Spectral and temporal cues for perception of material and action categories in impacted sound sources

Jens Hjortkjær[1,a)] and Stephen McAdams[2]
[1]*Oticon Centre of Excellence for Hearing and Speech Sciences, Department of Electrical Engineering, Technical University of Denmark, Ørsteds Plads 352, DK-2800 Kgs. Lyngby, Denmark*
[2]*Schulich School of Music, McGill University, 555 Sherbrooke Street West, Montreal, Quebec H3A 1E3, Canada*

(Received 31 July 2015; revised 17 June 2016; accepted 20 June 2016; published online 19 July 2016)

In two experiments, similarity ratings and categorization performance with recorded impact sounds representing three material categories (wood, metal, glass) being manipulated by three different categories of action (drop, strike, rattle) were examined. Previous research focusing on single impact sounds suggests that temporal cues related to damping are essential for material discrimination, but spectral cues are potentially more efficient for discriminating materials manipulated by different actions that include multiple impacts (e.g., dropping, rattling). Perceived similarity between material categories across different actions was correlated with the distribution of long-term spectral energy (spectral centroid). Similarity between action categories was described by the temporal distribution of envelope energy (temporal centroid) or by the density of impacts. Moreover, perceptual similarity correlated with the pattern of confusion in categorization judgments. Listeners tended to confuse materials with similar spectral centroids, and actions with similar temporal centroids and onset densities. To confirm the influence of these different features, spectral cues were removed by applying the envelopes of the original sounds to a broadband noise carrier. Without spectral cues, listeners retained sensitivity to action categories but not to material categories. Conversely, listeners recognized material but not action categories after envelope scrambling that preserved long-term spectral content. © *2016 Acoustical Society of America*. [http://dx.doi.org/10.1121/1.4955181]

[BLM] Pages: 409–420

I. INTRODUCTION

Sounds in a natural environment convey information about sound sources and sound-producing events. Several studies have examined the acoustical cues that allow listeners to identify the material of solid objects being struck together (Gaver, 1988, 1993; Lutfi and Oh, 1997; Kunkler-Peck and Turvey, 2000; Klatzky *et al.*, 2000; Lutfi, 2001; McAdams *et al.*, 2004; Giordano and McAdams, 2006; McAdams *et al.*, 2010). Most of this work has focused on single impact sounds where the object is freely vibrating after being struck. In this study, we investigated the similarity perception and categorization of materials across different types of impact actions (strike, drop, rattle), and of these actions across variation in the sound source material (glass, metal, wood). From a perspective of "ecological acoustics," Gaver (1993) argued that sound-generating actions are generally recognized via temporal cues, whereas material discrimination relies more on spectrotemporal information. However, there is still a lack of empirical evidence about the acoustic information used by listeners to recover either material or action information when different actions and materials are combined as in a natural context. Compared to single impact sounds, different types of impact introduce a much broader range of acoustic variation that is often encountered in ecological listening situations. For material discrimination, it is not clear that the acoustical features suggested in studies of single impact sounds are also used in the context of multiple impact sounds. In the context of combined actions and materials it is also not clear whether listeners might use particular features for particular combinations or instead favor potential cues that invariant across combinations. Metal and glass, for instance, may be well discriminated based on damping-related cues in the ideal situation of isolated single impacts where listeners have access to the full decay time. Spectral cues, however, may be more relevant if they can be used to discriminate materials with different patterns of envelope fluctuation introduced by different types of action. Similarly, it is unclear if there are invariant acoustical cues that allow listeners to discriminate action categories across the variation in the material of the manipulated object.

Previous work on material discrimination with *single* impact sounds has highlighted the relevance of damping-related cues. From a theoretical perspective, Wildes and Richards (1988) and Gaver (1993) have suggested that the frequency-specific decay of vibrations in struck solid bars or plates can be used to recover the material of the sound source independently of the objects' shape or manner of support. In later empirical studies, Klatzky *et al.* (2000), McAdams *et al.* (2004), McAdams *et al.* (2010), and Aramaki *et al.* (2011) used physical models of impacted plates or bars and found

[a)] Also at: Danish Research Centre for Magnetic Resonance, Centre for Functional and Diagnostic Imaging and Research, Copenhagen University Hospital Hvidovre, Denmark. Electronic mail: jhjort@elektro.dtu.dk that perceived similarity and categorization of wood and glass or wood and metal can indeed be described by damping parameters. The relevance of damping information has also been demonstrated with real impacted bars or plates (Giordano, 2003; Tucker and Brown, 2003; Giordano and McAdams, 2006). An influence of pitch or spectral cues has also been reported (Lutfi and Oh, 1997; Klatzky *et al.*, 2000; Giordano and McAdams, 2006; Avanzini and Rocchesso, 2001) but the role of frequency information in material discrimination is less clear. Although frequency content depends on the material of the sound source, spectral cues also vary with other important object properties such as object size and geometry. Investigating material identification of real impacted plates of variable size, Giordano and McAdams (2006) found that damping cues could account for identification of materials of vastly different mechanical properties (steel/glass vs wood/Plexiglas), but that listeners relied on frequency cues for fine-grained identification within these gross categories. Similarly, Lutfi and Oh (1997) showed that listeners made limited use of damping cues and relied on signal frequency for fine-grained material discrimination of simulated bars. McAdams *et al.* (2010) reported that perceptual ratings of similarity between impacted plates correlated with pitch, but also found that listeners ignored pitch information and relied on damping cues during material categorization of the same sounds. Thus, although a number of studies have demonstrated the relevance of acoustical measures related to damping, listeners appear to use different kinds of spectral or temporal information depending on the given task and stimulus context. It remains unclear, however, in which contexts the reliance on frequency information may represent an efficient perceptual strategy.

Spectral cues for material discrimination may become relevant when considering acoustic variation across a broader context of natural sound-producing events. In sounds that are not generated by a single impact, the decay time is typically a less efficient perceptual cue given that the object is not freely vibrating for a sustained period of time. Listeners may instead favor long-term frequency content if the associated spectral cues are more invariant across different types of action. Lemaitre and Heller (2012), however, found that material discrimination was generally poor for non-impact sounds (cylinders rolling, bouncing, or being scraped), arguing that material categories are only robustly identifiable when damping cues are available in single impact sounds. This led the authors to question whether material discrimination in general is based on auditory cues or whether the representations of material categories relies instead primarily on visual or haptic experience. Giordano *et al.* (2012), studying walked-upon materials, confirmed that material identification based on haptic feedback was indeed more accurate than in auditory perception. However, multiple impact sounds (rattling, bouncing, scattering, dropping, etc.) constitute a large and important part of natural sound events, and these sounds may still have characteristic spectral cues that could be used for material discrimination across different types of impacts. When different materials are combined with different types of impact as in the present study, listeners may either take advantage of the access to the frequency-dependent decay of partials in single impact sounds or they may favor spectral cues that are potentially more invariant across the variation in the sound-producing action.

Identification of sound-producing actions has received much less focus than material identification. This is despite the fact that action recognition has been reported to be notably accurate in auditory perception and potentially more robust than material identification (Lemaitre and Heller, 2012). Warren and Verbrugge (1984) investigated dropped objects and found that listeners accurately discriminated between bouncing and breaking events. Warren *et al.* (1987) found that listeners could estimate the elasticity of bouncing balls on the basis of the first onsets in a bouncing sequence. Gygi *et al.* (2004) using noise vocoding found that about 50% of a diverse set of environmental sounds could be identified without spectral cues altogether. Similarly, Warren and Verbrugge (1984) synthesized bouncing and breaking events using short spectrally averaged segments matching the impact onset pattern of the original sounds. Although a drop in categorization performance was observed, listeners continued to discriminate actions with high accuracy without spectral cues (86.7% for breaking and 93.0% for bouncing events). Although this confirms that temporal onset patterns can be used for action categorization, the potential role of spectral cues suggested by the drop in performance is less clear. It remains uncertain whether listeners can still use spectral information to infer information about action categories when temporal information is limited, e.g., in highly reverberant environments. Acoustical features that summarize temporal cues relevant for action recognition across variation in the sound object are also lacking.

