INTERNATIONAL JOURNAL OF PSYCHOLOGY, 1996, *31* (6), 255–278

Systematic Analysis of Verbalizations Produced in Comparing Musical Timbres

Elena Samoylenko
Maison des Sciences de l'Homme, Paris, France and Russian Academy of Sciences, Moscow, Russia

Stephen McAdams
Laboratoire de Psychologie Expérimentale (CNRS), Université René Descartes, and IRCAM, Paris, France

Valery Nosulenko
Université René Descartes, Paris, France and Russian Academy of Sciences, Moscow, Russia

A method of verbal protocol analysis is presented in which verbal units used to compare sound events are extracted from the protocols and analysed from three general points of view: (1) their logical sense, (2) their stimulus-relatedness, and (3) their semantic content. At each level, several analysis steps progressively label the verbal units from hierarchically superior to inferior levels (with increasing differentiation as one descends the hierarchy). The technique shows great reliability across analyses by independent experts. It was applied to an investigation of timbre comparisons in which Russian nonmusician subjects rated the dissimilarity of timbres presented in pairs and described verbally their similarities and differences. The technique demonstrates an effectiveness in revealing differences in focus on perceptual attributes of timbre that contribute to inter- and intra-subject discrepancies in dissimilarity ratings. Further, as concerns identification of the sound sources producing the events, there is evidence that basic categories are very general in nature in these nonmusician subjects.

Cet article présente une méthode d'analyse de protocoles verbaux qui extraie les unités verbales utilisées pour comparer des événements sonores et analyse ces unités sur trois niveaux généraux: (1) leur sens logique; (2) leur relation avec le stimulus; et (3) leur contenu sémantique. Pour chaque dimension, plusieurs étapes d'analyse permettent d'étiqueter progressivement les unités verbales en passant de niveaux hiérarchiquement supérieurs à des niveaux inférieurs (avec une différenciation croissante). La technique démontre une grande fiabilité pour des analyses faites par des experts indépendants. Elle a été appliquée dans une étude de comparaisons de timbre au cours de laquelle des sujets russes non musiciens évaluaient la dissemblance de timbres présentés en paires et décrivaient verbalement leurs similarités et différences. La technique se montre efficace pour révéler dans quelle mesure des différences dans l'importance accordée aux attributs perceptifs des timbres contribuent aux divergences intersujets et intrasujets dans l'évaluation de la dissemblance. De plus, les données suggèrent que les catégories de base utilisées par ces sujets non musiciens pour identifier les sources sonores produisant les événements sont très générales.

Requests for reprints should be addressed to Dr Stephen McAdams, Laboratoire de psychologie expérimentale, 28 rue Serpente, F–75006 Paris, France.

Portions of these results were presented at the Troisième Congrès Français d'Acoustique, Toulouse, and were reported in Nosulenko, Samoylenko, and McAdams (1994). This collaboration was made possible by financial support for the project "Homme-Environnement Acoustique" from the Fondation de la Maison des Sciences de l'Homme in Paris. We express our gratitude to Clemens Heller, Maurice Aymard, and Jean-Luc Lory for their continuing interest and support. Address correspondence to Valery Nosulenko at the Institute of Psychology, Russian Academy of Sciences, 13 Yaroslavskaya Street, 129366 Moscow, Russia (email:lena@rccc.msk.su, smc@ircam.fr, valery@rccc.msk.su).

INTRODUCTION

The technique for analysing verbal data reported in this paper seeks to explore the perceptual and cognitive processes involved in two activities: the perceptual analysis of sound and the activation of a subject's lexical and grammatical structures when comparing perceptual representations. Our analysis attempts in part to determine the nature of the lexicon and grammatical structures used by a group of subjects with similar language and sociocultural experience who are asked to compare complex sound events. This approach presupposes that subjects use a perceptual representation to activate a semantic and linguistic representation that is then embodied in a verbal report.

Two different approaches to exploring the link between verbal lexicon and perceptual representation have been used in the literature: verbal attribute scaling and free verbalization. In the first approach, a subject is presented with a list of verbal attributes selected beforehand by the experimenter. In one paradigm, the subject is asked to check those that are the most appropriate for each sound stimulus (e.g. Radocy & Boyle, 1979). In another, subjects are presented with bipolar scales (Osgood, Suci, & Tannenbaum, 1957) and are asked to rate each sound along the semantic continuum the scale represents (e.g. von Bismarck, 1974). A variant of this latter technique uses unipolar scales along which the magnitude of the verbal attribute is rated (see Kendall & Carterette, 1992a, b). These techniques often encounter problems when applied to musical timbre. One is that the chosen vocabulary may have little relation to sound and may thus be considered inappropriate. A second problem is that the studies are often not reproducible with different sets of sounds or different subject populations. A third is that the relations between classes of verbal attributes and acoustic properties are often quite weak (see Handel, 1989, chap. 8; Kendall & Carterette, 1992a, for critical reviews).

Since this article concerns the free verbalization approach, we will examine it in more detail both in terms of its underlying assumptions and in terms of its application to the perception of complex sound events. We will then present the verbal protocol analysis technique that was designed to make use of free verbalizations in comparing the timbres of musical sounds followed by the experiment and several analyses using the technique.

Free Verbalization

There are several groups of procedures to obtain verbalizations produced in a free form by subjects. The validity and appropriateness of using free verbalization procedures in the investigation of perception and cognition poses a methodological problem that has not been resolved in a completely satisfactory manner. Generally speaking, a theoretical prerequisite to solving the problem is the researcher's conception of whether it is possible to express the specific aspects of perception and cognition with language, a problem that raises the spectre of the scientific status of verbal reports.

There are three points of view on this issue. The first assumes an isomorphic relation between the cognitive processes involved in performing certain tasks and the processes identified in verbal reports produced during or after such tasks. The second (opposing) point of view rejects this kind of isomorphism on the grounds that introspection has an extremely limited potential of gaining access to the psychological processes involved in perceiving, acting, and thinking. It should be noted that this latter point of view has been widely criticized within the past few years (e.g. Ericsson & Simon, 1984; Smith & Miller, 1978). A third (intermediate) point of view on this problem seems to be the most fruitful. It seeks to determine the conditions under which verbal data may be used to study cognitive processes in a systematic and rigorous manner (e.g. Brommel, 1983; Caverni, 1988; Cuni, 1979; Ericsson & Simon, 1984; Hoc, 1984; Leplat & Hoc, 1981; Newell, 1977; Smith & Miller, 1978). Caverni (1988), for example, considers two of the main kinds of objections:

1. the verbalization process modifies the execution of the (perceptual or cognitive) task and therefore affects performance, and
2. mental processes are not accessible by verbalization, thus making verbal data:
   a. incomplete—not all components of a mental process can be accounted for verbally;
   b. epiphenomenal—verbalizations have nothing to do with the processes that are operational during the execution of the task being studied, and
   c. unexploitable—the analysis of verbal protocols, with the aim of subsequently exploiting the results, cannot be objective.

Caverni (1988) counters these objections as follows. Concerning (1), one can hypothesize that the deterioration of performance with concomitant verbalization may only be observed when the execution of the task is not naturally performed in a verbal code. In response to (2a,b), he notes that the origin of these objections lies in the putative impossibility of elaborating models of cognitive function on the basis of verbalizations, but that other research has refuted this objection by successfully comparing verbal protocols to task protocols obtained with other observable behaviours. Finally, the objection in (2c) may be refuted: (i) if one seeks indicators on the basis of a model of the process being studied, (ii) if the identification of the indicators proceeds on the basis of consensus or according to rules that are explicit and falsifiable, and (iii) if, in addition, the coding of the verbal protocol can be confronted with other protocols collected for the same kind of task. If these conditions are met, the verbal protocols can be considered to have the guarantees necessary for the elaboration of exploitable data.

In the methods using free verbalizations for the analysis of subjective representations, instructions require subjects to describe perceived objects (rather than thinking processes) step by step. This method has been applied to visual perception of objects and phenomena (Lange, 1893; Nikitin, 1905; Bartlett, 1932).

The method of free verbalization has also been used to study auditory representations. In one study (Wright, 1971), the absence of a readily available vocabulary for describing unfamiliar sounds of differing pitch, intermittence, and modulation was shown. In another study, verbalizations served to reveal the influence of occupational background on an individual's descriptions of unfamiliar auditory signals (Taylor, Gandy, & Dark, 1974). Those verbalizations did not have the form of detailed descriptions, but were simple word combinations or single utterances. The verbalizations produced by the subjects were divided into three types: onomatopoeic (e.g. "bleep," "buzz," "pop," "pip"), illustrative (e.g. "fog horn," "tug boat," "bass instrument," "siren"), and physical characteristics (e.g. "high," "low," "long," "short," "slow," "fast").

An analysis of the problem of using verbal descriptions for the study of sound perception was performed by Sokolov (1887). Of particular interest in his study was the attempt to establish a relation between verbalizations and the physical characteristics of sounds, such as intensity or spectral qualities. For example, for different intensities and spectra, the subjects might describe the sounds as "thin" or "thick," "sharp" or "dull," "liquid" or "viscous," "light" or "heavy," etc. (Sokolov, 1887, p. 398 [our translation]). Based on verbal descriptions, he concluded, for example, that the high harmonic frequencies associated with the fundamental frequency do not increase the (spatial) volume of the sound but rather reduce it, making the sound more "sharp, piercing, and definite" (Sokolov, 1887, p. 401 [our translation]).

An important role of verbal description in the process of memory and recognition of complex sounds has been demonstrated by the work of Bower and Holyoak (1973). The authors concluded from their study that in the recognition of sounds, verbalizations had a weight that was relatively greater than that of sensory parameters. Subjects recognized natural sounds by reconstructing the (most often visual) images of the corresponding sources from verbalizations. As such, the content of the verbalizations determined the cue for identifying the sound.

There remains, however, the problem of correlating acoustic parameters of sound and qualitative data from descriptions of these sounds. An important task consists of elaborating an approach that allows one to obtain reliable interpretations of sound events from their verbal descriptions that concord with exact measures based on physical methods. Some studies on the elaboration of a technical vocabulary concerning sound have shown this to be possible. One example is work by Kouznetsov (1981), who presented a list of terms characterizing the timbre of musical instrument sounds. The frequency of use of different terms was found to be a function of the spectral content of the sound: 200–900Hz—"succulent," "deep"; 800–2500Hz—"velvety"; 2500–8000Hz—"bright"; 3000–6300Hz—"piercing" [our translation].

In our own studies of auditory perception, we have integrated psychophysical and verbal communication methods within a single experimental paradigm (Lomov, Belyaeva, & Nosulenko, 1986; Nosulenko, 1988, 1989a; Nosulenko & Samoylenko, 1992; Samoylenko, 1986). The results of this research have led us to conclude that there exists a reasonably precise relation between data obtained by physical methods (subjective ratings of differences between sounds) and qualitative indices of the use of verbal units in the description of differences.

Our research programme seeks a new approach to the verbal characterization of musical timbre. The present article focuses on nonmusicians' verbal behaviour. The subjects were asked to listen to pairs of sounds and to describe in as much detail as possible the differences and similarities among their timbres. The principal feature of our approach consists of the idea that the verbalizations produced by subjects when solving a cognitive task should have a free form. Thus, the development of a method for analysing verbal protocols is an important task as it attempts to reveal the various linguistic devices (lexical, syntactic, semantic) the subjects used to compare the sounds.

