Animate Concepts, Inanimate Sources?

The Relative Relevance of Conceptual and Acoustical Information in Similarity Judgments of Animate and Inanimate Sounds

John McDonnell

Supervisors: Stephen McAdams, Bruno Giordano

Towards the satisfaction of the requirements of COGS 402.

Final Submission: 18. January, 2007

Abstract

Recent studies using behavioral measures of similarity between acoustic events have had interesting results, but have been inconclusive with regards to the basis on which event-class discrimination is made. In the present investigation, we suggest that much of the difficulty has resulted from confusion about the basis on which participants make similarity judgments. Judgments may well be made on the basis of acoustic similarity, but there may also be a strong effect of *a priori* conceptual knowledge in participants' assessments of similarity. It is unclear whether these influences can be separated through biasing. Also, it is possible that participants have a default scheme for similarity when no biasing instructions are given. Finally, it has been suggested that different bases underlie the processing of sounds with animate vs. inanimate sources, which might influence behavioral measures of similarity. We have devised an experiment involving similarity judgments of both acoustical and verbal-label stimuli, using animate and inanimate stimuli, in an attempt to answer these questions. Our results suggest that participants are more highly affected by conceptual knowledge when judging animate compared to inanimate stimuli. Also, biasing was shown to increase the difference between acoustic similarity judgments and similarity judgments of verbal stimuli.

Introduction

A great deal of research has been invested in studying the perceptual organization of both speech and musical sounds. Unfortunately, relatively little headway has been made in investigating the phenomena of everyday sound-source perception. This is ironic, because it is the everyday assessment of sound-producing events (for information such as source and location) that probably provided the survival advantage necessary for hearing to have evolved in the first place. Gaver (1993) suggested two reasons for our continued ignorance about everyday sound perception, one historical and one theoretical. The first was the historical concern for the perception of harmonic sounds generated by musical instruments (to which one might add speech sounds), which often have little to do with many of the kinds of sounds that people encounter on a day-to-day basis. The theoretical reason regarded the prevailing paradigms in the study of perception, wherein perceptual research is focused on "primitive physical dimensions" as their physical stimuli, such as frequency or phase. Working within this framework, source identification is said to be a function of memory and problem solving, therefore a higher cognitive function, and ignored by psychophysicists.

In seeking to open the field to inquiry, Gaver (1988) performed a modified version of a study carried out by Vanderveer (1979) in asking participants to respond in depth to a series of environmental sounds. Vanderveer asked 57 participants to identify 30 everyday sounds over the course of a group session. Gaver investigated the free-identification of 17 everyday sounds with five participants, who were asked the question "What do you hear?" and prompted to go into as much detail as possible. Both studies found that participants easily identified surprisingly subtle information about the source of the sounds (such as the difference between footsteps walking up and down stairs), but their descriptions amounted almost entirely to a description of the sound sources. As a result, participants' inferences proved unhelpful in decoding what kinds of elements or dimensions were relevant to the their judgments.

Gaver (1988) interpreted these results as indicating that an ecological approach to everyday sound perception was in order, studying the information people draw from sounds in context. Essentially, he proposed that humans' auditory world is arranged around sources and source properties, rather than around sounds and sound properties. If the purpose of everyday listening is to be informed about the world, then it makes sense that source properties should, in general, dominate the human auditory experience. This ‘everyday’ listening mode was contrasted to listening for the 'quality' of sounds, perhaps for their aesthetic value, as in the works of John Cage. A taxonomy of sound percepts based on source material and interaction type between materials involved was proposed as a putative model for human auditory perceptual space. This taxonomy was not entirely divorced from acoustics; sounds in the ‘splashing water’ category, for example, almost certainly share similar acoustic properties. The point is instead that the acoustical differences defining boundaries between classes of events (e.g., water vs. solid) are likely to be the most perceptually salient.

—

To ground research in sound-source perception, researchers have developed a more detailed theoretical picture of auditory perception, envisioning a multi-stage process (McAdams & Bigand, 1993; McAdams, 1993). The theory is summarized as follows: a signal generated by a sound source is: 1) transduced into neural information by the peripheral processing system, 2) subjected to scene analysis processes which group components most likely to originate from the same source into a unitary auditory object (see Griffiths & Warren, 2004, for a helpful discussion of auditory objects), and 3) subsequently analyzed in terms of basic auditory features (e.g. brightness, pitch). The extracted features are 4) fully matched to the elements of a mental lexicon, containing a representation of previously experienced sources. This matching process can be otherwise defined as categorization, the "mental operation by which the brain classifies objects and events" (Cohen & Lefebvre, 2005). Categorization permits the naming of an event and the selection of an appropriate motor program for interacting with the object.

The matching process is conceptualized in terms of a bottom-up process, based on sensory information. Top-down influences include selective attention (relevant features) and learning processes (Palmeri, Wond, & Gauthier, 2004).

It is widely accepted that the process of categorization involves the comparison of a stimulus with a mental representation of the category. The nature of this representation varies across theories (cf. Smits, Sereno, & Jongman, 2006): it is the 'prototypical' member of a category in prototype theory (Rosch, 1973); all previously experienced members of a category in exemplar theories (Nosofsky, 1986); the region in a hypothetical featural space dividing categories in decision-bound theories (Ashby & Perrin, 1988); or the across-members variability and average of the category defining features in distribution theories (Miller, 1994). Evidence concerning categorization of synthetic non-speech sound supports a hybrid position among these theories, where the categorization process would possess features predicted by a decision-bound approach and by an exemplar and distributional approach (Smits et al., 2006).

—

Within the framework of the multi-stage model of auditory perception, the fundamental

problem of everyday sound source perception lies in characterizing and quantifying the sensory information upon which the matching process is based (i.e., the information driving perception). There are two approaches that have been taken in this direction. One is to examine the important acoustical factors for sound sources on a 'molecular' level, varying the mechanical properties of a sound source on a restricted number of dimensions, and quantifying related perceptual changes. The other approach is to look on a 'molar' level, using large numbers of heterogeneous sounds to see how they are organized and differentiated in perceptual space.

Although the study presented here is focused on sound source perception at the molar level, it is worth noting the successes of a 'molecular' approach. Experiments follow along the lines of Freed (1990), who examined participants' perceptions of the hardness of a mallet striking various metal objects. The goal of such research is to identify acoustical aspects of the sounds generated that predict participants' judgments about the properties of sound sources. Many more experiments have been done along these lines (e.g. Lakatos, McAdams, & Caussé, 1997; Giordano & McAdams, 2006). Rather than artificially changing one dimension of the acoustical stimulus (e.g., frequency), the interacting objects are varied along different dimensions (e.g., hammer hardness), which have more ecological validity.

These studies are useful, but can only be conducted for restricted classes of events. The purpose of examining sound source perception on a 'molar' level is to take a wider view, examining the phenomena of sound source perception with a wider variety of stimuli. In particular, ‘molecular’ studies cannot bear on the acoustical criteria by which humans discriminate between classes of events. The 'molar' approach is more poorly developed than the 'molecular' approach, but there have been several important studies to note, besides those of Vanderveer (1979) and Gaver (1988) mentioned before.

Ballas (1993) conducted a general study of everyday sound perception using a large set of stimuli, examining causal uncertainty. He looked at the correlates of uncertainty, including ecological frequency. He was able to show that a measure of causal uncertainty (Hcu) correlates very well with identification time. That measure was also weakly correlated to ecological frequency, whereby the latter "would appear to enhance sound identification but is not necessary for fast and accurate sound identification."

These data may be helpful in sorting out the dynamics of sound identification, but do not speak to the central question of seeking the acoustical features that define a class of events. This question has been addressed by several other researchers, generally using a similarity rating task, or a sorting task used to derive estimates of similarity (see Coxon, 1999, for a review of sorting procedures). Bonebright (2001) conducted a free-sort of everyday sounds searching for acoustical correlates to participants' sorting patterns, when the participants were instructed to think about the sounds rather than their sources. The analysis for acoustical correlates based on multidimensional scaling (MDS) was inconclusive, because acoustical correlates for individual dimensions were highly complicated. Few acoustic measures projected solely onto a single dimension.

This introduces an important question. Bonebright asked participants 'not to identify' the source object in making their similarity judgments. Afterwards however, he was unable to discern an acoustical explanation for these judgments. It is possible that the techniques employed were simply inadequate to find existing correlates, or that the scheme used for similarity is

simply very complicated, so that complex acoustical explanations are indeed necessary to explain it. But from an ecological perspective, it is not valid to instruct participants 'not to identify' the source object. In this case, if participants are unable to follow the instructions given, it conceivable that participants would fall into a kind of default processing mode, and sort based on whatever criteria comes most naturally, perhaps on the basis of conceptual knowledge.