Previous sound source perception studies have used identification/categorization judgments (Warren and Verbrugge, 1984; Giordano and McAdams, 2006; Lemaitre and Heller, 2012), similarity ratings (Klatzky *et al.*, 2000; McAdams *et al.*, 2004; Giordano, 2005), or both (McAdams *et al.*, 2010; Gygi *et al.*, 2007). Multidimensional scaling (MDS) of similarity ratings with environmental sounds has been used to determine perceptual representations of sound source properties and their acoustical correlates (Klatzky *et al.*, 2000; McAdams *et al.*, 2004; Gygi *et al.*, 2007; McAdams *et al.*, 2010). Even for similarity ratings of musical tones, clusters in regions of MDS spaces have been shown to be occupied by sounds produced by the same instrument family or manner of excitation (Giordano and McAdams, 2010). Only a few studies have made quantitative comparisons between representations derived from similarity and categorization data. Gygi *et al.* (2007) compared similarity ratings and free grouping responses of a range of environmental sounds and found similar gross clusters related to sound source attributes in MDS representations derived from the similarity and grouping data. Although similar category information emerged, only a moderate correlation between the two representations was reported. The similarity representations generally enhanced the spacing between clusters and correlated better with acoustic descriptors than the categorization data. McAdams *et al.* (2010) found that similarity ratings of simulated impacted plates were described by two MDS dimensions closely related to mechanical properties of the sound source with one dimension being related to wave velocity and pitch and another to damping and duration. When asked to categorize the material of the same sounds, however, listeners ignored pitch information and relied exclusively on damping cues. This result could suggest that listeners focus on acoustically salient features when judging similarity but shift their focus to more task-relevant cues during categorization. It is, however, also possible that similarity and categorization data are related via a non-linear mapping that would only be revealed by a model describing this relationship.

The relationship between similarity and categorization has been discussed extensively from a general psychological perspective. Formal models of categorization often rely strongly on concepts of similarity although different measures of similarity have been discussed (Tversky, 1977; Goldstone, 1994). Categorization may be viewed as a function of similarity in the sense that objects perceived to be similar with respect to some feature are grouped together (Sloutsky, 2003). Shepard (1987) argued that categorization can be described formally as an exponentially decreasing function of distance in perceptual similarity space. This allows categorization performance to be predicted from a non-linear mapping of distances in MDS space (Shepard, 1987; Krumhansl, 1978; Nosofsky, 1986). Alternatively, similarity may be viewed as a function of identification or categorization confusion in the sense that objects that are grouped together become perceptually similar (Goldstone *et al.*, 2001). General recognition theory (GRT, Ashby and Perrin, 1988) assumes that subjects place decision boundaries in a multidimensional perceptual space (where MDS is the particular case of a metric space, Ashby and Perrin, 1988) and associate a category response with a particular region. Perceived similarity is then assumed to be proportional to the amount of confusion between stimulus categories.

In the current study, we investigated material and action perception in impact sounds that simultaneously varied across combinations of impact action and material. Using the same stimulus set, listeners rated similarity (experiment 1) and categorized (experiment 2) the same actions with variation in materials and the same materials with variation in actions. First, we modeled the similarity data using weighted MDS to investigate whether common acoustical features could describe the ratings across listeners. Next, we modeled category confusion patterns with general recognition theory (GRT), a multidimensional extension of signal detection theory (SDT). We used hierarchical model selection to test whether listeners' sensitivity in material categorization was independent of the sound-producing action, and conversely whether action sensitivity was independent of variation in materials. For material categorization, this allowed us to investigate whether listeners favored cues that are independent of the type of impact, or whether they selectively use decay-related information when this is available in single impact sounds. Modeling of the similarity and categorization data also allowed us to compare these representations quantitatively. We tested how well MDS dimensions and associated acoustical features predicted category confusions and vice versa. Finally, we examined action and material categorization with the sound stimuli synthesized to remove either long-term spectral cues (using time domain scrambling) or temporal envelope cues (using noise vocoding). Again, we modeled categorization responses with GRT to investigate whether the spectral and temporal manipulations affected category combinations differently. Materials with highly different damping properties such as metal and wood could potentially still be discriminated with single impact sounds solely based on the decay rate, whereas the material of multiple impacts would be indiscriminable. On the other hand, given that listeners have been reported to favor spectral cues for discrimination of fine-grained material categories, removal of temporal cues also allowed us to uncover whether spectral cues can be used for material perception and categorization more generally, and whether they can be used without temporal information altogether.

II. EXPERIMENT 1: SIMILARITY RATINGS

A. Methods

1. Subjects

Twenty listeners (10 males, 10 females, aged 21–40 yr) participated in experiment 1 and were paid for their participation. All participants reported having normal hearing.

2. Stimuli

Eighteen sounds were recorded to represent three different action categories (strike, rattle, drop) and three different material categories (wood, metal, glass). Two different exemplars were recorded for each of the nine category combinations. In each material category we used a small number (5–8) of objects of comparable range of sizes (5–30 cm). We used solid objects of varying shape (rods, plates, and cylinders) made of aluminum, pine, or glass. Rattle sounds were produced by manually rattling all of the objects and drop sounds were made by dropping them on a solid floor from a fixed distance. Strike sounds were made by impacting two of the objects within a material category.

All sounds were recorded at a sampling rate of 44.1 kHz in an acoustically isolated room using two Audio-technica AT4041 microphones connected to a remote computer via an RME Fireface sound card. The duration of the signals ranged from 0.9 to 1.6 s.

Subsequently, the stimuli were equalized in perceived loudness by having five expert listeners adjust the level of 17 of the 18 stimuli to the perceived level of the remaining sound (a single strike on metal). For each stimulus, the level was set using the median of the five loudness estimates.

3. Apparatus

During the experiment, listeners were seated in an IAC 1202 double-walled sound booth. Sounds were generated from a Macintosh computer through an RME Fireface sound card. Presentation of stimuli and the collection of responses were controlled using Matlab. Sounds were presented over a set of Sennheiser HD250 Linear II headphones. Sound pressure level was measured at the headphones on a Brüel & Kjaer 2209 sound level meter (A-weighting, fast response) with an IEC

60318–1 ear simulator (G.R.A.S. Type RA0039). The peak level of the stimuli equalized in perceived loudness ranged from 66 to 75 dB sound pressure level (SPL).

4. Procedure

Listeners were asked to rate the similarity between pairs of sounds by adjusting a continuous on-screen slider scale marked "very similar" and "very dissimilar" at the extremes. Listeners could replay the sounds as many times as they wished before making a rating. They were instructed to use the full range of the scale in their responses over the whole experiment. The 18 different sounds were presented in random order at the beginning of the experiment to give listeners a sense of the range of variation. In the subsequent rating trials, each of the 171 possible sound pairs was presented. Identical pairs were included to control for subjects not attending to the task (i.e., rating identical pairs as dissimilar). An experimental session took approximately 30 min to complete.

B. Results

1. Multidimensional scaling

We used multidimensional scaling to represent the rated similarity between each sound pair in a low-dimensional space. We used a weighted distance model (Carroll and Chang, 1970) where the similarity between stimuli is modeled in terms of Euclidean distances. The weighted model assumes common underlying perceptual dimensions but different subjects may weight these dimensions differently

$$d_{jin} = \left[\sum_{r=1}^{R} w_{nr} (x_{jr} - x_{ir})^2 \right]^{1/2}, \qquad (1)$$

where x_{jr} is the coordinate of sound stimulus j on dimension r and w_{nr} is the weight of subject n on the common dimension r. We performed model selection following the procedure described in McAdams *et al.* (1995) to determine the appropriate dimensionality of the common space.

Multidimensional scaling resulted in a two-dimensional space accounting for the rated similarity between the different sound pairs. As can be seen in Fig. 1, different categories occupy separate regions of the perceptual space and the two dimensions appear to relate to the actions and material categories. Dimension 1 separates wood from glass/metal, whereas dimension 2 separates the action categories. Glass and metal sounds are rated as being similar across action categories but dissimilar to wood sounds. For the action categories, listeners rated single impact strikes as being more similar to drop sounds than they were to rattle sounds.

