



**HAL**  
open science

## Words and Melody are intertwined in Perception of Sung Words: EEG and Behavioral Evidence

Reyna Gordon, Daniele Schön, Cyrille Magne, Corine Astésano, Mireille R Besson

► **To cite this version:**

Reyna Gordon, Daniele Schön, Cyrille Magne, Corine Astésano, Mireille R Besson. Words and Melody are intertwined in Perception of Sung Words: EEG and Behavioral Evidence. PLoS ONE, 2010, 5 (3), pp.e9889. 10.1371/journal.pone.0009889 . hal-02149830

**HAL Id: hal-02149830**

**<https://univ-tlse2.hal.science/hal-02149830v1>**

Submitted on 6 Jun 2019

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Words and Melody Are Intertwined in Perception of Sung Words: EEG and Behavioral Evidence

Reyna L. Gordon<sup>1,2\*</sup>, Daniele Schön<sup>2</sup>, Cyrille Magne<sup>3</sup>, Corine Astésano<sup>4,2</sup>, Mireille Besson<sup>2</sup>

**1** Center for Complex Systems and Brain Sciences, Florida Atlantic University, Boca Raton, Florida, United States of America, **2** Institut de Neurosciences Cognitives de la Méditerranée, CNRS and Université de la Méditerranée, Marseille, France, **3** Department of Psychology, Middle Tennessee State University, Murfreesboro, Tennessee, United States of America, **4** U.R.I. Octogone-Lordat 4156, Université de Toulouse II, Toulouse, France

## Abstract

Language and music, two of the most unique human cognitive abilities, are combined in song, rendering it an ecological model for comparing speech and music cognition. The present study was designed to determine whether words and melodies in song are processed interactively or independently, and to examine the influence of attention on the processing of words and melodies in song. Event-Related brain Potentials (ERPs) and behavioral data were recorded while non-musicians listened to pairs of sung words (prime and target) presented in four experimental conditions: same word, same melody; same word, different melody; different word, same melody; different word, different melody. Participants were asked to attend to either the words or the melody, and to perform a same/different task. In both attentional tasks, different word targets elicited an N400 component, as predicted based on previous results. Most interestingly, different melodies (sung with the same word) elicited an N400 component followed by a late positive component. Finally, ERP and behavioral data converged in showing interactions between the linguistic and melodic dimensions of sung words. The finding that the N400 effect, a well-established marker of semantic processing, was modulated by musical melody in song suggests that variations in musical features affect word processing in sung language. Implications of the interactions between words and melody are discussed in light of evidence for shared neural processing resources between the phonological/semantic aspects of language and the melodic/harmonic aspects of music.

**Citation:** Gordon RL, Schön D, Magne C, Astésano C, Besson M (2010) Words and Melody Are Intertwined in Perception of Sung Words: EEG and Behavioral Evidence. PLoS ONE 5(3): e9889. doi:10.1371/journal.pone.0009889

**Editor:** Antoni Rodriguez-Fornells, University of Barcelona, Spain

**Received:** August 19, 2009; **Accepted:** February 26, 2010; **Published:** March 31, 2010

**Copyright:** © 2010 Gordon et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** This research was supported by a grant from the Human Frontier Science Program “An interdisciplinary approach to the problem of language and music specificity” (HFSP#RGP0053) to M. Besson and was conducted at the Institut de Neurosciences Cognitives de la Méditerranée, while R.L. Gordon was a graduate student. D. Schön and C. Astésano were supported by the HFSP grant; C. Magne benefitted from a “Cognitive Science” Fellowship from the French Ministry of Research; and R.L. Gordon benefitted from a Fellowship from the American Academy of University Women. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

\* E-mail: reyna.gordon@alumni.usc.edu

## Introduction

Strong arguments have been made for both the opposing frameworks of modularity versus shared resources underlying language and music cognition (see reviews [1–5]). On the one hand, double dissociations of linguistic and musical processes, documented in neuropsychological case studies, often point to domain-specific and separate neural substrates for language and music [3,6–9]. On the other hand, results of brain imaging and behavioral studies have often demonstrated shared or similar resources underlying, for instance, syntactic and harmonic processing [10–14], auditory working memory for both linguistic and musical stimuli [15], and semantic or semiotic priming [16–21].

These conflicting results may stem from the use of different methods, but also from other methodological problems. The main disadvantage to comparing language and music processing by testing perception of speech and musical excerpts is that the acoustic properties, context, and secondary associations (e.g., musical style or linguistic pragmatics) between even the most carefully controlled stimuli may vary greatly between the two domains. One ecological alternative is to study the perception of song [22]. In this case, linguistic and musical information are

contained in one auditory signal that is also a universal form of human vocal expression. Furthermore, a better understanding of the neural basis of song is surely germane to the ongoing debate on the evolutionary origins of language and music, especially in view of propositions that the protolanguage used by early humans was characterized by singing [23,24] and that vocal learning was a key feature governing the evolution of musical and linguistic rhythm [25]. While most studies of music cognition have used non-vocal music stimuli, everyday music-making and listening usually involve singing. Moreover, from a developmental perspective, singing is also quite relevant for parent-infant bonding, as indicated by studies showing that babies prefer infant-directed singing to infant-directed speech [26,27].

Early studies of song cognition used dichotic listening paradigms to reveal lateralization patterns of left-ear (right hemisphere) advantage for melody recognition and right ear (left hemisphere) advantage for phoneme recognition in song [28] and in the recall of musical and linguistic content of sung digits [29]. Despite the lateralization tendencies, melody and lyrics appear to be tightly integrated in recognition [30] and priming experiments [31]. Indeed, the melody of a song may facilitate learning and recall of the words [32,33], though this advantage appears to be diminished

when the rate of presentation is controlled for, such that spoken lyrics are presented at the same rate as sung ones [34]. Furthermore, the segmentation of a pseudo-language into relevant units is facilitated for sung compared to spoken pseudowords [35], and infants learn words more easily when sung on melodies rather than when spoken [36].

The extent to which semantics and emotions are conveyed by song lyrics remains a controversial issue. One study showed that when participants were asked to listen to songs from a variety of popular music genres, they performed only at chance level when attempting to interpret the singer's intended message of each song [37]. Thus, while explicit literary interpretations of song lyrics do not appear consistent in this study, other work has suggested that sung lyrics have a greater influence over listeners' mood than the same melody played on an instrument [38]. However, this effect was amplified when the lyrics were sung with piano accompaniment, showing that the musical dimension retains importance. It has also been reported that lyrics intensify emotional responses to sad and angry music, yet mitigate the response to happy and calm music [39].

A key feature of several recent studies is the use of attentional focus to examine the interaction or independence of words and melodies in song, either by directing listeners' attention to language and music simultaneously [40–42], or to language only [41,43–47], or to music only [41,43]. Some of these studies have demonstrated interactive effects between the linguistic and musical dimensions of song, thereby suggesting that common cognitive processes and neural resources are engaged to process language and music. Bigand et al. [44] showed that a subtle variation in harmonic processing interfered with phoneme monitoring in the perception of choral music sung with pseudowords. In a follow-up study, the authors used a lexical decision task on sung sentence material to demonstrate that harmonic processing also interfered with semantic priming [46]. These observed interactions between semantics and harmony, measured through the implicit processing of the musical dimension, suggest that language and music in song are perceptually interwoven. Interestingly, data recently obtained by Kolinsky et al. [43] using a Garner paradigm [48] provides evidence that, while consonants remain separable from melody, vowels and melody are strongly integrated in song perception. This interaction may stem from integration of vowel and musical pitch in initial stages of sensory processing [49]. Sung sentences were also used by Fedorenko et al. [47] to demonstrate that the processing of syntactically complex sentences in language is modulated by structural manipulations in music, thereby indicating that structural aspects of language and music seem to be integrated in song perception.

By contrast, other studies of song perception and memory have shown evidence for independent processing of the linguistic and musical dimensions of song. Besson et al. [40] used the Event-Related brain Potential (ERP) method to study the relationship between words and melodies in the perception of opera excerpts sung without instrumental accompaniment. When musicians were asked to passively listen to the opera excerpts and pay equal attention to lyrics and tunes, results showed distinct ERP components for semantic (N400) and harmonic (P300) violations. Furthermore, the observed effects were well accounted for by an additive model of semantic and harmonic processing (i.e., results in the double violation condition were not significantly different from the sum of the simple semantic and melodic violations). Additional behavioral evidence for the independence of semantics and harmony in song was provided by a second experiment utilizing the same stimuli [41] and a dual task paradigm. When musician and non-musician listeners had to detect semantic and/or

harmonic violations in song, results showed that regardless of musical expertise, there was no decrease in performance when listeners simultaneously attended language and music, compared to attending only one dimension at a time. These results contrast with those recently obtained by van Besouw et al. [42], showing a detriment to performance in recalling pitch contour and recalling words when listeners had to simultaneously pay attention to the words and pitch in song, as well as a similar detriment when they were asked to pay attention to the words and pitch contour of speech. Singing was also used innovatively in a series of experiments by Levy et al. [50,51] that highlighted the influence of task demands and attentional focus on the perception of human voices in a non-linguistic context; the oddball paradigm generated a task-dependent positive ERP component (P320) in response to sung tones compared to instrumental tones.

