SESSION 2 LABORATORY SYSTEMS

Chaired by Robert W. Allan, Lafayette College

Convergent methods in psychomusical research based on integrated, interactive computer control

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

Much of what we learn from an experiment depends on what we ask of the subjects and how we ask it. Most music perception and cognition research is implemented using a single methodology. Recently, questions have been raised about the validity and generalizability of such a limited approach, particularly regarding such complicated theoretical constructs as musical tonality. We propose that a useful technique is to converge on the answer to experimental questions by applying multiple methods, in essence, simultaneously investigating the central research question as well as ancillary questions of method. Such an approach is facilitated by a computer hardware and software system that integrates experimental design, data collection, data analysis, and data dissemination, linking off-the-shelf components through standard file formats and data-exchange protocols. The list-based system, using an IBM-type PC running Windows 3.0, handles real-time natural sound sampling and playback, signal synthesis and analysis, MIDI input and output, digitized picture display, and various subject-response mechanisms. On the basis of graphical icons, the integrated system can be used to build custom experimental designs. Examples of convergent psychomusical experiments that were designed, implemented, and analyzed using the system are discussed.

An often overlooked principle of perceptual research is that *what we learn is dependent on what we ask and how we ask it*. In this paper, we explore the implications of this principle through an examination of psychomusical research methods. Recent experimentation in musical communication using multiple methods facilitated by a computer-based experiment-generating system provides evidence of the utility of a multifaceted, converging research methodology. Finally, we discuss a newly completed, object-oriented, free-form Musical Experiment Development System (MEDS), which facilitates multiple-methodology audio perception research. In order to provide the background that compelled the implementation of the MEDS, it is necessary to examine several important aspects of the empirical research paradigm.

PSYCHOMUSICAL RESEARCH

A major struggle in research design is that between reliability and validity; this is particularly true of psychomusical experiments. As the researcher approaches the natural musical setting of a live concert, for example, the empirical context becomes increasingly musically valid. At the same time, the variability of this context, both across time and across geographical regions, results in potentially unreliable data as a multiplicity of unidentified and uncontrolled variables interact. Conversely, as the researcher limits the musical universe to single, synthetic audio events, one gains reliability and replicability at the expense of musical validity. Until relatively recently, experimental research in music has erred in the direction of reliability, a fact bemoaned by Meyer (1956) in his discussion of the errors of atomism, hedonism, and universalism in psychological research. In fact, Mursell (1937) vehemently attacked the approach of Seashore (1938) on grounds of musical invalidity, with a specific focus on atomistic definitions of musical concepts such as timbre and musical talent.

UCLA Academic Senate Research Grants to both authors provided part of the resources for this work. For additional information on MEDS, contact R. A. Kendall at the Department of Ethnomusicology and Systematic Musicology, Schoenberg Hall, 405 Hilgard Ave., Los Angeles, CA 90024.

However, the proliferation of computer power has permitted the generation and manipulation of musical stimuli of increasing ecological validity. The validity of musical experiments can thus be enhanced among several dimensions: (1) Synthetic stimuli can have increasingly complex time-variant properties; computational speed has permitted precise stimulus generation using relatively inexpensive computers and software (e.g., CSOUND). (2) Natural-instrument signals can be sampled and controlled at musical-bandwidth rates permitting investigation of performer-initiated messages, orchestration, and related acoustical effects. (3) All types of stimuli can be manipulated flexibly by the computer as a control system. MIDI devices as well as natural and seminatural (e.g., sampling keyboard) signals are easily controlled, and any or all types of stimuli may be mixed in an experiment, allowing for comparison and contrast of results of differing levels of musical validity.

Although this computer power is now becoming increasingly widespread, the methodological flexibility at the researcher's command largely seems to be ignored. Several possible reasons can be put forward for this phenomenon: (1) Although computers permit easy manipulation of the components of an experiment, the computer skills required to prepare hardware and software to gener-

ate experiments, particularly experiments using natural, time-variant musical signals, were difficult and time-consuming to attain. (2) Most researchers did not build a system of hardware-software components that interfaced easily to other analytical tools, such as signal and statistical analysis software. This required still other programs to be written to perform mundane, but crucial, data translation tasks. (3) The hardware-software systems developed for individual research projects permitted single methodologies to be generated. New programs were written for each paradigm, requiring additional resource expenditures. This compelled researchers to tailor their research to the available resources.

There is perhaps no other more limiting effect on the field of music perception than the latter point. Research questions are often driven by the availability of method, rather than method being at the service of and driven by the problem. We take this latter perspective, in contrast to the APA Publication Manual admonition that research should use established methodologies. After all, how do methodologies develop and evolve if not in response to new research problems?

Another limiting aspect of current research orientation in music psychology can be broadly termed *the reification of concepts syndrome*. At its worst, this mind-set pre-

Figure 1. MicroEditor software screen display. The vertical bars are set at 1/10 of a SMPTE frame set for 30 frames per second, but may be set in a variety of different measurements and resolutions (all figures have been changed from color to monochrome with a white threshold of 69%).

