



**HAL**  
open science

# Conceptual priming for realistic auditory scenes and for auditory words

Aline Frey, Mitsuko Aramaki, Mireille R Besson

► **To cite this version:**

Aline Frey, Mitsuko Aramaki, Mireille R Besson. Conceptual priming for realistic auditory scenes and for auditory words. *Brain and Cognition*, 2014, 84 (1), pp.141-152. 10.1016/j.bandc.2013.11.013 . hal-01384055

**HAL Id: hal-01384055**

**<https://hal.science/hal-01384055v1>**

Submitted on 21 May 2020

**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.



## Conceptual priming for realistic auditory scenes and for auditory words



Aline Frey<sup>a,\*</sup>, Mitsuko Aramaki<sup>b</sup>, Mireille Besson<sup>c,d</sup>

<sup>a</sup> Laboratoire Cognitions Humaine & Artificielle, Université Paris 8, Saint-Denis, France

<sup>b</sup> Laboratoire de Mécanique et d'Acoustique, CNRS, Marseille, France

<sup>c</sup> Laboratoire de Neurosciences Cognitives, CNRS & Aix-Marseille Université, Marseille, France

<sup>d</sup> Cuban Neuroscience Center, Habana, Cuba

### ARTICLE INFO

#### Article history:

Accepted 26 November 2013

Available online xxx

#### Keywords:

Conceptual priming  
Auditory environmental scenes  
Explicit and implicit processing  
Event-Related Potentials (ERP)  
N400

### ABSTRACT

Two experiments were conducted using both behavioral and Event-Related brain Potentials methods to examine conceptual priming effects for realistic auditory scenes and for auditory words. Prime and target sounds were presented in four stimulus combinations: Sound–Sound, Word–Sound, Sound–Word and Word–Word. Within each combination, targets were conceptually related to the prime, unrelated or ambiguous. In Experiment 1, participants were asked to judge whether the primes and targets fit together (explicit task) and in Experiment 2 they had to decide whether the target was typical or ambiguous (implicit task). In both experiments and in the four stimulus combinations, reaction times and/or error rates were longer/higher and the N400 component was larger to ambiguous targets than to conceptually related targets, thereby pointing to a common conceptual system for processing auditory scenes and linguistic stimuli in both explicit and implicit tasks. However, fine-grained analyses also revealed some differences between experiments and conditions in scalp topography and duration of the priming effects possibly reflecting differences in the integration of perceptual and cognitive attributes of linguistic and nonlinguistic sounds. These results have clear implications for the building-up of virtual environments that need to convey meaning without words.

© 2013 Elsevier Inc. All rights reserved.

### 1. Introduction

Hearing someone knocking at the door typically triggers a response: “come in”, hearing a klaxon in a busy street prevents crossing in potentially dangerous conditions and hearing the sound of the cork of a bottle of champagne is most often the signature of happy events. These few examples illustrate that natural auditory scenes carry some meaning that we are able to understand. The question addressed in this paper is whether the processes that allow us to understand the meaning of auditory scenes are similar or different from the processes used to attribute meaning to words. The well-known priming paradigm, in which a prime is followed by a target that is conceptually related to the prime or not (Meyer & Schvaneveldt, 1971), was used to address this issue. In most previous studies, environmental sounds have been used as primes and printed words as targets and vice versa (Orgs, Lange, Dombrowski, & Heil, 2006; Orgs, Lange, Dombrowski, & Heil, 2007; Orgs, Lange, Dombrowski, & Heil, 2008; Van Petten & Rheinfelder, 1995) or pictures were used as primes and environmental sounds as targets (Cummings et al., 2006; Plante, Van Petten, & Senkfor, 2000).

In all experiments, results showed typical priming effects with higher error rates and/or slower reaction times (RTs) for unrelated than for related targets, independently of the linguistic or nonlinguistic nature of the targets. Moreover, the N400 component of the Event-Related brain Potentials (ERPs; Kutas & Hillyard, 1980; Kutas, Lindamood, & Hillyard, 1984) that is related to the processing of meaning was larger for unrelated than for related targets whether they were words or environmental sounds (Cummings et al., 2006; Orgs et al., 2006; Orgs et al., 2007; Orgs et al., 2008; Plante et al., 2000; Van Petten & Rheinfelder, 1995). Therefore, conceptual priming effects (i.e., the differences between unrelated and related targets) were taken to be similar for words and for environmental sounds.

However, some differences were also reported in the scalp distribution of the N400 priming effect that was larger over fronto-central regions for environmental sounds and more centro-parietally distributed for words (Cummings et al., 2006; Orgs et al., 2006). Moreover, the N400 effect (i.e. the difference between unrelated and related items) shows some “paradoxical lateralization” (Van Petten & Rheinfelder, 1995), with larger amplitude over the right than the left hemisphere for words and no inter-hemispheric differences for environmental sounds (Plante et al., 2000). Similar inter-hemispheric differences in the processing of words and environmental sounds were also reported by Lebrun et al. (2001) using the Event-Related Desynchronisation

\* Corresponding author.

E-mail address: [aline.frey@cognition-usages.org](mailto:aline.frey@cognition-usages.org) (A. Frey).

(ERD) method. Moreover, in at least two studies, the N400 priming effect showed an earlier onset for environmental sounds than for words (Cummings et al., 2006; Orgs et al., 2006). Based on these similarities and differences, one may conclude that a common conceptual system is involved for processing the meaning of words and environmental sounds. This conclusion is in line with results of experiments using different brain imaging methods (MEG; fMRI) and showing that similar networks of brain regions are involved in the semantic/conceptual processing of speech and musical sounds (Koelsch, 2005; Koelsch et al., 2004; Steinbeis & Koelsch, 2008). Moreover, the pre-attentive and attentive processing of the acoustic attributes of speech and non-speech auditory items are associated with similar responses in the primary and secondary auditory cortex (Chandrasekaran, Kraus, & Wong, 2012; Chobert, Marie, François, Schön, & Besson, 2011; Marie, Kujala, & Besson, 2012; Tervaniemi et al., 2009). Similarly, Musacchia, Sams, Skoe, and Kraus (2007) showed that the brainstem evoked responses that originate from the inferior colliculus were morphologically similar for syllables and for harmonic sounds although with some differences in the amplitude and latency of the different components (see also Bidelman, Gandour, & Krishnan, 2009 and Kraus & Chandrasekaran, 2010 for an extensive review). Thus, the representations of speech and non-speech auditory items share some similarities all along the auditory pathway up to cortical brain regions. However, specific aspects of speech processing (e.g., phonological processing) may rely on brain regions, such as the middle to posterior parts of the Superior Temporal Sulcus (Hickok & Poeppel, 2007), that are less activated by the presentation of non-speech sounds. Finally, differences in the perceptual attributes of speech and non-speech sounds may also account for differences in scalp distribution and onset latency of the priming effects (Lebrun et al., 2001).

In the studies reviewed above, environmental sounds were sounds of everyday life (e.g., ringing of a telephone, dog barking) that entertained a strong relationship with the source that produced the sound. Thus, when hearing a dog barking, participants may associate the verbal label to the sound. The similarity of the N400 conceptual priming effect for words and for environmental sounds may consequently result from the use of a common linguistic encoding strategy rather than from a common conceptual system. In an attempt to minimize linguistic mediation, Schön, Ystad, Kronland-Martinet, and Besson (2010) presented sounds with not easily identifiable sources (i.e., a verbal label could not easily be associated to the sounds). Results still revealed larger N400 to targets that were unrelated to the primes. However, as words were part of the experimental design, it remained possible that participants used a linguistic strategy not only for words but also for sounds.

This problem was possibly bypassed in two recent studies that only used environmental or harmonic sounds both as primes and as targets with no words (Aramaki, Marie, Kronland-Martinet, Ystad, & Besson, 2010; Grieser-Painter & Koelsch, 2011). Aramaki et al. (2010) used a homogeneous class of impact sounds from different materials (wood, metal and glass) and an explicit category membership task. Primes were always typical impact sounds and targets were impact sounds from the same or from a different category than the primes. Moreover, to allow for a finer-grained analysis of conceptual priming, a third type of targets was presented that were ambiguous sounds. They were built by using a morphing technique that allowed interpolating the acoustic parameters characteristic of two typical sounds (e.g., wood and glass). For comparison purposes, auditory linguistic stimuli were also presented but in separate conditions (i.e., not inter-mixed with impact sounds). Words were used as primes with words, pseudo-words (i.e., pronounceable nonsense items) and nonwords (words played backward) used as targets. Results in the explicit categorization task

showed highest error rates and largest N400 components for ambiguous sounds and for pseudo-words. Moreover, P300 components were largest to both typical sound targets from a different category than the primes and for nonwords. Thus, effects at the behavioral and electrophysiological levels (N400 and P300 effects) were similarly modulated by linguistic and nonlinguistic sounds. Most importantly, N400 components were elicited in a sound-sound design that minimized the possibility of linguistic mediation strategies.

Similar results were recently reported by Grieser-Painter and Koelsch (2011) using related and unrelated harmonic sounds with varying timbre and visually presented words in four stimulus combinations (Sound–Sound; Sound–Word; Word–Sound and Word–Word). When participants were asked to judge whether the prime and targets fit together (explicit task), results showed an enhanced N400 component to unrelated compared to related targets in all four stimulus combinations (no Stimulus Combination by Relatedness interaction), although the N400 effect was larger for word than for sound targets. Since words were presented visually and sounds were presented auditorily, these differences possibly reflect a modality of presentation effect (visual vs auditory). By contrast, no N400 effect was found in an implicit task when participants passively listened to the same stimulus pairs in order to decide whether a probe stimulus was previously presented (recognition test). The authors concluded that, at least with short sounds (2.7 s duration on average) meaningful representations are not automatically activated in a memory task that did not require deep, elaborate sound processing. These findings thus stand in contrast with previous results showing automatic semantic priming effects for words in Lexical Decision Task (LDT; e.g., Bentin, McCarthy, & Wood, 1985; Meyer & Schvaneveldt, 1971).

