stages of song perception, production, and memory. The present study examined the degree of integration (or separation) for song perception by means of a functional magnetic resonance (fMR)-adaptation paradigm. This method is based on the observation that the repetition of certain stimulus features reduces the activity in neuronal populations involved in representing these features (Grill-Spector, 2006; Krekelberg et al.,

Figure 1. Experimental design. The repetition or variation of lyrics and/or tunes within blocks of six songs was crossed in a 2 × 2 factorial design.

2006). This response reduction, also referred to as repetition suppression or neural priming, might reflect the dynamic tuning of the perceptual apparatus and represent the neurophysiological basis of the implicit build-up of perceptual memory representations (Henson, 2003). This approach has been successfully used to study a variety of higher cognitive functions, such as the processing of numbers (Naccache and Dehaene, 2001), voices (Belin and Zatorre, 2003), or language (Dehaene-Lambertz et al., 2006).

We applied a variant of the adaptation paradigm to induce selective adaptation effects for lyrics and tunes during passive listening to unfamiliar songs. Blocks of six short songs (sung by different singers to rule out repetition effects for voice) were presented. The repetition or variation of lyrics and/or tunes within blocks was crossed in a 2 × 2 factorial design. We predicted that brain regions sensitive to the respective component (lyrics or tunes) would be less strongly activated in blocks in which that property was repeated compared to when it varied. In addition, we hypothesized that a significant interaction between the adaptation effects for the two components would be shown by any brain regions that integrate the processing of lyrics and tunes. The lack of such an interaction would specify brain regions that process lyrics and/or tunes independently, along a continuum between integration and separation.

Materials and Methods

Participants. The study was conducted with 12 healthy French native speakers (6 women, 6 men, mean age: 29 years, mean education: 15.58 years). All participants were right handed (mean laterality quotient: 82.64%) as assessed by the Edinburgh Handedness Inventory (Oldfield, 1971), and reported to have normal hearing. None of the participants was a professional musician or actively playing an instrument at the time of testing (mean years of musical training: 1.92 years). Written informed consent was obtained from each participant before the study, which was approved by the local ethics committee.

Materials. One hundred sixty-eight short unfamiliar songs with different (meaningful) lyrics and tunes were created by a professional composer based on a collection of 19th century French folk songs (Robine, 1994). Each song had an average of 7.65 notes and 5.61 words. Major (A♭, E♭, B♭, F, C, G, D, A, E, F♯) and minor (b♭, f, c, g, d, a, e, b, f♯) mode and duple (2/4 or 4/4) and triple (3/4 or 6/8) time were balanced in the stimulus set.

All the songs were recorded by six trained singers (two sopranos, one alto, two tenors, and one bass; mean years of singing lessons: 5.3 years) in a sound studio, cut to 2500 ms, and normalized to −6 dB SPL using Adobe Audition 3 (Adobe Systems). Infrequent, slightly imprecisely sung pitches were adjusted using Celemony Melodyne Studio 3 (Celemony Software). Subsequently, 48 stimulus blocks were constructed consisting of six songs separated by 200 ms pauses resulting in a block duration of 16 s. To rule out potential adaptation to the singers' voices (Belin and Zatorre, 2003) or to simple pitch repetition, each song within a block was sung by another singer of varying age (range: 21–52 years) and sex (three men, three women), at an octave that best corresponded to the singer's voice (soprano, alto and tenor, bass). Consequently, pitch (i.e., octave) and voice-related parameters considerably varied within all 48 blocks, providing no basis for neuronal adaptation to singer's voice (in none of the four conditions described below) (Grill-Spector, 2006; Krekelberg et al., 2006). However, as a footnote it should be said that this manipulation does not completely exclude that the changing voices may differentially interact with the adaptation for lyrics or tunes. Across blocks, each singer's voice occurred with equal probability at any of the six song positions.

There were four types of blocks corresponding to the four experimental conditions: (1) 12 blocks containing songs with the same tunes and same lyrics ($S_T S_L$), (2) 12 blocks with the same tunes but different lyrics ($S_T D_L$), (3) 12 blocks with different tunes but same lyrics ($D_T S_L$), and (4) 12 blocks with different tunes and different lyrics ($D_T D_L$) (Fig. 1; stimulus examples are available at www.jneurosci.org as supplemental material). There were no significant differences in word/note number, word/note length, word frequency according to LEXIQUE 2 (New et al., 2004), duple and triple time, major and minor modes, interval size, and number of contour reversals between conditions as revealed by a multivariate one-way ANOVA with the fixed factor condition ($S_T S_L$ vs $S_T D_L$ vs $D_T S_L$ vs $D_T D_L$) calculated for all these variables (p values >0.220) (see supplemental Table 1, available at www.jneurosci.org as supplemental material). To avoid adaptation to phonology, semantic content, or syntactic structure (Noppeney and Price, 2004), lyrics within $S_T D_L$ and $D_T D_L$ blocks did not rhyme, were semantically distant, and differed with respect to syntactic structure.