Figure 2. The MEDS main design window before an experiment has been created.

sumes that the objects and concepts identified by noun words in theory, such as *tonality*, *pitch*, *timbre*, *loudness*, and *rhythm*, are orthogonal and correspond to something unique and unchanging. Instead, such concepts are identified in theory by the model volitionally accepted or developed by the researcher as a set of interrelated relationships between variables. The variables and the hypothetical relations between them are established by the researcher. We reject the premise that the properties of observable "reality" are "out there" and are in a fixed, immutable state, accepting instead the relativism of observer-based modeling: We create the experimental world in terms of our own theories. The value of a model is based on the predictive relations between variables established in theory and scientifically tested, and it must take into account issues of reliability and validity, discussed above. It is unlikely that any single model will completely account for all of the observables in the "real world," even with respect to a single concept, such as tonality, timbre, or pitch (for a detailed discussion of these issues, see Kendall, 1984, 1986, 1987). Recently, we have suggested (Kendall & Carterette, 1990) that psychomusical research must abandon the privileged status accorded single frames of reference, such as the notational and acoustical, and must include aspects of the complex interrelationships between composer, performer, and listener as the musical message is recoded into and from the notational and acoustical signal by these three behavioral systems. Thus, a given mu-

sical concept can have a different structure and definition at many levels in the process of musical communication.

The methodological implications of such a position are profound. Operational definitions of concepts identified in theory are directly reflected in the methodology. Different methodologies can, and often do, reflect different facets of a concept. It is not that one particular approach is inherently "true" to the exclusion of another, but rather that different methods produce different aspects of, and converge upon, the truth. Thus, we would argue that the myopic focus on singular methods often obscures the multidimensional nature of the musical communication process, a point to which we return below. We suggest that the research process outlined above mandates the use of multiple methods and the invention or derivation of new approaches, and that new computer systems must enhance the creation of such convergent methods.

The Principle of Convergence

Convergence can be thought of in terms of the necessary and sufficient conditions for a given result to be obtained. As a binary concept, $2^2 = 4$ states can be identified: sufficient but not necessary, necessary but not sufficient, both necessary and sufficient, and neither necessary nor sufficient. Let us take a simplified, theoretical example: An experiment employing two methods, A and B, differing in a single attribute, X, such that A lacks X and B possesses X. If both Methods A and B yield equiva-

Figure 3. The audio search function window (see text).

lent results, then the attribute X is sufficient, but not necessary (since A, without X, performs as well). If Method A yields an interpretable result but B does not, then X is neither sufficient nor necessary. If Method A yields an uninterpretable result but B is interpretable, then X is both sufficient and necessary. Finally, if neither method yields an interpretable result, the implication is that X is necessary, but not sufficient. Kendall (1986, p. 195) provides additional discussion of this logic, and we specifically apply this reasoning in our studies of verbal attributes of timbre, discussed below.

Of course, the example is a simple case, and the differences between methods may be multivariate. Also, Methods A and B may both yield interpretable results that differ in substantial and meaningful ways, providing evidence for a multidimensional approach to the experimental concept. Yet, in this case, we still converge on different facets of the truth, and additional collateral studies employing different methods are required.

Methods can be varied in terms of both stimulus and response properties and the temporal relations between them. It is often necessary, when converging on a multidimensional concept, to vary aspects of both stimulus and response and to translate subject data into formats that are comparable, for example, to change rating data into similarity data and vice versa. We turn now to examples of experiments that vary method in order to illustrate the preceding discussion, and we follow this with the presentation of the MEDS, which facilitates such studies.

Tonality

Of the many areas of psychomusical research, tonality has recently sparked perhaps the most lively controversy (Krumhansl, 1990b; see also Butler, 1990). Traditionally (and phenomenologically), tonality is "loyalty to tonic," a tendency for "active" tones in a melody to "move" toward "rest." Imagine the "amen" at the close of an Anglican hymn (a plagal cadence, IV-I). There is nearly a kinesthetic "feel" that the harmony on the "a" must resolve to the chord on "men." The Gestalt principle of closure is closely related to the concept of tonality in melodic and harmonic motion.

Krumhansl (1990a) describes her operational definition of what she calls the *tonal hierarchy*:

In the basic experiments, the elements of interest (tones or chords) are presented singly or in pairs following a musical unit, such as a scale or chord cadence. . . . The listeners are required to judge, using a numerical scale, how related the presented elements are. . . . The rationale for presenting tones in key-defining contexts is that key, or tonality . . . establishes hierarchies of structural significance for tones and chords. (p. 10ff)

This has been termed the *probe-tone* method and has been employed in a large number of studies (see Krumhansl, 1990a, for a complete review). In fact, this method measures the degree of closure. The scale or chord progression provides the context, and the rating provides an estimate of the degree to which the elements of interest close

Figure 4. Sound-list specification using the file selector (left) to obtain the files. Functions for auditioning and editing the list are specified by command buttons at the right of the window.

the pattern. Whether the method measures tonality is the source of the controversy.