In sum, both Aramaki et al. (2010) and Grieser-Painter and Koelsch (2011) have demonstrated similar conceptual priming effects for words and sounds in conditions that minimized the potential influence of linguistic mediation (sub-vocalization) strategies by using only environmental or harmonic sounds as primes and targets with no words included in the experimental conditions. Importantly, however, conceptual priming effects were only found in explicit categorization tasks and not in an implicit probe recognition task (Grieser-Painter & Koelsch, 2011). However, in a very recent study, Schirmer and collaborators found conceptual priming effects for related compared to unrelated sound pairs when participants listened to the sounds for a subsequent recognition test without explicitly focusing attention on the prime–target relationship (Schirmer, Soh, Penney, & Wyse, 2011). Thus, whether conceptual priming effects can be elicited by using implicit tasks therefore remains to be further explored.

To this aim, we conducted an experiment that included a LDT on linguistic stimuli and a modified version of the LDT on auditory environmental scenes. We also used an explicit categorization task to compare explicit and implicit conceptual priming. As in Grieser-Painter and Koelsch (2011), we used a  $4 \times 3$  design to compare, within subjects and within each of the four stimulus combinations (sound–sound (SS); word–sound (WS); sound–word (SW) and word–word (WW)), the effects of three types of conceptual relationships (Related, Ambiguous, Unrelated). However, in order to minimize carry-over effects of linguistic strategies (sub-vocalization), the SS condition was always presented first and followed by the WS, SW and WW conditions. Moreover, we used not only related and unrelated words/sounds, as in Grieser-Painter and Koelsch (2011), but also ambiguous sounds and pseudo-words as targets as in Aramaki et al. (2010). Having three types of targets allowed us to compute a finer-grained analysis of conceptual priming effects. Since the N400 has been shown to be larger to pseudo-words than to unrelated auditorily presented words (Holcomb & Neville, 1990), it was of interest to determine whether the N400

would be larger to ambiguous than to unrelated auditory scenes. Finally, an important aspect was to use a spatialized sound synthesizer to generate environmental sounds (Verron, Aramaki, Kronland-Martinet, & Pallone, 2010). This synthesizer allowed building up typical realistic auditory scenes of fire, wind and rain by manipulating a few calibrated control parameters related to the timbre and spatial extension of the sources. By listening to these sounds, listeners had a sense of strong immersion in the virtual environmental scenes. The synthesizer also allowed building up ambiguous auditory scenes by mixing the parameters of two typical auditory scenes.

## 2. Experiment 1: Explicit conceptual priming

### 2.1. Methods

#### 2.1.1. Participants

Eleven participants (mean age = 26.5 years; SD = 7.7; 4 women; all right-handed native French speakers) without hearing or neurological disorders were recruited for the experiment that lasted for about 1 h 30 min. All participants signed an informed consent to participate in the experiment that conformed to the rules of the local ethical committee and to the Declaration of Helsinki. They were paid for their participation.

#### 2.1.2. Stimuli

**2.1.2.1. Auditory environmental scenes.** Auditory scenes from 3 categories of environmental sources (fire, wind and rain) were generated with a spatialized sound synthesizer.<sup>1</sup> For each category, the synthesizer allowed the generation of calibrated sounds by manipulating intuitive control parameters linked to the physical attributes of the sources (Verron et al., 2010). Fire scenes were designed by defining the characteristics of the elementary sound sources that compose the fire (i.e., rate of crackling and combustion gain) and the overall intensity that determines the type of fire (from a small fireplace to a forest fire). Similarly to fire, rain scenes were designed by defining the characteristics of the elementary sound sources (gain of the background noise, drops rate and water streaming rate) and the overall intensity that differentiates light from heavy rains. Wind scenes were designed by defining the perceived strength and coldness of the blow (from a breeze to a storm). Together with these realistic typical scenes, the sound synthesizer also allowed us generating ambiguous auditory scenes by combining two typical auditory scenes (i.e., mixing fire/wind, fire/rain and rain/wind typical auditory scenes). All these scenes were also spatialized (spatially extended in a 3D virtual space) to increase their realism and the sense of immersion (adapted to headphone rendering). Sound examples are available at: [http://www.lma.cnrs-mrs.fr/~kronland/Environmental\\_Sounds/Priming/sounds.html](http://www.lma.cnrs-mrs.fr/~kronland/Environmental_Sounds/Priming/sounds.html).

Typical and ambiguous auditory scenes were pre-tested in a pilot experiment with 21 students who were tested collectively and who did not participate to the main experiments. They listened to 100 auditory scenes (50 a priori considered as typical and 50 a priori considered as ambiguous) and they were asked to write down all the words that came to mind within a 30 s time-limit. Results showed that 42 auditory scenes were typical in the sense that each participant wrote at least two words related to the semantic category of the scene (e.g., fire, wind or rain). Moreover, 36 scenes were ambiguous in the sense that half of the participants wrote at least two words associated to one of the component of the scene (e.g.,

related to rain in a rain/fire scene) while the other half wrote words related the other component (e.g., related to fire).

Thus, forty-two typical auditory scenes were selected with 14 auditory scenes in each category (i.e., fire<sub>(F)</sub>, wind<sub>(W)</sub> and rain<sub>(R)</sub>). In the Sound–Sound condition (SS) and for each category (Fire, Rain and wind), 2 typical scenes were used as primes ( $P_{F1}$  and  $P_{F2}$ ;  $P_{W1}$  and  $P_{W2}$ ;  $P_{R1}$  and  $P_{R2}$ ) and the 12 remaining scenes were used as targets. Two different lists were built. In list A, one of the two primes (e.g.,  $P_{F1}$ ) was paired with 6 targets to form related pairs and 6 targets were paired with a prime from a different category to form unrelated pairs (e.g., 3 with  $P_{W1}$  and 3 with  $P_{R1}$ ). In list B, the other prime (e.g.,  $P_{F2}$ ) was paired with 6 targets to form related pairs and 6 targets were paired with a prime from a different category to form unrelated pairs (e.g., 3 with  $P_{W2}$  and 3 with  $P_{R2}$ ). Thus, across the 3 categories (i.e., fire, wind and rain) and the two lists (List A and List B), 36 pairs were conceptually related (CR+) and 36 pairs were unrelated (CR–). Related and unrelated pairs were balanced across lists so that the targets that were related to the prime in List A were unrelated to the prime in list B and vice versa. Finally, a total of 36 ambiguous scenes were selected as targets (i.e., 12 fire/wind<sub>(F/W)</sub>, 12 fire/rain<sub>(F/R)</sub> and 12 rain/wind<sub>(R/W)</sub> scenes). In list A, half of the ambiguous targets were paired with one typical prime from each category ( $P_{F1}$   $P_{W1}$   $P_{R1}$ ), and the other half of the ambiguous targets with the other typical prime from each category ( $P_{F2}$   $P_{W2}$   $P_{R2}$ ). Thus, the 36 ambiguous pairs were built as follows: 6 targets<sub>F/W</sub> paired with  $P_{F1}$  and 6 targets<sub>F/W</sub> with  $P_{W1}$ ; 6 targets<sub>F/R</sub> with  $P_{F2}$  and 6 targets<sub>F/R</sub> with  $P_{R1}$  and 6 targets<sub>R/W</sub> with  $P_{R2}$  and 6 targets<sub>R/W</sub> with  $P_{W2}$ . In list B, the ambiguous targets that were paired with the typical prime  $P_{F1}$ ,  $P_{W1}$  or  $P_{R1}$  in List A, were paired with the typical prime  $P_{F2}$ ,  $P_{W2}$  or  $P_{R2}$ . Similarly, the ambiguous targets that were paired with the typical prime  $P_{F2}$ ,  $P_{W2}$  or  $P_{R2}$  in List A, were paired with the typical prime  $P_{F1}$ ,  $P_{W1}$  or  $P_{R1}$  (i.e., 6 targets<sub>F/W</sub> with  $P_{F2}$  and 6 targets<sub>F/W</sub> with  $P_{W2}$ ; 6 targets<sub>F/R</sub> with  $P_{F1}$  and 6 targets<sub>F/R</sub> with  $P_{R2}$  and 6 targets<sub>R/W</sub> with  $P_{R1}$  and 6 targets<sub>R/W</sub> with  $P_{W1}$ ). Thus, across the 3 categories (i.e., fire, wind and rain) and the two lists (List A and List B), 72 pairs were ambiguous (CR°). Each participant was presented with the two lists and the order of presentation was counter-balanced across-subjects.

In the Word–Sound condition (SW), typical and ambiguous target sounds were paired with prime sounds exactly as in the SS condition except that, rather than being a sound, the prime was one typical word of Fire ( $P_{F1}$  or  $P_{F2}$ ), of Rain ( $P_{R1}$  or  $P_{R2}$ ) or of Wind ( $P_{W1}$  or  $P_{W2}$ ).

Results of the pilot experiment also allowed us to test whether the duration of the auditory scenes influenced the results so that long auditory scenes (2000 ms) would be judged as more typical than short auditory scenes (700 ms). However, the effect of duration was not significant and prime duration was set to 2000 ms and target duration to 700 ms.

#### 2.1.3. Linguistic sounds

Based on the results of the pilot experiment, the words most frequently associated with typical auditory scenes were used in the main experiments. A total of 42 French spoken words were selected (i.e., 14 words related to fire<sub>(F)</sub>, 14 related to wind<sub>(W)</sub> and 14 related to rain<sub>(R)</sub>). In the Word–Word condition (WW) and as for auditory scenes (SS), 2 typical words from each category were used as primes ( $P_{F1}$  and  $P_{F2}$ ;  $P_{W1}$  and  $P_{W2}$ ;  $P_{R1}$  and  $P_{R2}$ ) and the 12 remaining words were used as targets. Again, two different lists were built following the same procedure as for auditory scenes so that across the 3 categories (i.e., fire, wind and rain) and the two lists (List A and List B), 36 pairs were conceptually related (CR+) and 36 pairs were unrelated (CR–). Moreover, a total of 36 pseudo-words (12 for each category) were constructed by changing one or two vowel(s) from the original words (e.g., “iverse” or

<sup>1</sup> This synthesizer was developed in the team “Modelling, Synthesis and Control of Sounds and Musical Signals”, at the Laboratory of Mechanics and Acoustics (Marseille, CNRS).

