The evocative power of sounds: conceptual priming between words and nonverbal sounds

Daniele Schön, Sølvi Ystad, Richard Kronland-Martinet, Mireille R Besson

The Evocative Power of Sounds: Conceptual Priming between Words and Nonverbal Sounds

Daniele Schön[1], Sølvi Ystad[2], Richard Kronland-Martinet[2], and Mireille Besson[1]

Abstract

■ Two experiments were conducted to examine the conceptual relation between words and nonmeaningful sounds. In order to reduce the role of linguistic mediation, sounds were recorded in such a way that it was highly unlikely to identify the source that produced them. Related and unrelated sound–word pairs were presented in Experiment 1 and the order of presentation was reversed in Experiment 2 (word–sound). Results showed that, in both experiments, participants were sensitive to the conceptual relation between the two items. They were able to correctly categorize items as related or unrelated with good accuracy. Moreover, a relatedness effect developed in the event-related brain potentials between 250 and 600 msec, although with a slightly different scalp topography for word and sound targets. Results are discussed in terms of similar conceptual processing networks and we propose a tentative model of the semiotics of sounds. ■

INTRODUCTION

Most research on the question of how we are able to derive meaning from the external world has been investigated by studies on language. Although this line of research turned out to be very fruitful, leading to models of semantic processing (see McNamara, 2005), it remains a highly debated question whether mechanisms for meaning allocation rely on cerebral resources that are specific to language or that are common to other domains. Understanding the meaning of language may require specific functional and anatomical pathways. Alternatively, similar neural networks may be involved for linguistic information and for other types of meaningful information such as objects, pictures, nonlinguistic sounds, or music. In this article, we will prefer the term concept to the term meaning, because the former is a general term, whereas the latter is often associated to semantics and linguistics. One way of studying conceptual processing is to look at context effects on the processing of a target stimulus. In a seminal study, Kutas and Hillyard (1980) showed that the amplitude of a negative component of the event-related potentials (ERPs) peaking around 400 msec postword onset, the N400 component, is larger for final words unrelated to the preceding sentence context than for related words (*The pizza was too hot to cry/eat*). Thereafter, the N400 has been widely used to investigate semantic processing in language, using the classical semantic priming paradigm, wherein one single word is used to create a context that influences the processing of a following target word (Bentin, McCarthy, & Wood, 1985). More recently, several researchers have become interested in studying whether an N400 can be elicited and modulated by the conceptual relation in a nonlinguistic context. Indeed, several studies have been published on conceptual processing with pictures (Holcomb & McPherson, 1994), odors (Castle, Van Toller, & Milligan, 2000; Sarfarazi, Cave, Richardson, Behan, & Sedgwick, 1999), and music (Daltrozzo & Schön, 2009; Frey et al., 2009; Koelsch et al., 2004).

Within the auditory domain, one way of comparing linguistic and nonlinguistic conceptual processing has been to use spoken words and environmental sounds. Environmental sounds are interesting in that they bear a direct relation with the source of the sound. They establish a reference to an object (bottle, cork, corkscrew) or an action (turn, pull, open). A number of studies have used the ERP method and the classical priming paradigm to study the conceptual processing of environmental sounds. To our knowledge, the first study was conducted by Van Petten and Rheinfelder (1995). They presented spoken words followed by environmental sounds and vice-versa. Words preceded by unrelated sounds evoked a larger N400 than those preceded by related sounds. This N400 effect (i.e., the difference between unrelated and related targets) was slightly lateralized to the right hemisphere. Sounds preceded by unrelated words also evoked a larger N400 than those preceded by related words but this effect was larger over the left hemisphere. Orgs, Lange, Dombrowski, and Heil (2006, 2007) used a similar design but with shorter stimuli (300 msec instead of 2500 msec)

[1]CNRS & Université de la Méditerranée, Marseille, France, [2]CNRS, Marseille, France

and also found similar effects on the N200 and N400 components. Finally, Cummings et al. (2006, p. 104) compared behavioral and electrophysiological responses to words, environmental sounds, and nonmeaningful sounds ("not easily associated with any concrete semantic concept") in semantically matching or mismatching visual contexts (photos). They found that words and environmental sounds mismatching the visual context evoked a larger N400 than words and environmental sounds matching the visual context. By contrast, no differences were found for the so-called nonmeaningful sounds. These sounds were selected so that they always fit either a smooth or a jagged category, and should, as such, have evoked concepts related to smoothness or roughness. However, the repetitive character (always smooth or jagged) might have greatly reduced the influence of the visual context on these "nonmeaningful" sounds.

Although the result of these experiments are most often interpreted as reflecting some form of conceptual priming between words or pictures and environmental sounds, they may also reflect linguistic mediated effects. For instance, looking at a picture of a cat and listening to the meowing of a cat may automatically activate the verbal label {cat}. This conceptual effect cannot be considered as purely nonlinguistic because there could be a semantic mediation between the (linguistic) label assigned to the drawing and the label assigned to the sound.

The purpose of the present study was to try to reduce, as much as possible, the chance that such labeling takes place. To this end, we generated, recorded, and, in some cases, also resynthesized sounds so that it was highly unlikely to identify a source (Ystad, Kronland-Martinet, Schön, & Besson, 2008). Thus, while people, when hearing a sound, may try to identify the source that produced it, our sounds should greatly reduce the likelihood of labeling compared to previous studies using environmental sounds.

We conducted two experiments. In Experiment 1, sounds were used as a context and were followed by visual words. In Experiment 2, visual words were used as a context and were followed by sounds. In Experiment 1, we predicted a larger N400 to words preceded by conceptually unrelated sounds compared to words preceded by related sounds. In Experiment 2, we predicted a larger N400 to sounds preceded by conceptually unrelated words compared to sounds preceded by related words.

EXPERIMENT 1

Methods

Participants

Sixteen nonmusician volunteers were tested in this experiment. All were right-handed, neurologically normal, had normal or corrected-to-normal vision, normal audition, and were native French speakers (age: $M = 27.5$ years, 7 women). All participants were paid for their participation to the experiment. Due to large drifts in EEG data, two participants were discarded from analyses.