Scavone, Lakatos, Cook, and Harbke (2001) approach the issue of perceptual space with a different task, using two conditions so as to address possible effects of context-based conceptual knowledge. Participants were asked to map sounds onto a two-dimensional visual space. They had two different sets of instructions: in one, participants were asked to sort based on a mental image of the scene conjured by the sound; in the other, participants were asked to sort based on the 'timbre,' or acoustical properties of the sounds. They reported that the mental image task produced markedly different results from the timbre condition: "In most cases, items clearly tend to group according to shared physical properties…in contrast to the timbre condition, where source-based clustering is not apparent." Neither detailed analysis nor discussion is provided, as the main purpose of the study was to provide an example of the usefulness of the experimental application for perceptual research, but they do seem to suggest that mapping by 'sound source' is reduced when biasing instructions are given. This would indicate that humans are not in fact 'locked into' conceptual hearing, but can attend to different aspects of the stimuli based on different instructions.

By far the most ambitious study bearing on these issues is that of Gygi, Kidd, and Watson (in press). Gygi et al. conducted an experiment in which participants were asked to judge the similarity of every possible stimulus pair among seventy stimuli, resulting in ten thousand comparisons for each participant. Three conditions were investigated: 1) acoustical stimuli, 2) imagined sounds based on typed verbal label stimuli, or 3) imagined events based on verbal stimuli. MDS solutions were similar for all three conditions, and Gygi et al. advanced several hypotheses for how this could be the case. It is possible that verbal descriptions of events elicit a strong acoustical memory component, causing participants in the verbal conditions to rate sounds according to acoustics. It is also possible that event properties influencing similarity judgments are likely to result in similar sounds, blurring the distinction between grouping by sound sources and grouping by sound properties. Finally, they suggest, contrary to Scavone et al. (2001), that participants may have judged sounds strictly based on their sources. It is not possible to distinguish between these hypotheses on the basis of the data and analyses in their study.

One issue here as well may be a theoretical confusion. Gygi et al. (in press) see these results as a confirmation Gaver (1993), in that subjects ‘heard sources’. It seems evident, however, that if subjects were unable to focus on acoustical properties of the sound stimuli as is suggested, they would be unable to extract sound source properties either. It would be more proper to say that these participants ‘heard concepts’. They were able to identify the sources, perhaps, but without acoustics, the sensory information would not have an opportunity to add anything more. For this reason, this conclusion will be taken to mean that subjects ‘hear concepts’, for the purposes of this paper, and may in fact be evidence against Gaver’s hypothesis.

An important factor to keep in mind with regard to Gygi et al. (in press) is that the participants in the sound condition were not given any kind of biasing instructions; Gygi et al. explicitly mention that "the instructions did not include any mention of a possible distinction

between event similarity and acoustic similarity." This leaves open the question of what the condition was actually measuring. Although the question of how participants judge sounds when given no biasing instructions is an interesting one, possibly related to a kind of 'natural processing' that humans fall into when given no further instructions, it does not answer the question as to whether participants are able to sort sounds based on sound properties, because they were not asked to do so.

A possible complication on the theme of 'natural processing' is the question of a possible distinction between animate and inanimate sounds. There is no *a priori* reason to believe that this is a meaningful division to draw. Nonetheless, a recent study points in this direction: Gèrard (1999) investigated semantic priming of animate and inanimate sounds. It was found that although priming sped identification of animate objects, it had no effect on the identification of inanimate objects. This was interpreted as implying a difference in the encoding of these types of events. Evidence for differential encoding of animate and inanimate sounds is also provided by neurological data from studies such as those of Lewis, Brefczynski, Phinney, Jannik, and DeYoe (2005) and Murray, Camen, Gonzalez Andino, Bovet, and Clarke (2006) showing temporal and spatial differences in the processing of animal vs. man-made sound sources. On the basis of Gèrard's study, these neurological differences in processing might underlie a difference in the extent to which conceptual knowledge is important in processing the two sound classes. It is interesting to note, therefore, that in Scavone et al. (2001), where different conditions were judged different, only inanimate sounds were used, and that in Gygi et al. (in press), where conditions did not have significantly differing results, a mix of animate and inanimate sounds was used. It is possible that in this latter study, the distinction between animate and inanimate sounds was so pronounced that it became important in all conditions, causing high similarity among conditions and possibly even biasing all participants to focus on conceptual distinctions in their ratings.

Finally, there is the question of the validity of performing an experiment in which each participant must make ten thousand different similarity judgments. Issues of fatigue become a major factor, as does the real possibility that participants did not retain consistent criteria for judgment over the course of the study. This issue makes interpretation difficult, because it is difficult to hypothesize what effect this might have on responses. For example, it is easy to imagine that the participants in the imagined sound and imagined source conditions could become increasingly lax with regard to their instructions, and end up sorting less consistently on the basis of conceptual or sensory similarity in general, perhaps defaulting to a kind of more natural 'general similarity'.

—

Given the failure of similarity measures of acoustical stimuli to yield strong acoustical correlates in Bonebright et al. (2001), combined with Gygi et al.'s suggestion that conceptual knowledge is of foremost importance, it is unclear to what extent participants' similarity ratings of acoustical stimuli are based on acoustical properties of the stimuli, as opposed to *a priori* conceptual knowledge about the events occurring. It is easy to imagine that sounds occurring consistently in the same context, e.g. farm sounds, might be grouped as more similar without respect to acoustical or visual representations. The same goes for semantic links. This makes these latter attributes of conceptual knowledge more interesting than the mnemonic sensory

representations of sound source events, because they are likely to diverge strongly from the acoustic information derived from listening.

These factors may operate at two levels. First, they may form the basis of a 'natural bias' in similarity ratings: participants who are instructed simply to rate or sort acoustical stimuli on the basis of an abstract notion of 'similarity' may find it most natural to rate on the basis of certain kinds of conceptual knowledge, rather than on the basis of acoustical properties. Second, and perhaps more importantly in terms of methodology, is the question of whether or not it is possible for participants to filter out different factors in their similarity judgments. If participants are incapable of making similarity judgments on the basis of acoustical information alone, this has a theoretical consequence: the conceptual representations of events heard is strong enough that acoustical properties may not have behavioral consequences at all. If this is so, it seems improbable for subjects that subjects ‘hear’ sound sources, as Gaver (1993) and the ecologists hypothesize, because source properties arise from acoustical properties. From a methodological standpoint, this would indicate that asking participants for similarity judgments may be infelicitous toward an understanding of how everyday sounds are processed by humans.

—

The experiments reported here were intended to test the effects of conceptual knowledge on acoustical similarity ratings for everyday sounds, controlling for the various factors mentioned.

In the first, preliminary experiment, 140 naturally-generated acoustical stimuli, half categorized as 'animate' and half as 'inanimate' and selected to be diverse in terms of sources while maintaining conceptual connectedness, were presented to participants. Participants were asked to provide concise identifying labels for the event causing the sound, using one verb and one or two nouns. The responses were assessed in terms of verbal agreement, conceptual agreement, and correctness. Data analysis provided a measure of identifiability and identification labels for each sound.

Highly identifiable sound events were investigated in a second experiment. Animate and inanimate events were investigated in separate experimental sessions, using a hierarchical sorting technique. A full similarity matrix was collected for a large set of stimuli in a relatively short time, thus increasing the likelihood of consistent judgmental criteria across the entire experimental session.

Three conditions were investigated. In one, referred to as *word-conceptual*, participants were asked to perform the sorting task on the basis of the similarity in the meaning of the identification labels collected during the free-identification experiment.. In another condition, *sound-unbiased*, acoustical stimuli were presented, and participants were instructed simply to group stimuli by similarity. In the third condition, *sound-acoustical*, participants were presented with acoustical stimuli and asked to sort them on the basis of their acoustical properties. These participants were also given a modified practice round to train participants to focus on the acoustical properties of sound sources independently of their associations to events or objects.

This arrangement was intended to answer both the question of natural bias, and the question regarding the ability of participants to separate acoustical and conceptual factors in

making similarity ratings. If the *sound-acoustical* condition is successful in causing participants to sort solely on the basis of the acoustical properties of sounds, then responses should be poorly correlated with the *word-conceptual* condition. With regard to natural bias, the association between the *sound-unbiased* condition and the other two should provide clues about the default processing scheme for everyday sounds. Another dimension is added to the analysis when the variation in results is considered between animate and inanimate sound sources. The counterbalancing of the presentation of the two sound sets also permitted an assessment of training effects.