“ivorse” from “averse”) and they were paired with one of the typical primes from each category in lists A and B, as for auditory targets. Each participant was presented with the two lists and the order of presentation was counter-balanced across-subjects.

In the Sound–Word condition (SW), word and pseudo-word targets were paired with prime sounds exactly as described above for the WW condition except that, rather than being a word, the prime was one typical sound of Fire ( $P_{F1}$  or  $P_{F2}$ ), of Rain ( $P_{R1}$  or  $P_{R2}$ ) or of Wind ( $P_{W1}$  or  $P_{W2}$ ).

Words and pseudo-words were spoken by a woman, recorded in stereo at 44.1 kHz sampling frequency and adjusted in intensity level using the software “Audacity” (Version 1.2.5). Spoken words were predominantly bisyllabic with a mean duration of 586 ms ( $SD = 129$ ). The mean duration of pseudo-words was 625 ms ( $SD = 109$ ) and was not significantly different from spoken words ( $p = .09$ ).

#### 2.1.4. Procedure

Each of the four stimulus combinations (SS, WS, SW, WW) was presented in a separate block of trials with a break between conditions. To minimize carry-over effects of linguistic strategies (subvocalization), the SS condition was always presented first. The order of presentation of the other conditions (i.e., WS, SW and WW) was identical across participants. Across the two lists, a total of 144 pairs were presented for each stimulus combination with 36 conceptually related pairs (CR+), 36 conceptually unrelated pairs (CR–) and 72 ambiguous pairs (CR°). Each participant listened to each target twice, once preceded by a conceptually related prime and once by an unrelated prime. Moreover, based on previous results by Aramaki et al. (2010), we expected ambiguous auditory scenes to generate as many “Yes” as “No” responses. Thus, to equate the number of “Yes” and “No” responses the number of ambiguous targets was twice the number of CR+ or CR– targets.

Timing within a trial included the presentation of the prime for 2000 ms for sound primes and for a variable duration (mean of 586 ms) for word primes. After an Inter-Stimulus Interval (ISI) of 300 ms sound targets were presented for 700 ms and linguistic targets for a variable duration (mean 606 ms). A row of 4 crosses (i.e., “XXXX”) was presented on the screen 2 s after target onset and for 1 s to allow participants time to blink. The prime of the next pair was then presented after 1 s. A training session comprising 6 trials (2 pairs CR+, 2 pairs CR– and 2 pairs CR°) preceded each experimental block to familiarize participants with the set of stimuli and with the task.

#### 2.1.5. Task

Participants were asked to decide whether the prime and the target fit together or not. They were informed that this decision could be difficult to make for some targets (ambiguous auditory scenes and pseudo-words) but that they should nevertheless try their best. They were asked to press one of two response buttons as fast as possible while making as few errors as possible. The association between responses (Yes/No) and buttons (Left/Right hand) was counterbalanced across participants.

#### 2.1.6. Recording ERPs

Participants were seated in a Faradized room, in front of a computer screen and all stimuli were presented through headphones. The electro-encephalogram (EEG) was recorded continuously with a .01–30 Hz bandpass and at a 512 Hz sampling rate, from 32 scalp electrodes (Biosemi Active system) mounted on an elastic headcap and located at standard positions (international extended 10/20 system; Jasper, 1958): Fz, Cz, Pz, Oz, Fp1, Fp2, AF3, AF4, F7, F8, F3, F4, Fc5, Fc6, Fc1, Fc2, T7, T8, C3, C4, Cp1, Cp2, Cp5, Cp6, P3, P4, PO3, PO4, P7, P8, O1, and O2. Both the horizontal and vertical electro-oculograms (EOG) were recorded from electrodes located

over the external canthi and below the right eye, respectively. Two additional electrodes were placed on the left and right mastoids and EEG was re-referenced offline relative to the averaged activity over the two mastoids. ERPs data were analyzed using Brain Vision Analyzer software (Version 01/04/2002; Brain Products, GmbH) segmented into 1400 ms epochs (from –200 ms until 1200 ms post-target onset).

#### 2.1.7. Data analyses

Error rates and RTs were analyzed using repeated-measures Analysis of Variance (ANOVAs) that included Stimulus Combination (SS, WS, SW and WW) and Relation (CR+, CR– and CR°) as within-subject factors. Only correct responses were considered in the analysis of RTs for CR+ and CR–. Due to their ambiguous nature, ambiguous auditory scenes (CR°) can be given “Yes” or “No” responses and there are no correct responses. To match with pseudo-words that required No responses, we considered “No” as correct responses and we expected 50% correct. Tukey tests were used for post hoc comparisons.

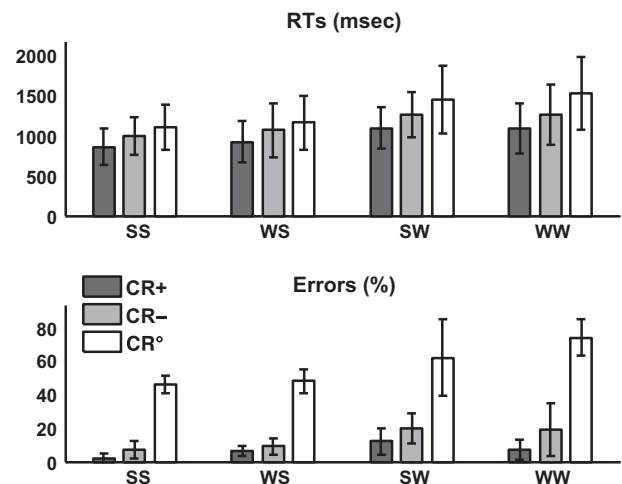
Similarly, ERPs data were analyzed for correct responses only. ANOVAs were conducted at midline and at lateral electrodes. At midline electrodes, the factors were: Stimulus Combination (4 levels), Relation (3 levels) and Electrode (Fz, Cz and Pz). At lateral electrodes, ANOVAs included the same Stimulus Combination and Relation factors plus Hemisphere (Left, LH and Right, RH), ROI (fronto-central, temporal and temporo-parietal) and Electrode (3 for each ROI: [F3–FC5–FC1/F4–FC6–FC2], [C3–CP5–CP1/C4–CP6–CP2] and [P3–P7–PO3/P4–P8–PO4]).

To simplify the presentation of the results and because our main interest was to compare the effect of explicit conceptual priming for the four stimulus combinations, we focused analyses on the Stimulus Combination  $\times$  Relation interaction. Effects were considered significant if the  $p$ -value was equal to or less than .05 ( $p$ -values were reported after the Greenhouse-Geisser correction for nonsphericity).

## 2.2. Results

### 2.2.1. Behavioral data

Reaction times (RTs) and error rates are illustrated on Fig. 1. RTs were longer for CR° (1315 ms), then for CR– (1147 ms) and then for CR+ (992 ms, all  $p < .01$ ; main effect of Relation:  $F(2,20) = 34.07$ ,



**Fig. 1.** Explicit task. Mean Reaction Times (RTs) and error rates are illustrated for each relation (CR+: conceptually related; CR–: conceptually unrelated; CR°: conceptually ambiguous) and for each stimulus combination (Sound–Sound (SS); Word–Sound (WS); Sound–Word (SW) and Word–Word (WW)).

$p < .001$ ). RTs were longer for SW (1272 ms) and WW (1294 ms), that did not differ from each other, than for WS (1051 ms) and SS (989 ms), that did not differ from each other (all  $p < .001$ ; main effect of Stimulus Combination:  $F(3,30) = 19.78, p < .001$ ).

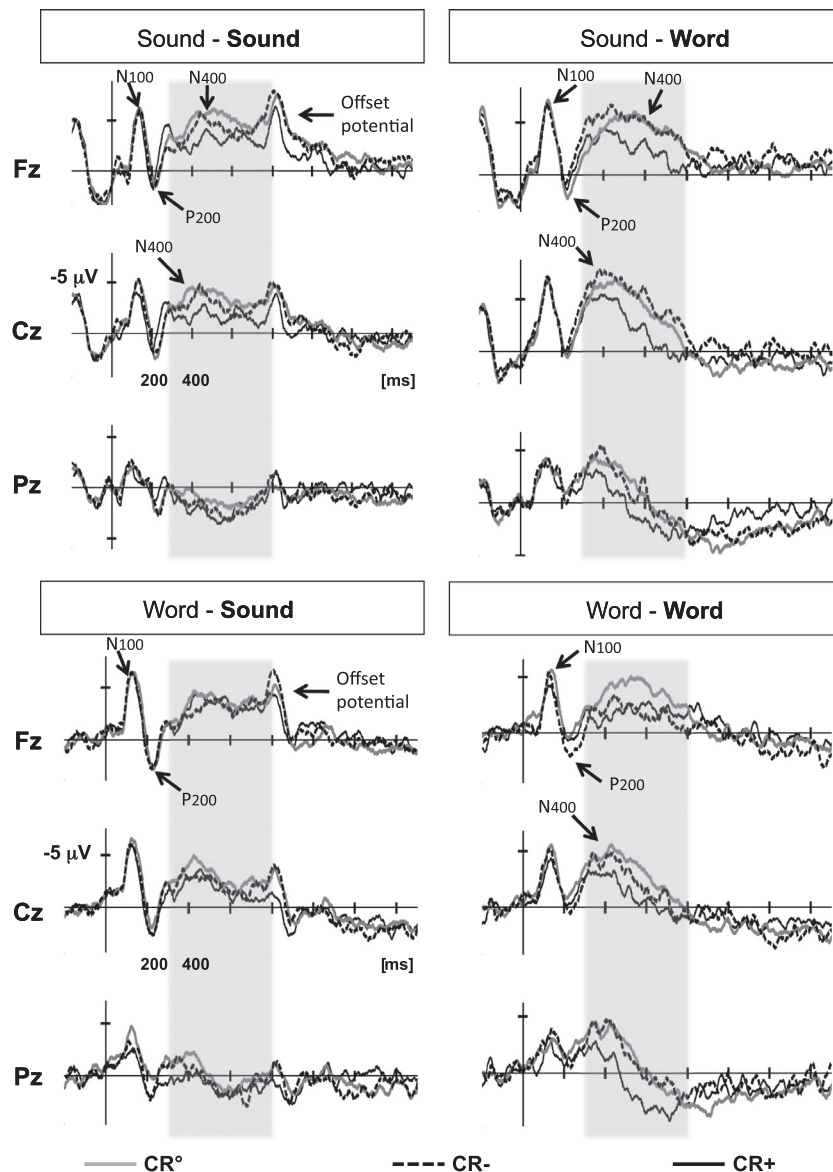
The error rate was higher for CR° (57%) than for both CR– (14%) and CR+ (7%; main effect of Relation:  $F(2,20) = 175.50, p < .001$ ). Moreover, the error rate was also higher for linguistic (32%) than for auditory scene targets (19%; main effect of Stimulus Combination  $F(3,30) = 37.59, p < .001$ ). The Stimulus Combination by Relation interaction was also significant ( $F(6,60) = 3.58, p < .05$ ): the difference in error rate between CR° and CR–/CR+ was larger in the WW condition (61%) than in the other three conditions (SS: 41.5%; WS: 40.5%; SW: 46%,  $p < .01$  in all cases).