C. Auditory features

In order to examine signal features relevant to listeners' perception of similarity, we analyzed the sound stimuli using a representative model of spectrotemporal processing in the auditory periphery. First, the average signal level of the recorded stimuli was scaled to ensure it matched the measured sound pressure level. The signal was filtered corresponding to minimum audible field measurements to account for outer- and middle-ear filtering (Killion, 1978). Frequency-to-place transformation along the basilar membrane was simulated using a gammatone filter bank (Patterson *et al.*, 1995). The center frequencies of the filters were uniformly spaced on the ERB-rate scale (Glasberg and Moore, 1990) that approximates the auditory filter bandwidths as derived from masked detection thresholds in normal hearing listeners. The ERB-rate scale returns the number of equivalent rectangular bandwidths (ERB) of the auditory filters approximated as $z = 21.4 * \log(1 + 0.00437 * f)$. The output of each filter channel $[\psi_z(z,t)]$ was raised to the power of 0.25 to account for basilar membrane compression (Plack *et al.*, 2008) and was half-wave rectified to simulate the transformation of basilar membrane oscillations into hair-cell potentials. Temporal adaptation and integration in the auditory nerve were simulated using a series of feedback loops as described by Dau *et al.* (1996). Finally, each channel was converted to dB and low-pass filtered at 150 Hz to account for further temporal integration at higher stages in the auditory system.

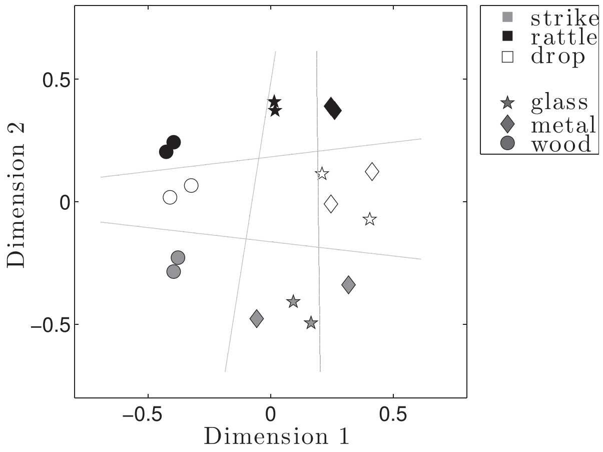

FIG. 1. Representation of perceptual similarity in two dimensions as found by multidimensional scaling. Different marker shapes indicate the material category and different shading indicate the action category. Solid lines visualize linear boundaries between material or action categories found by linear discriminant analysis.

1. Spectral centroid (SC)

First we investigated the potential correlation of long-term spectral content with MDS dimension 1 related to material categories. To do this, we used a measure of spectral centroid (SC) previously shown to be a robust predictor of timbre perception (Grey and Gordon, 1978; McAdams *et al.*, 1995; Marozeau *et al.*, 2003). Spectral centroid measures have also been related to viscoelastic properties of different impacted materials more generally (Avanzini and Rocchesso, 2004; McAdams *et al.*, 2004, 2010). We calculated the SC as the center frequency [in ERB-rate (z)] of a given filter weighted by the energy in that filter summed over time

$$\mathrm{SC} = \frac{\sum_z \left[z \sum_t \psi \right]}{\sum_z \sum_t \psi} \qquad (2)$$

calculated for the initial 200 ms of the sounds in order to avoid artifacts due to low signal levels at the end of the sound files.

The SC thus describes the spectral "center of gravity" in the distribution of energy over frequencies in the long-term spectrum. Impacts on glass or metal produce more high-frequency energy relative to impacts on wood and thus have higher SC values. The spectral envelopes calculated from the output of the auditory filters and SC values for different material categories are illustrated in Fig. 2.

We found a significant correlation between the SC and the material-related perceptual dimension of the weighted MDS model $[r(16) = 0.89, p < 0.0001]$. Like the perceptual similarity data, the auditory feature discriminates gross material categories (wood at low SC values vs glass-metal at higher SC values, see Fig. 4).

2. Temporal centroid (TC)

To examine the relationship between the similarity of actions along MDS dimension 2 and temporal information, we extracted a centroid measure in the temporal domain to quantify the dispersion of energy in the amplitude envelope over time. The temporal centroid (TC) describes the "center of gravity" of the temporal envelope as the sample times (in seconds) weighted by the envelope energy summed over frequencies

$$\mathrm{TC} = \frac{\sum_t \left[t \sum_z \psi \right]}{\sum_t \sum_z \psi}. \tag{3}$$

Rattle sounds have larger TCs than drop or strike sounds where the energy is concentrated around the initial impulse. The dispersion of material in drop sounds will also increase the TC relative to single impacts. The TCs for different actions are illustrated in Fig. 3. The correlation between the TC and the action related perceptual dimension was also significant $[r(16) = 0.91, p < 0.0001]$ and the descriptor effectively discriminates all action categories (see Fig. 4).

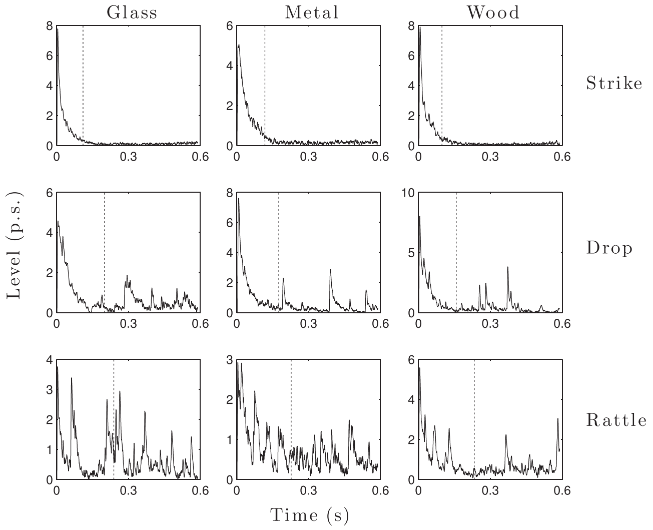

FIG. 3. Examples of temporal envelopes (in pseudosones) for different actions. The vertical lines indicate the position of the TCs.

3. Event density (ED)

The TC gives a global measure of the temporal position of envelope energy but it does not quantify individual impacts that may be perceptually salient with single and multiple impact sounds. For this reason we defined an alternative feature of event density (ED) to quantify similarity in the action-related dimension. To extract individual impacts, we summed the temporal envelope over frequency channels and extracted local peaks in the broad-band envelope. In order to make the peak detection reflect local impacts both in rattling and drop sounds, we removed faster envelope

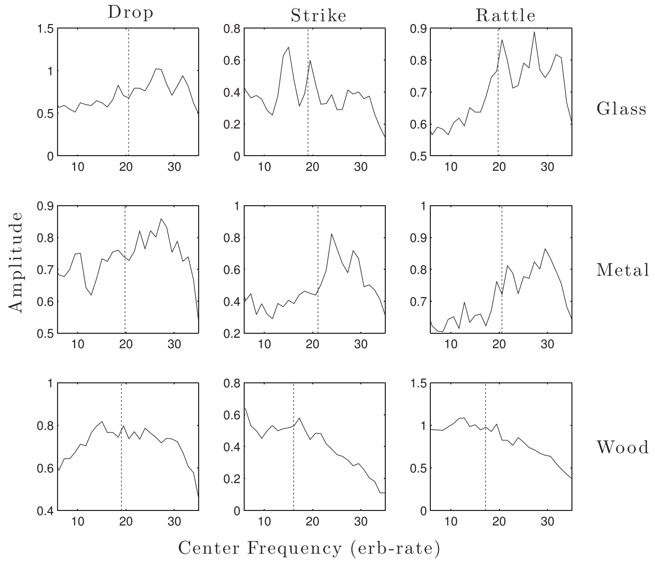

FIG. 2. Spectral envelopes and SC values (vertical lines) for different materials. The two columns show examples of spectral envelopes for different action categories.

