CENTER FOR COMPUTER RESEARCH IN MUSIC AND ACOUSTICS
JUNE 1991

Department of Music
Report No. STAN-M-71

RECOGNITION OF COMPLEX AUDITORY-SPATIAL PATTERNS

Stephen Lakatos

CCRMA
DEPARTMENT OF MUSIC
Stanford University
Stanford, California 94305

(to be published in a forthcoming issue of *Perception)*

© copyright 1991 by Stephen Lakatos
All Rights Reserved

Abstract

Two experiments were carried out to investigate the perception of complex auditory-spatial patterns. Subjects were asked to identify alphanumeric characters whose patterns could be outlined acoustically through the sequential activation of specific units in a speaker array. Signal bandwidths were varied systematically in both experiments. Signals in Experiment I had sharp onsets and offsets; envelope shapes in Experiment II were much more gradual. Subjects showed considerable ability in recognizing alphanumeric patterns traced with signals of varying acoustical composition. Reductions in the steepness of signal attack and decay produced limited declines in pattern recognition ability. Systematic trends in the relation between patterns and the distribution of incorrect responses suggest that subjects performed a pattern-matching task, in which identifications were made on the basis of component features The unexpected pattern recognition abilites that subjects demonstrated in both experiments suggest that spatial hearing, like vision, has access to mechanisms for generalized, multi-modal spatial representations.

Introduction

Auditory localization has been the focus of extensive research over the past century. Such research has demonstrated convincingly that localization is a function of both interaural time and amplitude differences, as well as monaural and binaural spectral cues. The vast majority of work on binaural hearing has been conducted from a psychophysical perspective. Unfortunately, emphasis on the effects of specific acoustical variables upon localization acuity has resulted in a neglect of more complex aspects of spatial hearing. Studies have dealt almost exclusively with static localization, with the aim of setting minimum conditions for the detection of changes in sound source location (e.g. Mills 1958). Only a handful of studies treat dynamic localization (Harris and Sergeant 1971; Harris 1972; Perrott and Musicant 1977), and these do not explore its aspects beyond standard measures of localization acuity (i.e., minimum audible angle). Since sensory systems seem especially sensitive to dynamic input (Gibson 1979) and since recognition of source movement is a necessary component of many auditory systems, it would seem prudent to gain a better understand the dynamic components of auditory localization. A strict functional distinction between horizontal and vertical localization has also been amplified in the literature. Since source elevation is derived from pinna-induced spectral filtering (Gardner 1973; Butler & Belendiuk 1977), while positions having nonzero azimuth angles are detected through interaural time differences and interaural amplitude differences (Mills 1972), as well as directional filtering by the external ears (Musicant & Butler 1985), localization mechanisms for the median and horizontal plane have been treated as separate phenomena. Although this distinction may be valid from a purely psychophysical perspective, practical localization situations are seldom limited to unique planes and there is no qualitative change in our sense of auditory space as localization conditions shift between horizontal and vertical.

Perhaps even more problematic is the fact that cognitive aspects of auditory localization have been overlooked almost entirely. Given recent neurophysiological evidence for extensive topographical coding of spatial information within the auditory systems of birds (Knudsen 1984) and mammals (Palmer & King 1982), it seems surprising that little consideration has been given in directional hearing research to possible cognitive representations of spatial information in humans. Spatial representation of visual information, on the other hand, has received considerable treatment. An extensive series of studies by Shepard and his colleagues, for example, has demonstrated that mental transformations of visual objects are analogical in nature and obey the constraints of kinematic geometry (Shepard & Metzler 1971; Cooper & Shepard 1973a, 1973b; Shepard and Cooper 1982). Studies of visual apparent motion (Corbin 1942; Farrell & Shepard 1981; Shepard and Zare 1983) and representational momentum (Freyd and Finke 1984) also indicate that prevailing physical constraints and tendencies in the external world are internalized to the degree that spatial relations in mental space are directly represented by the organization of the representation.

It remains an intriguing question whether such internalized constraints apply to spatial hearing. Positive findings would lend further support to the view that an analogical representation of spatial information is not restricted to the visual domain, but is rather a general property of all spatial representations, regardless of sensory modality. Research on tactile perception has indicated that many representational constraints found in vision apply to touch as well (Carpenter & Eisenberg 1978; Heller 1989). Rhodes (1987) has been the only one to address this issue directly from an auditory standpoint, by demonstrating that shifts of auditory attention during localization tasks are constrained by the spatial separation between consecutive auditory stimuli. Although the temptation to view auditory localization as one component of an integrative, multi-modal representational machinery is great, there is no *a priori* reason for such an assumption. Spatial information is not mapped topographically at the basilar membrane, as it is at the retinal and skin surface; rather, it must be computed at higher levels of the auditory system. Moreover, while the auditory system appears quite adept at tracking the spatial position and trajectory of moving sounds, it is unclear whether or not this ability requires the formation of a mental representation akin to that of visual images (In a similar vein, the spatial layout of senory receptors in both vision and touch does not necessary imply that similar representational frameworks exist, for example, for physical shape, although this appears to be the case [Garbin & Bernstein 1984]). Clearly, research into potential auditory-spatial representations along such lines would be useful for better understanding spatial representations, whether or not they are unique to spatial hearing.