### 2.2.2. Electrophysiological data

As can be seen on Fig. 2, following the exogenous N100–P200 components, a negative component, that we refer to as an N400 component and that was typically larger to CR° and CR– than to CR+, developed in the 300–800 ms range in all four conditions. Moreover, offset-potentials were clearly generated by auditory

scene targets that were of fixed duration (700 ms; see Fig. 2, left column). By contrast, no such offset-potentials were generated by linguistic targets that were of various durations (see Fig. 2, right column). Finally, it should also be noted that the baselines for targets are very different when the primes are sounds and when the primes are words (see Fig. 2, upper and lower rows). Again, this can be explained by the sound prime being of fixed duration and generating offset-potentials in the time interval within which the baseline is computed (200 ms before target onset). No such effects were found for word primes that were of various durations.

Analyses of the time-course of the ERPs effects in 100 ms time windows revealed significant Stimulus Combination  $\times$  Relation interactions (or Stimulus Combination  $\times$  Relation  $\times$  Electrodes/ROIs interactions) in different latency bands between 300 and 800 ms post-target onset at midline and/or lateral electrodes (see Table 1) thereby showing that the effect of the conceptual Relation differed as a function of Stimulus Combination. Separated analyses were thus conducted for each Stimulus Combination in these latency bands and results are presented in Table 2.



**Fig. 2.** Grand-averages ERPs to the targets in the explicit task and in the four experimental conditions. ERPs recorded at midline sites (Fz, Cz and Pz) are overlapped for ambiguous (CR°), unrelated (CR–) and related targets (CR+). Shaded area between 300 and 800 ms points to the latency band of most interest. In this and subsequent figures, the amplitude is in ordinate (in  $\mu\text{V}$ ) and the time is in abscissa (in milliseconds, ms).

**Table 1**

Explicit task. Summary of ERP time-course for global analyses including the four stimulus combinations in consecutive 100 ms latency bands within the 300–800 ms window, at midline and lateral electrodes. The Stimulus Combination  $\times$  Relation interactions and Stimulus Combination  $\times$  Relation  $\times$  Electrodes/ROIs interactions are reported (–: non-significant; \* =  $p < .05$ ; \*\* =  $p < .01$ ; \*\*\* =  $p < .001$ ).

	300–400	400–500	500–600	600–700	700–800
<i>Midline</i>					
Stim Comb. $\times$ Rel.	***	***	***	–	–
Stim Comb. $\times$ Rel. $\times$ Elect.	–	–	–	.07	–
<i>Lateral</i>					
Stim Comb. $\times$ Rel.	–	–	–	*	–
Stim Comb. $\times$ Rel. $\times$ ROI	–	–	–	–	–

**Table 2**

Explicit task. Summary of ERP time-course analyses in consecutive 100 ms latency bands within the 300–800 ms window for each stimulus combination (Sound–Sound (SS), Word–Sound (WS), Sound–Word (SW) and Word–Word (WW)). The main effect of the factor Relation (Rel) and the interactions with Electrodes (Elec) in midline analyses or with Region of Interest (ROI) in lateral analyses are reported (–: non-significant; \* =  $p < .05$ ; \*\* =  $p < .01$ ; \*\*\* =  $p < .001$ ).

	300–400	400–500	500–600	600–700	700–800
<i>ss</i>					
<i>Midline</i>					
Rel.	–	–	*	–	–
Rel. $\times$ Elect.	–	–	–	–	–
<i>Lateral</i>					
Rel.	–	–	–	–	–
Rel. $\times$ ROI	–	–	–	–	*
<i>ws</i>					
<i>Midline</i>					
Rel.	–	–	–	–	–
Rel. $\times$ Elect.	–	–	–	–	–
<i>Lateral</i>					
Rel. $\times$ ROI	–	–	–	–	–
<i>sw</i>					
<i>Midline</i>					
Rel.	–	*	***	–	–
Rel. $\times$ Elect.	–	–	–	–	–
<i>Lateral</i>					
Rel. $\times$ ROI	–	–	–	–	***
<i>ww</i>					
<i>Midline</i>					
Rel.	–	*	**	–	–
Rel. $\times$ Elect.	–	–	–	–	–
<i>Lateral</i>					
Rel. $\times$ ROI	–	–	$p = .05$	–	–

### 2.2.3. Sound–Sound

The N400 component was significantly larger to CR<sup>o</sup> (–2.55  $\mu$ V) and CR– (1.84  $\mu$ V) than to CR+ (–0.39  $\mu$ V) at midline electrodes between 500–600 ms (main effect of Relation:  $F(2,20) = 4.55$ ,  $p < .05$ ). Moreover, at lateral electrodes, the difference between CR<sup>o</sup>/CR– and CR+ was larger over fronto-central and temporal regions than over temporo-parietal regions between 700 and 800 ms (Relation by ROI: ( $F(4,40) = 3.67$ ,  $p < .05$ ).

### 2.2.4. Word–Sound

No significant differences were found at midline electrodes. By contrast, at lateral electrodes and in the 400–500 ms range, the N400s was larger to CR<sup>o</sup> than to CR+ at temporo-parietal sites (Relation by ROI: ( $F(4,40) = 5.25$ ,  $p < .05$ ).

### 2.2.5. Sound–Word

From 400 ms to 600 ms, the N400 component was significantly larger to CR<sup>o</sup> and CR– than to CR+ at midline electrodes (Main effect of Relation: 400–500 ms:  $F(2,20) = 4.25$ ,  $p < .05$ ; 500–600 ms:  $F(2,20) = 9.87$ ,  $p < .001$ ). While these differences were widespread across the scalp, they were larger over frontal than parietal sites in the 700–800 ms range (Relation by Electrode interaction:  $F(4,40) = 3.21$ ,  $p = .05$ ). Moreover, at lateral electrodes, the N400

was larger to CR– and CR<sup>o</sup> than to CR+ over fronto-central regions between 700 and 800 ms (Relation by ROI interaction:  $F(4,40) = 7.47$ ,  $p < .001$ ).

### 2.2.6. Word–Word

From 400 ms to 700 ms at midline electrodes, the N400 component was significantly different in the three relatedness conditions. Moreover, the N400 was larger to CR<sup>o</sup> than to both CR– and CR+ at frontal sites but larger to both CR<sup>o</sup> and CR– than to CR+ at parietal sites (Relation by Electrode interaction: 400–500 ms:  $F(4,40) = 3.35$ ,  $p < .05$ ; 500–600 ms:  $F(4,40) = 4.81$ ,  $p < .05$ ; 600–700 ms:  $F(4,40) = 4.12$ ,  $p < .05$  and at lateral electrodes: Relation by ROI in the 500–600 ms range:  $F(4,40) = 3.60$ ,  $p = .05$ ).

### 2.3. Discussion

In line with our hypothesis and with previous results (e.g., Aramaki et al., 2010; Holcomb & Neville, 1990), ambiguous targets in all four experimental conditions were associated with longer RTs, higher error rates and larger N400 components than related targets thereby showing conceptual priming effects for linguistic and non-linguistic sounds. Differences between unrelated and related targets were also present in all conditions except in the WS condition. Importantly, the Stimulus Combination by Relation or the Stimulus Combination by Relation  $\times$  Electrodes/ROIs interactions were significant on error rates, RTs and in different time windows between 300 and 800 ms post-target onset thereby also revealing differences in conceptual priming effects across conditions.

On average, the error rate was low for typical auditory scenes whether they were related (4%) or unrelated (8%) to the prime. This shows that participants were able to accurately decide whether auditory scene targets fit with the auditory prime or not. By contrast, the error rate was close to 50% for ambiguous auditory scene targets (47% on average). This was expected since ambiguous scenes were created by mixing-up the attributes of two typical scenes. Consequently, some of their attributes were conceptually related to the typical scenes used as primes and favored “Yes” responses (considered as incorrect). More surprisingly, the error rate was very high for pseudo-words (68%). However, upon debriefing participants mentioned that they were often trying to guess which real word the pseudo-word was issued from. This guessing strategy likely interfered with the explicit decision on the prime-target relationship. Note that if this interpretation is valid, the error rate to pseudo-words should be lower in the implicit task in which the use of guessing strategies should be minimized.

Turning to RTs and in line with the literature (e.g., Holcomb & Neville, 1990; Meyer & Schvaneveldt, 1971), CR– were associated with longer RTs than CR+ when the prime was a word (WS and WW conditions). Importantly, similar results were also found for auditory scenes (SW and SS conditions). Thus, conceptual priming was found for related compared to unrelated pairs in the explicit categorization task independently of whether the target or the prime was a linguistic or a nonlinguistic sound.