Procedure. Each participant was presented with one of four pseudo-randomizations of the 48 blocks. These were intermixed in a way that no more than two blocks of the same condition followed each other, and that transition probabilities between conditions were balanced. Interblock intervals were 10 ± 0.5 s to allow the hemodynamic response to return to baseline (Belin and Zatorre, 2003). Stimuli were presented using E-Prime 1.1 (Psychology Software Tools), and delivered binaurally through air pressure headphones (MR confon). The participants' task was to listen attentively with closed eyes, and to not hum or sing along with the melodies. After scanning, all participants rated on nine-point scales (1 = not at all, 9 = always) how attentively they had listened to the songs (mean: 7.75), and whether they had sung along overtly (mean = 0) or covertly (mean = 3.92) during the scan, confirming that they had followed the instructions. The duration of the experiment was ~30 min.

Scanning. Functional magnetic resonance imaging (fMRI) was performed on a 3T Siemens TRIO scanner (Siemens) at the Centre de Neuroimagerie de Recherche at the Salpêtrière Hospital in Paris. Before the functional scans, high-resolution T1-weighted images (1 × 1 × 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). Subsequently, one series of 595 blood oxygenation level-dependent (BOLD) images was acquired using a single-shot echo-planar gradient-echo (EPI) pulse sequence (TR = 2120

ms, TE = 25 ms, the first six volumes were later discarded to allow for T1 saturation). Forty-four interleaved slices (3 × 3 × 3 mm voxel size, 10% interslice gap) perpendicular with respect to the hippocampal plane were collected with a head coil. The field of view was 192 × 192 mm with an in-plane resolution of 64 × 64 pixels and a flip angle of 90°. Scanner noise was continuous during the experiment, representing a constant auditory background.

Data analysis. FMRI data were analyzed using SPM5 (Wellcome Department of Imaging Neuroscience). Preprocessing of the functional data included spatial realignment, coregistration of the functional and anatomical data, spatial normalization into the MNI stereotactic space, and spatial smoothing using a 3D Gaussian kernel with 8 mm full-width at half-maximum (FWHM). Low-frequency drifts were eliminated using a temporal high-pass filter with a cutoff of 200 s. Statistical evaluation was performed using the general linear model (GLM). Four regressors were modeled (one for each of the four conditions) using boxcar functions convolved with a hemodynamic response function (HRF). In addition, estimated motion parameters were included as covariates of no interest to increase statistical sensitivity.

The combined brain activations of all four listening conditions were contrasted against baseline (all > baseline). Linear contrasts pertaining to the main effect of lyric repetition, i.e., $[(S_TD_L + D_TD_L) - (D_TS_L + S_TS_L)]$, the main effect of tune repetition, i.e., $[(D_TS_L + D_TD_L) - (S_TD_L + S_TS_L)]$, and the interactions of the factorial design were calculated, i.e., $[(S_TD_L + D_TS_L) - (D_TD_L + S_TS_L)]$ and $[(D_TD_L + S_TS_L) - (S_TD_L + D_TS_L)]$. To identify brain regions that showed stronger adaptation for lyrics than for tunes and vice versa, both main effects were contrasted, i.e., $[2 \times (D_TS_L - S_TD_L)]$ and $[2 \times (S_TD_L - D_TS_L)]$. To illustrate the influence of the repetition/variation of one component (lyrics or tunes) on the adaptation for the other, we also computed the four contrasts pertaining to the simple effects, i.e., $[D_TD_L - D_TS_L]$, $[S_TD_L - S_TS_L]$, $[D_TD_L - S_TD_L]$, and $[D_TS_L - S_TS_L]$.