Two closely associated goals were set for the current study. First, it was hoped that previous experimental limitations could be bypassed to a degree by designing an approach in which more complex, integrative aspects of spatial hearing could be examined and, second, an acoustical medium was sought in which the validity of potential auditory-spatial representations could be tested in future research. For such purposes, a plan that included both a dynamic setting and a complex localization field, not restricted to either the horizontal or vertical plane, was deemed essential. Initial designs centering on recognition tasks, similar to those in visual pattern recognition studies, in which subjects would have to identify some auditory analog of a spatial "pattern", seemed the most suitable means for examining spatial hearing under these conditions. An obvious difficulty with such a proposal centered on what form such an auditory-spatial "pattern" might take; spatial patterns _per se_ are synonymous with vision, not spatial hearing. Even the simplest visual analogs would present logistical problems in an experimental setting. For one thing, patterns would have to be sequential in nature, since directional hearing has been shown to be ineffective in the discrimination of several simultaneous source locations [see Perrott (1984) for some of the difficulties of simultaneous localization with even two sources]. This limitation might be overcome with some type of tracing, in which the form of a pattern is outlined with sound. A further concern was that recognition of such abstract patterns would be a highly unusual task for subjects, requiring extensive training before satifactory performance levels could be obtained. Extensive learning periods might also obscure certain localization characteristics, through the creation of additional, artificial aids towards pattern recognition.

Such extended training might not be needed, however, if patterns have some visual concomitant. If, for instance, the auditory-spatial pattern resembles a visual one familiar to all subjects, then satifactory recognition proficiency may not require substantial practice, since subjects would presumably have a "mental representation" of the visual image that could be utilized in some type of cross-modal pattern comparison. This would be more likely if the corresponding visual pattern were highly familiar, such as an alphanumeric character; letters and numbers would be universally recognizable, and their patterns could conceivably be traced using a spatial arrangement of loudspeakers. Therefore, an unusual group of stimuli was adopted for this exploratory study of complex auditory localization and spatial representation: a set of alphanumeric characters whose patterns could be traced with sound. It was hoped that such patterns would provide some preliminary insight into auditory representations of spatial information, and would possess sufficient complexity to permit examination of complex localization under various conditions.

Methods

General Description

In two experiments, subjects seated before an array of speakers were asked to identify alphanumeric characters whose patterns could be traced acoustically by sequentially activating of specific speakers to form the desired patterns. Responses were selected from among a fixed, predetermined set of characters. Bandwidth was varied systematically as one of the variables of interest. Signal characteristics were identical in both experiments, except that envelope shapes in Experiment I had sharp onsets and offsets, and those in Experiment II were much more gradual. Patterns and bandwidths were combined in random fashion, and all such combinations occured an equal number of times for each subject.

Speaker Array Design

Sixteen small 4 1/2" speakers were mounted on a large wooden panel structure in an array configuration that permitted the tracing of several alphanumeric patterns (Figure 1A). Horizontally and vertically adjacent speakers were separated by a distance of 1'. With the subject's head located 6.5' from the midpoint of the array, angles between maximum horizontal and vertical speaker positions subtended 26^{o} and 50^{o}, respectively.

The panel structure served as a baffle for the speaker array. Rubber grommets were placed between individual speakers and panel to increase acoustic isolation between speakers. Acoustical foam panels were placed above and below the forward portion of the panel to minimize reflections from ceiling and floor; in addition, the entire panel structure was mounted on rubber cushions. The speaker array was located in the recording studio at the Center for Computer Research in Music and Acoustics. The room is satisfactorily isolated acoustically, and although not anechoic, it can be made relatively so by exposing absorptive panels located on all four sides. For the experiments, panels were positioned for maximum sound absorption.

Description of Patterns

Acoustic images of alphanumeric patterns were generated by outlining their shapes with sequential pulses at appropriate speakers. The resulting subject impression is that of a rapidly moving, pulsating sound source tracing the form of each pattern. Each pulse lasted 60 msec, with an additional 60 msec pause between consecutive pulses, resulting in a spacing of 120 msec between pulse onsets.

Characters were selected according to the following conditions: (i) each character could be presented with the array configuration described above; (ii) each character could be traced without any crossings or breaks; (iii) all characters could be traced with approximately the same number of speakers, and; (iv) all characters were of reasonably equivalent familiarity. A set of ten characters was chosen based on these constraints (Figure 1B).

Characters were similar, though not identical in the number of speakers required for tracing: C (11 speakers); P, S, U (12 speakers); 3 (13 speakers); 6, 9, G, O, R (14 speakers). Preliminary efforts at equalizing speaker requirements among all characters by repeating pulses at select speakers for those patterns composed of fewer than fourteen speaker locations (e.g. sounding two pulses at both the beginning and ending speakers in the letter U) proved distracting for pattern perception in pilot studies. Existing discrepancies were therefore maintained, with the assumption that they would not furnish a substantial cue in recognition tasks. In actual experiments, the starting point of each pattern trace was indicated to the subject, furnishing an additional identification cue.

Sound Generation

Sound pulses were created with four digital synthesizers (Yamaha TX802 FM Tone Generators) operating under the control of a microcomputer (Macintosh Plus). The synthesizers were programmed to provide 16 separate audio channels, one for each speaker. Each channel had 12 independent oscillators, with variable pitch, amplitude, and envelope control, permitting the generation of complex spectra with a maximum of 12 partials. Individual spectra could be stored in synthesizer memory banks and activated by specific computer commands. In addition, the computer could initiate speaker patterns by triggering consecutive synthesizer channels.

