Music Perception
Winter 1990, Vol. 8, No. 2, 129–164

The Communication of Musical Expression

ROGER A. KENDALL & EDWARD C. CARTERETTE
University of California, Los Angeles

This study focuses on the performer-listener link of the chain of musical communication. Using different perceptual methods (categorization, matching, and rating), as well as acoustical analyses of timing and amplitude, we found that both musicians and nonmusicians could discern among the levels of expressive intent of violin, trumpet, clarinet, oboe, and piano performers. Time-contour profiles showed distinct signatures between instruments and across expressive levels, which affords a basis for perceptual discrimination. For example, for "appropriate" expressive performances, a gradual lengthening of successive durations leads to the cadence. Although synthesized versions based on performance timings led to less response accuracy than did the complete natural performance, evidence suggests that timing may be more salient as a perceptual cue than amplitude. We outline a metabolic communication theory of musical expression that is based on a system of sequences of states, and changes of state, which fill gaps of inexorable time. We assume that musical states have a flexible, topologically deformable nature. Our conception allows for hierarchies and structure in active music processing that static generative grammars do not. This theory is supported by the data, in which patterns of timings and amplitudes differed among and between instruments and levels of expression.

WE explore the basic question, "How does the performer convey his ideas to the listener?" Other studies have investigated various aspects of expressive performance (e.g. Seashore, 1938; Bengtsson & Gabrielsson, 1983; Clarke, 1988; Clynes, 1983; Gabrielsson, 1988; Sundberg, 1988; Sundberg, Frydén, & Askenfelt, 1983), primarily focusing on acoustic measurements of timing and amplitude. Indeed, musical expression has been commonly defined in terms of deviations from mechanical performance of canonical notations. However, a distinction must be made between random performance variability and that attributable to expressive intent. Over 50 years ago, Seashore (1938) remarked

Requests for reprints may be sent to Roger A. Kendall, Department of Ethnomusicology and Systematic Musicology, or Edward C. Carterette, Department of Psychology, University of California, Los Angeles, Los Angeles, CA 90024.

> As a fundamental proposition we may say that the artistic expression of feeling in music consists in esthetic deviation from the regular—from pure tone, true pitch, even dynamics, metronomic time, rigid rhythms, etc. (p. 9)

In one study, Seashore's (1938, p. 247) pianist performed the first 25 measures of Chopin's "Nocturne," op. 27, no. 2 in "artistic" and "attempted metronomic" time. Patterns of accumulated measure and phrase durations were similar between the two renditions, and the dynamic range of the metronomic version was restricted. Seashore (1938) also noted the relative absence of intensity cues in accenting, saying ". . . time is always a rival of intensity in giving accent" (p. 243).

Some explorations of the role of "deviations" in expressive performance have come from measurement-qua-measurement and analysis-by-synthesis studies. In these studies the relationship between the performer and listener has been largely neglected. Musical communication is concerned not merely with a single frame of reference, but includes the complex relationships among composer, performer, and listener. Gabrielsson (1988) notes that

> To find truly general results in performance data is therefore difficult. *The generalities should rather be sought in the relations between performance and* (the listener's) *experience.* (italics his, p. 46)

There are a few relevant experimental studies on musical communication. Nakamura (1987) investigated the ability of the performer to communicate dynamics to the listener. In general, the performer's intentions were communicated, particularly for crescendos and across the differing dynamic ranges of violin, recorder, and oboe. Tro (1989) investigated perceptual differences in performed dynamics by using entire pieces. Perception varied among sex, singers and instrumentalists, performers and audience; perception of dynamic range depended on the contrast level to the low-intensity passages.

The semantic differential was used by Senju and Ohgushi (1987) to evaluate the ability of a violinist to convey her ideas to the audience. In playing the first movement of Mendelssohn's Violin Concerto in e minor, the performer tried to represent 10 musical feelings labeled with such words as "weak," "sophisticated," "bright," "powerful," and "fashionable," which referred to playing style. Generally, the semantic differential responses of musically trained listeners showed weak correspondence with the intent. One limitation was the wide range of intended styles in relation to the music. Another was the use of inappropriate verbal attributes to define the performance instead of simply using commonly notated expressive markings.

Campbell and Heller (1979) investigated the ability of a cellist and a violist to communicate detailed notational changes in dynamics, bowings, fingerings, and articulations. They found that musicians could identify 79%, 65%, and 50%, respectively, of normal, amplitude compressed, and synthesized notational interpretations. This useful attempt to explore musical communication relies on notational cues to manipulate performer interpretation rather than on behavioral modeling. Clarke (1988) also manipulated notational cues. He moved all the bar lines in a highly redundant composition (Satie's *Vexations*) and found that performances "show consistent deviations from strictly metrical [metronomical] timing that produce a profile of partially periodic timing curves" (p. 11). He states, ". . . expressive profile is generated at the time of performance from information specified in the musical structure . . . " (p. 11). We believe that Clarke minimizes the distinction between an interpretation and musical expression. All deviations from mechanical performance need not be "expressive"; expression's domain is the mind of the listener. Notational signs such as bar lines *can* be tied to the pitch/time structure, *but need not be* (q.v. Ives, "The Cage"; Webern, "Variations for Piano", second movement; Stravinsky, "Marche du soldat" from *L'Histoire du Soldat*). The question is, what defines the musical structure to which interpretation is subject? G. Houle (1987) says

> It seems to be the belief of most seventeenth- and eighteenth-century theorists that musical meter is naturally and adequately perceived by the listener and only secondarily heightened through performance techniques. Most performers today are aware of how crude it is to suggest that the measure is identified by regular accent of dynamic stress based on bar lines and time signature. (p. 84)

We shall return to the issue of the nature of musical expression in the section on theory.

A Metabolic Communication Theory of Musical Expression

Any model of communication involves the transmission and reception of messages. Our view of musical communication requires, generally, three components.[1] The process of musical communication begins with an intended musical message that is recoded from ideation to notation by the composer, then recoded from notation to acoustical signal by a performer, and finally recoded from acoustical signal to ideation by the listener (Figure 1).

1. For this paper, we deal only with traditional Western art music in which composer, performer, and listener are involved. We do not consider the degree of participation of each person, or indeed, the absence of components altogether, as in, for example, the indigenous music of India, some electronic musics, jazz, and other improvisational forms. In addition, feedback loops, both acoustical and gestural, are not discussed. Campbell and Heller (1981) discuss a similar model, differing considerably in detail.

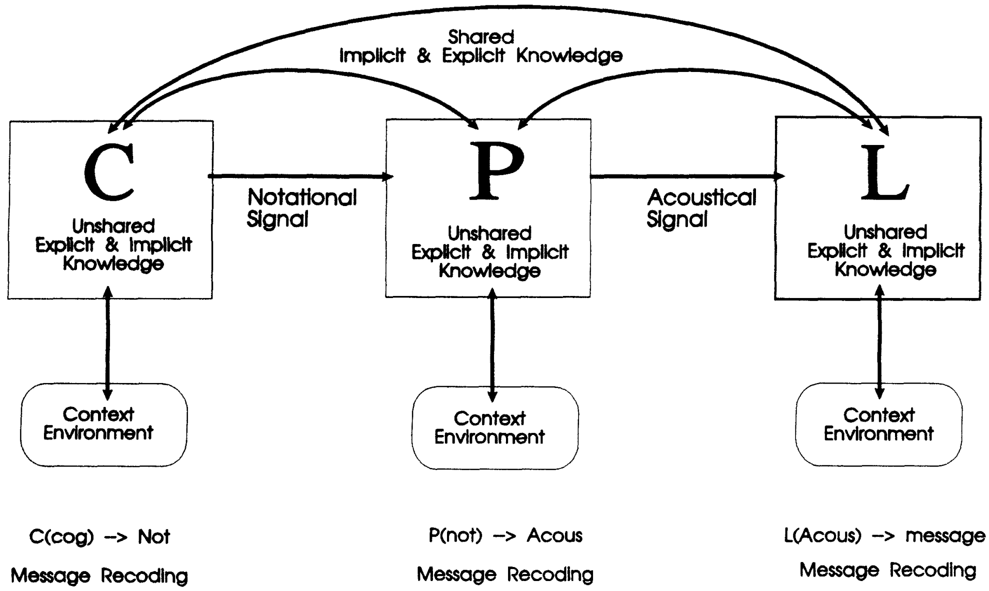


Fig. 1. A model of musical communication. C = composer; P = performer; L = listener.

We assume that a basic mental capability of the human is the grouping and parsing of elementary thought units (metasymbols).[2] In fact, this capacity is one basis for the survival of the organism. Human language systems are obvious examples of this capacity in respect of both referential and areferential elements. Our view is that music is but another manifestation of this capacity wherein the referential is largely suppressed. Abstract mathematics thus resembles music. From this perspective the answer to the question, "Is music necessary for survival?", is yes, insofar as human survival depends on the capacity for generating, synthesizing, and analyzing metasymbolic structures.

We conceive of music as based on a system of sequences of states that fill gaps of inexorable time. Depending on the frame of reference, these states can be, for example, sound states (perception), acoustic states (vibrational signal), symbol states (notational signal), and mental states (cognitive metasymbols). In fact, the transformation from one state sequence to another is the very core of the difficulty in musical communication. A sequence of states is not yet a pattern, not until some organization has

2. The mental representation, or form, of the idea that is manipulated by the creator varies by field and by person. Musicians "hear" or "feel" some sound structures, but others are vaguer or less immediate. Mathematicians say that they do not operate on symbols, but on indefinite (metasymbolic) mental forms and kinetic feelings. We assume that grouping and parsing by the composer, performer, or listener is carried out on lexical items stored as schemas in implicit and explicit long-term memories. Lexical items may be simple or complex, but generally they are stored as complex forms, in a phrasal lexicon. "Phrasal" is intended to convey the notion that the creator (composer) groups preformed structures, gestures, or schemas, like those of speech, into larger schemas. A composer may not select, order, group, and notate each individual note of a glissando or mordent.

been imposed on it by some mental process. It is important to note that what defines a state is relative to a frame of reference, and that within a frame of reference, various levels of state (e.g., microstates and macrostates) can exist at the same time

State changes are the basic building blocks for patterns. Short state sequences (protean units, as few as two states) can imply pattern, particularly in relation to the listener's schemata, as a single drawn line implies more than itself; this suggests expectation. In general, the structure of a state is defined relative to its context. We assume that musical states have a flexible, topologically deformable nature, and therein lies the source of frustration in grammatical or symbolic models of musical structure.

Consider a glissando for two octaves starting on C4 (Figure 2). This protean sequence consists, in the notational frame of reference, of 15 states (Figure 2a). Another representation simply uses three notational symbols (Figure 2b). Of course, Figure 2b is less specific, in that several interpretations are possible: a chromatic series, a diatonic series, or perhaps a portamento. Both examples rely on convention to aid interpretation of the relation of the glissando to the quarter notes. Part of the mapping from composer's ideation to notation depends on his explicit and implicit knowledge of performer/instrument properties.

Protean sequences may be termed "gesture"; the boundary between gesture, motif, and pattern is open to operational definition and empirical investigation. The potential pattern implied by the glissando (Figure 2) consists of structural states hung on inexorable time filled with a tissue of microstructural state changes, a gap-filling process.[3] A structural state is characterized by a degree of salience, this being defined relative to the frame of reference. It is possible to have a gap in any attribute of state; not only pitch, but temporal, dynamic, and timbral gaps are possible as well. The dynamics of gap-fill stem from constraints imposed by physiological, physical, perceptual, and cognitive limits. In the "real world," processes reach a limit and must change magnitude or direction to continue. In this way, quasiperiodic state changes mandate aspects of the form of the musical message.[4]

The recoding of musical messages depends on shared and unshared, implicit and explicit knowledge (Figure 1). Much musical activity emphasizes implicit over explicit knowledge: We "sense" tonal center, ex-

3. The term "inexorable" is meant to emphasize the imperative nature of the forward flow of time. The *perceptual* time window appears to widen and narrow with memory and attention, scanning events of the past and anticipating those to come. Backward motion of music can only be a forward reflection of the past. Attempts to map spatial relations onto musical structures as in fractal music lead to a paradoxical implication of negative time.