VERBAL PROTOCOL ANALYSIS PROCEDURE

In our previous research, we have shown that certain verbalization strategies used by subjects in comparison and rating tasks can indicate the perceived difference between described objects or events. The principles of classification of those strategies were also defined (Samoylenko, 1986; Nosulenko, 1988, 1989b; Nosulenko & Samoylenko, 1992). On the basis of these results, a detailed scheme for analysing verbal data was elaborated for comparisons of complex sounds (see Fig. 1). This method is summarized briefly here and a detailed description is included in the Appendix.

Each verbal text, produced by a given subject in comparing a pair of sounds, is considered at three main levels of analysis of the verbal units: their *logical sense*, their *stimulus-relatedness*, and their *semantic aspects*. At each main level, a series of hierarchical analysis steps are performed.

Logical Sense

The analysis of the logical sense of verbal units is performed in three steps. First, verbal units containing descriptions of stimulus *similarities* and *differences* are identified and marked accordingly. Then an analysis of the level of generality of similarity and difference is made, i.e. verbal units are labelled as representing terms corresponding to a *general* basis of comparison or as expressing *concrete* (or detailed) comparisons. Finally, verbal units are analysed according to the way sounds are opposed. They are consequently categorized as *classificational* or *gradual*. In classificational verbal units, the two sounds are identified as possessing different qualities, or a single sound of the pair is characterized. In gradual verbal units, the two stimuli are compared with respect to the same semantic class, but a degree of difference between the two objects is noted.

Stimulus-Relatedness

The analysis of stimulus-relatedness of verbal units is performed in two steps. Verbal units are labelled according to whether they are used to describe global aspects of the sounds (*sound Gestalts*) or *specific properties* of the sounds. Subsequently, the nature of specific properties is further classified as being *spatial*, *temporal*, *intensive*, or *spectral*.

Semantic Aspects

The analysis of semantic aspects of verbal units aims to determine the semantic categories used to describe the two sounds being compared. This level contains from three to five analysis steps, depending on the branch of the analysis tree (Fig. 1). First a division into two principal categories is made: *features* and *holistic meaningful entities*. A description of separate features means that distinguishing properties of sounds are represented without assigning a holistic meaning to those sounds.

Separate features are first divided according to whether they are *descriptive* or *attitudinal* (expressing an emotional or evaluative attitude to perceived sounds). The group of descriptive features is divided on the one hand into those normally used by people only to describe sound phenomena, that is, features referring only to the auditory modality (*unimodal*) and, on the other hand, into features that can be used to describe subjective representations for different sensory modalities (*polymodal*). The group of descriptions expressing *attitudes* to perceived sounds is always considered to be polymodal and is divided into those containing information about *emotional* relations to sounds and those where an aspect of *naturalness* is mentioned. The units containing information about emotional relations to sounds

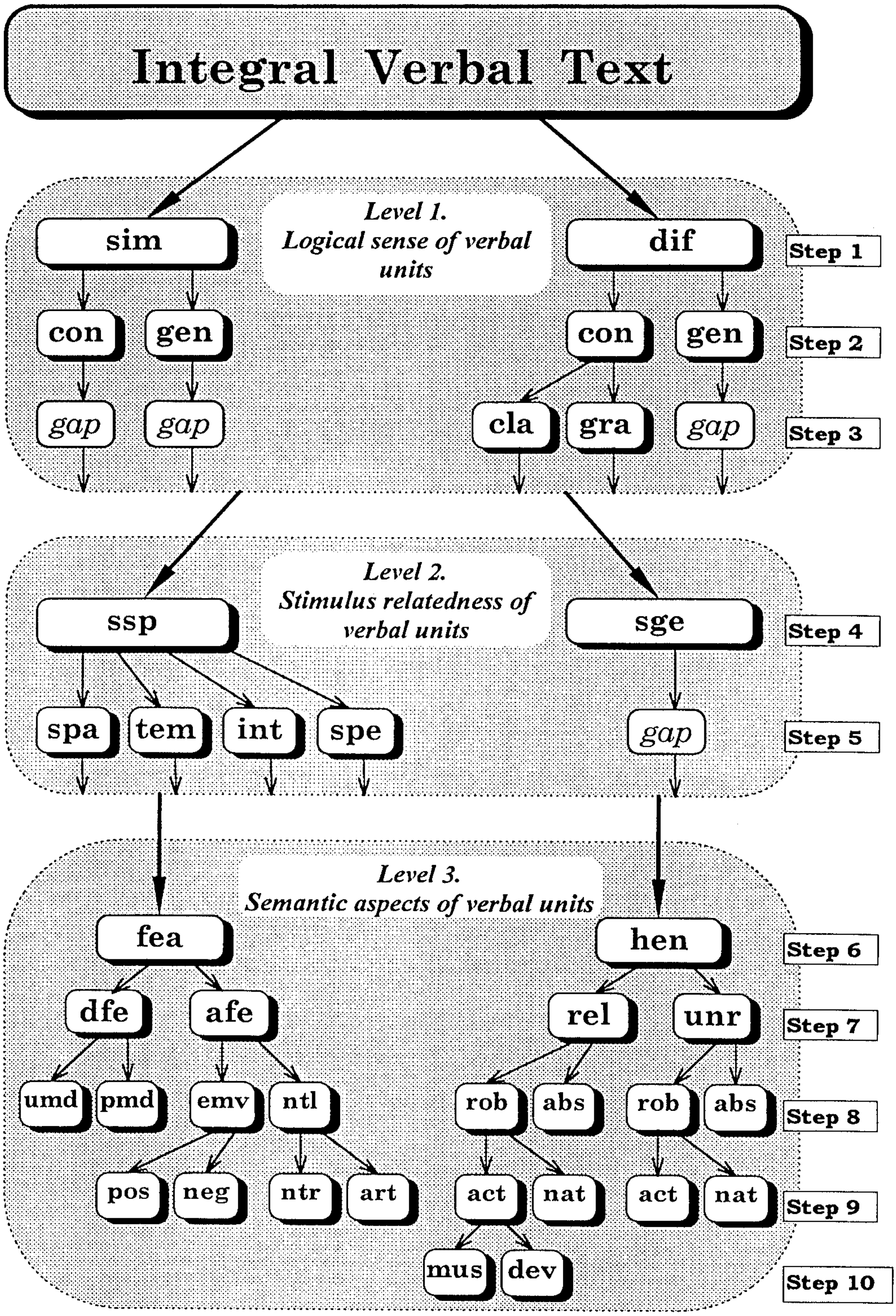

FIG. 1. Scheme of verbal data analysis. Arrows between levels indicate a possible branching to any class at the next level down (i.e. to \ssp\ or \sge\ at Step 4 and to \fea\ or \hen\ at Step 6). A full description of the analysis steps is found in the Appendix.

are further divided into those describing *positive* and *negative* aspects. The other group of descriptions is divided into those indicating the *naturalness* of a sound and those mentioning its *artificial* nature.

Holistic meaningful entities are divided first into the verbal units that are specifically *related* to the production of sound phenomena and those that are usually *unrelated* to sound phenomena. Then holistic meaningful entities are classified into units containing information either about *real objects* or about *abstract concepts*. Verbal units concerning concrete objects not designed to produce sound are finally divided into those concerning *natural objects* and phenomena and those concerning objects used in various *human activities*. As for descriptions concerning sound-related objects, they are also divided at this step into those concerning *natural objects* habitually involved with sound activity and those concerning *human-made objects* used in sound production. Finally, the means of sound reproducing activity are divided into *musical instruments* and other *sound reproducing devices*.

We assumed in constructing this scheme of verbal protocol analysis that descriptions of both distinctive features and references to sound-producing objects could be subdivided in several different ways. The chosen principles of subdivision were determined by the general goals and hypotheses of our experimental studies, as well as through preliminary attempts to apply the technique to real verbalizations.

METHOD

Subjects

The 14 Russian nonmusician subjects (7 males and 7 females) were recruited in Moscow. Their ages ranged from 21 to 40 years (M = 31.5). They produced verbal protocols in Russian. Each subject performed the experiment twice.

Stimuli

The sounds were synthesized on a Yamaha TX802 FM Tone Generator. The 17 "instruments" used were drawn from a set developed by Wessel, Bristow, and Settel (1987). The sounds were designed either to imitate closely many traditional Western instruments (e.g. clarinet, harpsichord, bowed string, trumpet, vibraphone) or to represent a chimaerical hybrid of known instruments (e.g. "vibrone" is a hybrid of vibraphone and trombone and "guitarnet" is derived from guitar and clarinet). All tones were produced at the same pitch (fundamental frequency = 311Hz, or E-flat4) and were perceptually equalized for duration (approximately 600 msec) and loudness (mezzo forte = approximately 50 on the MIDI velocity scale that controls intensity in commercial digital synthesizers). The instrument pairs were selected from those used in a study by McAdams and Cunibile (1992) which investigated the perception of timbral analogies. The instruments used in this experiment are listed in Table 1. In order not to confuse these synthetic sounds with the sounds of the acoustic instruments many of them were designed to imitate, their three-letter abbreviations (Table 1) will be used throughout the text.

Procedure

The stimuli were presented to subjects over a loudspeaker at a comfortable listening level in a sound-treated room. All of the sounds used were presented at the beginning of the session in ran-

TABLE 1
Synthesized Instruments Used in the Study

Instrument Name[a]	*Letter Code*[b]	*Numerical Code*[c]
French horn	HRN	1
Trumpet	TPT	2
Trombone	TBN	3
Trumpar (trumpet/guitar)	TPR	5
Vibraphone	VBS	7
Striano (bowed string/piano)	SNO	8
Sampiano (sampled piano)	SPO	9
Harpsichord	HCD	10
Bassoon	BSN	13
Clarinet	CNT	14
Vibrone (vibraphone/trombone)	VBN	15
Obochord (oboe/harpsichord)	OBC	16
Pianobow (bowed piano)	PBO	17
Guitar	GTR	18
Bowed string	STG	19
Piano	PNO	20
Guitarnet (guitar/clarinet)	GTR	21

[a] Names of the acoustic instruments that the synthesis instruments were meant to imitate, or the names of the hybrids they were meant to create (with the names of their progenitors).

[b] Letter code used in the text.

[c] Number code used in the database.

dom order to familiarize the subject with them and to establish a sense of the range of variation. Afterward 6 practice trials were presented, followed by the 46 experimental trials comprising the 23 timbre pairs in Table 2 presented in the 2 possible orders.

On each trial subjects listened individually to sound pairs that could be replayed as many times as necessary during the trial. A numerical rating of the dissimilarity of the two sounds was first made. The scale ranged from 1 to 8, where 1 signified very similar and 8 very dissimilar. Then subjects were asked to compare the sounds and to describe verbally the similarities and differences between them in as much detail as possible. These protocols were recorded on tape, transcribed verbatim, and analysed by the procedure described above.

TABLE 2
Instrument Pairs Selected from Timbral Analogies in McAdams and Cunibile (1992) for Presentation in the Present Study[a]

Numerical Code	*Instrument Names*
1–17	French horn—Pianobow
2–5	Trumpet—Trumpar
2–8	Trumpet—Striano
2–19	Trumpet—Bowed string
3–7	Trombone—Vibraphone
3–10	Trombone—Harpsichord
3–14	Trombone—Clarinet
3–15	Trombone—Vibrone
5–7	Trumpar—Vibraphone
5–17	Trumpar—Pianobow
5–18	Trumpar—Guitar
5–19	Trumpar—Bowed string
7–10	Vibraphone—Harpsichord
7–19	Vibraphone—Bowed string
8–9	Striano—Sampiano
8–17	Striano—Pianobow
9–13	Sampiano—Bassoon
9–17	Sampiano—Pianobow
9–18	Sampiano—Guitar
9–19	Sampiano—Bowed string
10–16	Harpsichord—Obochord
17–20	Pianobow—Piano
17–21	Pianobow—Guitarnet

[a] See Table 1 footnote for abbreviations.