Butler (1989, 1990) has suggested that the closure aspect of the method leads to invalid results. He states that "the crux of the matter is that probe-tone rating is a unidimensional measure, applied to a multidimensional stimulus" (Butler, 1989, p. 427). He notes that temporal features, the ordering of the tones in time, strongly influence the probe-tone ratings. In fact, Krumhansl's method deals with Markov chains of 0th order. Butler suggests a 1st-order, interval-based method (a change of stimulus property) modeled in terms of the principle that "any tone will suffice as a perceptual anchor—a tonal center—until it is supplanted by a better rival" (Butler, 1989, p. 429). Unfortunately, the response method he suggests, "vocalizing, whistling, or by using sol feggio syllables," would appear to be open to experimenter interpretation, and although his experiments with probe tones on time-reordered pitch sets clearly demonstrate the existence of problems with Krumhansl's method, he does not offer a substitute.

Although there is, no doubt, considerable musical invalidity in treating tonality as an orthogonal concept to that of time, which is exacerbated by atomization to a 0th order chain, Krumhansl's probe-tone results have produced profiles of pitch-class relations that are modestly confirmed in information-theory analysis of notation

(Krumhansl, 1990a, p. 157; however, no correlation accounts for more than 41% of the variance). Also, her work has included a number of variants of the probe-tone technique, and she acknowledges the role of time in her concept of contextual asymmetry (Krumhansl, 1990a, p. 149). In addition, much of her earlier work using similarity scaling is confirmed using the probe-tone method (Krumhansl, 1990a, p. 148).

What is needed is a carefully constructed methodological study, in which variations of stimulus and response methods are critically compared. We believe that all of these methods, probe tone, use of 0th- and 1st-order chains, and similarity scaling illuminate different aspects of the nature of tonality. Unless multiple methods are conducted in the same chronological and experimental context, it will be difficult to ascertain which procedures are providing the greatest ecological validity. In addition, various variables, such as rhythm, contour, and timbre, must be incorporated in order to determine the interactions between tonality and other musical materials. For example, it should be possible to alter the tonal implications of a sequence of pitch classes by varying the accenting of differing groups. Recently, a graduate student in our laboratory (Michael Frischkopf) applied similarity scaling to paired patterns of differing Markov chain orders with differing frequency distributions of pitch classes. Similarity scaling of patterns varying in pitch, time, and tonal ma-

Figure 5. Word-list specification via the text box at the left of the window. Word lists may be differentials (specified with a dash between words) or any text string.

materials would be a useful extension of previous work; such patterns can also be categorized, rated for semantic attributes, and used in matching procedures. In summary, to illuminate the multidimensional nature of tonality we need not necessarily use "multidimensional response measures," as implied by Butler, above. Instead, multiple experiments and methods that vary stimulus materials (chain order, timbre, pitch-time contour) and response measures (probe tone, similarity, rating scales) will gradually converge on the "truth" about tonality. We turn now to work that has applied such multiple methods.

Musical Communication

Recently, Kendall and Carterette (1990) employed multiple, converging methods in order to investigate the degree to which a performer-initiated expressive message was conveyed to the listener. Their experiments involved the composer, performer, and listener frames of reference as well as the notational and acoustical frames of reference. They state,

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. (Kendall & Carterette, 1990, p. 132)

To investigate the musical communication model, they had a professional pianist produce three categories of expressive performance of a phrase from "Thy Hand, Belinda" by Purcell: without expression (*senza espressione*), with appropriate expression (*con espressione*), and with exaggerated expression (*con troppo espressione*). Oboe, violin, clarinet, and trumpet instrumentalists listened to the piano models and performed their own renditions of the three expressive messages.

In order to assess the degree to which the intended level of expression was communicated to the listener, several methods were used. First, a computer-controlled categorization method (Kendall, 1988, p. 135) was employed: A set of three yellow rectangles, representing the model stimuli, was displayed on the computer screen. The set of wind-instrument stimulus choices ($4 \times 3 = 12$) was represented by uniform white rectangles at the bottom of the screen. All stimuli were randomly ordered. The subject could audition any stimulus by pointing to the rectangle and clicking a mouse button and could move choice stimuli underneath a model stimulus. The task was for the subject to group choice stimuli with the same expressive intent underneath their respective models. The subjects could freely move choice stimuli and audition any stimulus as often as desired. The computer recorded the final configuration, as well as the number of category decision changes, number of auditions, and listening duration.

Confusion matrix data from 19 subjects (10 nonmusicians and 9 musicians) showed correct matches, ranging

Figure 6. Picture-list specification using the file selector (left) to obtain bit-map (.BMP) files. The bit map may contain musical score, as shown.

from 81% (musicians, clarinet) to 43% (nonmusicians, violin). (Note—All results were confirmed with appropriate statistical tests, reported in the original article.) A composite matrix summed over instrument showed 70% correct matches for the musicians and 64% for the nonmusicians. Listening durations, number of times auditioned, and number of category changes indicated that there was a higher cognitive load in discriminating between Expressive Levels 2 and 3 than between Level 1 and the others.

The next method employed was matching, a change in response paradigm with retention of the original stimuli. The task was a three-alternative forced-choice paradigm with subject review. On each trial, 1 of the 12 choice stimuli was heard, followed by all three randomly ordered piano renditions. The subject then selected the piano rendition that best fit the level of expressiveness of the initial example. The subject could replay all stimuli. Results of 10 musicians showed a range of frequency of correct matches from 55% to 57%.