Communications from computer to synthesizer were accomplished through a MIDI (Musical Instrument Digital Interface) link (Opcode Systems). Software used to control both spectra and speaker activation patterns during experiments was written in Lisp (Lelisp - Act Informatique); a MIDI driver compatible with the Lisp software handled MIDI message transfers.

Characteristics of Sound Spectra

Seven different spectra were generated for the experiments. Each spectrum corresponded to a low- or high-pass filtered harmonic complex having a fundamental frequency of 1 kHz, but no actual filtering was performed; rather, a filtering effect was simulated by eliminating partials whose frequencies were higher or lower than certain absolute cutoff points. In this manner, all steady-state frequency components above or below certain values were removed, but any high-frequency transients resulting from signal onset and offset were preserved.

The first spectrum contained the full complement of twelve partials:

spectrum 1 --> partials 1-12

The following three spectra approximated increasing degrees of low-pass filtering:

spectrum 2 --> partials 1-8
spectrum 3 --> partials 1-6
spectrum 4 --> partials 1-4

In similar fashion, the remaining three simulated various high-pass filterings:

spectrum 5 --> partials 9-12
spectrum 6 --> partials 7-12
spectrum 7 --> partials 5-12

A comfortable listening level of 50 dB SPL was chosen for the full 12-partial complex. Amplitudes for the other spectra were adjusted by the experimenter, seated at the subject's listening position, in loudness matching tasks until all spectra were of approximately equal loudness (all speakers had previously been matched for equal acoustic output for a given MIDI input).

The same seven spectra were used in both experiments, except that a square-shaped envelope, with steep onset and offset, was imposed upon pulses in Experiment I, while those in Experiment II received a much smoother, exponentially rising and falling envelope shape (see Figure 2A, B). Spectra in Experiment II were matched in loudness to the 12-partial complex of Experiment I. Overall signal duration remained approximately constant at 60 msec in both experiments.

Experimental Procedure

Two groups of ten subjects (14 men, 6 women) were sampled from a group of researchers at CCRMA and from graduate students belonging to either the Stanford Music Department or the Neuroscience Division of Stanford Medical School. Most were musically sophisticated, although musical ability was not considered to be of any advantage in this study. None reported any hearing problems.

Prior to experimental trials, subjects were provided with an opportunity to associate the identity of each character with its spatial pattern. With room lights turned off, the experimenter called out the name of each letter and then repeated the corresponding pattern until the subject expressed confidence as to its general perceptual characteristics. After scrutinizing each pattern, subjects were allowed to compare any pairs or groups of characters which might still remain unclear. No minimum performance levels were sought in this warm-up period; its sole purpose was to acquaint subjects with pattern characteristics.

In Experiments I and II, no advance information about the pattern's identity was provided. The only difference beteen the two experiments lay in their contrasting envelope shapes for signals. Subjects were seated in the dark during both experiments. For each trial, the subject was presented with a specific pattern using one of the seven signal spectra. The task was to identify the pattern via a response board located to his or her left. The response board had a set of ten buttons, each pattern appearing as in Figure 1B above a button. Before entering a response, the subject had the option of repeating the pattern once by means of a "repeat" button on the response board. A subsequent pattern followed when the subject struck the key corresponding to his or her response a second time. Response rate was self-paced; subjects could pause indefinitely after initial pattern presentation, pattern repetition, or response selection. Responses were transmitted from response board to computer as MIDI messages, and were stored in individual subject files.

Trial blocks consisted of 70 character-spectrum combinations. Each of the ten characters appeared with each of the seven spectra once per block. Character-spectrum combinations were distributed randomly within blocks according to this restriction. Each subject completed four blocks of trials; short rest periods separated consecutive blocks.

Results

Experiment 1

The effects of different spectra on pattern discrimination ability were well above the chance level for any pair of spectra, except those comprising partials 1-8 and 5-12, or the comparison of 1-6 with 7-12. The main differences among the spectra are shown in Figure 3A. Here the means are pooled across subjects and patterns and are expressed as a percentage of correct responses. Each bar is based on 400 trials. There was a significant decrease in recognition ability with consistent removal of either upper or lower partials. Particularly striking is the 25.5% drop in recognition from patterns comprising partials 7-12 to those composed of 9-12. Even in the worst situation (spectrum with partials 9-12), however, recognition ability was well above the chance level (10%), and indicated that localization cues were still sufficient for limited pattern recognition.

Examination of individual subject performances with regard to spectra revealed significant differences as well. These are illustrated in Table 1A, where subject-spectrum values are pooled across patterns and given as a percentage of correct responses. Two-way analysis of variance of subject-spectrum values in Table 1A revealed highly significant differences among spectrum means ($F(6,54) = 21.44$, $p < .001$). This confirms that differences in discrimination ability among spectra were consistent across subjects, and that subjects' overall performance across averaged over all seven spectra varied considerably. Two of the individuals originally tested could not identify patterns correctly; recognition proficiencies for both were near chance level for all spectra. Since these two subjects differed so markedly from others in terms of their discrimination capabilities, the scores of both were considered statistical outliers and omitted from result calculations.

A comparison of performances for individual patterns and spectra indicated similar trends. Table 2A presents the percentage of correct pattern identifications pooled across subjects as a function of pattern and spectrum. Highly significant F-ratios were found for both pattern means ($F(9,54) = 9.16$, $p < .001$) and spectrum means ($F(6,54) = 26.12$, $p < .001$), indicating that individual patterns differed in their overall salience, while simultaneously exhibiting statistically reliable changes in recognizability as a function of spectrum.