Stimuli

Stimuli were built to favor what Pierre Schaeffer (1966) called "acousmatic listening" in his book *Traité des Objets Musicaux*. The term acousmatic relates to the ability of listening to a sound without considering the object(s) that created it, hence, reflecting the perceptual reality of a sound independently of the way it is produced or transmitted. By extension, sounds with no recognizable sources are "acousmatic sounds." These sounds are typically used as compositional resources in contemporary music such as "musique concrète" or electroacoustic music.

Stimuli included sounds originally intended for musical composition in electroacoustic music, as well as sounds specifically recorded for the experiment. Recordings aimed at decontextualizing the sounds to force listeners to pay attention to the sound itself. Some sounds were also obtained from traditional instruments, but their familiarity was altered by untraditional playing techniques or modified by signal processing techniques. To obtain a sound corpus representative of the main sound morphologies found in nature, we used the classification system proposed by Schaeffer (1966) and called "typology of sound objects," where sounds are mainly sorted as a function of their mass and shape. Schaeffer's typology of sound objects contains 35 classes of sounds, but only the nine main classes called "balanced sounds" (*sons équilibrés*) were used in this study. Two main aspects determine balanced sounds: maintenance (the way the energy is spread over time) and mass (linked to the spectral content of sounds and to the potential existence of pitch). Maintenance is used to distinguish sustained, iterative, and impulsive sounds. Mass distinguishes sounds with constant, varying, or indefinable pitch. The nine sound categories used here resulted from the combination of the three types of maintenances with the three types of masses (sustained with constant pitch, sustained with varying pitch, sustained with indefinable pitch; iterative with constant pitch, iterative with varying pitch, iterative with indefinable pitch; impulse with constant pitch, impulse with varying pitch, impulse with indefinable pitch).

We first selected 70 sounds representative of the nine categories of balanced sounds. Seven participants were then asked to listen to the sounds and to write down the first few words that came to mind. Although we did not measure the time participants needed to write the words evoked by each sound, this procedure lasted for almost 2 hr (i.e., more than one minute/sound in average). Participants were specifically asked to focus on the associations evoked by the sounds without trying to identify the physical sources that produced them. For instance, a particular sound evoked the following words: dry, wildness, peak,

winter, icy, polar, cold. Sounds that evoked identical or semantically close words for at least three of the seven participants were selected for the experiment resulting in a final set of 45 sound–word pairs. Each sound was paired with the proposed word of highest lexical frequency among the three words (e.g., "cold" was chosen between icy, polar, and cold). Finally, 45 unrelated pairs were built from this material by recombining words and sounds in a different manner. Average sound duration was 820 msec, standard deviation was 280 msec.

Procedure

Participants were comfortably seated in a Faraday box. Presentation of the sound was followed by the visual presentation of a word for 200 msec, with a stimulus onset asynchrony (SOA) of 800 msec (i.e., close to the average sound duration). Words were displayed in white lowercase on a dark background in the center of a 13-inches 88-Hz computer screen, set at about 70 cm from the participant's eyes. Participants were instructed to decide whether or not the sound and the target word fitted together by pressing one of two buttons. They were also told that the criterion for their relatedness judgment was of the domain of evocation rather than some direct relation such as the barking of a dog and the word "dog." It was also made clear that there were no correct or incorrect responses and participants were asked to respond as quickly as possible without much explicit thinking. A training session comprising 10 trials (with sounds and words different from those used in the experiment) was used to familiarize participants with the task.

Two seconds after word presentation, a series of "X" appeared on the screen signaling that participants could blink their eyes. A total of 45 related and 45 unrelated pairs were presented in pseudorandom order (no more than 5 successive repetitions of pairs belonging to the same experimental condition). Response side association (yes or no/left or right) was balanced across participants. A debriefing followed the experiment, questioning on possible strategies used by each participant (e.g., Do you have the feeling that you used a specific strategy? Do you have the feeling that the relation popped out from the stimuli or did you have to look for a relation? Did it happen that you gave a verbal label to sounds? Could you tell how sounds were generated? Did you try to find out?).

Data Acquisition and Analysis

Electroencephalogram (EEG) was recorded continuously at 512 Hz from 32 scalp electrodes (International 10–20 System sites) using a BioSemi Active Two system. Data were re-referenced off-line to the algebraic average of left and right mastoids. Trials containing ocular artifacts, movement artifacts, or amplifier saturation were excluded from the averaged ERP waveforms. Data were detrended and low-pass filtered at 40 Hz (12 dB/octave).

ERP data were analyzed by computing the mean amplitude, starting 100 msec before the onset of word presentation and ending 1000 msec after. Because there were no a priori correct responses, averages for related and unrelated pairs were based on the participants' responses. Repeated measures analyses of variance (ANOVAs) were used for statistical assessment of the independent variable (relatedness) To test the distribution of the effects, six regions of interest (ROIs) were selected as levels of two topographic within-subject factors (i.e., anteroposterior and hemisphere): left (AF3, F3, F7) and right (AF4, F4, F8) frontal; left (FC1, C3, CP1) and right (FC2, C4, CP2) central; and left (P3, PO3, P7) and right (P4, PO4, P8) parietal. Data were analyzed using latency windows of 50 msec in the 0 to 1000 msec range. Only results that were statistically significant in at least two successive 50-msec windows are reported. All p values were adjusted with the Greenhouse–Geisser correction for nonsphericity, when appropriate. Dunn–Sidak test was used in correcting post hoc multiple comparisons. The statistical analyses were conducted with Cleave (www.ebire.org/hcnlab) and Matlab.

Results

Behavioral Data

Even if the experimental conditions were defined by the participants' responses, we considered behavioral accuracies as the matching degree between the participants' responses and the related and unrelated pairs based upon the material selection procedure.

This analysis indicated an average accuracy of 77% that is significantly above the 50% chance level ($\chi^2 = 516$, $df = 25$, $p < .001$). No significant differences were found on RTs (mean $\pm$ SD = 1018 $\pm$ 170 and 1029 $\pm$ 220 msec, respectively, Wilcoxon matched pair test, $p = .97$).

Event-related Brain Potentials Data

As can be seen in Figure 1, visual word presentation elicited typical ERP components. An N1 component, peaking around 100 msec after stimulus onset, is followed by a P2 peaking around 150 msec. No differences are visible over these components for the related and unrelated targets. However, around 300 msec after stimulus onset, ERPs to related and unrelated words start to diverge with unrelated words associated to a larger negativity in the 300–700 msec latency window. This effect is maximal over central electrodes.

