



## Hearing living symbols and nonliving icons: Category specificities in the cognitive processing of environmental sounds

Bruno L. Giordano\*, John McDonnell, Stephen McAdams

Centre for Interdisciplinary Research in Music, Media and Technology, Schulich School of Music, McGill University, Montréal, Québec, Canada

### ARTICLE INFO

#### Article history:

Accepted 20 January 2010

Available online 25 February 2010

#### Keywords:

Environmental sounds  
Sound source perception  
Language  
Symbolism  
Hierarchical sorting

### ABSTRACT

The neurocognitive processing of environmental sounds and linguistic stimuli shares common semantic resources and can lead to the activation of motor programs for the generation of the passively heard sound or speech. We investigated the extent to which the cognition of environmental sounds, like that of language, relies on symbolic mental representations independent of the acoustic input. In a hierarchical sorting task, we found that evaluation of nonliving sounds is consistently biased toward a focus on acoustical information. However, the evaluation of living sounds focuses spontaneously on sound-independent semantic information, but can rely on acoustical information after exposure to a context consisting of nonliving sounds. We interpret these results as support for a robust iconic processing strategy for nonliving sounds and a flexible symbolic processing strategy for living sounds.

© 2010 Elsevier Inc. All rights reserved.

### 1. Introduction

Environmental sounds (ES from now on) have meaning by virtue of their connection with the events and objects that generated them (e.g., the sound of footsteps, Ballas & Howard, 1987). Research on human processing of ES has frequently revealed similarities with the processing of linguistic materials from a variety of perspectives: psychophysical (e.g., Ballas & Mullins, 1991; Gygi, Kidd, & Watson, 2004), neuropsychological (e.g., Saygin, Dick, Wilson, Dronkers, & Bates, 2003), and psychobiological (e.g., Noppeney, Josephs, Hocking, Price, & Friston, 2008; Thierry, Giraud, & Price, 2003). These similarities have recently been studied for a specific category of sounds: those generated by human actions (e.g., Galati et al., 2008; Pizzamiglio et al., 2005). Accordingly, the processing of both linguistic materials and action sounds includes an activation of the motor programs that could generate the sound, i.e., involves sensorimotor transformations (e.g., Lewis, 2006; Pizzamiglio et al., 2005; Pulvermüller, 2005; Wilson, Sereno, & Iacoboni, 2004). In the current study, we adopted a behavioral approach to quantify category specific differences in the similarity between the linguistic and nonlinguistic auditory domains. To do so, we focused on another fundamental property of language processing: the activation of mental contents arbitrarily connected with the sensory input, i.e., symbolism (cf., Bickerton, 2005; Deacon, 1997).

A basic difference between words and ES stands in their relationship to the referent (Ballas & Howard, 1987; Cummings et al., 2006; Dick et al., 2007; Saygin, Dick, & Bates, 2005). The acoustics of ES is determined by the mechanical properties of the sound source (e.g., sounds generated by small objects tend to have a higher pitch, Fletcher & Rossing, 1991). Because of this deterministic link, ES bear a similarity with the objects and actions at the sound source. As such, they tend to be icons of their referent (cf. Peirce, Houser, & Kloesel, 1998, p. 13). On the contrary, the word-referent relation is often arbitrary. A word is thus a symbol for its referent because the sensory input bears no resemblance to the mental representations of the message recipient (cf. Peirce et al., 1998, p. 9). Despite this basic difference, empirical investigations have frequently revealed an overlap in the neurocognitive resources dedicated to meaningful verbal and nonverbal auditory materials. A first piece of evidence in support of this notion is related to hemispheric asymmetries. ES are known to activate a large bilateral network of cortical areas, comprising the primary auditory, nonprimary temporal, and frontal cortices (e.g., Dick et al., 2007; Lewis, Phinney, Brefczynski-Lewis, & DeYoe, 2006). Consistently, right- and left-hemispheric lesions equally disrupt performance in a task matching pictures to sounds (Schnider, Benson, Alexander, & Schnider-Klaus, 1994). Nonetheless, right- and left-hemisphere lesions increase the rate of errors based on acoustical and semantic similarity, respectively (e.g., a crying baby sound is paired with the picture of a cat or with that of a laughing baby after right- and left-hemispheric damage, respectively, Faglioni, Spinnler, & Vignolo, 1969; Schnider et al., 1994; Spinnler & Vignolo, 1966; Vignolo, 1982). Further, unrecognizable and recognizable sounds trigger stronger left- and right-lateralized activations, respectively (e.g., Lebrun et al., 2001; Thierry & Price, 2006).

\* Corresponding author. Address: Schulich School of Music, McGill University, 555 Sherbrooke Street West, Montréal, QC, Canada H3A 1E3. Fax: +1 514 398 2962.  
E-mail address: [bruno.giordano@music.mcgill.ca](mailto:bruno.giordano@music.mcgill.ca) (B.L. Giordano).

Both brain imaging and neuropsychological studies thus appear to support a preferential involvement of the right and left (language-specific) hemispheres in the sensory and semantic analyses of ES, respectively.

Further ES/language similarities emerge from the activations of post-primary cortical areas (Adams & Janata, 2002; Bergerbest, Ghahremani, & Gabrieli, 2004; Dick et al., 2007; Engelen et al., 2006; Fecteau, Armony, Joannette, & Belin, 2004; Galati et al., 2008; Giraud & Price, 2001; Hocking & Price, 2008; Humphries, Willard, Buchsbaum, & Hickok, 2001; Kiefer, Sim, Herrnberger, Grothe, & Hoening, 2008; Lebrun et al., 2001; Lenz, Schadow, Thaerig, Busch, & Herrmann, 2007; Lewis, Brefczynski, Phinney, Jannik, & DeYoe, 2005; Lewis et al., 2004; Maeder et al., 2001; Murray, Camen, Gonzalez Andino, Bovet, & Clarke, 2006; Murray, Camen, Spierer, & Clarke, 2008; Noppeney et al., 2008; Specht & Reul, 2003; Thierry & Price, 2006; Thierry et al., 2003). These include the bilateral middle and superior temporal cortex, particularly the left posterior regions of the superior and middle temporal gyri (LpSTG and LpMTG, respectively, e.g., Dick et al., 2007; Humphries et al., 2001; Kiefer et al., 2008; Lewis et al., 2004; Murray et al., 2006; Specht & Reul, 2003; Thierry et al., 2003), and the left inferior frontal gyrus, LIFG (Humphries et al., 2001; Thierry et al., 2003; Dick et al., 2007; Giraud & Price, 2001; Lebrun et al., 2001; Murray et al., 2006; Lewis et al., 2004, 2005; Adams & Janata, 2002; Engelen et al., 2006; Thierry & Price, 2006; Maeder et al., 2001). Notably, these regions are activated by both ES and words (Humphries et al., 2001; Kiefer et al., 2008; Lebrun et al., 2001; Specht & Reul, 2003; Thierry et al., 2003) and are implicated in the comprehension of words and sentences (see Hasson and Small, 2008, for a review). These similarities likely originate in part from the activation in the LMTG of overlapping circuits for the retrieval of semantic knowledge (e.g., Binder & Price, 2001; Cabeza & Nyberg, 2000). Significant cross-priming effects are consistently observed between congruent ES and linguistic materials as measured by decreased reaction times and amplitudes of the semantic N400 component (Galati et al., 2008; Noppeney et al., 2008; Orgs, Lange, Dombrowski, & Heil, 2006; Orgs, Lange, Dombrowski, & Heil, 2007; Pizzamiglio et al., 2005; van Petten & Rieffers, 1995; although see Chiu, 2000; Chiu & Schacter, 1995; Stuart & Jones, 1995, for negative findings). Also consistent with the hypothesis of partially overlapped semantic networks is the study of left-hemispheric aphasics by Saygin et al. (2003). With these patients, impairments in the comprehension of language and ES were strongly correlated with LpMTG damage leading to comparable linguistic and nonlinguistic impairments and with LpSTG damage leading to a higher nonlinguistic impairment. A potential secondary source of ES/language similarity involves left posterior temporal regions in the integration of multisensory information (e.g., Beauchamp, Lee, Argall, & Martin, 2004; Lewis, 2006; Schneider, Debener, Oostenveld, & Engel, 2008; Senkowski, Saint-Amour, Kelly, & Foxe, 2007). Indeed, the multisensory convergence zone for ES partially overlaps the superior temporal sulcus (STS), an area implicated in audiovisual integration of speech (Calvert & Lewis, 2004). Finally, the LIFG is also involved in the semantic processing of verbal materials, i.e., of sentences (e.g., Humphries et al., 2001) and, more robustly, of isolated words (Hasson & Small, 2008). Differently from the LMTG, the LIFG has been linked not to semantic retrieval per se, but to the selection of semantic information from among competing alternatives (Moss et al., 2005; Thompson-Schill, 2005; Thompson-Schill, D'Esposito, Aguirre, & Farah, 1997). In summary, empirical evidence points towards a partial overlap in the ES- and language-specific neural networks: a semantic retrieval system in the left posterior temporal cortex and left inferior frontal circuits linked with the selection of semantic information.

Finally, more detailed ES/language similarities emerge from the study of the cortical selectivity for specific ES categories (Belin, Zatorre, Lafaille, Ahad, & Pike, 2000; Belin, Zatorre, & Ahad, 2002; von

Kriegstein, Eger, Kleinschmidt, & Giraud, 2003; Fecteau et al., 2004; von Kriegstein & Giraud, 2004; Fecteau, Armony, Joannette, & Belin, 2005; Lewis et al., 2005, 2006; Tranel, Grabowski, Lyon, & Damasio, 2005; Kraut et al., 2006; Pizzamiglio et al., 2005; Murray et al., 2006; Murray et al., 2008; Altmann, Doehrmann, & Kaiser, 2007; Altmann et al., 2008; Doehrmann, Naumer, Volz, Kaiser, & Altmann, 2008; Galati et al., 2008).<sup>1</sup> Notably, category specificities emerge in both hemispheres (e.g., Belin, Fecteau, & Bédard, 2004). Here, we focus on specificities in the language-specialized left hemisphere. Overall, these studies reveal a caudal-to-rostral progression of category specific temporal activations moving from nonliving action/tool sounds to animal vocalizations to human vocalizations. Indeed, the LpMTG, the multisensory convergence region, and the left inferior temporal cortex (LpITC) are more strongly activated by hearing and naming action/tool sounds (Lewis et al., 2005; Lewis et al., 2006; Tranel et al., 2005; Doehrmann et al., 2008); the left middle STG (LmSTG), the anterior superior temporal sulcus and gyrus (LaSTS/LaSTG) and the LaITC are more strongly activated when hearing or naming animal vocalizations (Fecteau et al., 2004; Lewis et al., 2005; Lewis et al., 2006; Tranel et al., 2005; Doehrmann et al., 2008); the middle and anterior portions of the STG, STS and MTG are more strongly activated by nonlinguistic human vocal stimuli (Belin et al., 2000; Belin et al., 2002; von Kriegstein et al., 2003; Fecteau et al., 2004; von Kriegstein & Giraud, 2004). Focusing on vocalizations, the LmSTG is thought to analyze sound features prominent in this class of ES, such as harmonic or phase-coupling content (Altmann et al., 2007; Lewis et al., 2005, see also Obleser, Eisner, & Kotz, 2008), and has also been associated with the phonemic analysis of speech (e.g., Liebenthal, Binder, Spitzer, Possing, & Medler, 2005). More importantly, vocalization-selective left anterior temporal regions, including the LaSTS, have also been associated with a variety of linguistic processes, from the extraction of phonemic representations to the analysis of syntax and the computation of the meaning of sentences, i.e., compositional semantics (e.g., Schlosser, Aoyagi, Fulbright, Gore, & McCarthy, 1998; Vandenberghe, Nobre, & Price, 2002; Narain et al., 2003; Humphries, Binder, Medler, & Liebenthal, 2006; Rogalsky & Hickok, 2008). Further category specific activations emerge for nonliving action/tool sounds in the left premotor cortex (Pizzamiglio et al., 2005; Murray et al., 2006) and specifically the ventral premotor cortex (LvPMC, Lewis et al., 2005; Lewis et al., 2006). Among the functions hypothesized for the vPMC are the comprehension of actions, the planning and preparation of arm and hand movements (Rizzolatti, Fogassi, & Gallese, 2002; Lewis, 2006) and, more importantly, the computation of direct sensorimotor transformations, which map sensory features to motor actions (Zatorre, Chen, & Penhune, 2007). Notably, sensorimotor transformations have also been documented in the linguistic domain (Wilson et al., 2004; Pulvermüller, 2005) and are considered necessary for language (e.g. Hauser, Chomsky, & Fitch, 2002). Finally, less clear category specificities emerge for another region involved in the processing of linguistic materials, the LIF cortex. Indeed, whereas Lewis et al. (2005), Lewis et al. (2006), Murray et al. (2006), Kaplan and Iacoboni (2007) and Galati et al. (2008) report stronger activations for action/tool sounds, Doehrmann et al. (2008) and Fecteau et al. (2005) observed stronger activations for animal and human vocalizations, respectively. It is worth noting that stronger superior LIFG activations have been documented for the overt and covert naming of visual objects characterized by a low, as opposed to high, naming agreement (e.g., Kan & Thompson-Schill, 2004). As such, the across-studies variation in category specific LIF activations might at least in part reflect uncontrolled differences in the number of