The data collected suggest that overall, responses in the *sound-unbiased* condition are best associated with the *sound-acoustical* condition for inanimate sound sources, and with the *word-conceptual* condition for animate sound sources. Furthermore, there is a training effect, whereby the component favored in the first set is bolstered in the second set. These factors combined indicate that we hear animate concepts and inanimate sources, or at least that these are the more favored representations in the unbiased condition. Finally, the lowest correlations are between the *sound-acoustical* and the *word-conceptual* conditions, indicating a greater relative independence of responses in these conditions.

Experiment 1: Free Identification

A large set of environmental sounds was investigated with a free-identification task. The objectives were to provide empirically derived verbal labels for each sound, to be used in the word-conceptual condition in Experiment 2, and to measure the sound events' identifiability.

Stimulus selection

140 stimuli were used in the free identification task: 70 judged 'animate' and 70 'inanimate' using criteria defined below. They were drawn primarily from a database of royalty-free sound effects (The General 6000 from Sound Ideas), complemented by additional online and published resources (Elliott, 2005), and with a database of musical instrument tones (Opolko & Wapnick, 1987).

Stimuli were selected to maximize acoustical diversity, and conceptual connectedness of sound events. All available sounds were labeled on the basis of a variety of context- and sound source-related properties. Labelings were based on previously published classifications of sound events (e.g., the taxonomy of inanimate sound sources found in Gaver 1993), but are not to be regarded as a rigorous attempt to build a comprehensive classification system for non-speech sound events.

Certain categories of sounds were excluded from consideration: speech samples, because of their ability to communicate conceptual information independent of their origin; synthetic sounds (e.g., computer alert sounds), because of the ambiguity of their source; Foley sounds; 'complex' sounds in the sense of Gaver (1993), resulting from multiple types of interaction (e.g., a ball rolling on the table which then falls off of it and hits the ground); and 'hybrid' sounds, also in the sense of Gaver (1993), resulting from the interaction of matter in multiple states (e.g.,

pouring a carbonated drink).

There are two bases by which one might define animate sounds. A source animacy criterion would classify a sound as animate if it results from the vibration of a substance or object that is part of a living animal. On this basis, shaving would be classified as an animate sound, because hair is part of a living animal, but a shod person's footsteps would be classified as inanimate, because shoes are not part of a living animal. A source agency criterion would classify a sound as animate if it resulted from the voluntary action of an animal. This would include human footsteps, but exclude natural sounds, such as the wind. It was decided that animate agency was far too uninformative (most sounds, such as 'balloon popping', were almost certainly created by an animate human agent, but it does not seem appropriate to include these as animate). It was ultimately decided that animate agent sounds involving locomotion (e.g., footsteps) or alimentation (e.g., dog lapping up water), which strongly suggested the presence of an animate agent, would be included in the animate set in addition to all animate source sounds.

Inanimate sounds were defined as those sounds resulting from the vibration of an object not part of a living animal. This definition included most animate agent sounds, and technically would also include some alimentation and locomotion sounds described above. These latter were included with the animate sounds instead, however, on an *ad hoc* basis, because they are highly suggestive of an animal. It was also decided that musical sounds would form their own class of inanimate sounds.

This produced three basic classes of sounds: animate source, inanimate-musical, and inanimate-non-musical. Further distinctions were made within these categories:

- **Animate source**
  - **Taxonomical class**: The source animal's taxonomical class. *Levels*: Amphibia, Aves, Insecta, Mammalia, and an extra classification for H. sapiens.
  - **Vocalization**: The generation of the sound involves the activity of the vocal apparatus of the animal. Levels: vocalization or non-vocalization.
  - **Communication**: the sound is used in communication. *Levels*: Communicative (e.g., rattlesnake's rattle) or non-communicative.

- **Inanimate-nonmusical sound source** (cf. Gaver, 1993)
  - **State of matter**: state of matter of vibrating object. *Levels*: aerodynamic, combustion, electric, liquid, solid.
  - **Interaction type**: nature of the interaction process which results in sound generation. An extension of Gaver (1993), each state of matter has several possible interaction types. *Levels*: *Aerodynamic*: continuous (e.g., horn honking), explosion, steam, whoosh (e.g., swinging golf club), wind. *Combustion*: simple (e.g., kitchen stove.), crackling (e.g., camp fire). *Electric*: explosive (e.g., thunder), continuous (e.g., electric light hum). *Liquid*: bubbling, dripping (e.g., water dripping into sink), flowing (constant volume of sounding liquid, e.g. river flowing), pouring (varying volume of sounding liquid, e.g., pouring a glass of wine), sloshing (shaking a liquid inside a closed container), splashing

(e.g., wave breaking), spraying (e.g., spray paint). *Solid*: deformation (e.g., paper crumpling), impact (e.g., rock falling), rolling (e.g., skateboard), scraping (e.g., removing sword from sheath).

- **Inanimate-musical sound source**
  - **Musical instrument family** (cf. von Hornbostel & Sachs, 1914). *Levels*: aerophone (wind instrument), chordophone (string instrument), idiophone (main body of instrument is sound-generator, e.g. vibraphone), membranophone (vibrating object is a membrane, e.g. a drum)
  - **Type of excitation**: nature of process by which sound is produced (cf. Hajda, Kendall, Carterette, & Harshberger, 1997). *Levels*: impulsive (e.g., plucked string), continuant (e.g., bowed cello), multiple impacts (e.g., tamborine).

Additional ad-hoc categories were created, to permit the inclusion of certain animate-agent sounds in the animate category:

- **Animate Agent** (all sounds classified in one of two ways)
  - **Locomotion**: Sound generated in agent movement (e.g., footsteps)
  - **Alimentation** Sound generated during agent's consumption (e.g., dog lapping up water.)

From here on, those sounds with animate sources, in addition to the ad hoc categories, will be referred to as 'inanimate' sounds. All other sounds will be referred to as 'animate'. Finally, several context-based distinctions were made:

- **Animate sources**: house (e.g., cat meowing), indoors (e.g., man snooring), toilet (e.g., brushing teeth), farm (e.g., chicken), sea (e.g., seagulls), wild (wild animal besides above), anywhere (none of the above).
- **Inanimate, non-musical sound sources**: casino (e..g, dice shaking), party (e.g., noisemaker), kitchen (e.g., kettle boiling), toilet (e.g., shower running), construction (e.g., sawing), office (e.g., stapler), store (e.g., operating cash register), indoors (indoor sounds besides the above, e.g., water dripping in sink), maritime travel (e.g., canoe paddling), bicycle travel, rail travel, travel (travel sounds besides the above, e.g., close luggage lid), military (e.g., gun firing), sea (e.g., ocean waves), wild (e.g., desert wind), sport (e.g., ping pong), anywhere (anything not mentioned above, e.g., keys jingling).

Some of these distinctions were *ad hoc*, when technically imprecise distinctions were deemed helpful for the goals of this experiment. For example, purely acoustical evidence would not support a distinction between continuous and wind aerodynamic sounds. Liquid sounds, similarly, are actually generated by the vibration of gaseous medium forming a changing surface with the liquid. The communicative and musical distinctions may have technically been more properly classified as contextual differences than acoustical differences. Finally, some exceptions to the rules were forced; for example, because of a dearth of identifiable fire sounds, a crackling combustion event was used, although this sound should be a hybrid event, as the crackling results from overheating water trapped inside the wood. It should be emphasized that these distinctions were drawn strictly for the sake of diversity of stimuli, and do not reflect an attempt to create a

rigorous taxonomy of sound events.

Seventy animate and seventy inanimate sounds were extracted randomly from the database, so that all cross-categories (e.g., reptile verbal communication) defined by the labeling scheme were evenly represented. Randomly extracted sounds judged unidentifiable or unrepresentative of their category were replaced by alternative sounds that better met those criteria. When adequate sounds could not be found for certain categories, these categories were eliminated.

Selected sounds were edited to a minimal duration, while still allowing the event to "unfold naturally" (Marcell et al., 2000; Port, Cummins, & McAuley, 1995). By this, we mean sounds were edited to the minimal duration necessary to maintain a signal representative of the generating sound source (e.g., a single footstep is not enough to represent the concept "footsteps".). Signal levels were left unmodified from the recordings.