To analyze in detail how ERP components were modulated by the independent variables manipulated in this experiment, we first computed repeated measure ANOVAs with Relatedness (related/unrelated) $\times$ Anteroposterior (frontal, central, and parietal ROIs) $\times$ Hemisphere (left/right) as within factors. An interaction between relatedness, anteroposterior, and hemisphere factors was significant in

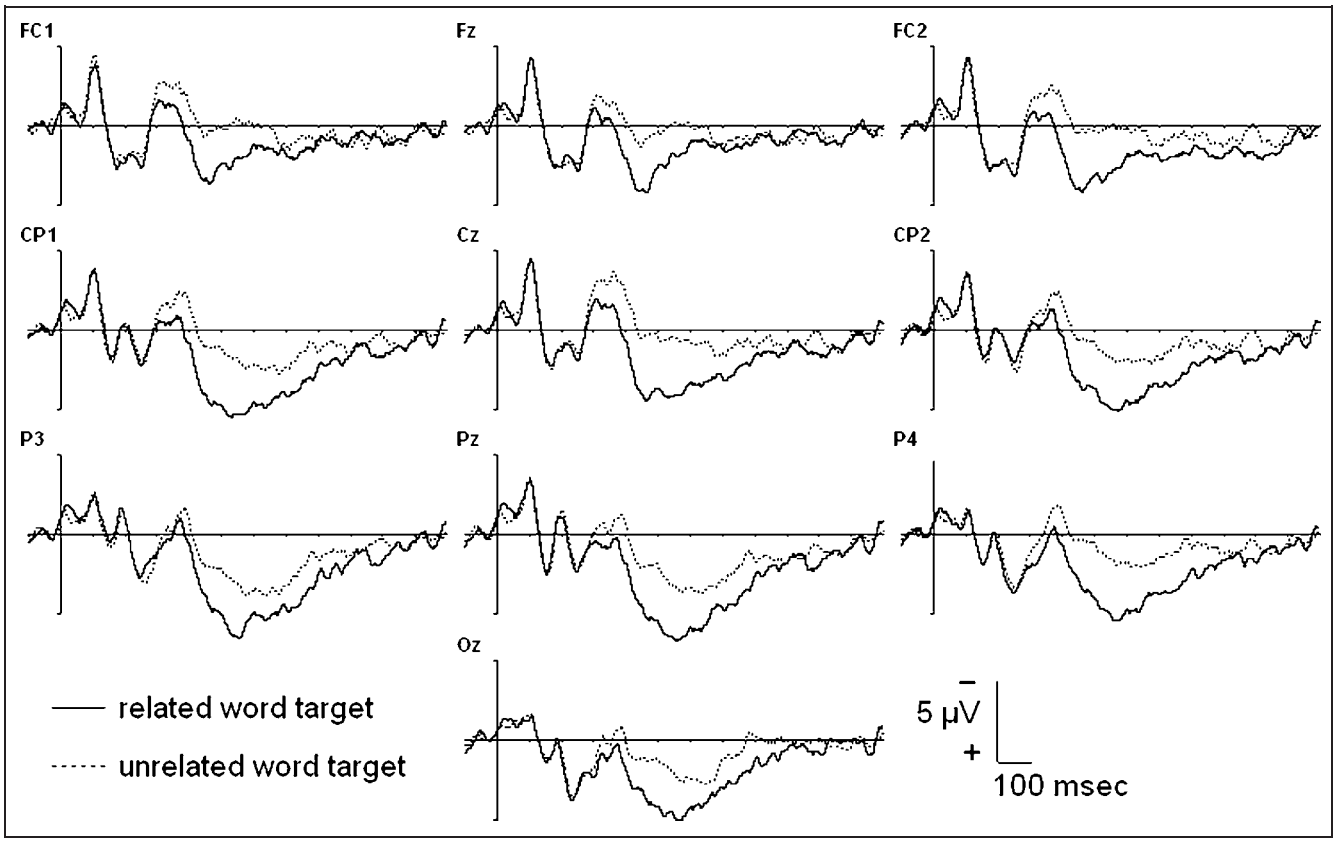

Figure 1. Grand-averaged ERPs to related and unrelated target words according the participants' responses (14 participants). Stimulus onset is the vertical calibration bar.

the 250–350 msec latency range [$F(2, 24) = 5.3, p < .05$]. Post hoc comparisons revealed a larger negativity to unrelated compared to related words over the left frontal region ($p < .0001$, effect size = 1.3 μV). In the 350–450 msec latency range, although the triple interaction was no longer significant, there was a significant Relatedness × Anteroposterior interaction [$F(2, 24) = 7.1, p < .01$] due to a larger effect over frontal and central regions compared to parietal regions, for both hemispheres (frontal: $p < .001$, effect size = 1.9 μV; central: $p = .001$, effect size = 2.0 μV; parietal: $p > .05$). Finally, in the 450–600 msec latency range, the relatedness effect was equally distributed over all ROIs and hemispheres [main effect of relatedness: $F(1, 12) = 6.83, p < .05$, effect size = 1.9 μV].

Discussion

Although, as stated above, there are no correct and incorrect responses in this experiment, the 77% participants' accuracy can be interpreted as a sign of a rather low intersubject variability between the experimental group and the group used in the pilot study. Moreover, this also shows that participants well understood the task, namely, to use the evocative features of a sound to judge its relation to a word. The fact that we did not find any significant difference on RTs between related and nonrelated responses is not so surprising insofar as the design is not a typical priming paradigm in that the task is more a value judgment than a categorization. This makes the task rather difficult, and RTs rather long. Indeed, priming experiments using lexical decision or categorical decision tasks report RTs between 500 and 800 msec (as compared to more than 1 sec here). Therefore, it might be the case that task difficulty, linked to the stimuli used in the present study, did override conceptual priming effects.

Electrophysiological results strongly resemble previous studies using target words preceded by semantically related or unrelated words (Chwilla, Brown, & Hagoort, 1995; Bentin et al., 1985), related or unrelated environmental sounds (Orgs et al., 2006, 2007; Van Petten & Rheinfelder, 1995), or musical excerpts (Daltrozzo & Schön, 2009; Koelsch et al., 2004). The similarity is mainly seen at the morphological level, with a negative component peaking between 300 and 400 msec. As in previous experiments, the amplitude of this N400-like component is modulated by the relatedness of the preceding sound. These results, together with the frontocentral distribution of the relatedness effect, will be further discussed in light of the results of Experiment 2.