<sup>1</sup> The reader interested in the literature on category specificities for the neural processing of ES in non-human primates is referred to Romanski and Averbeck (2009) for a review of the processing of vocalizations and to Kohler et al. (2002) and Cohen, Hauser, and Russ (2006), for two examples of studies on nonvocal sounds.

semantic representations activated by the particular stimuli at hand. In summary, nonliving action/tool sounds likely share with language part of the neural substrates for multisensory integration and, most importantly, the activation of sensorimotor transformations. By contrast, living human and animal vocalizations selectively activate neural substrates for the extraction of acoustical information and, most importantly, activate regions involved in the processing of compositional semantics and/or syntax.

Currently, the language/ES connection is emphasized only for the category of action/tool sounds. For example, the activation of sensorimotor transformations for this class of ES is argued to support the gestural hypothesis for the evolution of human communication and language (e.g., Lewis et al., 2005, p. 5156). Accordingly, language did not evolve from the vocal domain, but from communicative gestures (Hewes, 1973; Arbib & Rizzolatti, 1997; Arbib, 2005). Further, it has been argued that the frontal and premotor cortex are particularly well suited for the development of abstract representations (Kaplan & Iacoboni, 2005; Galati et al., 2008) and for the emergence of language (Aziz-Zadeh, Koski, Zaidel, Mazziot-ta, & Iacoboni, 2006), because they are activated by action events in a variety of modalities. Contrasting with this view, previous brain imaging studies appear to show that additional classes of ES share different neurocognitive processes with language. For example, living vocal sounds appear to activate temporal regions also implied in the analysis of syntax and compositional semantics. For a variety of reasons, however the evidence necessary to reconcile these inconsistencies is lacking. Firstly, in several cases neuroimaging investigations have contrasted only a limited number of subcategories of ES (e.g., living animal vocalizations vs. nonliving tool sounds, Murray et al., 2006; Lewis et al., 2004; Lewis et al., 2006) or have focused on a very low number of stimuli (e.g., eight in Pizz-amiglio et al., 2005). As such, it is unclear whether particular ES/language commonalities observed for specific categories of ES (e.g., sensorimotor transforms for nonliving tool/action sounds) hold for contrasts with different ES classes (e.g., nonliving nonaction sounds such as wind blowing) or when larger ES sets are considered. Secondly, it is not clear whether action/specific activations in the frontal cortex, hypothesized to mediate the action/language link, might instead be caused by uncontrolled variations in the semantic complexity of the stimuli (cf., above discussion). Thirdly, and most importantly, no previous study has to our knowledge systematically contrasted different ES classes with respect to the degree to which their neurocognitive processing overlaps with that of linguistic materials.

In this study, we address the problem of a language/ES commonality by focusing on category specific differences and by adopting a behavioral perspective. In Experiment 1, we investigate a large and heterogeneous set of living and nonliving sounds. We select stimuli from a large database of ES following a set of objective criteria designed to maximize the variety of the sound-generating events. Participants carry out a free identification task. In order to provide a partial answer to some of the methodological concerns raised above, we contrast various subclasses of ES (e.g., nonliving action vs. nonaction) relative to different measures of identification performance (e.g., naming agreement). Experiment 2 is conducted on two subsets of the stimuli investigated in Experiment 1. Participants evaluate the dissimilarity of the sounds or of the corresponding identification labels under three conditions: unbiased, semantic bias and acoustical bias. Based on these data, we measure the extent to which unbiased participants relied on sensory information or on the sensory-independent symbolic information activated by the identification labels. We contrast various ES subclasses with respect to eventual sensory or symbolic biases, among which living vs. nonliving, vocal vs. nonvocal and action vs. nonaction sounds. We expected a stronger reliance on language-specific symbolic information for nonliving action sounds.

Finally, we investigate whether eventual biases are affected by previous context so as to measure their robustness.

## 2. Experiment 1

We investigated the free identification of a large set of living and nonliving sounds. The stimulus set was derived from a database of ES following a set of objective rules. We computed different measures of identification performance: naming and conceptual agreement, and identification accuracy and time. For each of these measures, we tested for significant differences between living and nonliving, living vocal and nonvocal, living human and non-human and nonliving action and nonaction sounds. Measures of identification performance in this experiment were used to select a subset of the stimuli for Experiment 2. A subset of the identification labels derived from the free identification data of Experiment 1 were also used as stimuli in Experiment 2.

### 2.1. Method

#### 2.1.1. Participants

Participants ( $N = 20$ ; 9 females, 11 males; mean age = 21 yrs) were native English speakers and had normal hearing as assessed with a standard audiometric procedure (ISO 389-8., 2004; Martin & Champlin, 2000).

#### 2.1.2. Stimuli

Stimuli were selected from a large database of sound effects, see Appendix A and Tables 1 and 2. In contrast to previous studies, a selection process based on objective criteria was designed to maximize the variety of the sound-generating events and of the contexts in which the sounds are typically generated (e.g., kitchen). The living set ( $N = 71$ ) comprised sounds generated by the vibration of the body of living beings and also alimentation and locomotion sounds, and included vocal and nonvocal and human and non-human sounds. The nonliving set ( $N = 69$ ) comprised both action sounds generated as a result of the motor activity of a living agent either concurrently with or immediately preceding the generation of the sound and nonaction sounds. Note that according to this definition all living stimuli were also action sounds. All nonliving action sounds were generated using one or more artifacts, although not all of the sound-generating objects were tools used to perform or facilitate manual or mechanical work (e.g., crumpling paper bag as opposed to hammering nail in concrete).

The duration of the sounds (median value = 5.23 s) was reduced to the shortest possible value that still allowed identification of the sound-generating event (e.g., a walking sound contained more than two footsteps to avoid misidentification such as “hammer striking solid object”). The median RMS level was  $-27.4$  dB relative to the maximum possible value in the waveform. Based on the Wilcoxon rank sum test, we tested for significant differences in median duration and level between living and nonliving sounds, and between living human and non-human, living vocal and nonvocal, and nonliving action and nonaction sounds. Living and nonliving sounds had the same duration and level,  $p \geq 0.13$ . Human sounds were significantly shorter than non-human sounds, 3.1 and 6.1 s, respectively,  $p < 0.001$ . Vocal sounds were higher in level than nonvocal sounds,  $-26.0$  and  $-29.3$  dB, respectively,  $p = 0.001$ . Action sounds were shorter than nonaction sounds, 4.1 and 6.4 s, respectively,  $p < 0.001$ . No other significant differences emerged,  $p \geq 0.68$ .

#### 2.1.3. Apparatus

Sound stimuli were stored on the hard disk of a Macintosh G5 Workstation equipped with an M-Audio Audiophile 192 S/PDIF

**Table 1**  
Experimental stimuli investigated in Experiment 1, living set. The column labeled “Experiment 2” indicates which stimuli were also investigated in Experiment 2 (X), and in the training phase of Experiment 2 (training).

Sound event	Vocal	Human	Experiment 2	Identification	Naming agr.	Semantic agr.	p (correct)	IDT (s)
Bat screeching	X			Chirping bird	0.53	0.78	0.00	24.91
Alligator mississippiensis calling	X			Shooting video game	0.30	0.60	0.05	40.93
Human shaving		X		Brushing velcro	0.38	0.48	0.10	44.62
Chimpanzee panting heavily	X			Panting dog	0.33	0.58	0.15	32.19
Several cicadas buzzing in the forest				Buzzing insect	0.23	0.28	0.20	47.45
Otter sniffing	X			Sniffing dog	0.35	0.48	0.20	48.50
Rattlesnake rattling				Rattling sprinkler	0.20	0.25	0.25	40.30
Horse trotting on dirt				Running person	0.30	0.58	0.25	44.93
Timber rattlesnake rattling				Rattling snake	0.25	0.35	0.35	32.39
Herd passing by				Trotting train	0.18	0.45	0.35	42.67
Bull calf walking on hard ground				Walking horse	0.18	0.45	0.40	42.04
Chimpanzee panting	X			Panting dog	0.38	0.58	0.45	34.12
Man snoring loudly	X	X		Snoring man	0.63	0.70	0.45	16.12
Northern Leopard Frog calling				Croaking frog	0.28	0.60	0.50	34.34
Dog panting lightly	X			Panting dog	0.55	0.78	0.60	17.39
Long human kiss		X		Kissing kiss	0.33	0.60	0.65	23.03
Swarm of mosquitoes buzzing				Buzzing bee	0.50	0.73	0.70	20.60
Male sea elephants calling	X			Grunting animal	0.33	0.83	0.70	35.00
Italian crickets chirping at night				Buzzing insect	0.50	0.83	0.75	35.26
Adult sheep calling	X			Bleating sheep	0.60	0.98	1.00	13.52
Human brushing teeth				Brushing tooth	0.93	1.00	1.00	12.56
Crowd applauding heavily		X		Clapping hand	0.73	0.98	1.00	12.41
Horse whining indoors	X			Neighing horse	0.78	0.98	1.00	11.67
Elephant trumpeting	X			Trumpeting elephant	0.73	0.98	1.00	16.68
Happy man walking		X		Walking shoe	0.68	0.75	1.00	17.58
Sad man walking		X		Walking shoe	0.63	0.73	1.00	18.28
Man quietly snoring	X	X	X	Snoring man	0.70	0.75	0.50	15.85
Donkey braying	X		X	Braying donkey	0.28	0.48	0.55	27.99
Man breathing fast	X	X	X	Panting man	0.50	0.55	0.55	24.03
Camel chewing			X	Sucking mouth	0.20	0.33	0.55	62.93
Geese flapping wings			X	Fluttering bird	0.35	0.73	0.60	27.11
Turkey calling	X		X	Gobbling turkey	0.48	0.78	0.65	34.34
Horse snorting	X		X	Breathing horse	0.50	0.75	0.70	34.85
Man coughing	X	X	X	Coughing man	0.80	0.85	0.70	10.22
Small human fart		X	X	Farting person	0.48	0.70	0.70	34.27
Human swallowing		X	X	Swallowing liquid	0.40	0.65	0.70	18.53
Red tailed hawk calling	X		X	Calling eagle	0.38	0.85	0.75	16.35
Dog drinking water			X	Lapping water	0.48	0.75	0.75	18.10
Seal barks calling	X		X	Barking seal	0.60	0.85	0.80	28.62
Crow cawing	X		X	Cawing crow	0.60	0.88	0.80	10.97
Horse eating hay			X	Chewing animal	0.35	0.68	0.80	35.77
Single cricket chirping			X	Chirping cricket	0.83	0.90	0.85	9.99
Two horses galloping on dirt			X	Galloping horse	0.65	0.88	0.85	20.78
Human shushing	X	X	X	Shushing person	0.48	0.83	0.85	43.88
Human slowly eating apple		X	X	Biting apple	0.63	0.90	0.90	19.93
Honeybees foraging			X	Buzzing bee	0.75	0.95	0.90	11.75
Several chicken in barnyard	X		X	Clucking chicken	0.53	0.88	0.90	30.20
Woman giggling and laughing	X	X	X	Laughing woman	0.73	0.95	0.90	12.43
Dog panting heavily	X		X	Panting dog	0.93	0.93	0.90	11.59
Fly in cage			X	Buzzing fly	0.73	0.95	0.95	15.18
Calls of flying seagulls	X		X	Calling seagull	0.55	0.93	0.95	29.26
Pigeon landing on perch			X	Flapping bird	0.53	0.95	0.95	16.04
Pig snorting	X		X	Grunting pig	0.65	0.98	0.95	17.84
Woman screaming	X	X	X	Screaming woman	0.95	0.95	0.95	10.83
Person swimming in indoors pool		X	X	Splashing water	0.50	0.88	0.95	28.53
Large mixed breed dog whining	X		X	Whining dog	0.80	0.95	0.95	10.33
Medium human burp	X	X	X	Burping person	0.65	1.00	1.00	12.64
Crowd applauding moderately		X	X	Clapping hand	0.78	1.00	1.00	9.80
Various frogs croaking	X		X	Croaking frog	0.80	1.00	1.00	14.14
Morning call of rooster	X		X	Crowing rooster	0.85	1.00	1.00	17.16
Baby crying and whining	X	X	X	Crying baby	0.93	1.00	1.00	10.29
Wolves howling in chorus	X	X	X	Howling wolf	0.73	0.93	1.00	19.58
Cat meowing	X		X	Meowing cat	0.90	1.00	1.00	15.77
Cow mooing	X		X	Mooing cow	0.95	0.98	1.00	10.85
Throaty growls of a lion	X		X	Roaring lion	0.70	1.00	1.00	18.05
Happy woman walking		X	X	Walking shoe	0.60	0.73	1.00	17.11
Human slowly chewing potato chips		X	Training	Chewing food	0.45	0.88	0.85	14.57
Woman gasping	X	X	Training	Gasping woman	0.75	0.93	0.85	21.53
Two mallard ducks calling	X		Training	Quacking duck	0.68	0.95	0.95	20.62
Various kennel dogs barking	X		Training	Barking dog	0.95	0.98	1.00	10.05
Human blowing nose		X	Training	Blowing nose	0.98	0.98	1.00	10.12