Procedure:

Participants were first exposed to a familiarization procedure, in which each of the sounds was played in random order, separated by 100 ms intervals of silence. Afterwards, they carried out an identification task on each of the stimuli. Our study was concerned with proper sound source identification of the sound-producing events, rather than the type of language used by the participant to identify the event, as in Vanderveer (1979) or with 'acoustical plausibility,' as in Marcell et al. (2000). For this reason, concise responses demonstrating sufficient discerning of the event were sought. Participants were given three fields labeled "Verb", "Noun #1", and "Noun #2", and prompted to identify the sound as quickly and accurately as possible using a noun and a verb. An extra noun was permitted if necessary, so that it would be possible for participants to identify the agent, action, and object of an interaction. Participants were also told not to give responses that were vague (the example given being "thing") nor to give more than one word per field. Blank responses were not allowed. They advanced through the experiment at their own pace, repeating stimuli as desired using a "play" button. Participants were offered a 5-minute break one hour into the experiment.

Stimuli were stored on the hard disk of a Macintosh G5 Workstation, equipped with a M-Audio Audiophile 192 S/PDIF interface. Audio signals were amplified with a Grace Design m904 monitor system and presented through Sennheiser HD280 headphones. Participants sat inside an IAC double-wall soundproof booth. Signal peak level ranged from **XX** to **XX** dB SPL.

Participants:

Twenty-one musically-trained participants took part in the experiment (11 female, age range 18-42, median age: 21). An audiogram was administered to all participants. Both ears and frequencies from 125 - 8000 Hz were tested for threshold. Hearing thresholds never exceeded normative values by more than 15 dB (ISO 389-8, 2004; Martin & Champlin, 2000). Data from

one participant who consistently confused the noun and verb fields were not considered.

Results:

Previous studies on sound identification had focused on two different criteria for assessment. In some cases the modal response was used (Marcell et al., 2000), providing a measure of participants' agreement, while in other cases the known correct response was used (Vanderveer, 1979; Ballas, 1993), providing a measure of participant correctness. We were interested in participants' agreeing conceptually on the correct response, so three levels of agreement were assessed: 1) verbal, 2) conceptual, and 3) correctness.

Responses in the noun and verb fields were considered separately in the verbal part of the analysis and the conceptual analysis, but in the conceptual analysis the response as a whole was sometimes necessary for interpretation of conceptual equivalence. This is the case in the example 'baby crying' versus 'geese crying'; a different kind of 'crying' is implied by these two responses. Occasional confusions between the verb and noun fields were corrected.

For verbal agreement, responses involving the same root were considered to be equivalent. This included obvious misspellings, and alternate spellings of the same word (e.g., 'mew' and meow'). All verbs were converted to their corresponding gerunds (i.e., 'played' became 'playing), and all nouns were changed to the singular form. A measure of verbal agreement was computed for each response category (noun 1, noun 2, and verb). This was computed by dividing the number of participants using the modal root in each response category by the total number of participants responding, to derive the average number of agreeing responses. These data are presented in Figure 1.

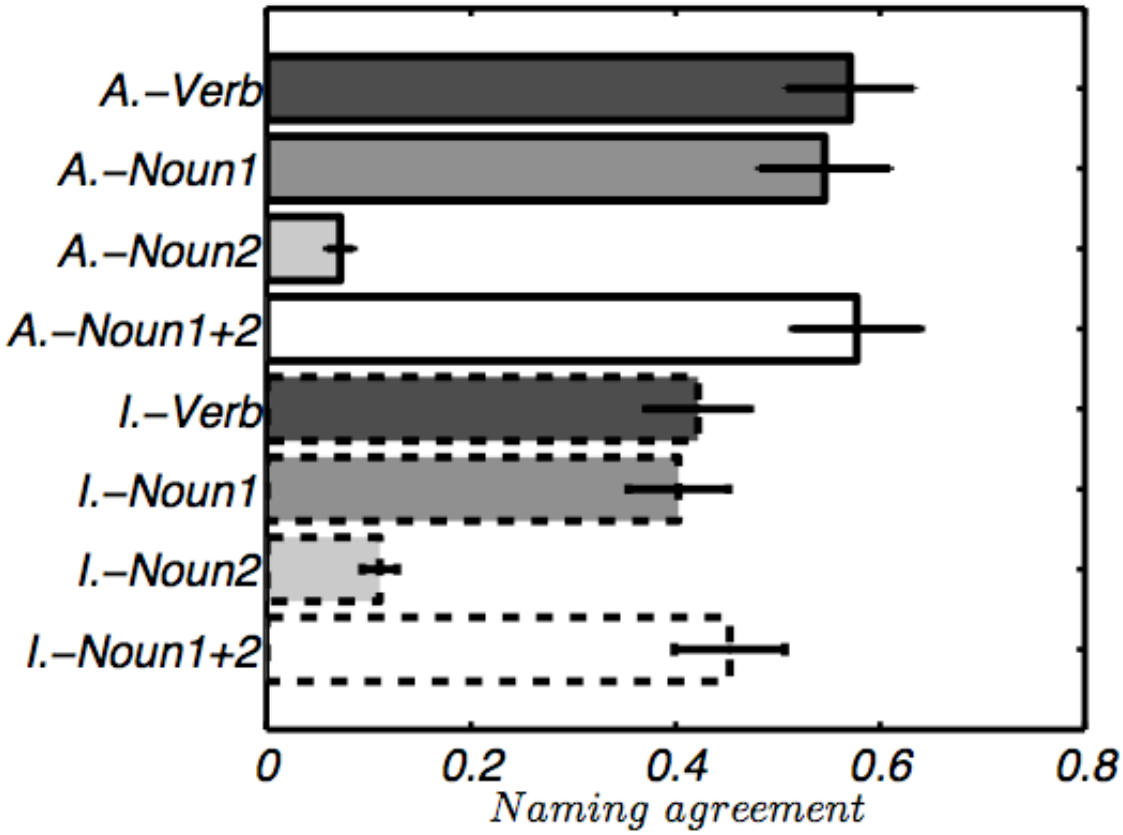

Figure 1. Average naming agreement for Animate (A.) and Inanimate (I.) sets, for the different input fields. A value of 1.0 would indicate total agreement among participants. Error bars bracket a 95% confidence interval for the average across-stimuli naming agreement.

Very low naming agreement for noun 2 reflects the fact that most participants left this field blank (75% of noun 2 responses; blanks were not permitted in the other fields). In addition, differences between the first and second noun fields could not be interpreted, because participants were not instructed to treat them differently. Therefore, the fields were assessed together in the following stages of this analysis.

For conceptual agreement, a conservative modification of Marcell et al., 2001 was used. For responses to be marked as correct, they needed to be at least at the level of specificity of the modal response. Therefore, if a plurality of respondents made reference to an 'eagle', 'bird' is not acceptable. Responses were compared to the modal verbal response, and were considered to be in conceptual agreement if they met any of the following criteria:

1) Synonyms: e.g., "Clapping" and "applauding."

2) Acoustically plausible conceptual coordinates: e.g., "hawk" and "eagle."

3) Conceptual subordinates: "eagle" may stand in for "bird", but not vice versa.

4) Part of the modal response: "mouth" is a part of a person, unless evidence is given that an animal mouth was intended (e.g. actual response: "chew" "carrot" "mouth" was judged to imply a human chewing a carrot in his/her mouth).

5) Modal response is implied: "brushing teeth" implies "toothbrush."

Criteria for conceptual agreement were stricter than those of Marcell et al. (2001). Marcell et al. (2001) permitted any sound judged as an "acoustically precise alternative" to be considered a correct response, using the example of "fish tank air pump" for "water bubbling." Since the concern in this study was to ensure that participants agreed conceptually, such responses were only considered correct when the two options were conceptual coordinates, so that acoustics and conceptual aspects were both preserved. Thus "fish tank air pump" would not have been an acceptable replacement for "water bubbling."

Responses were scored as correct if both noun and verb agreed conceptually with the correct answer, and were at least as specific as the modal noun and verb responses (i.e., the same level of specification was required for each sound in the conceptual and identification analysis). Identifiability scores were given by the proportion of participants providing a correct response. The distribution of identifiability levels for animate and inanimate sound sets is presented in Figure 2.