For random effect group analyses, the individual contrast images were submitted to one-sample t tests. All SPMs were thresholded at $p < 0.001$, cluster extent $k \geq 5$ voxels. In a first step, only activations that survived the FDR correction ($p < 0.05$) were considered significant; in a second step, data were examined at a less conservative, uncorrected threshold ($p < 0.001$, $k \geq 5$). Analyses were conducted within a "song-sensitive mask" to increase signal detection (Friston et al., 1994). This mask was created at the group level using the all > baseline contrast and included only voxels for which passive listening to songs (collapsed across all four conditions) elicited significantly greater activation than baseline (thresholded at $p < 0.001$, $k \geq 5$, FDR corrected at $p < 0.05$, whole brain). The resulting mask (volume: 6895 voxels) spanned an auditory–motor network (see supplemental Fig. 1, Table 2, available at www.jneurosci.org as supplemental material). Peak activations were localized by an experienced neuroanatomist via visual inspection of the averaged high-resolution anatomical scan of all participants. Auditory activations included Heschl's gyrus (HG) and the superior temporal gyrus and sulcus (STG/STS) bilaterally, extending into the pars triangularis of the left inferior frontal gyrus [IFG, Brodmann area (BA) 45] and the left inferior temporal gyrus (ITG, BA 20). Motor activa-
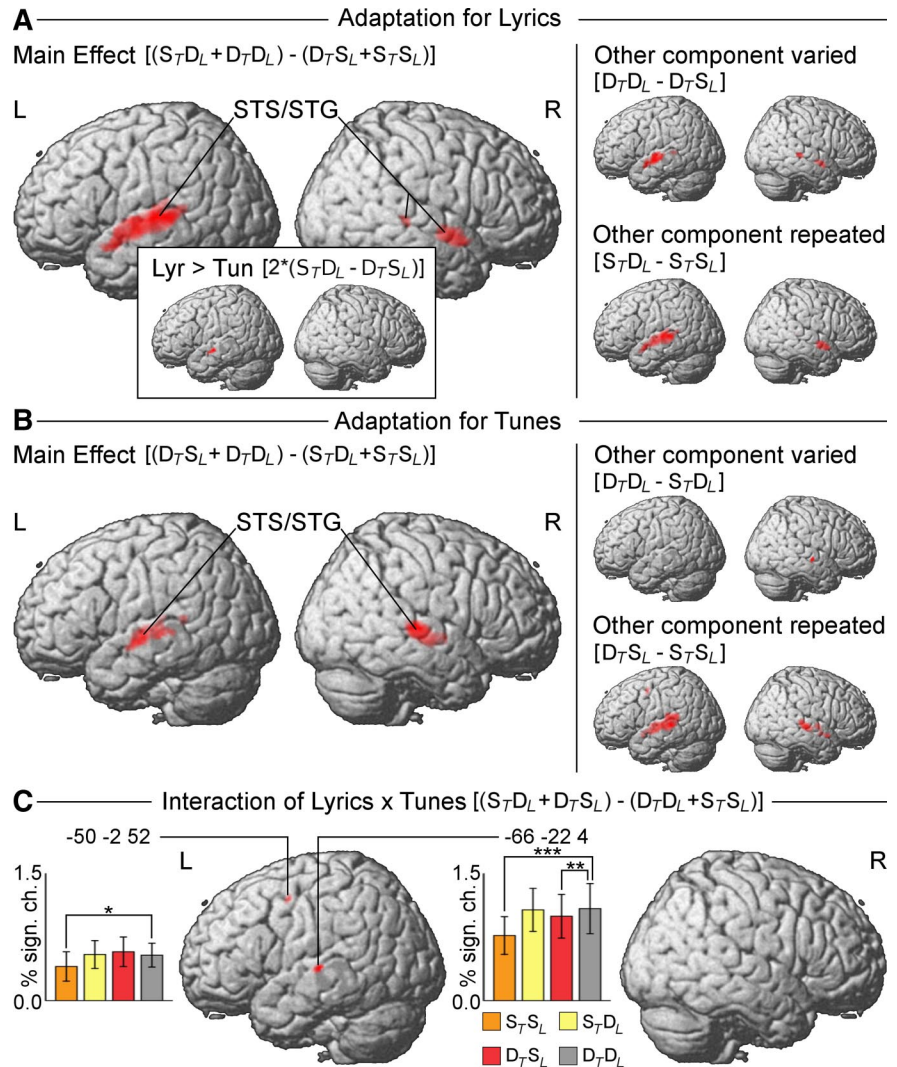

Figure 2. *A*, Adaptation effects for lyric repetition. Left, Main effect of the factorial design $[(S_TD_L + D_TD_L) - (D_TS_L + S_TS_L)]$. Right, Simple effects of lyric repetition when tunes varied $(D_TD_L - D_TS_L)$ (top), or when tunes were simultaneously repeated $(S_TD_L - S_TS_L)$ (bottom; see Results for details). The inset depicts stronger adaptation for the repetition of lyrics than of tunes $[2 \times (S_TD_L - D_TS_L)]$. *B*, Adaptation effects for tune repetition. Left, Main effect of the factorial design $[(D_TS_L + D_TD_L) - (S_TD_L + S_TS_L)]$. Right, Simple effects of tune repetition when lyrics varied $[D_TD_L - S_TD_L]$ (top), or when lyrics were simultaneously repeated $[D_TS_L - S_TS_L]$ (bottom). No brain region showed stronger adaptation for tunes than for lyrics $[2 \times (D_TS_L - S_TD_L)]$ (data not shown). *C*, Interaction of lyrics × tunes $[(S_TD_L + D_TS_L) - (D_TD_L + S_TS_L)]$. Bar diagrams depict the percentage signal change of the peak voxels in the four conditions relative to baseline. Error bars indicate one SEM. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$. For illustration, data are presented at a threshold of $p < 0.001$ (uncorrected, cluster size ≥ 5 voxels).

tions comprised the dorsal precentral gyrus (PrCG) and the cerebellum bilaterally, as well as parts of the right basal ganglia.

Results
Main effects
A main effect of lyric repetition was observed along the STG and STS with larger activations in the left (1147 voxels) than the right hemisphere (258 voxels) (Fig. 2*A*, left, and Table 1). These regions adapted to the repetition of lyrics even if the tunes varied (507 voxels) (Fig. 2*A*, top right), although the effect was spatially more than twice as extended if the tunes were simultaneously repeated (1063 voxels) (Fig. 2*A*, bottom right, and Table 1; for the BOLD percentage signal change of the peak voxels, see supplemental Fig. 2, available at www.jneurosci.org as supplemental material).

