

Deriving auditory features from triadic comparisons

FLORIAN WICKELMAIER AND WOLFGANG ELLERMEIER
Aalborg University, Aalborg, Denmark

A feature-based representation of auditory stimuli is proposed and tested experimentally. Within a measurement-theoretical framework, it is possible to decide whether a representation of subjective judgments with a set of auditory features is possible and how unique such a representation is. Furthermore, the method avoids confounding listeners' perceptual and verbal abilities, in that it strictly separates the process of identifying auditory features from labeling them. The approach was applied to simple synthetic sounds with well-defined physical properties (narrow-band noises and complex tones). For each stimulus triad, listeners had to judge whether the first two sounds displayed a common feature that was not shared by the third, by responding with a simple "yes" or "no." Because of the high degree of consistency in the responses, feature structures could be obtained for most of the subjects. In summary, the proposed procedure constitutes a supplement to the arsenal of psychometric methods with the main focus of identifying the type of sensation itself, rather than of measuring its threshold or magnitude.

One of the perennial unresolved problems in psychoacoustics is to find out which auditory sensations are elicited by complex acoustic stimuli. In applied research, the majority of studies have employed *direct* methods of eliciting verbal descriptors for the sensations that play a role in a given domain (e.g., the various qualitative changes produced by audio coding algorithms). With these methods, researchers have attempted to produce an exhaustive list of the relevant attributes by either guiding a panel of experts through a sequence of controlled exposures (descriptive analysis; Bech & Zacharov, 2006; Stone & Sidel, 2004, chap. 6) or by asking subjects to identify what similarities or differences emerge when comparing sets of sounds (repertory grid technique; Berg & Rumsey, 2006; Kelly, 1955).

The major problem with these direct elicitation methods is that they rely on a close correspondence between the (latent) auditory sensations and their verbal descriptors. It is conceivable, though, that listeners experience a sensation they do not have a word for in their lexicons. Likewise, it is by no means guaranteed that the verbal labels produced are meaningfully and consistently linked to the actual sensations.

This confounding of verbal and perceptual abilities is avoided when *indirect* methods of identifying the relevant sensations are employed. The pertinent psychometric methods have mostly focused on metric and dimensional representations of some measure of the psychological proximity of the stimuli, often derived from judgments of similarity or dissimilarity. Most notably, various versions of multidimensional scaling (MDS) have been developed

(e.g., by Carroll & Chang, 1970; Kruskal, 1964; Shepard, 1962; Torgerson, 1952; see Borg & Groenen, 1997, for an introduction and overview) and applied to uncover dimensions of auditory perception (e.g., Grey, 1977; Iverson & Krumhansl, 1993; Lakatos, McAdams, & Caussé, 1997). MDS seeks to represent the stimuli under study in some multidimensional space, such that the metric distances in that space correspond to the psychological proximities.

Beals, Krantz, and Tversky (1968) studied MDS from the viewpoint of representational measurement theory and formulated qualitative properties that the proximities must satisfy in order to be representable as metric distances. Tversky (1977) criticized metric and dimensional representations in general and demonstrated that observed proximities often systematically violate the metric conditions inherent in geometrical models. Instead, he proposed a feature-based representation, the so-called *contrast model*, that is able to explain many of the empirical findings. Formally, the contrast model predicts the similarity S between two stimuli a and b with the equation

$$S(a, b) = \vartheta f(A \cap B) - \alpha f(A \setminus B) - \beta f(B \setminus A) \quad (1)$$

(see Tversky, 1977, p. 332), where S and f are interval scales, $A \cap B$ denotes the features that are common to both a and b , $A \setminus B$ the features that belong to a but not to b , and $B \setminus A$ the features that belong to b but not to a ; the parameters ϑ , α , and β are nonnegative weighting factors. The contrast model expresses similarity between stimuli as a weighted function of their common and distinctive features. The main limitation of the contrast model as a method for revealing salient features results from the fact

that the features have to be explicitly specified in order for the model to be testable. Thus, from similarity data alone, the characterizing features cannot uniquely be identified. Sattath and Tversky (1987) provided further evidence for this lack of uniqueness inherent in the contrast model.

Heller (2000) concluded from the unsolved uniqueness problem of the contrast model that similarity data generally do not provide enough information to derive the characterizing feature structure. To overcome this problem, he introduced a theory of semantic features and an experimental paradigm for their assessment that are closely related to both knowledge space theory (Doignon & Falmagne, 1999; Falmagne, Koppen, Villano, Doignon, & Johannesen, 1990) and formal concept analysis (Ganter & Wille, 1999; Wille, 1982). For these so-called *semantic structures*, he formulated both representation and uniqueness theorems in the sense of representational measurement theory (Krantz, Luce, Suppes, & Tversky, 1971). Thus, a feature representation in this paradigm rests on qualitative, experimentally testable conditions, and its uniqueness can be stated explicitly and is determined empirically.

In this article, Heller’s (2000) semantic structures are applied to derive auditory features. In the following section, the theoretical notions needed to characterize auditory feature structures are briefly introduced, ideas that are in close correspondence with semantic structures. Subsequently, we will report on an experiment designed to test the proposed approach for revealing the auditory features elicited by simple synthetic sounds.

STRUCTURES OF AUDITORY FEATURES

Let X denote the total finite set of sounds under study, the so-called *domain*, and σ a collection of subsets of X , which will be interpreted as the set of auditory features of the sounds in X . In accordance with Heller (2000), $\langle X, \sigma \rangle$ is called an *auditory (feature) structure*. Furthermore, let $A \subseteq X$ denote a subset of X and $\sigma(A)$ the intersection of all sets in σ of which A is a subset,

$$\sigma(A) = \bigcap_{A \subseteq S, S \in \sigma} S, \tag{2}$$

which means that $\sigma(A)$ is the smallest set in σ that includes the sounds in A . In this case, a relation Q that relates the subsets of X to X can be defined in the following way: The sounds in A are said to be in relation to a sound $x \in X$ —formally, AQx —if and only if the subject answers “no” to the question “Do the sounds in A have something in common that makes them different from x ?” If the answer to that question is “yes,” the relation between A and x does not hold, which is denoted by $A\bar{Q}x$. Furthermore, Q is called a *quasi-ordinal relation on X* if

$$a \in A \Rightarrow AQa \text{ (reflexivity)} \tag{3}$$

and

$$AQb (\forall b \in B) \text{ and } BQc \Rightarrow AQc \text{ (transitivity)} \tag{4}$$

hold for all nonempty $A, B \subseteq X$, and $c \in X$. In the present application, the reflexivity of Q is assumed. Thus, only questions in which $x \notin A$ are presented to the subject. Transitivity will be illustrated in the example below.