The present study was developed to further investigate the interaction or independence of the linguistic and musical dimensions by examining the electrophysiological and behavioral correlates of words and melody in the perception of songs by individuals without formal musical training (and who are thus most representative of the general population). The choice to test non-musician participants was motivated by compelling evidence reviewed by Bigand & Poulin-Charronnat [52], in support of the idea that day-to-day normal exposure to music teaches non-musicians to implicitly process the structural aspects of music according to similar principles (although less explicitly) as individuals who have received extensive musical training. Results obtained with behavioral measures on non-musician participants demonstrate that pseudowords and intervals are processed interactively in song perception, regardless of whether listeners attend to the linguistic or to the musical dimensions [43]. Our goal was to determine whether the interactions between lyrics and tunes would also be observed when the linguistic and musical complexity of the sung stimuli was increased by using real words sung on short melodies.

The specific aim of the present experiment was two-fold: to determine the nature of the relationship (independent or interactive) between the linguistic and musical dimensions of sung words, and to specify how attention influences the dynamics of that relationship. To achieve these goals, we presented listeners with prime-target pairs of tri-syllabic words sung on 3-note melodies and recorded behavioral and electrophysiological data while they performed a same/different task. Compared to the prime, the melody and words of the sung target was manipulated orthogonally to create four experimental conditions: Same Word/Same Melody ( $W=M=$ ); Same Word/Different Melody ( $W=M\neq$ ); Different Word/Same Melody ( $W\neq M=$ ); Different Word/Different Melody ( $W\neq M\neq$ ; see Figure 1 for examples).

On the basis of previous findings that the N400 component is elicited by semantically unexpected or unrelated words in pairs of words [53,54], read and spoken sentences [55–57], and sung sentences [40], and from results showing decreased N400 amplitude with repetition [58], we predicted that different targets, semantically unrelated to the prime ( $W\neq$ ), would elicit larger N400 components, slower Reaction Times (RTs) and higher error rates than same, repeated targets ( $W=$ ) [59,60].

Besson et al. [40] also showed that an opera excerpt ending on an incongruous pitch evoked a positive component, P300/P600, typically associated with surprising events such as melodic incongruities [61–64]. Thus, we predicted that different melodies ( $M\neq$ ) would also elicit larger P300/P600 components, and slower RTs and higher error rates [65], compared to same melody ( $M=$ ).

Finally, if the perception of words and melodies in songs call upon independent processes, the Word effect (different – same word) should be similar, in behavioral measures and N400

	Prime	Target
a) $W=M$		
b) $W=M\neq$		
c) $W\neq M$		
d) $W\neq M\neq$		

**Figure 1. Stimuli examples.** Examples of stimuli in the four experimental conditions: same word, same melody (a); same word, different melody (b); different word, same melody (c); different word, different melody (d).

doi:10.1371/journal.pone.0009889.g001

amplitude, for same and different melodies. Likewise, the Melody effect (different – same melody) should be similar, in behavioral measures and P300/P600 amplitude, for same and different words. If the perception of words and melodies in sung words rely instead on interactive processes, the Word effect should be different for same and different melodies (interference effects) and vice-versa for the Melody effect. In addition, the use of an orthogonal design allows us to test the additive model following which the ERP in the double variations condition ( $W\neq M\neq$ ) should be equivalent to the sum of the ERPs in the simple variations conditions ( $W\neq M=$  plus  $W=M\neq$ ).

In order to determine how attention to one dimension or another modulates the processing of words and melody in song, we asked participants to perform a same/different task on the same set of stimuli and to focus their attention either on the linguistic dimension (Linguistic Task: are target words same or different as prime words?) or on the musical dimension (Musical Task: are target melodies same or different as prime melodies?). The same-different task has been used extensively in the literature to investigate the relationship between two dimensions of a stimulus in various modalities (e.g., melody recognition [66]; letter recognition [67]; meaningful environmental sounds [68]), and is particularly effective when participants are asked to attend to only one dimension at a time (see Thomas [69] for a review and in-depth analysis of the same-different task).

## Methods

### A. Participants

Twenty-one volunteers (15 females; mean age = 25 years old; age range 18–32) were paid 16 euros to participate in this experiment that lasted for about 90 minutes including preparation time. Informed consent was obtained from all participants, and the data was analyzed anonymously. Verbal consent was used because at the time of data collection, the local ethics committee did not require written consent for experiments using behavioral or ERP methods in healthy adult individuals. This study was approved by the CNRS - Mediterranean Institute for Cognitive Neuroscience and was conducted in accordance with local norms and guidelines for the protection of human subjects. All participants had normal hearing, no known neurological problems, and were native French-speaking, right-handed non-musicians (all had less than two years of formal music lessons).

### B. Stimuli

We created a set of 480 different pairs of stimuli (primes and targets). First, a list of 120 pairs of French tri-syllabic nouns was established. In each pair, the prime and target words were different and semantically unrelated. The phonological and phonetic characteristics of the words were controlled and we limited the use of certain phonemes with intrinsically longer durations (e.g. fricatives [70]), as well as consonant clusters, so that syllabic duration would be as consistent as possible between words. To increase task difficulty and to homogenize the linguistic and musical dimensions, the first syllable and the first note of the prime and target within a pair were always the same.

Next, 120 pairs of different 3-note isochronous melodies were created while controlling the harmonic content and using all 12 keys. All intervals up to the major sixth were used except the tritone. The melodic contour was also balanced across the stimuli. One quarter of the melodic pairs (30 melodies) consisted of a prime with rising contour (defined as two successive ascending intervals) paired with a target with falling contour (defined as two successive descending intervals) and vice versa for another  $\frac{1}{4}$  of the pairs. The other half of the pairs consisted of “complex” contours:  $\frac{1}{4}$  of the pairs had a prime made up of an ascending interval plus a descending interval, followed by a target with a descending plus an ascending interval, and vice-versa for the last  $\frac{1}{4}$  of the pairs. These different types of contours were evenly distributed among the experimental conditions. No melody was used more than three times, and any melody appearing more than once was always transposed into a different key and paired with a different prime melody. The melodies were written in a vocal range that was comfortable for the singer.

Finally, the pairs of melodies were randomly assigned to the pairs of words. Once the 120 different pairs had been created, they were distributed evenly over the four experimental conditions:  $W=M=$ ;  $W=M\neq$ ;  $W\neq M=$  and  $W\neq M\neq$  with 30 trials per condition (see Figure 1 and supporting materials Audio S1, Audio S2, Audio S3, Audio S4 for stimulus examples, and the Appendix S1 for a list of stimuli used). In order to control for specific stimulus effects, 4 lists were constructed so that each target appeared in all 4 conditions across the 4 lists (Latin square design).

The 120 targets and 480 primes were sung *a capella* by a baritone. Recording sessions took place in an anechoic room. In order to prevent listeners from making judgments based solely on lower-level acoustic cues, two different utterances of the sung words were selected to constitute the pairs in the  $W=M=$  conditions (in natural speech/song no two pronunciations of a segment by the same speaker are ever identical, but listeners normalize over perceived segments [71]). Although the singer sang at a tempo of 240 beats per minute to control syllable duration, natural syllabic lengthening always occurred on the last syllable/note, giving rise to an average duration of all stimuli of 913 ms ( $SD = 54$  ms). All words were normalized in intensity to 66 dB ( $SD$  across items = 1 dB).

### C. Procedure

Participants listened, through headphones, to 120 pairs of sung words from the four experimental conditions presented in pseudorandom order. The same pairs were presented twice in two attentional tasks: Linguistic and Musical. In the Linguistic task, participants were instructed to pay attention only to the language in order to decide, by pressing one of two response keys as quickly and accurately as possible, if the two words were the same or different. In the Musical Task, participants were instructed to pay attention only to the music in order to decide,

as quickly and accurately as possible, if the two melodies were the same or different.

Each session began with a block of practice trials. Each trial consisted of a prime sung word followed by a target sung word, with an SOA of 1800 ms. Participants were asked to avoid blinking until a series of X's appeared on the computer screen at the end of each trial. Response keys, order of tasks, and stimuli lists were counterbalanced across participants. The software Presentation (Neurobehavioral Systems, Albany, CA) was used to present stimuli and record behavioral responses (RTs and % errors).

#### D. Data acquisition

EEG was recorded continuously from 32 "active" (pre-amplified) Ag-AgCl scalp electrodes (Biosemi, Amsterdam) and located according to the International 10/20 system. The data were re-referenced offline to the algebraic average of the left and right mastoids. In order to detect eye movements and blinks, the horizontal electrooculogram (EOG) was recorded from electrodes placed 1 cm to the left and right of the external canthi, and the vertical EOG was recorded from an electrode beneath the right eye. The EEG and EOG signals were digitized at 512 Hz and were filtered with a bandpass of 0.1–40 Hz (post-analysis data were filtered with a lowpass of 10 Hz for visualization purposes only). Data were later segmented in single trials of 2200 ms starting 200 ms (baseline) before target onset. Trials containing ocular or movement artifacts or amplifier saturation (determined by visual inspection) were excluded from the averaged ERP waveforms (i.e., on average 12% of the trials, thereby leaving approximately 26 out of a possible 30 trials in each condition per participant). Individual data analysis and grand averages were computed using the Brain Vision Analyzer software (Brain Products, Munich).

#### E. Data Analyses

Behavioral data (RTs and arcsin-transformed Error Rates) were analyzed using a three-way ANOVA with within-subject factors: Attentional Task (Linguistic vs. Musical), Word (same vs. different), and Melody (same vs. different). A four-way ANOVA with factors Task Order, Attentional Task, Word, and Melody was computed to determine if results were influenced by the order in which participants performed the two tasks: Linguistic task first or Musical Task first. Although a main effect of Order was found, showing that the second task (whether Linguistic or Musical) was performed better than the first task (thereby reflecting increased familiarity with the experimental procedure), no significant interactions of Order with other factors were found, so this factor was not considered further.