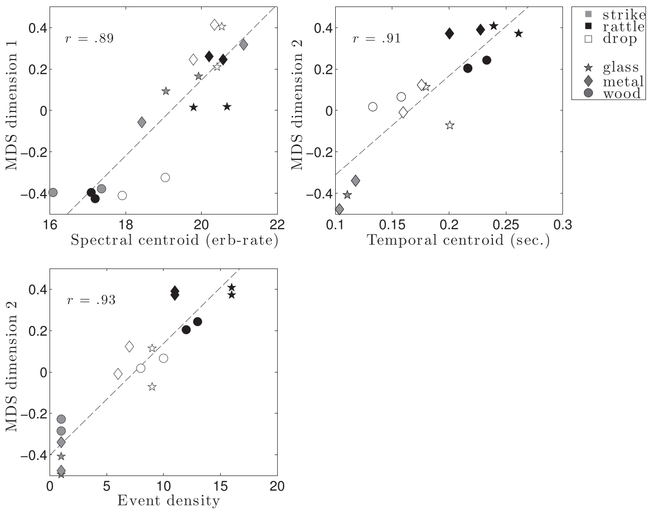

FIG. 4. Correlation between perceptual dimensions of the weighted MDS model and auditory features. Above left: the material-related dimension compared to the SC. Above right: the action-related dimension compared to the TC. Below: the action-related dimension compared to ED.

fluctuations by low-pass filtering at 50 Hz instead of 150 Hz. Event density was then defined as the sum of local peaks in the temporal envelope for the duration of the sound. We found that this feature also correlated significantly with MDS dimension 2 $[r(16) = 0.93, p < 0.0001]$.

4. Testing auditory features

The limited number of sounds used for pair-wise similarity comparison may represent a limited range of acoustic variation occurring naturally in the given category combinations. Although the auditory features appear to quantify the perceived similarity between categories, it may be difficult to generalize about category information because of the limited number of sounds used. For this reason, we tested the variation of the features on a larger number of sounds. We recorded ten novel exemplars of each of the nine category combinations yielding a total of 90 new sounds. We used a larger range of variation of object sizes and geometries to enhance the range of acoustic variation. We then calculated the auditory features for this larger sound bank to investigate the spread of categories in this feature space. Figure 5 shows the *SC* and *TC* values of each sound as extracted from the auditory model. As can be seen, the sound categories occupy similar regions of the feature space as found in the perceptual data. We then tested how well the SC and TC would distinguish the categories that were separated in the perceptual MDS space. The TC separated strike-drop and rattle-drop sounds with a mean error rate of 3.8%. The SC separated wood sounds with a mean error rate of 7.8%, whereas glass and metal sounds were highly overlapping as in the perceptual space.

D. Discussion

Multidimensional scaling of the similarity data suggested two perceptual dimensions related to category information. Gross material categories (wood vs metal/glass) were separated along dimension 1, whereas action categories were separated along dimension 2. The large similarity between glass and metal is consistent with classification studies showing confusion between them (Giordano and McAdams, 2006; Lemaitre and Heller, 2012). This could indicate that listeners rate similarity with respect to features that are relevant for categorization, even though they were not instructed to attend to category information in this experiment.

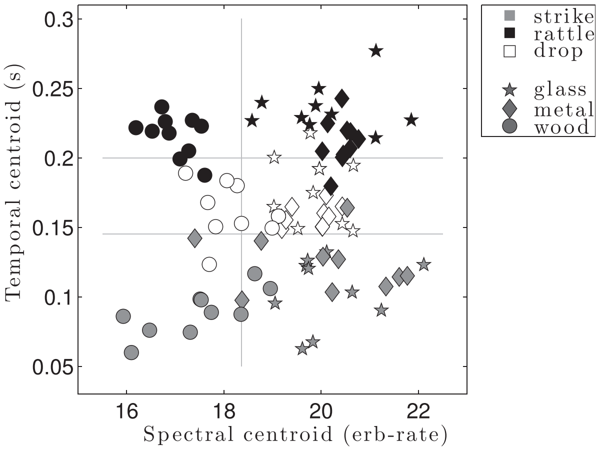

FIG. 5. Test of auditory features on a sound bank of 90 novel impact sounds. Horizontal and vertical lines visualize the separation value between relevant categories that minimize the error rate.

Analysis of auditory features suggested a relevance of spectral characteristics of materials across the different types of impact (MDS dimension 1). Previous studies with single impacts on different materials confirm a relation between perceptual similarity and SC measures, in particular, in the attack portion of the sounds (Grey and Gordon, 1978; McAdams *et al.*, 1995, 2004; Giordano, 2005; McAdams *et al.*, 2010). Our results suggest that this also applies across different sound producing actions. The spectral energy distribution may be a more effective cue for identifying material properties than pitch-related cues, for example, when also considering multiple impact sounds with less clear pitch quality.

Perception of actions and materials may involve loudness cues (Giordano and McAdams, 2006). In our stimulus set, loudness was perceptually equalized to avoid different stimuli having different audibility levels. However, this also means that potential effects of loudness were not investigated. We found a relatively large spread in the levels of the subjectively equalized stimuli, perhaps suggesting that listeners' loudness estimates were influenced by sound source information. When examining the acoustic features in a larger sound set (Sec. II C 4), we did not equalize the sounds subjectively. Similar variation within and between categories in this feature space suggest that these features are less influenced by the global level, but the precise role of loudness cues remained to be examined.

One interesting observation is the high degree of overlap between perceptual similarity and stimulus similarity that emerges at the level of the auditory periphery. We examined the same auditory features calculated at earlier stages in the auditory model. Using a model without the simulated neural adaptation in the auditory nerve, we observed a drop in the correlation between perceptual dimensions and both the SC (correlation with MDS dimension 1: $r = 0.84$) and TC (correlation with MDS dimension 2: $r = 0.54$). Adaptation did not influence peak detection so the correlation with ED was unchanged. Calculating the same features from an FFT-based spectrogram of the sounds instead of an auditory model resulted in considerably lower correlations (SC and MDS dimension 1: $r = 0.69$; TC and MDS dimension 2: $r = 0.72$). This suggests that processing in the auditory periphery may capture some features that are relevant for categorizing higher-order information, as indicated by previous studies (Lewicki, 2002). However, it also remains unclear whether listeners focus on salient spectrotemporal features that are implicitly carrying category-level information or whether the same features are used during behavioral categorization. In experiment 2, we examined whether the similarity ratings also predict performance in category discrimination by collecting categorization responses to the same set of sounds.

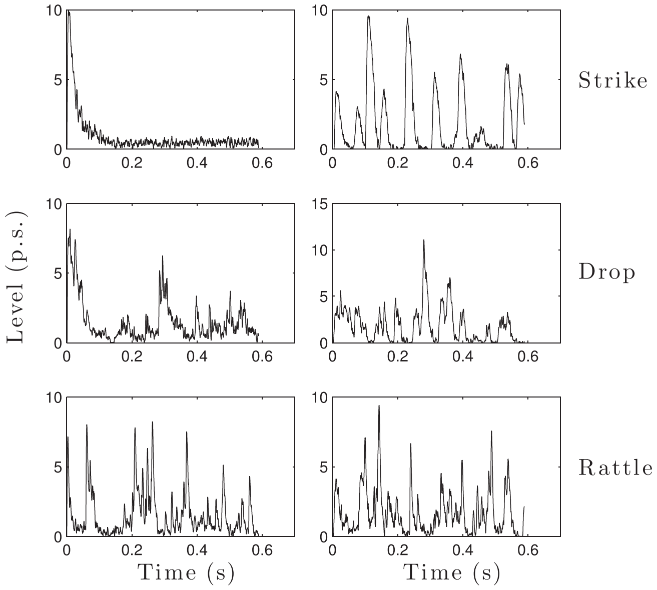

FIG. 6. Original (left) and scrambled (right) temporal envelopes for the different action categories.

III. EXPERIMENT 2: CATEGORIZATION JUDGMENTS

A. Methods

1. Subjects

Twenty subjects took part in the experiment (10 females, 10 males, aged 20–38). None of the subjects in experiment 2 had participated in experiment 1. All participants had normal hearing.

2. Stimuli

Subjects were presented with the same stimuli used in experiment 1. In order to verify the different influence of spectral and temporal cues suggested by the results of experiment 1, subjects were also presented with sounds manipulated to remove either spectral or temporal cues. Spectral cues were removed by applying the temporal envelopes of the original sounds to a broadband noise carrier. This resulted in a sound set with fixed SC (= 23.8 ERB, equivalent to that of glass/metal sounds in experiment 1) and TCs identical to the original sound set. As a second manipulation, we removed envelope cues by splitting the original sound signals into overlapping time windows (Hann-shaped, 40 ms length) and randomly permuting the windows in such a way that each of the scrambled sounds had a fixed long TC (= 300 ms, equivalent of rattle sounds) while preserving long-term spectral content. Figure 6 shows examples of the original and scrambled envelopes for each action category.