Note: Agr. = agreement; IDT = identification time.

**Table 2**

Experimental stimuli investigated in Experiment 1, nonliving set. The column labeled “Experiment 2” indicates which stimuli were also investigated in Experiment 2 (X), and in the training phase of Experiment 2 (training).

Sound event	Action	Experiment 2	Identification	Naming agr.	Semantic agr.	p (correct)	IDT (s)
Cuica playing	X		Calling animal	0.48	0.98	0.00	31.23
Pop up of toaster			Opening cash register	0.25	0.35	0.05	35.79
Several firecrackers popping			Popping firecracker	0.25	0.38	0.05	38.71
Bubbles inside large water cooler			Pouring water	0.45	0.58	0.05	28.37
Suitcase: lid closing and latches snatching	X		Clicking door	0.15	0.33	0.10	40.20
Bicycle pulling up and skidding to stop	X		Sliding animal	0.15	0.35	0.10	42.23
Small gas flame			Blowing wind	0.35	0.45	0.25	33.80
Pulling off the tops of several carrots	X		Breaking wrapper	0.18	0.45	0.25	97.82
Bubbles in mud boiling inside pot			Bubbling water	0.30	0.43	0.25	38.31
Cork popping out of champagne bottle			Popping cork	0.23	0.35	0.30	36.91
Bowed vibraphone note (F3)	X		Ringling glass	0.30	0.68	0.30	32.91
Cabasa playing	X		Shaking shaker	0.23	0.35	0.30	39.45
Dry weeds and grass burning and crackling			Crackling fire	0.25	0.30	0.35	23.40
Skateboard passing by	X		Rolling car	0.38	0.45	0.35	31.52
Rigid rake gravel gardening	X		Scraping gravel	0.15	0.63	0.35	39.83
Light paddling in canoe on the sea	X		Paddling water	0.38	0.50	0.40	25.58
Closing cupboard door	X		Closing door	0.23	0.35	0.45	29.81
Cutter cutting a medium-size sheet of paper	X		Cutting knife	0.33	0.58	0.45	23.57
Large vertical blinds: draw closed, several pulls	X		Cutting metal	0.20	0.33	0.45	35.10
Spray can: spraying and running out of paint	X		Spraying spray can	0.25	0.48	0.45	25.56
Hammer nailing through wood into concrete	X		Clapping nail <sup>a</sup>	0.33	0.58	0.55	29.35
Water slowly dripping into metal sink			Hitting pot	0.25	0.55	0.70	34.57
Desert wind blowing			Blowing wind	0.80	0.98	0.95	24.92
Large brass ship's bell slowly ringing			Ringling bell	0.63	1.00	1.00	19.03
Cutting slices of bread	X	X	Cutting bread	0.45	0.63	0.50	25.87
Classical oboe tone	X	X	Playing oboe	0.43	0.70	0.50	33.76
Small flag flapping in the wind		X	Flapping flag	0.25	0.68	0.55	30.06
Toboggan passing by	X	X	Passing ski	0.25	0.65	0.55	36.24
Classical guitar tone	X	X	Plucking guitar	0.38	0.63	0.55	33.48
Marimba tone	X	X	Ringling xylophone	0.33	0.68	0.55	26.27
Cello tone (C2, muted, vibrato)	X	X	Playing cello	0.43	0.68	0.60	25.49
Staggered burst of M-16 machine gun	X	X	Shooting machine gun	0.53	0.78	0.60	14.32
Various squeaks of a rubber balloon	X	X	Stretching balloon	0.23	0.38	0.60	30.17
Rock falling on dirt		X	Dropping ball	0.40	0.63	0.65	41.07
Single entry in cash register	X	X	Typing typewriter	0.48	0.63	0.65	47.97
Fluorescent light humming		X	Buzzing machine	0.30	0.60	0.70	23.25
Brush hits on marching snare drum	X	X	Hitting snare drum	0.55	0.80	0.70	18.80
Wood burning in fireplace		X	Crackling fire	0.40	0.75	0.75	24.37
Bulb bicycle horn honking twice	X	X	Honking bike horn	0.45	0.85	0.75	17.35
Dice shaking and rolling on backgammon board	X	X	Rolling dice	0.68	0.83	0.75	28.95
Lake waves coming in		X	Lapping water	0.38	0.83	0.80	22.21
Heavy rain on water		X	Raining rain	0.53	0.68	0.85	21.10
Swinging of golf wood	X	X	Swinging racket	0.28	0.85	0.85	18.88
Blowing up rubber balloon with a single breath	X	X	Blowing balloon	0.65	0.90	0.90	13.94
Tooting party horn	X	X	Blowing party whistle	0.38	0.93	0.90	34.11
Water dripping into full tub		X	Dripping water	0.75	0.93	0.90	14.30
Medium river flowing from a distance		X	Flowing water	0.38	0.95	0.90	15.21
Bicycle bell ringing twice	X	X	Ringling bike bell	0.75	0.95	0.90	12.33
Sword removed from sheath	X	X	Sharpening knife	0.58	0.80	0.90	18.46
Steam and whistling kettle with boiling water		X	Boiling kettle	0.58	0.98	0.95	36.21
Waves crashing heavily on the ocean seashore		X	Crashing waves	0.38	0.80	0.95	49.80
Crumpling paper bag	X	X	Crumpling paper	0.40	0.85	0.95	19.54
Several coins dropping into metal cash drawer		X	Dropping change	0.60	0.98	0.95	17.33
Wine pouring into glass	X	X	Pouring water	0.63	0.83	0.95	15.58
Wind blowing in ghost town		X	Blowing wind	0.85	1.00	1.00	14.26
Two players playing table tennis	X	X	Bouncing ping pong ball	0.45	0.95	1.00	15.58
Bubbles in water boiling at medium intensity		X	Bubbling water	0.60	1.00	1.00	21.18
House toilet flushing and tank filling		X	Flushing toilet	1.00	1.00	1.00	17.98
Long keys jingle	X	X	Jingling keys	0.65	0.98	1.00	14.38
Ringling bell in railroad crossing		X	Ringling bell	0.68	0.95	1.00	17.25
Shower water running		X	Running water	0.50	0.98	1.00	23.19
Hand coping saw cutting	X	X	Sawing wood	0.75	0.85	1.00	17.14
Thunder clap and rumble		X	Thundering thunder	0.53	0.95	1.00	24.97
Adding numbers in an electronic office calculator	X	X	Typing keyboard	0.78	0.98	1.00	12.78
Short blast of air horn	X	Training	Blowing air horn	0.38	0.78	0.55	19.91
Vacuum sucking air through hose	X	Training	Sucking air	0.30	0.58	0.65	39.46
Full tin can shaking	X	Training	Shaking liquid	0.35	0.68	0.75	23.14
Water slowly filling a porcelain sink		Training	Filling water	0.60	0.83	0.90	22.46
Rock slashing into water		Training	Splashing water	0.55	0.90	0.90	29.40

Note: Agr. = agreement; IDT = identification time.

<sup>a</sup> This stimulus, not presented in Experiment 2, was equally often identified as clapping hands or as hammering nail.

interface. Audio signals were amplified with a Grace Design m904 monitor system and presented through Sennheiser HD280 headphones. Participants sat inside an IAC double-walled soundproof booth. Peak sound level was 40 dB SPL on average (SD = 6 dB SPL).

#### 2.1.4. Procedure

On each trial, participants were presented with one stimulus and were asked to identify the sound-generating event using one verb and one or two nouns (e.g., “closing door”). They were instructed to maximize identification speed and accuracy. They could replay each of the stimuli as many times as necessary. A trial was ended, and the succeeding one was started, when the participant clicked on an on-screen button labeled “Next”. Stimuli were presented in random order. Each of the 140 stimuli was identified once by each of the participants. The experiment lasted approximately 2 h.

#### 2.2. Results

We extracted four measures of identification performance: naming and semantic agreement, identification accuracy and identification time (see Tables 1 and 2). Focusing on the root of the verbal responses (e.g., “dog” for “dogs”), we extracted the verb and noun most frequently used to identify each sound (modal responses). Because only 25% of all identifications included two names, we considered all name responses together. If for a given stimulus multiple verbs or nouns were used equally often, we selected the least specific alternative (e.g., “food” was chosen over “apple”). Naming and semantic agreement were computed independently for the verb and noun responses; the final scores were averaged across these two response categories. Naming and semantic agreement measured the proportion of participants who used the modal response and whose identification agreed semantically with the modal response, respectively. A verb or noun response was in semantic agreement with the modal response if it was a synonym, if its relation to the modal response was as a subordinate (e.g., “coffee” for “liquid”), if it was a component or implication of the modal response (e.g., “toothbrush” is implied by “brushing teeth”) or if it was an acoustically plausible alternative at least as specific as the modal response (e.g., “lapping” for a modal response “splashing”). A response that was a superordinate of the modal response (e.g., “metal” for “keys”) was not scored as its semantic equivalent. Finally, identifications were scored as accurate if they were in semantic agreement with the description of the actual sound-generating event and were at least as specific as the modal identification. In contrast to previous studies (Marcell, Borella, Greene, Kerr, & Rogers, 2000; Ballas, 1993), responses that were more generic than the modal identification were not scored as correct. The measures of naming and semantic agreement and of identification accuracy were positively correlated with each other, Spearman rank correlation  $\rho \geq 0.78$ ;  $df = 138$ ,  $p < 0.001$ ; all of these measures were in turn negatively correlated with identification time,  $\rho \leq -0.69$ ;  $df = 138$ ,  $p < 0.001$ .