Mean amplitude ERPs to the target words were measured in several latency bands (50–150, 150–300, 300–500, 600–800, 800–1000 ms) determined both from visual inspection and from results of consecutive analyses of 50-ms latency windows from 0 to

2000 ms. Eight regions of interest were defined by first separating the electrodes into two groups: midlines (8) and laterals (24), and then defining subsets of electrodes for analysis. The midlines were divided into two regions of interest: fronto-central: (Fz, FC1, FC2, Cz) and parieto-occipital (CP1, CP2, Pz, Oz). The lateral electrodes were separated into 6 regions of interest: left frontal (FP1, AF3, F3, F7), left temporal (FC5, T7, CP5, C3), left parietal (P3, P7, PO3, O1), right frontal (FP2, AF4, F4, F8), right temporal (FC6, T8, CP6, C4) and right parietal (P4, P8, PO4, O2). For the midline electrodes, an ANOVA with factors Attentional Task (Linguistic vs. Musical), Word (same vs. different), Melody (same vs. different) and Region (fronto-central vs. parieto-occipital) was computed on the mean amplitudes of the ERPs in each latency band. A similar ANOVA was computed for the lateral electrodes, with Attentional Task, Word, Melody, Hemisphere (left vs. right) and Region (frontal vs. temporal vs. parietal) as factors. Results of the ANOVAs are reported only when significant at  $p < 0.05$ . All  $p$  values for ERP results were adjusted with the Greenhouse-Geisser epsilon correction for nonsphericity when necessary. For both behavioral and ERP results, when interactions between two or more factors were significant, pairwise post-hoc comparisons between relevant condition pairs were computed and thresholded by Bonferroni correction. When post-hoc analysis revealed that none of the simple effects constituting an interaction reached the threshold for Bonferroni significance, the interaction was not considered further.

## Results

### Behavioral data

Mean Reaction times and Error rates are reported in Table 1.

The ANOVA on RTs showed that participants were slower in the Musical Task (1040 ms) than in the Linguistic Task (761 ms; main effect of Task [ $F(1,20) = 72.26, p < 0.001$ ]). Moreover, RTs were slower for  $W \neq$  (952 ms) than  $W =$  (849 ms; main effect of Word [ $F(1,20) = 88.46, p < 0.001$ ]). Finally, the Task x Word interaction was significant [ $F(1,20) = 22.76, p < 0.001$ ]: in the Musical Task participants were slower for  $W \neq$  (1119 ms) than for  $W =$  (961 ms; simple effect of Word: posthoc  $p < 0.001$ ) but this difference was not significant in the Linguistic Task. The Task x Melody interaction was not significant but the Word x Melody interaction was significant [ $F(1,20) = 18.44, p < 0.001$ ]: RTs were slower for  $M \neq$  (879 ms) than for  $M =$  (818 ms) only when words were same ( $W =$ ; posthoc  $p < 0.001$ ). By contrast, RTs were slower for  $W \neq$  than for  $W =$  regardless of whether melodies were same ( $M =$ ) or different ( $M \neq$ , both posthoc  $p$ 's  $< 0.001$ ).

The ANOVA on Error rates showed that participants made more errors in the Musical Task (4.21%) than in the Linguistic Task (0.87%) [main effect of Task:  $F(1,20) = 20.95, p < 0.001$ ]. Moreover, both the Task x Word and the Task x Melody interactions were significant [ $F(1,20) = 9.53, p = 0.006$  and

**Table 1.** Behavioral data.

Condition	Linguistic Task				Musical Task			
	W = M =	W = M ≠	W ≠ M =	W ≠ M ≠	W = M =	W = M ≠	W ≠ M =	W ≠ M ≠
RTs	718 (151)	756 (162)	786 (131)	783 (153)	919 (168)	1003 (153)	1129 (221)	1109 (255)
% Err	0.8 (1.5)	0.6 (1.3)	1.0 (2.1)	1.1 (1.9)	0.8 (1.8)	3.7 (5.3)	3.5 (4.5)	8.9 (9.1)

Mean Reaction Times (RTs) and errors rates (in %) for each of the 4 experimental conditions ( $W = M =$ : same word, same melody;  $W = M \neq$ : same word, different melody;  $W \neq M =$ : different word, same melody;  $W \neq M \neq$ : different word, different melody), in the Linguistic and Musical tasks. SD is indicated in parentheses.

doi:10.1371/journal.pone.0009889.t001

**Table 2.** ANOVA results on mean amplitudes of ERPs.

		Latency (ms)									
		50–150		150–300		300–500		600–800		800–1000	
Factors	df	F	p	F	p	F	p	F	p	F	p
<b>Midlines</b>	W	1,20		5.89	0.025	50.10	<0.001	5.51	0.029		
	M	1,20				6.78	0.017	10.99	0.004	7.58	0.012
	T×W	1,20	4.9†	0.039†			5.53	0.029			
	W×R	1,20				21.31	<0.001				
	M×R	1,20						6.52	0.019		
	W×M	1,20					7.14	0.015			
	T×W×R	1,20					4.65	0.044			
<b>Laterals</b>	W	1,20		4.40	0.049	28.08	<0.001				
	M	1,20				7.06	0.015	6.08	0.023		
	T×W	1,20	8.85†	0.008†	4.90†	0.039†	15.60	<0.001		4.52†	0.046†
	W×R	2,40				26.01	<0.001				
	W×M	1,20				7.19	0.014				
	W×H	1,20								5.79	0.026
	W×H×R	2,40	6.33	0.007	4.76	0.018	3.88	0.036			
	T×W×R	2,40					3.98	0.046			
	T×M×R	2,40							3.58	0.048	

Results of ANOVAs computed on midline and lateral electrodes for main effects, 2-way and 3-way interactions. Only significant effects ( $p < 0.05$ ) are shown. Abbreviations: df, degrees of freedom; T, Attentional Task; W, Word; M, Melody; H, Hemisphere; R, Region.

† Pairwise comparisons of interest did not meet the criteria for Bonferroni significance and thus the interaction is not discussed further in the text.

doi:10.1371/journal.pone.0009889.t002

$F(1,20) = 9.21$ ,  $p = 0.006$ , respectively]. In the Musical Task participants made more errors for  $W \neq$  (6.19%) than for  $W =$  (2.22%; simple effect of Word: posthoc  $p < 0.001$ ) and for  $M \neq$  (6.27%) than for  $M =$  (2.14%; simple effect of Melody: posthoc  $p < 0.001$ ), but these differences were not significant in the Linguistic Task. The Word  $\times$  Melody interaction was not significant.

### ERP data

Results of the ANOVAs on ERP data in the different latency ranges are presented in Table 2. When the main effects or relevant interactions were significant, results of pairwise posthoc comparisons are reported in the text (except for posthoc results of the Word by Melody interaction, which are reported in Table 3). The Word effect and the Melody effect in each task are illustrated on Figures 2 and 3, respectively.

Between 50 and 150 ms, different words ( $W \neq$ ) elicited a larger N100 component than Same words ( $W =$ ) over the right frontal region (Word  $\times$  Hemisphere  $\times$  Region interaction;  $p < 0.001$ ). This effect was larger in the Linguistic Task than in the Musical Task at lateral electrodes ( $p = 0.021$ ; see Figure 2), but this result did not reach significance after Bonferroni correction.

Between 150 and 300 ms,  $W \neq$  elicited a smaller P200 component than  $W =$  (main effect of Word at both midline and lateral electrodes). This effect was more prominent over bilateral frontal and left parietal regions (Word  $\times$  Hemisphere  $\times$  Region; all  $p < 0.001$ ). Again, this effect was larger in the Linguistic than in the Musical Task at lateral electrodes ( $p = 0.011$ ; see Figure 2) but this result was only marginally significant with the Bonferroni correction.

Between 300 and 500 ms,  $W \neq$  elicited a larger N400 component than  $W =$  at both midline and lateral electrodes (main effect of

Word), with larger differences over parieto-occipital than fronto-central midline electrodes (Word  $\times$  Region interaction: both  $p < 0.001$ ), and over parietal and temporal lateral regions (Word  $\times$  Region, both  $p < 0.001$ ), with a slight right hemisphere predominance (Word  $\times$  Hemisphere  $\times$  Region, both  $p < 0.001$ ). The N400 effect ( $W \neq$  minus  $W =$ ) was larger at lateral electrodes in the Linguistic ( $p < 0.001$ ) than in the Musical Task ( $p = 0.004$ ; Task  $\times$  Word) and at midlines (both  $p < 0.001$ ), with a centro-parietal scalp distribution in the Linguistic Task and a parietal distribution in the Musical Task (Task  $\times$  Word  $\times$  Region at midline and lateral electrodes, all  $p < 0.001$ ).