DATABASE OF VERBAL UNITS

Each verbalization obtained in the experiment was tape-recorded and transcribed into a text file which was independently analysed by three experts. Differing opinions concerning a particular verbal unit were subsequently resolved by consensus, but all such discrepancies were noted.

In the first phase of analysis, the meaningful verbal units were extracted from the texts. Each unit corresponded to a separate characteristic of a sound. For example, in the description "a very rich, high sound," there are two verbal units: "very rich" and "high," whereas in the description "resembles a hammer which falls and returns to its place" the whole phrase is one verbal unit. An analysis of the identification of verbal unit boundaries revealed that the number of discrepancies represented less than 0.5% of the total number of verbal units. The discrepancies concerned primarily the level of detail of a certain sound characteristic: for example, "very constant in pitch" versus "very constant," "steady part spectrally richer" versus "spectrally richer." In a small number of cases, the discrepancies concerned the number of verbal units identified. For example, "a bad imitation of a plucked string, harpsichord-like sound" was finally divided into two verbal units: "bad imitation" (meaning an artificial aspect of the sound), and "a plucked string, harpsichord-like sound" (meaning a complex referent). Finally, a small number of verbal units were not taken into consideration if it was impossible to determine whether the sound being referred to was the first or the second of the pair.

In the second phase, the verbal units were entered into a database[1] (see Fig. 2) along with the subject's number (Sb), the number of the trial in which the timbre pair was presented (St), the instrument codes for each timbre of the pair in their order of presentation (T1 and T2, see Table 1), the dissimilarity rating (Ev), and the instrument code for the timbre to which the verbal unit referred (Td).

In the third phase of database construction, three experts analysed the verbal units (using the scheme of verbal data analysis described ear-

[1] The database described in this article is available to researchers in psychology, acoustics, and linguistics who have scientific questions that they feel may be addressed by probing these data. A programme for specifying flexible searches according to the analysis scheme is written in Paradox 4.5 and runs on IBM PC-compatible machines. The database may be obtained by writing to Valery Nosulenko, Maison Suger, 16–18 rue Suger, F–75006 Paris, France.

N°	Sb	St	T1	T2	Ev	Td	Verbal units	St1	St2	St3	St4	St5	St6	St7	St8	St9	St10
11	4	3	16	10	2	16	***pitch***	sim	cgb	gap	ssp	spe	fea	dfe	pmd	gap	gap
12	4	3	16	10	2	10	***pitch***	sim	cgb	gap	ssp	spe	fea	dfe	pmd	gap	gap
13	4	3	16	10	2	16	***guitar***	sim	con	gap	sge	gap	hen	rel	rob	act	mus
14	4	3	16	10	2	10	***guitar***	sim	con	gap	sge	gap	hen	rel	rob	act	mus
15	4	4	17	8	1	17	***very artificial***	sim	con	gap	sge	gap	fea	afe	ntl	art	gap
16	4	4	17	8	1	8	***very artificial***	sim	con	gap	sge	gap	fea	afe	ntl	art	gap
52	4	10	2	19	2	2	***more sharp***	dif	con	gra	sge	gap	fea	dfe	pmd	gap	gap
53	4	10	2	19	2	19	***"less sharp***	dif	con	gra	sge	gap	fea	dfe	pmd	gap	gap
54	4	10	2	19	2	19	***blunt***	dif	con	cla	sge	gap	fea	dfe	pmd	gap	gap
120	4	20	18	5	7	5	***very boring***	dif	con	cla	sge	gap	fea	afe	emv	neg	gap
121	4	20	18	5	7	5	***natural***	dif	con	cla	sge	gap	fea	afe	ntl	ntr	gap
122	4	20	18	5	7	18	***cleanness***	dif	cgb	gap	sge	gap	fea	dfe	pmd	gap	gap
123	4	20	18	5	7	5	***cleanness***	dif	cgb	gap	sge	gap	fea	dfe	pmd	gap	gap

FIG. 2. Example of listing from the database. Sb = subject number, St = trial number for stimulus pair, T1 = first timbre, T2 = second timbre (numbers for timbres refer to instrument number in the Yamaha synthesizer; see Table 1), Ev = dissimilarity rating, Td = instrument number of sound being described, Stl–St10 = analysis steps (see Fig. 1). Verbal units are translated from Russian.

lier) and filled in the corresponding cells for each analysis step. The possible codes for each step are shown in Fig. 1. In the case of a skip across an analysis step (for example, across Step 5 for units labelled as \sge\ in Step 4), the corresponding cell in the database was coded with \gap\ (see Fig. 2).

For verbal units corresponding to similarity descriptions \sim\ and general difference descriptions \dif\gen\, identical entries for each of the timbres were made in the database. For instance, in the utterance, "Sounds are close in their pitch; both resemble a guitar but both are very artificial," the verbal units "pitch," "guitar," and "artificial" were doubled in the database, once for each of the 2 timbres 16 and 10 (see entries No. 11, 12, 13, and 14 in Fig. 2).

A different kind of doubling was performed for the verbal units referring to a gradual comparison \gra\. For instance, from the utterance, "The first one is more sharp," produced for the pair of sounds (2–19), two verbal units were entered into the database: "more sharp" (for Td2) and "less sharp" (for Td19) (see entries No. 52 and 53 in Fig. 2). The double quote preceding the verbal unit for timbre 19 indicates that it is an inverted copy of the unit actually produced.

The database entries for the 10 analysis steps were prepared independently by 3 experts. After a joint discussion of these versions and consensus adjustments made in cases of discrepancy, the final version of the database was constructed.

It should be noted that during the construction of the database, the global rate of discrepancies between the experts was less than 2%. This rate was calculated as the ratio between the total number of discrepancies and the total number of judgements made by the 3 experts for all 10 analysis steps (fields in the database), applied to all verbal units. However, since the classification of a verbal unit in later steps of the analysis depends in many cases on earlier steps, discrepancies were only included in the total if all prior steps were labelled in the same way by all three experts. An examination of this rate for the different steps of analysis showed that most of them concerned decisions in Steps 4 to 6. Figure 3 shows the proportion of discrepancies in labelling at a given step with respect to the total number of discrepancies across all steps. In spite of the negligible quantity of discrepancies, an analysis of their nature was of particular interest, in allowing us to specify more precisely the rules by which the verbal protocols should be analysed.

In the final stage of constructing the database, the entries for all subjects were collected into a single file in order to allow comparative analyses. More than 7300 entries, each with 17 fields, were thus obtained in this study. A comparative analysis of the databases for each subject revealed important individual differences in the use of verbal units. These differences were first mani-

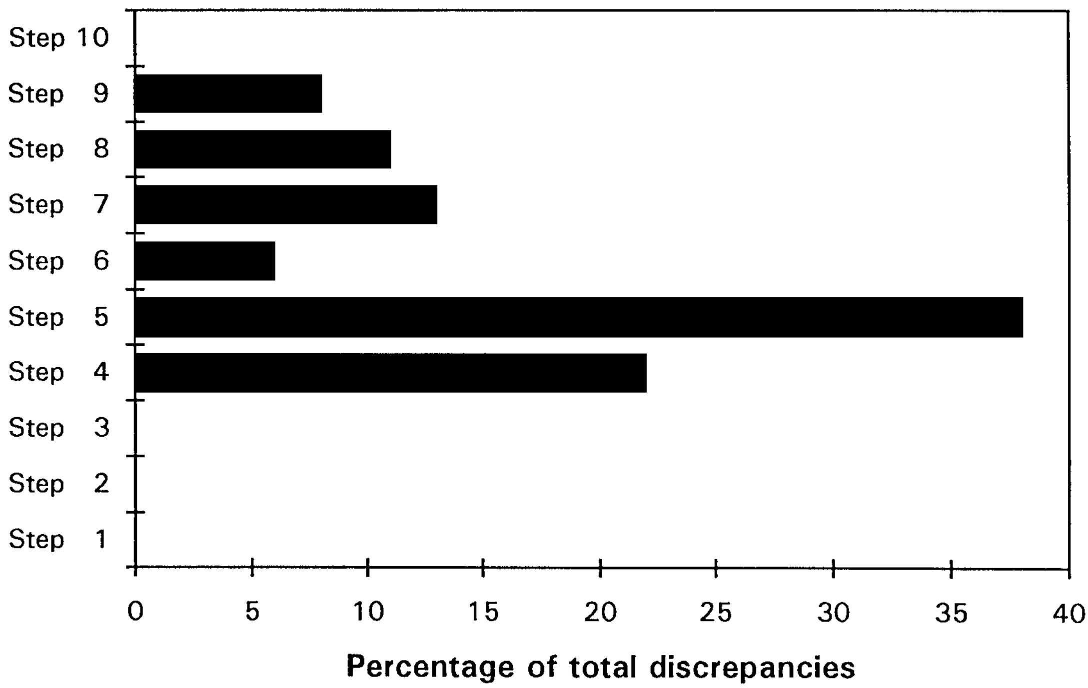

FIG. 3. Proportion of entries (expressed as percentage of total number of discrepancies) giving rise to discrepancies in labelling between the three experts for the different steps of the analysis.

fested in the global quantity of verbal units extracted by the experts. For example, the number of entries produced by each subject varied from 136 (subject 8) to 470 (subject 14).

In order to compare the relative use of the different descriptive categories in our analysis procedure, the rate of use of different categories was normalized with respect to the total number of units produced by each subject within each subject group. A rate coefficient (Kp_s) was calculated with respect to the maximum number of verbalizations (N_{max}) produced by a single subject. The coefficient for a given subject s was calculated as the ratio between N_{max} and the total number of verbal units, N_s, that he or she produced:

$$Kp_s = N_{max} / N_s.$$

All data on the frequency of use of verbal units will be presented in terms of this normalized coefficient, both for computations based on the integral data for each subject and for verbal units in a single category of the database for a given subject.

We next analysed the variability of use of verbal units corresponding to the different labelling steps in the database. The coefficient of variation (K_v) was used as an index of variability. K_v is the standard deviation divided by the mean computed on the frequency of use of verbal units in different categories by all subjects. It is a unit-independent measure of relative dispersion in the data. We determined which steps of the analysis scheme were characterized by a maximal group stability (low K_v) and which steps or subject differences had the greatest variation (high K_v)[2]. Figure 4 shows a histogram of the coefficient of variation for data corresponding to the different analysis steps for the 14 subjects with 2 repetitions.

Due to the fact that no variation occurred between subjects in Step 8 for a given category of Step 7 nor in Step 10 for a given category in Step 9, Steps 8 and 10 do not appear in Fig. 4.

[2] We will discuss the general results in this section: The analysis of a specific step does not take into account the particularities of verbal unit classification determined at earlier steps, and thus indicates only the general variability of verbal unit classification at that step.