The main difference, from a theoretical point of view, between semantic and auditory structures is that the hyponymy relation that exists between two words if they are sub- and superordinate concepts (for example, *dog* is hyponymous to *animal*, which implies that $\{\text{animal}\} Q \text{dog}$) is not expected to exist between sounds. Therefore, Q is assumed not to hold between any two single sounds a and b , and thus $\{a\} \bar{Q} b \forall a, b \in X, a \neq b$. Consequently, singleton subsets of X are not presented to the subject. This, together with reflexivity (Equation 3), corresponds to the assumption that the sounds in X are perceptually distinct (i.e., have at least one characteristic feature). Therefore, a *minimal feature structure*, $\sigma_0 = \{\emptyset, \{a\}, \{b\}, \{c\}, \dots, X\}$ including the empty set, the singletons, and the domain, is assumed a priori. We are interested in whether additional common features can be derived from the collected data. It can be shown that transitivity as defined in Equation 4 is necessary and sufficient for an auditory structure to exist (Heller, 1991, Appendix A.2; Heller, 2000, Theorem 2).

Example

As an example, consider a set of four sounds $X = \{a, b, c, d\}$, as well as a hypothetical auditory structure $\sigma = \{\emptyset, \{a\}, \{b\}, \{c\}, \{d\}, \{a, b, c\}, X\}$ defined on it; this auditory structure may also be denoted as the minimal structure σ_0 and one additional feature, or $\sigma = \sigma_0 \cup \{\{a, b, c\}\}$. It follows from Equation 2 that $\sigma(\{a, b\}) = \{a, b, c\}$, indicating that sounds a and b share all their common features also with sound c , whereas $\sigma(\{a, d\}) = X$ denotes that a and d display no other common features than the ones shared by all sounds in X . Figure 1 shows the lattice graph of σ (without the empty set \emptyset). In such a graph, the features are represented as nodes connected by lines, so that lower nodes are subsets of connected higher nodes. For example, the set $\{a, b, c\}$, which is a subset of the domain X , represents a feature shared by the sounds a, b , and c . To illustrate, let us assume $\{a, b, c\}$ denotes a clarinet-like timbre, which the sound d does not have. Suppose now that the relation Q has been established by querying a listener: In line with the assumed structure, the listener answered with a “no”

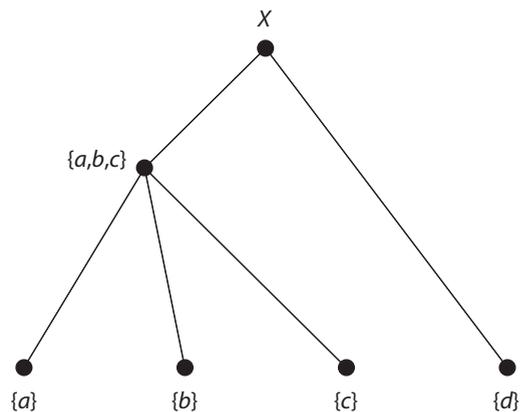


Figure 1. Lattice graph of the auditory feature structure $\sigma = \{\emptyset, \{a\}, \{b\}, \{c\}, \{d\}, \{a, b, c\}, X\}$ on the domain $X = \{a, b, c, d\}$. The empty set is omitted.

the question of whether a and b had something in common that distinguished them from c , and thus $\{a, b\}Qc$. Furthermore, let $\{a, b\}\bar{Q}d$ (“yes”) but $\{b, c\}Qd$ (“no”). From reflexivity, it follows trivially that $\{a, b\}Qb$. Given this pattern of responses, the relation Q is nontransitive, since transitivity would require that

$$\{a, b\}Qb, \{a, b\}Qc, \text{ and } \{b, c\}Qd \Rightarrow \{a, b\}Qd,$$

and consequently the structure σ cannot be derived from Q . Obviously, the listener was not able to consistently identify the timbral feature $\{a, b, c\}$ because it played a role only in some of the triadic comparisons but was irrelevant in others (e.g., in $\{b, c\}Qd$). Note that a test of the transitivity condition is possible only if the left side of Equation 4 is true. Therefore, the more “no” responses that have been observed, the more transitivity tests will be possible. In the following representation theorem, transitivity is the critical empirical condition for a representation.

Representation

Let Q be a relation on $2^X \times X$ that is reflexive and transitive ($2^X \times X$ denotes the Cartesian product of the power set 2^X of X and the set X). There then exists an auditory structure $\langle X, \sigma \rangle$, such that

$$AQb \text{ if and only if } b \in \sigma(A) \quad (5)$$

for all nonempty $A \subseteq X$ and $b \in X$.

In order to construct an auditory structure, it is convenient to define the set of all sounds that are in relation to A as

$$AQ = \{x \in X : AQx\}. \quad (6)$$

In the example above, for $A = \{a, b\}$, AQ is given by $\{a, b\}Q = \{a, b, c\}$, since $\{a, b\}Qa$ and $\{a, b\}Qb$ are implied by reflexivity (Equation 3), and $\{a, b\}Qc$ and $\{a, b\}\bar{Q}d$ represent the observed responses. From Equation 5, it follows that $AQ = \sigma(A)$. The goals of our experiment are to determine all $\sigma(A)$ from the collected responses and to construct the auditory structure σ that represents the quasi-ordinal relation Q .

Uniqueness

Heller (2000) has demonstrated that it is not necessary to establish the relation Q on *all* possible subsets of X . Instead, the antecedent subsets can be restricted to containing only two or fewer elements. Since in the present application singletons are never presented, the subsets are further constrained to include only pairs of elements, resulting in triadic comparisons among the sounds. In doing so, not only is the number of questions drastically reduced, but also the load on the subject’s working memory for each question is kept at a reasonable level. Such a reduction of questions comes at the cost of a potential loss of information, yielding a nonunique representation. In general, the equivalence in Equation 5 holds for more than one structure σ , given a set of triadic comparisons.

Formally, let $\varphi^{(2)}$ denote the smallest structure (with respect to number of elements) representing a quasi-ordinal

relation Q that is based on triadic comparisons. The structure $\varphi^{(2)}$ is constructed by

$$\varphi^{(2)} := \{\{a, b\}Q : \{a, b\} \subseteq X\}. \quad (7)$$

It denotes the collection of all sets $\{a, b\}Q$, where $\{a, b\}$ is a subset of X and $\{a, b\}Q$ is defined as in Equation 6. Furthermore, let $\sigma^{(2)}$ denote the largest possible representing structure (containing the most elements), which is defined as

$$S \in \sigma^{(2)} \text{ if and only if } (\{a, b\}Qs \Rightarrow s \in S), \quad (8)$$

for all $S \subseteq X$, $s \in X$, and all $\{a, b\} \subseteq S$. Any structure σ is then a representing structure of Q if and only if

$$\varphi^{(2)} \subseteq \sigma \subseteq \sigma^{(2)}. \quad (9)$$

In particular, the representation is unique if $\varphi^{(2)} = \sigma^{(2)}$. The restriction to triadic comparisons, therefore, does not necessarily result in a loss of information, but it will depend on the complexity of Q whether a unique representation can be obtained.