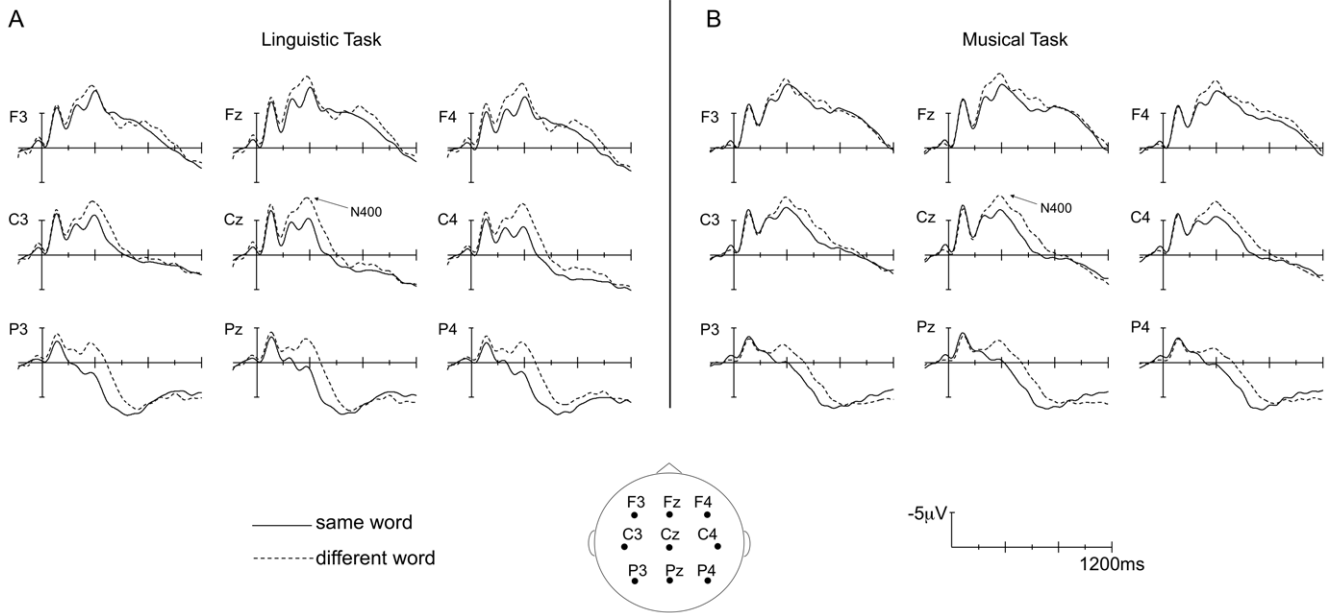
$M \neq$  elicited larger N400-like components than  $M =$  (main effect of Melody at midline and lateral electrodes; see Figure 3). Moreover, the Word  $\times$  Melody interaction was significant at midline and at lateral electrodes: the Melody effect ( $M \neq$  vs.  $M =$ )

**Table 3.** Posthoc comparisons for Word  $\times$  Melody interaction.

Pairwise Comparison	Midlines	Laterals
$W = M =$ vs. $W = M \neq$	0.006*	0.004*
$W = M =$ vs. $W \neq M =$	<0.001*	<0.001*
$W = M =$ vs. $W \neq M \neq$	<0.001*	<0.001*
$W = M \neq$ vs. $W \neq M =$	0.004*	0.032
$W = M \neq$ vs. $W \neq M \neq$	0.02	0.092
$W \neq M =$ vs. $W \neq M \neq$	0.493	0.596

Results of pairwise posthoc comparisons for the Word  $\times$  Melody interaction, in the 300–500 ms latency band. Pairs that meet the criteria for significance with the Bonferroni threshold ( $p = 0.0083$ ) are indicated with \*.

doi:10.1371/journal.pone.0009889.t003

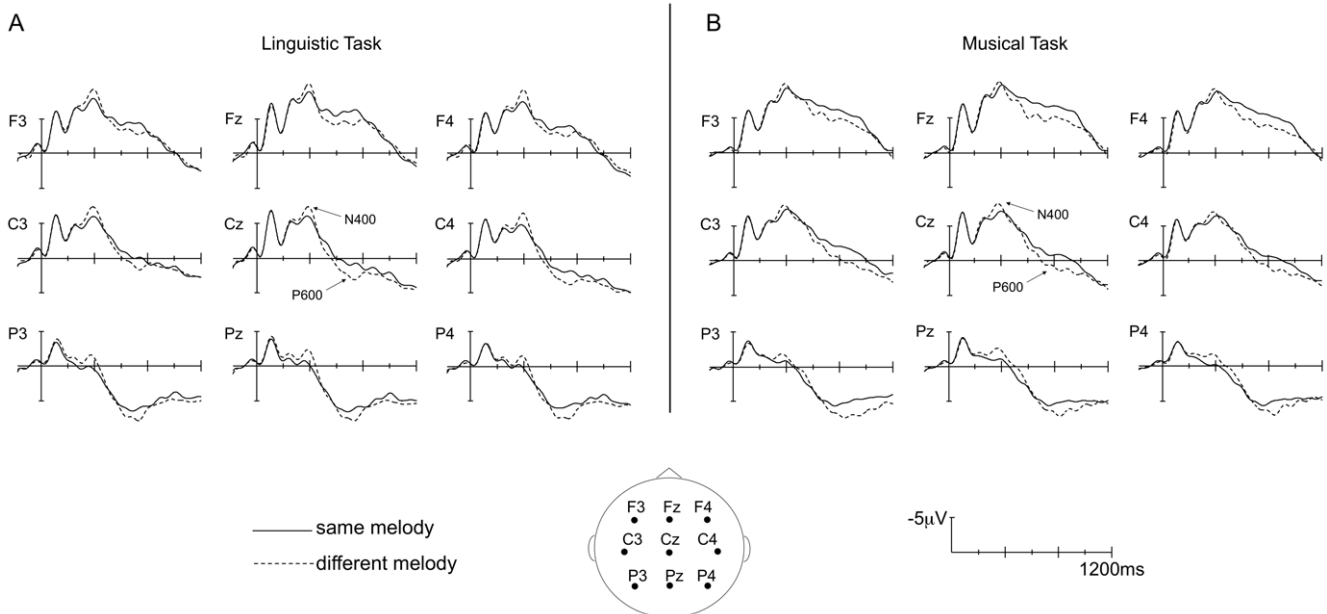


**Figure 2. Word effect.** Grand average ERPs timelocked to the onset of targets with the same word as the prime (solid line) or a different word than the prime (dashed line), in the Linguistic Task (A) and Musical Task (B). Selected traces from 9 electrodes are presented. In this figure, amplitude (in microvolts) is plotted on the ordinate (negative up) and the time (in milliseconds) is on the abscissa. doi:10.1371/journal.pone.0009889.g002

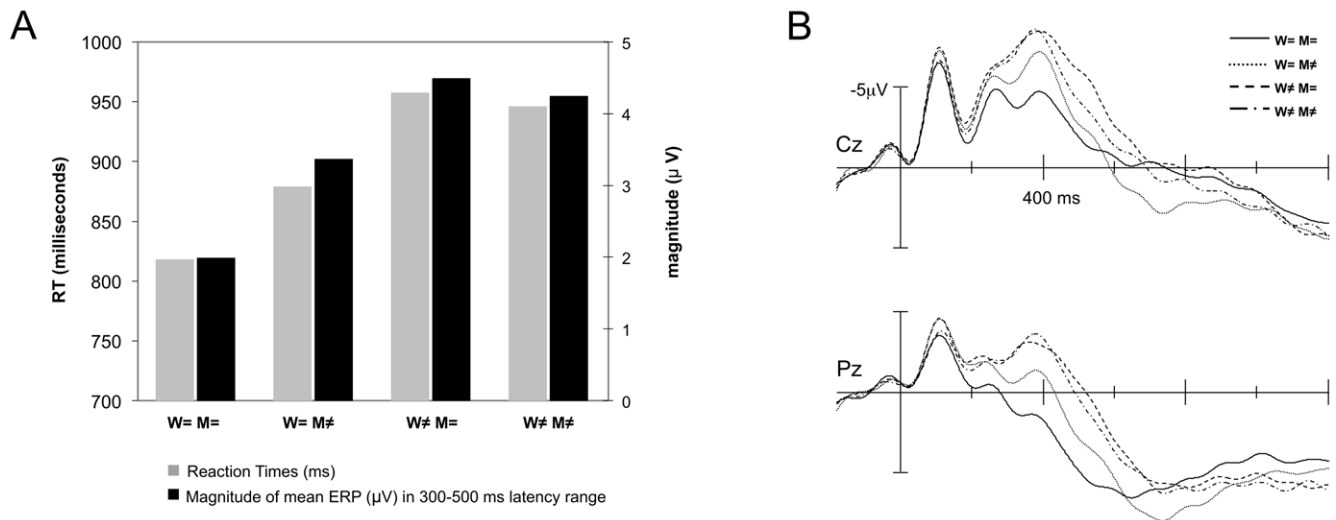
was only significant when Word was same ( $W =$ ) but not when Word was different ( $W \neq$ ; see Table 3 for all posthoc p-values for the Word x Melody interaction). Likewise, the Word effect was only significant when Melody was same ( $M =$ ) but not when Melody was different ( $M \neq$ ; see Figure 4, which shows the four orthogonal conditions averaged over both tasks). Furthermore, negative components in the  $W = M \neq$ ,  $W \neq M =$ , and  $W \neq M \neq$  conditions were larger than in  $W = M =$  condition. At the midline

electrodes, negative components were also larger in the  $W \neq M =$  than in the  $W = M \neq$  conditions.