A main effect of tune repetition was found in similar areas along the STG/STS bilaterally (left: 362 voxels, right: 448 voxels)

Table 1. Main effects (top rows) and simple effects (middle and bottom rows) of lyric and tune repetition

Gyrus or region	Adaptation for lyrics					Adaptation for tunes				
	Size	x	y	z	Z	Size	x	y	z	Z
Main effects	$(S_TD_L + D_TD_L) - (D_TS_L + S_TS_L)$					$(D_TS_L + D_TD_L) - (S_TD_L + S_TS_L)$				
Left hemisphere										
STS/STG	1147	−66	−26	4	5.14	342	−54	−16	0	5.71
		−50	−30	4	4.60		−54	−24	2	4.56
		−56	8	−16	4.29		−66	−12	−6	4.29
STS/STG						20	−56	−42	10	3.83
Right hemisphere										
STS/STG	232	56	2	−8	4.49	448	52	−26	2	5.02
		62	−4	−8	4.07		60	−16	−4	4.42
STS/STG	26	68	−32	0	3.52		68	−20	4	4.38
Simple effects: variation of the other modality	$D_TD_L - D_TS_L$					$D_TD_L - S_TD_L$				
Left hemisphere										
STS/STG	352	−64	−18	−2	4.49					
		−56	−10	−6	4.47					
		−64	−10	0	4.32					
STS/STG	38	−46	−38	4	4.00					
STS/STG	7	−60	−34	6	3.41					
Right hemisphere										
STS/STG	36	70	−32	0	4.57	**11**	**58**	**−14**	**−4**	**3.46**
STS/STG	74	58	0	−12	3.66					
Simple effects: repetition of the other modality	$S_TD_L - S_TS_L$					$D_TS_L - S_TS_L$				
Left hemisphere										
STS/STG	857	−64	−30	4	5.03	975	−58	−32	0	5.38
		−48	−32	4	4.83		−54	−24	2	4.74
		−60	−22	2	4.50		−66	−36	14	4.44
PrCG (BA 6)						39	−42	0	46	3.74
							−50	−2	52	3.59
Right hemisphere										
STS/STG	201	58	−2	−8	4.05	482	56	6	−12	4.28
STS/STG	5	50	−34	6	3.25		60	−24	0	4.20
							64	−6	2	4.06

Plain type values, Thresholded at $p < 0.001$, cluster size ≥5 voxels, significant after FDR correction at $p < 0.05$; bold values, $p < 0.001$, cluster size ≥5 voxels, uncorrected. Brain atlas coordinates (MNI stereotactic space) are indicated in millimeters along left–right (x), anterior–posterior (y), and superior–inferior (z) axes.

(Fig. 2B, left, and Table 1). In FDR-corrected SPMs, no region adapted to the repetition of tunes if the lyrics varied, although a small effect was found in the right mid-STS at a less conservative, uncorrected threshold of $p < 0.001$, $k \geq 5$ (11 voxels) (Fig. 2B, top right). The adaptation effect emerged bilaterally (FDR corrected) if the lyrics were simultaneously repeated (1457 voxels, spatial extent enhanced by a factor of 132) (Fig. 2B, bottom right, and Table 1).

The direct comparison of the two main effects yielded no significant difference after FDR correction, probably due to the relatively low number of subjects. However, as several studies claim a dissociated processing of lyrics and tunes (e.g., Besson et al., 1998; Bonnel et al., 2001), possible differences between the adaptation for lyrics and tunes were tested directly at an uncorrected threshold ($p < 0.001$, $k \geq 5$). This test indicated a stronger adaptation effect for lyrics than for tunes in an anterior portion of the left STS ($x = -54$, $y = -12$, $z = -6$, cluster size = 35 voxels, $Z = 3.88$) (Fig. 2A, inset; for the BOLD percentage signal change of the peak voxel, see also supplemental Fig. 2, available at www.jneurosci.org as supplemental material). No brain regions showed stronger adaptation for tunes than for lyrics at this threshold.