We consider the "gap-fill" process to be a general architectonic principle. In this sense, the term is quite different from that proposed by Meyer (1973).

4. Here we can only touch on the ramifications and possibilities of our position.

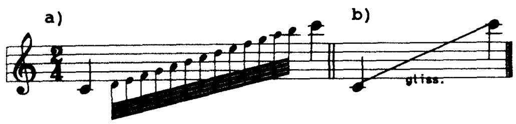


Fig. 2. Two ways of notating a glissando.

ecute micromuscular embouchure adjustments, achieve timbral balances and fusion, and parse the acoustical signal into a musical message.

The relationship between implicit and explicit operations is represented schematically in Figure 3 as an information processing model. In essence, the right-hand side of the figure deals with symbols; the left-hand side with metasymbols. Conscious awareness has direct access only to working memory and, through working memory, to explicit procedures and explicit long-term memory. Although conscious awareness can direct a problem or query to implicit procedures, it cannot access directly their content or form. External world inputs are parsed (differentiated and categorized) by implicit procedures, albeit under the potential direction of conscious awareness, and are conditioned by schemas in implicit long-term memory. The translator maps metasymbolic to symbolic units, and vice versa. What is perceived at the conscious level via working memory is a translated version of implicit knowledge; one never can know to what extent the mapping is isomorphic. In a similar way, explicit knowledge is, through rehearsal, mapped in a translated form to the implicit. In the figure, pairs of arrows in opposite directions imply parallel processing; the double-headed arrow implies serial processing.

It would be erroneous to interpret the model in terms of transformational grammars. Explicit and implicit procedures are not equivalent to deep and surface structures and transformational rules. Hierarchies are clearly a part of the musical message and its representation in any frame of reference, as any redundancy implies structural organization. The nature of hierarchies, whether strict, symmetrical, and unique or flexible, asymmetric, and ambiguous, must be taken into account. The role of hierarchy in our approach is flexible, asymmetric, and ambiguous. Therefore generative and transformational procedures are neither necessary nor sufficient in message recoding. Our processing concept is one of manifold procedures and multiple strategies, not of grammars. Indeed, generative theories have been found wanting in a number of domains. Roads (1985, p. 429) notes that ". . . the rewrite rule by itself has been shown to be insufficient as a representation in music." Minsky holds that grammars are static representations that distract researchers from studying music as a cognitive *process* (Roads, 1980). Further, Winograd (1972) found it necessary to eschew transformational grammars in favor of case grammars for the parsing of natural language.

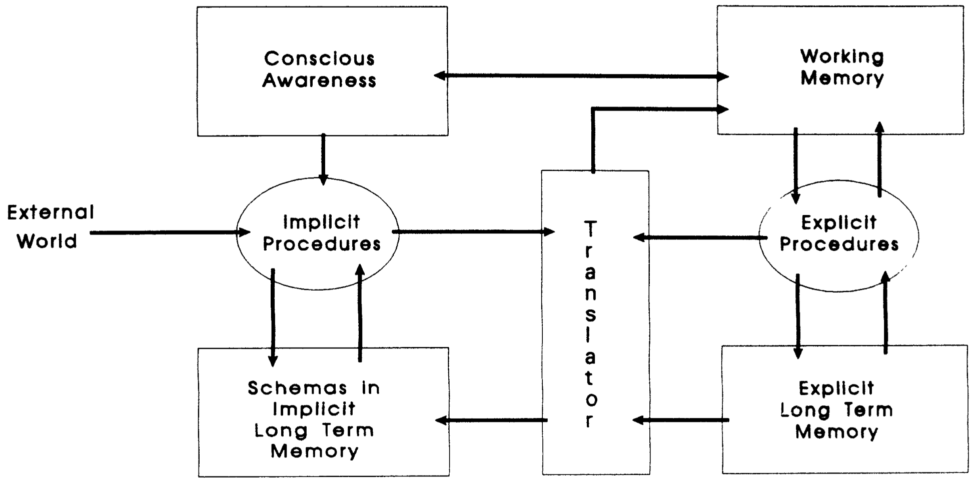


Fig. 3. Information-processing model showing some details of the relations among implicit and explicit knowledge and procedures of the persons depicted in Figure 1.

One of the problems in studying musical expression is that its chief domain lies in implicit procedures. Evidence for this can be seen in the difficulty of verbalizing about the generation and perception of expressiveness. It is no accident that most performance instruction is conducted one-on-one in an interactive, modeling environment. Despite its largely covert nature, *something* is being imparted by the performer, as is evident from the large number of multiple recordings and performances of the same music notation. In our study, expressiveness is the intended message generated by the performer and directed at the listener. When and if the intended message is received, communication has occurred. The performer's message is a synchronous modulation of the composed states that serve as the carrier. In the process of message parsing and recoding, the listener may impose meanings unintended by either the composer or performer. Indeed, it is in the nature of the listener always to seek meaning. Even if the composer's message is unparsable, the performer-generated message may be grasped by the listener. In this case, the composer's work serves as little more than a framework for the dialogue between performer and listener.[5]

The work reported here is concerned with some of these issues of musical communication. We propose to integrate performer-generated levels of expressive intent with perceptual and acoustical analyses in a series of

5. We admit the possibility of asynchronicity between expressive (performed) and composed state structures, particularly for generating performer hallmarks or novelty. The relationship between the two forms part of the dynamics of information content in the message, as is the case for pitch and time in the cognitive coding of melodies (Monahan, Kendall, & Carterette, 1987). Strict dodecaphonicism, pointillism, and use of sound structures à la Stockhausen may exceed cognitive limits, yet the performance may be parsed for meaning.

studies, which investigate the communication of musical expression between performer and listener as moderated by compositional style. In this, we go beyond previous research by considering at the same time many aspects of musical communication.

Experimental Rationale

As mentioned earlier, most previous investigations have focused on analysis and manipulation of acoustical aspects of musical performance. In contrast, we begin our study by conducting perceptual experiments, whose outcomes are then correlated with acoustical data. The initial research questions are directed at the performer-listener part of the communication model of Figure 3.

A number of methods were employed in order to converge on an answer to the question "to what extent are the performer's expressive intentions conveyed to the listener?" Different methods induce different listening strategies, a fact often ignored. For this reason, we adapted or invented techniques to explore the variety of listener responses. We used categorization, matching, and rating methods and two distinct classes of listeners, musicians and nonmusicians, and, in one experiment, we used both natural and artificial music. Categorization reveals the ability to assign an instance to a model when the whole range of possibilities can be reviewed. In contrast, matching compels the listener to remember the possibilities and respond only in the presence of a single model. Ratings require that the listener provide a measure of the degree to which an attribute is possessed by an instance. The use of artificial [Musical Instrument Digital Interface (MIDI) synthesized] performances was meant to remove some of the effects of the degrees of freedom at the disposal of the performer. Thus, data produced using these methods allow comparison and contrast.

Materials and Methods

Musical materials were drawn from the vocal literature of four stylistic periods, baroque, classical, romantic, and twentieth century, in order eventually to assess the influence of compositional style on the generation and communication of expressiveness. Because instruments were used, biases potentially present in instrumental music were reduced, we hope, by the use of vocal music.[6] Selection of phrases was guided by tessitura considerations for the five instruments used: piano, clarinet, oboe, violin, and trumpet. The tonal center for all selections was standardized by transposition to g minor; the meter was $\frac{3}{4}$. The melodies were "Thy Hand, Belinda" from *Dido and Aeneas* by Purcell (converted from $\frac{3}{2}$ to $\frac{3}{4}$), "Der Wanderer" by Haydn, "Der Müller und der Bach" by Schubert and "Weise im Park" from *Vier Lieder* by Webern. This paper reports only on results obtained using

6. Whatever music might be chosen, the problem of appropriateness arises. Music written specifically for each instrument would be ideal; however, the resultant confounding of expression and materials would cloud comparisons. We have chosen to take a middle ground, providing the same musical materials across instruments. We note the pervasive transcription of vocal melodies to instruments and the relative rarity of the inverse.

"Thy Hand, Belinda" (Figure 4). The bracketed phrase is that used for stimulus material; the entire phrase was performed and recorded, however.

Performances were recorded digitally (Sony PCM-601ES) on stage in a moderately reverberant concert hall (Schoenberg Hall, reverberation time ca. 1.6 sec). We used a coincident microphone (AKG Model 422) feeding a matrix box (Audio Engineer Associated Model MS 38) set for a crossed orthogonal figure eight with an axis of 45 degrees to the source. The height of the microphone was 1.6 m; the distance of microphone to piano center was 1.24 m, to wind instrument chair, 1.52 m, to violin chair, 1.19 m, and to trumpet chair, 1.75 m. These placements were established by a professional recording engineer, David Cloud, to optimize sound quality.

As a model, a professional concert pianist, Johana Harris-Heggie, performed each of the four melodies in their entirety at three intended levels of expressiveness: Without expression (*senza espressione*), with appropriate expression (*con espressione*), and with exaggerated expression (*con troppo espressione*). We instructed the pianist to interpret *without expression* as mechanical, *with expression* as appropriate and *with exaggerated expression* as "too much." Dynamic range was established in the following way: Players were asked to imagine a space of dynamics with *mezzo forte* at the center; then play the mechanical version statically at *mezzo forte*, and the expressive versions relative to this level. The authors and recording engineer monitored the recording session throughout. No other instructions were given to the performers.

Four professional instrumentalists (oboe, clarinet, violin, trumpet) listened with headphones to the recorded model performances of each of the three levels of expressiveness. They auditioned each model only once. The instrumentalists were asked to mark their scores *ad libitum*. Following this, the instrumentalist played the piece in his interpretations of the intended level of expression. The order of pieces was randomly set.

Digital tape recordings were reproduced in analog form and resampled onto computer hard disk for experimental use. Sampling was monophonic (stereo channels were mixed) at 28,571.4 samples/sec, 16-bit, using Canetics' PC-DMA model 16 analog-to-digital and digital-to-analog interface. Krohn-Hite model 3202 filters were used for anti-aliasing and smoothing. Playback from hard-disk was through a Sony TA-AX4 amplifier into Sennheiser HD-222 headphones. All experiments were controlled, stimuli randomized, and data collected by computer program.

Experiment 1

PROCEDURE

A computer-based categorization paradigm was used (Kendall, 1988). Model stimuli (piano) were represented by three colored bars on the computer screen (Figure 5). The selections to be categorized were displayed in a different region of the screen as undifferentiated white bars. Both models and selections were randomly ordered. The subject used a mouse to point at a bar and either select it for playing or for moving. The goal was to match a selection to an appropriate model by placing its bar underneath that of

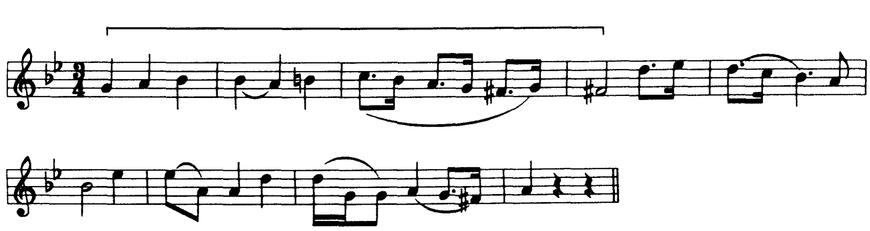


Fig. 4. Notation of excerpt from "Thy Hand, Belinda" by Purcell. The bracketed phrase was the musical material of the experiments.

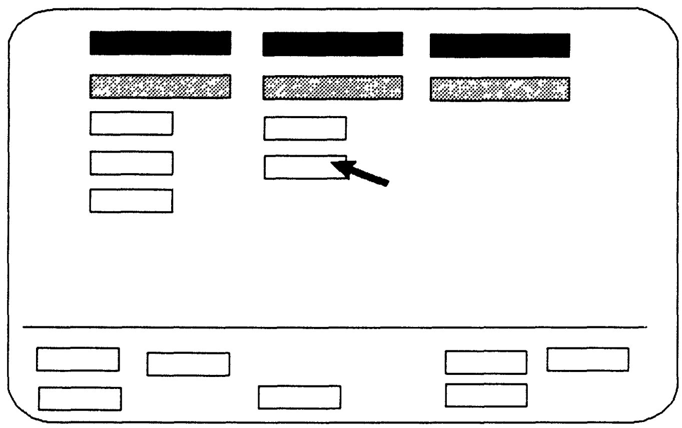


Fig. 5. Schematic diagram of the computer screen seen by a subject during categorizing.

the model. A practice session with different stimuli familiarized subjects with the procedure and equipment. Neither here nor in the main experiment was feedback provided regarding correctness of response.