To further test the Word by Melody interaction, difference waves were computed (on mean amplitudes) for each of the following comparisons:  $d1 = W \neq M =$  minus  $W = M =$  (effect of Word when Melody is same);  $d2 = W = M \neq$  minus  $W = M =$  (effect of Melody when Word is same);  $d3 = W \neq M \neq$  minus  $W = M =$  (effect of different Word and different Melody). If words



**Figure 3. Melody effect.** Grand average ERPs timelocked to the onset of targets with the same melody as the prime (solid line) or a different melody than the prime (dashed line), in the Linguistic Task (A) and Musical Task (B). Selected traces from 9 electrodes are presented. In this figure, amplitude (in microvolts) is plotted on the ordinate (negative up) and the time (in milliseconds) is on the abscissa. doi:10.1371/journal.pone.0009889.g003



**Figure 4. Word by Melody interaction.** (A) For each of the 4 experimental conditions (averaged across both tasks because there was no Task  $\times$  Word  $\times$  Melody interaction): the reaction time in milliseconds (gray bars, left Y-axis) and the magnitude ( $\mu\text{V}$ ) of the mean amplitude of the ERPs in the 300–500 ms latency range, averaged across all electrodes (black bars, right Y-axis). (B) ERPs associated with the 4 experimental conditions (averaged across both tasks because there was no Task  $\times$  Word  $\times$  Melody interaction) for electrodes Cz (top) and Pz (bottom). Solid line: same word, same melody; dotted line: same word, different melody; dashed line: different word, same melody; dashed-dotted line: different word, different melody. doi:10.1371/journal.pone.0009889.g004

and melodies are processed independently, then  $d1+d2$  should be equal to  $d3$ . ANOVAs with factor Data (double variation condition [ $d3$ ] vs. additive model [ $d1+d2$ ]) together with the other factors of interest (for midlines: Attentional Task and Region and for laterals: Attentional Task, Hemisphere, and Region) were carried out. Results showed that the sum of the ERP effects of the simple variations ( $d1 + d2$ ) was significantly larger than the ERP effects in the double variations condition [ $d3$ ; midline electrodes,  $F(1,20) = 7.14$ ,  $p = 0.015$ ; lateral electrodes,  $F(1,20) = 7.19$ ,  $p = 0.014$ ]; see Figure 5.

Between 600 and 800 ms,  $W \neq$  still elicited more negative ERPs than  $W =$  (main effect of Word at midline electrodes) but  $M \neq$  elicited larger late positive components than  $M =$  (main effect of Melody at midline and lateral electrodes, see Figure 3). At the midline electrodes, this effect was larger over the fronto-central region than the parieto-occipital region (both  $p < 0.001$ ; Melody  $\times$

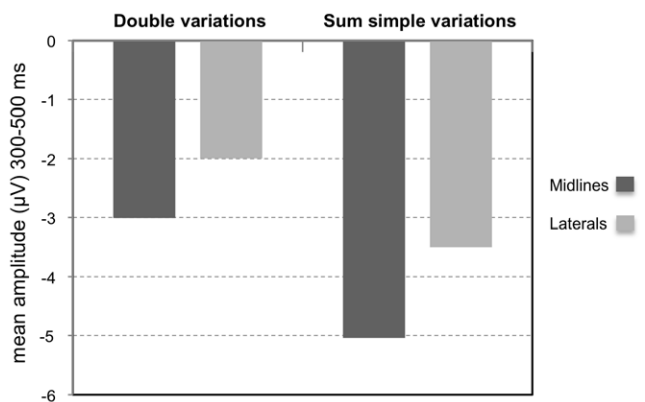
Region); furthermore, at lateral electrodes, the effect was larger over temporal and parietal regions (both  $p < 0.001$ ) in the Linguistic Task but was larger over frontal regions ( $p < 0.001$ ) in the Musical Task (Task  $\times$  Melody  $\times$  Region).

Between 800 and 1000 ms,  $W \neq$  still elicited larger negativities than  $W =$  over the right hemisphere ( $p = 0.002$ ; Word  $\times$  Hemisphere). This effect was larger in the Linguistic than in the Musical Task ( $p = 0.017$ ) but this difference did not reach significance with the Bonferroni correction. Finally,  $M \neq$  still elicited larger positive components than  $M =$  (main effect of Melody at midline electrodes).

### Scalp distribution of the N1, P2, and N400 components (Word effects)

ERPs in the N1, P2, and N400 latency bands were more negative for different word than for same word. These effects may consequently reflect an early onset of the N400 effect, or three distinct components. Since different scalp distributions were found in each of the three latency bands tested separately, it was therefore of interest to directly compare the Word effect ( $W \neq$  minus  $W =$ ) across latency bands. To this end, we conducted additional ANOVAs on the difference waves, with factors: Latency Band (50–150 ms vs. 150–300 ms vs. 300–500 ms), Hemisphere (left vs. right), and Region (frontal vs. temporal vs. parietal). Results showed a significant Latency band  $\times$  Region interaction [ $F(4,80) = 43.15$ ,  $p < 0.001$ ]. While there were no significant differences in scalp distribution between the effect of Word in the 50–150 ms (N1) and in the 150–300 ms (P2) latency bands, the topography of the N400 (300–500 ms) was different from both the N1 and the P2. Pairwise posthoc comparisons showed that the N400 had a more parietal distribution compared to the N1 ( $p < 0.001$ ) and the P2 ( $p < 0.001$ ). The Latency  $\times$  Hemisphere  $\times$  Region interaction was not significant.

In order to prevent the topographical shape of the ERPs from being potentially confounded by the amplitude of ERP effects, the same statistical analysis was then repeated on data that had undergone vector scaling (c.f. [72], but see also [73] for a



**Figure 5. Additive model test.** Mean amplitude (in  $\mu\text{V}$ ) of ERP difference waves in the 300–500 ms latency band, for double variations observed ( $W \neq M \neq$  minus  $W = M =$ ) and the modeled sum of simple variations ( $W \neq M =$  minus  $W = M =$ ) + ( $W = M \neq$  minus  $W = M =$ ), at midline electrodes (dark gray bars) and lateral electrodes (light gray bars). doi:10.1371/journal.pone.0009889.g005



discussion of the limitations of this method). The Latency x Region interaction was again significant [ $F(4,80) = 21.22, p < 0.001$ ], and pairwise posthoc tests showed the same pattern of results as in the unscaled data. This analysis therefore confirmed that the frontal distribution of the early negativities (N1/P2 complex) is significantly different from the parietal distribution of the N400.

## Discussion

### Processing the words

As predicted on the basis of several results in both the behavioral (e.g., [59]) and neurolinguistic literatures (e.g., [40,54,55,57]), sung word targets that were different from sung word primes ( $W \neq$ ) were associated with lower levels of performance (more errors and slower RTs) and with larger N400 components than same words ( $W =$ ). Thus, as noted in [40], similar processes seem to be involved in accessing the meaning of spoken and sung words. One could argue that access to word meaning was not necessary to perform the Linguistic Task and that participants could have based their decision on phonological cues. However, this is unlikely as previous work on spoken words has demonstrated that word meaning is processed automatically in phonological tasks [74,75], prosodic tasks [76–78], during passive listening in the waking state [74,75], and even during sleep [79].

Moreover, the finding that an N400 word effect also developed in the Musical Task, with similar onset latency and duration (until around 800 ms post-target onset), and a similar scalp distribution in the 300–500 ms latency range as in the Linguistic Task (centro-parietal for language and parietal for music; see Figure 2), also provides evidence in favor of the automatic processing of sung word meaning regardless of the direction of attention. The smaller size of the N400 effect in the Musical than in the Linguistic Task was most likely due to fewer attentional resources being available for processing words in the Musical Task (attention focused on the melody) than in the Linguistic Task (attention focused on words), as has been argued previously [75,76,78].

Early Word effects were also found with larger N100 components in the 50–150 ms latency band and smaller P200 components in the 150–300 ms latency band over frontal regions to different ( $W \neq$ ) than same words ( $W =$ ; see Figure 2). Even though both same and different words started with the same first syllable, which lasted for 250 ms on average, subtle articulation differences (in particular, in vowel quality and pitch of the sung syllable) were most likely present in the first syllable of different target words (e.g., the “me” in “messenger” does not sound identical to the “me” in “mélodie”). Moreover, even though the post-hoc comparison for the Task by Word interaction was not significant after Bonferroni correction between 50–150 ms and between 150–300 ms (probably because task differences were too small), it is clear from Figure 2 that the N100 and P200 effects were primarily present when participants attended to the words. Attending to the linguistic dimension may have amplified participants’ sensitivity to small differences in co-articulation, which in turn influenced the early perception of sung words, just as subtle phonetic differences modulate the N100 in speech perception [80]. This interpretation is supported by the vowel harmony phenomenon described by Nguyen & Fagyal [81], in which the pronunciation of the vowel of the first syllable assimilates to the anticipated vowel of the second syllable, which was indeed different in the  $W \neq$  conditions. We also considered the idea that the early N100 and P200 effects were the leading edge of the N400 component, in light of previous reports demonstrating the early onset of the auditory N400 effect [82], possibly reflecting the fact that lexico-semantic processing starts before the spoken

word can be fully identified [83]. However, this interpretation seems unlikely in view of the results of the scalp distribution analysis that demonstrated a significant difference between the frontally-distributed early negativities and parietally-distributed N400.

### Processing the melody

Different melodies ( $M \neq$ ) compared to same melodies ( $M =$ ) elicited larger negative components between 300 and 500 ms, followed by larger late positive components in the 600–1000 ms latency band.

The P600 component was expected based on previous reports showing that unexpected melodic/harmonic variations (e.g., [61–64,84]) elicit effects belonging to the P300 family of components. These effects are generally interpreted as reflecting the processing of surprising and task-relevant stimuli [85–87] and are indicative of the allocation of attention and memory resources (see Polich [88] for a recent review and discussion of functionally divergent P3 subcomponents). The longer onset latency of the positive effect in the present experiment than in previous studies is probably due to the fact that the first note of the melody was the same in both the  $M \neq$  and  $M =$  conditions, with the second note being sung at around 250 ms post-onset of the target. Interestingly, the task did influence the scalp distribution of the late positivity, which was frontal when the melodies were explicitly processed (Musical Task) and parietal when the melodies were implicitly processed (Linguistic Task). The frontal scalp distribution of the positive component in the Musical Task is consistent with the scalp distribution of the P3a component reported for chord sequences ending with dissonant harmonies [84] and harmonically acceptable chords with deviant timbre [89]. The parietal scalp distribution of the positive component in the Linguistic Task is consistent with previous results when participants were asked to pay attention to both lyrics and tunes [40].