3. Apparatus

The apparatus used was identical to that of experiment 1.

4. Procedure

On different trials, listeners were presented with the sound stimuli and asked to identify either the material (wood, metal, or glass) or the action (drop, strike, or rattle) category. Word labels for the three different response categories were presented on-screen and subjects were asked to choose the appropriate label via a key press. The 18 original sounds were presented intermingled with the (2 × 18) scrambled sounds in random order. The action and material categorization was presented in separate blocks and both blocks were presented twice. In each categorization block, all (3 × 18) sound stimuli were presented. Response time was not limited. After a response, the interface indicated the chosen category but no feedback on accuracy was given. Subjects performed a short trial round before the experiment to familiarize themselves with the task. An experimental session took approximately 20 min to complete.

B. Categorization performance

Table I reports the categorization confusion data for all participants. The confusion scores suggest accurate categorization of both actions and materials for the original sounds. Spectral scrambling biases material categorization towards metal responses, whereas most action responses remain correct with the exception of struck glass, which is more often categorized as being dropped. For the temporal scrambling, material categorization remains accurate whereas actions are categorized as rattles.

TABLE I. Observed categorization confusion matrices for the population of subjects in each experimental condition. Rows correspond to stimuli and columns to response categories. Boldface indicates correct responses. W = Wood; G = Glass; M = Metal; S = Strike; D = Drop; R = Rattle.

	Experimental condition																	
	Original sounds						Spectral scrambling						Temporal scrambling					
	W	M	G	S	D	R	W	M	G	S	D	R	W	M	G	S	D	R
WS	**80**	0	0	**80**	0	0	**12**	58	10	**47**	32	1	**76**	3	1	**3**	7	70
WD	**79**	1	0	1	**78**	1	**9**	51	20	11	**67**	2	**75**	5	0	3	**1**	76
WR	**80**	0	0	1	13	**66**	**3**	62	15	0	26	**54**	**79**	0	1	0	2	**78**
MS	0	**73**	7	**80**	0	0	5	**47**	28	**21**	58	1	11	**41**	28	**5**	1	74
MD	0	**58**	22	5	**75**	0	2	**63**	15	5	**73**	2	2	**60**	18	0	**2**	78
MR	0	**59**	21	0	5	**75**	4	**70**	6	0	18	**62**	0	**52**	28	0	0	**80**
GS	1	11	**68**	**80**	0	0	4	48	**28**	**22**	58	0	2	11	**67**	**12**	0	68
GD	0	9	**71**	2	**77**	1	7	56	**17**	5	**71**	4	3	16	**61**	1	**3**	76
GR	0	6	**74**	1	6	**73**	4	66	**10**	2	29	**49**	0	14	**66**	0	0	**80**

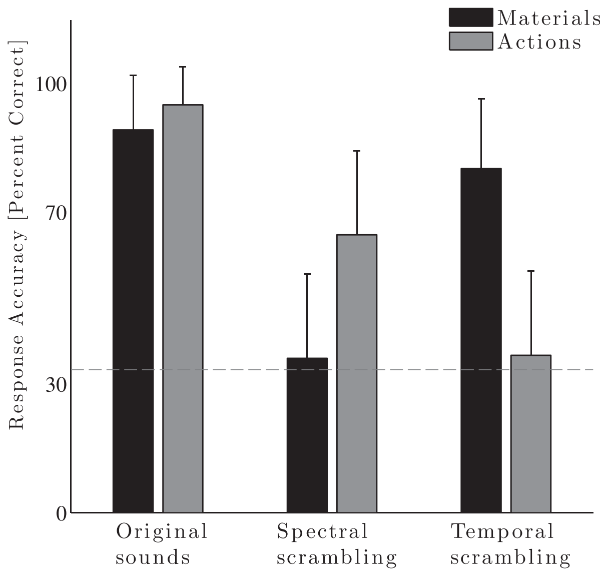

FIG. 7. Average categorization performance for material and action categories in the different experimental conditions. Chance performance is 33.3% correct (stippled line). Error bars indicate 1 S.E.M.

Figure 7 shows the mean categorization accuracy for each condition averaged over the individual action and material categories. We analyzed the categorization performance using a repeated measures analysis of variance (ANOVA) that included factors for the category types (materials, actions) and manipulation types (none, spectral scrambling, temporal scrambling). The analysis of variance (ANOVA) revealed an interaction between the two factors $[F(1,2) = 32.46, p < 0.0001]$, demonstrating that the manipulation of temporal and spectral cues affect the categorization of actions and materials differently.

C. Category sensitivity: General recognition theory

We used GRT (Ashby and Townsend, 1986; Ashby and Lee, 1991), a multivariate generalization of SDT, to model listeners' category discrimination performance. With two different perceptual dimensions (materials and actions), the GRT model assumes that a given stimulus S_i elicits a perceptual effect $f_i(x,y)$ that follows a two-dimensional normal distribution $\mathcal{N}$ with mean μ_i and co-variance matrix Σ_i. In a categorization task, the listener is assumed to divide the perceptual space into regions each associated with a given response. Decision bounds between these regions can be modeled as linear functions. The probability of responding R_j to stimulus S_i is then the density of the perceptual effect in the associated response region

$$P(R_j|S_i) = \int_{\mathbf{R_j}}\int f_i(x,y)\,dxdy, \tag{4}$$

where

$$f_i(x,y) = \mathcal{N}(\mu_i, \Sigma_i). \tag{5}$$

GRT allows us to examine quantitatively whether different combinations of action and material categories are perceptually separable (Ashby and Townsend, 1986). For instance, a given material category such as glass is perceptually separable if the perceptual effects of glass sounds do not vary depending on the type of impact producing the sound (i.e., the perceptual effect of glass sounds will have identical means across actions). Similarly, if decision bounds are identical for all glass sounds then listeners' tendency to respond glass is not biased by the type of impact (they are "decisionally separable," cf. Ashby and Townsend, 1986). Here, we fit the model with categorization data where responses occur to one dimension at a time (unlike identification experiments where there is a unique response for each stimulus). In this case, model parameters are estimated separately for each dimension and it is not meaningful to estimate co-variance between the dimensions.

We estimated a GRT model with the categorization confusion data of the original sound set for all participants by minimizing the negative log-likelihood of the model. In the most general model, the means, variances, and decision bounds for each category combination are free to vary. Models with identical means or variances across dimensions are special cases of the more general model. Because of this nesting, we use a hierarchical model selection procedure to find the appropriate number of free parameters using a likelihood ratio test (cf. Ashby and Lee, 1991). We first test the general model against a model with equal means by computing the ratio of the likelihoods of the two models. The log-likelihood ratio is compared to a chi-square distribution with degrees of freedom equal to the difference in number of free parameters between the general and the restricted model in order to determine whether the extra free parameters of the general model provide a significantly better fit to the data. We then proceed to examine whether models with fixed bounds and variances account for the categorization responses (Ashby and Lee, 1991).

This model selection procedure resulted in a model with equal means, decision bounds, variances (12 free parameters) along both dimensions. This model accounted for 99.0% of the variance in the categorization responses (the full model explaining all of the variance). Figure 8 shows the fitted GRT model (a), and the initial full model with unequal means, bounds, and variances (b).

Having a very good fit with equal parameters along both dimensions suggests that categorization of both materials and actions are separable with respect to perception and decision bias. Listeners recognized the sound source material similarly across impact types and, conversely, were not affected by the material in their ability to discriminate the impact action category. Interestingly, this means that listeners discriminate material categories in multiple impact sounds and that the discrimination sensitivity is similar to what is found with single impact strikes. Across actions, glass and metal sounds were confused more often than metal/glass and wood, as often reported in studies of single impact sounds (Lutfi and Oh, 1997; Giordano and McAdams, 2006). For the action discrimination, strike sounds were occasionally confused with drop sounds but not with rattle sounds, whereas rattle sounds could be confused with drops.