Subjects were told that each piano model represented a different level of expression and that the remaining 12 musical selections were four instruments each playing at three levels of expression. Subjects were informed that each of the four instruments occurred only once for a given level of expression.

Subjects were allowed to replay and move stimuli until they informed the experimenter that they were finished. Therefore, at the end of the experiment, four different instruments were associated with a given model on the screen (see Appendix). This final procedure was the distillation of results obtained from 18 subjects, nine musicians and nine nonmusicians, who participated in various pilot studies.

SUBJECTS

There were 19 subjects in all, 10 nonmusicians and 9 musicians. Nonmusicians had less than 1 year of formal music instruction; most of the musicians were graduate students in musicology and had a minimum of 10 years of formal instruction. None had participated in the pilot experiments.

RESULTS AND DISCUSSION

Tables 1A and 1B show a set of eight confusion matrices; four for nonmusicians and four for musicians. The three columns represent the three expressive levels of the model performance (none, appropriate, exaggerated); each row represents the corresponding expressive level of an instrument (oboe, clarinet, violin, or trumpet). Each entry represents the sum over all subjects who categorized the selection with a level of the model. For example, musicians placed eight clarinet Level 1 performances under the piano Level 1, one under piano Level 2, and none under the exaggerated Level 3. (P1, P2, and P3 refer to piano models' levels of expression. Likewise, E1, E2, and E3 refer to expressive levels for each of the other instruments.) Therefore, the upper-left to lower-right diagonals of each matrix represent the "correct" (intended) matches for a level

TABLE 1
Categorization Frequencies

A. Musicians

	Clarinet			Oboe			Violin			Trumpet		
Expressive Level	P1	P2	P3	P1	P2	P3	P1	P2	P3	P1	P2	P3
E1	8	1	0	8	1	0	9	0	0	8	1	0
E2	1	6	2	1	6	2	0	2	7	1	6	2
E3	0	1	8	0	3	6	0	7	2	0	2	7
Mean hits (%)			81			74			48			78

B. Nonmusicians

	Clarinet			Oboe			Violin			Trumpet		
Expressive Level	P1	P2	P3	P1	P2	P3	P1	P2	P3	P1	P2	P3
E1	6	3	1	6	2	2	8	1	1	8	0	2
E2	1	7	2	1	6	3	1	2	7	2	6	2
E3	2	2	6	2	2	6	1	6	3	2	3	5
Mean hits (%)			63			60			43			70

of expression. Figures 6a and 6b are bar graphs of the data in Tables 1A and 1B.[7] These data are collapsed over groups to produce Table 2. Chi-square analyses of Table 2 row frequencies were all statistically significant ($p < .025$, df = 2). Data were collapsed across instruments to produce Table 3. Chi-square analyses of row frequencies were again all statistically significant ($p < .025$, df = 2). Note that an analysis was performed on each row and that the alpha level was adjusted for these repeated analyses. A two-way analysis is inappropriate, because there is lack of independence of observations in columns.

Note that the intended expression level matches the subject categorization well beyond chance level. However, for violin the categorization of expressive Levels 2 and 3 was poor, with the two categories being reversed. We will return to these findings later.

In addition to subject categorization data, the computer recorded the amount of time (in sec) the subject spent listening to a given selection, the number of hearings, and the number of category changes. Analysis of variance using multiple general linear hypotheses (ANOVA using MGLH) on repeated measures indicated that the difference in the mean listening durations for musicians (mean = 48.2 sec) and nonmusicians (mean = 72.1 sec) was not statistically significant [$F(1,17) = 3.14$, $.05 < p < .082$].

7. Because the expected frequencies were too low, and because of repeated measures, chi-square analysis was not performed on the frequencies of Tables 1A and 1B. Tables 2, 3, and 4 had adequate minimum expected frequencies, and were analyzed by row only, because each row (but not column) was independent.

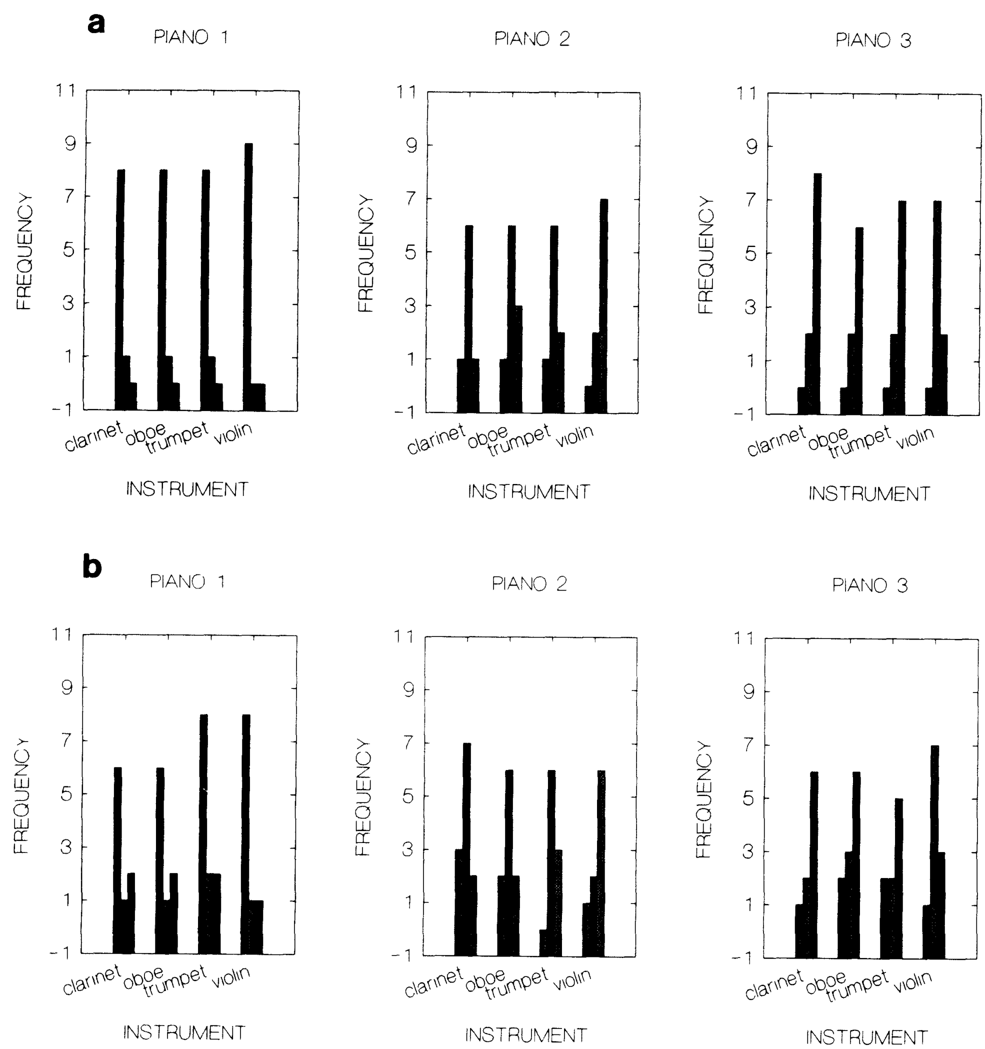


Fig. 6. Graphs of the data of Tables 1A and 1B.

TABLE 2

Categorization Frequencies for Combined Musicians and Nonmusicians

	Clarinet			Oboe			Violin			Trumpet		
Expressive Level	P1	P2	P3	P1	P2	P3	P1	P2	P3	P1	P2	P3
E1	14	4	1	14	3	2	17	1	1	16	1	2
E2	2	13	4	2	12	5	1	4	14	3	12	4
E3	2	3	14	2	5	12	1	13	5	2	5	12
Mean hits (%)			72			67			46			74

However, across all combinations of instruments and levels of expression (Table 4), the listening time was significantly different [$F(11,187) = 4.70$, $p < .0009$]. Figure 7, based on the means in Table 4, is a line graph that

TABLE 3
Categorization Frequencies Summed over Instruments

	Musicians			Nonmusicians		
Expressive Level	P1	P2	P3	P1	P2	P3
E1	33	3	0	28	5	7
E2	3	20	13	6	21	13
E3	0	13	23	6	14	20
Mean hits (%)			70			64

TABLE 4
Mean Listening Durations (sec)

Expressive Level	Clarinet	Oboe	Violin	Trumpet
E1	49.74	53.11	36.37	45.05
E2	73.05	68.90	77.32	59.37
E3	71.00	67.63	70.37	57.90

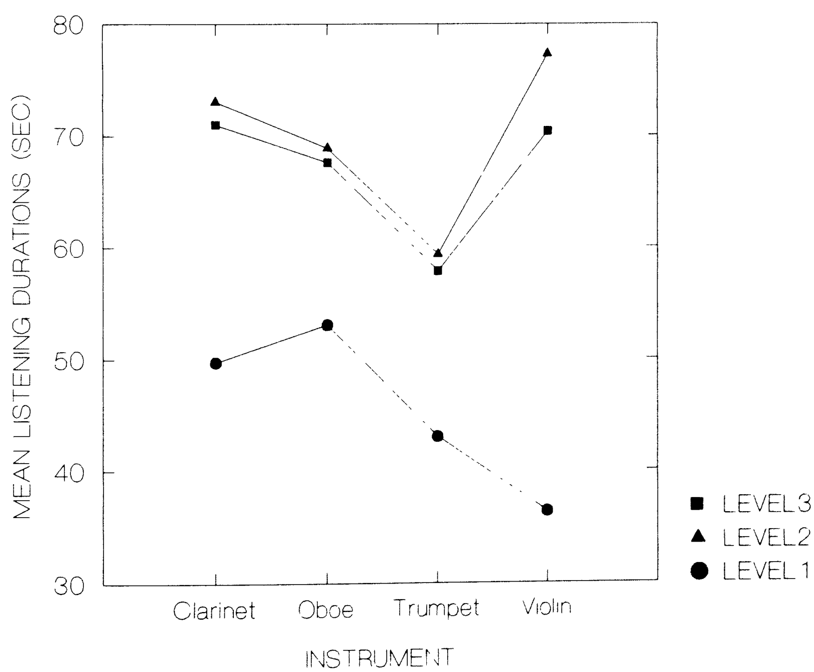


Fig. 7. Mean listening durations during categorization, across instruments and expressive levels.

shows that decision-making required more listening time in the case of nonmechanical expression. In addition, the appropriate level of expressiveness, Level 2, consistently had the highest listening mean durations. It is notable that the violin Level 2 listening mean durations were the highest in the group; this level of expressiveness was often interchanged with Level 3 in the categorization task (Table 1). Musicians listened a fewer number of times (mean = 5.7) to the stimuli than did the nonmusicians (mean = 9.1)[$F(1,17) = 5.11$, $p < .037$]. Across instruments and levels of expression the mean number of times a stimulus was auditioned was significantly different [$F(11,187) = 2.85$, $p = .002$] (Table 5, Figure 8). Except for the oboe, the second level of expressiveness yielded a higher frequency of listening than the other levels. Comparison of the mean durations across instruments and levels of expression (Figure 7) with

TABLE 5
Mean Number of Times Auditioned

Expressive Level	Clarinet	Oboe	Violin	Trumpet
E1	7.05	7.58	5.68	6.32
E2	8.47	8.47	8.68	7.32
E3	7.74	9.11	7.68	7.00