Interaction

No voxel survived the FDR correction, but as previous studies reported an interaction between the processing of verbal and musical information in song (e.g., Bigand et al., 2001; Lidji et al., 2009; Schön et al., 2005) and guided by our hypothesis (see Introduction), a possible interaction of lyrics and tunes was tested at a less conservative threshold ($p < 0.001$, $k \geq 5$ voxels, uncorrected). This analysis revealed an interaction of lyrics × tunes $[(S_TD_L + D_TS_L) - (D_TD_L + S_TS_L)]$ in the left mid-STS ($x = -66$, $y = -22$, $z = 4$, cluster extent = 16 voxels, $Z = 3.47$) and the left dorsal PrCG ($x = -50$, $y = -2$, $z = 52$, cluster extent = 5 voxels, $Z = 3.47$) (Fig. 2C). This indicates that in these regions, the combined repetition of lyrics and tunes (S_TS_L) induced significantly stronger adaptation (compared to D_TD_L) than the simple repetition of lyrics (D_TS_L) and tunes (S_TD_L) summed up, suggestive of an integrated processing of both components. No interactions were found in the right hemisphere and the reverse contrast $[(D_TD_L + S_TS_L) - (S_TD_L + D_TS_L)]$. Both clusters were distant to typical voice areas (Belin et al., 2000; Belin and Zatorre, 2003), indicating that the present effect was (as expected) not grounded on an interaction between lyrical/melodic and voice information. To further explore these effects, percentage signal change values were extracted from the peak voxels of each cluster in each participant using the MarsBaR SPM toolbox (http://marsbar.sourceforge.net). These values were subjected to *post hoc* paired-samples t tests, evaluating the adaptation effects when only lyrics (D_TD_L vs D_TS_L), only tunes (D_TD_L vs S_TD_L), or both components (D_TD_L vs S_TS_L) were repeated (Fig. 2C, bar diagrams). In line with the interaction, the combined repetition of lyrics and tunes induced the strongest adaptation effects in both regions (left mid-STS: $t_{(11)} = 6.53$, $p < 0.001$; left PrCG: $t_{(11)} = 2.92$, $p < 0.015$). The adaptation effect for the simple repetition of lyrics was significant, but considerably weaker in the left mid-STS ($t_{(11)} = 3.43$, $p < 0.007$), and nonsignificant in the left PrCG ($p > 0.314$).
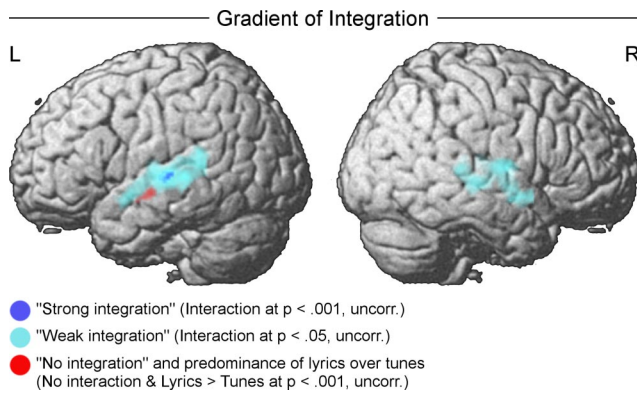
Gradient of Integration

● "Strong integration" (Interaction at p < .001, uncorr.)

● "Weak integration" (Interaction at p < .05, uncorr.)

● "No integration" and predominance of lyrics over tunes
(No interaction & Lyrics > Tunes at p < .001, uncorr.)

Figure 3. Posterior–anterior gradient of integration.

No cluster showed significant changes during the simple repetition of tunes (p values >0.717).

Gradient

To capture a possible gradient between integration and separation of lyrics and tunes, the interaction (taken as an index for integrated processing) was examined at different statistical thresholds ($p < 0.001$ and $p < 0.05$, uncorrected, $k \geq 5$ voxels). These clusters were compared with the regions that exhibited no interaction but a significantly stronger adaptation for lyrics than for tunes ($p < 0.001$, uncorrected, $k \geq 5$; no regions showed stronger adaptation for tunes than for lyrics; see above), indicating an independent (perhaps separate) processing of lyrics. Figure 3 illustrates that the interaction was confined to the left mid-STS at $p < 0.001$ (blue cluster), suggesting a relatively "strong integration" of both components. The interaction extended more anteriorly and posteriorly, and emerged also in the right STS/STG at a lowered threshold of $p < 0.05$ (cyan cluster), taken as a "weaker" form of integration. Anteroventral to this, the left STS showed no more interaction ($p > 0.05$), but a significantly stronger adaptation effect for the repetition of lyrics compared to tunes (red cluster) (see also Fig. 2*A*, inset), suggesting "no integration" and a predominance of lyrics over tunes in this region. Altogether, these findings appear to constitute a gradient from more to less integrated processing along the posterior–anterior axis of the left STS (Fig. 3, left).

Discussion

The present study demonstrates that lyrics and tunes of unfamiliar songs are processed at different degrees of integration along the axis of the superior temporal lobe and the left precentral gyrus (PrCG). This is consistent with the idea of a different weighting of integration (and separation) at different stages of the processing of unfamiliar songs.

Main adaptation effects were found along the superior temporal lobe bilaterally. These results are consistent with studies reporting activations of the STG/STS during listening to songs (Schön et al., 2005; Callan et al., 2006), and the processing of various aspects of language (Scott and Johnsrude, 2003; Vigneau et al., 2006) and music (Stewart et al., 2006). This suggests that the observed adaptation effects reflect the facilitated processing of the repeated lyrical and melodic information. Most importantly, the voice sensitivity of the STS (Belin et al., 2000; Belin and Zatorre, 2003) and pitch processing cannot account for the observed adaptation effects because singers' voices and octave varied in all four conditions (see Materials).