As can be seen, the configuration of category means is similar to the perceptual structure inferred from the similarity

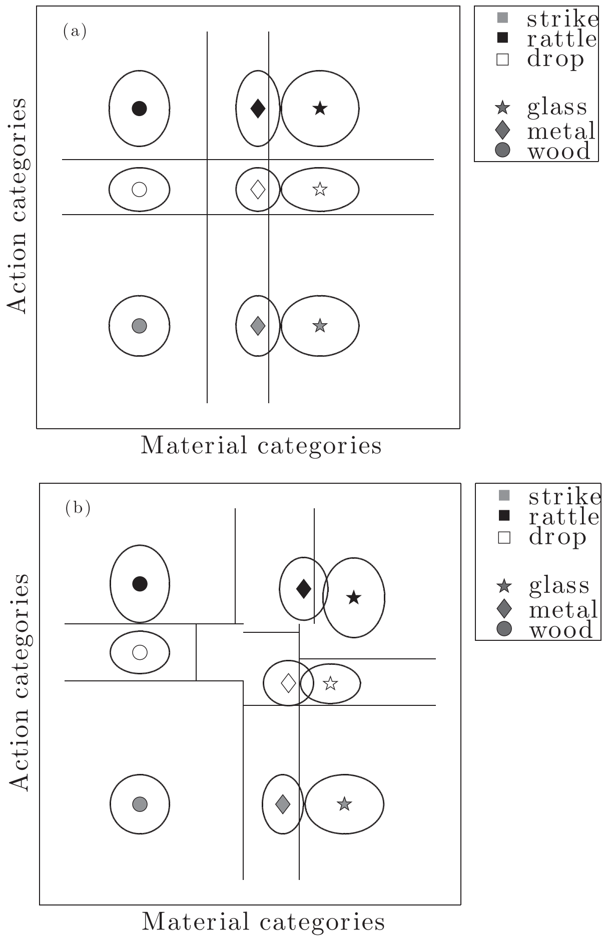

FIG. 8. GRT models fitted to the categorization data for the original sound set. Circles show the contours of equal probability associated with the joint distributions of the perceptual effects. Lines indicate decision bounds between response regions. The means of perceptual effects all fall in regions associated with a correct response. (a) Fitted model with equal means, variances, and decisions bounds across category combinations, and (b) full model where all parameters are free to vary.

ratings in experiment 1. In effect, categories perceived as more similar are also more likely to be confused during categorization. As a heuristic measure to compare the modeled similarity and categorization responses, we calculated the correlation between the distribution of means for a given category in the fitted GRT model and the MDS coordinates. For the MDS representation, we used the average of the two exemplars of the same category combination. We found high correlations between both MDS dimension 1 and the material-related dimension in the GRT model $[r(7) = 0.85, p < 0.036]$ and between MDS dimension 2 and the action-related GRT dimension $[r(7) = 0.97, p < 0.0001]$.

D. Predicting perceptual similarity from categorization performance

The correlation between the similarity structure and categorization sensitivity also suggests the possibility of making quantitative predictions about categorization performance from perceptual similarity, or inversely of deriving similarity measures from category sensitivity. As mentioned, GRT views similarity as being proportional to the probability of category confusion. The similarity s between stimulus S_i and stimulus S_j may be defined as the amount of the perceptual effect of stimulus S_i that falls into the response region associated with S_j (Ashby and Perrin, 1988),

$$s(S_i, S_j) = \int_{\mathbf{R_j}}\int f_i(x,y)\,dxdy. \tag{6}$$

This yields a bias-free measure (analogous to d' in SDT), where the distance between two stimuli in perceptual space is only determined by the parameters of the normal distribution defining the perceptual effects.

We used this measure to calculate a similarity matrix for each exemplar of the different category combinations from the fitted GRT model. We then compared these matrices to the pair-wise similarity ratings obtained in experiment 1. Overall, the GRT-derived similarities corresponded poorly to observed similarities, accounting for only 56.7% of the variation of the observed similarity ratings. The GRT model appears to underestimate the similarity between stimuli where there is little confusion. Because the categorization accuracy was very high for many category combinations, the normal distribution of perceptual effects will result in very small similarity estimates for these stimulus combinations. The fact that we observed a similar configuration of perceptual effects in the GRT model and the MDS space suggests that the relation between similarity and categorization sensitivity is not accurately captured for our data by the shape of the Gaussian distribution on which the GRT similarity measure relies.

E. Predicting categorization performance from perceptual similarity

It is also possible, however, to consider the inverse relation in which categorization is explained as a function of perceived similarity in multidimensional perceptual space. As mentioned above, this approach is suggested by different variants of the choice model (Shepard, 1957; Luce, 1963). Unlike the context-free GRT measure of similarity [Eq. (6)], choice models weight the context of other stimulus exemplars in the experiment and do not make assumptions about parametric distribution of perceptual effects. In a categorization experiment, the probability of a stimulus S_i being identified as belonging to category C_j may simply be formulated as a function of the summed similarity η_{ij} between S_i and all other stimuli in the category normalized by the summed similarity between all stimuli η_{ik} (cf. the generalized context model, Nosofsky, 1986; here we ignore possible response bias),

$$P(R_j|S_i) = \frac{\sum_{j\in C_j}\eta_{ij}}{\sum_{k}\eta_{ik}}, \tag{7}$$

where η_{ij} is a function of the distance d_{ijr} between stimuli along a particular common dimension r in weighted MDS space [Eq. (1)]

$$\eta_{ij} = -c e^{d_{ijr}^{p}}. \tag{8}$$

This formulation means that the probability of a categorization confusion falls off monotonically with the distance in MDS space. The nonnegative parameters c and p determine the rate and shape of the decay and thus define the general stimulus discriminability. If $p = 2$, then η takes the shape of a Gaussian function, and the model becomes similar to a context-sensitive formulation of the GRT similarity measure (where the perceptual effects are integrated across all members of a category; Ashby and Perrin, 1988). Gaussian similarity functions with $p = 2$ may better explain category performance in individual well-practiced subjects (Nosofsky, 1986), whereas exponential decay with $p = 1$ may generalize better across subjects and experiments (Shepard, 1987).

Using the context model in Eq. (7), we estimated category responses from the two common dimensions of the MDS space inferred from the similarity data in experiment 1. We then compared the predicted categorization responses from the MDS model to the observed categorization data from experiment 2. We set the parameter $p = 1$, because we compared the MDS common space to categorization performance in the population of untrained participants. The scaling parameter c was set to a high value of 10 due to the general high discriminability between sound exemplars. With these parameters, we found that the model accounted for a large portion of the observed categorization performance. With similarities along MDS dimension 1, the model explained 78.8% of the variation in the material categorization performance, whereas similarities along MDS dimension 2 accounted for 93.8% of the variation in action categorization. In comparison, MDS dimension 1 explained only 10.7% of the variation in the action categorization, and MDS dimension 2 explained 15.4% of the variation in material categorization. With even sharper decay of the similarity function [$p = 0.5$ in Eq. (8)], we found that the model explained 95.0% and 99.1% of the material and action categorization data, respectively. This also implies that the auditory features correlated with the MDS dimensions (SC, TC, ED) effectively predict categorization sensitivity.

F. Effects of temporal and spectral scrambling

The averaged categorization performance (Fig. 7) showed that removal of either spectral or temporal cues lowered categorization performance but affected the categorization of actions and materials differently. As indicated by the results of experiment 1, removal of temporal cues with preservation of long-term spectral content (and thus identical SCs) resulted in categorization at chance level for actions but not for materials. Conversely, removal of spectral information with preserved envelopes resulted in chance level performance for material categories but not for actions.

We fitted GRT models for the categorization data obtained with spectrally or temporally scrambled stimuli (Fig. 9). In the spectral scrambling condition, model selection resulted in a configuration with equal bounds but unequal variances and means on the material dimension, and equal means and variances but unequal bounds on the action dimension. Removal of spectral cues resulted in loss of sensitivity even for gross material categories (wood vs glass/metal). Listeners categorized sounds primarily as metal and sometimes as glass. This is consistent with the fact that these noise vocoded sounds have a fixed high SC value that correspond to glass/metal sounds as shown in experiment 1. On the other hand, listeners continued to discriminate action categories with relatively high sensitivity without spectral cues. Model selection suggested that the spectral manipulation introduced a decisional bias toward drop responses that was stronger for glass and metal. This bias followed the pattern of confusion observed with the original sound set: strike or rattle sounds became more confused with drop sounds but the manipulation did not increase the confusion between strike and rattle sounds.