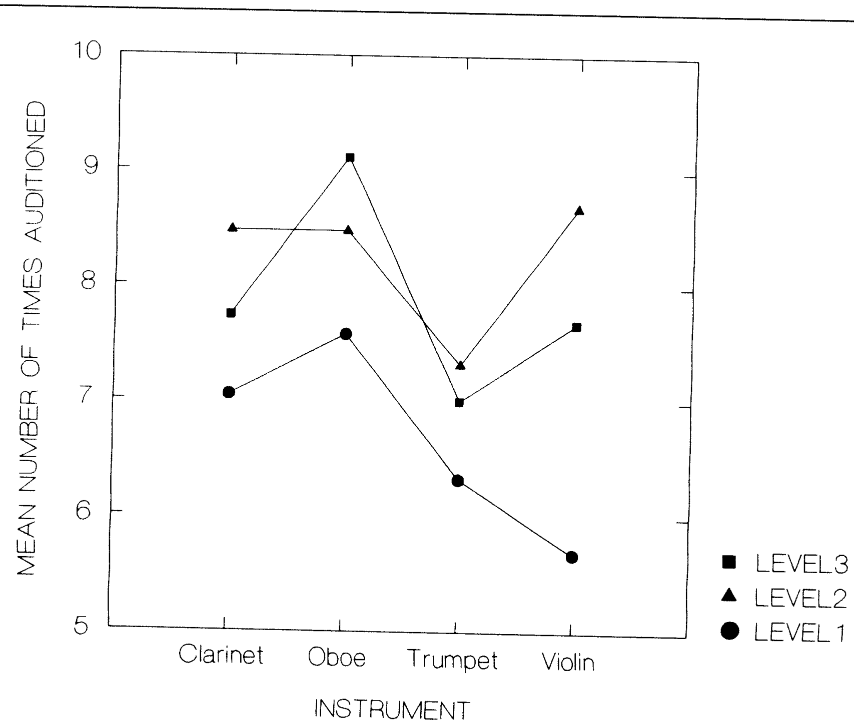


Fig. 8. Mean number of times that a subject listened during categorization, across instruments and expressive levels.

the means for times auditioned (Figure 8) shows a similar profile. The only significant outcome for category changes was across all combinations of instruments and levels of expression [$F(11,187) = 2.54, p < .005$] (Table 6). A line graph of these means is shown in Figure 9, which indicates that more category changes were made for Levels 2 and 3 expressive performances. We suggest that these data indicate a higher cognitive load in discriminating between expressive Levels 2 and 3.

Experiment 2

In the categorization paradigm, subjects can review any and all stimuli. In this way, the listener can build a perceptual space. In contrast, matching

TABLE 6
Mean Number of Category Changes

Expressive Level	Clarinet	Oboe	Violin	Trumpet
E1	1.32	1.16	1.11	1.37
E2	1.42	1.53	1.68	1.53
E3	2.05	1.63	1.42	1.37

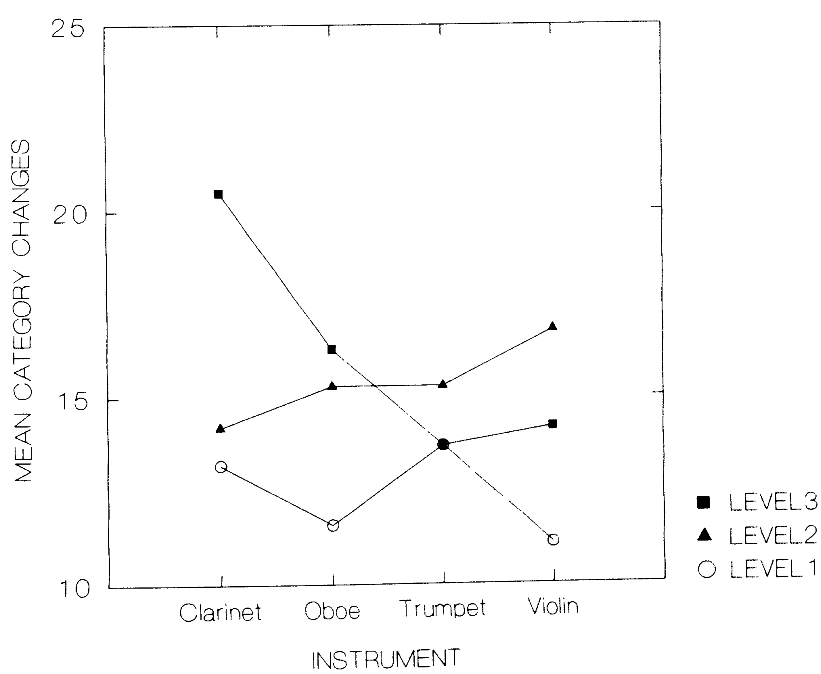


Fig. 9. Mean number of category changes during categorization, across instruments and expressive levels.

constrains the listener in developing a perceptual space; emphasis is placed on the immediate model and choice set. As an independent measure of a listener's ability to discern intended levels of expressiveness, we used a contrasting, matching method.

PROCEDURE

The task was a three-alternative forced-choice paradigm with subject review. On a given trial, one of the 12 selections (four instruments × three levels of expression) was heard, followed by all three piano renditions. The subject then selected the piano rendition that best fit the level of expressiveness of the initial example. The subject could review both the example and piano choices. Stimuli were randomized. The subject response was made by highlighting a letter label (A, B, or C) and pressing a mouse button. Two series of the 12 selections were presented.

SUBJECTS

Ten musicians, each paid for participation, served as subjects. In order to qualify, they had to have 10 or more years of formal musical training, and could not have participated in categorization Experiment 1.

RESULTS AND DISCUSSION

Table 7 is a confusion matrix like those shown in Table 1. Although the hit rates are well above chance (33%), they are considerably below those obtained in Experiment 1, except for the case of the violin, which fared poorly there. Figure 10 is a bar graph of Table 7 data. In order to determine differences in correct responses across instruments, we analyzed all scores within diagonals of Table 7. We scored subject responses as correct when the model and choice were matched, averaged replications, and submitted the data to a 12-level repeated-measures ANOVA. Across instruments, correct matches were statistically equivalent [$F(11,99) = 1.135$, $p < .343$], as the mean hits of Table 7 show. In order to compare correct and incorrect matches for a given level of expressiveness, matrices were summed across the four instruments, creating a single 3 × 3 matrix. Because row sums were equal (80), leading to singularities, the data were perturbed by adding a noise value that randomly ranged between $+/-1.00$ (Table 8).[8]

Three two-factor 3 × 3-level ANOVAs on repeated measures were performed, one for each column of the resulting summed matrix; alpha levels were adjusted accordingly. In all cases, the correct match (diagonal of

8. For a discussion of perturbation of data in singular matrices, see Press, Flannery, Teukolsky, and Vetterling (1986). While not identical, our procedure embodies some of their ideas. In our case, the lack of proportional weighting of the perturbation of off-diagonal values works against the null hypothesis, a conservative approach.

Table 8) produced higher mean scores than off-diagonals [without expression: $F(2,6) = 34.55$, $p < .001$; with appropriate expression: $F(2,6) = 7.97$, $p < .02$; with exaggerated expression: $F(2,6) = 20.25$, $p < .002$].

TABLE 7
Match Frequencies (Musicians)

	Clarinet			Oboe			Violin			Trumpet		
Expressive Level	P1	P2	P3	P1	P2	P3	P1	P2	P3	P1	P2	P3
E1	11	7	2	10	6	4	14	2	4	14	2	4
E2	4	14	2	6	13	1	3	10	7	7	6	7
E3	2	9	9	3	5	12	1	10	9	5	2	13
Mean hits (%)			57			58			55			55

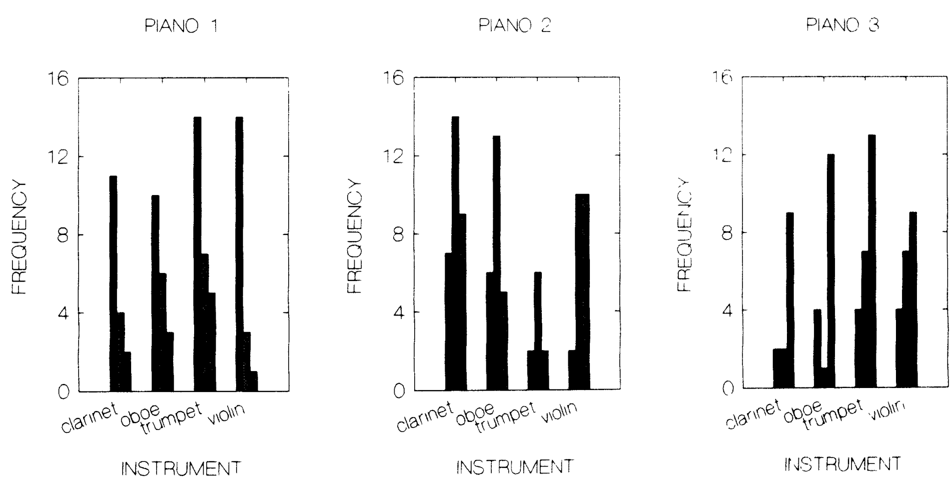


Fig. 10. Graphs of data of Table 7.

TABLE 8
Perturbed Match Frequencies (Table 7) Summed over Instruments

Expressive Level	P1	P2	P3
E1	12.63	4.84	3.98
E2	4.99	10.72	4.12
E3	2.75	6.17	10.83

Musical Signal Analysis

We explored various signal correlates to musical expression, including timing, amplitude rates of change, vibrato and time-variant spectral characteristics. In terms of the musical expression communication model (Figure 4), we investigated some aspects of performer and listener recoding in the acoustical signal frame of reference.

TIMING

Computer algorithms aided in partitioning the acoustical signals. Signals were parsed in 100-sample windows (3.5 msec). First, the absolute value of the average of minimal and maximal amplitudes within the window was obtained. A second-order forward-difference operator was applied to these values in sequence (Press et al., 1986). A threshold was set for determining the onset of a note. Figure 11 illustrates the output of this procedure for the piano expression Level 1 performance. Vertical bars represent tracked note onsets, which were based on the filter function output shown as a dotted line above the signal envelope. An arrow indicates the tracking of a subtle legato transition.

Figure 12 illustrates timings for clock-time performance of the rhythms of "Thy hand, Belinda." The first row of timings (msec) is cumulative; individual note durations are shown in the second row. All clock times were based on an M.M. quarter note equal to 63 beats per minute. Figure 13, based on timings extracted by the computer algorithm, plots deviations of performed durations (msec) from the clock times of Figure 12. Thus, from Figure 13, note 4 of clarinet Level 2 (with appropriate expression), it can be seen that the clarinetist lengthened the note almost 600 msec. As this quarter note was 952 msec in clock time (Figure 12), the total length was about 952 + 600 = 1552 msec.

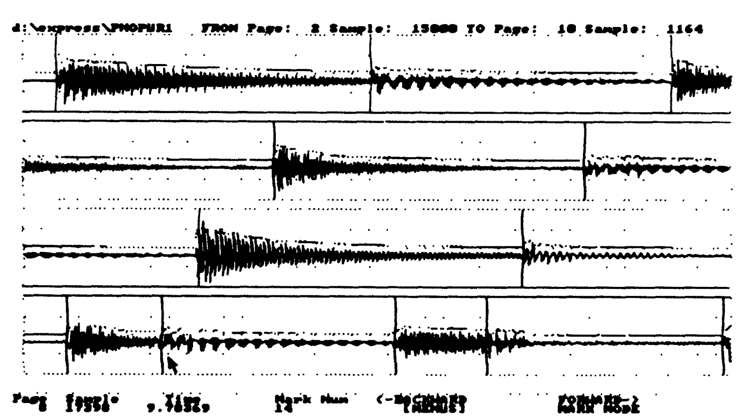


Fig. 11. Screen display shows the segmented acoustical signal of Piano 1. The arrow indicates the tracking of a subtle legato transition.