In the categorization paradigm, a subject can review any and all stimuli. Thus, the listener can build a perceptual space. In contrast, matching constrains intercategory comparisons and emphasizes the immediate model and choice set. In terms of necessary and sufficient conditions (discussed above), the intercategory comparison is necessary and sufficient to achieve relatively high hit rates (ca. 70%+). However, since both methods produced the same profile of hit rates that were above chance (33%), both demonstrated the communication of performer intent.

The next convergent method employed involved a change in stimulus. All 15 digitally recorded performances were acoustically analyzed in 100-sample (3.5-msec) windows, producing both timing and amplitude (RMS) data. The timing data were used to drive a musical instrument digital interface (MIDI) synthesizer programmed with emulations of the original instruments, but without time-variant amplitude characteristics. The categorization paradigm was employed with 12 nonmusicians; the “synthetic” stimuli comprised the model and choice sets. Mean correct matches ranged from 33% (violin) to 69% (clarinet). The same profile of responses by instrument category was obtained as in the first method, which employed the natural signals. Thus, we found that timings alone carried enough of the expressive message to be sufficient for reasonable categorization accuracy.

The final method employed was ratings. In the experiment, both natural and artificial stimuli were randomly presented and rated by 8 musician subjects on a scale from 0 (*without*) to 100 (*great*) expressiveness presented on the computer screen. This method has the same lack of intercategory review possessed by matching, and results “produced data parallel to those of matching” (Kendall & Carterette, 1990, p. 159). Thus, the method was neither necessary nor sufficient to produce the relatively high accuracy of categorization; it was, however, sufficient to produce above-chance differentiation of expressiveness for both natural and artificial performances (Kendall & Carterette, 1990, pp. 157-160). Also, the change of re-

Figure 7. A moderately complex experimental research design using play lists and differentials and containing a practice and main experiment.

sponse method to ratings, using the verbal attribute *expressive*, gave a preliminary indication that the “appropriate” levels of expression had higher expressivity than did the “exaggerated” level (Kendall & Carterette, 1990, p. 159), although there was lack of statistical significance due to high variability. This shows the value of collateral methods not only in confirming previous results, but in suggesting the course for additional research—namely, experiments with the use of rating scales. We turn to a series of experiments that culminates in the exploration of this issue.

Timbre

Kendall and Carterette (1991, in press-a, in press-b, in press-c) conducted an extensive investigation using multiple converging methods on the perceptual, verbal, and acoustical attributes of two wind instruments sounding together. The first method (Kendall & Carterette, 1991) involved that of rating timbral similarity among all pairs of saxophone, oboe, flute, trumpet, and clarinet dyads (thus, OC = a dyad of oboe and clarinet). Musical contexts were unisons (B $\flat$ 4), unison melodies (D5, Eb5, F5, and D5), major thirds (B $\flat$ 4–D5), and harmonized melodies (using I, IV $\flat$, V $\flat$, I harmonic progression). Homophonic (harmonized) conditions had each member of the dyad as the soprano (top), resulting in six musical contexts. Music majors ($n = 19$) and nonmusic majors

($n = 16$) rated similarity of all possible pairs of dyads ($N = 10$; total of 55 items for each context) on a 100-point scale presented on the computer screen. Rating was accomplished by moving a red pointer along the scale with a mouse and pressing a mouse button when the pointer was at the desired position. Individual triangles of subject dissimilarity ratings were averaged and subjected to classical multidimensional scaling (MDS), using Kruskal’s method of minimizing stress.

The resulting three-dimensional configurations had stress that ranged from a high of 14.8% for inverted instruments on the harmonized melody for nonmusicians to a low of 2.0% for major thirds rated by musicians. Two of the dimensions were stable and interpretable, with extrema associated as follows: Dimension 1, oboe (nasal) versus trumpet and saxophone; Dimension 2, saxophone (reedy, rich) versus trumpet (brilliant). The third dimension was interpreted as “simple” versus “complex.”

Average dissimilarity triangles were treated as “subjects” in an individual differences scaling (INDSCAL) analysis, producing an amalgamation across contexts. The resulting two-dimensional configuration was circumplacial, with extrema as above.

Kendall and Carterette (1991) note that “one weakness of MDS generally is that, once the scaling is complete, the number and ‘meaning’ of dimensions are subject to interpretation . . . [this] often has been lax” (p. 372). Ad-

Figure 8. Setting the message text property of the message box object.

ditional experiments, using different methods, were conducted to converge on the "meaning" of the dimensions, as well as the fundamental nature of timbral combination.

For example, one question that arose is whether the individual instrument timbres linearly sum to form the composite timbre of the dyad combination. To answer this, a "mental imaging" method was devised. The subject, an expert, was given a space marked with the dimensional interpretation (nasal, rich, brilliant) suggested by the composite MDS two-dimensional configuration. The task of the expert musician subject was to *imagine* the timbre of the *individual* instruments (clarinet, oboe, saxophone, flute, trumpet) and mark their positions, relative to one another, in a space in which the horizontal axis represented "nasal" versus "not nasal" and the vertical axis represented "brilliant" versus "rich." Each of the five instrument points was treated as a vector in two-dimensional space, such that the location of the combination of instruments in the same space was the mean, unweighted vector sum of the coordinates of each of the constituents of the dyad. The dyad space thus mathematically obtained from the imaging experiment could be compared with the space that was based on dyad timbral similarity: Pearson correlations between distances were .938 and .983 for the first and second dimensions, respectively (Kendall & Carterette, 1991, p. 396). By using this method, initial evidence for a linear combination of timbres and evidence of the validity of the verbal descriptions of the scaling dimensions were obtained.