Based on the Wilcoxon rank sum test, we tested for significant differences in each of the four performance measures between living and nonliving sounds, living vocal and nonvocal sounds, living human and non-human sounds, and nonliving action and nonaction sounds. When compared to nonliving sounds, living sounds were identified faster, medians of 25.5 and 18.5 s, respectively. They also had higher median scores of semantic agreement, 0.70 and 0.85, respectively, naming agreement, 0.40 and 0.60, respectively, and identification accuracy, 0.70 and 0.85, respectively,  $p \leq 0.05$ . When compared to living nonvocal sounds, living vocal sounds had higher median scores of naming agreement, 0.40 and 0.65, respectively, and semantic agreement, 0.73 and 0.93, respectively,  $p \leq 0.009$ . No other significant differences emerged,  $p \geq 0.06$ .

We finally computed the correlation between the durations and levels of the sound stimuli, on the one hand, and each of the four measures of identification performance, on the other. Only the correlation between RMS level and semantic agreement was significant, although rather low in value,  $\rho = 0.18$ ,  $p = 0.04$ . None of the other correlations was significant,  $|\rho| \leq 0.16$ ,  $p \geq 0.06$ .

#### 2.3. Discussion

We investigated the free identification of a large set of ES. Based on a set of objective criteria, we maximized the variety of the sound-generating events included in the stimulus set. As such, we minimized selection biases and maximized the generality of our findings.

A variety of performance measures were extracted from the identification data: naming and semantic agreement, and identification accuracy and time. Strong positive correlations emerged between the first three of these four measures: more accurate identifications were also associated with higher levels of naming and semantic agreement among participants. All of these three measures increased for decreasing identification time. For example, consistently with the results of Ballas (1993) and Marcell et al. (2000), accurate identifications were made more quickly; the Spearman rank correlations between identification accuracy and time in these two studies were  $-0.75$  and  $-0.74$ , respectively. Note that the influence of two low-level sound features, duration and level, on identification performance appears to have been secondary at best.

We quantified differences in identification performance between various ES subcategories. As compared to living sounds, nonliving sounds were identified more slowly, less accurately, and with a vocabulary that was more heterogeneous and more semantically diverse. This result is consistent with the report by Fabiani, Kazmerski, Cycowicz, and Friedman (1996), in which human, animal and bird sounds were characterized by naming agreement scores that were higher than those for nonliving and synthetic sounds. Part of these results appears to be due to strong differences between vocal and nonvocal living sounds. Indeed, despite equally fast and accurate identifications for these two ES subcategories, naming and semantic agreement scores were significantly higher for vocal than for nonvocal sounds. However, no significant difference emerged between living human and non-human sounds or between nonliving action and nonaction sounds. When related to brain imaging studies of ES, these results point towards a need to control various measures of identification performance or to include them as covariates in the data analysis process, because uncontrolled differences in identification performance between ES subcategories might be sufficient to produce patterns of neural selectivity. For example, higher LIF activations might emerge for sounds characterized by a low, rather than high, naming/semantic agreement (Kan & Thompson-Schill, 2004), independently of the subcategory to which the various sounds belong. More specifically, an LIF specificity for nonliving tool/action sounds as compared to animal vocalizations (e.g., Murray et al., 2006; Lewis et al., 2004) might arise not because of a preferential link between action events and language (e.g., Aziz-Zadeh et al., 2006), but because of a higher number of verbal and semantic representations activated by nonliving action sounds.

### 3. Experiment 2

We investigated a subset of the living and nonliving stimuli from Experiment 1. In separate experimental conditions, we collected behavioral judgments of: (1) the acoustical dissimilarity of the sound stimuli, (2) the dissimilarity of the meaning of the iden-

tification labels for the sound stimuli, and (3) the dissimilarity of the sound stimuli in absence of biasing instructions. Based on these data, we modeled the extent to which listeners differentiate between ES by spontaneously focusing on their acoustical properties or on the symbolic mental representations activated by their identification. We measured differences in judgment biases between various ES subcategories. Based on previous hypotheses, we expected a stronger reliance on symbolic information for the category of nonliving action sounds. In order to increase the generality of the findings, we measured the extent to which a previous context of living or nonliving sounds influenced eventual cognitive biases. Based on previous studies of ES identification, we expected no effects of context.

### 3.1. Method

#### 3.1.1. Participants

All participants ( $N = 60$ ; 42 females, 18 males; mean age = 22 yrs) were native English speakers and had normal hearing and normal or corrected-to-normal vision. None of them had participated in Experiment 1.

#### 3.1.2. Stimuli and apparatus

We selected two sets of stimuli (40 living, 40 nonliving) and the corresponding modal identification labels from those investigated in Experiment 1 (see Tables 1 and 2 and Appendix A). The selection procedure aimed to maximize the identifiability and the semantic and acoustical heterogeneity of the sounds. It also sought to equalize the identifiability of living and nonliving sounds. Each of the stimuli was correctly identified by at least 50% of the participants in Experiment 1 (median = 90%). Living and nonliving sounds had the same median identification time, semantic agreement and identification accuracy, as measured by the Wilcoxon rank sum test,  $p \geq 0.07$ . Living sounds were characterized by a higher naming agreement,  $p = 0.007$ . Living and nonliving sounds had the same median duration and RMS level,  $p \geq 0.82$ , grand medians of 5.1 s and  $-27$  dB, respectively. Human sounds were shorter than living non-human sounds,  $p = 0.007$ ; vocal sounds had a lower RMS level than living nonvocal sounds,  $p = 0.02$ ; nonliving action sounds were shorter than nonaction sounds,  $p < 0.001$ .

For the training phase, we selected an additional group of ten highly identifiable stimuli (five living, five nonliving) from those investigated in Experiment 1. In one of the experimental conditions, the acoustical condition (see below), the training sounds were manipulated so as to make the sound-generating event difficult to identify, while still preserving part of the acoustical properties. We adopted a technique similar to that presented in Gygi et al. (2004). The modified sounds were random noises with the same amplitude envelope, same spectral mode (the frequency of the most intense spectral component) and same spectral center of gravity (the amplitude-weighted average frequency) as the original signals.

The apparatus was the same as for Experiment 1.

#### 3.1.3. Procedure

We used the method of hierarchical sorting to collect estimates of the dissimilarities among the stimuli (Coxon, 1999). Sounds or identification labels were represented by randomly numbered on-screen icons. They could be heard or viewed by clicking on the icon. Identification labels were presented at the screen center for a duration of 5 s, approximately the median duration of the sound stimuli. Participants were initially presented with each of the stimuli in sequential random order, each separated by a 100-ms silence or pause. They were then asked to create 15 groups of similar stimuli. They did so by dropping the icons into one of 15 boxes, each representing a group. Groups had to contain at least

one stimulus. Once the groups had been formed, participants were asked to merge together the two most similar groups. A binary merging was made at each of the subsequent steps, until all stimuli were merged into one group. Throughout the procedure, participants were free to listen or view the stimuli individually or in a group as many times as needed. The stimuli in the group were presented in sequence, successive sounds or words being separated by a 100-ms silence or pause. It took approximately 1 h to evaluate one stimulus set. The sorting task was initially practiced with a different set of 10 training stimuli.

An equal number of participants ( $N = 20$ ) was assigned to each of three experimental conditions. In a first (unbiased) condition, participants were presented sound stimuli and were asked to focus on their similarity. No further specification of the response criterion was given. In a second (acoustical) condition, they were presented sound stimuli and were instructed to focus on the similarity of their acoustical properties. For this condition, the training stimuli were unidentifiable manipulations of the training stimuli for the unbiased condition. In a final (semantic) condition, participants were presented with the linguistic identification (words on the screen) of the sound-generating event. They were asked to focus on the similarity of their meaning. The same participants in the unbiased condition evaluated the living and nonliving sets in separate sessions. Set order was counterbalanced across participants. Separate equally sized groups of participants in the acoustical and semantic conditions evaluated either the living or the nonliving set.

### 3.2. Results

Participants were asked to group similar stimuli at an earlier stage than dissimilar stimuli. The stage of the hierarchical sorting procedure at which two stimuli first merged was thus considered as an ordinal estimate of their dissimilarity (cf. Rao & Katz, 1971). All analyses were carried out on population data from the different experimental conditions, defined as the median dissimilarity across participants. Table 3 shows the Spearman rank correlation  $\rho$  between the dissimilarities from the different conditions.

In order to validate the behavioral estimates of acoustical and semantic dissimilarity, we compared them with objective acoustical and semantic dissimilarities. Objective semantic dissimilarities were computed based on a latent semantic analysis (LSA; Landauer, McNamara, Dennis, & Kintsch, 2007) of the identification labels. Roughly speaking, the dissimilarity between two identification labels is computed based on the co-occurrence of the words within meaningful units of text (e.g., sentences, paragraphs) in a reference database. We used the LSA resources available at <http://lsa.colorado.edu>. The reference database comprised general reading texts up to the first year of college. Objective acoustical dissimilarities were computed based on the Fast Fourier Transform (FFT) of the sound stimuli, filtered digitally so as to simulate the filtering that takes place in the outer and middle ear (cf. Giordano & McAdams, 2006). The FFT was calculated using the entire signal duration and was normalized so as to account for between-stimuli variations in the FFT size.<sup>2</sup> From the FFT, we calculated the average level within each of 30 third-octave spectral bands equally spaced on a logarithmic scale between 16 and 16000 Hz. The objective acoustical dissimilarity was finally defined as the Euclidean distance between the third-octave spectra of the sounds. For both the living and nonliving sets, the objective acoustical dissimilarity was more strongly correlated with the subjective acoustical dissimilarity,  $\rho = 0.24$  and 0.30 for living and nonliving

<sup>2</sup> Given a 1-kHz pure tone, the magnitude of the 1-kHz component of the discrete Fourier transform for a 2-s signal is the double of the magnitude of that for a 1-s signal.

**Table 3**  
Correlation between behavioral dissimilarities (lower and upper triangular matrix = polychoric and Spearman rank correlation, respectively), and Spearman rank correlation between behavioral and objective dissimilarities.