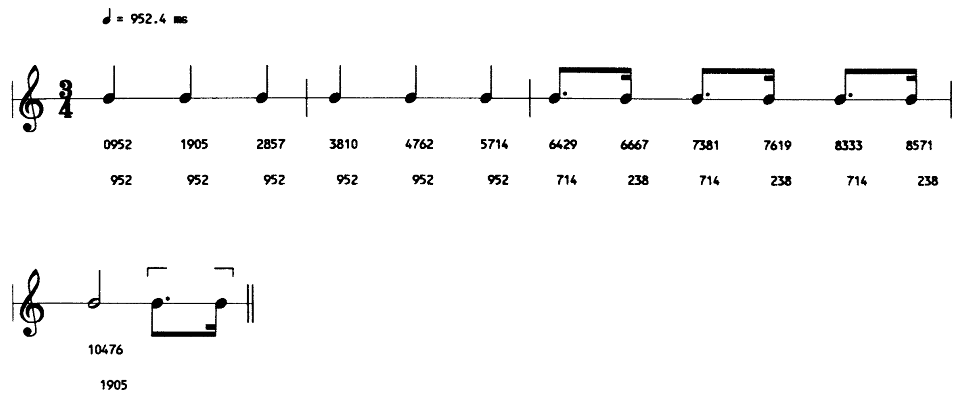


Fig. 12. Clock-time durations, cumulative (top) and individual notes (bottom) in msec.

In Figure 13, the three levels of expressiveness are the columns from left to right. It is notable that, for a given column, the timing contours are similar; across expression levels (rows), however, there are striking contour differences. To quantify this relationship, we submitted the timings to a paired-correlation analysis (Table 9). These correlations were then submitted to cluster analysis with complete linkage (Figure 14). A few features deserve comment: Mechanical expressive level timings cor-

TABLE 9
Correlations of Time Deviations

	Piano 1	Oboe 1	Trumpet 1	Violin 1	Clarinet 1
Piano 1	1.000				
Oboe 1	.658	1.000			
Trumpet 1	.167	.443	1.000		
Violin 1	.038	.374	.222	1.000	
Clarinet 1	.007	.136	.371	.441	1.000
	Piano 2	**Oboe 2**	**Trumpet 2**	**Violin 2**	**Clarinet 2**
Piano 2	1.000				
Oboe 2	.831	1.000			
Trumpet 2	.677	.824	1.000		
Violin 2	.156	.389	.271	1.000	
Clarinet 2	.628	.726	.475	.545	1.000
	Piano 3	**Oboe 3**	**Trumpet 3**	**Violin 3**	**Clarinet 3**
Piano 3	1.000				
Oboe 3	.421	1.000			
Trumpet 3	.135	.621	1.000		
Violin 3	.442	.661	.721	1.000	
Clarinet 3	.523	.707	.803	.922	1.000

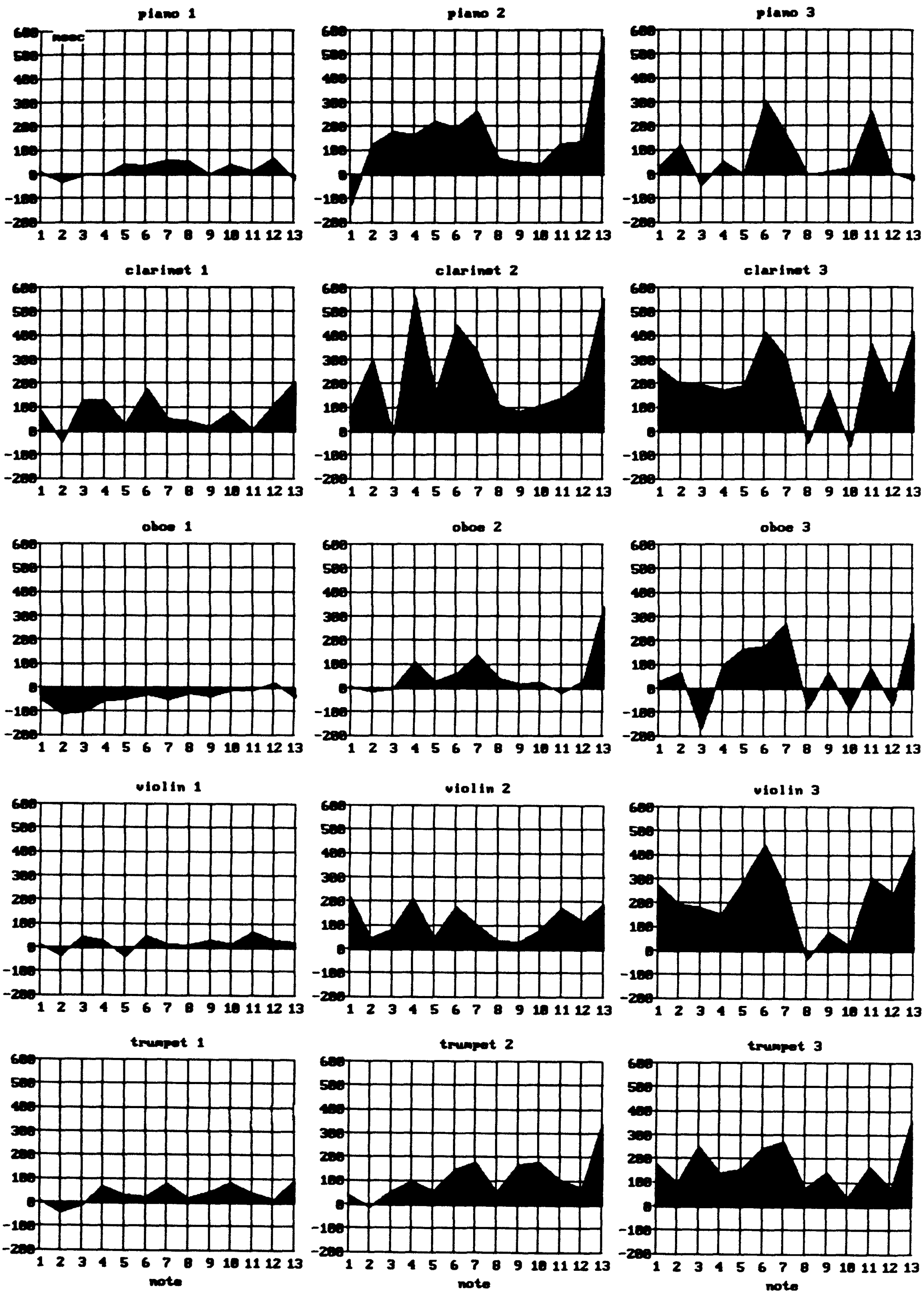


Fig. 13. Performance deviations from ontological time (msec) for three expressive levels and five instruments.

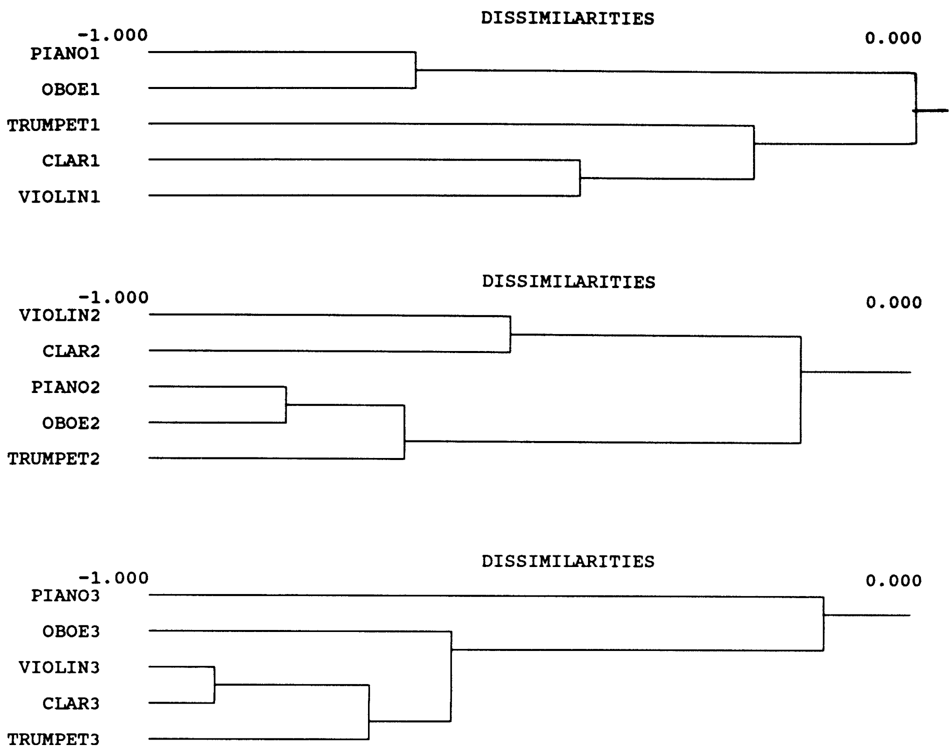


Fig. 14. Cluster analysis (complete linkage) of correlations of timing deviations for the three expressive levels.

relate less well (mean = .287, r = .007–.658) than those of expressive Level 2 (mean = .553, r = .156–.831) and expressive Level 3 (mean = .589, r = .135–.922). Cluster analyses with complete linkage (Figure 14) show the close timing correlation of piano and oboe across expressive levels. Similarly, clarinet and violin timing correlations form a pair. Trumpet timings consistently appear as a separate branch.

Within instruments and across levels of expression, the piano performances each have a distinctive timing profile. The timing correlation of expressive Levels 1–3 for piano averaged only −.045. The four imitators were much less able to create distinctive separations between expressive level timings, with mean correlations of .213, .433, .412, and .560, respectively for oboe, clarinet, violin, and trumpet. Indeed, a comparison of timing correlations for the first six notes versus the last six notes (omitting note 13) supports the above finding. In general, timing correlations are greater for the second half of the phrase with increasing expressive level (unfortunately, the data sets are too numerous to report here).

AMPLITUDE

For each of the five instruments at each of the three expression levels, root-mean square (RMS) amplitude values were calculated within consecutive 100-sample (3.5-msec) windows. Figure 15 presents the result of this RMS analysis. Note that time vs. amplitude relations can be discerned from the vertical position of note-onsets (triangles). In general, the RMS values by note are smaller for the middle, appropriate expressive level than for the other two; the dynamic range is greatest for Level 3, followed by Level 1. One notes that, for instruments commonly played with vibrato in this country (oboe, violin, and trumpet), amplitude vibrato extent increases with each expressive level (although the dynamic range of the oboe vibrato is barely visible in Figure 15). Total duration of the phrases is shown by the position of the last plotted point in each of the 15 graphs. Without exception, the expressive performances are longer. Indeed, the graphs of deviations from ontological time (Figure 13) show that, with few exceptions, note values get longer with increasing expressiveness.

The RMS values based on 100-sample segments were averaged for each note, producing a single value. Correlations of RMS means within instruments and between expressive levels are high in the case of piano, clarinet, and trumpet, with mean correlations of .85, .90, and .76, respectively. The oboe RMS means are high between expressive Levels 1 and 2 (.67) and 2 and 3 (−.43); however, almost no correlation is found between Levels 1 and 3 (−.05). Violin RMS correlations are consistently low (mean = .12) across expressive levels.

INTERACTIONS OF TIMING AND AMPLITUDE

We correlated time deviations and mean RMS values for each note within instruments and across expressive levels. Figure 16 plots z-scores of mean RMS and time deviations for piano, clarinet, trumpet. RMS values (dotted lines) are provided for the three levels of expressiveness; time deviations (solid lines) are only plotted for expressive Levels 2 and 3. Correlations of time deviations and RMS values are moderately high and largely negative (Table 10). This indicates a general tendency for notes longer than ontological time to be lower in RMS value, and vice versa. Sixteenth notes in particular have a tendency to be higher in RMS and shorter than ontological time, in other words, the sixteenth notes are emphasized in time and loudness.