Behavioral data also revealed that auditory scenes were overall easier to process than linguistic targets. Upon debriefing participants mentioned that because they were spatialized, auditory scenes created a strong sense of immersion that facilitated their decision. However, differences in the level of processing required to make the decision may also account for this difference. While semantic processing was required for linguistic targets, lower levels of processing were possibly involved for auditory scenes. In that case, the decision could be based upon the match between perceptual and physical attributes of the sound sources (e.g., rate of crackling, rate of drops or presence of blast of wind) used as primes and targets. In line with this view, recent results by Schirmer et al. (2011) showed similar perceptual priming effects for meaningful and non-meaningful sound pairs.

The electrophysiological data revealed that the N400 was larger for ambiguous than for related targets in all four Stimulus Combinations. Thus, analysis of the ERPs also provided evidence that conceptual priming effects were generated not only by linguistic stimuli but also by nonlinguistic auditory scenes. Importantly, such priming effects were found when auditory scenes were presented both as primes and as targets, and before the other experimental conditions, in an attempt to minimize the influence of linguistic strategies (e.g., sub-vocalization, verbal labeling). While using a fixed presentation order possibly introduced practice effects, in that participants were more familiar with the stimuli at the end than at the beginning of the experiment, the relatively high number of stimulus repetition within each stimulus combination, is likely to have counter-balanced the effects of familiarity. In this respect, it should be noted that the high number of prime repetition (each prime was presented 24 times) is unlikely to have influenced target processing because for each stimulus combination, each prime was paired with an equal number of related (6) unrelated (6) and ambiguous targets (6 Yes and 6 No). Thus, participants could not guess the status of the target upon presentation of the prime. Taken together, the present results therefore allow generalizing to auditory scenes previous results with homogeneous classes of impact sounds (wood, metal and glass; Aramaki et al., 2010).

Importantly, the N400 component was also larger to unrelated (CR–) than to related (CR+) targets in all conditions except in the Word–Sound condition. Interestingly, this finding replicates the non-significant difference between unrelated and related targets in the Word–Sound condition of Grieser-Painter and Koelsch (2011). In their experiment, the absence of priming in the Word–Sound condition was found whether data were averaged as a function of pre-experimentally defined categories, as in the present experiment, or according to participants' own judgments. Thus, the two types of averages led to similar results. Moreover, in both Grieser-Painter and Koelsch's and in the present experiment, priming effects were significant in the Sound–Word condition. Thus, a word seems less able to prime a sound, be it a harmonic sound or an auditory scene, than a sound can prime a word. This may reflect a human bias to label the elements, including sounds that are surrounding us.

Fine-grained analyses of the time-course of conceptual priming also revealed spatio-temporally localized differences between the WW and SS conditions. In the WW condition, the N400 was significantly larger for CR– than for CR+ between 400 and 700 ms over centro-parietal midline sites and between 500 and 600 ms at lateral electrodes. While the CR– vs CR+ difference was also significant in the SS condition, it was localized over fronto-central electrodes (with no difference over parietal electrodes). Thus, and as reported in previous studies (Aramaki et al., 2010; Cummings et al., 2006; Grieser-Painter & Koelsch, 2011; Orgs et al., 2006), conceptual priming effects were frontally distributed for auditory scene targets and centro-parietally distributed for linguistic targets (see Kutas & Federmeier, 2011, for a review). Finally, conceptual

priming effects were temporally more extended for linguistic targets than for auditory scenes.

In sum, results of this experiment revealed differences in scalp distribution and duration of conceptual priming effects for auditory scene and linguistic targets. However, they also showed evidence for conceptual priming on both behavioral and electrophysiological measures (higher error rates, longer RTs and larger N400 for ambiguous and unrelated targets than for related targets). As these results were found when the relationship between the prime and the target was explicitly processed, the next step was to determine whether similar effects would develop when the prime-target relationship was implicitly processed.

### 3. Experiment 2: Implicit conceptual priming

The aim of Experiment 2 was to determine whether conceptual priming effects are elicited in an implicit task in which the relationship between the prime and the target is not the focus of attention. To compare results between Experiments 1 and 2, the stimuli, conditions and procedure were the same but the task varied. We used a Lexical Decision Task (LDT) for linguistic targets and a variant for auditory scene targets. If the conceptual relationship between the prime and the target is implicitly processed, conceptually unrelated and ambiguous targets should generate more errors, longer RTs and larger negativity in the 300–600 ms latency band (N400 components) than conceptually related targets.

#### 3.1. Methods

##### 3.1.1. Participants

Seventeen participants (mean age = 24.7 years; SD = 2.1; 9 women; all right-handed native French speakers), without hearing or neurological disorders, did participate to the experiment that lasted for about 1 h 30 min. None had participated in Experiment 1. All participants signed an informed consent to participate in the experiment that conformed to the rules of the local ethical committee and to the Declaration of Helsinki. They were paid for their participation. Three participants were excluded from the analyses because too many artifacts (e.g., blinks and muscular movements) were present in the electrophysiological data.

Stimuli and Procedure as well as ERP recordings and Data analyses were the same as in Experiment 1.

##### 3.1.2. Task

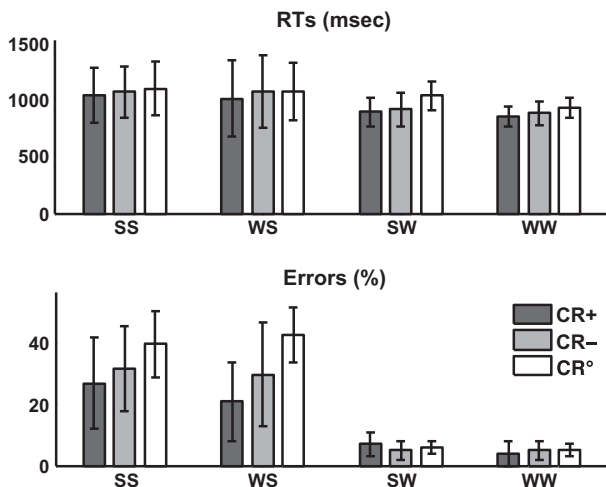
Participants were asked to focus their attention on the targets. For linguistic targets (SW and WW conditions), they had to decide whether the target was a word or a pseudo-word and for nonlinguistic targets (SS and WS conditions), they had to decide whether the target was a typical or an ambiguous auditory scene. In both cases, they had to indicate their choice by pressing one of two response buttons as fast as possible while making as few errors as possible. The association between responses (Yes/No) and buttons (Left/Right hand) was counterbalanced across participants.

#### 3.2. Results

##### 3.2.1. Behavioral data

RTs and error rates are illustrated in Fig. 3. RTs for CR° (1042 ms) were longer than for CR– (991 ms) and CR+ (956 ms; main effect of Relation:  $F(2,26) = 14.07$ ,  $p < .001$ ). The Stimulus Combination by Relation interaction was also significant ( $F(6,78) = 2.7$ ,  $p = .05$ ): the CR° vs CR– and CR+ differences were largest in the SW condition. Moreover, RTs were longer for auditory scene targets (1069 ms) than for linguistic targets (925 ms; main effect of Stimulus Combination:  $F(3,39) = 7.13$ ,  $p < .01$ ).





**Fig. 3.** Implicit task. Mean Reaction Times (RTs) and error rates are illustrated for each Relation (CR+: conceptually Related; CR-: conceptually unrelated; CR°: conceptually ambiguous) and for each stimulus combination (Sound-Sound (SS); Word-Sound (WS); Sound-Word (SW) and Word-Word (WW)).

The error rate was higher for CR° (23%) than for CR- (18%) and CR+ (15%; main effect of Relation:  $F(2,26) = 10.94$ ,  $p < .001$ ). The Stimulus Combination by Relation interaction was also significant ( $F(6,78) = 7.52$ ,  $p < .001$ ): conceptual priming was significant for auditory scenes (higher error rate for CR° (41%) than for CR+ (24%),  $p < .01$ ) but not for linguistic targets ( $p > .10$ ). Moreover, the error rate was also higher auditory scene (32%) than for linguistic targets (5%; main effect of Stimulus Combination:  $F(3,39) = 83.77$ ,  $p < .001$ ).

### 3.2.2. Electrophysiological data

As can be seen on Fig. 4, negative components that we refer to as N400 components developed in the 300–800 ms range in all four conditions following the N100–P200 components. Overall, the amplitude of the negativity was largest to CR°, intermediate to CR- and smallest to CR+. As in Experiment 1 (since the stimuli are the same), offset-potentials were generated by auditory scene targets (see Fig. 4, left column) but not by linguistic targets (see Fig. 4, right column). Also as in Experiment 1, baselines were different for sound and word primes (see Fig. 4, upper and lower rows). Both effects can be explained by sound primes being of fixed duration (and generating offset-potentials) and by word primes being of various durations (and not generating such potentials).

Analyses of the time-course of the ERPs effects in 100 ms time windows revealed significant Stimulus Combination  $\times$  Relation interactions (or Stimulus Combination  $\times$  Relation  $\times$  Electrodes/ROIs interactions) in different latency bands between 400 and 800 ms post-target onset at midline and/or lateral electrodes (see Table 3) thereby showing that the effect of the conceptual Relation differed as a function of Stimulus Combination. Separated analyses were thus conducted for each Stimulus Combination in these latency bands and results are presented in Table 4.

### 3.2.3. Sound-Sound

At midline fronto-central electrodes, the N400 was significantly larger to CR° ( $-5.54 \mu\text{V}$ ) than to both CR- ( $-3.92 \mu\text{V}$ ) and CR+ ( $-3.94 \mu\text{V}$ ) in the 400–500 ms latency range (Relation by Electrode:  $F(4,52) = 3.56$ ,  $p < .05$ ). By contrast, in both the 600–700 and the 700–800 ms ranges and still at fronto-central electrodes, the N400 was significantly larger to CR° ( $-4.99 \mu\text{V}$ ) and CR- ( $-3.76 \mu\text{V}$ ) than to CR+ ( $-2.61 \mu\text{V}$ ; Relation by Electrode: 600–700 ms:  $F(4,52) = 3.29$ ,  $p < .05$ ; 700–800 ms:  $F(4,52) = 3.98$ ,

$p < .05$ ). Results showed no significant differences at lateral electrodes.