A different view of pattern discrimination techniques used by subjects emerges when identification errors are inspected. Table 3A displays the distribution of erroneous identifications among remaining patterns, pooled across subjects and spectra, as a function of the actual presented pattern. Values are expressed in both graphical and numerical terms as percentages of the total number of erroneous responses for a specfic pattern. It is clear that for a given pattern, certain patterns were chosen in error more frequently than others. The number of incorrect responses varies considerably among patterns, and is quite large in a few instances. For example, when G was incorrectly identified, it was mistaken

81.0% of the time as an O, and suggested that, although G was generally quite differentiable from patterns other than O, G and O were confused often. Interestly, however, O was not confused as often with G as in the reverse case, perhaps because of a greater tendency to select the simpler pattern O in uncertain situations. Similarly, although 3 and 9, with similar shapes and identical ending points, were confused quite often (31.3% among incorrect responses for 3), 3 and C were never mistaken for each other. In the case of other patterns, error distribution was much more even. The letter C, when identified incorrectly, was confused with the remaining patterns quite uniformly, with a slightly greater percentage for those patterns that had notably similar outlines (e.g. 6, G, and S). In general, the error distributions in Table 3A seem to suggest some form of pattern identification process based on a variety of pattern cues, such as overall similarity of shape, identical partial components (e.g. the lower right-to-left sweep of both 9 and S) and beginning and ending points. These points will be further amplified in the discussion section.

No significant increase in recognition proficiency, pooled across subjects and patterns, was found for the four consecutive blocks composing the experimental sessions. Subjects, on the whole, did not improve in their ability to identify patterns as the experiment progressed. This finding, along with the equal distribution of spectrum/pattern combinations in the four blocks, assured that that the learning of cues associated with specific pattern did not factor in the study.

Experiment 2

With the substitution of an exponentially rising and falling envelope function, pattern discrimination ability appeared to decrease, although not markedly for all spectra, with the exception of the signals consisting of partials 9-12. Overall subject ability to recognize patterns presented with partials 9-12 remained statistically unchanged. Correct response percentages organized according to spectra are shown in Figure 3B. The proportions among the different spectra found in Figure 3A were essentially preserved (excluding partials 9-12), with a concomitant reduction in correct identifications ranging from 6.3% to 13.0%. Again, even the poorest response percentages (partials 1-4 and 9-12) were above the chance level.

Table 1B shows the percentage of correct responses for individual subject-spectrum combinations pooled across patterns. As in Table 1A, spectrum differences ($F(6,54) = 24.47$, $p < .001$) were highly significant. One individual was unable to identify patterns above chance level; his scores were not considered in any overall results. Similar significance levels were found for the pattern-spectrum values in Table 2B: Pattern means ($F(9,54) = 3.99$, $p < .001$) and spectrum means ($F(6,54) = 11.86$, $p < .001$). Evidently, changes made to the signal envelope had no critical effect upon any of these measurements.

The distribution of erroneous responses among individual patterns remained essentially unchanged. Distributions are given in

Table 3B. Patterns that were frequently mistaken for presented ones in Experiment 1 were also prominent in Experiment 2. Comparison between Table 3A and 3B of standard errors for response percentages of a given pattern revealed no significant differences. Subjects as a whole showed no signficant increase in recognition performance over the span of experimental sessions. Again, learning was not a component of the recognition task.

Discussion

The two experiments reported here represent a preliminary investigation of complex directional hearing and its capacity for integrating individual source localizations into a coherent spatial image. Results indicate a considerable ability to hear complex spatial patterns even under less than ideal localization conditions. The integrative capacity of the mechanisms associated with spatial hearing seems remarkable when one considers that patterns in these experiments were traced with a limited number of speakers arranged in a highly discrete array configuration. Even more impressive is the fact that the maximum total duration of a pattern (composed of 14 speakers) was only 1.62 seconds, and that adjacent speakers were closely spaced in terms of angular separation from the listener's vantage point (ranging from approximately 8^{o} to 9^{o}). Informal post-experimental questioning indicated that subjects under these conditions heard the equivalent of a single sound moving around the array, rather than separate sounds.

The apparent ability of the auditory system to perceive spatial images derived from visual concomitants lends support to the view that spatial representations are not necessarily sensory-specific, but rather are multi-modal. Since patterns of the type used here have no conceivable auditory equivalent in the natural environment, it would seem difficult to justify the existence of complex mechanisms for spatial representation unique to hearing. It could be argued, on the other hand, that auditory-spatial patterns can be processed through representational machinery normally associated with vision. In this way, patterns could be mapped into a strictly visual representation of space, at which point spatial relations and distance information are derived.