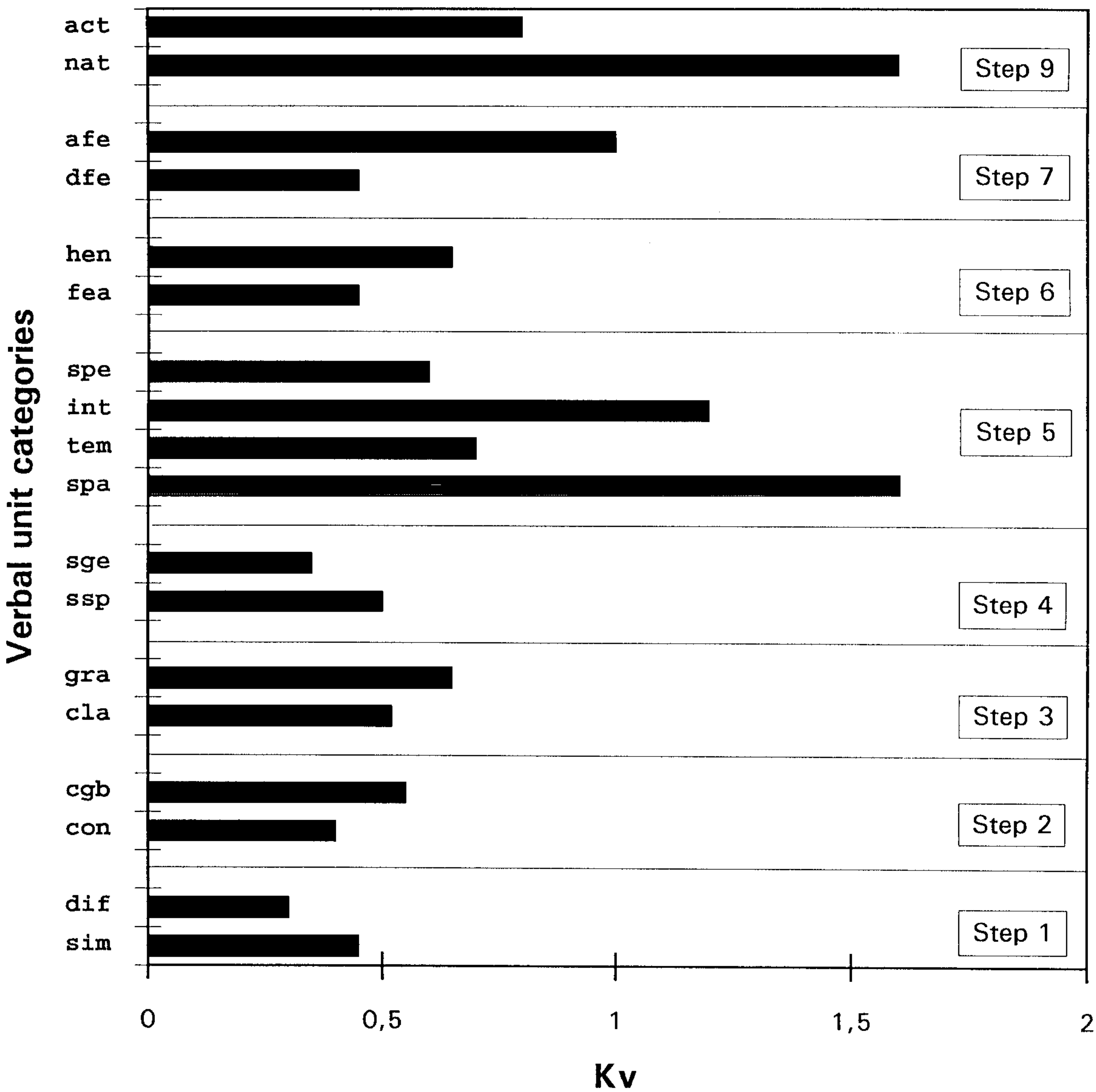

FIG. 4. Coefficient of variation (Kv) for normalized frequency of use of verbal unit categories in Steps 1–7 and Step 9.

Note that the greatest variation among subjects occurred from the fifth analysis step onward. The highest variation concerned the specific stimulus attributes chosen by subjects to compare the sounds. Intersubject differences in the variability of use of spatially related verbalizations \spa\ were especially high, as were those for \int\ and \nat\. At the seventh analysis step, the largest differences between individuals concerned the use of emotional-evaluative verbal units \afe\.

The results presented here used the means computed over the whole set of stimuli. This analysis did not take into account specific effects of a given stimulus on the nature of the verbalizations. By studying only the means within a given category at a specific level, we did not take into account the possible influence of the different sorting paths of the individual verbal units. As such these results are best viewed as indications of general tendencies. Nonetheless, in the following section, we examine a subset of the data in greater detail in order to analyse the specificities of certain stimulus comparisons.

In order to demonstrate the usefulness of this technique for studying timbre perception, we next present a few analysis examples using the database: (1) relation of verbal behaviour to differences in dissimilarity judgements between and within subjects, and (2) spontaneous identification of sound sources.

VERBALIZATION ANALYSIS OF DISCREPANT DISSIMILARITY RATINGS

The effectiveness of a new procedure can be judged by the analysis of experimental results that have not been easily interpreted with other procedures. In particular, we were interested in the extent to which large differences in dissimilarity ratings of pairs of timbres would be reflected in the way similarities and differences between the timbres were verbalized by subjects. These differences were examined both across repetitions by individual subjects as well as across subjects. More specifically, we analysed differences in verbalizations between groups that gave either relatively SMALL (more similar) or relatively LARGE (more dissimilar) ratings. Such differences in rating may indicate either an artefact or weakness in the experimental procedure or the existence of divergent criteria and cues used to judge the timbral differences that change from subject to subject or between trials for a given subject.

Our approach was based on the latter hypothesis (diversity and evolution of rating criteria) and sought to demonstrate the ability of verbalization analysis to reveal the nature of the rating disparities by way of quantitative and qualitative differences in the categories of verbal descriptors.

Details of the Analysis

The verbal data were sorted on the basis of subjects' numerical dissimilarity ratings. Ratings were classed as LARGE if they exceeded 5 (6–8) and as SMALL if they were less than 4 (1–3) on the 8-point scale. Nine timbre pairs for which at least 25% of the subjects gave small ratings and at least 25% gave large ratings were selected for analysis. We analysed a total of 94 protocols that corresponded to large ratings and 60 protocols that corresponded to small ratings. The list of timbre pairs and of subjects corresponding to each group are presented in Table 3.

For some timbre pairs, certain subjects gave a large rating in one experimental session and a small rating in the other. For example, the following descriptions [translated from the Russian] of differences between two timbres (STG–TPR) were given by the same subject (#4) in the two experimental sessions.

Session 1: Rating = 2. "The two sounds are very similar . . . because they are both very artificial."
Session 2: Rating = 7. "These are . . . different instruments. Further . . . they differ in pitch."

The analysis consisted of a comparison of the verbal characteristics accompanying LARGE and SMALL ratings. We compared the frequency of use of different types of verbal units as classified according to the analysis scheme mentioned earlier. To achieve this aim, we first computed the mean frequency of use of each type of verbal unit for each group and for each pair of timbres. The maximum mean number of ratings that could be

TABLE 3
Subjects Giving Small (1–3) and Large (6–8) Dissimilarity Ratings for Each Instrument Pair[a]

Timbre Pairs	*Small Dissimilarity Ratings*	*Large Dissimilarity Ratings*
CNT—TBN	1,2[b],3,4[b],5,6,7,8,9,10,11,13,14	3,9,12,13
PBO—SNO	1,2,3,4[b],5,10,12[b],14	3,5,6,7,8,9,11,12,13,14
SPO—STG	1,2,5,6,7,8,11,14	1,2,3,4,9,10
HRN—PBO	1,2,3,4,6,8,10,11,14	3,6,7,9,11,13
STG—TPR	1,2,4,5,6,7,9,10,11	1,2,3,4,8,9,13
TBN—VBN	1,5,6,7,8,10[b],11,12	1,2,3,4,6,7,9,12,13,14
SPO—PBO	5,6[b],7,8,12,14	3,4,7,9,11,13
PBO—PNO	1,2,4,5,6,7,8,10,11,12,14	2,4,7,9,13,14
VBS—HCD	1,2[b],3,4[b],5,6,7,8[b],9,10[b],11,14	2,7,8,12,13

[a] See Table 1 for key to instrument codes.
[b] Cases where a given subject made ratings of a given category (large or small) in both experimental sessions.

obtained for a given pair of timbres was four (two presentation orders for each of two sessions).

All data on the frequency of use of verbal units are presented in terms of normalized coefficient (Kp_s). Means for the small and large rating groups were computed on the selected set of data across subjects and the nine pairs of timbres. The number of verbal units extracted from the protocols that corresponded to the established rating constraints was large enough to perform statistical analyses only for the first six steps of the verbal unit analysis scheme.

Results and Discussion

The analysis of the use of different types of verbal units required a sorting of the database according to the scheme described earlier (Fig. 1).

Step 1. Units Referring to Similarity or Difference. The distribution of verbal units used to express similarities \sim\ and differences \dif\ between the timbres in the groups of SMALL and LARGE dissimilarity ratings are presented in Fig. 5. This figure shows that although verbal units corresponding to difference identifications were more frequent overall in both groups [for rating < 4: $t(16) = 4.0$, $P < 0.001$; for rating > 5: $t(16) = 13.3$, $P < 0.001$], similarity identifications were used more frequently when small dissimilarity ratings were given [$t(16) = 3.6$, $P < 0.005$], and a greater frequency of use of difference units corresponded to large dissimilarity ratings [$t(16) = 4.24$, $P < 0.001$]. However, it should be noted that this fact reflects the difference in the amount of all types of verbal units used to describe the similarity between the timbres. In order to understand which concrete type of these units is discriminative, the frequencies of their utilization in the two given groups (small and large dissimilarity ratings) are compared. Our analysis was run separately for verbal units used to describe similarities \sim\ and differences \dif\ between the timbres.

Verbal Units Describing Similarities

Step 2. Level of Generality of Similarity. There were no significant differences between the LARGE and SMALL groups in the frequency of use of concrete \sim\con\ or general \sim\gen\ forms of verbal units [\con\: $t(16) = 1.56$, ns; \gen\: $t(16) = 0.28$, ns]. There were also no differences for either group between the frequencies of use of verbal units classed as \sim\con\ or \sim\gen\ [SMALL: $t(16) = 1.31$, ns; LARGE: $t(16) = 1.24$, ns]. As such, our analysis of identification of similarity at subsequent levels will be made across Step 2 (grouping \con\ and \gen\ descriptions of similarities).

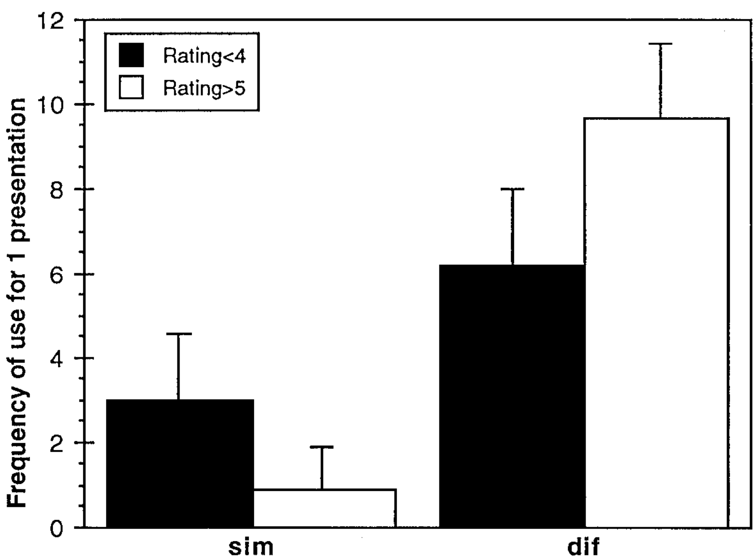

FIG. 5. Average normalized frequency of use of verbal units expressing similarities \sim\ or differences \dif\ between stimuli. The data are classed according to groups of subjects giving SMALL (1–3) and LARGE (6–8) dissimilarity ratings. Vertical bars represent 1SD.