### 3.2.4. Word-Sound

The three types of conceptual relationship significantly differed from each other at Fz electrode in the 600–800 ms latency range with the largest negativity to CR° ( $-5.19 \mu\text{V}$ ), intermediate to CR+ ( $-3.96 \mu\text{V}$ ) and smallest to CR- ( $-2.41 \mu\text{V}$ ; Relation by Electrode: 600–700 ms:  $F(4,52) = 4.77$ ,  $p < .005$ ; 700–800 ms:  $F(4,52) = 6.98$ ,  $p < .005$ ). Results showed no significant differences at lateral electrodes.

### 3.2.5. Sound-Word

At midline electrodes, the N400 was significantly larger to CR° ( $-4.69 \mu\text{V}$ ) than to CR+ ( $-2.23 \mu\text{V}$ ) between 500 and 800 ms (main effect of Relation: 500–600 ms:  $F(2,26) = 8.29$ ,  $p < .001$ ; 600–700 ms:  $F(2,26) = 21.29$ ,  $p < .001$ ; 700–800 ms:  $F(2,26) = 5.24$ ,  $p < .05$ ). Moreover, between 500–600 ms and 600–700 ms, the N400 component was larger to CR° ( $-4.64 \mu\text{V}$ ) than to both CR- ( $-1.95 \mu\text{V}$ ) and CR+ ( $-1.12 \mu\text{V}$ ). In the 600–700 ms range, the three conditions differed from each other at Cz (Relation by Electrode:  $F(4,52) = 2.85$ ,  $p = .05$ ). Finally, in the 700–800 ms latency band, the N400 was still larger to CR° than to both CR- and CR+ at Cz and Pz (Relation by Electrode:  $F(4,52) = 3.20$ ,  $p = .05$ ).

At lateral electrodes, the N400 was significantly larger to CR° ( $-3.49 \mu\text{V}$ ) than to both CR- ( $-1.73 \mu\text{V}$ ) and CR+ ( $-1.04 \mu\text{V}$ ) between 500 and 800 ms (main effect of Relation: 500–600 ms:  $F(2,26) = 10.77$ ,  $p < .001$ ; 600–700 ms:  $F(2,26) = 27.18$ ,  $p < .001$ ; 700–800 ms:  $F(2,26) = 12.06$ ,  $p < .001$ ).

### 3.2.6. Word-Word

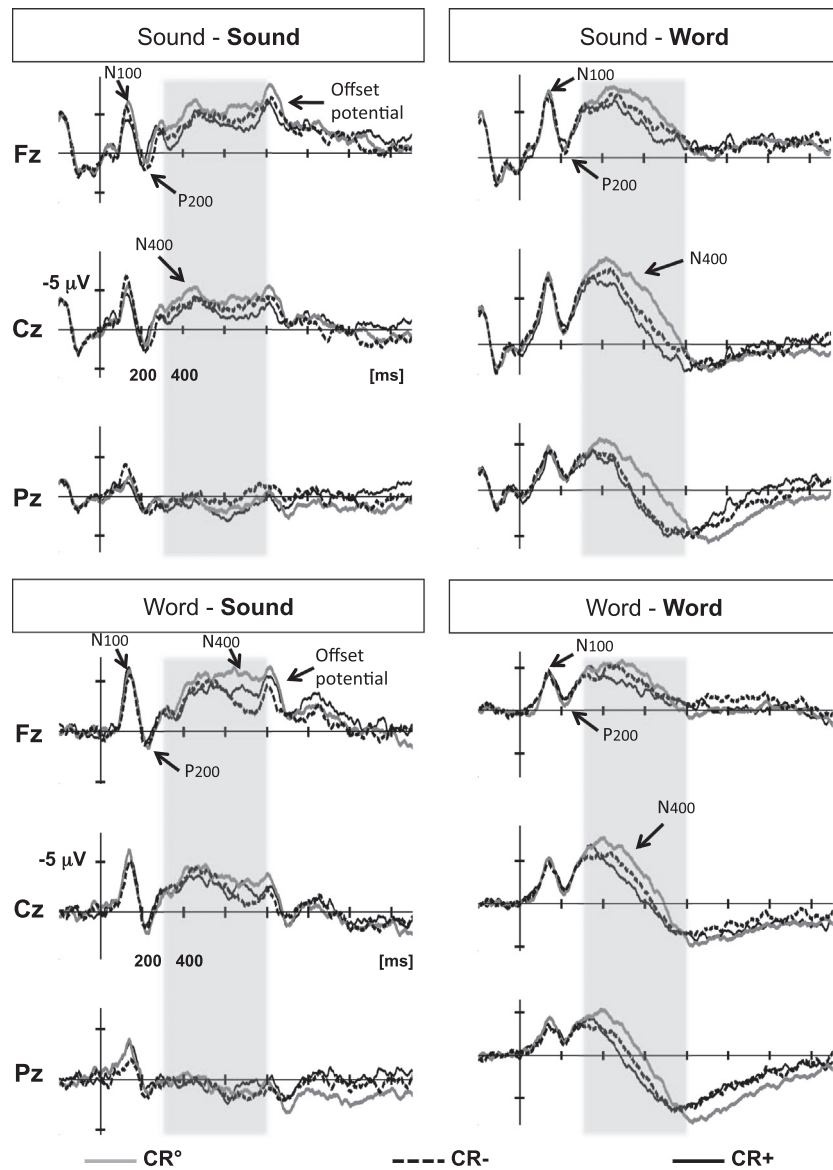
At parietal lateral electrodes and in the 400–500 ms latency band, the three types of conceptual relationship significantly differed from each other, with largest N400 to CR° ( $-4.45 \mu\text{V}$ ), intermediate to CR- ( $-3.34 \mu\text{V}$ ) and smallest to CR+ (lat:  $-2.39 \mu\text{V}$ ; Relation by ROI interaction:  $F(4,52) = 6.04$ ,  $p < .01$ ). In the 500–600 ms and 600–700 ms and at both midline and lateral electrodes, the N400 was significantly larger to CR° (mid:  $-2.26 \mu\text{V}$ ; lat:  $-1.84 \mu\text{V}$ ) than to both CR- (mid:  $-0.50 \mu\text{V}$ ; lat:  $-0.35 \mu\text{V}$ ) and CR+ (mid:  $0.34 \mu\text{V}$ ; lat:  $0.35 \mu\text{V}$ ; main effect of Relation: mid: 500–600 ms:  $F(2,26) = 14.85$ ,  $p < .001$ ; 600–700 ms:  $F(2,26) = 7.33$ ,  $p < .005$ ; lat: 500–600 ms:  $F(2,26) = 16.08$ ,  $p < .001$ ; 600–700 ms:  $F(2,26) = 9.99$ ,  $p < .01$ ).

### 3.2.7. Between-task comparisons

To directly compare conceptual priming effects between the two experiments, Task (explicit vs implicit) was included as a between-subject factor in the following analyses. We focus on the Task by Relation and Task by Relation by Stimulus Combination interactions that are of main interest here.

In the explicit task, RTs to CR+ (992 ms) were faster than to CR- (1147 ms) that were also faster than to CR° (1315 ms). By contrast, in the implicit task, RTs to CR+ (956 ms) and CR- (991 ms) did not differ and RTs to CR+ were faster than to CR° (1042 ms; Task by Relation interaction: ( $F(2,46) = 18.45$ ,  $p < .001$ ). The Task by Relation by Stimulus Combination was not significant ( $F(6,138) = 1.72$ ,  $p = .17$ ).

In the explicit task, the error rate was lower to both CR+ (7%) and CR- (14%) than to CR° (57%). By contrast, in the implicit task, only CR+ (15%) was lower than CR° (23%; Task by Relation interaction: ( $F(2,46) = 95.18$ ,  $p < .001$ ). The Task by Relation by Stimulus Combination was also significant ( $F(6,138) = 8.78$ ,  $p < .001$ ). The differences between CR° and CR-/CR+ were larger in the explicit than in the implicit task specifically for the two auditory scene conditions (SS and WS).



**Fig. 4.** Grand-averages ERPs to the targets in the implicit task and in the four experimental conditions. ERPs recorded at midline sites (Fz, Cz and Pz) are overlapped for ambiguous (CR°), unrelated (CR-) and related targets (CR+). Shaded area between 300 and 800 ms points to the latency band of most interest.

**Table 3**

Implicit task. Summary of ERP time-course for global analyses including the four stimulus combinations in consecutive 100 ms latency bands within the 300–800 ms window, at midline and lateral electrodes. The stimulus combination × Relation interactions and stimulus combination × Relation × Electrodes/ROIs interactions are reported (-: non-significant; ° =  $p < .05$ ; °° =  $p < .01$ ; °°° =  $p < .001$ ).

	300–400	400–500	500–600	600–700	700–800
<i>Midline</i>					
Stim Comb. × Rel.	-	-	°	°	°
Stim Comb. × Rel. × Elect.	-	-	-	°°°	°°°
<i>Lateral</i>					
Stim Comb. × Rel.	-	°°	°°	°°	°°
Stim Comb. × Rel. × ROI	-	-	°	°°°	°°°

Finally, between-tasks comparisons in the 300–800 ms latency range showed that in the WW condition, differences between CR+ and CR- at Fz were larger in the explicit than in the implicit task (Task by Relation by Stimulus Combination by Electrode:  $F(12,276) = 3.12, p < .01$ ; see Fig. 5). By contrast, in the SW condition, differences between CR° and CR+/CR- were larger in the implicit than in the explicit task (Task by Relation:  $F(2,46) = 5.14,$

$p < .009$ ; see Fig. 5). These interactions were not significant for nonlinguistic targets (always  $p > .08$ ).