EXPERIMENT 2

The same experimental design and stimuli as in Experiment 1 were used in Experiment 2, except that the order

of stimulus presentation was reversed with words presented as primes and sounds as targets.

Methods

Participants

In order to reduce stimuli repetition effects, we tested a new group of 18 nonmusician volunteers. All were right-handed, neurologically normal, had normal or corrected-to-normal vision, normal audition, and were native French speakers (age: $M = 26$ years, 8 women). All participants were paid for their participation in the experiment. Due to large drifts in the EEG data, five participants were discarded from analyses.

Stimuli

Same materials were used as in Experiment 1.

The procedure was identical to Experiment 1, except that words were used as primes and sounds as targets. This has a direct effect on the SOA. In Experiment 1, wherein sounds were used as primes, we used an 800-msec SOA because the average sound duration was 820 msec and the end of the sound is generally not very informative due to natural damping. In Experiment 2, visual target words are used as primes. The use of 800 msec SOA would be too long as words do not require 800 msec to be read and also by comparison to typical durations used in priming experiments with visual words. Thus, a 500-msec SOA was used in Experiment 2.

Procedure

Data acquisition and analysis. Same as in Experiment 1.

Results

Behavioral Data

Participants showed an average accuracy of 78% that is significantly above the 50% chance level ($\chi^2 = 681$, $df = 35$, $p < .001$). These results were not significantly different from those of Experiment 1 for both unrelated and related trials (Mann–Whitney U test: $p = .70$ and $p = .67$, respectively). No significant relatedness effect was found on RTs (1320 ± 230 msec and 1295 ± 217 msec, respectively, Wilcoxon matched pair test, $p = .3$). However, RTs were slower than in Experiment 1, for both related and unrelated responses (Mann–Whitney U test: $p = .001$).

Event-related Brain Potentials Data

As can be seen in Figure 2, sound presentation elicited an N1 component, peaking around 130 msec after stimulus onset, followed by a P2 peaking around 220 msec. No

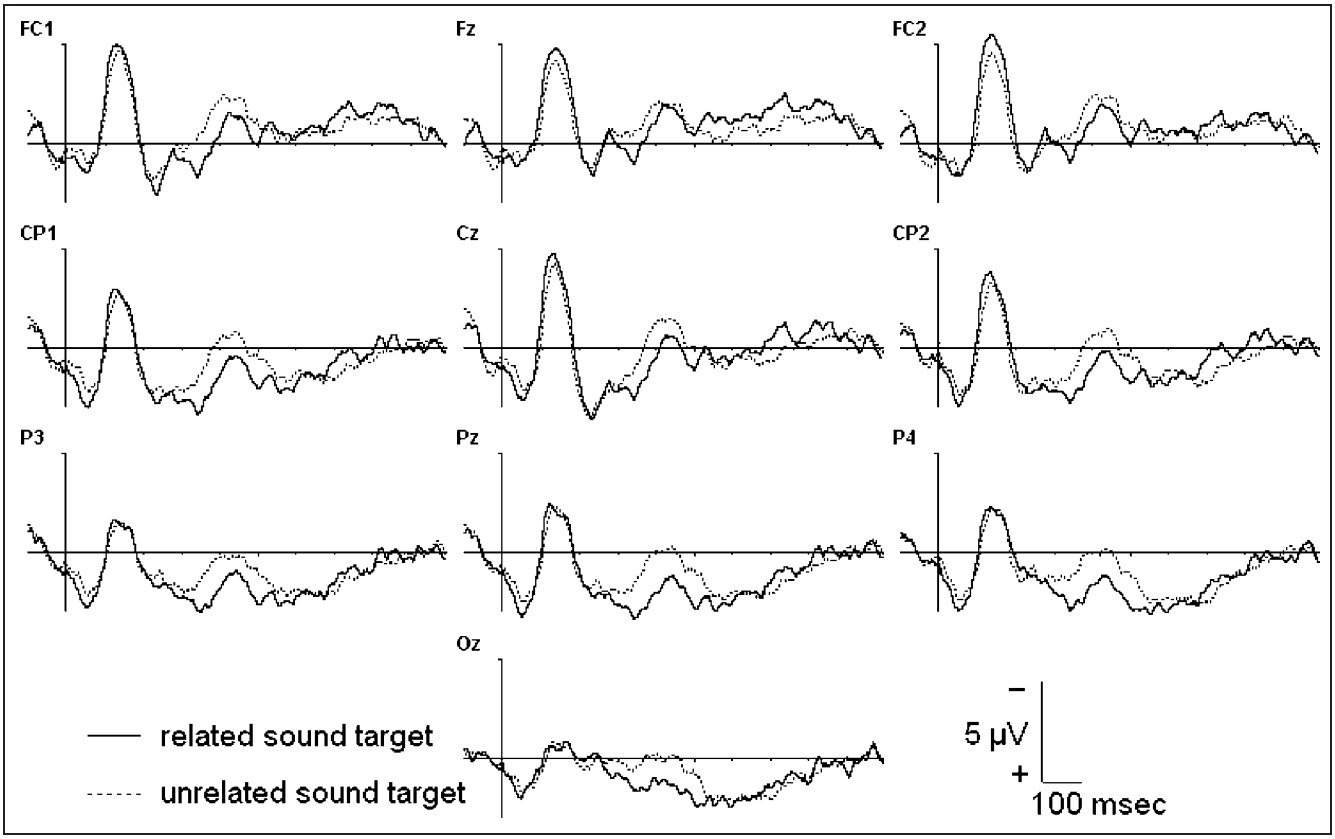

Figure 2. Grand-averaged ERPs to related and unrelated target sound according the participants' responses (13 participants). Stimulus onset is the vertical calibration bar.

differences are visible over these components for related and unrelated targets. However, around 300 msec after stimulus onset, ERPs to related and unrelated sounds start to diverge: Compared to related sounds, ERPs to unrelated sounds elicit a larger negativity in the 300–500 msec latency window. This effect is maximal over parietal electrodes.