	Behavioral dissimilarities				Objective dissimilarities			
<i>Living set</i>								
Sem	–	Sen	Un1	Un2	Sem	Aco	Dur	RMS
Sen	0.33**	–	0.45**	0.61**	0.15**	0.24**	0.03	0.11**
Un1	0.61**	0.51**	–	0.51**	0.27**	0.03	0.13**	0.08**
Un2	0.39**	0.68**	0.59**	–	0.14**	0.19**	0.04	0.11**
<i>Nonliving set</i>								
Sem	–	0.36**	0.28**	0.35**	0.32**	0.09**	0.04	0.00
Sen	0.45**	–	0.56**	0.71**	0.32**	0.30**	0.01	0.14**
Un1	0.36**	0.65**	–	0.57**	0.26**	0.14**	0.13**	0.03
Un2	0.44**	0.79**	0.68**	–	0.27**	0.32**	0.09**	0.09*

Note: Sem = semantic; Sen = sensory; Un1 = unbiased-first condition; Un2 = unbiased-second condition; Aco = acoustical; Dur = duration; RMS = RMS sound level. \*  $p < 0.05$  and \*\*  $p < 0.01$ ;  $df = 778$ .

sets, respectively,  $df = 778$ ,  $p < 0.001$ , than with the subjective semantic dissimilarity,  $\rho = 0.07$  and  $0.19$  for the living and nonliving sets, respectively,  $df = 778$ ,  $p \geq 0.07$ . For both sets,  $p < 0.001$  for the difference between the correlations (Meng, Rosenthal, & Rubin, 1992). For the living set, the objective semantic dissimilarity was more strongly correlated with the subjective semantic than with the subjective acoustical dissimilarity,  $\rho = 0.37$  and  $0.15$ , respectively,  $df = 778$ ,  $p < 0.001$ ,  $p < 0.001$  for the difference. For the nonliving set, however, the objective semantic dissimilarity was equally strongly correlated with the subjective semantic and acoustical dissimilarities,  $\rho = 0.32$ ,  $df = 778$ ,  $p < 0.001$  for both dissimilarities,  $p = 0.95$  for the difference. This latter result mapped a significantly stronger correlation between subjective semantic and acoustical dissimilarities for nonliving than for living sounds (see below). Overall, the objective dissimilarities explained 15% of the variance in the ranks of the subjective dissimilarities at best. These values could be improved, e.g., by considering a higher number of acoustical features (e.g., Gygi, Kidd, & Watson, 2007) or by optimizing the reference database of texts for the LSA (Landauer et al., 2007). Overall, we took these results as validating the subjective dissimilarities.

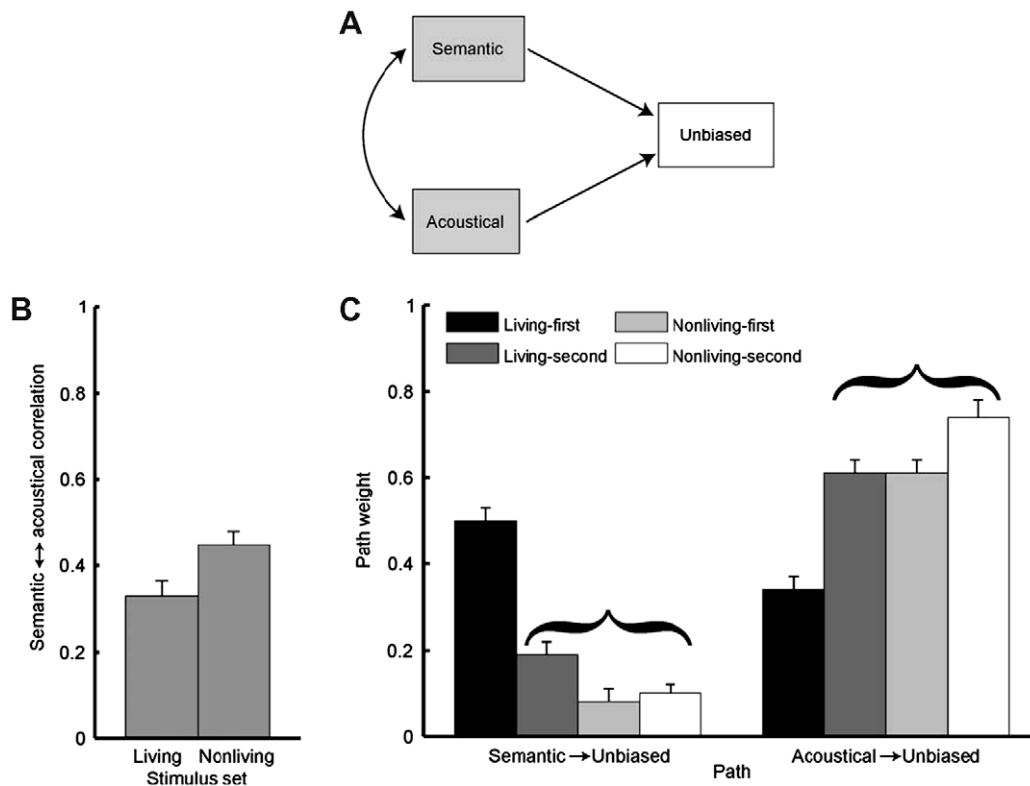
Further modeling was based on the polychoric correlation  $\rho_p$  between behavioral data from the different conditions (Jöreskog, 1990, see Table 3). A polychoric correlation measures the association between the continuous latent variables assumed to underlie the observed ordinal variables (e.g., Olsson, 1979). Both the nonliving and living semantic-acoustical correlations were significant,  $\rho_p = 0.45$ ,  $0.33$ , respectively,  $p < 0.001$ . The semantic-acoustical correlation was significantly higher for the nonliving set,  $p = 0.01$  for the bootstrap hypothesis test based on percentile intervals, number of bootstrap samples = 10,000 (Efron & Tibshirani, 1993). We modeled the polychoric correlation between dissimilarities from the different experimental conditions using a multigroup path analysis (Jöreskog & Sörbom, 2001, see Fig. 1). The path weights can be thought of as regression coefficients that estimate the effect of the predictors partialing out their correlation. In particular, in the case of two predictors  $x_1$  and  $x_2$  and a dependent variable  $y$ ,  $\rho(x_1, y) = pw(x_1, y) + pw(x_2, y)pw(x_1, x_2)$ , where  $pw()$  = path weight. Here the predictors were the semantic and acoustical dissimilarities, and the dependent variable was the unbiased dissimilarity from each of four different experimental conditions comparing living and nonliving sounds in the first or second block: living-first, living-second, nonliving-first and nonliving-second. As such, the path weights measured the extent to which the unbiased dissimilarities were influenced by information specific to either the semantic or acoustical dissimilarity. Different experimental conditions were considered as different groups in the multigroup path model. The model selection procedure involved two sequential stages. Firstly, we tested for significant between-groups differences

in the acoustical and semantic path weights altogether (Do the semantic and acoustical paths for different unbiased conditions differ?). Based on the results of this analysis, path weights for groups that were not significantly different were constrained to the same value in a second model. Secondly, within the simplified model we tested for significant within-group differences in the path weights (Do the semantic and acoustical paths for a given unbiased condition differ?). The first stage of the model selection procedure revealed that the weight of the semantic and acoustical paths did not differ significantly between the nonliving-first, nonliving-second and living-second conditions; pairwise group comparisons,  $\chi^2(2) \leq 2.65$ ;  $p \geq 0.26$ . The path weights for these three conditions were thus set to the same value in a refined model;  $T$ -values for path weights  $\geq 8.43$ ;  $p < 0.001$ , goodness-of-fit: Satorra-Bentler scaled  $\chi^2(4) = 7.37$ ;  $p = 0.12$ ; RMSEA = 0.03. Within this model, the weight of the semantic path for the living-first condition was significantly larger than that of the acoustical path,  $\chi^2(1) = 4.86$ ;  $p = 0.03$ . In all of the other experimental conditions, the weight for the acoustical path was higher than that of the semantic path,  $\chi^2(1) \geq 122.94$ ;  $p < 0.01$ . We finally note that we repeated the analyses on two subsets of stimuli, 37 living and 37 nonliving, equalized in all of the measures of identification performance, naming agreement included. We obtained essentially the same results as with the full dataset.

We repeated the path analysis by focusing on dissimilarities within and between specific subcategories of living and nonliving sounds. In a first set of models, we considered the dissimilarities within: vocal, living nonvocal, human, living non-human, nonliving action and nonliving nonaction sound sets,  $N = 300$ ,  $105$ ,  $91$ ,  $325$ ,  $300$  and  $105$ , respectively. For all of the living subcategories, the weights of the acoustical and semantic paths in the living-first condition were not significantly different,  $\chi^2(1) \leq 0.47$ ;  $p \geq 0.26$ , whereas in the living-second condition the acoustical path had a heavier weight,  $\chi^2(1) \geq 13.67$ ;  $p < 0.001$ . For both of the nonliving subcategories, and independently of previous context, the acoustical path was heavier than the semantic path,  $\chi^2(1) \geq 11.12$ ;  $p < 0.001$ . In a second set of models, we considered the dissimilarities between subcategories of sounds: living vocal vs. nonvocal, living human vs. non-human and nonliving action vs. nonaction,  $N = 375$ ,  $364$  and  $375$ , respectively. No significant effect of context emerged. For both of the sets of living dissimilarities, the semantic and acoustical paths did not have reliably different weights,  $\chi^2(1) \leq 0.48$ ;  $p \geq 0.49$ . For the action-nonaction dissimilarities, the acoustical path had a higher weight than the semantic path,  $\chi^2(1) \geq 59.87$ ;  $p < 0.001$ .

We finally tested the extent to which between-sound differences in duration and RMS level accounted for the behavioral dissimilarities. A feature-based dissimilarity was defined as the





**Fig. 1.** Experiment 2: (A) Path analysis model for unbiased dissimilarity. Double-headed arrow = correlation; one-headed arrow = path. (B) Polychoric semantic-acoustical correlation. (C) Path weights for the unbiased similarity in each of the four experimental conditions (living or nonliving, presented first or second) as estimated in the saturated model (perfect fit). Path weights grouped with curly braces are not significantly different ( $p \geq 0.26$ ). Error bar = standard error of the estimate.

absolute difference in the value of a feature between paired sounds. Table 3 reports the Spearman rank correlation between feature-based dissimilarities and behavioral dissimilarities. Several significant positive correlations emerged, showing that pairs of stimuli judged as more dissimilar tended to be characterized by larger differences in duration and RMS level. Note however that these parameters explained 2% of the variance in the data at best. Further, none of these correlations was higher than that between the subjective acoustical dissimilarity and the objective acoustical dissimilarity computed on the basis of the spectra of the sounds. As such, although participants were significantly influenced by duration and level differences, the cognitive role of these low-level features was secondary.

### 3.3. Discussion

With this experiment, our aim was to measure the extent to which listeners differentiate between living or nonliving ES by spontaneously focusing either on their acoustical properties or on sensory-independent symbolic representations activated by the ES. We measured acoustical and symbolic biases for various ES subcategories and tested whether biases were robust to changes in previous context. Participants judged the dissimilarity of the sounds under three conditions: in the absence of experimenter-specified response criteria (unbiased condition), focusing on the acoustical properties of the sounds (acoustical condition), or focusing on the meaning of the words identifying the sound-generating event (semantic condition). An acoustical or symbolic bias was inferred if unbiased data correlated most strongly with the acoustical or semantic dissimilarities, respectively. In particular, the measurement of biases focused on the proportion of variance specific to the acoustical and semantic dissimilarities, and discarded the common acoustical-semantic variance.