Step 4. Specificity with Respect to the Sound Stimulus. The group making SMALL dissimilarity ratings produced more verbal units to describe similarities for both single attributes of sounds \sim\ . . . \ssp\ and the holistic nature of sounds \sim\ . . . \sge\ than did the LARGE group [\ssp\: $t(16) = 2.99$, $P < 0.01$; \sge\: $t(16) = 2.33$, $P < 0.05$].

Step 5. Specificity with Respect to Single Stimulus Parameters. Since the number of verbal units used to describe similarities of timbres in terms of spatial \spa\, temporal \tem\, and intensity \int\ parameters were negligible (less than 2%), we will discuss data concerning only the spectral \spe\ class of descriptors. This type of verbal unit was used more frequently by the SMALL group than by the LARGE group [$t(16) = 2.52$, $P < 0.05$]. Thus, the similarities between the timbres in the pair were verbalized predominantly on the basis of spectral parameters.

Step 6. Features and Holistic Meaningful Entities.

1. *Similarities in single stimulus parameters* (sorting \sim\ . . . \ssp\ . . . \fea\ or \sim\ . . . \ssp\ . . . \hen\). For both groups, the average frequency of use of the holistic meaningful entities for describing the similarities of single sound attributes \sim\ . . . \ssp\ . . . \hen\ was less than 5% of the total for this step. As such, our analysis will be confined to descriptions of features. Verbal units describing features of attributes \sim\ . . . \ssp\ . . . \fea\ were used more frequently in the SMALL group than in the LARGE group [$t(16) = 3.15$, $P < 0.01$]. Thus, we can conclude that the two groups are distinguished only in their use of features of single attributes for expressing the similarities between the timbres.

The following list of verbal units was used to describe similarity in terms of features (95 verbal units analysed) [our translation]: "similar in tonality" (43%); the "same note" (31%); "similarity in pitch" or "high-pitched sound" (20%); "static" or "long beginning" or "long end" (< 4%); "voluminous" (< 2%).

2. *Similarities in sound Gestalts* (sorting: \sim\ . . . \sge\ . . . \fea\ or \sim\ . . . \sge\ . . . \hen\). There were no significant differences between the SMALL and LARGE groups in their use of holistic meaningful entities (\sim\ . . . \sge\ . . . \hen\) for describing similarities in Gestalts [$t(16) = 1.45$, ns]. Verbal units describing features of Gestalts \sim\ . . . \sge\ . . . \fea\ were used more frequently by the SMALL group than by the LARGE group [$t(16) = 4.29$, $P < 0.001$]. This result suggests the existence of a correspondence between the use of verbal features of Gestalts to describe similarities and subjective distances separating the timbres. There was no such correspondence in the use of holistic meaningful entities for either attributes or Gestalts.

The following list of verbal units was used to describe similarity in terms of features of Gestalts (56 verbal units analysed) [our translation]: "well-coloured sounds" (25%); "saturated sounds" (18%); "ringing sounds" (12%); "jingling sounds" (9%); "muddy sounds" (8%); "clean sounds" (7%); "harsh sounds" (7%); "noisy, nasal sounds" (7%); "rich sounds" (7%).

Summary

The main findings of this section are summarized in Fig. 6, which shows the points of significant difference between the SMALL and LARGE dissimilarity rating groups in their descriptions of similarities. The SMALL rating group produced more verbal units describing similarities than did the LARGE group at four steps of the analysis. These verbal units concern primarily spectral and Gestalt features.

Verbal Units Used to Describe Differences between the Timbres

Step 2. Level of Generality of Difference. This analysis step revealed that units corresponding to LARGE ratings were more numerous than those corresponding to SMALL ratings [$t(16) = 2.7$, $P < 0.05$]. On the contrary, the difference between the groups in the frequency of use of general \dif\gen\ verbal units was not significant [$t(16) = 1.7$, ns].

The analysis shows that the concrete verbal units \dif\con\ were used more often than verbal units of general form \dif\gen\ in both groups [SMALL: $t(16) = 17.9$, $P < 0.0001$; LARGE: $t(16) = 12.8$, $P < 0.0001$]. Since the relative number of general verbal units \dif\gen\ was negligible (< 8%), in subsequent steps we will discuss only the data concerning the concrete \dif\con\ class of descriptors.

Step 3. Gradual and Classificational Units. The relative use of gradual verbal units \dif\con\gra\ was similar for both groups [$t(16) = 1.27$, ns]. Therefore, the specifics of the cognitive

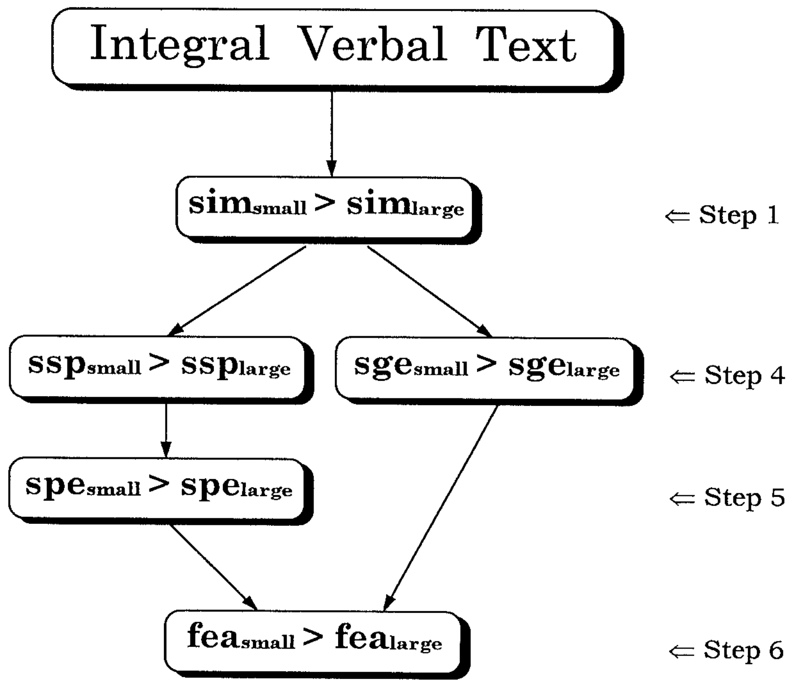

FIG. 6. Steps in the analysis scheme at which significant differences were found between groups of subjects giving small and large dissimilarity ratings in descriptions of similarities \sim\. The direction of difference in average frequency of use of verbal units of a given category is indicated.

operation of graduation were not reflected in corresponding evaluations. On the contrary, classificatory descriptors \dif\con\cla\ were used more often with LARGE ratings than with SMALL ratings [$t(16) = 4.23$, $P < 0.001$]. As such, we can conclude that the sounds judged as dissimilar were reported more often as belonging to different semantic classes.

Step 4. Specificity with Respect to the Sound Stimulus.

1. *Gradual differences* (sorting: \dif\con\gra\ssp\ or \dif\con\gra\sge\). There were no significant differences between the groups for either the gradual single attributes \dif\con\]gra\ssp\ [$t(16) = 1.67$, ns], or the gradual descriptions of holistic nature of sounds \dif\con\gra\sge\ [$t(16) = 0.08$, ns].

2. *Classificational differences* (sorting: \dif\con\cla\ssp\ or \dif\con\cla\sge\). The group giving LARGE dissimilarity ratings produced more verbal units describing differences for both classificational single attributes of sounds \dif\con\cla\ssp\ and the holistic nature of sounds \dif\con\cla\sge\ than did the SMALL group [\ssp\: $t(16) = 3.9$, $P < 0.005$; \sge\: $t(16) = 2.61$, $P < 0.05$]. We can thus use this analysis step to help define the perceptions corresponding to the subjective distances between the timbres.

Step 5. Specificity with Respect to Single Stimulus Parameters. Since the number of verbal units used to describe similarities of timbres in terms of intensity \int\ parameters were negligible (< 2%), we will discuss data concerning only spatial \spa\, temporal \tem\, and spectral \spe\ classes of descriptors. Since the portion of these verbal units was very limited, we performed the sorting across Step 3 (common for both \cla\ and \gra\ descriptions of dissimilarities). There were no significant differences between the groups in the use of either spatial or spectral characteristics [\spa\: $t(16) = 1.0$, ns; \spe\: $t(16) = 1.03$, ns]. On the contrary, the temporal units were used more frequently in the LARGE group than in the SMALL group [$t(16) = 3.96$, $P < 0.005$]. Thus, the perceived differences between the timbres in the pair were described predominantly on the basis of temporal parameters.

Step 6. Features and Holistic Meaningful Entities.

1. *Gradual differences in sound attributes* (sorting: \dif\con\gra\ssp\. . .\fea\ or \dif\con\gra\ssp\. . .\hen). For both SMALL and LARGE groups, no subjects used holistic meaningful entities for describing the differences of gradual, single sound attributes \dif\con\gra\ssp\. . .\hen\. As such, our analysis was confined to descriptions of features. The difference between the groups for the features in gradual single attributes \dif\con\gra\ssp\. . .\fea\ was not significant [$t(16) = 1.69$, ns].

2. *Gradual differences in sound Gestalts* (sorting: \dif\con\gra\sge\fea\ or \dif\con\gra\sge\hen\). As in the preceding example, the average frequency of use of the holistic meaningful entities for describing the differences of gradual sound Gestalts \dif\con\gra\sge\hen\ was negligible for both groups (less than 4% of the total for this step). As such, our analysis was confined to descriptions of features. Further, there were no significant differences between the groups for the features of gradual Gestalts \dif\con\gra\sge\fea\ [$t(16) = 0.46$, ns].

3. *Classificational differences in sound attributes* (sorting: \dif\con\cla\ssp\fea\ or \dif\con\cla\ssp\hen\). There were no significant differences between the groups for the holistic meaningful entities in classificational sound attributes \dif\con\cla\ssp\hen\ [$t(16) = 0.93$, ns]. Units in the form of verbal features in attributes \dif\con\cla\ssp\fea\ were used more frequently in the LARGE group than in the SMALL group [$t(16) = 4.42$, $P < 0.0005$]. Thus, we can conclude that two groups of data can be distinguished only in the use of features of single attributes for expressing the differences between the timbres \dif\con\cla\ssp\fea\.

Here, we present the list of verbal units used to describe differences in terms of features of single attributes (206 verbal units analysed) [our translation]:

a. Spatial characteristics (\spa\fea\): "localizable at one point in space" or "localizable inside the head" (6%); "voluminous" (6%); "near sound" or "far sound" (4%); "broad sound" (3%); "large space" (3%); "flat sound" (2%); "limited in volume" (2%);

b. Temporal characteristics (\tem\fea\): "short sound" or "fast sound" or "limited in duration" (19%); "very perceptible attack" (11%); "dynamic sound" (9%); "long sound" or "slow sound" (8%); "stable sound" (5%); "continuous sound" (4%); "vibrating sound" (4%); "double sound" (2%);

c. Intensity characteristics (\int\fea\): "change from soft to loud" (2%);

d. Spectral characteristics (\spe\fea\): "lot of high frequencies" (5%); "the second octave" (3%); "very high pitch" or "very low pitch" (2%).