Such a model seems unlikely for several reasons, however. First, neurophysiological evidence (e.g., Knudsen 1984) suggests that even when modality-specific spatial maps of different sensory origin are found, they tend to be superimposed in such a way that adjacent neural regions respond to the same regions of space. Second, during informal post-experimental questioning, listeners did not report any visual images accompanying pattern presentations. If patterns were processed visually, one could expect some strong visual impressions or images. The lack of such images suggests, although not conclusively, that visual mechanisms were not employed exclusively in spatial representations. It may be, however, that some type of comparison occurs between fully-formed, distinct auditory and visual representations of a given pattern; presumably, this could be tesed in a future chronometric study by comparing reaction times for cross-modal matching of visual and auditory-spatial patterns with those for visual ones only. A further difficulty in accepting the primacy of vision in potential auditory-spatial integration is that visual and auditory mechanisms differ in their extent of spatial coverage. While vision provides precise information about a relatively limited spatial area at any given time, directional hearing gives complete coverage, encompassing all possible radial directions. While visually derived spatial maps of such extent can be built up from sucessive views, it would seem that a model of spatial processing based primarily on a visual representation of space would seem ill-equiped to handle rapid 360° spatial input. Work is currently underway by the author to assess whether auditory-spatial patterns of the type described here can be represented successfully when traced within a particular directional plane with the subject seated at the center of the plane, as opposed to the projection of patterns on an array placed before each subject. Finally, it seems inherently cumbersome to justify independent spatial processing mechanisms for specific senses from a theoretical, as well as an evolutionary standpoint. Successful adaptation to the multitude of spatial information impinging on the senses would require a consistent, comprehensive treatment of this combined sensory data. The need for simultaneous channels of input for various types of spatial information (e.g., object and pattern recognition, spatial orientation, mental maps) makes the notion of separate spatial representations ecologically redundant. A theory of generalized, multi-modal spatial representation is eminently more satifying in this respect.

The results of the two experiments also revealed the effects of several acoustical variables on the perception of complex patterns. Performance in both experiments decreased systematically as greater numbers of upper or lower partials were removed from the signal. The deterioration of pattern coherence with restrictions upon signal bandwidth generally concurs with findings from previous experiments on localization in the vertical and horizontal planes. It is known, for example, that frequencies above 7 kHz are needed for accurate median-plane localization (Roffler & Butler 1968; Gardner 1973); signals with frequency cut-offs below this value are localized randomly. Consequently, it was hypothesized that image clarity, and thus performance would degrade considerably as partials above this frequency were removed (e.g. 1-12 -> 1-8), and then level off. Expectations were only partly confirmed. Performance levels did decrease in the expected range, but kept on decreasing steadily as partials below 7 kHz were removed. This suggests that some relatively low-frequency information remained that provided limited elevation cues. It would be useful in a future study to test whether different vertical, as well as horizontal displacements of the entire image would produce significant changes in performance levels for upper partial removal. Such changes seem likely, since cues for different general areas of the median plane, such as "front", "above", and "behind" have different high-frequency requirements (Hebrank & Wright 1974).

The decrease in performance levels observed for the three signals with upper partials removed was somewhat surprising, particularly the dramatic drop induced by the removal of partials 7-8, since both horizontal and vertical localization mechanisms have been shown to operate in these frequency ranges. Horizontal cues for complex signals above 4 kHz can be extracted from intensity differences at the two ears, and vertical cues exist primarily at frequencies above 7 kHz. Small, but significant drops in performance for spectra with partials 5-12 and 7-12 could be explained on the basis that interaural time differences in this frequency range were no longer useful as localization cues. However, there should have been little change in recognition proficiency from spectra with partials 7-12 to those composed of partials 9-12, since cues from interaural amplitude differences and pinna assymetries were present in both; yet performances were poorest for this latter spectrum. A possible explanation may lie with the presentation of the signal itself; the quick succession of short pulses with rapidly changing spatial location may not provide enough time for spatial hearing mechanisms to extract relevant information at such frequencies. This hypothesis is difficult to confirm from past localization research, since practically all studies have tested static localization conditions with considerably longer stimulus durations.

The removal of sharp attack and decay from all signal spectra reduced performance levels significantly, if not dramatically across all spectra, except for that composed of partials 9-12; in this latter case, the high-frequency attack components presumably did not contribute much to localization, since their removal had no significant effect. Reductions in performance for the remaining spectra did not affect performance rates for the perception of individual patterns. Overall, however, it appears from the results that changes in envelope function on the scale tested offer some degree of localization information.

Systematic trends in the distribution of incorrect identification responses among patterns offer strong evidence that listeners in both experiments used some form of pattern-matching scheme in which specific portions or components of a pattern were organized into a compound image and then compared in some manner with spatial representations of all possible patterns until a positive identification was made. Patterns were most frequently confused with those that had similar component features, irrespective of the starting point of the pattern or its trace direction (see Tables 3A and 3B). For example, the distribution of errors among patterns for the number '3' is directly related to the number of common features that specific patterns share with '3': '9' is confused most often with '3', followed by 'R', '6', 'S', 'P', 'U', 'O', 'G' and 'C'. Similarly, in the case of 'S', confusion levels for patterns are proportional to their number of shared features. Such trends hold for the other patterns, although there are a handful of individual exceptions: for instance, 'O' is rarely confused with 'U', or vice-verse, even though they are quite similar in component structure. Despite some anomalies, confusion rates for all patterns decrease in a quite consistent way with an increase in shared features. Such results suggest the employment of feature extraction and comparison processes for character recognition proposed by several authors (e.g., Selfridge 1959; McClelland and Rummelhart 1981). Such models may offer the best paradigm in this case for the perception of a familiar pattern through unfamiliar sensory inputs. Unfortunately, the current experimental design did not yield enough data samples to permit a comparison of spectral profile with errors in pattern identification; a more complete analysis in this regard could reveal the effect of specific forms of image degradation upon recognition ability.