Finally, it is interesting to note that late positivities, i.e., the late positive potential (LPP), have also been observed during the evaluation of affective stimuli [90,91], such as tones sung with a sad voice presented simultaneously with sad pictures [92]. In the present study, the musical dimension of the sung words, although minimal, may have called upon emotional processes, reflected by the late positivities. Further work on the emotional response to singing may clarify these issues.

One of the most interesting findings of the present study is that, prior to the late positive components,  $M \neq$  also elicited widely distributed, larger negative components than  $M =$  in the 300–500 ms latency band in both the Linguistic and Musical tasks (see Figure 3). This negativity bears the scalp distribution and peak latency typically seen for the N400 component. Indeed, N400’s have been recently associated with musical incongruities related to memory and emotional meaning, such as in familiar melodies containing an unexpected but harmonically congruous note [66], or when a mismatch ensues between musical chords and emotion words (e.g., a dissonant chord target primed by the visually presented word “love”) [18]. However, the N400 Melody effect in the present study was slightly smaller in amplitude than the N400 Word effect at the midline electrodes. The difference between these effects may be due to an overlap with the subsequent late positive component generated in the  $M \neq$  but not in the  $W \neq$  condition, but could also result from greater intrinsic salience of the linguistic dimension in songs [30,31].

Thus, in both attentional tasks, words sung on different melodies ( $M \neq$ ) were associated with larger N400 components than words sung on same melodies ( $M =$ ). Since the intonational contour of lyrics in song is provided by the musical melody, it has

been suggested that the variations in prosodic-like effects for sung lyrics could explain why words in song are better recognized with their original melodies than with a different melody [93]. In fact, several recent studies show that words spoken with prosodically incongruous patterns are associated with increased amplitudes of the N400 component followed by late positivities [78,94,95]. Thus, words sung on different melodies may hinder lexical access in a similar manner as unexpected prosodic patterns in spoken language. If familiarity is established through repeated listening to a song, which may reinforce prosodic representations of the words that are created by the melody, then the present findings may be better understood in light of results obtained by Thompson & Russo [45]. They showed that participants perceived the meaning of song lyrics as enhanced when familiarity with the songs was increased (see section 6.4 in [5] for an interesting discussion of those results). We could thus speculate that our participants' *lexico-semantic expectations for sung words* were violated not only when the target word was different from the prime ( $W \neq M =$  condition) but also when the target melody was different from the prime ( $W = M \neq$ ). This interpretation accounts for the N400 effects associated with differences on each dimension as they stand in contrast to the tight perceptual combination of repeated words and melodies ( $W = M =$ ). Further work is needed to differentiate how variations in the musical dimension of songs affect lexical access [96], general semantic memory [97], and conceptual relatedness [20,21,98]. For instance, future studies using pairs of sung words that are semantically related to each other, or sung word targets primed by other meaningful stimuli (e.g. pictures, environmental sounds, or meaningful musical excerpts), could elucidate the dynamics of the N400 component in song.

Overall, results showed that N400 components are generated when the target does not match the prime in pairs of sung words on either dimension (linguistic or musical). It must be emphasized here that these results were found regardless of the direction of attention, thereby reflecting the automatic processing of the linguistic and musical dimensions when words are sung. This pattern of results may also reflect the inability of participants to selectively focus their attention on the words or on the melodies, precisely because the two dimensions cannot be separated. We explore this possibility next.

### Interactive processing

Both behavioral and ERP data in the N400 latency band clearly revealed interactive processing of the linguistic and musical dimensions in song, which occur simultaneously in sung words. This interaction was found independently of the direction of attention (i.e., in both the Linguistic and Musical tasks and furthermore in the absence of a Task by Word by Melody interaction). Moreover, results of an ANOVA on the difference waves did demonstrate that the theoretical sum of the ERPs for simple linguistic and musical variations was significantly larger than the actual ERP in the double variation condition (see also Figure 5). Therefore, an additive model did not account for the data reported here. Furthermore, the pattern of interaction is strikingly symmetric between the two dimensions. The N400 word effect (different vs. same words) only occurs when melodies are the same; likewise, the N400 melody effect (different vs. same melodies) and the effect on RTs (slower for  $M \neq$  than  $M =$ ) only occur when words are same but not when words are different, as illustrated in Figure 4. These findings coincide with previous studies of sung and spoken language that have documented an influence of the musical dimension on linguistic processing, even when attention is directed to the linguistic aspect [43,44,46,47,99]. Thus, the main conclusion that can be drawn from these results is

that words and melody are closely interwoven in early stages of cognitive processing. This outcome is compatible with a recent report by Lidji et al. [49] of ERP evidence for interactive processing between vowel and pitch in song perception. The spatio-temporal brain dynamics of this integrated response could be responsible for interactive effects between word and melody in song, observed in a growing number of behavioral studies on perception [43,44,46,47], learning [35,36], and memory [30–33].

Some important differences between our protocol using sung word pairs and previous studies using opera excerpts [40,41] can provide an explanation for why we did not find the same tendency toward independence of neural and behavioral correlates associated with the perception of words and melodies. First, the type of same-different task employed in the present study on stimulus pairs, but not in [40] and [41], has been previously used by Miranda & Ullman [66] to show that notes that are tonally congruous (in-key) but incorrect in familiar melodies elicit both the N400 and P600 components, even when participants' attention was directed away from pitch. Furthermore, the violation paradigm used by Besson et al. [40] and Bonnel et al. [41], in which the last note of the sung phrase of the opera excerpt was not only unexpected in the context but also out-of-key, may have made wrong notes more salient for the listener than the more subtle different melody targets used in the present experiment. Indeed, even when the target melody was different than the prime, it contained tonal intervals in a reduced harmonic context. In fact, subtle stimulus variations have been used in several studies reporting interaction of linguistic and musical processing, such as the interference of harmony on phonological and semantic processing [44,46] or the interaction of semantics and harmony [17].

Nevertheless, it should be noted that the present results also provide some evidence for separate effects associated with the linguistic and musical dimensions. First, RTs were slower for different than same words regardless of whether melodies were same or different (but, as mentioned above, RTs were slower for different than for same melodies only when words were same). This slightly asymmetric pattern of interferences may be related to the fact that our non-musician participants were less accustomed to making explicit judgments about melodic information than linguistic information, as demonstrated by slower RTs in the Musical Task than in the Linguistic Task. These results correspond to those obtained in the first of a series of experiments on non-musicians by Kolinsky et al. [43] showing slower reaction times in the melodic than phonological task, in addition to an enhanced interference effect between phonology and intervals in the melodic task.

Second, while early differences were found in the 50–150 and 150–300 ms latency bands were found between same and different words (independently of the melodies), no such early differences were observed between same and different melodies. As discussed above, these early differences mostly likely reflect an effect of co-articulation caused by phonetic differences already present in the first syllable of different words rather than an early onset of the N400 word effect.

Finally, differences in the late positivity were found between same and different melodies but not between same and different words. As mentioned above, results of several experiments have shown increased P3 components to unexpected variations in melody or harmony [40,61–64,84,89], typically interpreted as reflecting the allocation of attention and memory resources to task-relevant stimuli [85–88]. The late positivity in the present study may also be related to the LPP, which is associated with the processing of affective stimuli [90–92]. Based on these accounts,

the absence of a difference in late positive components for words may reflect the fact that they were easier to process than melodies (thereby requiring fewer attentional and memory resources) or that they did not elicit an emotional response. This last interpretation could be tested in further experiments by using affective sung words as targets.

To summarize, the present results show that N400 components were elicited not only by different words but also by different melodies, although the effect of melody began later and was followed by a late positive component. Moreover, the effects of melody and word were interactive between 300 and 500 ms, thereby showing that lyrics and tunes are intertwined in sung word cognition. A companion study conducted in our lab with the fMRI method, using the same stimuli and attentional tasks, also yielded robust interactions between words and melody in songs in a network of brain regions typically involved in language and music perception [100]. These results are consistent with a growing number of studies establishing that language and music share neural resources through interactive phonological/semantic and melodic/harmonic processing (cf. [5]).

The present findings, along with other recent work on song perception and performance, are beginning to respond to the question of why song is, and has been since prehistoric times [23,24], so prevalent in the music perception and performance activities occurring in most humans' daily lives. Intrinsic shared mechanisms between words and melody may be involved in a number of song-related behaviors that have shaped human nature, although we do not yet know if the linguistic-musical interactions are the cause or effect of these tendencies. For example, it appears that infants' preference for singing over speech [26] cannot be merely attributed to the presence of the musical dimension [27] and may reflect a specific proclivity for singing-based mother-infant interactions. In early humans, adding melody to speech may have fostered parent-infant bonding and thus given an evolutionary advantage to individuals possessing more highly developed musical traits [101]. Singing to children fosters language acquisition, perhaps because exaggerated prosody aids segmentation [102] and the added musical information provides redundant cues for learning [35,36]. Melody in song may also serve as a mnemonic for storage of words in long-term memory (e.g., [33]). Research along these lines may also begin to shed light on the mechanisms responsible for the benefits of Melodic Intonation Therapy and other singing-based music therapy techniques in the speech rehabilitation process [103].