In summary, the experimental procedure for deriving auditory structures can be outlined as follows: First, establish Q based on triadic comparison judgments. Then, test the transitivity of Q . If transitivity holds, construct both $\varphi^{(2)}$ (Equation 7) and $\sigma^{(2)}$ (Equation 8). If $\varphi^{(2)}$ and $\sigma^{(2)}$ are equal, they form the uniquely representing auditory structure. Otherwise, all structures that satisfy Equation 9 are representing structures of Q . Finally, in order to obtain a full structure including the singleton subsets, which can be represented by a lattice diagram, the structures satisfying Equation 9 are united with the minimal structure σ_0 .

The presented approach is able to uncover the underlying auditory features if and only if at least some stimuli have features in common and are thereby distinct from some other stimuli. If all sounds under study are nondiscriminable or each one is entirely different from all the others (i.e., they possess only unique features), the method will not provide further insight into the auditory organization of the sounds. More precisely, the presented approach can be considered a method to derive *common* as well as *distinctive* auditory features. Situations in which stimuli are perceived as entirely unique entities are, however, potentially rare. Often the context provided by a set of sounds would initiate processes of categorization and organization. We hypothesize that such processes are feature based, and our proposed method aims at deriving the features signifying these auditory categories.

This is—to our knowledge—the first experimental attempt to apply semantic structures (Heller, 2000) to perceptual stimuli rather than to verbal concepts. Therefore, to increase the chance of finding interpretable auditory structures, highly discriminable sounds were presented to naive listeners in two stimulus sets. The first set consisted of four sounds varying in center frequency and amplitude envelope. The second set was somewhat larger, allowing representation by more complex structures: Here, the physical variables manipulated were fundamental frequency and the number of overtones. We hypothesized

that different auditory sensations of pitch, brightness, timbre, or loudness changes could be evoked by these sounds and captured using auditory structures.

METHOD

Subjects

The sample consisted of 18 listeners (9 male, 9 female), who were between 21 and 30 years of age (median, 23.5 years). None of the subjects reported any hearing problems. Normal hearing of the subjects was assessed using pure-tone audiometry. The highest threshold found was at 25 dB hearing level (re. ISO 389-1, 1998) for 1 subject (M.L.) in one ear at two out of the ten audiometric frequencies between 250 and 8000 Hz.

Stimuli and Apparatus

Two different sets of synthetic sounds constituted the two experimental conditions: In the first condition, the stimuli consisted of four third-octave-band Gaussian noises with a center frequency of 500 or 2000 Hz and either a linearly rising (denoted by +) or falling (-) amplitude envelope; in order to increase audibility of the envelope, each noise had a duration of 2 sec. In the second condition, six periodic complex tones served as stimuli, with fundamental frequencies of 220 Hz (denoted by A), 277 Hz (C#), and 349 Hz (F) separated by at least a major third (400 cents), and composed of either 4 or 20 harmonics in random phase. The amplitude of a given harmonic was proportional to the inverse of its number. Each complex tone had a duration of 1 sec. All stimuli had cosine-shaped rise and fall times of 10 msec. Figure 2 depicts examples of the stimuli schematically. In the remainder of this article, stimuli are labeled by their two components: For example, "500+" denotes the 500-Hz narrow-band noise with rising envelope, and "A4" refers to the complex tone having a fundamental frequency of 220 Hz and four harmonics.

The stimuli were rendered digitally in MATLAB at a sampling frequency of 44.1 kHz and exported as 16-bit .wav files. They were played back by a personal computer using a digital sound card (RME DIGI96/8 PST) connected to an external D/A converter (RME ADI-8

DS) and delivered to the headphones (Beyerdynamic DT990) by a power amplifier (Rotel RB-976 Mk II).¹ The presentation software was implemented in LabView. The subjects entered their responses by clicking the buttons for "yes" or "no" on a computer screen using the mouse. The experiment was conducted in a sound-insulated, double-walled listening cabin.

The instrumentally measured loudness of the four noises and the six complex tones was aligned such that their mean loudness in sones matched approximately. In order to do this, the stimuli were recorded binaurally using a head and torso simulator (Brüel & Kjør 4128) and a measurement system (Brüel & Kjør PULSE 3560C), and the gains of the signals were adjusted to compensate for the measured loudness differences. After loudness alignment, the noises varied within a range of 0.5 sones and the complex tones within a range of 1 sone. On average, the equivalent sound pressure levels (L_{eq}) after loudness alignment were 59.3 dB for the noises and 60.6 dB for the complex tones.

Procedure

The experimental procedure consisted of two parts: In the familiarization segment, the subjects were presented with all four or six stimuli and asked to listen to them and try to recognize features that the sounds might share or that distinguished them from other sounds. The sounds could be repeated as often as the subject desired. The familiarization was completed in a self-paced manner.

In the data collection segment, on each trial, the subjects were presented with a stimulus triple and the question "Do sounds A and B have something in common that makes them different from sound C?" The subjects were to answer "yes" if they heard that the first and second sounds displayed a common feature that was not shared by the third sound. Otherwise, the answer was to be "no." Any of the three sounds could be repeated as often as necessary to reach a decision.

Generally, in order to establish the relation Q for n stimuli, $\binom{n}{2}(n-2)$ questions—corresponding to all unordered pairs of stimuli combined with any of the remaining stimuli—have to be asked; this requirement gives rise to 12 and 60 questions for the noises and tones, respectively. These questions were presented twice in two separate blocks. If the responses in the two blocks were identical,

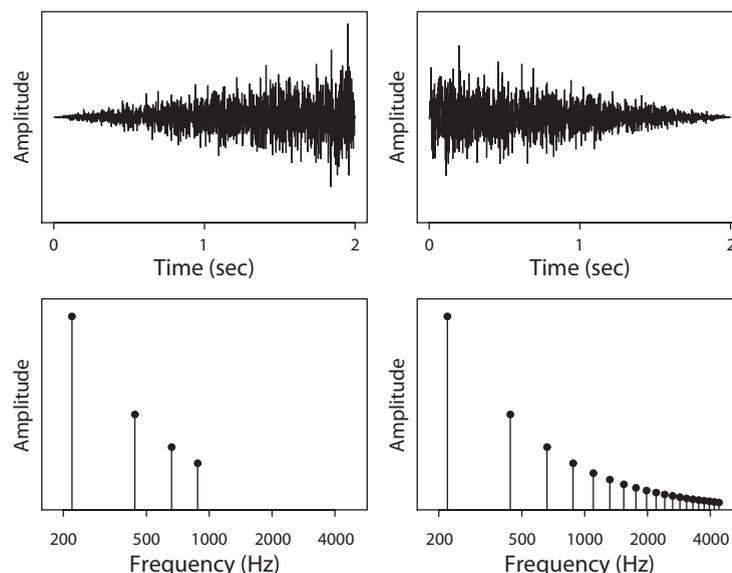


Figure 2. Illustration of the stimuli. (Upper panels) Condition I: Rising or falling third-octave-band noise. (Lower panels) Condition II: Periodic complex tones with 4 or 20 harmonics. Only tones with a fundamental frequency of 220 Hz are displayed.

data collection was completed. In case of contradictory responses, a third block was presented consisting of only those questions that had been answered differently in the first and second blocks. The order of the triples was randomized in each block. The order of sounds A and B was balanced across subjects, as was the order of the stimulus conditions: Half of the sample started with noises, the other half with complex tones.