To analyze in detail how these components were modulated by the independent variables manipulated in this experiment, we computed repeated ANOVAs with Relatedness (related/unrelated) × Anteroposterior (frontal, central, and parietal ROIs) × Hemisphere (left/right) as within factors using 50-msec latency windows.

The main effect of relatedness was significant in the 300–400 msec latency range [$F(1, 13) = 6.95$, $p < .05$; effect size = 1.2 μV]. In the 400–600 msec latency range, there was a significant Relatedness × Anteroposterior interaction [$F(2, 26) = 6.03$, $p < .05$] due to a larger relatedness effect over central and parietal regions compared to frontal regions (400–500 msec: frontal, $p > .7$; central, $p < .01$, effect size = 1 μV; parietal, $p < .001$, effect size = 1.1 μV; 500–600 msec: frontal and central, $p > .1$; parietal, $p < .05$, effect size = 1.0 μV).

In order to compare the relatedness effect found in the two experiments, we computed a four-way ANOVA including the same within-subject factors and target modality (word/sound) as between-subjects factor. Results showed a significant triple interaction of target modality, relatedness, and anteroposterior factors between 250 and 600 msec [$F(2, 50) = 5.2$, $p < .05$]. Post hoc analyses showed that this interaction was due to a lack of relatedness effect over frontal regions when sounds were used as targets (see Figure 3).

Discussion

Participants' accuracy was similar to what found in Experiment 1, which again shows that participants were sensitive to the conceptual word–sound relation. Moreover, the similarity of results in both experiments also points to an equivalent level of task difficulty. However, RTs were longer in Experiment 2 than in Experiment 1. We interpret this difference in terms of stimulus modality and familiarity. Although in Experiment 1 the target word was presented in the visual modality and lasted for 200 msec, in Experiment 2 the target sound average duration was around 800 msec. Therefore, although for words all information necessary to make a relatedness decision was available within 200 msec for sounds, information only gradually becomes available over time and the relatedness decision cannot be made as fast as for words. Moreover, whereas target words (Experiment 1) were highly familiar, target sounds (Experiment 2) were unfamiliar, as they were specifically created for the purpose of the experiments in order to minimize linguistic mediation. Thus, both modality and familiarity may account for longer RTs to sound than to word targets.

Electrophysiological data strongly resemble previous studies using target sounds preceded by semantically related or unrelated words (Orgs et al., 2006; Van Petten & Rheinfelder, 1995). The amplitude of a negative component, peaking between 300 and 400 msec, was modulated

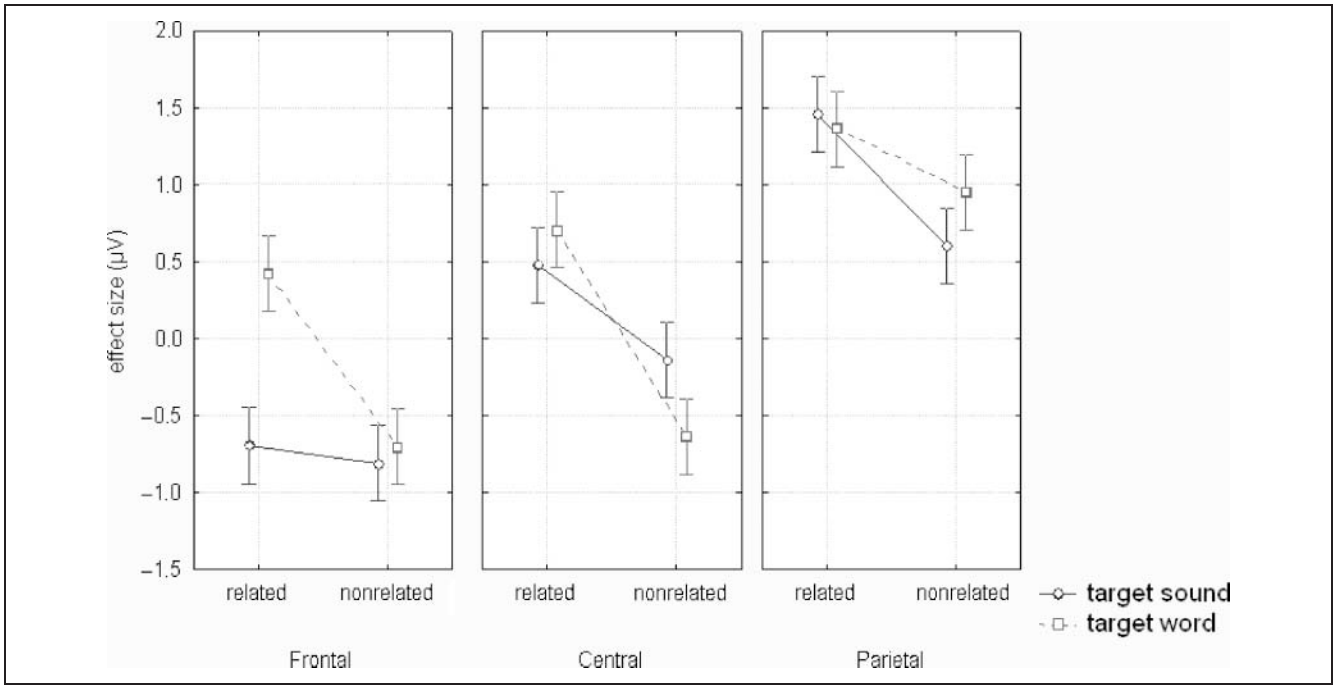

Figure 3. Interaction of Target modality × Relatedness × ROI (Anteroposterior factor). The relatedness effect is almost absent at frontal sites for sound targets. Vertical bars denote 95% confidence interval.

by its relatedness to the preceding sound. However, the centro-parietal distribution of this effect differed from the fronto-central distribution of the effect observed in Experiment 1. These differences are discussed in the General Discussion in view of the literature.