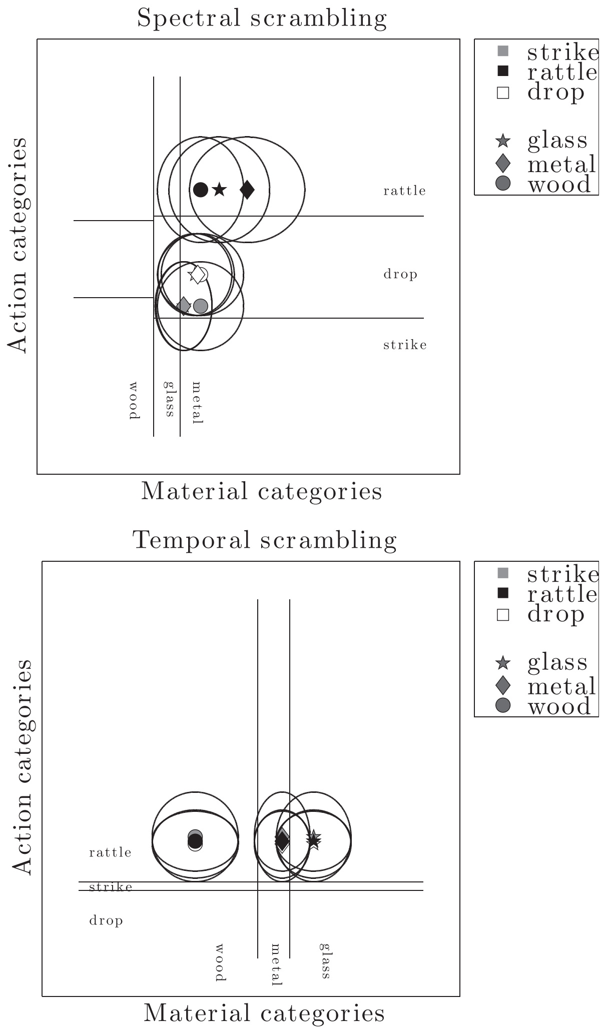

FIG. 9. GRT models fitted to the categorization data with spectral scrambling (above) and temporal scrambling (below).

For the temporally scrambled sounds a model with equal means, decision bounds, and variances on both dimensions was found to account for the categorization performance. Removal of envelope information resulted in complete loss of action category sensitivity. Nearly all envelope-scrambled sounds were categorized as rattle sounds, as expected by the high TC or ED values. Material categories, on the other hand, continued to be discriminated with a similar sensitivity as observed with the original sounds. Metal/glass sounds continued to be confused similarly as with temporal cues, although metal appeared to be categorized less accurately in the case of strike sounds.

IV. GENERAL DISCUSSION

GRT model selection suggested that actions were perceptually separable across the material of the object, and that materials were separable across sound-generating actions. Combined with the acoustical analysis derived from the similarity data, these results suggest the relevance of cues that are invariant across the large range of acoustic variation introduced by the different impact actions. Although single impact studies have suggested cues that allow fine-grained discrimination of materials, our results suggest the relevance of considering the cues that are invariant across a greater variation in context.

The material-related dimension derived from the similarity ratings was described by spectral content as quantified by the SC, effectively predicting sensitivity in material categorization. Removal of temporal cues showed that listeners were able to discriminate gross material categories based only on spectral information also for multiple impact sounds. Without temporal cues, listeners discriminated gross categories (wood vs glass/metal) but also continued to discriminate glass and metal sounds at a level similar to that of the original sounds. When we removed spectral cues, on the other hand, the material of single impact sounds could potentially have been discriminated based on the amplitude decay rate but we did not see this. Listeners showed no sensitivity between any of the material categories without spectral cues, suggesting that they cannot use envelope cues for material discrimination independently of frequency information (cf. Wildes and Richards, 1988; Avanzini and Rocchesso, 2004). With single impacts, we saw a small tendency for improved discrimination of metal sounds and impaired discrimination of metal when temporal cues were removed, but GRT model selection suggested similar sensitivity across actions.

These findings indicate that spectral content is favored as a more invariant cue to material properties when considering different impact types. The strong focus on the relevance of damping-related cues for material discrimination in the sound source perception literature should be viewed in relation to the fact that most studies have used the same sound generating action, single impacts. This constraint may have created a bias towards the relevance of damping cues. We suggest that spectral cues may be more general and robust when considering cues for material discrimination in a broader context of different sound-generating events. Our results also question the conclusions drawn by Lemaitre and Heller (2012) arguing that material discrimination *per se* has only limited relevance in audition given that damping cues are only efficient with single impact sounds. To the contrary, we find that reliance on spectral cues allow listeners to pick up material information across impact categories, even if it may result in lower sensitivity for single impacts. Weak material discrimination was reported by Lemaitre and Heller (2012) in particular, with rolling and scraping cylinders that lack both the temporal and spectral cues characteristic of impacted sound sources. However, impact actions generating vibration of solid materials are still a major part of natural acoustic events. It seems likely that listeners favor spectral cues that are both invariant across contexts and potentially faster to compute as they rely less on slow temporal information. Our temporal scrambling resulted in sounds with relatively uniform spectral content over the duration of the sound, suggesting discrimination based on global frequency content. Given that the material composition of objects is often not optimally inferred with visual perception, efficient auditory cues for material categories are highly valuable in a natural environment. The perceptual relevance of spectral cues reported in a number of studies (Lutfi and Oh, 1997; Giordano, 2003; Giordano and McAdams, 2006; Klatzky *et al.*, 2000; Avanzini and Rocchesso, 2001) should also be viewed in relation to natural settings where the use of more efficient and context-invariant spectral features may be traded off for accuracy.

Our results confirm a remarkable robustness of auditory action perception investigated in only a few previous studies (Warren and Verbrugge, 1984; Lemaitre and Heller, 2012). Action discrimination accuracy remained high even after removal of spectral cues. However, the loss of spectral information caused listeners to bias their responses toward drop sounds, particularly for strikes on metal/glass. This result supports those results of Warren and Verbrugge (1984), who found that listeners discriminated bouncing and breaking events without spectral cues, although their absence reduced discrimination sensitivity. However, in our study, temporal scrambling destroyed action sensitivity entirely suggesting that listeners are not able infer action information from spectral cues without temporal information. We quantified the perceived similarity between action categories via the centroid of the temporal envelope or the density of impacts, effectively also describing sensitivity in action discrimination. These features summarizing the temporal evolution of the envelope may be sufficient for recognition of events generated by continuous excitation such as rattle sounds. Our temporal scrambling produced realistic rattle sounds suggesting that for these types of continuous "textural" sound events, action perception may not rely on the exact temporal envelope pattern, but can be captured by summary features (McDermott and Simoncelli, 2011).

The similarity and categorization data revealed related representations. Description of this relationship, however, relied on accurate model assumptions. As predicted by the formal categorization models considered in this study, categorization performance was described via a nonlinear mapping of the similarities and normalization. Since we used stimuli that where highly discriminable for many category combinations, we found that a sharp decay of the mapping function yielded better predictions. It was also important to retrieve the correct dimensionality of the data. For the similarity ratings, subject weighting in the MDS procedure captured two separate dimensions related to the material and action category information and these dimensions could then be used to predict categorization performance. The fact that categorization can be deduced from similarity with these model assumptions also suggests that they reflect related but not identical processes. This may also be suggestive as to why previous studies have derived qualitatively similar sound source features from similarity and categorization data but sometimes with only moderate linear correlations between them (Gygi *et al.*, 2007; McAdams *et al.*, 2010; Giordano and McAdams, 2010).

Aramaki, M., Besson, M., Kronland-Martinet, R., and Ystad, S. (**2011**). "Controlling the perceived material in an impact sound synthesizer," IEEE Trans. Audio, Speech, Lang. Process. **19**, 301–314.