The blocks, each containing familiarization and subsequent data collection, were distributed over two sessions of approximately 45 min each on two separate days. A short training block of six triadic-comparison trials preceded the data collection. The experiment was concluded with an informal debriefing in which the subjects were asked to name the auditory features upon which they had based their answers.

Data Analysis

The derivation of auditory feature structures from the collected triadic-comparison judgments rested on two conditions: First, the responses should reflect a certain degree of reliability; that is, the number of answers changing between the blocks should be small. Second, the judgments had to be consistent enough to allow for a representation. Consistency was operationalized in terms of the number of transitivity violations. In principle, only a single systematic violation would prevent representability, but in an experiment, it is well conceivable that a subject could carelessly give a wrong response, resulting in a random violation. In particular, when one response accounts for several violations, it may be regarded as a careless error. Therefore, when reliability and consistency were judged to be sufficient, and only few transitivity violations were still present after the third block, an attempt was made to find a structure that fit the data as closely as possible out of several plausible candidates. How this was done in practice will be illustrated in the following section. The discrepancy $\delta_Q(\sigma)$ between the experimentally determined relation Q and the proposed structure σ served as a lack-of-fit measure; $\delta_Q(\sigma)$ amounts to the number of responses it would be necessary to reverse in order to resolve all transitivity violations and, thus, for Q to become consistent with a potential structure σ .

RESULTS

Stimulus Set I: Narrow-Band Noises

Reliability and consistency. Table 1 displays the indexes of reliability and consistency in the narrow-band noise condition for the 18 subjects. Overall, the number of responses that changed between the blocks was low. In 12 cases, no third block needed to be presented because of perfect reliability (indicated by a dash in the third column). In the few cases in which a third block was necessary, at most one answer was changed back.

Transitivity was checked for the first and second blocks, and for all answers when a third block was presented; in doing so, the contradictory answers in the first and second blocks were replaced by the ones given in the third block. Thus, the number of violations reported in the sixth column of Table 1 denotes the *residual* transitivity violations based on all collected responses. For 15 of the 18 subjects, we found no violations after the last block; this means that the condition for representability of their judgments by an auditory feature structure was fulfilled without restriction. The remaining 3 subjects showed one (A.G. and V.H.) and three (A.M.) violations, respectively.

For these final 3 subjects, the violations had to be classified as random or systematic inconsistencies, with the latter preventing representation by an auditory structure.²

Table 1
Indexes of Reliability (Response Changes Between Blocks) and Consistency (Number of Transitivity Violations) in the Third-Octave-Band Noise Condition

Subject	Response Changes		Transitivity Violations			$\delta_Q(\sigma)$
	I-II	II-III	I	II	III	
A.G.	1	0	2	1	1	1
A.K.	0	–	0	0	–	0
A.M.	5	1	0	4	3	(2)
C.G.	0	–	0	0	–	0
F.M.	4	1	0	1	0	0
G.B.	0	–	0	0	–	0
J.S.	0	–	0	0	–	0
K.G.	4	0	2	0	0	0
K.P.	0	–	0	0	–	0
M.A.	0	–	0	0	–	0
M.B.	0	–	0	0	–	0
M.L.	0	–	0	0	–	0
N.C.	0	–	0	0	–	0
N.L.	0	–	0	0	–	0
O.K.	0	–	0	0	–	0
S.J.	0	–	0	0	–	0
S.R.	3	1	1	1	0	0
V.H.	3	0	2	1	1	1

Note—The discrepancy $\delta_Q(\sigma)$ indicates the number of response reversals necessary to resolve the remaining transitivity violations in order for the data to be consistent with the closest representing auditory structure. Parentheses indicate that no representation was attempted.

The following example illustrates the procedure. By applying Equation 6, A.G.'s responses can be summarized as

$$\begin{aligned}
 \{500+, 500-\}Q &= \{500+, 500-\} \\
 \{500+, 2000+\}Q &= \{500+, 2000+, 2000-\} \\
 \{500+, 2000-\}Q &= X \\
 \{500-, 2000+\}Q &= X \\
 \{500-, 2000-\}Q &= \{500-, 2000-\} \\
 \{2000+, 2000-\}Q &= \{2000+, 2000-\},
 \end{aligned}$$

where $X = \{500+, 500-, 2000+, 2000-\}$. It follows from transitivity (Equation 4) that $B \subseteq AQ \Rightarrow BQ \subseteq AQ$, for all nonempty subsets A, B of X . This implication is violated once in the data, since $\{500+, 2000-\} \subseteq \{500+, 2000+\}Q$, but $\{500+, 2000-\}Q \not\subseteq \{500+, 2000+\}Q$. In order to resolve this violation, either $2000-$ (indicated above in italics) has to be removed from $\{500+, 2000+\}Q$ or $500-$ has to be added. These two options correspond to assuming two different potentially underlying auditory structures: one structure containing the feature $\{500+, 2000+\}$ evoked by tones having a rising amplitude envelope, and one without this feature. Here, the first option was considered more plausible, because removing $2000-$ would mean assuming that A.G. *erroneously* responded with “no” when asked whether sounds $500+$ and $2000+$ had a common feature not shared by $2000-$. Also, without prior hypotheses about elicited features, it is often more plausible to conjecture that a “no” response occurred by mistake than that a “yes” response did (Heller, 2000, p. 29). Taking into account the overall good reliability and consistency of A.G.'s judgments, it seems likely that on this trial only he missed the otherwise salient feature. Therefore, the response was reversed from

“no” to “yes,” resulting in a discrepancy of $\delta_O(\sigma) = 1$. On the basis of the corrected responses, an auditory structure $\varphi^{(2)} = \{\{500+, 500-\}, \{500+, 2000+\}, \{500-, 2000-\}, \{2000+, 2000-\}, X\}$ is obtained using Equation 7. Since $\varphi^{(2)} = \sigma^{(2)}$ (Equation 8), this representation is unique. After uniting with the minimal structure σ_0 , its lattice graph can be drawn, and this graph is displayed in the right panel of Figure 3 (see the following section).