Indeed, it would seem that further work with complex auditory-spatial patterns can offer new insight into elementary feature detection mechanisms and pattern-recognition networks.

It seems curious that despite its fundamental role as a source of spatial information, hearing has not been examined in the context of current theories of complex perception and mental representations (Gibson 1979; Marr 1982; Shepard 1984). Sources of sound, whether static or moving, provide invariants in the flux of energy available to the sense of an individual freely moving in the environment. The position of a sound source, for example, undergos shifts as a listener orients himself with respect to it that are predictable by the laws of Euclidean geometry, and that conceivably produce transformational invariants similar those in the visual domain. Following this line of reasoning, it seems logical that such constraints might be internalized to a high degree, permitting internal simulations of the kind Shepard has elicited for vision (see Shepard & Cooper 1982). In this sense, the notion of some form of "ecological phonics" might not seem so far-fetched at all! Moreover, if spatial representations are indeed multi-modal, analogical mechanisms for mental rotations and transformation of spatial images that have been found for visual stimuli could apply to spatial hearing as well, albeit in a more rudimentary fashion. Such possibilities merit additional investigation.

References

Butler R A, Belendiuk K, 1977 "Spectral cues utilized in the localization of sound in the median sagittal plane" *Journal of the Acoustical Society of America* **61** 1264-1269

Carpenter P A, Eisenberg P, 1978 "Mental rotation and the frame of reference in blind and sighted individuals" *Perception & Psychophysics* **23** 117-124

Cooper L A, Shepard R N, 1973 "Chronometric studies of the rotation of mental images" in *Visual Information Processing* Ed W. G. Chase (New York: Academic Press)

Cooper L. A, Shepard R N, 1973b "The time required to prepare for a rotated stimulus" *Memory & Cognition* **1** 246-250

Corbin H H, 1942 "The perception of grouping and apparent motion in visual depth" *Archives of Psychology* No 273

Farrell J E, Shepard R N, 1981 "Shape, orientation, and apparent rotational motion" *Journal of Experimental Psychology: Human Perception and Performance* **7** 477-486

Freyd J J, Finke R A, 1984 "Representational momentum" *Journal of Experimental Psychology: Learning, Memory, and Cognition* **10** 126-132

Garbin C P, Bernstein I H, 1984 "Visual and haptic perception of three-dimensional solid forms" *Perception and Psychophysics* **36** 104-110

Gardner M B, 1973 "Some monaural and binaural facets of median plane localization" *Journal of the Acoustical Society of America* **54** 1489-1495

Gibson J J, 1979 *The Ecological Approach to Visual Perception* (Boston: Houghton Mifflin)

Harris J D, Sergeant R L, 1971 "Monaural/binaural minimum audible angle for a moving sound source" *Journal of Speech and Hearing Research* **14** 618-629

Harris J D, 1972 "A florilegium of experiments on directional hearing" *Acta Oto-Laryngol* Suppl 298

Hebrank J, Wright D, 1974 "Spectral cues used in the localization of sound sources on the median plane" *Journal of the Acoustical Society of America* **56** 1829-1834

Heller M A, 1989 "Texture perception in sighted and blind observers *Perception & Psychophysics* **45** 49-54

Knudsen E I, 1984 "Synthesis of a neural map of auditory space in the owl" in *Dynamic Aspects of Neocortical Function* Eds G M Edelman, W M Cowan, & W E Gall (NewYork: Wiley)

Marr D, 1982 *Vision* (San Francisco: W H Freeman and Company)

McClelland J R, Rummelhart D E, 1981 "An interactive model of context effects in letter perception: I. An account of basic findings" *Psychological Review* **88** 375-407

Mills A W, 1958 "On the minimum audible angle" *Journal of the Acoustical Society of America* **30** 237-246

Mills A W, 1972 "Auditory localization" in *Foundations of Modern Auditory Theory* (Vol 2) Ed J V Tobias ((New York: Academic Press)

Musicant A D, Butler R A, 1985 "Influence of monaural spectral cues on binaural localization" *Journal of the Acoustical Society of America* **77** 202-208

Palmer A R, King A J, 1982 "The representation of auditory space in the mammalian superior colliculus" *Nature* **299** 248-249

Perrott D R, 1984 "Concurrent minimum audible angle: A re-examination of the concept of auditory spatial acuity" *Journal of the Acoustical Society of America* **75** 1201-1206

Perrott D R, Musicant A D, 1977 "Minimum auditory movement angle: Binaural localization of moving sound sources" *Journal of the Acoustical Society of America* **62** 1463-1466

Rhodes G, 1987 "Auditory attention and the representation of spatial information" *Perception & Psychophysics* **42** 1-14

Roffler S K, Butler R A, 1968 "Factors that influence the localization of sound in the vertical plane" *Journal of the Acoustical Society of America* **43** 1255-1259

Shepard R N, 1984 "Ecological Constraints on Internal Representation: Resonant Kinematics of Perceiving, Imagining, Thinking, and Dreaming" *Psychological Review* **91** 417-447

Shepard R N, Cooper L A, 1982 *Mental images and their transformations* (Cambridge, MA: MIT Press)

Shepard R N, Zare S L, 1983 "Path-guided apparent motion" *Science* **220** 632-634

Stratton G M, 1897 "Vision without inversion of the retinal image" *Psychological Review* **4**

Thomas G J, 1941 "Experimental study of the influence of vision on sound localization" *Journal of Experimental Psychology* **28** 163-175

List of Figure Legends

Figure 1. (a) Form of speaker array. (b) Ten patterns selected for the experiment. The starting point for each trace is indicated by an 'X'. Arrows clarify trace direction.