## References

- Besson M, Schön D (2001) Comparison between language and music. *Annals of the New York Academy of Sciences* 930: 232–258.
- Patel AD, Peretz I (1997) Is music autonomous from language? A neuropsychological appraisal. In: Deliege I, Sloboda J, eds. *Perception and cognition of music*. London: Erlbaum Psychology Press. pp 191–215.
- Peretz I, Coltheart M (2003) Modularity of music processing. *Nature Neuroscience* 6: 688–691.
- Koelsch S (2005) Neural substrates of processing syntax and semantics in music. *Current Opinion in Neurobiology* 15: 207–212.
- Patel AD (2008) *Music, Language, and the Brain*. New York: Oxford University Press.
- Peretz I, Kolinsky R, Tramo M, Labrecque R, Hublet C, et al. (1994) Functional dissociations following bilateral lesions of auditory cortex. *Brain* 117 (Pt 6): 1283–1301.
- Hébert S, Peretz I (2001) Are text and tune of familiar songs separable by brain damage? *Brain and Cognition* 46: 169–175.
- Racette A, Bard C, Peretz I (2006) Making non-fluent aphasics speak: sing along! *Brain* 129: 2571–2584.
- Schmithorst VJ (2005) Separate cortical networks involved in music perception: preliminary functional MRI evidence for modularity of music processing. *NeuroImage* 25: 444–451.
- Patel AD, Gibson E, Ratner J, Besson M, Holcomb PJ (1998) Processing syntactic relations in language and music: an event-related potential study. *Journal of Cognitive Neuroscience* 10: 717–733.
- Maess B, Koelsch S, Gunter TC, Friederici AD (2001) Musical syntax is processed in Broca's area: an MEG study. *Nature Neuroscience* 4: 540–545.
- Levitin DJ, Menon V (2003) Musical structure is processed in "language" areas of the brain: a possible role for Brodmann Area 47 in temporal coherence. *NeuroImage* 20: 2142–2152.
- Koelsch S, Friederici AD (2003) Toward the neural basis of processing structure in music. Comparative results of different neurophysiological investigation methods. *Annals of the New York Academy of Sciences* 999: 15–28.
- Gelfand JR, Bookheimer SY (2003) Dissociating neural mechanisms of temporal sequencing and processing phonemes. *Neuron* 38: 831–842.
- Hickok G, Buchsbaum B, Humphries C, Muftuler T (2003) Auditory-motor interaction revealed by fMRI: speech, music, and working memory in area Spt. *Journal of Cognitive Neuroscience* 15: 673–682.
- Koelsch S, Kasper E, Sammler D, Schulze K, Gunter T, et al. (2004) Music, language and meaning: brain signatures of semantic processing. *Nature Neuroscience* 7: 302–307.

## Supporting Information

**Audio S1** Example of stimulus pair in condition same word/same melody (W = M=).

Found at: doi:10.1371/journal.pone.0009889.s001 (0.22 MB WAV)

**Audio S2** Example of stimulus pair in condition same word/different melody (W = M≠).

Found at: doi:10.1371/journal.pone.0009889.s002 (0.22 MB WAV)

**Audio S3** Example of stimulus pair in condition different word/same melody (W≠ M=)

Found at: doi:10.1371/journal.pone.0009889.s003 (0.22 MB WAV)

**Audio S4** Example of stimulus pair in condition different word/different melody (W≠ M≠).

Found at: doi:10.1371/journal.pone.0009889.s004 (0.22 MB WAV)

**Appendix S1** Pairs of sung words in each of the four experimental conditions, in one list of the Latin Square design (the first author can be contacted to obtain the other three lists), with each trisyllabic French word and the 3-note melody on which it was sung (one note per syllable). The melodies are represented in standard MIDI codes, where: C4 = 60, C#4 = 61, D4 = 62, D#4 = 63, E4 = 64, F4 = 65; F#4 = 66; G4 = 67; G#4 = 68; A4 = 69; A#4 = 70; B4 = 71, C5 = 72, and so on.

Found at: doi:10.1371/journal.pone.0009889.s005 (0.24 MB DOC)

## Acknowledgments

The authors gratefully acknowledge Vanina Luigi, Elaine Ne and Monique Chiambretto for their technical assistance; Solvi Ystad for assistance with recording the stimuli; Serge Charon for singing the stimuli; the Laboratoire de Mécanique et d'Acoustique in Marseille for allowing us to use their anechoic room for sound recordings; Jill Cuadra for proofreading; and Edward Large and three anonymous Reviewers for their helpful comments on previous versions of the manuscript.

## Author Contributions

Conceived and designed the experiments: RLG DS CM CA MB. Performed the experiments: RLG CA. Analyzed the data: RLG DS. Contributed reagents/materials/analysis tools: RLG DS CM CA. Wrote the paper: RLG DS CM CA MB.