A similar argument applies to V.H.’s data for classifying the single transitivity violation as a random error. For A.M., however, no representation was attempted. This is partly due to the subject’s overall lack of reliability and consistency, and partly to a discrepancy of $\delta_O(\sigma) = 2$ with the closest structure, which seems high for a total of 12 questions.

Auditory structures. On the basis of the reported results concerning consistency and reliability, an auditory feature structure could be derived for 17 of the 18 subjects. The left panel of Figure 3 displays a lattice diagram of the structure obtained from the judgments of 4 subjects. The nodes in the graph denote features common to all stimuli connected to the node. The four noises (500+, 500-, 2000+, and 2000-) have one unique feature each, represented by the lowest nodes in the graph. The top node represents a feature common to all four stimuli. These features, however, already result from the assumptions that the sounds are perceptually distinct and comparable; the features are contained in the minimal structure σ_0 , and therefore in any representing auditory structure. More interesting, however, are the two additional features, one common to the 500-Hz stimuli and one to the 2000-Hz stimuli, indicating that noises with the same center frequency share an auditory feature.

It is worth noting that one node in the lattice diagram can represent one or a combination of auditory features. More specifically, one node denotes a salient auditory category characterized by one or more features. The identifiability of single features depends on the choice of the stimuli and on how the features covary in these stimuli. For a method of how to derive a minimal set of features, see Heller (2000, p. 17).

The right panel of Figure 3 shows the auditory structure derived for 12 subjects. It contains four nontrivial features. In addition to the categories for noises with the

same center frequency, two features are assigned to noises having the same amplitude envelope. Note that already such a relatively simple structure is too complex to be represented by a rooted tree graph, which allows for only one possible pathway from the top to the terminal node. Therefore, tree graphs, which are frequently used to depict the outcomes for metric distance models (e.g., cluster analysis), are inadequate here; lattice graphs, on the other hand, provide sufficient flexibility to represent the interrelations between features.

For 1 subject (S.R.), an additional auditory structure including three nontrivial features was obtained. Its lattice graph can be inferred from Figure 5. (See also the section below on comparing feature structures.)

Stimulus Set II: Complex Tones

Reliability and consistency. Table 2 displays the indexes of reliability and consistency in the complex-tone condition. Generally, more within- and between-subjects variability was observed here than for the narrow-band noises. One subject (M.L.), however, still displayed perfect reliability, and 6 more answered at most 4 of the 60 questions differently when queried the second time, indicating a high degree of reliability in their judgments. For 3 subjects (O.K., J.S., and V.H.), not only the number of changes between the first and second blocks, but also the fact that they reversed about half of these answers when queried again, suggests that their judgments were unreliable.

Overall, transitivity was found to hold without violation for 8 listeners, which corresponds to perfect representability. Seven more subjects displayed a discrepancy of at most four answers ($\delta_O(\sigma) \leq 4$) with an auditory structure. For these subjects, a representation was attempted as well. In doing so, the remaining transitivity violations were classified as random errors. This appeared to be justified when several violations were resolved by reversing only a few responses. For example, six violations were left for F.M. after the last block; only a single response reversal was needed to resolve them all. This makes it likely that the subject had carelessly given the response. For the 3 subjects O.K., J.S., and V.H., consistency was not judged sufficient for a representation; the discrepancy of their judgments with the closest fitting structure was at least five answers.

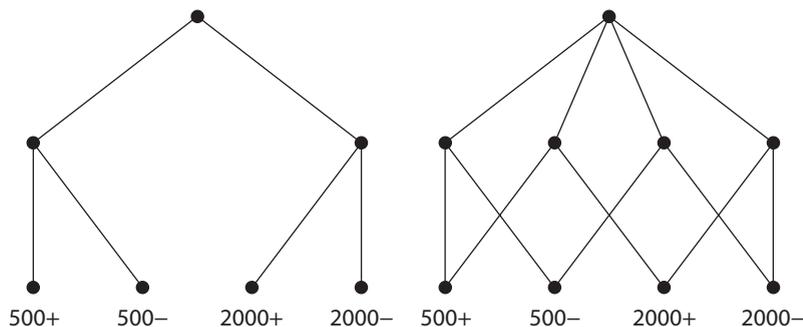


Figure 3. Auditory feature structures derived from the triadic-comparison judgments of 4 (left) and 12 (right) subjects. The stimuli consisted of narrow-band noises having center frequencies (in hertz) indicated by the numbers and rising (+) or falling (-) amplitude envelopes.

Auditory structures. A representation was derived for 15 of the 18 subjects. The left panel of Figure 4 displays the auditory structure representing the judgments of 4 of the participating subjects. It includes two nontrivial auditory categories, one for the 4-harmonic and one for the 20-harmonic complex tones, indicating that overtone content elicited a common auditory sensation. The right panel of Figure 4 shows the structure derived for 6 other subjects. It contains three additional features common to tones of the same fundamental frequency (A, C#, or F). For the remaining 5 subjects, auditory structures were obtained that included from 4 to 14 features. Their lattice graphs can be inferred from Figure 6. (See also the following section on comparing feature structures.)

All auditory feature structures for which perfect representability held were found to be unique in the sense of the uniqueness theorem, which means that $\varphi^{(2)}$ and $\sigma^{(2)}$ were equal (see Equation 9)—that is, all features in a given structure followed directly from the relation Q that was established by the triadic comparisons. This was true with one exception: Subject M.A. displayed a rather complex relation Q . Consequently, the triadic comparisons did not provide enough information to decide whether two features were or were not included in his auditory structure. These two features were the ones characterizing the 4- and 20-harmonic tones. The resulting nonuniqueness might be resolved in two ways: One possibility would be to ask the questions that could provide the necessary information. One such question, for example, could have been “Do the sounds A4, C#4, and F4 share a feature that A20 does not have?” Thus, quadruple comparisons would have resolved the nonuniqueness. The second, less elegant but more practical, solution would be to rely on the debriefing to provide the missing information—for instance, by asking the subject to name the involved features *after* the data collection had been completed. In this case, it was inferred from the descriptions obtained in the debriefing session that both features were included in this subject’s auditory structure. (See the section below on labeling auditory features.)