Finally, although N400-like components have been reported in most studies using environmental sounds, Cummings et al. (2006) failed to report such an effect in what they called the "nonmeaningful sound condition." In their study, the authors compared ERPs evoked during the processing of words, environmental sounds, and nonmeaningful sounds in semantically matching or mismatching visual contexts. However, there are two major differences between our acousmatic sound stimuli and the one used in the nonmeaningful sound condition by Cummings et al. that may explain why we found an effect while they did not. First, their criterion in choosing sounds was very different from ours. They chose sounds from an internet database (www.sounddog.com) in order "to portray either a smooth sound (e.g., a harmonic tone), or a jagged sound (e.g., a cracking sound)." Therefore, they used a binary categorical criterion, which most probably generated a rather repetitive sound corpus. By contrast, we chose sounds that could be used in electroacoustic music composition and we tried to maximize sound variability. Second, the context was not linguistic as in our study but comprised abstract visual patterns: "colorful, non-object-looking patterns chosen to represent one of two categories," smooth or jagged. Therefore, the conceptual relation between context and sound target was very different in the two studies: each sound matching one out of two categories (smooth/jagged) versus our choice of looking for an "optimal" verbal descriptor for each given sound.

These differences in sound selection, context, and context–sound relations may possibly explain the different results. Thus, the lack of relatedness effect for the nonmeaningful sound condition in Cummings et al.'s study could be due to a weaker conceptual processing of their sound database compared to our more "musical" sounds, a weaker conceptual processing of the abstract images compared to words used in our experiment and a weaker relation between context and sounds. Unfortunately, not enough details concerning the nonmeaningful sound condition are given in the manuscript in order to clearly understand the reasons of the different results.

GENERAL DISCUSSION

The general aim of the present experiments was to compare conceptual priming effects when either words or sounds were used as targets. The originality of our approach was to create sounds whose sources were, in most cases, impossible to identify (i.e., "acousmatic sounds") in order to reduce the influence of linguistic mediation. In both experiments, behavioral data showed that participants were able to evaluate the sound–word or word–sound relations with relative low intersubject variability and good consistency. No relatedness effect was found on RTs. However, electrophysiological data revealed an enhanced negativity in the 250–600 msec latency range to unrelated compared to related targets in both experiments, although with a more fronto-central distribution to word targets in Experiment 1 and more centro-parietal distribution to sound targets in Experiment 2. These findings are discussed relative to the linguistic versus amodal theories of concepts.

To Label or Not to Label

The reason for using nonverbal stimuli in previous studies was to determine whether behavioral effects such as priming effects and electrophysiological effects such as the N400 effect are specific to language or not. This is not a trivial question to answer because a recurrent question in the behavioral and electrophysiological literature on conceptual priming with nonverbal stimuli is whether nonverbal items are given a verbal label or not (Koelsch et al., 2004). Indeed, if labeling takes place, then behavioral or electrophysiological differences between related and unrelated items may simply reflect a linguistic relatedness effect. Such results would therefore support a linguistic theory of concepts. By contrast, if the effects are found independently of language mediation, such results would support a more general and amodal theory of concepts.

In this respect, the relevant aspect of our study is the low probability that labeling takes place. Indeed, although labeling the picture or line drawing of a cat (Holcomb & McPherson, 1994), the odor of a lemon (Sarfarazi et al., 1999), the barking of a dog (Van Petten & Rheinfelder, 1995), or the ringing of a telephone (Orgs et al., 2006, 2007) is easy and rather automatic (i.e., we cannot avoid labeling), the stimuli used in the present experiment are rather difficult to label because they are uncommon sounds and it is difficult to identify the source that produced them (sound examples are available at www.sensons.cnrs-mrs.fr/Schaeffer/Schaeffer.html). Not surprisingly, in the pilot study, when participants were asked to find a related verbal label for each sound (see Stimuli section), they asked to listen to the sound several times and they needed 1 min rather than 1 sec to find an appropriate label. Of course, this does not completely prevent the possibility that participants of the two experiments still imagine and label a possible source of sounds (although incorrect), or attach verbal associations evoked by the sounds, for instance, by using adjectives to describe or characterize them. The argument is not that verbal labeling is completely prevented (which is probably impossible to do), but that it was strongly reduced in our experiments compared to previous experiments using, for instance, environmental sounds.

Another related issue is that the strength of the conceptual relation between sounds and words is probably weaker in our experiments compared to studies using environmental sounds. Indeed, although a barking sound

immediately and strongly calls for a precise label {dog}, the sounds we used rather evoked a larger set of concepts and feelings, but in a weaker manner (see Figure 4). This might explain why the relatedness effect sizes in our experiment are smaller than those found in some studies using environmental sounds (Orgs et al., 2006, 2007; Cummings et al., 2006), although they do not seem to greatly differ from other studies (Plante, Petten, & Senkfor, 2000; Van Petten & Rheinfelder, 1995).

Differences in scalp topography between Experiments 1 and 2 and, more generally, between the several experiments that used environmental sounds can be taken to argue that the N400 effect encompasses different processes. Indeed, it may be influenced by both the high-level cognitive processing of the conceptual relation between two stimuli (that may be similar in all experiments) and the lower-level perceptual processes linked with the specific acoustic features of the sounds (that would be different across experiments). In this respect, it is interesting to note that the scalp distribution of the relatedness effect associated to unrelated and related targets changes dynamically over time, which reflects the spatio-temporal dynamics and potential interactions in the underlying processes (Experiment 1: left frontal in the early latency band, fronto-central in the 350–450 msec range, and spread across scalp sites in the 450–600 msec range; Experiment 2: widely distributed across scalp sites in the 300–400 msec range and spatially more localized over centro-parietal sites in the 400–600 msec range).

Finally, it is also important to note that all negativities in the 300–600 msec latency band are not necessarily N400 components. For instance, what Orgs et al. (2006, 2007) consider as an early onset of the N400 for sounds (between 200 and 300 msec) may also be an N200 component reflecting a physical mismatch between the sound expected on the basis of the context and the sound actually presented or an offset potential trigged by a constant sound offsets (all sounds had 300 msec duration). In conclusion, one way to reconcile these different results is to consider that the cognitive processing of the conceptual relation, as reflected by the amplitude modulation of the N400, is present in all the abovementioned experiments, but that other perceptive processes, which differ depending upon the specific characteristics of the stimuli, are also involved and influence the scalp distribution of the N400 effect.