Figure 16 shows that the overall patterns of RMS and time deviations are, at the macrostate level, similar, sometimes with a displacement (slight phase shift) on the time axis. However, at a microstate level, patterns homogeneous within instruments can be quite different across them. A general tendency at the microstate level is for the performer to alternate

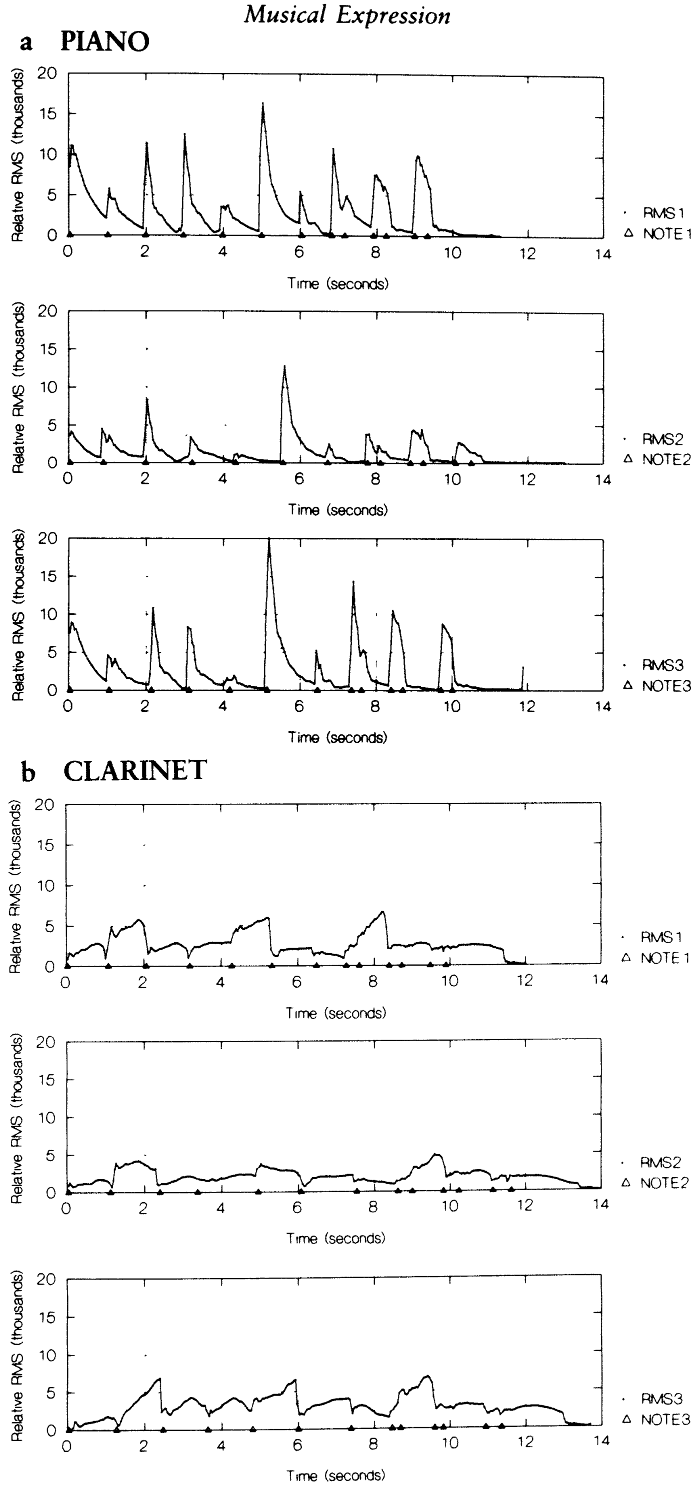
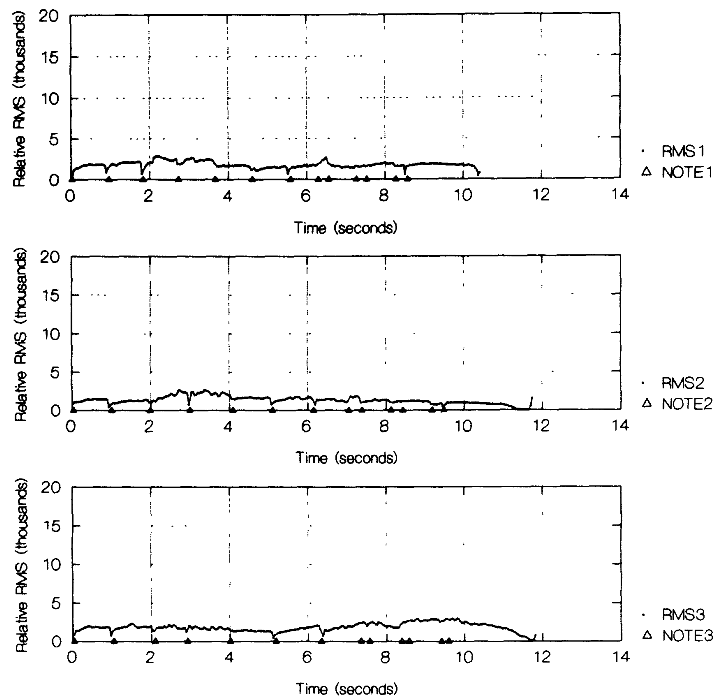


Fig. 15. Relative root-mean-square (RMS) graphs for each performance, grouped by instrument in order of intended expressive level (RMS 1–3; NOTE 1–3). Open triangles mark note onsets. (a) Piano, (b) clarinet, (c) oboe, (d) violin, (e) trumpet.

c OBOE



d VIOLIN

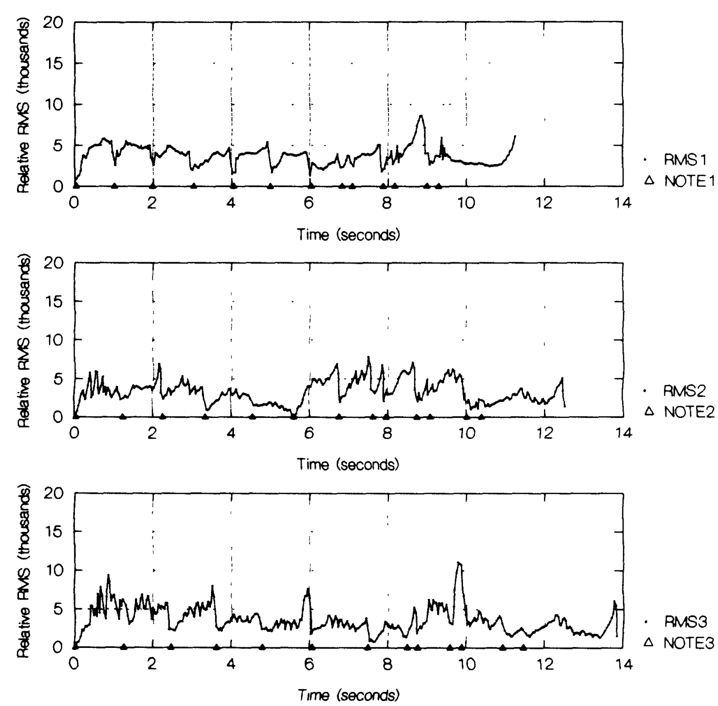


Fig. 15 *continued*

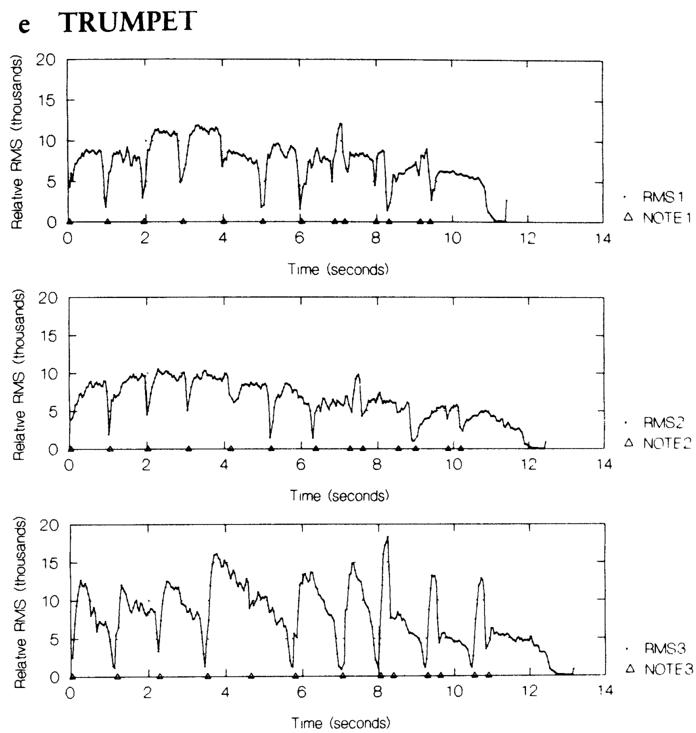


Fig. 15 *continued*

TABLE 10
Correlations of Time Deviations and RMS (Z-Scores)

Expressive Level	Piano	Clarinet	Oboe	Violin	Trumpet
E1	.60	−.63	.04	.12	−.21
E2	−.35	.06	.31	−.17	−.67
E3	−.145	−.22	−.49	−.57	−.50

time deviations and RMS changes in patterns of twos or threes. For example, the pattern short-long in time paired with more-less in RMS value is pervasive. But rule-writing from such observations would fail to capture the wide range of microstate tissues that are observed. We will return in the General Discussion to the time/RMS interaction.

Experiment 3

In an initial attempt to explore the relative contributions of variables such as timing and RMS to expressive musical communication, we synthesized a new set of signals. Experiment 3 deals only with timing deviations.

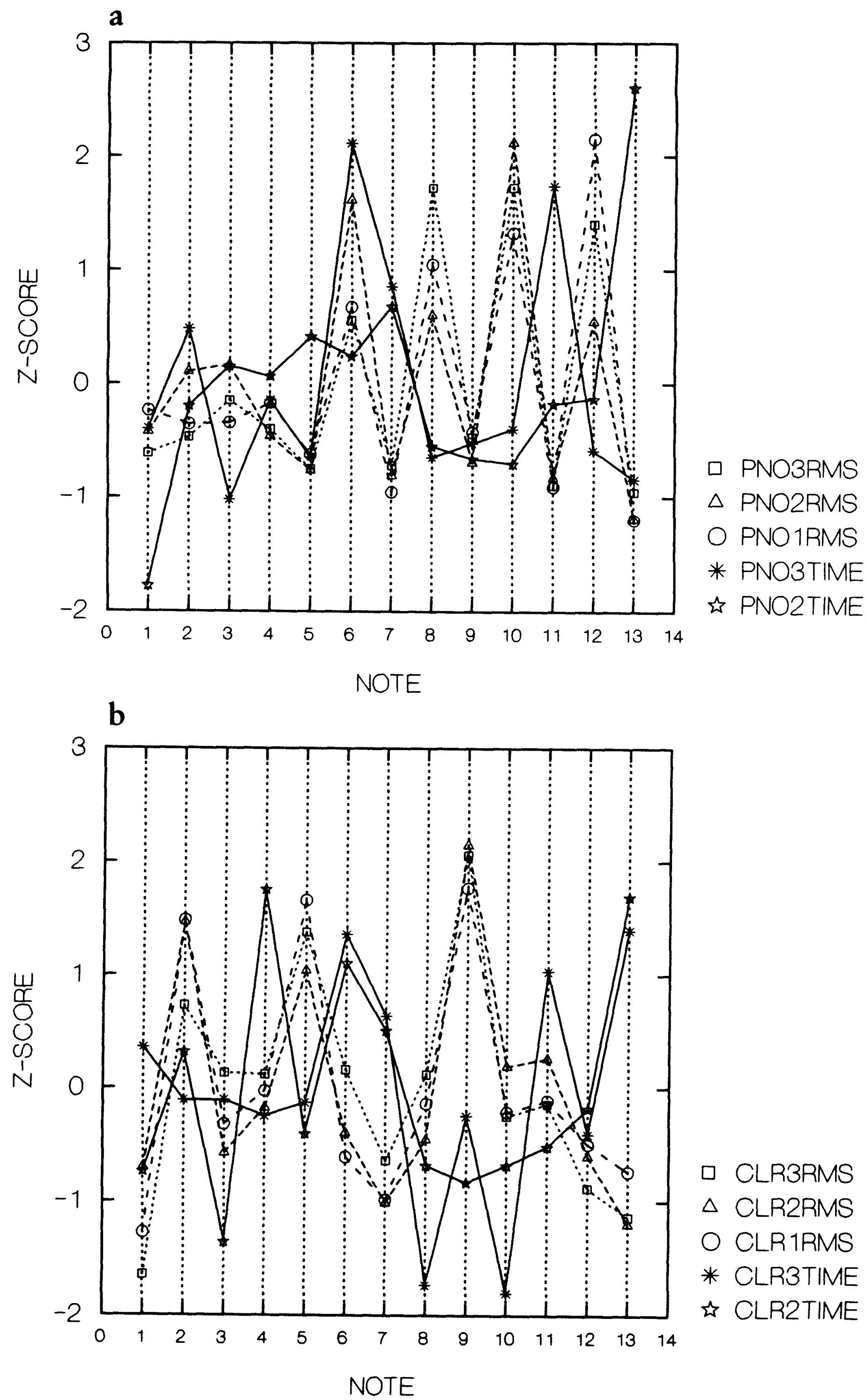


Fig. 16. Plots of root-mean-square (RMS) averages and time deviations as standardized (Z) scores for (a) piano, (b) clarinet, and (c) trumpet.