Comparing Feature Structures

So far, the results we have presented are strictly individual. Indeed, one of the strengths of the proposed method is

Table 2
Indexes of Reliability and Consistency (Over Blocks)
in the Complex-Tone Condition

Subject	Response Changes		Transitivity Violations			$\delta_Q(\sigma)$
	I–II	II–III	I	II	III	
A.G.	2	1	8	0	0	0
A.K.	9	1	31	4	4	1
A.M.	9	4	21	25	17	3
C.G.	1	0	6	0	0	0
F.M.	2	2	6	18	6	1
G.B.	9	1	25	3	0	0
J.S.	11	6	31	26	20	(6)
K.G.	3	1	8	6	0	0
K.P.	18	0	0	0	0	0
M.A.	8	0	12	0	0	0
M.B.	6	5	17	17	14	4
M.L.	0	–	0	0	–	0
N.C.	4	4	10	0	10	4
N.L.	11	0	39	0	0	0
O.K.	20	10	36	37	24	(5)
S.J.	3	1	20	6	9	3
S.R.	13	7	23	14	11	4
V.H.	32	17	24	56	33	(7)

Note—The rightmost column gives the discrepancy between the responses and the closest representing auditory structure (see Table 1). Parentheses indicate that no representation was attempted.

that it does not rely on aggregated or averaged data, but allows for individual differences to become apparent. On the other hand, a researcher might be interested in questions like “How salient is a given auditory feature in a sample of subjects?” or “How (in)homogeneous with respect to auditory perception is the sample under study?” which require a certain level of aggregation. Such questions can be answered within the framework of auditory structures. In order to do so, the individual structures of all subjects were arranged in a common lattice graph in which each node represents a possible auditory structure.

Figure 5 shows the lattice graph of all extracted individual structures in the narrow-band noise condition. Solid circles denote structures that actually represent the judgments of one or more listeners, and open circles indicate *potential* structures that were not implied by the actual judgments. The top node denotes the minimal structure σ_0 that includes only the trivial features that are—per

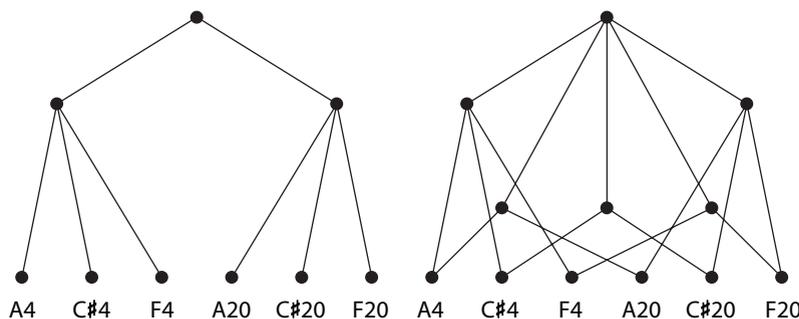


Figure 4. Auditory feature structures derived from the triadic-comparison judgments of 4 (left) and 6 (right) subjects. The stimuli consisted of complex tones having a fundamental frequency of 220 Hz (A), 277 Hz (C#), or 349 Hz (F) and 4 or 20 harmonics (denoted by the numbers).

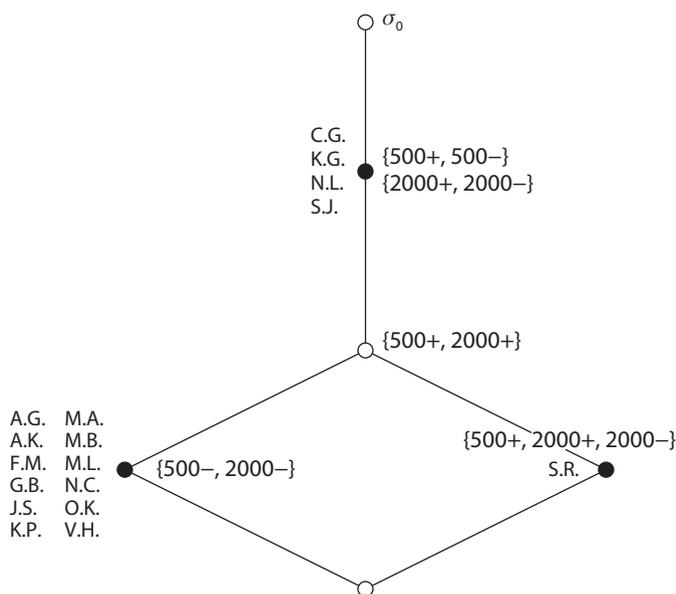


Figure 5. Combined lattice graph of the 17 individual auditory structures representing the narrow-band noises. Each node denotes an auditory structure including the features written next to the node, plus all features of nodes that can be reached by following the ascending lines. Initials indicate the subjects from whom the auditory structures were derived.

assumption—contained in any structure. The second node from the top shows the simple structure displayed in the left panel of Figure 3; only the nontrivial features {500+, 500-} and {2000+, 2000-} are indicated. To the left of the node are the initials of the 4 subjects for whom this structure was derived. Lower level nodes represent structures that include the features of *all* higher level nodes that can be reached by following ascending lines. Therefore, the lower the node, the more complex the structure. For example, the node labeled {500-, 2000-} also contains the three features from higher level nodes; it represents the structure shown in the right panel of Figure 3 and was derived for 12 subjects.

From Figure 5, it is obvious that the two features elicited by noises of the same center frequency ({500+, 500-} and {2000+, 2000-}) were the most salient in the sample, because they are included in all resulting auditory structures. Thirteen listeners used the feature {500+, 2000+}, 12 listeners the feature {500-, 2000-}. Only 1 subject (S.R.) displayed an extra feature shared by three noises ({500+, 2000+, 2000-}), which was therefore the least salient in the sample. The fact that only three different structures were derived argues for a strong agreement between the subjects about the auditory features emerging from this stimulus set.

Figure 6 displays the common lattice graph for the complex-tone condition. The two structures shown in Figure 4 are denoted by the two top-left nodes in the graph. The most salient features were the ones assigned to tones with the same number of harmonics; 14 of the 15 subjects for whom structures were derived used the two features {A4, C#4, F4} and {A20, C#20, F20} to distinguish among

sounds. The features elicited by tones of the same fundamental frequency ({A4, A20}, {C#4, C#20}, and {F4, F20}) were included in the structures of 8 subjects. Five of the subjects perceived an auditory feature when two tones with the same number of harmonics were not more than one third apart from each other—for example, {A20, C#20} or {C#4, F4}. This might indicate that fundamental frequency and number of harmonics interact in order to create a new feature. (Note, however, that A.M.'s structure is missing the feature {A4, C#4}.) The additional features found by S.R. seem to be rather idiosyncratic, and the informal debriefing did not provide further information as to how they might be labeled appropriately. In general, however, the simple shape of Figure 6 indicates good agreement between the subjects.