4. *Classificational differences in sound Gestalts* (sorting: \dif\con\cla\sge\fea\ or \dif\con\cla\sge\hen\). There was no difference between the groups in the use of holistic meaningful entities \dif\con\cla\sge\hen\ for describing the differences in Gestalts [$t(16) = 0.34$, ns]. Verbal units describing features of Gestalts \sim\ . . . \sge\ . . . fea\ were used more frequently in the LARGE group than in the SMALL group [$t(16) = 2.82$, $P < 0.05$]. There exists, therefore, a correspondence between the use of features of Gestalts to describe the differences and subjective distances separating the timbres. There was no such correspondence for the use of holistic meaningful entities of either attributes or Gestalts.

The following list of verbal units was used to describe differences in terms of features of Gestalts (43 verbal units analysed) [our translation]: "harsh sound" (20%); "rich sound" (18%); "piercing sound" (16%); "jingling sound" (14%); "artificial sound" (11%); "clean sound" (9%); "noisy-nasal sound" (8%); "natural sound" (4%).

Summary

The main results of this section are summarized in Fig. 7, which shows the points of significant difference between the SMALL and LARGE rating groups in their descriptions of differences \dif\. The verbal units expressing differences were produced more frequently by the LARGE rating group than by the SMALL rating group. These differences primarily concerned temporal and Gestalt features of a concrete and classificational nature.

Our results concerning the correspondence of the numerical ratings of dissimilarities between sounds to the way their similarities and differences are described are important in the context of those studies where attempts have been made to compare the results obtained with methods using verbalizations and other methods for analysing subjective representations. It should be

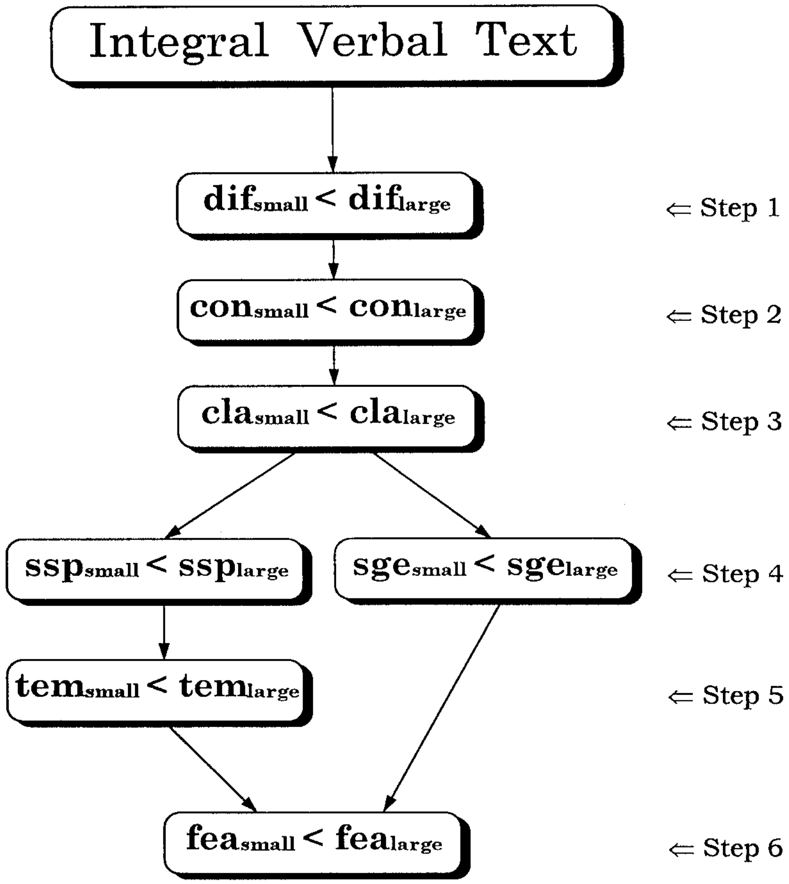

FIG. 7. Steps in the analysis scheme at which significant differences were found between groups of subjects giving small and large dissimilarity ratings in descriptions of differences \dif\. The direction of difference in average frequency of use of verbal units of a given category is indicated.

mentioned that although a number of studies using verbal descriptions have been performed, few of them were directed at making this kind of direct comparison. Although there seems to be no previous work on this correspondence for the auditory modality, a number of studies have examined the visual modality. The studies of Artemieva (1980), for example, employed a special procedure of reproduction (drawing) of complex visual stimuli on the basis of verbal features produced beforehand in situations of referential communication. These were used to support the claim that the verbalized features were the real "psychological coordinates" of the stimuli. The adequacy of verbal descriptions of objects has been demonstrated by means of simultaneous registration and collation of these descriptions with trajectories of visual tracing of these objects (Urvantsev, 1979). Tversky (1977) found a correlation between similarity ratings between objects and the number of identical verbal features used to describe each of them.

In summary, subjects giving small dissimilarity ratings for specific timbre pairs produce a greater proportion of verbal units expressing similarity than do subjects giving large ratings. Complementary to this result is the fact that subjects giving large dissimilarity ratings produce a greater proportion of verbal units expressing differences. This result is not surprising in itself and merely serves to show that the verbalizations are coherent with the dissimilarity ratings. Nonetheless, the verbalization analysis allows us to go one step further in revealing the nature of the sound properties that are focused on. The types of verbal units that are used to express similarity more often by the SMALL group than by the LARGE group are primarily related to spectral and Gestalt features of the sounds. On the contrary, the units used more often by the LARGE group than by the SMALL group to express differences are more related to temporal and Gestalt features of a classificational nature. Therefore this analysis technique provides us with more information concerning the aspects of sounds that are used in comparing them than does the dissimilarity rating.

IDENTIFICATION OF TIMBRES AS MUSICAL INSTRUMENTS

This analysis concerns the recognition of the timbres as concrete sound objects as revealed in the last analysis steps (7–10). A further classification of Step 10 allowed us to classify the timbres according to musical instrument categories (a kind of "identification portrait" of the sounds described).

For this analysis, we selected and sorted the data according to one branch of the analysis tree (. . . \hen\ . . . \mus) and calculated the number of concrete references indicating specifically musical instruments (e.g. "guitar," "brass instrument," "vibraphone," "piano") or qualities that unequivocally evoked these instruments ("brass-like," "percussive"). A content analysis of the selected verbal units suggested that it was possible to group them into the following general classes by musical instrument families: "Winds," "Strings," and "Percussion." Keyboard instruments (e.g. piano, harpsichord) were classified as Strings. A fourth class ("Others") corresponded to all other references to musical instruments. The verbal units assigned to this latter class were often of a general nature (e.g. "musical instrument" without specifying the type, or various references to electronic instruments such as "electric organ"). They also included negative identifications (e.g. "not brass," "less percussive," "backwards percussive," "approaches less the sound of a string"). We calculated the frequency of use of referents in each class for each type of timbre presented to the subjects. A total of 720 verbal units corresponding to the appropriate branch of the analysis tree were selected for these comparisons across subjects. The global percentage of units that were further classified into instrument family categories were as follows: Winds (33%), Strings (28%), Percussion (9%), and Others (30%).

Next we examine the use of these classes of referents during comparisons of TPR (trumpar—trumpet/guitar hybrid) with each of the timbres TPT (trumpet), VBS (vibraphone), PBO (piano-bow—bowed piano), GTR (guitar), and STG (bowed string). Figure 8 presents the "identification portraits" of these synthetic timbres. The values for TPR are computed across all five pairs.

The classification of TPT by the subjects in the context of TPR was very vague: Winds, Strings, and mostly Others (this latter identification often being simply "musical instrument" or "electronic

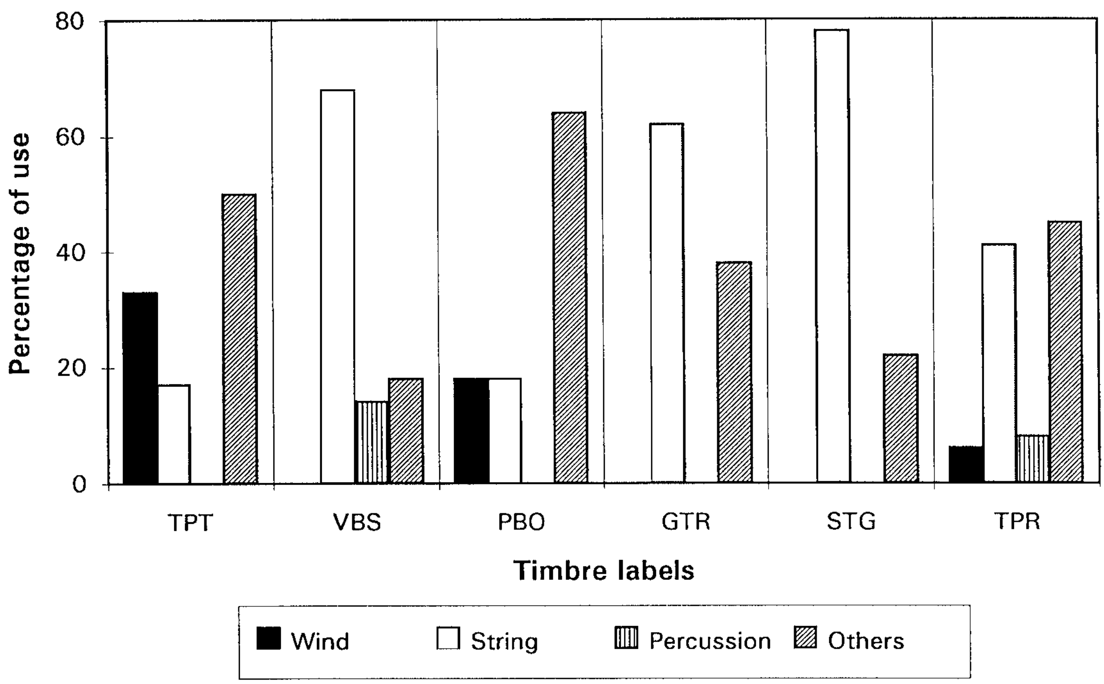

FIG. 8. Rate of identification of timbres TPT, VBS, PBO, GTR, STG, and TPR as specifically belonging to the Wind, String, or Percussion families of musical instruments. The category "Others" indicates identifications belonging to different categories or identifications too general in nature to be categorized at this level (see Table 1 for timbre abbreviations).

wind instrument''). The percussion instrument VBS was identified as String or Percussion or often as ''instrument'' or ''organ.'' PBO, a simulation of an unusual manner of playing a piano (by bowing) is most often identified as ''instrument,'' ''electronic wind instrument,'' or ''electric organ,'' and rarely as Wind or String. GTR was more unambiguously identified as String with a few ''instrument,'' ''organ,'' or ''approaches less a string sound'' labels. On the other hand, STG was identified most often as a String with a few additional vague ''instrument'' labels. Finally, TPR, the most unusual hybrid of the set, was labelled as ''instrument,'' ''synthesizer,'' ''electric organ,'' followed by String, and a rare Wind or Percussion label. These latter labels, particularly ''synthesizer,'' testify to the subjects' inability to match this hybrid of a plucked string and a blown air column to a plausible physical object.

This analysis led us to the conclusion that the subjects used labelling categories that were not very well defined compared to what one might expect from professional musicians, for example. The descriptors tended to be at broader levels of a hierarchical categorization scheme in which musical instrument is a category containing wind instrument, which contains brass instrument, which in turn contains trumpet and trombone, for example. It would be very interesting to extend this research to see if these levels of category structure could be linked to Rosch's notion of superordinate, basic, and subordinate levels (Rosch et al., 1976) as a function of musical training. The hybrid sounds that did not strongly evoke acoustic instruments may have seemed merely ''musical'' without a much finer distinction for these nonmusician subjects. We would like, nevertheless, to emphasize the fact that some of these synthetic imitations, and the two hybrid timbres, reveal a certain ambiguity in terms of the identifications they evoked.