The novel finding is that within these superior temporal regions, specifically in the left hemisphere, lyrics and tunes are processed at varying degrees of integration, with some indication of an independent processing of lyrics in the left anterior STS. The left mid-STS, inferior to Heschl's gyrus, showed an interaction of the adaptation effects for lyrics and tunes, indicating that the combined repetition of both components (S_TS_L) induced a significantly stronger response reduction (compared to D_TD_L) than the simple repetition of lyrics (D_TS_L) and tunes (S_TD_L) summed up. This overadditive effect demonstrates an integrated processing of both components within the left mid-STS. The interaction (and thus integration) decayed in regions anterior to this cluster. A more anteroventral portion of the left STS exhibited no more interaction, but a stronger adaptation for lyrics than tunes, suggesting a predominant processing of lyrics in this region (see below for a discussion why no region showed a predominance for tunes). Taking these findings together, the picture of a posterior–anterior gradient emerges along the axis of the left STS, from an integrated processing of lyrics and tunes in the mid-STS to the rather independent processing of lyrics in more anterior temporal regions.

This posterior–anterior gradient is reminiscent of the functional (Binder, 2000; Davis and Johnsrude, 2003; Liebenthal et al., 2005; Scott and Johnsrude, 2003) and temporal (Patterson et al., 2002; Kiebel et al., 2008; Overath et al., 2008) hierarchy of auditory (speech) perception in the superior temporal lobe. These models posit a rostral stream running from primary auditory areas to more lateral and anteroventral areas in the (left) STG and STS, comprising consecutive levels of processing that deal with increasingly abstract representations of the auditory information within growing temporal windows: spectrotemporal features in the millisecond range within the primary auditory cortices, prelexical phonemic information within the surrounding left mid-STG/STS (for an overview, see Obleser and Eisner, 2009), and sentential structure and meaning spanning several hundred milliseconds in more anterior temporal regions (Vandenberghe et al., 2002; Scott and Johnsrude, 2003; Crinion et al., 2006; Spitsyna et al., 2006).

Against this theoretical background, the localization of the lyrics–tunes interaction in the left mid-STS suggests an integration of musical and linguistic aspects of song during an intermediate, phonemic processing stage in this rostral auditory pathway (although the current study neither contains time course information, nor specifically manipulates acoustic, phonemic, or structural–semantic processing). No integration of lyrics and tunes was observed for early nonspecific sound analysis within primary auditory areas, although the present data do not exclude integration at this level. First, these pitch-sensitive regions were most likely blind to the repetition of the songs sung by different voices at different octaves, and second, their temporal integration window was probably too narrow to perceive the repetition of the 2.5 s songs (Kiebel et al., 2008). The localization of the interaction effect in the left mid-STS suggests that lyrics and tunes are particularly integrated at prelexical, phonemic processing levels (Obleser and Eisner, 2009). This observation is consistent with previous behavioral and EEG studies showing an interaction between the processing of melodic/harmonic information and nonsense syllables or vowels (Serafine et al., 1986; Crowder et al., 1990; Bigand et al., 2001; Lidji et al., 2009). Beyond that, the data suggest a separate processing of lyrics at subsequent levels of structural analysis and lexical–semantic representation or access in the left anterior STS. Note that this view would not contradict the ability of music to convey meaning (Koelsch et al., 2004) but propose a predominance and greater autonomy of linguistic (compared to musical) meaning in songs. In sum, it may be sug-

gested as a working hypothesis that the degree of integration of lyrics and tunes decreases as the processing of (unfamiliar) songs proceeds along the rostral auditory stream. Note that although the current study does not address song memory and production (precluding a direct comparison between our data and the majority of the prevailing studies), we speculate that also beyond auditory perceptual processing, the degree of integration/separation depends on the specific cognitive processes targeted by an experimental task (e.g., recognition vs recall or production of familiar vs unfamiliar songs), perhaps accounting for some of the conflicting results.

The profile of adaptation effects argues in favor of bidirectional connections between lyrics and tunes, as the adaptation for one component (lyrics or tunes) was modulated by the simultaneous repetition/variation of the other (Fig. 2*A*,*B*, right). However, it appears that the strength of these connections differs depending on their direction, in a way that tunes are tightly bound to lyrics, whereas the processing of lyrics exhibits a considerable autonomy [for converging behavioral data, see Serafine et al. (1984), Samson and Zatorre (1991), and Schön et al. (2005)]. Consistent with this notion, the left anterior STS showed stronger adaptation for lyrics than tunes (Fig. 2*A*, inset), whereas no reverse effects (tunes > lyrics) were found. It remains to be specified to what extent this imbalance of lyrics and tunes depends on the settings of the present experiment. Listeners may have paid particular attention to the lyrics (as they convey the message), probably boosting the adaptation effect (Chee and Tan, 2007). Correspondingly, deeper lexical–semantic processing (see above) may account for the more robust adaptation effects for lyrics. Alternatively, the predominance of lyrics might be due to the higher linguistic than musical expertise of the listeners (French native speakers, but musically untrained), consistent with the sensitivity of left STS activations to the expertise of listeners with the employed stimulus material (Leech et al., 2009). Future studies with trained musicians (i.e., balanced linguistic and musical expertise), focused listening to the melodies, and/or the use of nonsense lyrics could address these issues.