Labeling Auditory Features

The labeling of the obtained features does not directly follow from the triple-comparison judgments. So far, when deriving feature structures, the stimuli have been organized into categories (or sets), which—due to the absence of other (e.g., semantic) information—may be assumed to be *auditory* categories. If one is only interested in the *behavior* a subject displays when categorizing a set of auditory stimuli, data analysis might stop here. The result is a collection of sets, with each set containing sounds sharing a common feature. Therefore, the derivation of features from the observed judgments can be completed without requiring additional verbal information. In the absence of such information, feature labels might be inferred by inspecting the lattice graph of a derived structure. This is not unlike in MDS, in which the researcher inspects the

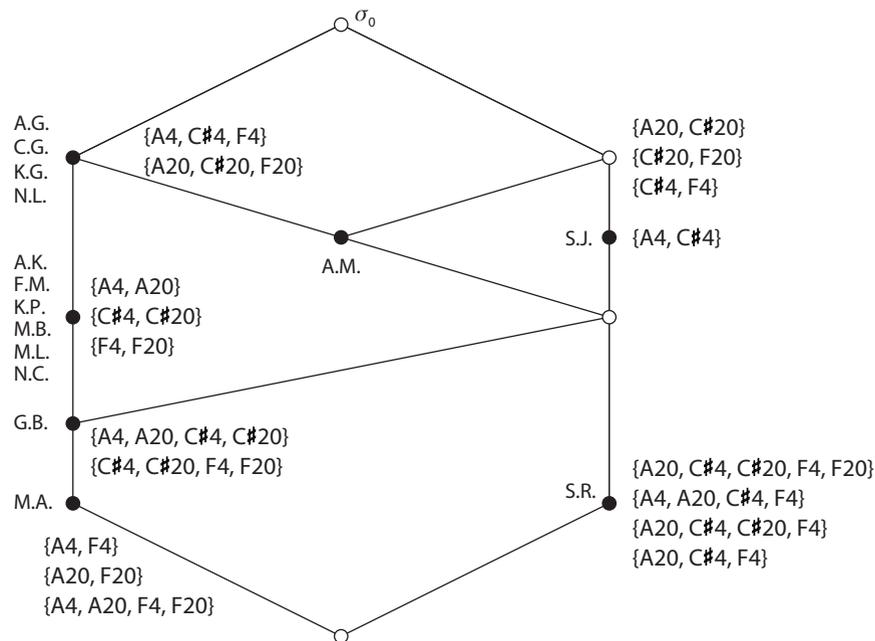


Figure 6. Combined lattice graph of the 15 individual auditory structures representing the complex tones. (See the Figure 5 caption for details of how to read the structure.)

MDS space and assigns verbal labels to its dimensions based on his or her own interpretation of that space. In the case of carefully designed synthetic stimuli (as in the present case), an educated guess might be attempted as to what forms the perceptual basis of the categories obtained, which might then be labeled accordingly.

Alternatively, category labels may be derived from the additional information acquired in the debriefing session, in which the subjects were requested to name the involved features *after* the data collection had been completed. This strategy potentially leads to more accurate labeling and is less biased by the experimenter’s expectations. In the narrow-band noise condition, descriptions like “crescendo”/“decrescendo” or “fade in”/“fade out” corroborate the hypothesis that noises of the same amplitude envelope (+ or -) elicited auditory sensations related to their dynamic loudness. Descriptions like “low”/“high” or “thick”/“thin” indicate that noises of the same center frequency (500 or 2000 Hz) shared the same pitch or brightness feature. In the complex-tone condition, subjects described the 4- and 20-harmonic tones as being played on two different instruments or as “smooth” versus “scratchy,” indicating that by manipulating the number of harmonics, two timbral auditory features were elicited. The sensations evoked by sounds of the same fundamental frequency were described as different musical notes or as being “low,” “medium,” or “high,” suggesting that pitch varied as fundamental frequency was manipulated.

DISCUSSION

The major advantage of the method we have presented is that it aims at a feature representation of auditory stim-

uli on the basis of simple qualitative judgments. In spite of the simplicity of the answer (“yes”/“no”), one should not overlook the fact that the decision process to reach such an answer is potentially rather complex. In order to reduce the demand for the subject, it is crucially important to include a familiarization segment in the experimental session, together with instructions for the subject to structure the stimuli and to recognize, identify, and organize the auditory features. With more complex stimuli than the ones used in the present study, it might be fruitful to split up the familiarization into an introduction to the *task*, potentially in the form of a tutorial using pictures of simple geometric shapes and eye-catching visual features, and an actual familiarization with the *sounds* as described earlier.

We hypothesized that the success of the method—the highly consistent judgments and the large number of representations obtained—was partly due to the simple acoustical structure of the stimuli. The high demands of the task for the subject become more obvious as the sounds under study become more complex. In such a case, eventually labeling the auditory features might require more accurate information than could be obtained in the present study by an informal debriefing. A possible strategy for a more formal debriefing would be to present a listener with the sets of sounds from his or her own auditory structure, along with the request “Describe briefly what feature(s) these *n* sounds share with each other, but not with the remaining sounds” (Choisel & Wickelmaier, 2006). The answers to such questions might also provide valuable hints for selecting the most appropriate structure in the case of a nonunique representation (see the Results section above).

A restriction of the presented approach is that it requires a certain independence from the *local* context provided by

a given triad of sounds. Instead, the decision whether or not a feature belongs to a sound has to be based on the *global* context provided by all sounds under study. Features based on local context effects will most likely result in inconsistent judgments. It is possible, however, that a subject may learn to appreciate new features during the data collection part of the experiment that were not identified during the familiarization part. Consider, as an example, subject K.P. in the complex-tone condition. Throughout the experiment, her judgments were perfectly consistent, indicating that she could clearly identify the salient features. There were, however, 18 response changes between the first and second blocks. In combination with the perfect consistency, this should not be taken as unreliable behavior, but rather as evidence that learning of new features occurred after the data collection part of Block I, and these features were then consistently judged in the remaining blocks.

As this example suggests, a subject might not identify all relevant features at once when the sounds are presented for the first time during familiarization. Therefore, we recommend running several repetitions of familiarization and subsequent data collection, as described earlier. One might argue that data should be collected until the responses are clearly stable; this might be done by including additional blocks of data collection until block-to-block response changes no longer occur. In doing so, the consistency of the judgments would presumably increase as well. Care has to be taken, however, not to exhaust the subjects by posing an unreasonably large number of questions. Future experimental work should address whether and how increased exposure to the stimuli affects individual responses and structures. Future development of the method might focus on how a probabilistic framework could be adapted to feature structures, which would allow for the identification of a feature to be modeled as probabilistic event and for the estimation of probabilities of careless errors.

With more complex stimuli, the classification of transitivity violations as random or systematic would also become more difficult. Unfortunately, as with other applications of axiomatic measurement theory, there are no simple criteria for such a classification. Rather, the indexes of reliability and consistency, their development over time (i.e., blocks), and the discrepancy $\delta_Q(\sigma)$ must be considered together in order to decide whether there is enough evidence in the data to allow violations to be classified as random, and consequently whether the transitivity axiom holds. A statistical test would certainly remedy the problem, but such a test has not yet been developed. In the absence of such a test, it is common practice to relate the number of violations of an axiom to the number of possible tests of that axiom, implying that a higher violation *ratio* is indicative of stronger evidence against that axiom holding. Such a strategy, however, cannot be advocated for perceptual structures, since the more transitivity tests that are possible, the more frequently a subject must have responded with a “no,” which in turn would imply that only a few features had to be considered. The more features a subject has in mind, the more complex Q will become, and the *fewer* transitivity tests will be possible. For that reason,

the number of possible tests can be misunderstood and was therefore not reported in the present study.