CONCLUSIONS

We have elaborated a scheme of verbal protocol analysis consisting of the progressive differentiation of freely produced verbal units into a hierarchy of classes adapted to the description of complex sound events. We have shown the analysis scheme to be precise and relatively unambiguous as concerns reliability among experts in the selection of verbal units from the integral protocol and their classification according to the scheme.

It should be noted that the most elaborated psychological classifications of verbal units have been created for visual stimuli. One of the most interesting examples of such a classification was proposed in the referential communication study of Fussell and Krauss (1989), where expressions created by subjects for referring to abstract figures were first divided into three general types: ''literal'' messages describing what the figure is; ''figurative'' ones describing what figure is like; and ''symbolic'' ones limited to letters and numbers. Figurative messages were further subdivided into nine major categories: people, body parts, animals, plants, small objects, large objects, features of the environment, buildings/parts of buildings, and ''others.'' As for expressions referring to auditory stimuli, there is, to our knowledge, no existing systematic classification system aside from the one proposed here.

The analysis scheme proposed in our studies demonstrates that verbalizations can be adequate for studying perceptual representations and for revealing specific aspects of the perception of similiarities and differences among musical timbres. To illustrate this adequacy, we examined two points on the basis of the database constructed from verbalizations comparing the timbres of synthetic musical instrument sounds: (1) the inter- and intra-subject discrepancies in numerical dissimilarity ratings and their correspondence with descriptions of similarity and difference, and (2) the specificity of identification of different timbres as musical instruments. The main findings are summarized next.

Correspondence between Dissimilarity Ratings and Verbalizations

The peculiarities of perception of dissimilarities between musical timbres were revealed in each of three levels of analysis of verbal units: their *logical sense*, their *stimulus-relatedness*, and their *semantic aspects*. We have shown that, at the level of the *logical sense* of verbal units, numerical dissimilarity ratings of sounds presented in pairs were found to correspond to the relative number of verbal units used to describe their similarities and differences when the ratings were classed as small (more similar) and large (more dissimilar). Verbalizations of similar characteristics of sounds were used more often by the group that gave small dissimilarity ratings than by the group giving large ratings to the same timbre pairs, and the reverse was true for verbalizations of difference. So numerical dissimilarity ratings are globally coherent with the proportion of verbal units produced to express both similarity and difference across groups.

It is important to note that this comparison concerns the same timbre pairs as judged by different groups of subjects. In spite of the fact that all subjects received the same stimuli, their judgements varied a great deal for certain pairs. This result suggests that listeners' perceptions can vary a great deal, as judged by the kinds of features they choose to report in comparing similarities and differences among the timbres. At the level of *stimulus-relatedness*, the correspondence between verbal descriptors and numerical ratings concerned both specific sound attributes (for example: "attack") and the characteristics of sound Gestalts (for example: "a rich sound"). Further, the similarities between the timbres in the pair were verbalized predominantly on the basis of spectral parameters. However, at the level concerning *semantic aspects* of verbal units, a significant correspondence between the verbal and psychophysical data was found only for verbal units classed as sound features. The frequency of use of verbal units related to holistic meaningful entities does not differ between the two subject groups. This finding suggests to us that in the cognitive operation of comparing timbres, the characteristics that are important for deciding on the similarity between timbres are found in categories concerning specific features of the perceived sounds.

The verbalization of dissimilar characteristics of sounds were used more often by the group of subjects giving larger dissimilarity ratings than by the group giving smaller ratings. At the level of the *logical sense* of the verbal units, a correspondence between the verbal descriptors and numerical ratings of differences concerned only the verbal units that had a classificatory nature, reporting compared sounds in different semantic categories. The use of a strategy of graduation in the verbalization of dissimilarities did not covary with the judged dissimilarity between the sounds. At the level of *stimulus-relatedness*, the correspondence of verbal characteristics to numerical ratings concerned both sound attributes \dif\con\cla\ssp\ and the characteristics of sound Gestalts \dif\con\cla\sge\. As for single stimulus parameters, the dissimilarities between the timbres in the pairs were verbalized predominantly on the basis of temporal parameters \dif\con\cla\ssp\tem\. As with the verbalization of similarities, at the level of analysis concerning the *semantic aspects* of verbal units, a significant correspondence between the verbal data expressing dissimilarities and the psychophysical data was found only for descriptions of sound features \fea\. The holistic meaningful entities \hen\ had no relation to the dissimilarity ratings. We may thus conclude that our analysis reveals a correspondence between the numerical rating of dissimilarities between sound and the way that subjects describe their similarities and differences. It further suggests that the kinds of features emphasized by subjects are different for similarity and difference. It is at this level that we have shown the technique to be a potentially powerful tool in refining dissimilarity rating analyses.

Spontaneous Identification of Musical Instruments

The nonmusician subjects often used categories that were not very differentiated in attempting to identify the object that produced the sound ("wind instrument," "musical instrument"). Further, they were not always in agreement when more differentiated categories were used. Since these synthetic sounds were more or less successful in their imitations of traditional instruments, some were identified ambiguously across our subject pool. This ambiguity was also evidenced with the sounds meant to be hybrids of traditional instruments, as one would expect.

In closing, we would like to suggest that the proposed procedure of verbal protocol analysis may be used not only for the analysis of descriptions of auditory stimuli, but could also be adapted for other modalities. In this case the steps corresponding to referential analysis (Steps 7–10) would need to be specified. In the auditory domain, one research perspective that we hope to follow with this technique involves the study of perceptual processing of complex acoustic events and sources for which it may be difficult to define a clear link between stimulus properties and perceptual reactions. In particular, we intend to continue this line of research addressing more directly the differences in description that reveal sociocultural differences in complex auditory perceptual phenomena and that reveal the semantic categories systematically associated with the underlying perceptual dimensions of the timbre of complex sounds.

Manuscript first received June 1994
Revised manuscript accepted January 1996

REFERENCES

Artemieva, E. Yu. (1980). *Psikhologiya subektivnoj semantiki* [*Psychology of subjective semantics*]. Moscow: MGU.

Bartlett, F.C. (1932). *Remembering*. Cambridge: Cambridge University Press.

Bower, G.H., & Holyoak, K. (1973). Encoding and recognition memory for naturalistic sounds. *Journal of Experimental Psychology*, *101*(2), 320–366.

Brommel, R. (1983). "Understanding texts" as heuristics for the analysis of thinking-aloud protocols. *Communication and Cognition*, *16*(3), 215–231.

Caverni, J.-P. (1988). La verbalisation comme source d'observables pour l'étude du fonctionnement cognitif [Verbalization as a source of observables for the study of cognitive functions]. In J. Caverni, C. Bastien, P. Mendelsohn, & G. Tiberghien (Eds.), *Psychologie cognitive: Modèles et méthodes* (pp. 253–273). Grenoble: Presses Universitaires de Grenoble.

Cuni, X. (1979). Different levels of analyzing process control tasks. *Ergonomics*, *22*, 415–425.

Ericsson, K.S., & Simon, H.A. (1984). *Protocol analysis. Verbal reports as data*. Cambridge, MA: MIT Press.

Fussell, S.R., & Krauss, R.M. (1989). The effects of intended audience on message production and comprehension: Reference in a common ground framework. *Journal of Experimental Social Psychology*, *25*, 203–219.

Handel, S. (1989). *Listening: An introduction to the perception of auditory events*. Cambridge, MA: MIT Press.

Hoc, J.M. (1984). La verbalisation provoquée pour l'étude du fonctionnement cognitif [Evoked verbalization for the study of cognitive function]. *Psychologie Française*, *29*, 231–234.

Kendall, R.A., & Carterette, E.C. (1992a). Verbal attributes of simultaneous wind instrument timbres. I: von Bismarck's adjectives. *Music Perception*, *10*, 445–468.

Kendall, R.A., & Carterette, E.C. (1992b). Verbal attributes of simultaneous wind instrument timbres II: Adjectives induced from Piston's Orchestration. *Music Perception*, *10*, 469–502.

Kouznetsov, L.A. (1981). *Osnovy teorii konstruirovaniya, proizvodstva i remonta elektromuzykalnykh insrumentov* [*Theoretical bases for the conception, production and restoration of musical instruments*]. Moscow: Legkaya i Pischevaya Promyshlennost.

Lange, N. (1893). *Psikhologicheskie issledovaniya. Zakon vospriyatiya. Teoriya vnimaniya* [*Psychological investigations. The law of perception. Theory of attention*]. Odessa: Psikhologicheskoe Obschestvo.

Leplat, J., & Hoc, J.M. (1981). Subsequent verbalization in the study of cognitive processes. *Ergonomics*, *24*, 743–755.

Lomov, B.F., Belyaeva, A.V., & Nosulenko, V.N. (1986). *Verbalnoe kodirovanie v poznavatelnykh protsessakh* [*Verbal coding in cognitive processes*]. Moscow: Nauka.

McAdams, S., & Cunibile, J.-C. (1992). Perception of timbre analogies. *Philosophical Transactions of the Royal Society, London*, *B336*, 383–389.

Newell, A. (1977). On the analysis of human problem solving protocols. In P.M. Johnson-Laird & P.C. Wason (Eds.), *Thinking: Readings in cognitive science* (pp. 46–61). Cambridge: Cambridge University Press.

Nikitin, M.P. (1905). Problema zritelnogo vospriyatiya [The problem of visual perception]. *Byulleten po Psikhologii Kriminalnoj Antropologii i Gipnotizmu*, *2*(2), 112–122.

Nosulenko, V.N. (1988). *Psikhologiya slukhovogo vospriyatiya* [*Psychology of auditory perception*]. Moscow: Nauka.

Nosulenko, V.N. (1989a). The psychophysics of a complex signal: Problems and perspectives. *Soviet Psychology*, *1*, 62–78.

Nosulenko, V.N. (1989b). Prostranstvo-vremya v slukhovom vospriyatii [Space-time in auditory perception]. *Psikhologicheskij journal*, *2*, 22–32.

Nosulenko, V.N., & Samoylenko, E.S. (1992). Perception of sound environment changes determined by technological development of society. *International Journal of Psychology*, *27*, 20(A).

Nosulenko, V.N., Samoylenko, E.S., & McAdams, S. (1994). L'analyse de descriptions verbales dans l'étude des comparaisons de timbres musicaux [Analysis of verbal descriptions in the study of comparisons of musical timbre]. *Journal de Physique*, *4*(C5), 637–640.

Osgood, C.E., Suci, G.J., & Tannenbaum, P.H. (1957). *The measurement of meaning*. Urbana, IL: University of Illinois Press.

Radocy, R., & Boyle, J. (1979). *Psychological foundations of musical behavior*. Springfield, IL: Thomac.

Rosch, E., Mervis, C.B., Gray, W.D., Johnson, D.M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology*, *7*, 573–605.

Samoylenko, E.S. (1986). *Operatsiya sravneniya pri reshenii kognitivno-kommunikativnykh zadach* [*Comparison operations in cognitive-communicative tasks*]. Unpublished doctoral dissertation. Moscow: Institout Psikhologii AN SSSR.

Smith, E.R., & Miller, F.D. (1978). Limits on perception of cognitive processes: A reply to Nisbett and Wilson. *Psychological Review*, *85*, 355–362.