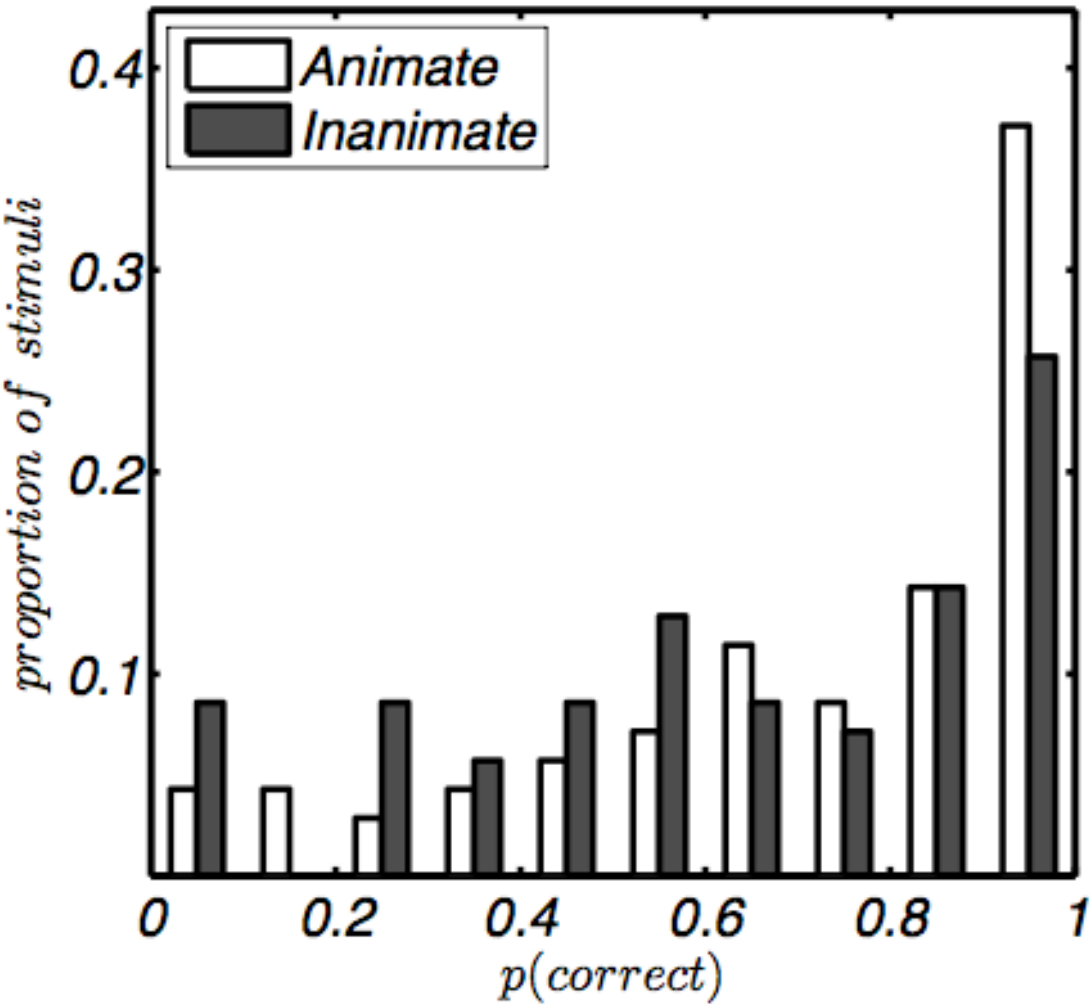

Figure 2. The distribution of identification scores for sounds in the animate and inanimate sound sets.

A Wilcoxon rank sum test for equal medians was computed to test for differences between the animate and inanimate sets in each of the three performance measures. Averages using the proportion of responses in agreement with the modal response were used for naming and conceptual agreement. Naming and conceptual agreement scores were averaged across stimuli and response fields; identification performance scores were averaged across stimuli. The proportion of responses judged 'correct' based on the criteria described above was used for the identification measure. Results can be found in Figure 3. These tests revealed a significant difference ($p<.001$) in means of verbal agreement between animate ($p_{agreement}=.58$) and inanimate ($p_{agreement}=.45$) stimuli, and a less significant difference for conceptual agreement ($p=.032$). In both cases, agreement was lower in inanimate sounds than in animate sounds. These results were interpreted to mean that the vocabulary for describing inanimate sounds was more diverse than

for describing animate sounds. This seems likely both because they were probably more heterogenous, and because identification performance for animate ($p_{correct}$=.71) and inanimate ($p_{correct}$=.65) groups did not vary significantly (p=.069), although there is a tendency for animate performance to be higher than inanimate performance.

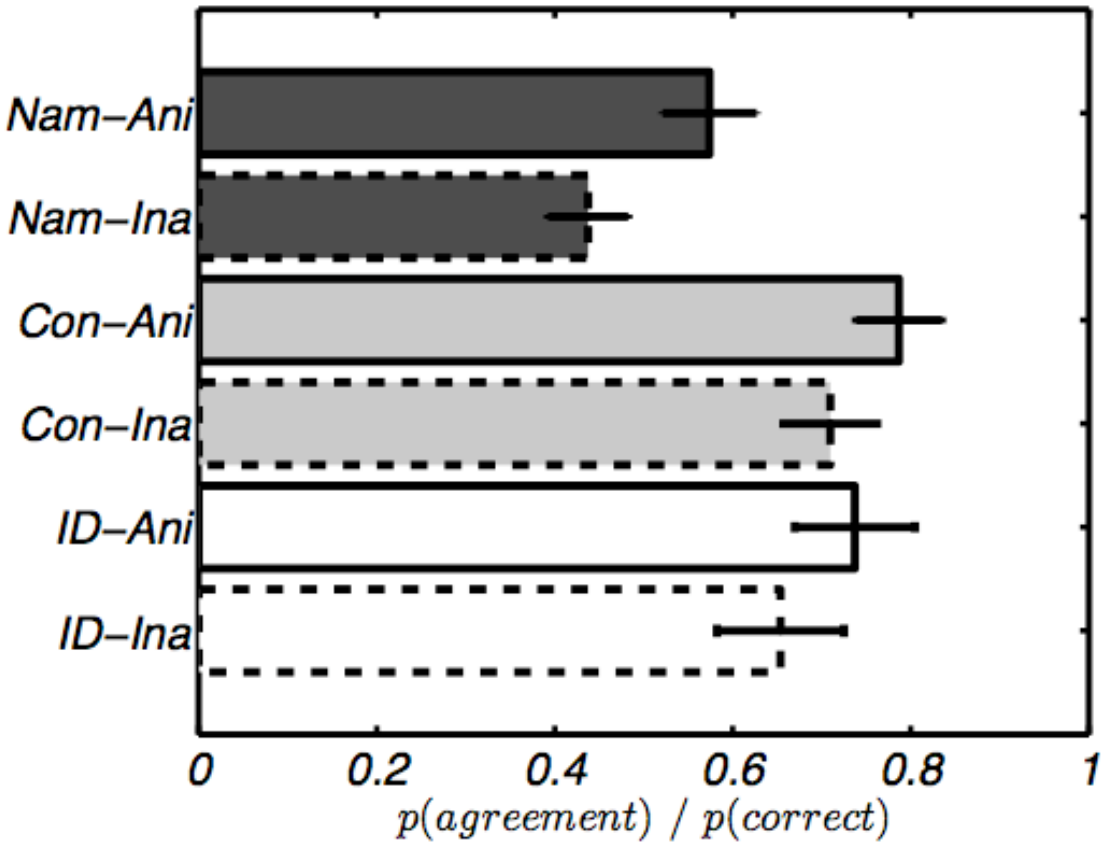

Figure 3. The agreement or correctness scores in the three analyses for the two sound sets. Naming (Nam) and conceptual (Con) agreement are averaged across response fields and stimuli, and identification performance scores are averaged across stimuli. Error bars bracket a 95% confidence interval for the mean value.

Discussion:

On the basis of participant responses, identifiability ratings were assessed and verbal labels derived for every sound event. Importantly, the identifiability distributions of the relatively identifiable animate and inanimate sounds were close enough to permit their harmonization during stimulus selection for Experiment 2. Overall relatively low identifiability scores are attributable both to the strictness of criteria used (with conceptual agreement required), and to the number of stimuli and their heterogeneity; sometimes it was difficult to find a sound from a certain category that seemed plausibly identifiable. The strong difference in verbal agreement

between animate and inanimate sounds may be an indication of the heterogeneity of the inanimate sound sources; participants simply had larger vocabularies for inanimate sources than for animate sources. This finding could also result from animate sounds being more iconic: there were many more animate than inanimate sounds whose verbal label was unanimously agreed upon. These could be cases where a concept invoked by an iconic sound was strongly associated with a single verbal description.

Experiment 2: Sorting

The sorting task was designed to respond to two basic questions. The first is whether it is possible for participants to make similarity judgments in a hierarchical sorting task solely on the basis of acoustical information from a sound-source, without using conceptual information. For this, two conditions were used. One, the sound-acoustical, was designed to facilitate a focus on the acoustical properties of the sound by means of a special training set. In the other, the word-conceptual condition, participants judged the similarity of the identification labels derived from the analysis of data in Experiment 1. A difference between these two conditions was assumed to measure the extent to which conceptual information and acoustical information could be separated by participants. Support for the effect of instructions is provided by Lakatos et al. (2001), who found such differences between conditions on their two-dimensional mapping task. We hypothesized that such differences would arise for both animate and inanimate sound sets.