The next method investigated additional aspects of the timbral combination—specifically, the nature of the timbral "fusion" (Kendall & Carterette, in press-a). The degree of "blend" of the two instruments was operationalized in terms of the extent of "oneness" versus "two-ness." The same dyads used in the previous experiments were randomly played for each of 9 musicians whose task was to move a red pointer on a 100-point scale to the position that best described his feeling of oneness or two-ness. Results indicated that blend ratings separated instruments along the "nasal" versus "not nasal" axis; "nasal" combinations blended less well than did those that were "brilliant" or "rich."

An additional method was employed, designed to determine a subject's ability explicitly to identify the two instruments sounding. Eight musicians were presented with randomized soundings of the 55 dyads in six musical context blocks. The subject then selected from a list presented on a computer screen, the correct pair of instruments then sounding from the 10 possibilities. As might be expected, a general inverse relationship was found between identification and blend; the greater the blend, the worse the identification (for unisons, Pearson $r = -.731$).

Next, blend ratings were decomposed into a triangle in which the blend rating for OF was an entry in row O column F. The triangular matrix was analyzed using Kruskal's MDS. This produced a two-dimensional space (stress < .00031) for the five individual instruments. These individual points were combined by vector summing (see above) to produce a hypothetical dyad space based on blend. This blend space was then correlated with the composite similarity space; correlations were .893 for Dimension 1 unisons and .843 for Dimension 2 unisons, showing "a remarkable convergence between differing methods and models" (Kendall & Carterette, in press-a).

Figure 9. Play-list stimulus object properties (see text).

Figure 10. Differential response object properties.

All of these results demonstrate how rating, identification, imaging, matching, and similarity-scaling methods converge on the truth in the study of timbral combinations. The following experiments required analysis of a single method: the semantic differential. The experiments were conducted to attach specific appropriate verbal labels to the timbral similarity scaling dimensional axes.

Kendall and Carterette (in press-b) used traditional semantic differential scaling on differentials previously employed by von Bismarck (1974a, 1974b), which they se-

lected for their high factor loadings. Eight adjectives were used to rate the 10 dyad unison timbres used in previous investigations (hard-soft, sharp-dull, loud-soft, complex-simple, compact-scattered, pure-mixed, dim-brilliant, heavy-light). The subject judged the verbal attribute of the timbre by using the mouse to position a pointer on a 100-point scale. The eight scales were presented and the ratings were collected before the next dyad was played. This method differs from paper-based procedures where the subject has all of the scales and responses visually before him. In the computer-based procedure, each scale is randomly presented and erased from the screen after the response.

Results clearly demonstrated the failure of this method to discriminate among the 10 unison dyad timbres. Kendall and Carterette (in press-b) state that "it occurred to us that a central problem in the use of bi-polar opposites is that the 'opposite' is, in fact, not an antipode." They changed the method to one termed *Verbal Attribute Magnitude Estimation* (VAME). In this method, an adjective is paired with its negation (e.g., "sharp" vs. "not sharp"). The same procedure was used, except that the timbre was sounded before every scale. Thus, the subjects rated the degree of "sharpness," not the degree of "sharp" versus "dull."

Results ($N = 11$) demonstrated "a far greater degree of differentiation of ratings by instrument" (Kendall & Carterette, in press-b) and were statistically significant for the dyad $\times$ scale interaction in the analysis of variance (ANOVA) for the VAME method; statistical significance was lacking for the semantic differential method. Thus, the semantic differential method was neither necessary nor sufficient to yield interpretable results, but the VAME method was both necessary and sufficient to yield interpretable results.

Figure 11. Checking an experiment results in this display, where a message indicating the status of each object is shown. During this process, memory for the experiment is allocated, and structural aspects of the experiment are resolved.

Figure 12. The subject-response screen for differential response. The example number at the top of the window (Example 3b) indicates which stimulus has been sounded. In this case, pair comparisons are being presented, since the suffix "a" versus "b" is used to indicate an element of the pair.

Numerous additional computational transformations of the ratings data, including assigning to a dyad a vector of attributes and then applying MDS, yielded a space of verbal attributes. The configuration of dyads based on VAME ratings mapped weakly into the composite timbral similarity circumplex, however. An additional experiment (Kendall & Carterette, in press-c) was conducted using adjectives derived from Piston's (1955) *Orchestration*, in order to increase potentially the ecological validity of the scales. In addition, a rating scale of 0 to 100, rather than the negation of the attribute, was used. Using principal-components analysis, cluster analysis, MDS, and attribute vector analysis, a superior and interpretable mapping of word to sound was obtained.