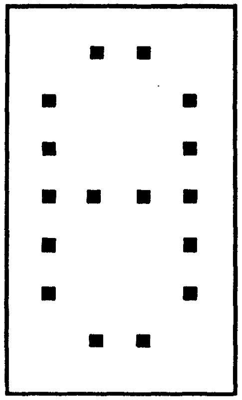
Figure 2. (a) Envelope shapes of signals used in Experiment 1 and (b) Experiment 2. The signal onset of (a) is shown enlarged at (c).

Figure 3. (a) Results of Experiment 1 and (b) Experiment 2 showing the effects of bandwidth on pattern recognition. The percentage of correct responses pooled across subjects and patterns is plotted for each of the seven spectra. Bars around percentage values represent standard errors for each spectrum.

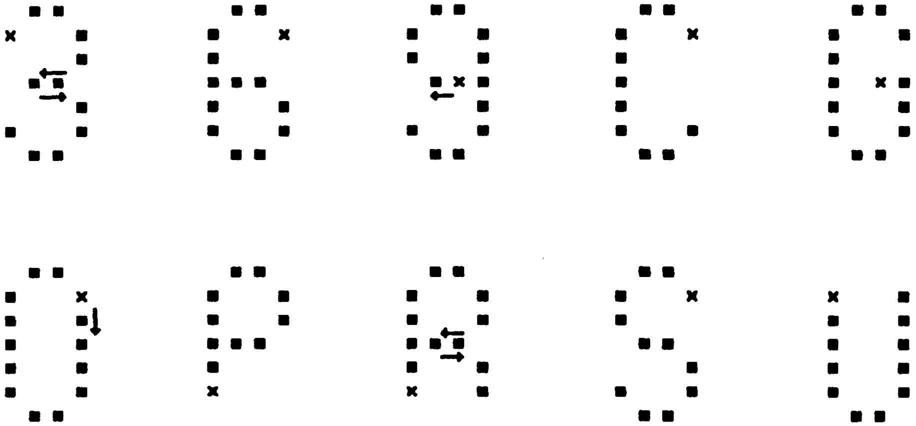
Table 1. (a) Results of Experiment 1 and (b) Experiment 2. Correct response percentages for individual subjects, pooled across patterns, are given for each spectrum.

Table 2. (a) Results of Experiment 1 and (b) Experiment 2 showing correct response percentages, pooled across subjects, for individual characters as a function of spectrum.

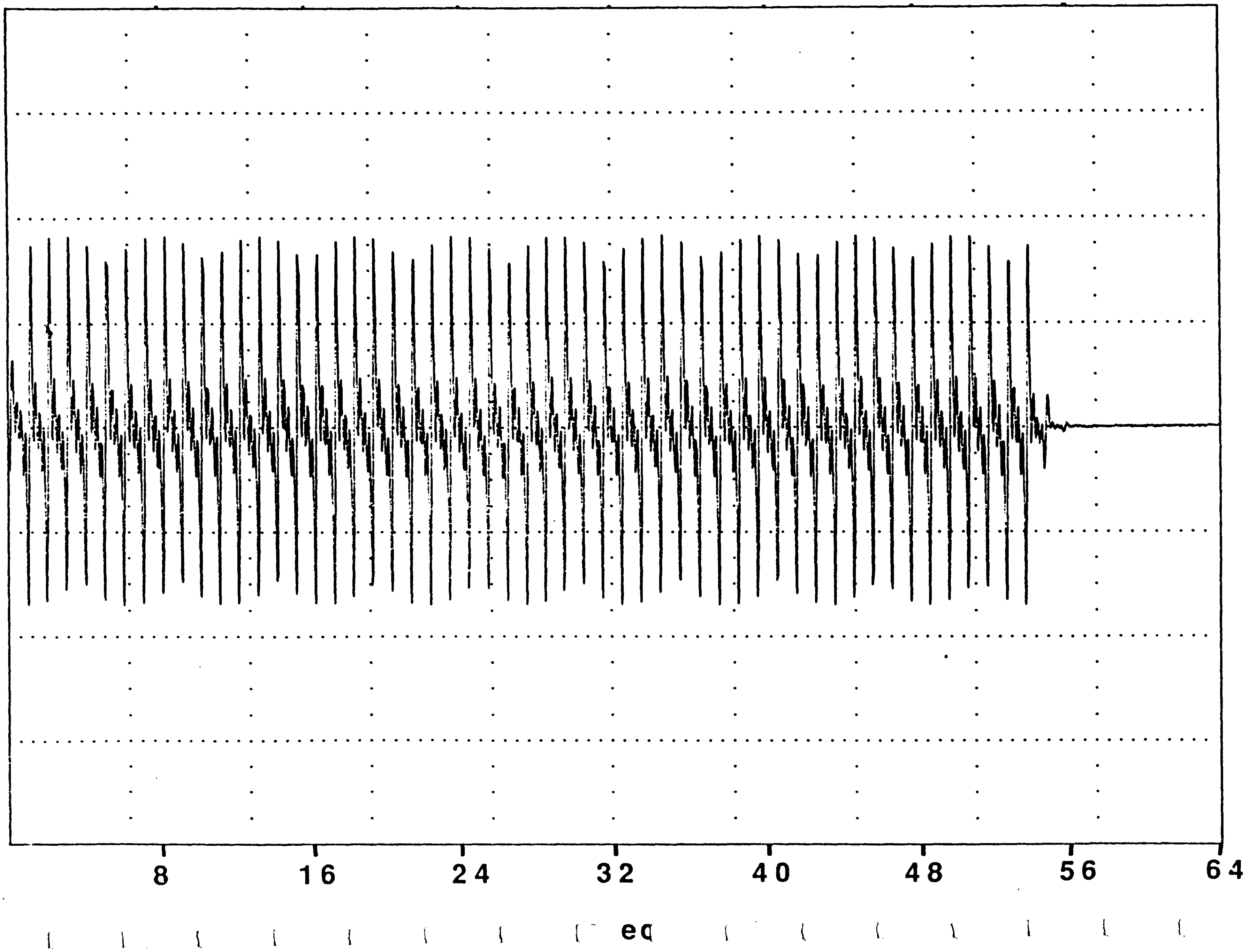
Table 3. (a) Erroneous responses pooled across subjects and spectra in Experiment 1 and (b) Experiment 2 are plotted according to pattern, as a function of the actual pattern presented. Values for each pattern are given in both graphical and numerical form as the percentage of erroneous responses corresponding to that specific pattern. Values in bold squares give the total percentage of correct responses for the character in each row.