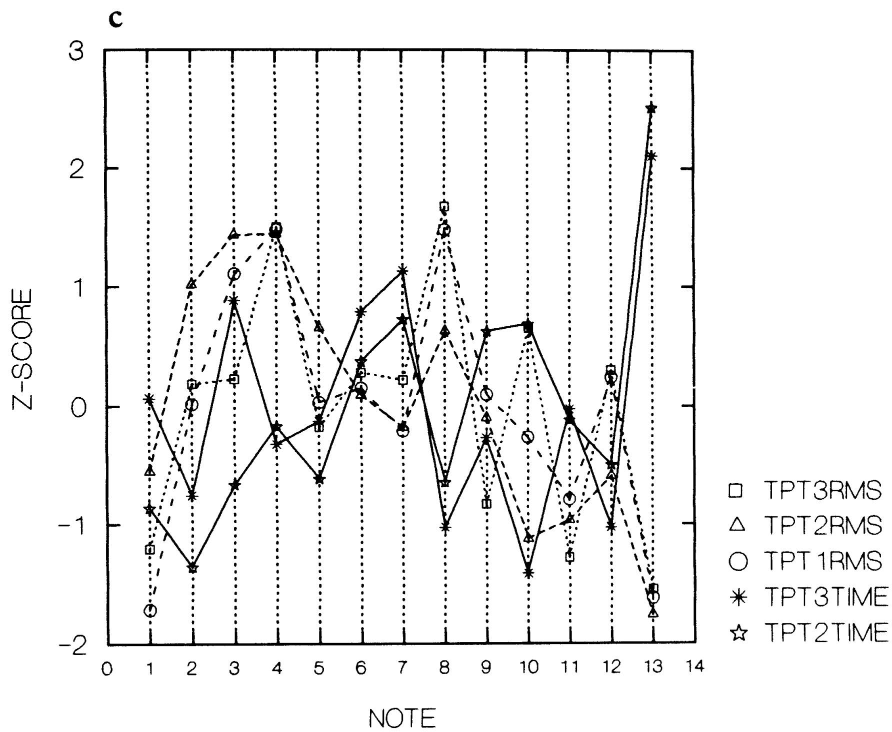


Fig. 16 *continued*

PROCEDURE

We used a Kwai K1-M digital synthesizer module driven by MIDI programmed in assembly code in order to achieve maximum timing accuracy conservatively gauged at 3 msec (monophonic). The factory internal signals, based on single periods of sampled instruments, were used modified as follows: The "sustain level" portions of the envelopes were adjusted to achieve approximately equal loudnesses. No amplitude or frequency modulations were permitted; the amplitudes were fixed across notes. The oboe was created *de novo*. Note-on events alone were initiated at the clock times of the real performances for the three levels of expressiveness. Data were collected under the same categorization procedure as described in Experiment 1.

SUBJECTS

Twelve nonmusicians served as subjects. None of these subjects participated in any previous expression experiment.

RESULTS AND DISCUSSION

The categorization outcomes are presented in Table 11. A comparison of these data with those of Table 1B shows notable similarity of the pattern of hits. Consider, for example, violin matrices: The same inversion of expressive Level 2 and 3 categorization is evident. Timings alone, with

TABLE 11
Nonmusician Categorization Frequencies: Artificial Performances

	Clarinet			Oboe			Violin			Trumpet		
Expressive Level	P1	P2	P3	P1	P2	P3	P1	P2	P3	P1	P2	P3
E1	7	3	2	7	0	5	5	5	2	8	2	2
E2	4	8	0	2	8	2	2	2	8	2	6	4
E3	1	1	10	3	4	5	5	5	2	2	4	6
Mean hits (%)			69			56			33			56

reasonably natural signals, seem to be sufficient for categorization accuracy that is nearly equal to that obtained with real instruments. This does not imply that these synthetic performances were "expressive," but only that they contained some aspect of expressiveness sufficient to permit reasonable categorization accuracy.

Experiment 4

In this experiment, we applied a third convergent method to examine the communication of musical expression. It was intended to assess the relationship between expressive level and the perceived magnitude of expression gauged by ratings.

PROCEDURE

Both natural and artificial sets of mechanical, appropriate, and exaggerated expressive performances were used. Each set of stimuli, natural and artificial, was presented in a group. A given subject received a randomly determined order within sets and was randomly assigned to hear either natural or artificial signals first, followed by the remaining group. The subject's task was to rate a stimulus on a scale from 0 (without) to 100 (great) expressiveness. The subjects were instructed as follows:

> We want to know how musicians communicate expression to listeners. Musical expression can be likened to the expression of an actor in speaking his part: He may speak in a monotone, in a manner appropriate to the idea, or he might exaggerate.
>
> In this experiment, five instrumentalists played the same music with different levels of expression. I will now play some examples which are in order of increasing expression. [Examples were played.]
>
> You will hear a set of 15 interpretations. Rate each example along the scale according to your judgement of the degree of expression.

A labeled scale was presented on the computer screen. Subjects responded by moving a pointer on the computer screen by using a mouse. The position of the pointer, from 0 to 99, was read by the computer and stored as score data.

SUBJECTS

Eight musicians served as subjects. None had participated in previous experiments. Only musicians were used in order to parallel the matching Experiment 2.

RESULTS AND DISCUSSION

ANOVA on repeated measures (n = 8) indicated significant differences between mean ratings of the natural performances [$F(14,98) = 4.125$, $p < .001$] but not of the synthetic performances. Mean expressive ratings as a function of instrument are plotted for both natural and synthetic performances in Figures 17a and 17b, respectively. Tukey-*a* ($\alpha = .05$) post-hoc paired-comparisons revealed no significant differences between any expressive Level 2 or 3 means. Post-hoc analyses (Tukey-*a*, $\alpha = .05$) indicate that those means in Figure 17a that are 25.86 units apart were statistically different. This value clearly differentiates the mechanical performance ratings from those of expressive performance in the case of natural renditions.

We combined values for expressive Levels 2 and 3 on the basis of post-hoc analysis. These data were subjected to a Group (Synthetic/Natural) by instrument/expression (mechanical/expressive) by subjects repeated measures ANOVA, with adjustment of α for multiple tests. These results indicate that the profile of means (Figure 17) is different for natural and synthetic (time-deviation-based) performances [$F(9,126) = 1.99$, $p < .046$].

Intersubject variability for the rating task is very high and clearly accounts for the general lack of significant differences. We strongly suspect that data from more subjects would differentiate the general trends shown in Figures 17a and 17b, namely that appropriate expression (Level 2) would be rated higher in expressiveness than either mechanical or exaggerated performances.

Cluster analysis with single linkage (nearest neighbor) was performed on natural performance ratings (Figure 18). There is a general tendency for ratings to cluster according to expressive level. Violin Level 3 is a conspicuous exception, being displaced into association with expressive Level 2. You will recall that subjects tended to switch violin Level 3 with violin Level 2 in categorizing.

General Discussion and Conclusions

We conclude that, in general, both musicians and nonmusicians can discern expressive intent. Greater than chance responses were obtained for all cases other than artificial violin performances (Table 11). Differing response tasks, however, yielded somewhat different results.

In particular, matching yielded lower overall hit rates than did categorization. There was a tendency for adjacent intended expressive levels to be confused by the listener in a matching task. However, statistical analysis confirmed that, across instruments, intended expressive levels were matched with the models.

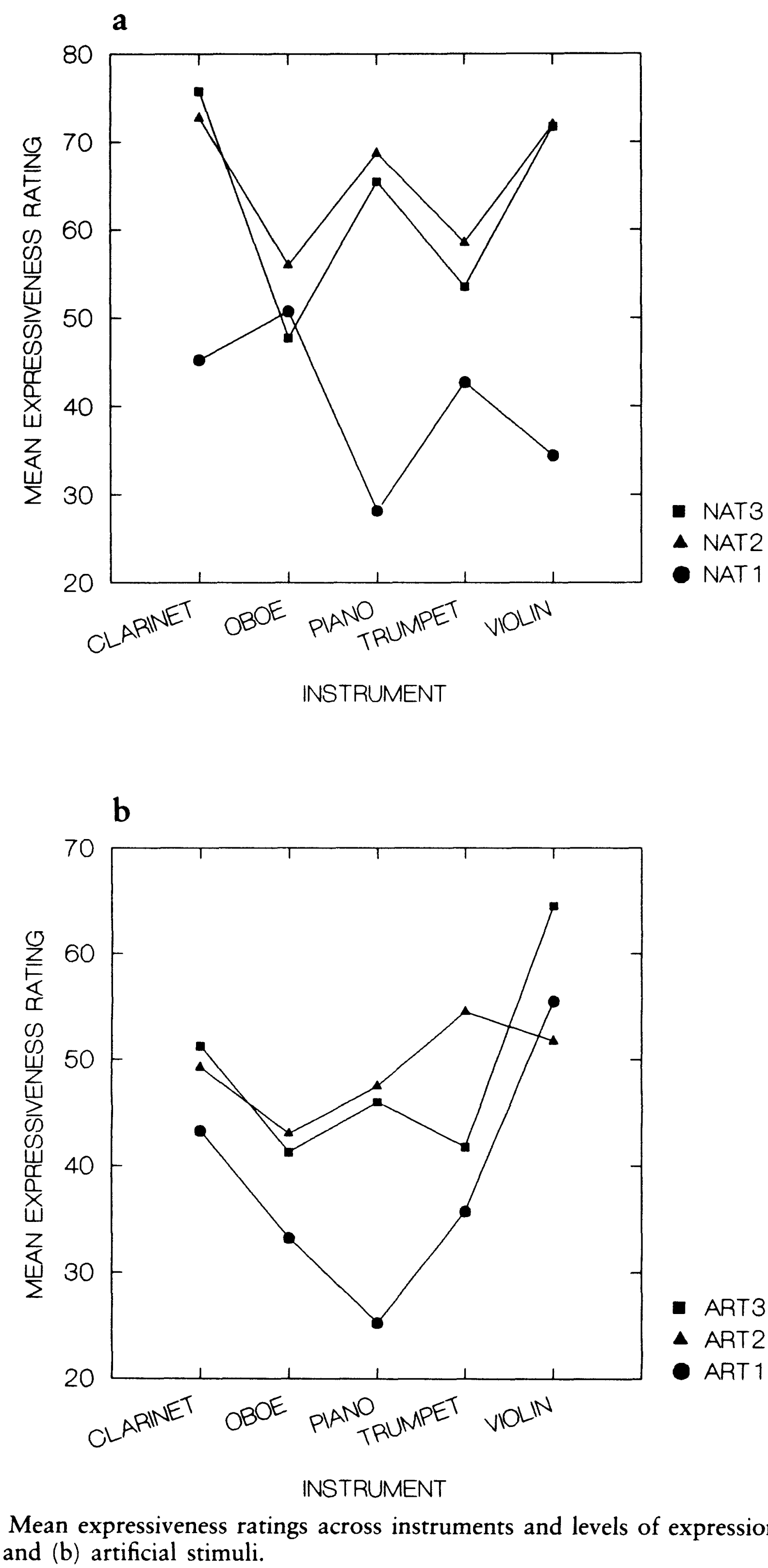


Fig. 17. Mean expressiveness ratings across instruments and levels of expression for (a) natural and (b) artificial stimuli.