In all of these studies, the same stimuli were subjected to different methods designed to converge on an interpretation of the timbral similarity circumplex. If it were necessary to program a method from scratch, such method variations would be prohibitively time-consuming to produce. An additional attribute of these studies is the flexible manipulation and analysis of data. Entering data into the spreadsheet format of most statistical programs by hand would have been prohibitively costly. We turn now to the means by which these studies were conducted.

THE MUSICAL EXPERIMENT DEVELOPMENT SYSTEM

Previously, Kendall (1988) presented a computer-based system for psychomusical research based on direct-to-disk sampling of natural-instrument musical signals. This system was used to conduct the experiments in musical communication and timbre described above. The system also provided data manipulation routines that interfaced to Systat 5.0.

However, even greater flexibility in experimental design is desired than that afforded by a menu-driven system, in which one is at the mercy of the available menu choices, no matter how comprehensive in scope. In addition, recent developments in sampling technology, catalyzed by the commercial use of digital editing in the recording and film industries, supersede traditional analog-to-digital (A/D) and digital-to-analog (D/A) converters attached directly to the computer system. The Musical Experiment Development System (MEDS) was developed to address these issues.

Hardware and Software Platform

The MEDS was designed to run under Windows 3.0 on an IBM-type 80286 or on a higher computer system hav-

Figure 13. The picture-list response object. In this example, the subject must indicate which chord voicing was just heard.

ing 2 MB of RAM and any type of large hard drive. Software development was with Microsoft Visual BASIC, with assembler code for MIDI functions and precise timing written in in-line assembler using Borland's Turbo Pascal for Windows. MIDI, timing, and sample-to-disk functions reside in dynamic link libraries (DLL), which are independent code modules with common interfaces to all Windows 3.0 programs. In this manner, additional hardware resources can be effortlessly added to MEDS; the appropriate DLL driver is the main requirement.

The principal signal manipulating device in MEDS is the Micro Technology Unlimited (Wind Chime Court, P.O. Box 21601, Raleigh, NC 27619) Microsound system. This system provides an interface to the PC bus with a Motorola 5001 digital-signal-processing chip and an external input/output (I/O) subsystem containing up to four channels of A/D and D/A and two channels of direct digital I/O in either AES-EBU or SPDIF format. Recording and digital editing of source signals are accomplished by using the Microsound MicroEditor software (Figure 1); a DLL for playback under Windows 3.0 is used in MEDS. Signals are stored in 16-bit PCM format, stereo or mono.

In MEDS, signals are recorded on digital audiotape (DAT) using any DAT machine. Instead of resampling signals, the direct digital I/O is connected from the DAT machine to the Microsound I/O subsystem. The Microsound system interprets the DAT format and synchronizes to its sampling rate of 32,000, 44,100, or 48,000 samples per second per channel, recording the digitized signal onto a standard hard-disk file. Using this system, the originally digitized signal remains in digital form until it is played back during an experiment. Playback is through an amplifier with digital inputs and outputs or through a DAT machine connected to a standard amplifier. MEDS also has a custom DLL for rudimentary, four-voice polyphonic MIDI sequencing with a Roland MPU-401-type interface to a synthesizer or sampling keyboard.

Experiment Design Software

The graphical, icon-oriented nature and multitasking capabilities of Windows are exploited in MEDS. Figure 2 shows the main experiment-development window. Two sets of icons at the bottom of the screen form the stimulus and response objects of an experiment. Stimulus objects

Figure 14. The word-list response object, used in this example to request identification of the sounding orchestral instrument.

Figure 15. The data editor, showing the source experiment-data file to the left and the transformed, rectangular data file to the right.

have properties that present and order stimuli and also provide timers, messages, and input of subject and grouping informational text. Response objects have properties that control the subject-response method, such as rating scales, picture (image) boxes, differentials, word lists, and response sound lists. The blank rectangles above the icons hold the objects specified in the experiment (see below). The rectangle above these displays the properties of a selected (mouse-clicked) object. At the top of the window are the menus for additional functions. We now examine the procedure for creating an experiment in some detail.

Sound is the primary stimulus (although even response objects can become stimulus objects), and recorded signals, synthesized signals (using CSOUND), and MIDI sequences may be used as properties of objects. Digitally recorded stimuli often must be adjusted for equal durations, or they must be adjusted to remove unwanted silence or noise from the start and end. The MEDS audio search function, accessed from the Signal menu, provides the means for making this simple edit (Figure 3). The left signal window in Figure 3 is 500 samples in length from the start position set by the top sliding control. The right signal window is 500 samples forward from the position of the bottom sliding control. Adjusting the vertical sliders on each window changes the amplitude scale; adjusting

the horizontal sliders underneath the window changes the length of signal presented in the window. When the "Off" button control is pressed (Figure 3, bottom left), it toggles "On," and the signal from the start to the end position is played back in a continuous loop. Position sliders (Figure 3, above the control buttons) can be moved until the signal is aurally or visually free from unwanted elements. Pressing the "Save Mark" button (Figure 3, bottom right) saves the marked position in the header of the signal file.