### 3.3. Discussion

Results of Experiment 2 showed that ambiguous targets were associated with higher error rates, longer RTs and larger N400

**Table 4**

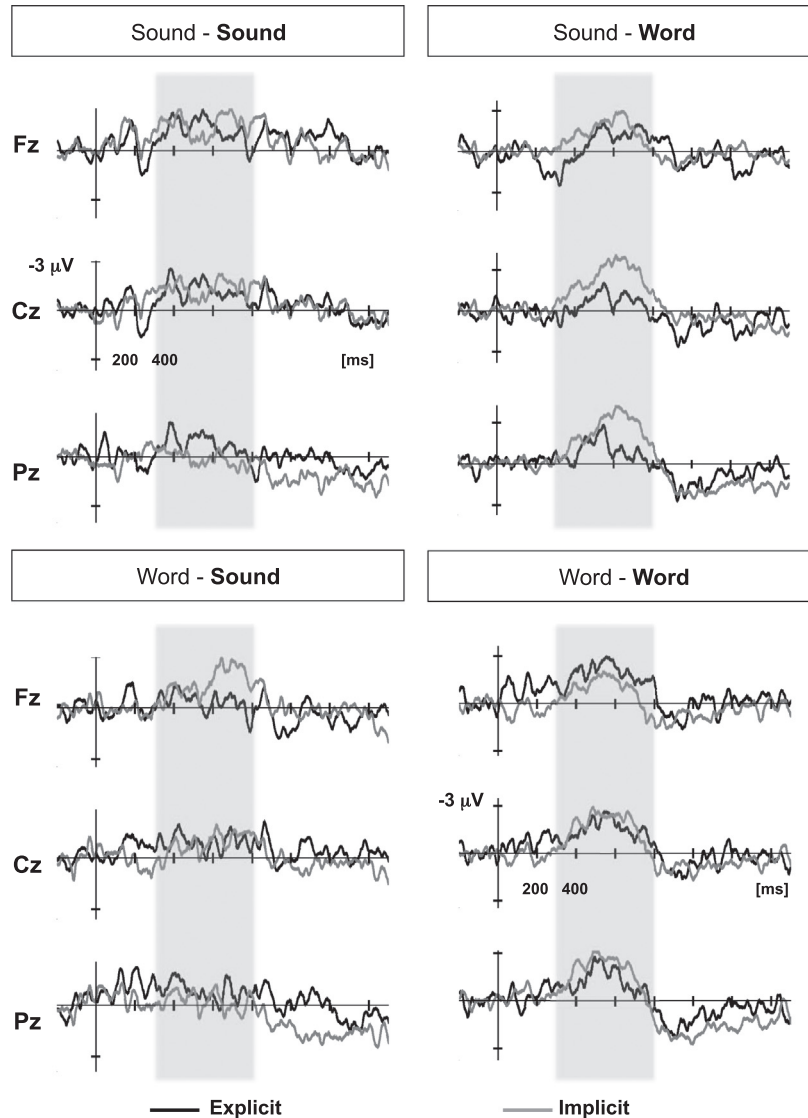
Implicit task. Summary of ERP time-course analyses in consecutive 100 ms latency bands within the 300–800 ms window for each stimulus combination (Sound–Sound (SS), Word–Sound (WS), Sound–Word (SW) and Word–Word (WW)). The main effect of the factor Relation (Rel) and the interactions with Electrodes (Elec) are reported (-: non-significant; \* =  $p < .05$ ; \*\* =  $p < .01$ ; \*\*\* =  $p < .001$ ).

	300–400	400–500	500–600	600–700	700–800
<i>ss</i>					
Rel.	-	-	-	*	*
Rel. × Elect.	-	-	-	*	*
<i>ws</i>					
Rel.	-	-	-	**	**
Rel. × Elect.	-	-	-	**	**
<i>sw</i>					
Rel.	-	-	***	***	**
Rel. × Elect.	-	-	-	$p = .05$	$p = .05$
<i>ww</i>					
Rel.	-	-	***	**	-
Rel. × Elect.	-	-	-	-	-

components than related targets. Thus, conceptual priming effects were elicited even when the relationship between the prime and ambiguous target was not the focus of attention (implicit task).

However, conceptual priming effects on RTs were larger for linguistic (SW) than for auditory scene targets. Without an explicit processing of the prime, it was probably difficult to decide whether the auditory scene target was typical or ambiguous. This interpretation is supported by the longer RTs to auditory scene targets compared to linguistic targets in the implicit task (while the reverse was found in the explicit task). Note, however, that in contrast to RTs, conceptual priming on error rate was significant for auditory scenes (higher error rate for CR° than for CR+) but not for linguistic targets. As the error rate to linguistic targets was very low, the lack of conceptual priming probably reflects a floor effect.

Importantly, the differences between unrelated and related targets were not significant either on RTs or on error rate. However, fine-grained analyses of the time-course and scalp distribution of the N400 effect revealed significant differences not only between CR° and CR+ but also between CR- and CR+. Although these differences were localized in time and space, the unrelated vs related differences were found in all four conditions. In the WW condition, the three conditions significantly differed from each other across all electrodes in the 400–500 ms latency band, with largest N400 to CR°, intermediate to CR- and smallest to CR+. These differences were also present in the SW condition but they were significant



**Fig. 5.** Difference waves ( $CR^\circ - [(CR^+) + (CR^-)]/2$ ) in the Explicit task (black line) and in the implicit task (gray line), at representative midline electrodes (Fz, Cz, Pz), for the four conditions (SS, WS, SW, WW).

later, in the 600–700 ms range and centrally localized. In the WS condition, they developed in the 600–700 ms and 700–800 ms ranges, and were frontally localized. Finally, in the SS condition, the N400 was significantly larger fronto-centrally to CR° than to both CR– and CR+ in the 400–500 ms latency range but larger to CR° and CR– than to CR+ in the 600–800 ms range.

This pattern of results differ from [Grieser-Painter and Koelsch \(2011\)](#) who report no significant differences in the N400 to related and unrelated targets in any of their four experimental conditions. In their implicit task, participants were asked to pay attention to the stimuli for a later recognition memory test. The authors argued that three different possibilities can account for the lack of N400 effects in their implicit task: the recognition test may require too shallow level of processing, stimuli duration may be too short to automatically activate meaningful associations or the effects of these two factors may interact. However, the auditory scenes (700 ms in duration) and the linguistic stimuli (around 600 ms) used in the present experiment were of shorter duration than the stimuli used by [Grieser-Painter and Koelsch \(2011\)](#); 2700 ms on average). Similarly, [Daltrozzo and Schön \(2009\)](#) reported significant N400 effects with musical excerpts that were 1 s in duration. Thus, stimulus duration *per se* does not seem to be the most important factor. As [Daltrozzo and Schön \(2009\)](#) also used a lexical decision task, one could argue that implicit N400 effects are generated when items are processed for lexical decision but not for later recognition. This interpretation is unlikely, however, since in a seminal study, [Koelsch et al. \(2004\)](#) reported N400 effects in a recognition memory task. Moreover, [Schirmer et al. \(2011\)](#) found clear N400 priming effects when participants listened to pairs of environmental sounds for later recognition without explicitly focusing attention on the sound relationship within the pairs. Importantly, however, conceptual priming effects were only found for meaningful, recognizable sounds (e.g., door unlocking, telephone dialing) but not for unrecognizable shuffled version of these same sounds. Based upon this contrastive set of results, a possible interpretation is that N400 effects are generated in implicit tasks (whether lexical decision or recognition memory) when the meaning conveyed by the stimuli (e.g., words, musical excerpts, environmental sounds, auditory scenes, etc.) is sufficiently strong to automatically activate their associated representations. This may not be the case when sounds only differ in timbre as in the experiment by [Grieser-Painter and Koelsch \(2011\)](#).

Importantly, direct comparisons of results in Experiments 1 and 2, also revealed interesting differences in explicit and implicit conceptual priming effects. These are discussed below.

### 3.4. General discussion

At the behavioral level, the main difference between the two experiments was that conceptual priming effects were found for related (CR+) compared to unrelated targets (CR–) on both RTs and error rates in the explicit task but not in the implicit task. Thus, in line with previous results in the literature (e.g., [Bentin, Kutas, & Hillyard, 1993](#); [Shiffrin & Schneider, 1977](#)), the present results showed larger priming effects in the explicit than in the implicit task. Nevertheless, in both tasks, conceptual priming effects were found for related compared to ambiguous targets on both RTs and error rates. Thus, when between-stimuli differences are large (between CR° and CR+) they trigger significant conceptual priming effects in both tasks but when they are smaller (between CR– and CR+) they trigger significant priming effects only when they are explicitly processed. One could argue, however, that because ambiguous and related targets are very different from each other, other differences than conceptual priming may account for the results. It is indeed the case that in some experiments, for instance using non-pronounceable pseudo-words (e.g., [Holcomb & Neville,](#)

[1990](#)) or words played backward ([Aramaki et al., 2010](#)), these ambiguous targets are so different that they entail different processes than unrelated targets words. This is reflected in the ERPs by the occurrence of P300 components. However, in previous experiments ([Aramaki et al., 2010](#); [Holcomb & Neville, 1990](#)) as well as in the present experiments, pseudo-words are associated with N400 components similar to unrelated words but with larger amplitude. This is taken to indicate that similar conceptual processes of graded amplitude are involved in both cases. This issue is more complicated for ambiguous auditory scenes because they are indeed very different from the typical auditory scenes. However, in the present experiments, as in [Aramaki et al. \(2010\)](#), ambiguous scenes also elicited N400-like components rather than P300 components. Thus, we are confident that conceptual processes are also involved for ambiguous auditory scenes. Taken together, these results illustrate the importance of testing several types of prime-target relationships and not only related vs unrelated pairs.

Large differences between the explicit and implicit tasks were found for linguistic targets on error rate (32% vs 5%). However, as mentioned in the discussion of Experiment 1, participants possibly adopted a guessing strategy (e.g., from which word is the pseudo-word issued from?) in the explicit task that may have interfered with the task at hand and led to the CR° high error rate. Such a strategy was of no use in the lexical decision task since participants were asked to focus attention on the target to decide whether the target item was a word or a pseudo-word without processing the prime-target relationship. As a consequence, the difference in error rate likely reflects differences in the strategies used to perform the task.