The second question concerns the nature of a default response strategy in unbiased listeners (strictly conceptual vs.acoustical vs. conceptual+acoustical). This was assessed through a third condition, sound-unbiased, in which participants were asked to focus on the similarity among sounds, without further orientation of response criteria. On the basis of data from Gèrard (1999), we expected the default response strategy to vary between sound sets involving stimuli from animate vs. inanimate sources, with a possible preference for conceptually-based strategies in the animate sound set. Examining the correlations between the different conditions, a strong association between the sound-unbiased condition and the word-conceptual condition would indicate a default response strategy based on conceptual information. A similar association with the sound-acoustical condition would indicate greater employment of acoustic criteria in a default response strategy.

Stimulus selection criteria:

80 stimuli were selected from those investigated in Experiment 1: 40 animate and 40 inanimate sound events. All had an identifiability score of at least .5, meaning half or more of participants in Experiment 1, were able to identify them correctly. The use of stimuli with low identifiability would make the use of verbal labels invalid, since the labels would not be representative of the conceptual information, if any, conveyed by the sound, which may vary greatly among participants. The requirement was met by 49 of the inanimate and 58 of the animate sounds investigated in Experiment 1. The 40 most identifiable sounds meeting these criteria were

selected, while maintaining an even distribution of identifiability between the two sound sets. The distributions of identifiability in the animate and inanimate sets were not significantly different when compared using a Kolmogorov-Smirnov test (p=.893). The distributions are displayed in Figure 4.

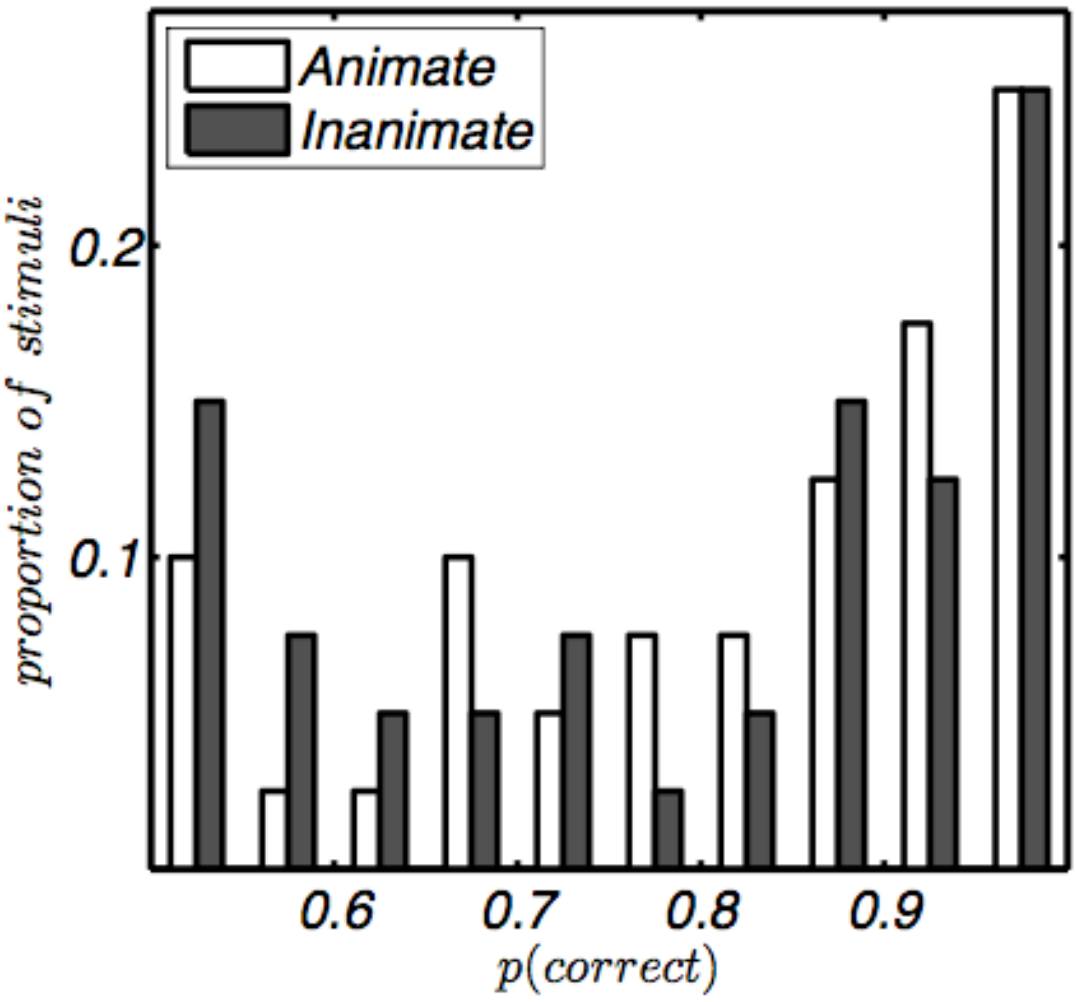

Figure 4. Distribution of identification performance scores for the animate and inanimate sound sets investigated in Experiment 2.

Verbal labels consisted of the modal verb and noun from the responses to Experiment 1. The verb was always in its gerundival form and first, followed by the noun in the singular or plural, depending on which was more often used by participants answering correctly (e.g. "panting dog"). Stimuli for which label extraction produced more than one noun or verb (as in cases of a tie) were excluded.

For the training round, 5 animate and 5 inanimate stimuli from Experiment 1, which were not selected for either of the primary blocks, were used. These were also identified by at least

50% of participants in Experiment 1. These were used in an unmodified form in the sound-unbiased condition, but were modified for use in the sound-acoustical condition. For this conditon, the stimuli were manipulated to diminish conceptual information they might contain through reducing their identifiability, while preserving some of their acoustical properties. This was designed to force an estimation based on acoustical properties rather than on the basis of evoked conceptual knowledge. For the word-conceptual condition, verbal labels were used. These were derived using the same process described above.

Signal manipulation was carried out using a method similar to the Event Noise Modulation technique presented in Gygi, Kidd, and Watson (2004). The amplitude envelope E(t) of the original sound was defined as:

$$E\,(t) = |x(t) + iH[x(t)]|$$

Where *H* is the Hilbert transform of *x* (t) (Hartmann, 1997). Hearing range amplitude fluctuations were attenuated by forward-reverse filtering *E* (t) using a third-order Butterworth filter with a low-pass cutoff frequency of 50 Hz (cf. McAdams, Chaigne, & Roussarie, 2004). Further acoustical properties of the sound event were estimated on the basis of the fast Fourier transform (FFT) of the entire signal (Hanning window). The spectral center of gravity (*SCG*) was defined as the linear-amplitude-weighted frequency average (assessed between 16 and 16000 Hz); the spectral mode (*SM*) was defined as the frequency of the highest amplitude FFT bin. The lower and upper spectral slope, respectively *LL* and *UL*, were estimated as the slope of the least squares line of the dB spectrum, from the lowest frequency to the *SM* and from the *SM* to the Nyquist frequency, respectively.

A random-phase signal was synthesized with the same duration as the original signal. Its *SM* corresponded to that of the original signal, with spectral level decaying linearly in dB, as a function of the distance from the *SM*. The *UL* and *LL* of the synthetic signal were recursively adjusted, starting from the value estimated in the original signal. At each step of the iterative procedure the current synthetic signal was multiplied by the amplitude envelope of the original signal, and the *SCG* of the resulting sound was calculated. If the *SCG* of the synthetic signal differed from the *SCG* of the original signal by more than 0.1%, the current *LL* and *UL* values were multiplied by the ratio of the *SCG* of the original with the *SCG* of the synthetic signal, and by the inverse of this quantity, respectively. If the SCG of the synthetic and original signals differed by less than 0.1%, the procedure was terminated. Convergence was reached in less than 27 recursive steps across all sounds.

Procedure:

For each of three blocks of stimuli, participants performed two tasks on a given set of stimuli. In the first task, after being presented all the stimuli in random order, they were presented with a set of stimuli represented as a set of numbered buttons on one side of the screen, and a set of boxes on the other side, signifying groups. Stimuli were presented when participants clicked on the

buttons. Participants grouped sounds together by dragging the buttons into the boxes. Empty boxes were not permitted. Once all the buttons were grouped and the participants confirmed their choices, the second task began. In this task, participants were presented the groups they had created in the first task as buttons, and were instructed to merge the two most similar groups. Once two groups were merged, a new group was formed containing the stimuli from the two merged groups. This new group, along with the other groups, was available to perform another merger. Participants merged groups together until there were only two groups in total.