17. Steinbeis N, Koelsch S (2008) Shared neural resources between music and language indicate semantic processing of musical tension-resolution patterns. *Cerebral Cortex* 18: 1169–1178.
18. Steinbeis N, Koelsch S (2008) Comparing the processing of music and language meaning using EEG and fMRI provides evidence for similar and distinct neural representations. *PLoS One* 3: e2226.
19. Frey A, Marie C, Prod'Homme L, Timsit-Berthier M, Schön D, et al. (2009) Temporal Semiotic Units as Minimal Meaningful Units in Music? An Electrophysiological Approach. *Music Perception*. pp 247–256.
20. Schön D, Ystad S, Kronland-Martinet R, Besson M (in press) The evocative power of sounds: EEG study of conceptual priming between words and nonverbal sounds. *Journal of Cognitive Neuroscience*.
21. Daltrozzo J, Schön D (2009) Is conceptual processing in music automatic? An electrophysiological approach. *Brain Res* 1270: 88–94.
22. Gordon RL, Racette A, Schön D (2006) Sensory-Motor Networks in Singing and Speaking: a comparative approach. In: Altenmüller E, ed. *Music, Motor Control and the Brain*. New York: Oxford University Press.
23. Brown S (2000) The “musilanguage” model of music evolution. In: Wallin N, Merker B, Brown S, eds. *The Origins of Music*. Cambridge MA: MIT Press. pp 271–300.
24. Mithen SJ (2005) *The Singing Neanderthals: the origins of music, language, mind and body*. London: Weidenfeld & Nicolson.
25. Patel A (2006) Musical Rhythm, Linguistic Rhythm, and Human Evolution. *Music Perception* 24: 99–104.
26. Nakata T, Trehub SE (2004) Infants' responsiveness to maternal speech and singing. *Infant Behavior & Development* 27: 455–464.
27. de l'Etoile SK (2006) Infant behavioral responses to infant-directed singing and other maternal interactions. *Infant Behavior & Development* 29: 456–470.
28. Bartholomew B (1974) Effects of task requirements on ear superiority for sung speech. *Cortex* 10: 215–223.
29. Goodglass H, Calderon M (1977) Parallel processing of verbal and musical stimuli in right and left hemispheres. *Neuropsychologia* 15: 397–407.
30. Serafine ML, Crowder RG, Repp BH (1984) Integration of melody and text in memory for songs. *Cognition* 16: 285–303.
31. Peretz I, Radeau M, Arguin M (2004) Two-way interactions between music and language: evidence from priming recognition of tune and lyrics in familiar songs. *Memory & Cognition* 32: 142–152.
32. Wallace WT (1994) Memory for music: Effect of melody recall on text. *Journal of Experimental Psychology* 20: 1471–1485.
33. Rainey DW, Larsen JD (2002) The Effect of Familiar Melodies on Initial Learning and Long-term Memory for Unconnected Text. *Music Perception* 20: 173–186.
34. Kilgour AR, Jakobson LS, Cuddy LL (2000) Music training and rate of presentation as mediators of text and song recall. *Memory & Cognition* 28: 700–710.
35. Schön D, Boyer M, Moreno S, Besson M, Peretz I, et al. (2008) Songs as an aid for language acquisition. *Cognition* 106: 975–983.
36. Thiessen ED, Saffran JR (2009) How the melody facilitates the message and vice versa in infant learning and memory. *Ann N Y Acad Sci* 1169: 225–233.
37. Konečni VJ (1984) Elusive effects of artists' “messages”. In: Crozier WR, Chapman AJ, eds. *Cognitive Processes in the Perception of Art*. Amsterdam: North Holland. pp 71–93.
38. Stratton VN, Zalanowski AH (1994) Affective impact of music vs. lyrics. *Empirical Studies of the Arts* 12: 173–184.
39. Ali SO, Peynircioğlu ZF (2006) Songs and emotions: are lyrics and melodies equal partners? *Psychology of Music* 34: 511–534.
40. Besson M, Faïta F, Peretz I, Bonnel A-M, Requin J (1998) Singing in the brain: Independence of Lyrics and Tunes. *Psychological Science* 9: 494–498.
41. Bonnel AM, Faïta F, Peretz I, Besson M (2001) Divided attention between lyrics and tunes of operatic songs: evidence for independent processing. *Perception and Psychophysics* 63: 1201–1213.
42. van Besouw RM, Howard DM, Ternstrom S (2005) Towards an understanding of speech and song perception. *Logopedics Phoniatrics Vocology* 30: 129–135.
43. Kolinsky R, Lidji P, Peretz I, Besson M, Morais J (2009) Processing interactions between phonology and melody: Vowels sing but consonants speak. *Cognition* 112: 1–20.
44. 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: 11–20.
45. Thompson WF, Russo FA (2004) The attribution of emotion and meaning to song lyrics. *Polskie Forum Psychologiczne* 9: 51–62.
46. Poulin-Charronnat B, Bigand E, Madurell F, Peereman R (2005) Musical structure modulates semantic priming in vocal music. *Cognition* 94: 67–78.
47. Fedorenko E, Patel A, Casasanto D, Winawer J, Gibson E (2009) Structural integration in language and music: evidence for a shared system. *Memory & Cognition* 37: 1–9.
48. Garner WR, Felfoldy GL (1970) Integrality of stimulus dimensions in various types of information processing. *Cognitive Psychology* 1: 225–241.
49. Lidji P, Jolicoeur P, Moreau P, Kolinsky R, Peretz I (2009) Integrated preattentive processing of vowel and pitch: a mismatch negativity study. *Ann N Y Acad Sci* 1169: 481–484.
50. Levy DA, Granot R, Bentin S (2001) Processing specificity for human voice stimuli: electrophysiological evidence. *Neuroreport* 12: 2653–2657.
51. Levy DA, Granot R, Bentin S (2003) Neural sensitivity to human voices: ERP evidence of task and attentional influences. *Psychophysiology* 40: 291–305.
52. Bigand E, Poulin-Charronnat B (2006) Are we “experienced listeners”? A review of the musical capacities that do not depend on formal musical training. *Cognition* 100: 100–130.
53. Bentin S, Kutas M, Hillyard SA (1993) Electrophysiological evidence for task effects on semantic priming in auditory word processing. *Psychophysiology* 30: 161–169.
54. Holcomb PJ, Neville HJ (1990) Auditory and Visual Semantic Priming in Lexical Decision: A Comparison Using Event-Related Brain Potentials. *Language and Cognitive Processes* 5: 281–312.
55. Kutas M, Hillyard SA (1980) Reading senseless sentences: brain potentials reflect semantic incongruity. *Science* 207: 203–205.
56. Kutas M, Hillyard SA (1984) Brain potentials during reading reflect word expectancy and semantic association. *Nature* 307: 161–163.
57. McCallum WC, Farmer SF, Pocock PV (1984) The effects of physical and semantic incongruities on auditory event-related potentials. *Electroencephalography and Clinical Neurophysiology* 59: 477–488.
58. Besson M, van Petten CV, Kutas M (1992) An Event-Related Potential (ERP) Analysis of Semantic Congruity and Repetition Effects in Sentences. *Journal of Cognitive Neuroscience* 4: 132–149.
59. Meyer DE, Schvaneveldt RW (1971) Facilitation in recognizing pairs of words: evidence of a dependence between retrieval operations. *Journal of Experimental Psychology* 90: 227–234.
60. Neely JH (1977) Semantic Priming and Retrieval from Lexical Memory: Roles of Inhibitionless Spreading Activation and Limited-Capacity Attention. *Journal of Experimental Psychology: General* 106: 226–254.
61. Besson M, Macar F (1987) An event-related potential analysis of incongruity in music and other non-linguistic contexts. *Psychophysiology* 24: 14–25.
62. Besson M, Faïta F (1995) An Event-Related Potential study of musical expectancy: Comparison of musicians with non-musicians. *Journal of Experimental Psychology: Human Performance & Perception* 21: 1278–1296.
63. Paller KA, McCarthy G, Wood CC (1992) Event-related potentials elicited by deviant endings to melodies. *Psychophysiology* 29: 202–206.
64. Verleger R (1990) P3-evoking wrong notes: unexpected, awaited, or arousing? *The International Journal of Neuroscience* 55: 171–179.
65. Tillmann B, Janata P, Bharucha JJ (2003) Activation of the inferior frontal cortex in musical priming. *Brain Research Cognitive Brain Research* 16: 145–161.
66. Miranda RA, Ullman MT (2007) Double dissociation between rules and memory in music: an Event-Related Potential study. *NeuroImage* 38: 331–345.
67. Pachella RG, Miller JO (1976) Stimulus probability and same-different classification. *Perception and Psychophysics* 19: 29–34.
68. Gregg MK, Samuel AG (2009) The importance of semantics in auditory representations. *Atten Percept Psychophys* 71: 607–619.
69. Thomas RD (1996) Processing time predictions of current models of perception in the classic additive factors paradigm. *Journal of Mathematical Psychology* 50: 441–455.
70. Astésano C (2001) Rythme et accentuation en français. Invariance et variabilité stylistique. Paris: L'Harmattan.
71. Nguyen N (2005) La perception de la parole. In: Nguyen N, Wauquier-Gavelines S, Durand J, eds. *Phonologie et Phontique*. Paris: Hermès. pp 425–447.
72. McCarthy G, Wood CC (1985) Scalp distributions of event-related potentials: an ambiguity associated with analysis of variance models. *Electroencephalogr Clin Neurophysiol* 62: 203–208.
73. Urbach TP, Kutas M (2002) The intractability of scaling scalp distributions to infer neuroelectric sources. *Psychophysiology* 39: 791–808.
74. Perrin F, García-Larrea L (2003) Modulation of the N400 potential during auditory phonological/semantic interaction. *Brain Res Cogn Brain Res* 17: 36–47.
75. Relander K, Rämä P, Kujala T (2009) Word Semantics Is Processed Even without Attentional Effort. *J Cogn Neurosci* 21: 1511–1522.
76. Astésano C, Besson M, Alter K (2004) Brain potentials during semantic and prosodic processing in French. *Brain Research Cognitive Brain Research* 18: 172–184.
77. Hohlfeld A, Sommer W (2005) Semantic processing of unattended meaning is modulated by additional task load: evidence from electrophysiology. *Brain Res Cogn Brain Res* 24: 500–512.
78. Magne C, Astésano C, Aramaki M, Ystad S, Kronland-Martinet R, et al. (2007) Influence of syllabic lengthening on semantic processing in spoken French: behavioral and electrophysiological evidence. *Cerebral Cortex* 17: 2659–2668.
79. Ibáñez A, López V, Cornejo C (2006) ERPs and contextual semantic discrimination: degrees of congruence in wakefulness and sleep. *Brain Lang* 98: 264–275.
80. Digeres FM, Wohlbered T, Hoppe U (2009) Contribution of spectrotemporal features on auditory event-related potentials elicited by consonant-vowel syllables. *Ear Hear* 30: 704–712.
81. Nguyen N, Fagyal Z (2008) Acoustic aspects of vowel harmony in French. *Journal of Phonetics* 36: 1–27.
82. Hagoort P, Brown CM (2000) ERP effects of listening to speech: semantic ERP effects. *Neuropsychologia* 38: 1518–1530.

83. Van Petten C, Coulson S, Rubin S, Plante E, Parks M (1999) Time course of word identification and semantic integration in spoken language. *J Exp Psychol Learn Mem Cogn* 25: 394–417.
84. Janata P (1995) ERP Measures Assay the Degree of Expectancy Violation of Harmonic Contexts in Music. *J Cogn Neurosci* 7: 153–164.
85. Kutas M, McCarthy G, Donchin E (1977) Augmenting mental chronometry: the P300 as a measure of stimulus evaluation time. *Science* 197: 792–795.
86. Johnson R, Jr., Donchin E (1978) On how P300 amplitude varies with the utility of the eliciting stimuli. *Electroencephalography and Clinical Neurophysiology* 44: 424–437.
87. Donchin E (1981) Presidential address, 1980. Surprise!...Surprise? *Psychophysiology* 18: 493–513.
88. Polich J (2007) Updating P300: an integrative theory of P3a and P3b. *Clin Neurophysiol* 118: 2128–2148.
89. Carrion RE, Bly BM (2008) The effects of learning on event-related potential correlates of musical expectancy. *Psychophysiology* 45: 759–775.
90. Cunningham WA, Espinet SD, DeYoung CG, Zelazo PD (2005) Attitudes to the right- and left: frontal ERP asymmetries associated with stimulus valence and processing goals. *NeuroImage* 28: 827–834.
91. Pastor MC, Bradley MM, Low A, Versace F, Molto J, et al. (2008) Affective picture perception: emotion, context, and the late positive potential. *Brain Res* 1189: 145–151.
92. Spreckelmeyer KN, Kutas M, Urbach TP, Altenmüller E, Münte TF (2006) Combined perception of emotion in pictures and musical sounds. *Brain Res* 1070: 160–170.
93. Serafine ML, Davidson J, Crowder RG, Repp BH (1986) On the Nature of Melody-Text Integration in Memory for Songs. *Journal of Memory and Language* 25: 123–135.
94. Mietz A, Toepel U, Ischebeck A, Alter K (2008) Inadequate and infrequent are not alike: ERPs to deviant prosodic patterns in spoken sentence comprehension. *Brain Lang* 104: 159–169.
95. Schmidt-Kassow M, Kotz SA (2009) Event-related brain potentials suggest a late interaction of meter and syntax in the P600. *J Cogn Neurosci* 21: 1693–1708.
96. Lau E, Almeida D, Hines PC, Poeppel D (2009) A lexical basis for N400 context effects: evidence from MEG. *Brain Lang* 111: 161–172.
97. Kutas M, Federmeier KD (2000) Electrophysiology reveals semantic memory use in language comprehension. *Trends Cogn Sci* 4: 463–470.
98. Aramaki M, Marie C, Kronland-Martinet R, Ystad S, Besson M (in press) Sound Categorization and Conceptual Priming for Nonlinguistic and Linguistic Sounds. *Journal of Cognitive Neuroscience*.
99. Slevc LR, Rosenberg JC, Patel AD (2009) Making psycholinguistics musical: self-paced reading time evidence for shared processing of linguistic and musical syntax. *Psychon Bull Rev* 16: 374–381.
100. Schön D, Gordon R, Campagne A, Magne C, Astésano C, et al. (in press) Similar cerebral networks in language, music, and song perception. *NeuroImage*.
101. Dissanayake E (2008) If music is the food of love, what about survival and reproductive success. *Musicae Scientiae Special Issue*. pp 169–195.
102. Bergeson TR, Trehub SE (1999) Mothers' Singing to Infants and Preschool Children. *Infant Behavior & Development* 22: 51–64.
103. Norton A, Zipse L, Marchina S, Schlaug G (2009) Melodic Intonation Therapy: shared insights on how it is done and why it might help. *Ann N Y Acad Sci* 1169: 431–436.