In both experiments and in contrast to behavioral measures, the amplitude of the N400 component was not only larger for ambiguous targets than for related targets but also for unrelated than for related targets. Thus, the N400 proved to be a more sensitive measure than RTs or error rate. However, the N400 effect to auditory scenes showed a frontally localized scalp distribution, as was previously reported for environmental sounds ([Cummings et al., 2006](#); [Orgs et al., 2006](#)) and for impact sounds of wood, metal and glass ([Aramaki et al., 2010](#)). By contrast, the distribution of the N400 effect to linguistic stimuli was more widespread and centro-parietally localized (cf. [Kutas & Federmeier, 2011](#), for a review). Also, the N400 to CR– and CR+ in the WW condition was larger over the right than the left hemisphere, with no such differences in the other conditions (see [Cummings et al., 2006](#), and [Lebrun et al., 2001](#) for similar results). Thus, the scalp distribution of the N400 component as well as the duration of the conceptual priming effects did vary across conditions. Since linguistic and nonlinguistic sounds are perceptually different, these differences possibly reflect different integration of their perceptual and cognitive attributes ([Schirmer et al., 2011](#)).

In sum, while results of both experiments revealed differences in scalp distribution and duration of the conceptual priming effects for auditory scenes and linguistic targets, results also showed that the conceptual relationship between the prime and the target was both explicitly and implicitly processed and influenced both behavioral and electrophysiological measures. In this respect, it is interesting to note that the N400 developed earlier in the explicit task (between 400 and 600 ms) than in the implicit task (between 500 and 800 ms; see [Tables 2 and 4](#)). Previous results by [Orgs, Lange, Dombrowski, and Heil \(2007\)](#) also showed N400 effects in implicit (a physical task in which participants decided whether the sound was presented to the right or to the left ear) and explicit tasks (a semantic task in which participants decided whether the prime word and the target sound were conceptually related) with small differences in time-course and scalp distribution. Similar findings were reported when the physical task was always

presented first in order to avoid carry-over semantic effects and when processing of the meaning of the sound was not required (Orgs, Lange, Dombrowski, & Heil, 2008). However, these results and ours, showing similar conceptual priming effects in explicit and implicit tasks stand in contrast with the absence of conceptual priming in an implicit recognition task reported by Grieser-Painter and Koelsch (2011). The recent report by Schirmer and colleagues (2011) is revealing in this respect as they found conceptual priming effects for recognizable environmental sounds but not for perceptually similar but unrecognizable sounds in an implicit recognition task. Thus, rather than reflecting differences between explicit and implicit conceptual priming, the different results reported across experiments may reflect the increased power of environmental sounds (Orgs, Lange, Dombrowski, and Heil, 2008; Orgs et al., 2006; Schirmer et al., 2011) and auditory scenes to implicitly evoke meaning compared to single harmonic sounds.

In conclusion, by showing behavioral and electrophysiological conceptual priming effects between two auditory scenes, the present results add to a growing dataset showing that the processing of meaning is not tight to the processing of words. Processing the meaning of auditory scenes, whether explicitly or implicitly, entails a complex set of computations and interactions, some of them possibly domain-specific and others domain general and shared with the processing of word meaning. Importantly, these results have strong implications for the building-up of virtual environments that need to convey meaning without words.

## Acknowledgments

This research was supported by a Grant from the ANR-07-NEURO (#024-01) to Mireille Besson and was conducted while Dr. Aline Frey was a half-time post-doctoral researcher in Dr Besson's laboratory (at the "Institut de Neurosciences Cognitives de la Méditerranée") and working half-time at the Laboratory "Cognitions Humaine et Artificielle". We would like to thank Hugues Perrin for his help in building up the stimuli and for running part of the experiments as well as Dr. Charles Verron for his help in recording the auditory scenes.

## References

- Aramaki, M., Marie, C., Kronland-Martinet, R., Ystad, S., & Besson, M. (2010). Sound categorization and conceptual priming for nonlinguistic and linguistic sounds. *Journal of Cognitive Neuroscience*, 22(11), 2555–2569.
- Bentin, S., Kutas, M., & Hillyard, S. A. (1993). Electrophysiological evidence for task effects on semantic priming in auditory word processing. *Psychophysiology*, 30(2), 161–169.
- Bentin, S., McCarthy, G., & Wood, C. C. (1985). Event-related potentials, lexical decision, and semantic priming. *Electroencephalography and Clinical Neurophysiology*, 60, 343–355.
- Bidelman, G. M., Gandour, J. T., & Krishnan, A. (2009). Cross-domain effects of music and language experience on the representation of pitch in the human auditory brainstem. *Journal of Cognitive Neuroscience*, 23(2), 425–434.
- Chandrasekaran, B., Kraus, N., & Wong, P. C. M. (2012). Human inferior colliculus activity relates to individual differences in spoken language learning. *Journal of Neurophysiology*, 107(1325–1336), 2012.
- Chobert, J., Marie, C., François, C., Schön, D., & Besson, M. (2011). Enhanced passive and active processing of syllables in musician children. *Journal of Cognitive Neuroscience*, 23(12), 3874–3887.
- Cummings, A., Ceponiene, R., Koyama, A., Saygin, A. P., Townsend, J., & Dick, F. (2006). Auditory semantic networks for words and natural sounds. *Brain Research*, 1115, 92–107.
- Daltrozzo, J., & Schön, D. (2009). Conceptual processing in music as revealed by N400 effects on words and musical targets. *Journal of Cognitive Neuroscience*, 21(10), 1882–1892.
- Grieser-Painter, J. G., & Koelsch, S. (2011). Can out-of-context musical sounds convey meaning? An ERP study on the processing of meaning in music. *Psychophysiology*, 48(5), 645–655.
- Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience*, 8(5), 393–402.
- Holcomb, P. J., & Neville, H. J. (1990). Auditory and visual semantic priming in lexical decision: A comparison using event-related brain potentials. *Language and Cognitive Processes*, 5, 281–312.
- Jasper, H. H. (1958). The ten–twenty system of the International Federation. *Electroencephalography and Clinical Neurophysiology*, 10, 371–375.
- Koelsch, S. (2005). Neural substrates of processing syntax and semantics in music. *Current Opinion in Neurobiology*, 207–212.
- Koelsch, S., Kasper, E., Sammler, D., Schulze, K., Gunter, T., & Friederici, A. D. (2004). Music, language and meaning: Brain signatures of semantic processing. *Nature Neuroscience*, 7(3), 302–307.
- Kraus, N., & Chandrasekaran, B. (2010). Music training for the development of auditory skills. *Nature Reviews Neuroscience*, 11, 599–605.
- Kutas, M., & Federmeier, K. D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event-related brain potential (ERP). *Annual Review of Psychology*, 62, 621–647.
- Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science*, 207, 203–204.
- Kutas, M., Lindamood, T. E., & Hillyard, S. A. (1984). Word expectancy and event-related brain potentials during sentence processing. In S. Kornblum & J. Requin (Eds.), *Preparatory states and processes* (pp. 217–237). Hillsdale, New Jersey: Lawrence Erlbaum.
- Lebrun, N., Clochon, P., Étévenon, P., Lambert, J., Baron, J. C., & Eustache, F. (2001). An ERD mapping study of the neurocognitive processes involved in the perceptual and semantic analysis of environmental sounds and words. *Cognitive Brain Research*, 11, 235–248.
- Marie, C., Kujala, T., & Besson, M. (2012). Musical and linguistic expertise influence pre-attentive and attentive processing of non-speech sounds. *Cortex*, 48(4), 447–457.
- Meyer, D. E., & Schvaneveldt, R. W. (1971). Facilitation in recognizing pairs of words: Evidence of a dependence between retrieval operations. *Journal of Experimental Psychology*, 90, 227–234.
- Musacchia, G., Sams, M., Skoe, E., & Kraus, N. (2007). Musicians have enhanced subcortical auditory and audiovisual processing of speech and music. *Proceedings of the National Academy of Sciences of the United States of America*, 104, 15894–15898.
- Orgs, G., Lange, K., Dombrowski, J., & Heil, M. (2006). Conceptual priming for environmental sounds and words: An ERP study. *Brain and Cognition*, 62, 267–272.
- Orgs, G., Lange, K., Dombrowski, J., & Heil, M. (2007). Is conceptual priming for environmental sounds obligatory? *International Journal of Psychophysiology*, 65, 162–166.
- Orgs, G., Lange, K., Dombrowski, J. H., & Heil, M. (2008). N400-effects to task-irrelevant environmental sounds: Further evidence for obligatory conceptual processing. *Neuroscience Letters*, 436, 133–137.
- Plante, E., Van Petten, C., & Senkfor, A. J. (2000). Electrophysiological dissociation between verbal and nonverbal semantic processing in learning disabled adults. *Neuropsychologia*, 38, 1669–1684.
- Schirmer, A., Soh, Y. H., Penney, T. B., & Wyse, L. (2011). Perceptual and conceptual priming of environmental sounds. *Journal of Cognitive Neuroscience*, 23(11), 3241–3253.
- Schön, D., Ystad, S., Kronland-Martinet, R., & Besson, M. (2010). The evocative power of sounds: Conceptual priming between words and nonverbal sounds. *Journal of Cognitive Neuroscience*, 22(5), 1026–1035.
- Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending and a general theory. *Psychological review*, 84(2), 127.
- Steinbeis, N., & Koelsch, S. (2008). Shared neural resources between music and language indicate semantic processing of musical tension-resolution patterns. *Cerebral Cortex*, 18(5), 1169–1178. <http://dx.doi.org/10.1093/cercor/bhm149> (New York, NY: 1991).
- Tervaniemi, M., Kruck, S., De Baene, W., Schröger, E., Alter, K., & Friederici, A. (2009). Top-down modulation of auditory processing: Effects of sound context, musical expertise, and attentional focus. *European Journal of Neuroscience*, 30, 1636–1642.
- Van Petten, C., & Riefelder, H. (1995). Conceptual relationships between spoken words and environmental sounds: Event-related brain potential measures. *Neuropsychologia*, 33, 485–508.
- Verron, C., Aramaki, M., Kronland-Martinet, R., & Pallone, G. (2010). A 3D immersive synthesizer for environmental sounds. *IEEE Transactions on Audio, Speech and Language Processing*, 18(6), 1550–1561.