Similarity is a central construct for the study of various cognitive processes, among which categorization (e.g., Goldstone, 1994), identification, choice, preference (Ashby, 1992) and memory (e.g., Baddeley, 2003). The neurobiological significance of behavioral measures of dissimilarity resides in the role they play in the data-interpretation process. In a PET study, Zatorre, Bouffard, and Belin (2004) observed a positive correlation between right aSTS blood flow levels, on the one hand, and the dissimilarity of mixtures of simultaneous ES, on the other. It was hypothesized that the right aSTS processes the sound features that differentiate between ES. Halpern, Zatorre, Bouffard, and Johnson (2004) measured the behavioral dissimilarity of heard and imagined musical timbres and their neural correlates. Perceptual and imagery dissimilarities were strongly correlated. As emerging from a conjunction analysis of BOLD responses, this behavioral result was paralleled by a number of common perceptual-imagery activations (e.g., right pSTG). Differently from activation-based methods, information-based analyses focus on what information about the experimental conditions is encoded in the patterns of neural activity (Kriegeskorte, Goebel, & Bandettini, 2006). Information-based methods rely on the comparison of dissimilarity matrices computed within different domains (Kriegeskorte, Mur, & Bandettini, 2008): neural (e.g., spatial patterns of activation), computational (e.g., models of sensory processing), conceptual (e.g., living vs. non-living distinction), and behavioral (e.g., dissimilarity ratings). Within this framework, a significant neural-behavioral correlation provides evidence for which cortical areas are involved in the behavioral task (e.g., Weber, Thompson-Schill, Osherson, Haxby, & Parsons, 2009). Importantly, behavioral dissimilarities that are strongly correlated with each other would likely correlate equally strongly with the neural dissimilarities from the same cortical area. Related to this study, a correlation between dissimilarities from different conditions might potentially indicate common cognitive

representations, and also common neural substrates for the judgments in the different conditions.<sup>3</sup>

Acoustical and semantic similarities were significantly correlated for both the nonliving and living set. This result is in line with the hypothesis of partially overlapping semantic networks for ES and linguistic materials (e.g., Saygin et al., 2003). The semantic-acoustical correlation was significantly stronger for nonliving than for living sounds. This result might parallel the fact that nonliving sounds preferentially activate a region that emerges as a common neural substrate for words and ES: the left posterior temporal cortex (Lewis et al., 2005; Kiefer et al., 2008). Based on the results of our study, it might be argued that a neurobiological connection between nonliving ES and language arises because of a stronger coherence between their semantics and their acoustics: similar words describe acoustically similar ES, and particularly nonliving ES.

Further analyses measured the influence of acoustical and semantic dissimilarities on unbiased dissimilarities, independently of the semantic-acoustical correlation. Participants differentiated living and nonliving sounds focusing more on symbolic information for the former and acoustical information for the latter. We can advance three different psychophysical explanations for this result. Firstly, listeners focused on the source of dissimilarity that better differentiated the stimuli (cf. Tversky's, 1977, diagnosticity principle). Vocalizations are consistently reported as more acoustically similar than are nonliving tool sounds (Lewis et al., 2005). Secondly, listeners focused on the most reliable source of information, i.e., that which remains constant across repeated presentations of the same stimulus (cf. Ernst & Bühlhoff, 2004). In line with this, living sounds evoke linguistic representations that are more constant across subjects, and likely across repeated representations for the same participant (Experiment 1). Finally, the participants in Experiment 2 might simply focus on the information they exploit most frequently outside the laboratory. Notably, a focus on the acoustics of a sound would optimize the recognition of the mechanics of the sound source (e.g., Is this the sound of liquid flowing?), because acoustics and mechanics are deterministically connected (Fletcher & Rossing, 1991). As such, it might be argued that in everyday listening we are concerned with recognizing the source of a nonliving sound and the symbolic meaning of living sounds: estimating precisely the length of the vocal tract of a crying baby is not as important as recognizing the need for help. From the neurobiological standpoint, the relevance of symbolic information for living sounds parallels the left anterior temporal involvement with the processing of vocalizations (e.g., Belin et al., 2000) and with the highly symbolic linguistic processes of syntactic and computational semantic analysis (e.g., Humphries et al., 2006). Notably, the acoustical (and not symbolic) bias for nonliving sounds redefines previous speculations about a preferential link between nonliving action/tool sounds and language. In contrast with the hypothesis of abstract processing for action sounds (e.g., Kaplan & Iacoboni, 2005; Galati et al., 2008), our study shows that sensory-independent representations play a largely secondary role in the cognitive processing of the entire class of nonliving ES.

We carried out more detailed analyses by quantifying the symbolic and acoustical biases for subcategories of living and nonliving ES. Vocal, nonvocal, human and non-human sounds were evaluated giving the same weight to symbolic and acoustical information. Action and nonaction living sounds were evaluated focusing on acoustical information. We also focused on the dissimilarities between vocal and nonvocal, human and non-human and action

and nonaction sounds. We observed the same cognitive biases as for the dissimilarities within subcategories of ES. It should be noted that the majority of the studies on neural specificities for nonliving action/tool sounds involved a contrast with living vocalizations, but not with other subclasses of nonliving sounds (e.g., Murray et al., 2006). Our findings point to a larger uniformity in the processing strategies for the nonliving ES category as a whole. The generality of future brain imaging studies will thus be increased by considering nonliving action and nonaction sounds (e.g., thunder, rain).

Participants evaluated the living or nonliving sounds before or after the other stimulus set. Our aim was to assess the extent to which response biases were robust with respect to the influence of previous context. Nonliving sounds, including the action and nonaction subcategories and their dissimilarities, were robustly evaluated focusing on acoustical information. Thus the processing of nonliving sounds was independent of the influence of previous context. This result is consistent with a previously reported null effect of context on the identification of acoustically ambiguous nonliving sounds (Ballas & Mullins, 1991). Previous context did, however, affect the evaluation of living sounds. Indeed, after a preceding nonliving context, the evaluation of living sounds and the subcategories of vocal, nonvocal, human and non-human sounds focused on acoustical information. As such, the processing of living sounds generally appeared to be more flexible than that of nonliving sounds: depending on context, listeners are capable of extracting either symbolic information or, when focusing on the acoustical surface, source properties such as the identity of a speaker (cf. Belin et al., 2000; Belin et al., 2002). Interestingly, the evaluation of the dissimilarity between vocal and nonvocal living sounds and between human and non-human living sounds was not affected by previous context. Listeners always assigned the same weight to acoustical and symbolic information. A context-independent robust strategy for these important distinctions (vocal vs. nonvocal and human vs. non-human) might serve the purpose of maximizing the likelihood of a correct perceptual discrimination, based on all the available information (e.g., Kellman, 1996). Two hypotheses could be formulated to explain the observed context effects. Firstly, an "inertial" tendency to continue adopting the same response strategy as for the first sound set. Secondly, there is a generalized increase in the weight of acoustical information as listeners become more skilled with the task or are exposed to a larger number of sounds. None of these explanations are able to fully account for the observed data. Most importantly, to our knowledge no brain imaging study has assessed the extent to which the neural processing of a specific category of ES changes when it is presented along with ES from another category. Indeed, the vast majority of our knowledge of cortical specificities relies on experiments conducted with mixed living-nonliving sound sets. In the light of our results, this methodology might have resulted in an increased activation of areas devoted to the processing of the sensory features of living sounds, at the expense of a reduced activation in areas devoted to the processing of symbolic representations.

#### 4. Conclusions

Environmental sounds and language share a referential function: they both can activate mental representations of objects and events in the environment. Previous studies have demonstrated a number of neurocognitive similarities between the processing of ES and of language, among which partially overlapping systems of semantic retrieval located in the left temporal lobe (e.g., Kiefer et al., 2008). Previous studies have argued for a preferential language/ES link for the category of nonliving action/tool

<sup>3</sup> Correlation-based methods must be taken as indications of a potential causal relationship and not as proof of an actual causal result. Further experimental manipulations are necessary to confirm a causal link between cortical activity and the processes resulting in an observed behavior (e.g., TMS studies).

sounds. Indeed, similarly to language, this ES subcategory activates sensorimotor transformations in the premotor cortex, which reenact the motor programs for the generation of the sound being heard (e.g., Lewis et al., 2005; Lewis et al., 2006). Most importantly, the action ES/language link has also been emphasized because of a left frontal and premotor response to action events in multiple sensory modalities, which has been argued to favor the development of sensory-independent symbolic representations (Kaplan & Iacoboni, 2005; Galati et al., 2008), because the manipulation of symbolic representations underlies the language ability (e.g., Deacon, 1997; Bickerton, 2005). Notably, however, no previous study has assessed the extent to which processing of environmental sounds in general, or of specific categories of environmental sounds (e.g., nonliving action sounds), requires the activation of symbolic representations.

We investigated the free identification and dissimilarity estimation of large sets of nonliving and living sounds. The results of Experiment 1 showed that the identifications for nonliving action and nonaction sounds use a vocabulary that is more heterogeneous and semantically varied than is the case with living sounds, particularly vocal sounds. As such, we hypothesized that the LIFG selectivity for nonliving action sounds observed in previous studies might have emerged not because of a preferential action/language link, but because of a larger number of semantic/linguistic representations activated by these sounds. Most importantly, the results of Experiment 2 showed that listeners differentiate nonliving action and nonaction sounds using a robust iconic strategy that focuses on acoustical information, independent of context. On the other hand, living sounds were evaluated using a flexible cognitive strategy, which focuses spontaneously on symbolic information and, after a nonliving context, relies heavily on acoustical information. Overall, these results redefine previous hypotheses on the neurocognitive similarities between ES and language: whereas for nonliving sounds the strongest point of similarity appears to be the activation of sensorimotor transformations, for living sounds it is the ability to focus on symbolic representations.

## Acknowledgments

Portions of this research were reported at the International Conference on Auditory Display (ICAD 2007) in Montréal, Québec, Canada, in June 2007, and at the 48th Annual Meeting of the Psychonomic Society in Long Beach, California, USA, in November 2007. This work was supported by Stephen McAdams' Canada Research Chair in Music Perception and Cognition, by a grant to him from the Natural Sciences and Engineering Research Council of Canada (RGPIN 312774-05), and by a Special Research Opportunity Grant from the Natural Sciences and Engineering Research Council of Canada (NSERC).

The authors wish to thank two anonymous reviewers for helpful comments on previous versions of this manuscript.

## Appendix A

We created a database of roughly 6500 sounds drawing primarily from commercially available databases of sound effects (Sound Ideas, 2004; Opolko & Wapnick, 1987), but also from online and published resources (Elliott, 2005) and from a previous study on the perception of walking sounds (Giordano & Bresin, 2006). Overall, the database did not include speech, synthetic sounds, Foley effects, sounds generated by multiple types of interaction between sound-generating objects, such as the rolling and impact in bowling sounds, and sounds generated by vibrating materials of multiple states, such as liquid and solid in the sound of coffee stirring (cf. Gaver, 1993).

We classified each sound in relation to several known properties of the sound-generation process. The classification system guided the process of selecting the stimuli for Experiment 1. All of the sounds were classified based on the following three criteria: living vs. nonliving: the sound-generating object is or is not part of the body of a living being; action vs. nonaction: the sound is or is not generated as a result of the motor activity of a living agent that is either concurrent or immediately precedes the generation of the sound; musical vs. nonmusical: the sound source is or is not commonly referred to as a musical instrument.

Living sound sources were further classified based on: vocal vs. nonvocal: the sound is produced by the vibration of vocal folds; taxonomical class: amphibians, birds, insects, humans, non-human mammals; communicative vs. noncommunicative: the sound has or does not have a communicative function.

Nonliving, nonmusical sound sources were further classified based on Gaver (1993): material class: solid, liquid, gaseous (e.g., wind), combustion (e.g., fire) and electric sounds; interaction type: deformation, impact, rolling and scraping for solids; bubbling, dripping, flowing, pouring, sloshing, and splashing for liquids; continuous, steam, whoosh and wind for gases; simple and crackling for combustion; explosive and continuous for electric. Musical sounds were further classified based on: musical instrument family: aerophone, chordophone, idiophone, membranophone (von Hornbostel & Sachs, 1914); interaction type: impulsive, continuant, multiple impacts (Hajda, Kendall, Carterette, & Harshberger, 1997).

Action sounds were further classified based on the following criteria: locomotion vs. nonlocomotion: the sound is generated by the locomotion of a living being or not; alimentation vs. nonalimentation: the sound is generated by a feeding living being or not.

All sounds were finally classified based on their context, i.e., the location where they are typically generated. For living sounds we distinguished between six contexts: anywhere, indoors-generic, toilet, farm, sea, wild (e.g., forest). For nonliving sounds we distinguished between 19 contexts: anywhere, casino, party, indoors-generic, kitchen, toilet, construction field, military, office, outdoors-anywhere, sea, wild, sport, shopping, travel-generic, bicycle travel, marine travel, rail travel. Musical sounds were assumed as potentially generated anywhere.