As a final footnote, the simple repetition of lyrics induced a bilateral response reduction with left hemisphere preponderance, whereas a small cluster in the right hemisphere tended to adapt to the simple repetition of tunes. This differential hemispheric weighting is consistent with prevailing models of a relative specialization of the left and right hemisphere for linguistic and musical stimulus features respectively, like temporal and spectral (Zatorre et al., 2002) or segmental and suprasegmental information (Friederici and Alter, 2004). Interestingly, lyrics and tunes appeared to be more strongly integrated in the left than in the right hemisphere. This might be due to the predominance of lyrics over tunes in the present study and, thus, a stronger involvement of the left hemisphere.

The interaction of adaptation effects in the left precentral gyrus (BA 6) also suggested an integrated processing of lyrics and tunes. The PrCG is the seat of primary motor and premotor areas, and its involvement in the present experiment may be associated either with (voluntary) internal singing or humming (Hickok et al., 2003; Callan et al., 2006), or with a more general (involuntary) coupling between the auditory and the motor system as proposed by models of auditory–motor integration in language (Scott and Johnsrude, 2003; Hickok and Poeppel, 2007) and music (Warren et al., 2005; Zatorre et al., 2007). These models posit a direct matching between the perception of an auditory signal like a speech sound or a piano tone and a stored (pre)motor code for its production. Along these lines, it may be speculated that the ad-

aptation of the neural activity in the left PrCG reflects either the increasing efficiency of subvocal rehearsal, i.e., vocal learning (Rauschecker et al., 2008), or the facilitated mirroring of articulatory gestures during passive listening. It appears that lyrical and melodic features must be integrated in a vocal code for singing as they are simultaneously articulated via the vocal tract.

To conclude, the present study is the first demonstration that lyrics and tunes of songs are processed at different degrees of integration (and separation) through the consecutive processing levels allocated along the posterior–anterior axis of the left superior temporal lobe and the left PrCG. While both components seem to be integrated at a prelexical, phonemic stage of the auditory analysis in the left mid-STS, and the preparation of a motor output in the left PrCG, lyrics may be processed independently at levels of structural and semantic integration in the left anterior STS. Overall, the findings demonstrate an anatomical and functional gradient of integration of lyrics and tunes during passive listening to unfamiliar songs.

References

Belin P, Zatorre RJ (2003) Adaptation to speaker's voice in right anterior temporal lobe. Neuroreport 14:2105–2109.

Belin P, Zatorre RJ, Lafaille P, Ahad P, Pike B (2000) Voice-selective areas in human auditory cortex. Nature 403:309–312.

Besson M, Faïta F, Peretz I, Bonnel AM, Requin J (1998) Singing in the brain: independence of lyrics and tunes. Psychol Sci 9:494–498.

Bigand E, Tillmann B, Poulin B, D'Adamo DA, Madurell F (2001) The effect of harmonic context on phoneme monitoring in vocal music. Cognition 81:B11–B20.

Binder J (2000) The new neuroanatomy of speech perception. Brain 123:2371–2372.

Bonnel AM, Faita F, Peretz I, Besson M (2001) Divided attention between lyrics and tunes of operatic songs: evidence for independent processing. Percept Psychophys 63:1201–1213.

Callan DE, Tsytsarev V, Hanakawa T, Callan AM, Katsuhara M, Fukuyama H, Turner R (2006) Song and speech: brain regions involved with perception and covert production. Neuroimage 31:1327–1342.

Chee MWL, Tan JC (2007) Inter-relationships between attention, activation, fMR adaptation and long-term memory. Neuroimage 37:1487–1495.

Crinion JT, Warburton EA, Lambon-Ralph MA, Howard D, Wise RJS (2006) Listening to narrative speech after aphasic stroke: the role of the left anterior temporal lobe. Cereb Cortex 16:1116–1125.

Crowder RG, Serafine ML, Repp B (1990) Physical interaction and association by contiguity in memory for the words and melodies of songs. Mem Cognit 18:469–476.

Davis MH, Johnsrude IS (2003) Hierarchical processing in spoken language comprehension. J Neurosci 23:3423–3431.

Dehaene-Lambertz G, Dehaene S, Anton JL, Campagne A, Ciuciu P, Dehaene GP, Denghien I, Jobert A, LeBihan D, Sigman M, Pallier C, Poline JB (2006) Functional segregation of cortical language areas by sentence repetition. Hum Brain Mapp 27:360–371.

Friederici AD, Alter K (2004) Lateralization of auditory language functions: A dynamic dual pathway model. Brain Lang 89:267–276.

Friston KJ, Worsley RSJ, Frackowiak JC, Mazziotta JC, Evans AC (1994) Assessing the significance of focal activations using their spatial extent. Hum Brain Mapp 1:214–220.

Grill-Spector K (2006) Selectivity of adaptation in single units: implications for FMRI experiments. Neuron 49:170–171.

Hébert S, Peretz I (2001) Are text and tune of familiar songs separable by brain damage? Brain Cogn 46:169–175.

Hébert S, Racette A, Gagnon L, Peretz I (2003) Revisiting the dissociation between singing and speaking in expressive aphasia. Brain 126:1838–1850.

Henson RNA (2003) Neuroimaging studies of priming. Prog Neurobiol 70:53–81.

Hickok G, Poeppel D (2007) The cortical organization of speech processing. Nat Rev Neurosci 8:393–402.