Sokolov, P. (1887). Fakty i teoriya "tsvetnogo sloukha" [Facts and theory of "colour vision"]. *Problemy Filosofii i Psikhologii*, *37*, 378–412.

Taylor, L.J., Gandy, L.J., & Dark, G. (1974). Linguistic description and auditory perception. *Perceptual and Motor Skills*, *38*, 703–707.

Trubetskoy, N.S. (1960). *Osnovy fonologii* [*Basis of phonology*]. Moscow: In. Iaz.

Tversky, A. (1977). Features of similarity. *Psychological Review*, *84*, 327–352.
Urvantsev, L.P. (1979). Experimentalnoe issledovanie protsessov vospriyatia i interpretatsii rentgenogramm [Experimental investigations of perceptive processes and interpretations of radiographs]. *Psykhologicheskye problemy ratsionalizatsii deyatelnosti, svyazannoj s opoznaniem obrazov* (pp. 60–97). Yaroslavl: Yar Gyu.
von Bismarck, G. (1974). Timbre of steady sounds: A factorial investigation of its verbal attributes. *Acustica*, *30*, 146–159.
Wessel, D.L., Bristow, D., & Settel, Z. (1987). Control of phrasing and articulation in synthesis. *Proceedings of the 1987 International Computer Music Conference* (pp. 108–116), San Francisco, CA: Computer Music Association.
Wright, P. (1971). Linguistic description of auditory signals. *Journal of Applied Psychology*, *55*, 244–250.

APPENDIX: DETAILED DESCRIPTION OF THE VERBAL PROTOCOL ANALYSIS PROCEDURE

The Appendix describes the detailed scheme for analysing verbal data that was elaborated for comparisons of complex sounds. The various steps of the analysis are presented in Fig. 1. Each verbal text, produced by a given subject in comparing a pair of sounds, is considered at three main levels of analysis of the verbal units: (1) their *logical sense* (Steps 1–3), (2) their *stimulus-relatedness* (Steps 4–5), and (3) their *semantic aspects* (Steps 6–10).

Level 1. Logical Sense of Verbal Units Used in a Comparison

Step 1. Identification of Verbal Units Referring to Similarity or Difference. In an integral verbal protocol produced by a subject, the parts containing descriptions of stimulus similarities \sim\[3] and differences \dif\ are marked accordingly. In order to distinguish the given parts, corresponding syntactic and lexical constructions existing in a given language (Russian in the present case) are taken into consideration. For expressing similarity, one uses various syntactic constructions like complex sentences with a copulative conjunction "and," or simple sentences with homogeneous parts connected by repeating and double conjunctions ("both . . . and," "this and that," etc.) along with lexical constructions representing the identity of objects ("same," "to the same extent," "common," "equal," "also," "corresponds," "do not differ," etc.). For expressing difference, one uses syntactic constructions like complex sentences with an adversative conjunction or lexical forms expressing difference ("differs," "difference in," etc.).

For example, consider the following integral verbal protocol (". . . " indicates a pause in the speech stream): "Similarity in that . . . both attack times are long . . . although the second one is shorter and there is at the end some sort of noise like a damping or something and it's also brighter in the spectral quality." The verbal units corresponding to the similarity \sim\ identification are represented in the following part of the protocol: "Similarity in that . . . both attack times are long . . . ." The verbal units corresponding to the identification of differences \dif\ are represented in the rest of this protocol.

Step 2. Analysis of the Level of Generality of Similarity and Difference. The verbalizations are categorized according to whether similarities and differences are expressed in a general or a concrete form. Within parts of a verbal text identified as referring either to similarity or to difference, verbal units representing terms corresponding to a general basis of comparison \gen\ and those expressing concrete (or detailed) comparisons \con\ are labelled. For instance, in the statements "the sounds have a very similar spectrum" or "the sounds have different spectrum qualities," there is a lack of verbal units expressing how the spectra are similar or how the spectrum qualities are different. The verbal units "spectrum" and "spectrum qualities" are labelled as general bases of comparison \gen\ of similarity and difference, respectively.

On the contrary, in the statements like "similarity in that . . . both attack times are long" or "the second sound is shorter and there is, at the end, some sort of noise, like a damping or something, and it's also brighter in the spectrum qual-

[3] Letters between back-slashes "\" correspond to codes related to different levels of the verbalization analysis schema. These can be concatenated to indicate a descent through several levels of analysis, e.g. \dif\con\cla\ (see Fig. 1). The notation \dif\ . . . \sge\ indicates all permissible paths from \dif\ to \sge\.

ity,'' there is a concrete specification \con\ of attack times as long or of their spectrum quality as bright, etc. Thus, the corresponding verbal units are labelled as concrete representations of similarity and difference.

Step 3. Gradual and Classificational Oppositions in the Representations of Difference. This analysis is performed only for concrete difference verbalizations \dif\con\. Verbal units are analysed according to the way sounds are opposed and are categorized as "classificational" \cla\ or "gradual" \gra\. The following categorization is based on notions from the theory of oppositions elaborated by Trubetskoy (1960) for the phonological level of language.

In *classificational* verbal units, the two sounds are identified as possessing different qualities, or a single sound of the pair is characterized. For example:

"The first is a harpsichord, the other is a violin."
"The first sound is very bright."

In *gradual* verbal units, the two stimuli are compared with respect to the same semantic class. These units refer to the degree of difference between the two objects. For example:

"The second sound is more rough [than the first one]."
"The first one resembles a plucked string more than the second one."

Level 2. Stimulus-Relatedness of Verbal Units Used in Comparison

At this level of analysis the verbal units are labelled according to whether they are used to describe global aspects of the sounds or their specific (single) properties (Steps 4–5).

Step 4. Specificity with Respect to the Sound Stimulus. The verbal units can be used to describe the holistic nature of the sound (sound Gestalts—\sge\) or be limited to a single attribute of the heard sound (single sound properties—\ssp\).

The reference of a verbal unit to the whole sound event means that it is impossible to identify a simple, concrete sound parameter being compared. For example:

"First we have the plucked string. The second one is a horn sound" (concrete difference in classificational Gestalts \dif\con\cla\sge\);
"The first one resembles more a plucked string" (concrete difference in gradual Gestalts \dif\con\gra\sge\);
"They are produced by different musical instrument" (general difference in Gestalts \dif\gen\gap\sge\);
"They both are harpsichords" (concrete similarity in Gestalts \sim\con\gap\sge\);
"They are produced by the same musical instrument" (general similarity in Gestalts \sim\gen\gap\sge\).

The reference of a verbal unit to a single property \ssp\ of a sound allows a further identification of the specific sound parameters being compared. These parameters are labelled in Step 5.

Step 5. Specificity with Respect to Single Stimulus Parameters. Single sound parameters are divided into four major classes: *spatial* \spa\, *temporal* \tem\, *intensive* \int\, and *spectral* \spe\. Verbal units used to describe spatial characteristics generally refer either to the space occupied by the sound ("large space," "broad sound," "narrow," "voluminous") or to its localization in space with respect to the listener ("near," "far," "localizable"). Units corresponding to temporal characteristics refer to temporal parts of a sound ("beginning,' "attack,' "steady part,' "body of the sound,' "end," "decay," "damping,' "resonance part'), to its duration ("long," "short") or to its temporal behaviour, texture, or form ("fast," "slow," "modulation," "roughness,' "evolution,' "stable"). Verbal units used to describe the intensity characteristics refer to intensity directly or to loudness ("intensity,' "energy," "loud,' "soft," "amplified,' "mezzo piano'). Verbal units used to characterize spectral aspects of a sound refer in general to frequency content ("sinusoidal sound,' "lot of high frequencies,' "inharmonic," "has noise components," "rich spectrum'), to pitch ("the second octave,' "a fifth above,' "medium register"), or to properties related to the spectral envelope ("spectral profile,' "spectral envelope,' "nasal," "head tone'). Reference to correlates of space, intensity, and spectrum that change over time are classified under \tem\.

Here are a few examples of temporal attributes:

"The second one has an evolution in the spectrum, the first one does not have it" (\dif\con\cla\ssp\tem\);
"In the second case we have a slower attack" (\dif\con\gra\ssp\tem\);
"The main difference here is in the attack" (\dif\gen\gap\ssp\tem\);
"Both of them have a quite soft attack" (\sim\con\gap\ssp\tem\);
"They are quite similar in the steady part" (\sim\gen\gap\ssp\tem\).

Level 3. Semantic Aspects of Verbal Units

This level aims to analyse semantic categories used to describe the two sounds being compared (Steps 6–10). A division into two principal categories is first made in Step 6: features \fea\ and holistic meaningful entities \hen\. Features are then further subdivided in Steps 7–9, with holistic entities subdivided in Steps 7–10.

Features

A description of separate features means that distinguishing properties of sounds are represented without assigning a holistic meaning to those sounds.

Step 7. Separate features are first divided into those having descriptive features \dfe\ ("loud," "intense," etc.) or attitudinal features \afe\ by means of which a person expresses an emotional or evaluative attitude to perceived sounds ("a very pleasant sound," etc.).

Descriptive Features

Step 8. The group of descriptive features \dfe\ is divided on the one hand, into those normally used by people only to describe sound phenomena, that is, features referring only to the auditory modality (unimodal) \umd\ ("loud," "high-pitched," "noisy"), and, on the other hand, into features that can be used to describe subjective representations for other sensory modalities (polymodal) \pmd\ ("colourful," etc.).

Attitudes

Step 8. The group of descriptions expressing emotional–evaluative attitudes \afe\ to perceived sounds is always considered to be polymodal and is divided first into those containing information about emotional \emv\ relations to sounds ("pleasant sound") and those where an aspect of artificiality or naturalness \ntl\ is mentioned ("it is very far from natural").

Step 9. The \emv\ units are further divided into those describing positive and negative aspects (e.g. "pleasant sound" and "ugly sound"—\pos\ and \neg\, respectively). The \ntl\ descriptions are divided into those indicating the naturalness \ntr\ of a sound and those mentioning its artificial nature \art\ ("a synthesized sound").

Holistic Meaningful Entities

Step 7. Holistic meaningful entities are divided first into the verbal units that are specifically related to the production of sound phenomena (related) \rel\ ("sound of a piano") and those that are not usually associated specifically with sound phenomena (unrelated) \unr\ ("sound resembling happiness").

Step 8. Holistic meaningful entities can contain information about real objects \rob\ or abstract concepts \abs\. For example, "piano" or "loudspeaker" would represent a real object designed to produce sound, whereas a "jazzy sound" would be an abstract sound concept. As for the group of holistic verbal representations not specifically related to sound phenomena, "hammer" would be an example of a real object that is not designed to produced sound (unless it was clearly a piano hammer or a similar component of musical instrument), whereas the expression "sound resembling happiness" would contain an abstract concept not related to sound.

Step 9. Verbal units concerning concrete objects not designed to produce sound are finally divided into those concerning natural \nat\ objects and phenomena ("sound produced by an ocean" or "sound produced by a woodpecker," for example) and those concerning objects used in various human activities \act\ ("sound produced by a hammer"). As for descriptions concerning sound-related objects, they are also divided at this step into those concerning natural \nat\ objects habitually involved with sound activity ("voice of a singing bird") and those concerning objects used in sound reproducing activity \act\ ("sound of musical instrument," "loudspeaker").

Step 10. Finally, the means of sound-reproducing activity are divided into musical instruments \mus\ ("brass instrument sound") and other sound reproducing devices \dev\ ("loudspeaker").