For the purpose of the selection of the stimuli for Experiment 1 and Experiment 2, we defined a set of categories of interest, given by the intersection of the above-defined classes (e.g., a vacuuming sound belonged to the “nonliving – action – nonmusical – aerodynamic – continuous – indoors-anywhere” category). For Experiment 1, the selection procedure sought to maximize the number of categories of interest. Experiment 2 investigated a subset of the stimuli in Experiment 1. The selection procedure sought to: maximize the identifiability of the sound stimuli, minimize identifiability differences between living and nonliving stimuli, maximize the number of categories of interest, and maximize the diversity of the identification labels investigated in the semantic condition.

## References

- Adams, R. B., & Janata, P. (2002). A comparison of neural circuits underlying auditory and visual object categorization. *Neuroimage*, 16, 361–377.
- Altmann, C. F., Doehrmann, O., & Kaiser, J. (2007). Selectivity for animal vocalizations in the human auditory cortex. *Cerebral Cortex*, 17, 2601.
- Altmann, C. F., Nakata, H., Noguchi, Y., Inui, K., Hoshiyama, M., Kaneoke, Y., et al. (2008). Temporal dynamics of adaptation to natural sounds in the human auditory cortex. *Cerebral Cortex*, 18, 1350–1360.
- Arbib, M. A. (2005). From monkey-like action recognition to human language: An evolutionary framework for neurolinguistics. *Behavioral and Brain Sciences*, 28, 105–124.
- Arbib, M. A., & Rizzolatti, G. (1997). Neural expectations: A possible evolutionary path from manual skills to language. *Communication & Cognition*, 29, 393–423.
- Ashby, F. G. (1992). *Multidimensional models of perception and cognition*. Hillsdale, NJ: Lawrence Erlbaum Associates.

- Aziz-Zadeh, L., Koski, L., Zaidel, E., Mazziotta, J., & Iacoboni, M. (2006). Lateralization of the human mirror neuron system. *Journal of Neuroscience*, 26, 2964.
- Baddeley, A. (2003). Working memory: Looking back and looking forward. *Nature Reviews Neuroscience*, 4, 829–839.
- Ballas, J. A. (1993). Common factors in the identification of an assortment of brief everyday sounds. *Journal of Experimental Psychology: Human Perception and Performance*, 19, 250–267.
- Ballas, J. A., & Howard, J. H. Jr. (1987). Interpreting the language of environmental sounds. *Environment and Behavior*, 19, 91.
- Ballas, J. A., & Mullins, T. (1991). Effects of context on the identification of everyday sounds. *Human Performance*, 4, 199–219.
- Beauchamp, M. S., Lee, K. E., Argall, B. D., & Martin, A. (2004). Integration of auditory and visual information about objects in superior temporal sulcus. *Neuron*, 41, 809–823.
- Belin, P., Fecteau, S., & Bédard, C. (2004). Thinking the voice: neural correlates of voice perception. *Trends in Cognitive Sciences*, 8, 129–135.
- Belin, P., Zatorre, R. J., & Ahad, P. (2002). Human temporal-lobe response to vocal sounds. *Cognitive Brain Research*, 13, 17–26.
- Belin, P., Zatorre, R. J., Lafaille, P., Ahad, P., & Pike, B. (2000). Voice-selective areas in human auditory cortex. *Nature*, 403, 309–312.
- Bergerbest, D., Ghahremani, D. G., & Gabrieli, J. D. E. (2004). Neural correlates of auditory repetition priming: Reduced fMRI activation in the auditory cortex. *Journal of Cognitive Neuroscience*, 16, 966–977.
- Bickerton, D. (2005). Beyond the mirror neuron U-the smoke neuron? Commentary on Arbib (2005). *Behavioral and Brain Sciences*, 28, 126.
- Binder, J., & Price, C. J. (2001). Functional neuroimaging of language. In R. Cabeza & A. Kingstone (Eds.), *Handbook of functional neuroimaging of cognition* (pp. 187–251). Cambridge, MA: MIT Press.
- Cabeza, R., & Nyberg, L. (2000). Imaging cognition II: An empirical review of 275 PET and fMRI studies. *Journal of Cognitive Neuroscience*, 12, 1–47.
- Calvert, G. A., & Lewis, J. W. (2004). Hemodynamic studies of audiovisual interactions. In G. Calvert, C. Spence, & B. Stein (Eds.), *Handbook of multisensory processes* (pp. 483–502). Cambridge, MA: MIT Press.
- Chiu, C. Y. P. (2000). Specificity of auditory implicit and explicit memory: Is perceptual priming for environmental sounds exemplar specific? *Memory and Cognition*, 28, 1126–1139.
- Chiu, C. Y. P., & Schacter, D. L. (1995). Auditory priming for nonverbal information: Implicit and explicit memory for environmental sounds. *Consciousness and Cognition*, 4, 440–458.
- Cohen, Y. E., Hauser, M. D., & Russ, B. E. (2006). Spontaneous processing of abstract categorical information in the ventrolateral prefrontal cortex. *Biology Letters*, 2, 261–265.
- Coxon, A. P. M. (1999). *Sorting data: Collection and analysis*. Thousand Oaks, CA: Sage Publications.
- Cummings, A., Cèponienė, R., Koyama, A., Saygin, A. P., Townsend, J., & Dick, F. (2006). Auditory semantic networks for words and natural sounds. *Brain Research*, 1115, 92–107.
- Deacon, T. (1997). *The symbolic species: The co-evolution of language and the brain*. New York, NY: W.W. Norton & Company Ltd.
- Dick, F., Saygin, A., Galati, G., Pitzalis, S., Bentrovato, S., D'Amico, S., et al. (2007). What is involved and what is necessary for complex linguistic and nonlinguistic auditory processing: Evidence from functional magnetic resonance imaging and lesion data. *Journal of Cognitive Neuroscience*, 19, 799.
- Doehrmann, O., Naumer, M. J., Volz, S., Kaiser, J., & Altmann, C. F. (2008). Probing category selectivity for environmental sounds in the human auditory brain. *Neuropsychologia*, 46, 2776–2786.
- Efron, B., & Tibshirani, R. J. (1993). *An introduction to the bootstrap*. New York, NY: Chapman & Hall.
- Elliott, L. (2005). *A guide to wildlife sounds*. Mechanisburg, PA: Stackpole Books.
- Engelien, A., Tüscher, O., Hermans, W., Isenberg, N., Eidelberg, D., Frith, C., et al. (2006). Functional neuroanatomy of non-verbal semantic sound processing in humans. *Journal of Neural Transmission*, 113, 599–608.
- Ernst, M. O., & Bühlhoff, H. H. (2004). Merging the senses into a robust percept. *Trends in Cognitive Sciences*, 8, 162–169.
- Fabiani, M., Kazmerski, V. A., Cycowicz, Y. M., & Friedman, D. (1996). Naming norms for brief environmental sounds: Effects of age and dementia. *Psychophysiology*, 33, 462–475.
- 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.
- Fecteau, S., Armony, J. L., Joannette, Y., & Belin, P. (2004). Is voice processing species-specific in human auditory cortex? An fMRI study. *Neuroimage*, 23, 840–848.
- Fecteau, S., Armony, J. L., Joannette, Y., & Belin, P. (2005). Sensitivity to voice in human prefrontal cortex. *Journal of Neurophysiology*, 94, 2251–2254.
- Fletcher, N. H., & Rossing, T. D. (1991). *The physics of musical instruments*. New York, NY: Springer-Verlag.
- Galati, G., Committeri, G., Spironi, G., Aprile, T., Di Russo, F., Pitzalis, S., et al. (2008). A selective representation of the meaning of actions in the auditory mirror system. *Neuroimage*, 40, 1274–1286.
- Gaver, W. W. (1993). What in the world do we hear? An ecological approach to auditory event perception. *Ecological Psychology*, 5, 1–29.
- Giordano, B. L., & Bresin, R. (2006). Walking and playing: What's the origin of emotional expressiveness in music? In M. Baroni, A. R. Addessi, R. Caterina, & M. Costa (Eds.), *Proceedings of the 9th international conference on music perception and cognition (ICMPC9)*. Italy: Bologna.
- Giordano, B. L., & McAdams, S. (2006). Material identification of real impact sounds: Effects of size variation in steel, glass, wood and plexiglass plates. *Journal of the Acoustical Society of America*, 119, 1171–1181.
- Giraud, A. L., & Price, C. J. (2001). The constraints functional neuroimaging places on classical models of auditory word processing. *Journal of Cognitive Neuroscience*, 13, 754–765.
- Goldstone, R. L. (1994). The role of similarity in categorization: Providing a framework. *Cognition*, 52, 125–157.
- Gygi, B., Kidd, G. R., & Watson, C. S. (2004). Spectral-temporal factors in the identification of environmental sounds. *Journal of the Acoustical Society of America*, 115, 1252–1265.
- Gygi, B., Kidd, G. R., & Watson, C. S. (2007). Similarity and categorization of environmental sounds. *Perception & Psychophysics*, 69, 839–855.
- Hajda, J. M., Kendall, R. A., Carterette, E. C., & Harshberger, M. L. (1997). Methodological issues in timbre research. In I. Deliège & J. Sloboda (Eds.), *The perception and cognition of music* (pp. 253–306). London: L. Erlbaum.
- Halpern, A. R., Zatorre, R. J., Bouffard, M., & Johnson, J. A. (2004). Behavioral and neural correlates of perceived and imagined musical timbre. *Neuropsychologia*, 42, 1281–1292.
- Hasson, U., & Small, S. L. (2008). Functional magnetic resonance imaging “(fMRI)” research of language. In B. Stemmer & H. A. Whitaker (Eds.), *Handbook of the neuroscience of language* (pp. 81–89). London, UK: Academic Press.
- Hauser, M. D., Chomsky, N., & Fitch, W. (2002). The faculty of language: What is it, who has it, and how did it evolve? *Science*, 298, 1569.
- Hewes, G. W. (1973). Primate communication and the gestural origin of language. *Current Anthropology*, 14, 5.
- Hocking, J., & Price, C. J. (2008). The role of the posterior superior temporal sulcus in audiovisual processing. *Cerebral Cortex*.
- Humphries, C., Binder, J. R., Medler, D. A., & Liebenthal, E. (2006). Syntactic and semantic modulation of neural activity during auditory sentence comprehension. *Journal of Cognitive Neuroscience*, 18, 665–679.
- Humphries, C., Willard, K., Buchsbaum, B., & Hickok, G. (2001). Role of anterior temporal cortex in auditory sentence comprehension: An fMRI study. *NeuroReport*, 12, 1749.
- Ideas, Sound. (2004). *Series 6000 DVD combo sound effects library*. Ontario, Canada: Richmond Hill.
- ISO 389-8. (2004). *Acoustics – Reference zero for the calibration of audiometric equipment – Part 8: Reference equivalent threshold sound pressure levels for pure tones and circumaural earphones (Tech. Rep.)*. International Organization for Standardization, Geneva.
- Jöreskog, K. G. (1990). New developments in LISREL: Analysis of ordinal variables using polychoric correlations and weighted least squares. *Quality and Quantity*, 24, 387–404.
- Jöreskog, K. G., & Sörbom, D. (2001). *LISREL 8: User's reference guide*. Lincolnwood, IL, USA: Scientific Software International.
- Kan, I. P., & Thompson-Schill, S. L. (2004). Effect of name agreement on prefrontal activity during overt and covert picture naming. *Cognitive, Affective, & Behavioral Neuroscience*, 4, 43–57.
- Kaplan, J. T., & Iacoboni, M. (2005). Listen to my actions! Commentary on Arbib (2005). *Behavioral and Brain Sciences*, 28, 135–136.
- Kaplan, J. T., & Iacoboni, M. (2007). Multimodal action representation in human left ventral premotor cortex. *Cognitive Processing*, 8, 103–113.
- Kellman, P. J. (1996). The origins of object perception. In R. Gelman & T. Kit-Fong Au (Eds.), *Handbook of perception and cognition*, *Perceptual and cognitive development* (Vol. 8, pp. 3–48). San Diego, CA: Academic Press.
- Kiefer, M., Sim, E. J., Herringer, B., Grothe, J., & Hoenig, K. (2008). The sound of concepts: Four markers for a link between auditory and conceptual brain systems. *Journal of Neuroscience*, 28, 12224–12230.
- Kohler, E., Keyers, C., Umiltà, M. A., Fogassi, L., Gallese, V., & Rizzolatti, G. (2002). Hearing sounds, understanding actions: Action representation in mirror neurons. *Science*, 297, 846–848.
- Kraut, M. A., Pitcock, J. A., Calhoun, V., Li, J., Freeman, T., & Hart, J. Jr. (2006). Neuroanatomic organization of sound memory in humans. *Journal of Cognitive Neuroscience*, 18, 1877–1888.
- Kriegeskorte, N., Goebel, R., & Bandettini, P. (2006). Information-based functional brain mapping. *Proceedings of the National Academy of Sciences*, 103, 3863–3868.
- Kriegeskorte, N., Mur, M., & Bandettini, P. (2008). Representational similarity analysis – connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience*, 2.
- Landauer, T. K., McNamara, D. S., Dennis, S., & Kintsch, W. (2007). *Handbook of latent semantic analysis*. Mahwah, NJ: Lawrence Erlbaum Associates.
- 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.
- Lenz, D., Schadow, J., Thaeig, S., Busch, N. A., & Herrmann, C. S. (2007). What's that sound? Matches with auditory long-term memory induce gamma activity in human EEG. *International Journal of Psychophysiology*, 64, 31–38.
- Lewis, J. W. (2006). Cortical networks related to human use of tools. *The Neuroscientist*, 12, 211.
- Lewis, J. W., Brefczynski, J. A., Phinney, R. E., Jannik, J. J., & DeYoe, E. D. (2005). Distinct cortical pathways for processing tool vs. animal sounds. *The Journal of Neuroscience*, 25, 5148–5158.
- Lewis, J. W., Phinney, R. E., Brefczynski-Lewis, J. A., & DeYoe, E. A. (2006). Lefties get it “right” when hearing tool sounds. *Journal of Cognitive Neuroscience*, 18, 1314.