Hickok G, Buchsbaum B, Humphries C, Muftuler T (2003) Auditory-

motor interaction revealed by fMRI: speech, music, and working memory in area Spt. J Cogn Neurosci 15:673– 682.

Kiebel SJ, Daunizeau J, Friston KJ (2008) A hierarchy of time-scales and the brain. PLoS Comput Biol 4:e1000209.

Koelsch S, Kasper E, Sammler D, Schulze K, Gunter T, Friederici AD (2004) Music, language and meaning: brain signatures of semantic processing. Nat Neurosci 7:302–307.

Krekelberg B, Boynton GM, van Wezel RJA (2006) Adaptation: from single cells to BOLD signals. Trends Neurosci 29:250–256.

Leech R, Holt LL, Devlin JT, Dick F (2009) Expertise with artificial non-speech sounds recruits speech-sensitive cortical regions. J Neurosci 29:5234–5239.

Lidji P, Jolicoeur P, Moreau P, Kolinsky R, Peretz I (2009) Integrated pre-attentive processing of vowel and pitch: a mismatch negativity study. Ann N Y Acad Sci 1169:481– 484.

Liebenthal E, Binder JR, Spitzer SM, Possing ET, Medler DA (2005) Neural substrates of phonemic perception. Cereb Cortex 15:1621–1631.

Naccache L, Dehaene S (2001) The priming method: imaging unconscious repetition priming reveals an abstract representation of number in the parietal lobes. Cereb Cortex 11:966–974.

New B, Pallier C, Brysbaert M, Ferrand L (2004) Lexique 2: A new French lexical database. Beh Res Meth Instr Comp 36:516–524.

Noppeney U, Price CJ (2004) An fMRI study of syntactic adaptation. J Cogn Neurosci 16:702–713.

Obleser J, Eisner F (2009) Pre-lexical abstraction of speech in the auditory cortex. Trends Cogn Sci 13:14–19.

Oldfield RC (1971) The assessment and analysis of handedness: the Edinburgh inventory. Neuropsychologia 9:97–113.

Overath T, Kumar S, von Kriegstein K, Griffiths TD (2008) Encoding of spectral correlation over time in auditory cortex. J Neurosci 28:13268–13273.

Patterson RD, Uppenkamp S, Johnsrude IS, Griffiths TD (2002) The processing of temporal pitch and melody information in auditory cortex. Neuron 36:767–776.

Peretz I (1996) Can we lose memories for music? A case of music agnosia in a nonmusician. J Cogn Neurosci 8:481–496.

Poulin-Charronnat B, Bigand E, Madurell F, Peereman R (2005) Musical structure modulates semantic priming in vocal music. Cognition 94:B67–B78.

Racette A, Bard C, Peretz I (2006) Making non-fluent aphasics speak: sing along! Brain 129:2571–2584.

Rauschecker AM, Pringle A, Watkins KE (2008) Changes in neural activity associated with learning to articulate novel auditory pseudowords by covert repetition. Hum Brain Mapp 29:1231–1242.

Robine M (1994) Anthologie de la chanson française— des trouvères aux grands auteurs du XIXe siècle. Paris: Albin Michel.

Samson S, Zatorre RJ (1991) Recognition memory for text and melody of songs after unilateral temporal lobe lesion: evidence for dual encoding. J Exp Psychol Learn Mem Cogn 17:793– 804.

Schön D, Gordon RL, Besson M (2005) Musical and linguistic processing in song perception. Ann N Y Acad Sci 1060:71– 81.

Scott SK, Johnsrude IS (2003) The neuroanatomical and functional organization of speech perception. Trends Neurosci 26:100–107.

Serafine ML, Crowder RG, Repp BH (1984) Integration of melody and text in memory for songs. Cognition 16:285–303.

Serafine ML, Davidson J, Crowder RG, Repp B (1986) On the nature of melody-text integration in memory for songs. J Mem Lang 25:123–135.

Spitsyna G, Warren JE, Scott SK, Turkheimer FE, Wise RJS (2006) Converging language streams in the human temporal lobe. J Neurosci 26:7328– 7336.

Stewart L, von Kriegstein K, Warren JD, Griffiths TD (2006) Music and the brain: disorders of musical listening. Brain 129:2533–2553.

Vandenberghe R, Nobre AC, Price CJ (2002) The response of left temporal cortex to sentences. J Cogn Neurosci 14:550–560.

Vigneau M, Beaucousin V, Hervé PY, Duffau H, Crivello F, Houdé O, Mazoyer B, Tzourio-Mazoyer N (2006) Metaanalyzing left hemisphere language areas: phonology, semantics, and sentence processing. Neuroimage 30:1414–1432.

Warren JE, Wise RJS, Warren JD (2005) Sounds do-able: auditory-motor transformations and the posterior temporal plane. Trends Neurosci 28:636–643.

Zatorre RJ, Belin P, Penhune VB (2002) Structure and function of auditory cortex: music and speech. Trends Cogn Sci 6:37–46.

Zatorre RJ, Chen JL, Penhune VB (2007) When the brain plays music: auditory-motor interactions in music perception and production. Nat Rev Neurosci 8:547–558.