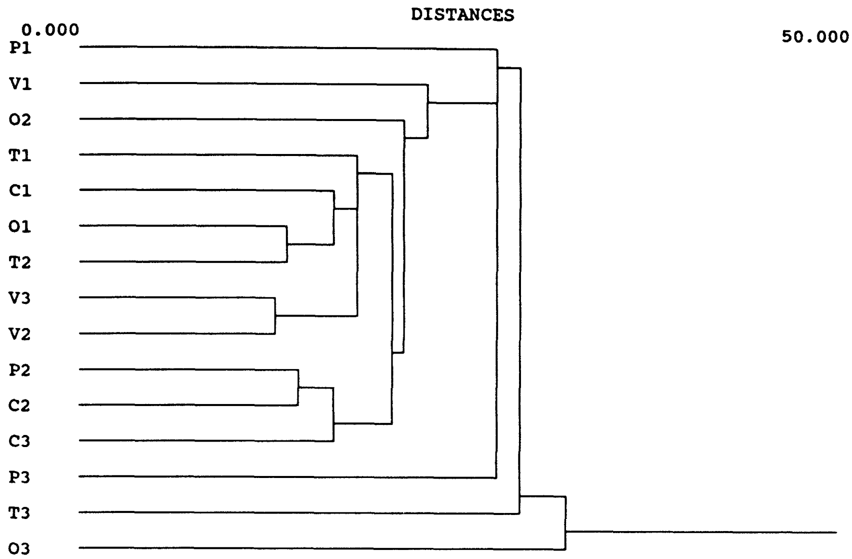


Fig. 18. Cluster analysis (single linkage) of expressiveness ratings for natural performances.

Ratings of expressiveness produced data parallel to those of matching. Expressive Levels 2 and 3 produced nearly identical rating means. It is worth noting that for both natural and artificial renditions the mean expressiveness ratings were highest for oboe, piano, and trumpet at expressive Level 2, appropriate expression. There was no suggestion to the subject that higher expressiveness ratings went with appropriateness. Indeed, it is likely that our instructions worked against a stronger result.

Cluster analysis of the ratings of natural performances differentiated well between levels of expressiveness (Figure 18). Violin Level 2 and 3 ratings were reversed in the cluster diagram. A similar reversal was found in the categorization task.

This reversal of expressiveness levels has its root in the categorization task, which permits extensive review of the entire range of models and choices, whereas matching does not. Extensive review in the categorization task leads the listener further astray. Confirming evidence for this reversal was found in acoustical signal cross-correlations. Piano Level 2 and violin Level 3 amplitudes and time-deviations were positively correlated (r = .530, .542, respectively). Piano Level 2 and violin Level 2 amplitudes were not correlated. Piano Level 3 and all violin time-deviations were moderately correlated (.349, .306, .442, respectively). In all other cases, piano models and intended choices had time-deviations that were positively correlated.

These data suggest the existence of conditions conducive to the confusion of violin categories. Apparently, repeated listenings exposed salient features for reversing categories; reversal was not as evident in matching data. Such a feature may be the peculiar pattern of longer and shorter time deviations around note 11 for violin Levels 2 and 3, which is not evident in the cases of other instruments (Figure 13).

An interesting finding was the lack of timing correlation among expressive levels for the piano, in contrast to moderate, positive correlations among levels for other instruments. This world-class pianist, given the task of creating three distinctive expressive levels, was able to do so in terms of timing-deviation profiles, yet the other, less-experienced instrumentalists did not do so. Correlations among RMS values within instruments were high for piano, clarinet, and trumpet, and less homogeneous for oboe and violin. This finding generally supports the proposition that RMS values were less useful as cues for categorizing than timing-deviation profiles. This conclusion is further supported by the fact that, for nonmusicians with artificial renditions based on timing-deviations, the categorization hit levels were comparable with the natural performances.

The relation of "musical structure" to deviations from canonical notation is emphasized by a number of researchers, including Todd (1985) and Clarke (1988).[9] Clynes (1983) suggests the idea of "composers pulses," which consist of periodic patterns of deviations, a generalization strongly disputed by Repp (1989).

Our data fail to support something as strict and invariant as *the* musical grammar, performer grammar, or listener grammar. Piano Levels 1 and 3 of Figure 15, for example, show the following pattern of relative RMS values for the first six quarter notes (high = H, low = L): HLH-HLH. Piano Level 2 (appropriate) is, on the other hand, LLH-LLH. Clarinet RMS (Figure 15), across all levels, is in the pattern: LHL-LHL. In Figure 15e, trumpet, a symmetrical arch of RMS values is observed. Similar patterns and variations of patterns likewise can be discerned in timing-deviations.

It is clear that all performers signal salient structural points by temporal and dynamic contrast. Examples are the sixth to seventh note transition, signaling pitch-time contour direction change, and the final cadence. The number of ways to signal these structural features is very large. Contrast is a key operative principle, not merely "accent" in terms of increased magnitude. Contrast patterns form the microstructural tissue that fills structural gaps.

9. Relevant to our conceptual position is that of Clarke (1985), who acknowledges that ". . . musical structures may be thought of as possessing a double aspect: A relatively fixed canonical representation . . . in a score and a more flexible and indeterminant representation that is evident in expressive performance." (p. 211). In addition he notes the variability in performance attributable to performer intent. Philosophically, he admits flexibility, but chooses the more formal approach in his experimental work (Clarke, 1988), focusing on "the structure."

There is a very large number of possibilities at the disposal of the performer for solving the problems of musical communication. What distinguishes the great performer from the merely skilled is the richness and invention of the solutions. In this sense, the composer and performer play similar roles: They both solve musical problems.[10,11]

References

Bengtsson, I., & Gabrielsson, A. Analysis and synthesis of musical rhythm. In J. Sundberg (Ed.), *Studies of music performance*. Stockholm: Royal Swedish Academy of Music, 1983, #39, pp. 27–60.

Campbell, W. C., & Heller, J. J. Judgements of interpretation in string performance. Paper presented at the Research Symposium on the Psychology and Acoustics of Music, University of Kansas, Lawrence, Kansas, February 21–22, 1979.

Campbell, W. C., & Heller, J. J. Psychomusicology & psycholinguistics: parallel paths or separate ways. *Psychomusicology*, 1981, *1*(2), 3–14.

Clarke, E. C. Structure and expression in rhythmic performance. In P. Howell, I. Cross, & R. West (Eds.) *Musical structure and cognition*. London: Academic Press, 1985.

Clarke, E. Generative principles in music performance. In J. A. Sloboda (Ed.), *Generative processes in music*. Oxford: Clarendon Press, 1988, pp. 1–26.

Clynes, M. Expressive microstructure in music, linked to living qualities. In J. Sundberg (Ed.), *Studies of music performance*. Stockholm: Royal Swedish Academy of Music, 1983, #39, pp. 76–181.

Gabrielsson, A. Timing in music performance and its relations to music experience. In J. A. Sloboda (Ed.), *Generative processes in music*. Oxford: Clarendon Press, 1988, pp. 27–51.

Houle, G. *Meter in music, 1600–1800*. Bloomington, IN: Indiana University Press, 1987.

Kendall, R. A sample-to-disk system for psychomusical research. *Behavior Research Methods, Instruments, & Computers*, 1988, *20*(2), 129–136.

Meyer, L. B. *Explaining music: Essays and explorations*. Berkeley, CA: University of California Press, 1973.

Monahan, C., Kendall, R., & Carterette, E. The effect of melodic and temporal contour on recognition memory for pitch change. *Perception and Psychophysics*, 1987, *41*, 576–600.

Nakamura, T. The communication of dynamics between musicians and listeners through musical performance. *Perception and Psychophysics*, 1987, *41*, 525–533.

Press, W. H., Flannery, B. P., Teukolsky, S. A., & Vetterling, W. T. *Numerical recipes*. Cambridge, England: Cambridge University Press, 1986, pp. 133–135.

---

10. A preliminary interpretation of some of the data reported in this study was presented as invited papers to the First International Conference on Music Perception and Cognition, Kyoto, Japan, 17–19 October 1989, and to The Institutes of Eastern and Western Music, Seoul National University, Seoul, Korea, 24 October 1989.

11. We thank our many subjects, musicians and nonmusicians alike; David Cloud for arranging and conducting the recording sessions; Johana Harris-Heggie for performing the piano models of expressive levels, and the graduate musicians for their interpretations of these models: Amanda Walker, clarinet; Margaret Gilinksy, oboe; Caroline O'Keefe, trumpet; and Jacke Carrasco, violin. We are grateful to our graduate research associates, Glenn Cornett, Scott Lipscomb, Kathryn Vaughn, and Suk Won Yi. Partial financial support was provided by the UCLA Academic Senate Committee on Research and by Canetics, Incorporated, Pasadena, CA.

Repp, B. Perceptual evaluations of four composers' "pulses." In *Proceedings of the First International Conference on Music Perception and Cognition*. Kyoto, Japan: The Japanese Society of Music Perception and Cognition, 1989, pp. 23–28.

Roads, C. Interview with Marvin Minsky. *Computer Music Journal*, 1980, *4*, 25–39.

Roads, C. Grammars as representations for music. In C. Roads & J. Strawn (Eds.), *Foundations of computer music*. Cambridge, MA: MIT Press, 1985, pp. 401–442.

Seashore, C. E. *Psychology of music*. New York: McGraw-Hill, 1938. (Reprinted, New York: Dover, 1967).

Senju, M., & Ohgushi, K. How are the player's ideas conveyed to the audience? *Music Perception*, 1987, *4*, 311–323.

Sundberg, J. Computer synthesis of music performance. In J. A. Sloboda (Ed.), *Generative processes in music*. Oxford: Clarendon Press, 1988, pp. 52–69.

Sundberg, J., Frydén, L., & Askenfelt, A. What tells you the player is musical? An analysis-by-synthesis study of music performance. In J. Sundberg (Ed.), *Studies of music performance*. Stockholm: Royal Swedish Academy of Music, 1983, #39, pp. 61–67.

Todd, N. A model of expressive timing in music. *Music Perception*, 1985, *3* (1), 33–58.

Tro, J. How loud is music? Experience with the evaluations of musical strength. In *Proceedings of the First International Conference on Music Perception and Cognition*. Kyoto, Japan: The Japanese Society of Music Perception and Cognition, 1989, pp. 353–358.

Winograd, T. *Understanding natural language*. New York: Academic Press, 1972.

Appendix

Subject Instructions for the Categorization Procedure

We are interested in musical communication; the ability of the performer to send a message to the listener. None of the parts of our study is a test of musical talent or a test of your ability to hear.

Before we run the main experiment, a practice session has been designed to acquaint you with the procedure and equipment.

You will notice on the screen bars of various colors in different positions. The white and yellow bars stand for musical selections.

Use the mouse to move the red pointer to a bar. Press the LEFT button. You will hear the musical selection associated with that bar. Try another bar by moving the pointer and playing the selection.

The screen is divided into two sections by a white line. The bars at the bottom of the screen are to be moved to one of the columns headed by yellow and purple bars. The purple bar defines a group. The yellow bar underneath the purple bar is the model (prime example) for the group. The model can be played by pointing and pressing the LEFT button.

The task is to listen to the models, and assign choices which best fit within a group. To move a white bar to a group, point to the bar and press the RIGHT button. The bar will turn red. Now, point to the purple group bar and press a button. The white bar will move from the bottom of the screen to the selected group.

Try it.

A bar can be moved from one group to another in exactly the same manner. Try it. Also, you can replace a bar to the selection area at the bottom of the screen. Select the bar and point anywhere below the line and press the LEFT button. The bar will return to its original position.

Try it.

Now, as practice, listen to the models and then move choices to the group it best fits. You will have two in each group.

Are there any questions?

We want to know how musicians communicate expression to listeners. Musical expression can be likened to the expression of an actor in speaking his part: He may speak in a monotone, in a manner appropriate to the idea, or he might exaggerate.

In this experiment, five instrumentalists played the same music with three different levels of expression: with little expression, with moderate expression, and with exaggerated expression.

The yellow bar under each group is the model example. The order of the groups is random as is the order of the choices. Listen to the model examples. Move choice examples to the group they best fit.

NOTE that there are FOUR choices which fit under EACH model for a total of 12. EACH CHOICE in EACH GROUP is a different instrument.

You may listen to any choice or model as often as you like. You may change your mind and move choices from one group to another until you are satisfied.

Remember—group on the basis of similar level of musical expression, not instrument type.

Are there any questions?