These tasks were repeated for three blocks of stimuli. The first block was a practice run, with an experimenter present to ensure the instructions were clear. Ten animate and inanimate stimuli were sorted into 4 groups, which were then merged. The remaining two blocks consisted only of animate and inanimate stimuli, respectively, with the order counterbalanced between participants. In each block, 40 stimuli were sorted into 15 groups, which were then merged. At the beginning of each block, participants were presented with all the stimuli in random order, separated by a silence of 100 ms.

Each participant was placed in one of three conditions. In the sound-unbiased condition, participants were simply asked to group similar sounds together. The practice block consisted of 10 unmodified animate and inanimate sounds. The two primary blocks used unmodified acoustical stimuli drawn from Experiment 1, as described in the "Stimulus Selection" section of this report. The sound-acoustical condition was similar, but the practice round and instructions were modified. Participants were instructed to group sounds based on their acoustical properties, and the stimuli in the practice round were modified versions of the acoustical stimuli used in the sound-unbiased condition, as described above.

Stimuli were stored on the hard disk of a Macintosh G5 Workstation, equipped with an M-Audio Audiophile 192 S/PDIF interface. Audio signals were amplified with a Grace Design m904 monitor system and presented through Sennheiser HD280 headphones. Participants sat inside an IAC double-wall soundproof booth. Signal peak level ranged from XX to XX dB SPL.

Participants:

Sixty individuals took part in the experiment (XX females; age: XX-XX; median age: XX). Twenty were placed in each condition, for three instruction conditions: sound-acoustical, sound-unbiased, and word-conceptual. Presentation of animate and inanimate sound sets first was also counterbalanced between participants. As in Experiment 1, an audiogram was administered to all participants in the sound conditions. Both ears and frequencies in octave-spaced frequencies (125 - 8000 Hz) were tested for threshold. Hearing thresholds never exceeded normative values by more than 15 dB (ISO 389-8, 2004; Martin & Champlin, 2000). All participants in the conceptual condition reported having normal hearing and normal or corrected-to-normal vision.

Animate (presented first)

	Word-Conceptual	Sound-Acoustical	Sound-Unbiased
Word-Conceptual	1.0	.27	.63
Sound-Acoustical	.27	1.0	.46
Sound-Unbiased	.63	.46	1.0

Animate (presented second)

	Word-Conceptual	Sound-Acoustical	Sound-Unbiased
Word-Conceptual	1.0	.29	.48
Sound-Acoustical	.29	1.0	.61
Sound-Unbiased	.48	.61	1.0

Inanimate (presented first)

	Word-Conceptual	Sound-Acoustical	Sound-Unbiased
Word-Conceptual	1.0	.36	.28
Sound-Acoustical	.36	1.0	.60
Sound-Unbiased	.28	.60	1.0

Inanimate (presented second)

	Word-Conceptual	Sound-Acoustical	Sound-Unbiased
Word-Conceptual	1.0	.54	.54
Sound-Acoustical	.54	1.0	.82
Sound-Unbiased	.60	.82	1.0

Table 1. Spearman robust rank correlations between conditions.

Results:

Sorting produced participant-specific estimates of the dissimilarity of the investigated stimuli. The dissimilarity between two stimuli was considered to be given by the merging level (**1-15**) at which they were joined. Overall dissimilarities were assessed using the median dissimilarity between two stimuli in the same condition to ensure a robust measure of dissimilarity. Conditions were compared using a robust rank correlation. The results are shown in Table 1.

Three important effects were found. First, the lowest overall agreement between conditions is that between the word-conceptual and sound-acoustical conditions, indicating a degree of relative independence between the two conditions, especially with the animate sounds. Second, there was a training effect with two components. One component was that agreement between all conditions increased in the second block of stimuli, suggesting that participants became more proficient. The other component was that whatever two conditions were best correlated in the first condition became disproportionately more correlated than would be expected in the second block. This means that when participants had the animate set first, the sound-unbiased responses for the inanimate block were in greater agreement with the word-conceptual condition than would be expected. Similarly, when participants were given the inanimate set first, sound-unbiased responses in the animate set were in greater agreement with the sound-acoustical condition than would otherwise be expected. . The final effect is that, independent of set order, the highest correlations for the unbiased condition are word-conceptual for the animate set, and sound-acoustical for the inanimate set. Statistical tests confirming these data have not yet been performed; the discussion here is based on qualitative assessment of the correlation data presented in Table 1. Further statistical verification is needed to validate these conclusions.

Discussion:

These data suggest an intriguing difference between similarity ratings of acoustical stimuli with animate and inanimate sound sources. When participants are not given biasing instructions, they tended to group more like they would when given words instead of sounds, despite the greater methodological similarity of the sound-acoustical condition to the sound-unbiased condition. Furthermore, this trend is reversed for inanimate sounds. Both effects were strong enough to bias participants on a second sorting task, **making the distinctions less strong or even reversing them**. Since acoustic properties probably coincide more strongly with sound-source properties than with conceptual information, these data suggest that in the 'real world', we hear inanimate sources, but animate concepts. Further study would be necessary to validate and fine-tune this claim.

Another important finding was that the lowest correlation was always between the sound-acoustical and word-conceptual conditions. This suggests that instructions and biasing were

successful in drawing participants' attention away from conceptual factors, so that judgment was more strongly based on the acoustical properties of the acoustical stimuli. The two were never completely independent, but because events with similar conceptual representations would be expected to have some acoustical similarity, we would not necessarily expect these conditions to have complete independence. Analysis of acoustical correlates would be necessary to confirm the validity of these conclusions.

The hierarchical sorting method employed was shown to be effective at retrieving rich similarity data quickly and easily, without the same risks of participant fatigue associated with pair-wise similarity ratings.

These results indicate that different processing systems are involved in the similarity judgment of sounds in the animate and inanimate groups, and that the method presented here of biasing participants to attend to the acoustical elements of the signal seems to have been at least an improvement over instructions asking simply for similarity.

General Discussion:

The data from the present study indicate, perhaps unsurprisingly, that responses to a similarity-based task involving a large set of everyday sounds vary significantly depending on the set of sounds investigated and the instructions used to guide participants' responses.

A distinction is suggested between animate and inanimate sound sources. It is not clear why a distinction should be made between these sets, but further study could help clarify the exact distinction. Two possible important differences between the groups are the possible symbolic nature of animate sounds, and the homogeneity of animate sounds as a group. Animate sounds have a much more narrow definition than inanimate sounds, and are thus probably more homogeneous. If animate sounds were processed as symbols rather than being analyzed for their content, this might explain the close relationship between the sound-unbiased condition and the word-conceptual condition, which is presumably a model for similarity of conceptual content. However, there were also symbolic sounds included in the inanimate set, such as train-crossing bells. With regards to homogeneity, animate sounds may be less acoustically homogenous than the inanimate sounds, but should be at least as conceptually heterogeneous, thanks to the stimulus selection process. In this case, conceptual distinctions may become much more salient than acoustical ones.

One area of great interest is that of neuroanatomical correlates. A functional analysis of tool sounds versus animal sounds (largely vocalizations) performed by Lewis et al. (2005), found that many areas are preferentially activated in the tool sounds, but only three were preferred for the animal sounds. Of these, one was the left Brodmann's area 22, home to Wernicke's area, which is believed to underlie language comprehension. This result mirrors the result presented in the present study, whereby unbiased participants respond to animate stimuli similarly to their

verbal labels. Lewis et al. suggest that this has to do with animate vocalizations having similar properties to speech sounds, causing them to be processed along similar pathways. This neuroanatomical relationship between animal vocalizations and language comprehension could be a major piece of the puzzle in explaining a behavioral animate-inanimate distinction.

A helpful strategy for assessing the effect might be to refine the exact boundaries of the distinction. For example, the true distinction may be between iconic sounds and non-iconic sounds. This is intuitively plausible. To test this against the animate-inanimate distinction, a set of iconic inanimate sounds could be run against a set of less-iconic animate sounds in a task similar to that presented in Experiment 2. Further work could also be done to refine the animate-inanimate distinction. For example, one could legitimately ask the question whether animate agent sounds should have been included with the animate source sounds. Also, related to the question of iconic sounds, it is possible that the effect is most clearly demonstrated with animal communications, such as cats meowing, and less important in non-communicative sounds, such as people snoring. Gygi et al. (in press) found a major distinction between vocalizations and non-vocalizations in the MDS solution of their similarity ratings. This might support the hypothesis of Lewis et al. that the difference is caused by a similarity between human speech and animal vocalizations; however, a finding that sounds such as footsteps retained a conceptual bias would weaken this interpretation.