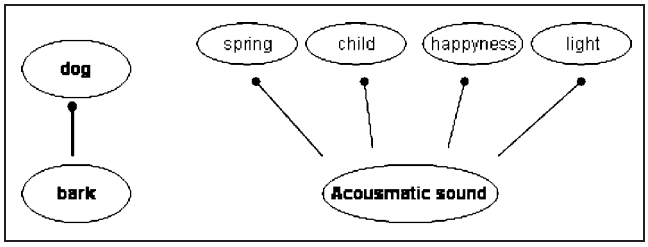

Figure 4. Different strengths between an environmental sound and a concept, and an acousmatic sound and the related concepts.

Modeling the Semiotics of Sounds

It might be the case that, after signal analysis, taking place in the brainstem and in the primary auditory regions, sound representations automatically spread to a whole range of concepts. With this respect, an attractive view is that of distributed network models (Anderson, 1993; McClelland, Rumelhart, & the PDP Research Group, 1986). According to these models, concepts are represented as patterns of activation of an interconnected set of units. Similar concepts share similar patterns of activation. What is interesting in these models is the fact that the units can be thought of as representing aspects of a given object (e.g., sound, word). Most importantly, these aspects "need not be nameable or correspond in any obvious way to the features people might list in a description of the entity" (McNamara, 2005, p. 29). Within such a theoretical framework, sound features, such as attack time, spectral centroid, spectral variation, energy modulation, inharmonicity, and others, might become input units shared by several patterns of activation for a set of concepts. This means that a combination of sound features, so-called invariants, might be used by the listeners to determine specific aspects of sounds. In Figure 5, we propose a model that may explain how conceptual priming takes place in our studies. Depending on whether we read it from left to right or from right to left, the model explains the effects of Experiment 1 or Experiment 2, respectively. We will quickly go through it beginning from the left, that is, for Experiment 1, wherein the probe is a sound and the target is a word. Once a sound is presented, acoustic features are extracted and represented in a sparse manner at the acoustic feature level. In the case of an environmental sound, these feature representations may feed forward to activate a precise item in the sound lexicon and possibly find a good match. If no good match is found in the lexicon, as in the case of the present experiment using acousmatic sounds, competition might be rather high, slowing down sound processing. In such cases, a direct path to an amodal concept representation level may take over, possibly influenced by the emotional connotation of the sound, carried by the signal features (Juslin & Västfjäll, 2008; Koelsch & Siebel, 2005; Juslin & Laukka, 2003). This amodal representation of concepts would be the link between concepts evoked by sounds and concepts evoked by words. Indeed, activation of the concepts in the amodal concept lexicon would spread to the semantic level (and possibly to the lexical and letter level), therefore priming the semantic processing of a following presented word.

Of course, this is just a tentative model and several issues need clarification. First, the existence of a direct pathway from the feature extraction level to the amodal concept lexicon needs to be proved. Indeed, it might be the case that processing always transits via a sound lexicon. Second, although we believe that sound features can be automatically directed to an "emotional parser" without transiting via the sound lexicon, this needs to be

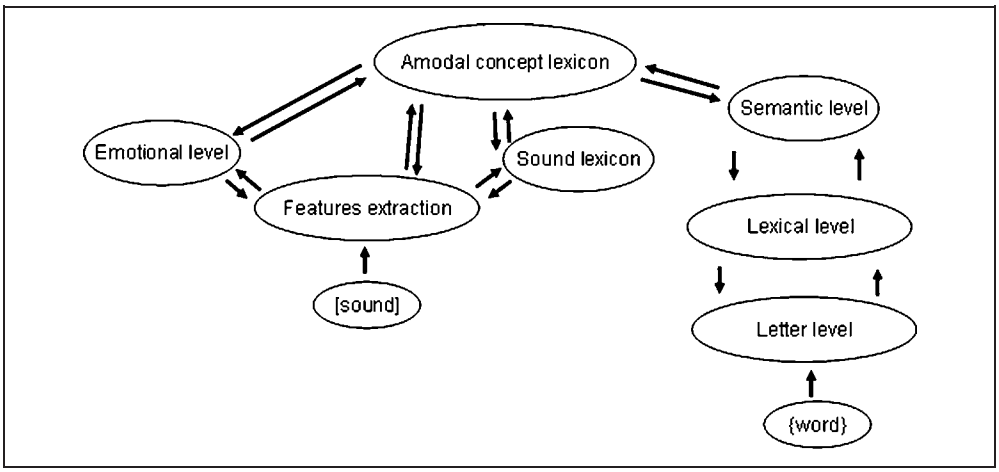

Figure 5. Tentative model describing how sounds can evoke concepts.

demonstrated. These two issues could possibly be addressed by studying patients with nonverbal auditory agnosia (Vignolo, 1982). These patients are particularly interesting because they do not have linguistic deficits, but they cannot anymore recognize environmental sounds. It is interesting to note that these patients can have difficulties in discriminating acoustically related sounds or semantically related sounds, often depending upon the lesion site (right or left, respectively; Schnider, Benson, Alexander, & Schnider-Klaus, 1994; Faglioni, Spinnler, & Vignolo, 1969). Unfortunately, testing procedures used in previous studies do not allow for qualitative error analysis, which would be most informative in order to understand whether these patients have a deficit at the sound lexicon level, at the amodal concept lexicon or both. Moreover, little is also known concerning whether these patients, experiencing difficulties in discriminating acoustically or semantically related sounds, can still attribute the correct emotion to these sounds. This could be an elegant way of showing that sound features can be processed by an "emotional parser" without transiting via the sound lexicon.

From Sound to Music

We previously said that the sound stimuli used in the present study are acousmatic sounds intended for "musique concrète." Of course, by no means would we claim that we studied music processing in this study, insofar as music goes well beyond a single sound. However, for a theoretical purpose, it is interesting to think about the relation between a single "acousmatic sound" and music.