MEDS uses lists of signals, pictures, and words to specify the properties of objects. The Lists menu choice (Figure 2) provides windows for creating lists. Figure 4 shows the window for creating sound lists. The left side provides the file selector system. A file name is selected (as shown, example3.sf) and appears as the next element in the list box immediately to its right. The "Include Markers" radio button (set for "No") determines whether the audio search function marked segment is included in the list. The buttons below this provide flexible means of editing the list and listening to one or more items in the list. Items can be inserted into the list by simply clicking the list entry above which the insertion is to take place and selecting additional files. Figures 5 and 6 show similar windows for creating word lists (which may contain any string of

Figure 16. The data editor, showing a triangular data transformation.

characters) and lists of pictures, which are bit-mapped (.BMP) images, including musical scores from programs such as FINALE.

Once the signals are recorded, transferred to hard disk, and edited, and lists are created, the task of designing an experiment is to order the stimulus and response objects and to set their properties. Ordering of objects takes place in the rectangular outlines above the icons (Figure 2). Clicking an icon places that object in the next available blank box. Experiment execution is from left to right, top to bottom.

Figure 7 shows a moderately complex design for an experiment involving differentials, and it demonstrates many of MEDS' important characteristics. The first object in the experiment is a message box. This icon is dark because it has been selected with a single mouse click, and its properties (a message that is blank) appears above it. Figure 8 shows the window that appears if the message object is double clicked, requesting the properties window for that object. A message concerning the title of the experiment and some instructions about preparation for the experiment is written. The window has all of the functions of a word processor, including word wrap. When the window is closed, the message is added to the message object and will be displayed in a window at the start of the experiment as the first event of the experiment.

The second object (Figure 7) in the experiment is the input-data object, which has properties for a message about the type of data to provide (e.g., a subject's or group's identifier). When the experiment is run, this object will

create a window requesting the data; the provided information is then written to the experiment-data file.

The third object (Figure 7) is a start repeat, which is always paired with an end repeat (Object 7), both of which enclose the object(s) in a loop (the repeat objects are similar to FOR-NEXT or DO-LOOP in other programming languages). As can be seen, the start-end repeat structures can be nested; they may also be chained. The number of times the repeat takes place is determined automatically as the count of the list(s) in the enclosed object(s), or it can be set manually to any value. The fourth object is a stimulus play list, the properties window of which is shown in Figure 9. Figure 9 indicates a list called TEST.SLT (selecting the "Change" button brings up a file-selection window). The TEST.SLT list has three entries; however, because the paired-comparisons property is selected, the three entries become six pairs (listed after the heading "Total Count"). The "Random" radio button is selected, indicating that the stimuli should be randomly presented. The "Time" box shows a 0.5-sec pause between the pairs of stimuli in the pair comparisons. These properties are saved in the play-list object (#4).

The fifth object is another start repeat that, with its end repeat, encloses another list-oriented object, the differentials object (#7). Figure 10 shows the properties window of this object. TEST.WLT is the name of the word list that contains three differentials (i.e., not good-good). The other properties are the type of response scale "scroll" bar (horizontal or vertical) and whether or not the differentials are randomized.

Figure 17. The complete software development environment during the creation of the audio search function.

The eighth and ninth objects are the end repeats for the repeat structure. With this structure, a randomly selected play-list item will sound, followed by all three randomly ordered differentials; this sequence would be repeated for the remaining play-list stimuli.

Row 2 of the design begins with a message object and is followed by a nearly identical object structure to row 1. This is because the first row of objects defines a practice session, and the second row the main experiment. In the main experiment, there is a pause defined by a timer object after the differentials are rated.

After an experiment is designed, it is checked (an option under the Run menu). Figure 11 shows the result of checking the experiment defined in Figure 7. If there were stray repeats, lists not specified, or structural inconsistencies, a fatal error would be generated; other less serious omissions are flagged as warnings. The experiment, once checked, is saved using the File menu. Figures 12, 13, and 14 show the appearance of various response objects during the execution of an experiment.

Altering an existing experiment is one of the most powerful features of MEDS. Figure 7 shows editing buttons to the top right of the window. The "Insert" button

opens an object position to the left of a selected object. "Delete" removes a selected object, and "Consolidate" removes all blank rectangles in the middle of a design, moving objects to the top left, filling in blank space. With these functions, editing of the experiment is fast and easy. For example, the design in Figure 7 can be changed by selecting the play-list icons (headphones), deleting them, and clicking on the play-one icon (musical staff) under "Stimulus Objects." This object would replace the two play lists. Since the play-one object has a count of one, the outer repeat structures are meaningless, and would be ignored. Alternatively, they could be deleted, and "Consolidate" could be employed.

Data Collection and Analysis

After a subject has been run in the experiment, the response data are saved in an ASCII file. Figure 15 shows the window for experiment-data manipulation. The raw response data appear in the left text box; the transformation data appear in the right text box. Transformations can be as simple as removing the data type identifiers (DATA START, DATA END) or can involve rectangular or triangular formatting, which is controlled by the "For-

mat" radio button and "Columns" setting. Figure 16, for example, shows the transformation of the data set to a triangular matrix. The "Skip Data Sets" setting permits data to be partitioned into subsets. For example, the practice session data can be removed and formatted independently of the main experiment. Functions that do not appear in the illustrations include data set rotation and averaging. Both text boxes in Figures 15 and 16 provide word-processor-type editing functions, permitting data errors to be corrected or new data to be added.