Because behavior varies by instructions given and this variation interacts with the sound-set given, being aware of specific variations may be critical to researchers studying heterogeneous sounds. Such information would permit the selection of a set of acoustical stimuli most likely to avoid conceptual interference and permit a more valid measure of perceptual similarity. One such distinction not addressed in the present study is the difference between sounds with higher or lower identifiability. Sounds with low identifiability have promise for doing away with conceptual baggage while retaining ecological validity. Such an analysis may be challenging, because participants presented with sounds with low identifiability may simply make an incorrect guess about the identity of the sound source, possibly producing conceptual effects anyway. Additionally, the experimental method presented here would not lead to interpretable results in sounds of low identifiability, because the verbal labels would not be representative of the conceptual content evoked by the sound events in listeners who could not identify them accurately. Despite these challenges, the potential of low-identifiability sounds in eliminating conceptual bias might make it worth the effort to run a well-controlled study.

One last point of interest might be inter-cultural variability. Chrea et al. (2005) conducted a similarity study of olfactory stimuli using a sorting task. Participants from France, the United States, and Vietnam were involved. Like the study presented here, some were placed in conditions where olfactory stimuli were presented, some in conditions involving verbal labels (translated to the appropriate language). Strong commonalities were demonstrated among the three cultures in the olfactory condition, whereas the verbal labels diverged. The inter-cultural nature of the study added a degree of validity, in that conceptual associations were probably

different cross-culturally, while responses to the stimulus remained relatively constant, suggesting a human perceptual basis rather than a cultural conceptual one. Such studies might prove to be of interest in the auditory domain as well.

Finally, the results presented here validate the methodology used in Experiment 2. The sound-acoustical condition was effective in achieving a greater degree of independence from the word-conceptual condition than the sound-unbiased condition, and is recommended for use in future studies. Similarly, the use of a hierarchical sorting task was effective at retrieving rich similarity data from participants without heroic efforts like those of Gygi et al. (in press).

Conclusions:

The results reported suggest a difference in processing between sounds indicating animate vs. inanimate sources. Essentially, listeners appear to hear inanimate sources but animate concepts. Also, the sound-acoustical condition seems to have been effective in drawing participants' attention to the acoustical properties of the acoustical stimuli, a claim which can be verified through analysis of acoustical correlates. Finally, hierarchical sorting was determined to be an effective measure of participants' similarity judgments among a wide variety of verbal and acoustical stimuli. These findings will be of interest to future researchers in the field both for their methodological value and theoretical repercussions.

References:

Ashby, F.G., & Perrin, N.A. (1988). "Toward a unified theory of similarity and recognition." *Psychological review*, 95(1): 124-150.

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(2): 250-267.

Bonebright, T.L. (2001). "Perceptual structure of everyday sounds: a multidimensional scaling approach," in *Proceedings of the 7th international conference on auditory display* (pp 73-78).

Chrea, C., Valentin, D., Sulmont-Rossé, C., Hoang Nguyen, D., & Abdi, H. (2005). "Semantic, typicality, and odor representation: a cross cultural study." *Chem. Senses*, 30: 37-49.

Cohen, H. & Lefebvre, C. (Eds) (2005). *Handbook of categorization in cognitive science*. Oxford, UK: Elsevier.

Coxon, A.P.M. (1999). *Sorting data: Collection and analysis*. Thousand Oaks, CA: Sage Publications.

Elliot, L. (2005). *A Guide to Wildlife Sounds*. Machanisburg, PA: Stackpole Books.

Freed, Daniel J (1989). "Auditory correlates of perceived mallet hardness for a set of recorded percussive sound events," J. Acoust. Soc. Am., 87(1), 311-322.

Gaver, William W.W. (1988). "Everyday listening and auditory icons," Unpublished doctoral dissertation, University of California, San Diego.

Gaver, William W. (1993). "What in the World Do We Hear," *Ecological Psychology*, 5(1) 1-29.

Gérard, Y. (1999). "Mémoire sémantique et sons de l'environment." Unpublished doctoral

dissertation, Université Paris 6.
Giordano B.L. & McAdams S (2006). "Material identification of real impact sounds: Effects of size variation in steel, glass, wood, and plexiglass plates," J. Acoust. Soc. Am. 119 (2), 1171-1181.
Griffiths, T.D., & Warren, J.D. (2004). "What is an auditory object?" *Nature Reviews Neuroscience*. 5: 887-892.
Gygi, Kidd, and Watson (in press). "Similarity and categorization of environmental sounds." Accepted for publication in *Perception and Psychophysics*.
Hajda, J.M., Kendall, R.A., Carterette, E.C., & Harshberger, M.I. (1997). "Methodological issues in timbre research," in I. Deliege & J. Sloboda (Eds.), *The Perception and Cognition of Music* (pp 253-306). London: L. Erlbaum.
Hartmann, W.M. (1997). *Signals, Sound and Sensation*. Woodbury, NY: AIP Press.
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.
Lakatos S, McAdams S, & Caussé R (1997). "The representation of auditory source charactaristics: simple geometric form," Percept. Psychophys, 59(8): 1180-1190.
Lewis, J.W., Brefczynski, J.A., Phinney, R.E., Jannik, J.J., & DeYoe, E.D. (2005). "Distinct cortical pathways for processing tool versus animal sounds." *The Journal of Neuroscience*, 25(21): 5148-5158.
Marcell, M.E., Borella, D., Greene, M., Kerr., E., & Rogers, S. (2000). "Confrontation naming of environmental sounds." *Journal of Clinical and Experimental Neuropsychology*, 22(6): 830-864.
Martin, F.N., & Champlin, C.A. (2000). "Reconsidering the limits of normal hearing." *Journal of the American Academy of Audiology*, 11(2): 64-66.
McAdams, S. (1993). "Recognition of sound sources and events," in McAdams, S. & Bigand, E. (Eds.), *Thinking in sound: the cognitive psychology of human audition* (pp 146-198). Oxford University Press.
McAdams, S., & Bigand, E. (1993). "Introduction to auditory cognition," in McAdams, S. & Bigand, E. (Eds.), *Thinking in sound: the cognitive psychology of human audition* (pp 1-9). Oxford University Press.
McAdams, S., Chaigne, A., & Roussarie, V. (2004). "The psychomechanics of simulated sound sources: Material properties of impacted bars." *J. Acoust. Soc. Am.*, 115(3): 1306-1320.
Miller, J.L. (1994)."On the internal structure of phonetic categories: a progress report." *Cognition*, 50: 271-285.
Murray, M.M., Camen, C., Gonzalez Andino, S.L., Bovet, P. & Clarke, S. (2006). "Rapid brain discrimination of sounds of objects," Journal of Neuroscience, 26(4): 1293-1302.
Nosofsky, R.M. (1986). "Attention, similarity, and the identification-categorization relationship." *Journal of Experimental Psychology: General*, 115(1), 39-57.
Opolko, F., & Wapnick, J. (1987). *McGill university master samples, [Compact Disc]*. Montréal, Québec: McGill University.
Palmeri, T.J., Wond, A.C., & Gauthier, I. (2004). "Computational approaches to the development of perceptual expertise." *Trends in Cognitive Sciences*, 8(8), 378-386.
Port, R.F., Cummins, F., & McAuley, J.D. (1995). "Naive time, temporal patterns, and human audition," in R.F. Port & T. Van Gelder (Eds.), *Mind as motion: explorations in the dynamics of cognition*, (pp. 339-371). Cambridge, MA: Massachusetts Institute of Technology.

Rosch (1973). "Natural categories." *Cognitive Psychology*, (4): 328-350.
Scavone, G.P., Lakatos, S., Cook, P., & Harbke, C.R. (2001). "Perceptual spaces for sound effects obtained with an interactive similarity rating program," in *Proceedings of the international symposium on musical acoustics*, Perugia, Italy.
Smits, R., Sereno, J., & Jongman, A. (2006). "Categorization of sounds." *Journal of Experimental Psychology: Human Perception and Performance*, 32(3): 733-754.
Vanderveer, N. J. (1979). "Ecological acoustics: Human perception of environmental sounds." Unpublished doctoral dissertation, Cornell University. *Dissertation Abstracts International*. 40/09B, 4543. (University Microfilms No. 8004002).
Van Hornbostel, E.M., & Sachs, C. (1914). "Systematik der Musikinstrumente: ein Versuch." *Zeitschrift für Ethnologie*, 4-5, 553-590.