Indeed, the fact that conceptual processing can take place for a single sound, independently of its source, is also of interest for the understanding of the meaning of music. Surprisingly, although timbre variations are consciously used by composers and by musicians during performance (Barthet, Kronland-Martinet, & Ystad, 2008), the sound structure or "sound matter" (having a quasi-physical connotation) is marginal or not considered at all in the taxonomy of musical meanings (see Patel, 2008). The musically meaningful elementary unit is, most of the time, considered to be a set of sounds composing a motif, a sentence, a theme, and so on. Of course, the way sounds combine in music is of utmost importance and, indeed, most theories on the meaning of music focus on the relation between musical events (e.g., Jackendoff, 1991; Meyer, 1956, see also Frey et al., 2009, for experimental evidence). However, if a single sound, out of a musical context, can generate meaning, we should question the possibility that, in music, elementary units, much shorter than motifs or themes, may also convey part of the musical meaning, via the property of the "sound matter" they carry at each single lapse of time. With respect to this hypothesis, and extending the work of Koelsch et al. (2004), we recently used a similar design to show that 1 sec of music can communicate concepts and influence the processing of a following target word (Daltrozzo & Schön, 2009). Most importantly, we also showed that when music is preceded by a verbal context, the amplitude of an N400 component to music is modulated by the degree of conceptual relation between the context and the musical excerpt, as soon as 300 msec after music onset. The fact that concepts carried by words can influence the processing of a following musical excerpt can be interpreted as a strong sign that the time window of elementary meaningful units in music might be very small, well below the time window of a motif or a theme. Therefore, the model we propose here for conceptual processing of sounds might also be at work in music listening. The meaning of music will, therefore, be the result of a rather complex process, taking into account the structural properties of music, the personal and cultural background of the listener, the aesthetic and emotional experience, and also the structure or matter of the sounds whereof a given excerpt is composed.

Acknowledgments

This project has been supported by the French National Research Agency (ANR 05-BLAN-0214 "Music and Memory" to D. Schön; ANR, JC05-41996, "senSons," www.sensons.cnrs-mrs.fr/ to S. Ystad).

Reprint requests should be sent to Daniele Schön, INCM-CNRS, 31 Ch J Aiguier, Marseille 13402, France, or via e-mail: schon@incm.cnrs-mrs.fr.

REFERENCES

Anderson, J. A. (1993). The BSB Model: A simple nonlinear autoassociative neural network. In M. Hassoun (Ed.), *Associative neural memories* (pp. 77–103). New York, NY: Oxford University Press.

Barthet, M., Kronland-Martinet, R., & Ystad, S. (2008). Improving musical expressiveness by time-varying brightness shaping. In *Lecture notes in computer science* (Vol. 4969, pp. 313–337). Berlin: Springer-Verlag.

Bentin, S., McCarthy, G., & Wood, C. C. (1985). Event-related potentials, lexical decision and semantic priming. *Electroencephalography and Clinical Neurophysiology, 60,* 343–355.

Castle, P. C., Van Toller, S., & Milligan, G. J. (2000). The effect of odour priming on cortical EEG and visual ERP responses. *International Journal of Psychophysiology, 36,* 123–131.

Chwilla, D. J., Brown, P. M., & Hagoort, P. (1995). The N400 as a function of the level of processing. *Psychophysiology, 32,* 274–285.

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,* 1882–1892.

Faglioni, P., Spinnler, H., & Vignolo, L. A. (1969). Contrasting behavior of right and left hemisphere-damaged patients on a discriminative and a semantic task of auditory recognition. *Cortex, 5,* 366–389.

Frey, A., Marie, C., Prod'Homme, L., Timsit-Berthier, M., Schön, D., & Besson, M. (2009). Temporal semiotic units as minimal meaningful units in music? An electrophysiological approach. *Music Perception, 26,* 247–256.

Holcomb, P. J., & McPherson, W. B. (1994). Event-related brain potentials reflect semantic priming in an object decision task. *Brain and Cognition, 24,* 259–276.

Jackendoff, R. (1991). Musical parsing and musical affect. *Music Perception, 9,* 199–230.

Juslin, P. N., & Laukka, P. (2003). Emotional expression in speech and music: Evidence of cross-modal similarities. *Annals of the New York Academy of Sciences, 1000,* 279–282.

Juslin, P. N., & Västfjäll, D. (2008). All musical emotions are not created equal: The cost of neglecting underlying mechanisms. *Behavioral and Brain Sciences, 31,* 559–575.

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,* 302–307.

Koelsch, S., & Siebel, W. (2005). Towards a neural basis of music perception. *Trends in Cognitive Sciences, 9,* 578–584.

Kutas, M., & Hillyard, S. A. (1980). Reading senseless sentences: Brain potentials reflect semantic incongruity. *Science, 204,* 203–205.

McClelland, J. L., Rumelhart, D. E., & the PDP Research Group. (1986). *Parallel distributed processing: Explorations in the microstructure of cognition* (Vol. II). Cambridge, MA: MIT Press.

McNamara, T. P. (2005). *Semantic priming: Perspectives from memory and word recognition*. New York: Psychology Press.

Meyer, L. (1956). *Emotion and meaning in music*. Chicago: University of Chicago Press.

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. H., & Heil, M. (2007). Is conceptual priming for environmental sounds obligatory? *International Journal of Psychophysiology, 65,* 162–166.

Patel, A. (2008). *Music, language, and the brain*. New York: Oxford University Press.

Plante, E., Petten, C. V., & Senkfor, A. J. (2000). Electrophysiological dissociation between verbal and nonverbal semantic processing in learning disabled adults. *Neuropsychologia, 38,* 1669–1684.

Sarfarazi, M., Cave, B., Richardson, A., Behan, J., & Sedgwick, E. M. (1999). Visual event related potentials modulated by contextually relevant and irrelevant olfactory primes. *Chemical Senses, 24,* 145–154.

Schaeffer, P. (1966). *Traité des Objets Musicaux*. Paris: Editions du Seuil.

Schnider, A., Benson, D. F., Alexander, D. N., & Schnider-Klaus, A. (1994). Non-verbal environmental sound recognition after unilateral hemispheric stroke. *Brain, 117,* 281–287.

Van Petten, C., & Rheinfelder, H. (1995). Conceptual relations between spoken words and environmental sounds: Event-related brain potential measures. *Neuropsychologia, 33,* 485–508.

Vignolo, L. A. (1982). Auditory agnosia. *Philosophical Transactions of the Royal Society of London, Series B, Biological Sciences, 298,* 49–57.

Ystad, S., Kronland-Martinet, R., Schön, D., & Besson, M. (2008). Vers une approche acoustique et cognitive de la sémiotique des objets sonores. In E. Rix & M. Formosa (Eds.), *Vers une sémiotique générale du temps dans les arts* (pp. 73–83). Paris: Ircam-Centre Pompidou/Delatour.