- Lewis, J. W., Wightman, F. L., Brefczynski, J. A., Phinney, R. E., Binder, J. R., & DeYoe, E. A. (2004). Human brain regions involved in recognizing environmental sounds. *Cerebral Cortex*, *14*, 1008–1021.
- Liebenthal, E., Binder, J. R., Spitzer, S. M., Possing, E. T., & Medler, D. A. (2005). Neural substrates of phonemic perception. *Cerebral Cortex*, *15*, 1621–1631.
- Maeder, P. P., Meuli, R. A., Adriani, M., Bellmann, A., Fornari, E., Thiran, J. P., et al. (2001). Distinct pathways involved in sound recognition and localization: A human fMRI study. *Neuroimage*, *14*, 802–816.
- Marcell, M. E., Borella, D., Greene, M., Kerr, E., & Rogers, S. (2000). Confrontation naming of environmental sounds. *Journal of Clinical and Experimental Neuropsychology*, *22*, 830–864.
- Martin, F. N., & Champlin, C. A. (2000). Reconsidering the limits of normal hearing. *Journal of the American Academy of Audiology*, *11*, 64–66.
- Meng, X. L. I., Rosenthal, R., & Rubin, D. B. (1992). Comparing correlated correlation coefficients. *Psychological Bulletin*, *111*, 172–175.
- Moss, H. E., Abdallah, S., Fletcher, P., Bright, P., Pilgrim, L., Acres, K., et al. (2005). Selecting among competing alternatives: Selection and retrieval in the left inferior frontal gyrus. *Cerebral Cortex*, *15*, 1723–1735.
- Murray, M. M., Camen, C., Gonzalez Andino, S. L., Bovet, P., & Clarke, S. (2006). Rapid brain discrimination of sounds of objects. *The Journal of Neuroscience*, *26*, 1293–1302.
- Murray, M. M., Camen, C., Spierer, L., & Clarke, S. (2008). Plasticity in representations of environmental sounds revealed by electrical neuroimaging. *Neuroimage*, *39*, 847–856.
- Narain, C., Scott, S. K., Wise, R. J. S., Rosen, S., Leff, A., Iversen, S. D., et al. (2003). Defining a left-lateralized response specific to intelligible speech using fMRI. *Cerebral Cortex*, *13*, 1362–1368.
- Noppeney, U., Josephs, O., Hocking, J., Price, C. J., & Friston, K. J. (2008). The effect of prior visual information on recognition of speech and sounds. *Cerebral Cortex*, *18*, 598.
- Obleser, J., Eisner, F., & Kotz, S. A. (2008). Bilateral speech comprehension reflects differential sensitivity to spectral and temporal features. *Journal of Neuroscience*, *28*, 8116.
- Olsson, U. (1979). Maximum likelihood estimation of the polychoric correlation coefficient. *Psychometrika*, *44*, 443–460.
- Opolko, F., & Wapnick, J. (1987). *McGill University Master Samples [Compact Disc]*. Montréal, Québec, Canada: McGill University.
- Orgs, G., Lange, K., Dombrowski, J. H., & 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.
- Peirce, C. S., Houser, N., & Kloesel, C. J. W. (1998). *The essential Peirce: Selected philosophical writings (1893–1913)* (Vol. 2). Bloomington, IN: Indiana University Press.
- Pizzamiglio, L., Aprile, T., Spitoni, G., Pitzalis, S., Bates, E., D'Amico, S., et al. (2005). Separate neural systems for processing action- or non-action-related sounds. *Neuroimage*, *24*, 852–861.
- Pulvermüller, F. (2005). Brain mechanisms linking language and action. *Nature Reviews Neuroscience*, *6*, 576–582.
- Rao, V. R., & Katz, R. (1971). Alternative multidimensional scaling methods for large stimulus sets. *Journal of Marketing Research*, *8*, 488–494.
- Rizzolatti, G., Fogassi, L., & Gallese, V. (2002). Motor and cognitive functions of the ventral premotor cortex. *Current Opinion in Neurobiology*, *12*, 149–154.
- Rogalsky, C., & Hickok, G. (2008). Selective attention to semantic and syntactic features modulates sentence processing networks in anterior temporal cortex. *Cerebral Cortex*, *12*, 315–316.
- Romanski, L. M., & Averbach, B. B. (2009). The primate cortical auditory system and neural representation of conspecific vocalizations. *Annual Review of Neuroscience*, *32*, 315–346.
- Saygin, A. P., Dick, F., & Bates, E. (2005). An online task for contrasting auditory processing in the verbal and nonverbal domains and norms for college-age and elderly subjects. *Behavior Research Methods*, *37*, 99–110.
- Saygin, A. P., Dick, F., Wilson, S. W., Dronkers, N. F., & Bates, E. (2003). Neural resources for processing language and environmental sounds: Evidence from aphasia. *Brain*, *126*, 928.
- Schlösser, M. J., Aoyagi, N., Fulbright, R. K., Gore, J. C., & McCarthy, G. (1998). Functional MRI studies of auditory comprehension. *Human Brain Mapping*, *6*, 1–13.
- Schneider, T. R., Debener, S., Oostenveld, R., & Engel, A. K. (2008). Enhanced EEG gamma-band activity reflects multisensory semantic matching in visual-to-auditory object priming. *Neuroimage*, *42*, 1244–1254.
- Schneider, A., Benson, D. F., Alexander, D. N., & Schneider-Klaus, A. (1994). Non-verbal environmental sound recognition after unilateral hemispheric stroke. *Brain*, *117*, 281–287.
- Senkowski, D., Saint-Amour, D., Kelly, S. P., & Foxe, J. J. (2007). Multisensory processing of naturalistic objects in motion: A high-density electrical mapping and source estimation study. *Neuroimage*, *36*(3), 877–888.
- Specht, K., & Reul, J. (2003). Functional segregation of the temporal lobes into highly differentiated subsystems for auditory perception: An auditory rapid event-related fMRI-task. *Neuroimage*, *20*(4), 1944–1954.
- Spinnler, H., & Vignolo, L. A. (1966). Impaired recognition of meaningful sounds in aphasia. *Cortex*, *2*, 337–348.
- Stuart, G. P., & Jones, D. M. (1995). Priming the identification of environmental sounds. *The Quarterly Journal of Experimental Psychology Section A*, *48*(3), 741–761.
- Thierry, G., Giraud, A. L., & Price, C. (2003). Hemispheric dissociation in access to the human semantic system. *Neuron*, *38*(3), 499–506.
- Thierry, G., & Price, C. J. (2006). Dissociating verbal and nonverbal conceptual processing in the human brain. *Journal of Cognitive Neuroscience*, *18*(6), 1018–1028.
- Thompson-Schill, S. L. (2005). Dissecting the language organ: A new look at the role of Broca's area in language processing. In A. Cutler (Ed.), *Twenty-first century psycholinguistics: Four cornerstones* (pp. 173–189). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Thompson-Schill, S. L., D'Esposito, M., Aguirre, G. K., & Farah, M. J. (1997). Role of left inferior prefrontal cortex in retrieval of semantic knowledge: A reevaluation. *Proceedings of the National Academy of Sciences*, *94*(26), 14792–14797.
- Tranel, D., Grabowski, T. J., Lyon, J., & Damasio, H. (2005). Naming the same entities from visual or from auditory stimulation engages similar regions of left inferotemporal cortices. *Journal of Cognitive Neuroscience*, *17*(8), 1293–1305.
- Tversky, A. (1977). Features of similarity. *Psychological Review*, *84*, 327–352.
- van Petten, C., & Rieffelder, H. (1995). Conceptual relationships between spoken words and environmental sounds: Event-related brain potential measures. *Neuropsychologia*, *33*(4), 485–508.
- Vandenberghe, R., Nobre, A. C., & Price, C. J. (2002). The response of left temporal cortex to sentences. *Journal of Cognitive Neuroscience*, *14*(4), 550–560.
- Vignolo, L. A. (1982). Auditory agnosia. *Philosophical transactions of the royal society of London: Series B, Biological Sciences*, *298*(1089), 49–57 [1934–1990].
- von Hornbostel, E. M., & Sachs, C. (1914). Systematik der musikinstrumente Ein versuch. *Zeitschrift für Ethnologie*, *4–5*, 553–590.
- von Kriegstein, K., Eger, E., Kleinschmidt, A., & Giraud, A. L. (2003). Modulation of neural responses to speech by directing attention to voices or verbal content. *Cognitive Brain Research*, *17*(1), 48–55.
- von Kriegstein, K., & Giraud, A. L. (2004). Distinct functional substrates along the right superior temporal sulcus for the processing of voices. *Neuroimage*, *22*(2), 948–955.
- Weber, M., Thompson-Schill, S. L., Osherson, D., Haxby, J., & Parsons, L. (2009). Predicting judged similarity of natural categories from their neural representations. *Neuropsychologia*, *47*, 859–868.
- Wilson, S. M., Sereno, M. I., & Iacoboni, M. (2004). Listening to speech activates motor areas involved in speech production. *Nature Neuroscience*, *7*(7), 701–702.
- Zatorre, R. J., Bouffard, M., & Belin, P. (2004). Sensitivity to auditory object features in human temporal neocortex. *Journal of Neuroscience*, *24*, 3637.
- Zatorre, R. J., Chen, J. L., & Penhune, V. B. (2007). When the brain plays music: Auditory-motor interactions in music perception and production. *Nature Reviews Neuroscience*, *8*, 547–558.