From the beginning, the authors have avoided reinventing the wheel. Data analysis is a case in point, where comprehensive statistical analysis programs abound. At this writing, no statistical analysis program for Windows 3.0 is available. The Analyze function under the Data main menu (Figure 2) provides functions for taking the output of the data editor and preparing command files for Systat 5.0 running in a DOS window.

Of greater utility is the Dynamic Data Exchange (DDE) capability of Windows. This protocol establishes a software mechanism for applications to send data and commands to one another. MEDS uses DDE to make Microsoft EXCEL, a spreadsheet and data graphing program, operate as a client. MEDS has functions that activate EXCEL, send data to the spreadsheet, perform statistical analyses, and graph the results, all from within the MEDS environment. In the future, a statistical program, containing a comprehensive suite of advanced procedures and graphing options, will become available for Windows; with DDE, it will become part of MEDS as well.

Future Directions

MEDS itself does not contain program logic for making new objects. However, the Visual BASIC environment is so simple to use and ideal for this purpose, it precludes such a need. To illustrate this, Figure 17 shows an actual session during which the Audio Search function was designed. At the bottom of the figure, the actual window as it will appear while running is being designed. Objects from the Tool Box window (to the left) are simply clicked and then placed and sized on the design window using a mouse. Properties, such as font type, font size, or caption text, are added from the window at the very top of the figure. Beneath the properties window is a color window; to the right of that is a window containing all of the components of the program. Beneath the color window is the code window, containing the code for the second vertical scroll bar (VScroll2); simply clicking on an object in the design window brings up the code associated with that object. Thus, a Windows program in Visual BASIC is very similar to Hypercard, except that a very complete programming language, rather than a limited script language, is available. To create a new object

for MEDS, one need simply design a new window and return data in a global variable.

We have described Version 1.0 of MEDS. In the immediate future, a package of signal-analysis routines, including fast Fourier transforms, filters, and synthesis modules, will be incorporated into MEDS. Additional stimulus and response objects will be added for laser and compact disc control under the Windows Multimedia protocol. Finally, the structural logic of MEDS will be expanded to include parallel experiment processes and decision-making objects.

REFERENCES

- BUTLER, D. (1989). Mapping the listener's recognition of tonality: A delineation of the theory of intervallic rivalry. In *Proceedings of the First International Conference on Music Perception and Cognition* (pp. 425-430). Kyoto: Japanese Society of Music Perception and Cognition.
- BUTLER, D. (1990). Response to Carol Krumhansl. *Music Perception*, 7, 325.
- KENDALL, R. A. (1984). *The role of transients in listener categorization of musical instruments: An investigation using digitally recorded and edited musical phrases*. Doctoral dissertation, University of Connecticut. (University Microfilms No. 8225754).
- KENDALL, R. A. (1986). The role of acoustic signal partitions in listener categorization of musical phrases. *Music Perception*, 4, 185-214.
- KENDALL, R. A. (1987). Model-building in music cognition and artificial intelligence. In *Proceedings of the First Artificial Intelligence & Advanced Computer Technology Conference/East* (pp. 183-196). Wheaton, IL: Tower Conference Management.
- KENDALL, R. A. (1988). A sample-to-disk system for psychomusical research. *Behavior Research Methods, Instruments, & Computers*, 20, 129-136.
- KENDALL, R. A., & CARTERETTE, E. C. (1990). The communication of musical expression. *Music Perception*, 8, 129-162.
- KENDALL, R. A., & CARTERETTE, E. C. (1991). Perceptual scaling of simultaneous wind instrument timbres. *Music Perception*, 8, 369-404.
- KENDALL, R. A., & CARTERETTE, E. C. (in press-a). Identification and blend of timbres as a basis for orchestration. *Contemporary Music Review*.
- KENDALL, R. A., & CARTERETTE, E. C. (in press-b). Verbal attributes of simultaneous wind instrument timbres: Part I. von Bismarck's Adjectives. *Music Perception*.
- KENDALL, R. A., & CARTERETTE, E. C. (in press-c). Verbal attributes of simultaneous wind instrument timbres: Part II. Adjectives induced from Piston's *Orchestration*. *Music Perception*.
- KRUMHANSL, C. (1990a). *Cognitive foundations of musical pitch*. Oxford: Oxford University Press.
- KRUMHANSL, C. (1990b). Tonal hierarchies and rare intervals in music cognition. *Music Perception*, 7, 309.
- MEYER, L. (1956). *Emotion and meaning in music*. Chicago: University of Chicago Press.
- MURSELL, J. (1937). *Psychology of music*. New York: W. W. Norton.
- PISTON, W. (1955). *Orchestration*. New York: W. W. Norton.
- SEASHORE, C. E. (1938). *Psychology of music*. New York: McGraw-Hill.
- VON BISMARCK, G. (1974a). Timbre of steady sounds: A factorial investigation of its verbal attributes. *Acustica*, 30, 146-159.
- VON BISMARCK, G. (1974b). Sharpness as an attribute of the timbre of steady sounds. *Acustica*, 30, 159-192.