Concluding Remarks

It appears that auditory structures provide a viable method for identifying the auditory features relevant for a given domain. They promise to be particularly useful when—in contrast to the present proof-of-concept experiment—the relevant features are unknown a priori, such as in investigations of the audio quality of loudspeaker systems, in studies of music perception, or when the sound properties constituting good product sound have to be identified. In such situations as these, finding the relevant attributes is a prerequisite for subsequently scaling their perceived magnitude, and a method achieving this goal in a tractable way is highly desirable.

On the basis of the present initial study, the following conclusions may be drawn: (1) Subjects were able to produce reliable and consistent judgments about common auditory features of simple synthetic sounds. (2) The proposed measurement-theoretically founded approach for deriving auditory features was shown to have the advantages that it (a) can reveal a failure to represent a participant’s judgments when they are inconsistent (the representation is falsifiable); (b) provides an opportunity to test the identifiability of auditory features; and (c) does not require labeling of the features encountered, and therefore distinguishes the perceptual and verbal abilities of subjects. (3) The results from the present study encourage the application of the method to more complex auditory stimuli (Choisel & Wickelmaier, 2006), and potentially to investigating features within other perceptual modalities.

AUTHOR NOTE

This research was carried out as part of the Centerkontrakt on Sound Quality, which establishes participation in and funding of the Sound Quality Research Unit (SQRU) at Aalborg University. The participating companies are Bang & Olufsen, Brüel & Kjær, and Delta Acoustics & Vibration. Further financial support came from the Ministry for Science, Technology, and Development (VTU) and from the Danish Research Council for Technology and Production (FTP). The authors are grateful to Sylvain Choisel for numerous helpful discussions of the theory and implementation of auditory feature structures, and to Ville Sivonen for helping with the instrumental analyses of the stimuli. We also thank Ragnar Steingrímsson and Jürgen Heller for their helpful comments on an earlier draft of the manuscript. W.E. is now at Technische Universität Darmstadt (e-mail: ellermeier@psychologie.tu-darmstadt.de). Correspondence concerning this article should be sent to F. Wickelmaier, Department of Psychology, University of Tübingen, Friedrichstrasse 21, 72072 Tübingen, Germany (e-mail: florian.wickelmaier@uni-tuebingen.de).

REFERENCES

- BEALS, R., KRANTZ, D. H., & TVERSKY, A. (1968). Foundations of multidimensional scaling. *Psychological Review*, *75*, 127-142.
- BECH, S., & ZACHAROV, N. (2006). *Perceptual audio evaluation: Theory, method and application*. Chichester, U.K.: Wiley.
- BERG, J., & RUMSEY, F. (2006). Identification of quality attributes of spatial audio by repertory grid technique. *Journal of the Audio Engineering Society*, *54*, 365-379.
- BORG, I., & GROENEN, P. (1997). *Modern multidimensional scaling: Theory and applications*. New York: Springer.
- CARROLL, J. D., & CHANG, J.-J. (1970). Analysis of individual differences in multidimensional scaling via an n -way generalization of “Eckart–Young” decomposition. *Psychometrika*, *35*, 283-319.
- CHOISEL, S., & WICKELMAIER, F. (2006). Extraction of auditory features

- and elicitation of attributes for the assessment of multichannel reproduced sound. *Journal of the Audio Engineering Society*, **54**, 815-826.
- DOIGNON, J.-P., & FALMAGNE, J.-C. (1999). *Knowledge spaces*. Berlin: Springer.
- FALMAGNE, J.-C., KOPPEN, M., VILLANO, M., DOIGNON, J.-P., & JOHANNESSEN, L. (1990). Introduction to knowledge spaces: How to build, test, and search them. *Psychological Review*, **97**, 201-224.
- GANTER, B., & WILLE, R. (1999). *Formal concept analysis: Mathematical foundations* (C. Franzke, Trans.). Berlin: Springer.
- GREY, J. M. (1977). Multidimensional perceptual scaling of musical timbres. *Journal of the Acoustical Society of America*, **61**, 1270-1277.
- HELLER, J. (1991). *Eine experimentelle und theoretische Untersuchung zur Begriffsbildung* [An experimental and theoretical investigation of concept formation]. Unpublished doctoral thesis, University of Regensburg.
- HELLER, J. (2000). Representation and assessment of individual semantic knowledge. *Methods of Psychological Research*, **5**, 1-37.
- ISO 389-1 (1998). Acoustics—Reference zero for the calibration of audiometric equipment—Part 1: Reference equivalent threshold sound pressure levels for pure tones and supra-aural earphones. Geneva: International Organization for Standardization.
- IVERSON, P., & KRUMHANSL, C. L. (1993). Isolating the dynamic attributes of musical timbre. *Journal of the Acoustical Society of America*, **94**, 2595-2603.
- KELLY, G. A. (1955). *The psychology of personal constructs*. New York: Norton.
- KRANTZ, D. H., LUCE, R. D., SUPPES, P., & TVERSKY, A. (1971). *Foundations of measurement: Vol. 1*. New York: Academic Press.
- KRUSKAL, J. B. (1964). Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. *Psychometrika*, **29**, 1-27.
- LAKATOS, S., MCADAMS, S., & CAUSSÉ, R. (1997). The representation of auditory source characteristics: Simple geometric form. *Perception & Psychophysics*, **59**, 1180-1190.
- SATTATH, S., & TVERSKY, A. (1987). On the relation between common and distinctive feature models. *Psychological Review*, **94**, 16-22.
- SHEPARD, R. N. (1962). The analysis of proximities: Multidimensional scaling with an unknown distance function. Part I. *Psychometrika*, **27**, 125-140.
- STONE, H., & SIDEL, J. L. (2004). *Sensory evaluation practices*. Amsterdam: Elsevier Academic Press.
- TORGERSON, W. S. (1952). Multidimensional scaling: I. Theory and method. *Psychometrika*, **17**, 401-419.
- TVERSKY, A. (1977). Features of similarity. *Psychological Review*, **84**, 327-352.
- WILLE, R. (1982). Restructuring lattice theory: An approach based on hierarchies of concepts. In I. Rival (Ed.), *Ordered sets* (pp. 445-470). Dordrecht: Reidel.

NOTES

1. The influence of the headphones on the signal was neglected in this study. We expected that equalizing for the headphone transfer functions (which imitate a diffuse field) would not improve the identifiability of the auditory features elicited by the stimuli.
2. Heller (2000) and Choisel and Wickelmaier (2006) have developed software tools to provide assistance in finding a feature structure potentially underlying the responses in the presence of transitivity violations.

(Manuscript received July 6, 2005;
revision accepted for publication June 16, 2006.)