Ashby, F. G., and Lee, W. W. (**1991**). "Predicting similarity and categorization from identification," J. Exp. Psychol. **120**, 150–172.

Ashby, F. G., and Perrin, N. A. (**1988**). "Toward a unified theory of similarity and recognition," Psychol. Rev. **95**, 124–150.

Ashby, F. G., and Townsend, J. T. (**1986**). "Varieties of perceptual independence," Psychol. Rev. **93**, 154–79.

Avanzini, F., and Rocchesso, D. (**2001**). "Controlling material properties in physical models of sounding objects," in *Proc. Int. Computer Music Conf.*, pp. 91–94.

Avanzini, F., and Rocchesso, D. (**2004**). "Physical modeling of impacts: Theory and experiments on contact time and spectral centroid," in *Proceedings of the Conference on Sound and Music Computing*, pp. 287–293.

Carroll, J., and Chang, J. (**1970**). "Analysis of individual differences in multidimensional scaling via n-way generalization of Eckart-young decomposition," Psychometrika **35**, 283–319.

Dau, T., Püschel, D., and Kohlrausch, A. (**1996**). "A quantitative model of the 'effective' signal processing in the auditory system. I. Model structure," J. Acoust. Soc. Am. **99**, 3615–3622.

Gaver, W. (**1993**). "What in the world do we hear?: An ecological approach to auditory event perception," Ecol. Psychol. **5**, 1–29.

Gaver, W. W. (**1988**). "Everyday listening and auditory icons," Ph.D. thesis, University of California.

Giordano, B. L. (**2003**). "Material categorization and hardness scaling in real and synthetic impact sounds," in *The Sounding Object*, edited by D. Rocchesso and F. Fontana (Mondo Estremo, Firenze), pp. 73–93.

Giordano, B. (**2005**). "Sound source perception in impact sounds," Ph.D. thesis, University of Padova.

Giordano, B., and McAdams, S. (**2006**). "Material identification of real impact sounds: Effects of size variation in steel, glass, wood, and Plexiglas plates," J. Acoust. Soc. Am. **119**, 1171–1181.

Giordano, B., and McAdams, S. (**2010**). "Sound source mechanics and musical timbre perception: Evidence from previous studies," Music Percept. **28**, 155–168.

Giordano, B., Visell, Y., Yao, H.-Y., Hayward, V., Cooperstock, J. R., and McAdams, S. (**2012**). "Identification of walked-upon materials in auditory, kinesthetic, haptic, and audio-haptic conditions," J. Acoust. Soc. Am. **131**, 4002–4012.

Glasberg, B. R., and Moore, B. C. (**1990**). "Derivation of auditory filter shapes from notched-noise data," Hear. Res. **47**, 103–138.

Goldstone, R. L. (**1994**). "The role of similarity in categorization: Providing a groundwork," Cognition **52**, 125–157.

Goldstone, R. L., Lippa, Y., and Shiffrin, R. M. (**2001**). "Altering object representations through category learning," Cognition **78**, 27–43.

Grey, J., and Gordon, J. (**1978**). "Perceptual effects of spectral modifications on musical timbres," J. Acoust. Soc. Am. **63**, 1493–1500.

Gygi, B., Kidd, G. R., and Watson, C. S. (**2004**). "Spectral-temporal factors in the identification of environmental sounds," J. Acoust. Soc. Am. **115**, 1252–1265.

Gygi, B., Kidd, G. R., and Watson, C. S. (**2007**). "Similarity and categorization of environmental sounds," Percept. Psychophys. **69**, 839–855.

Killion, M. C. (**1978**). "Revised estimate of minimum audible pressure: Where is the 'missing 6 dB?,' " J. Acoust. Soc. Am. **63**, 1501–1508.

Klatzky, R. L., Pai, D. K., and Krotkov, E. P. (**2000**). "Perception of material from contact sounds," Presence: Teleoperators Virtual Environ. **9**, 399–410.

Krumhansl, C. L. (**1978**). "Concerning the applicability of geometric models to similarity data: The interrelationship between similarity and spatial density," Psychol. Rev. **85**, 445–463.

Kunkler-Peck, A. J., and Turvey, M. T. (**2000**). "Hearing shape," J. Exp. Psychol. **26**, 279–294.

Lemaitre, G., and Heller, L. M. (**2012**). "Auditory perception of material is fragile while action is strikingly robust," J. Acoust. Soc. Am. **131**, 1337–1348.

Lewicki, M. S. (**2002**). "Efficient coding of natural sounds," Nat. Neurosci. **5**, 356–363.

Luce, R. D. (**1963**). "Detection and recognition," in *Handbook of Mathematical Psychology* (John, New York), Vol. 1, pp. 103–190.

Lutfi, R. A. (**2001**). "Auditory detection of hollowness," J. Acoust. Soc. Am. **110**, 1010–1019.

Lutfi, R. A., and Oh, E. L. (**1997**). "Auditory discrimination of material changes in a struck-clamped bar," J. Acoust. Soc. Am. **102**, 3647–3656.

Marozeau, J., de Cheveigné, A., McAdams, S., and Winsberg, S. (**2003**). "The dependency of timbre on fundamental frequency," J. Acoust. Soc. Am. **114**, 2946–2957.

McAdams, S., Chaigne, A., and Roussarie, V. (**2004**). "The psychomechanics of simulated sound sources: Material properties of impacted bars," J. Acoust. Soc. Am. **115**, 1306–1320.

McAdams, S., Roussarie, V., Chaigne, A., and Giordano, B. L. (**2010**). "The psychomechanics of simulated sound sources: Material properties of impacted thin plates," J. Acoust. Soc. Am. **128**, 1401–1413.

McAdams, S., Winsberg, S., Donnadieu, S., De Soete, G., and Krimphoff, J. (**1995**). "Perceptual scaling of synthesized musical timbres: Common dimensions, specificities, and latent subject classes," Psychol. Res. **58**, 177–192.

McDermott, J. H., and Simoncelli, E. P. (**2011**). "Sound texture perception via statistics of the auditory periphery: Evidence from sound synthesis," Neuron **71**, 926–40.

Nosofsky, R. M. (**1986**). "Attention, similarity, and the identification-categorization relationship," J. Exp. Psychol. Gen. **115**, 39–61.

Patterson, R. D., Robinson, K., Holdsworth, J., McKeown, D., Zhang, C., and Allerhand, M. (**1995**). "Complex sounds and auditory images," in *Auditory Physiology and Perception*, edited by Y. Cazals, L. Demany, and K. Horner (Pergamon, Oxford), pp. 429–446.

Plack, C. J., Oxenham, A. J., Simonson, A. M., O'Hanlon, C. G., Drga, V., and Arifianto, D. (**2008**). "Estimates of compression at low and high frequencies using masking additivity in normal and impaired ears," J. Acoust. Soc. Am. **123**, 4321–4330.

Shepard, R. N. (**1957**). "Stimulus and response generalization: A stochastic model relating generalization to distance in psychological space," Psychometrika **22**, 325–345.

Shepard, R. N. (**1987**). "Toward a universal law of generalization for psychological science," Science **237**, 1317–1323.

Sloutsky, V. M. (**2003**). "The role of similarity in the development of categorization," Trends Cogn. Sci. **7**, 246–251.

Tucker, S., and Brown, G. J. (**2003**). "Modelling the auditory perception of size, shape and material: Applications to the classification of transient sonar sounds," in *Audio Engineering Society Convention*, Audio Engineering Society, p. 114.

Tversky, A. (**1977**). "Features of similarity," Psychol. Rev. **84**, 327–352.

Warren, W. H., and Verbrugge, R. R. (**1984**). "Auditory perception of breaking and bouncing events: A case study in ecological acoustics," J. Exp. Psychol. **10**, 704–712.

Warren, W. H., Jr., Kim, E. E., and Husney, R. (**1987**). "The way the ball bounces: Visual and auditory perception of elasticity and control of the bounce pass," Perception **16**, 309–336.

Wildes, R. P., and Richards, W. A. (**1988**). "Recovering material properties from sound," in *Natural Computation*, edited by W. A. Richards (MIT Press, Cambridge), pp. 356–363.