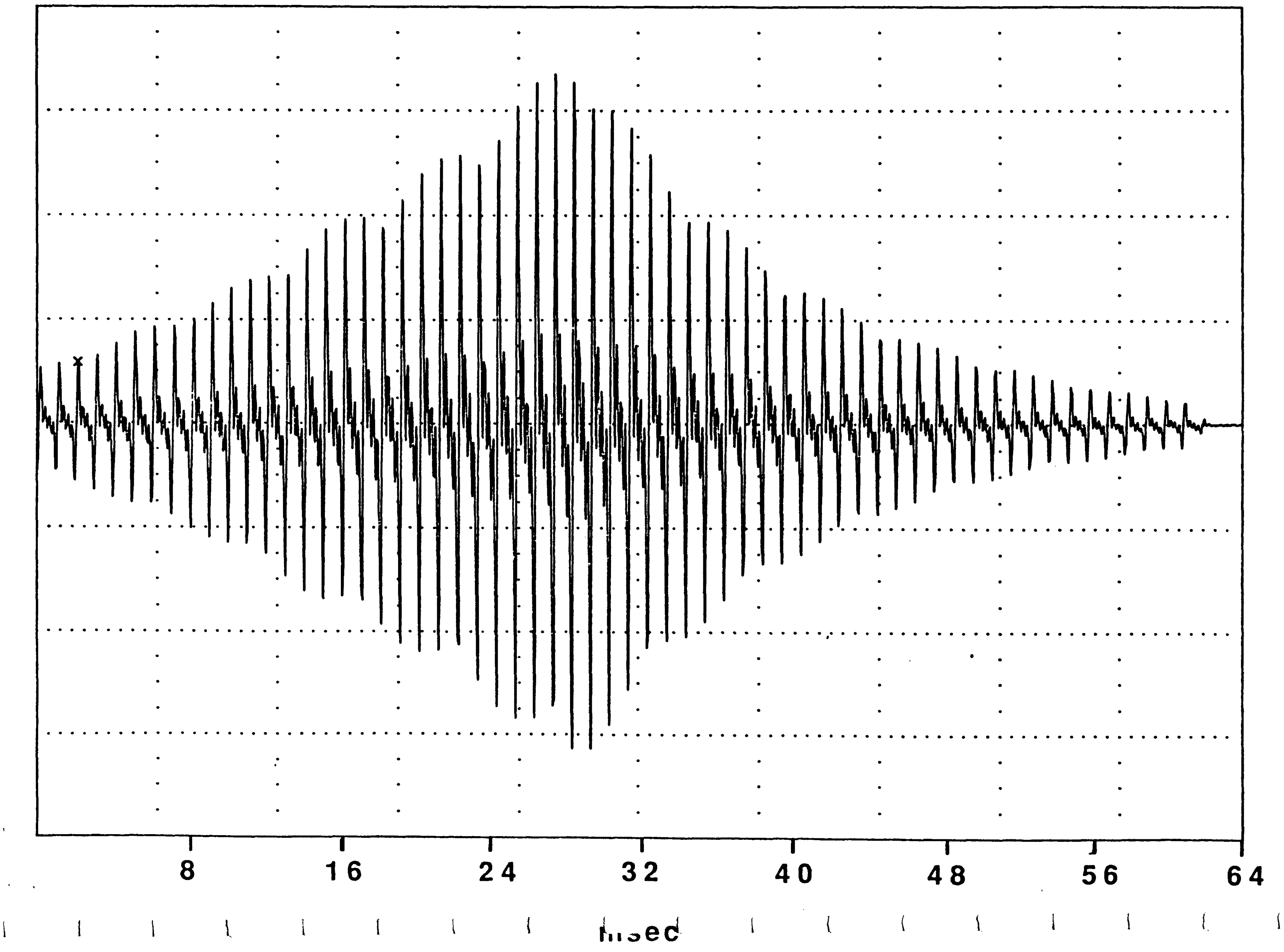


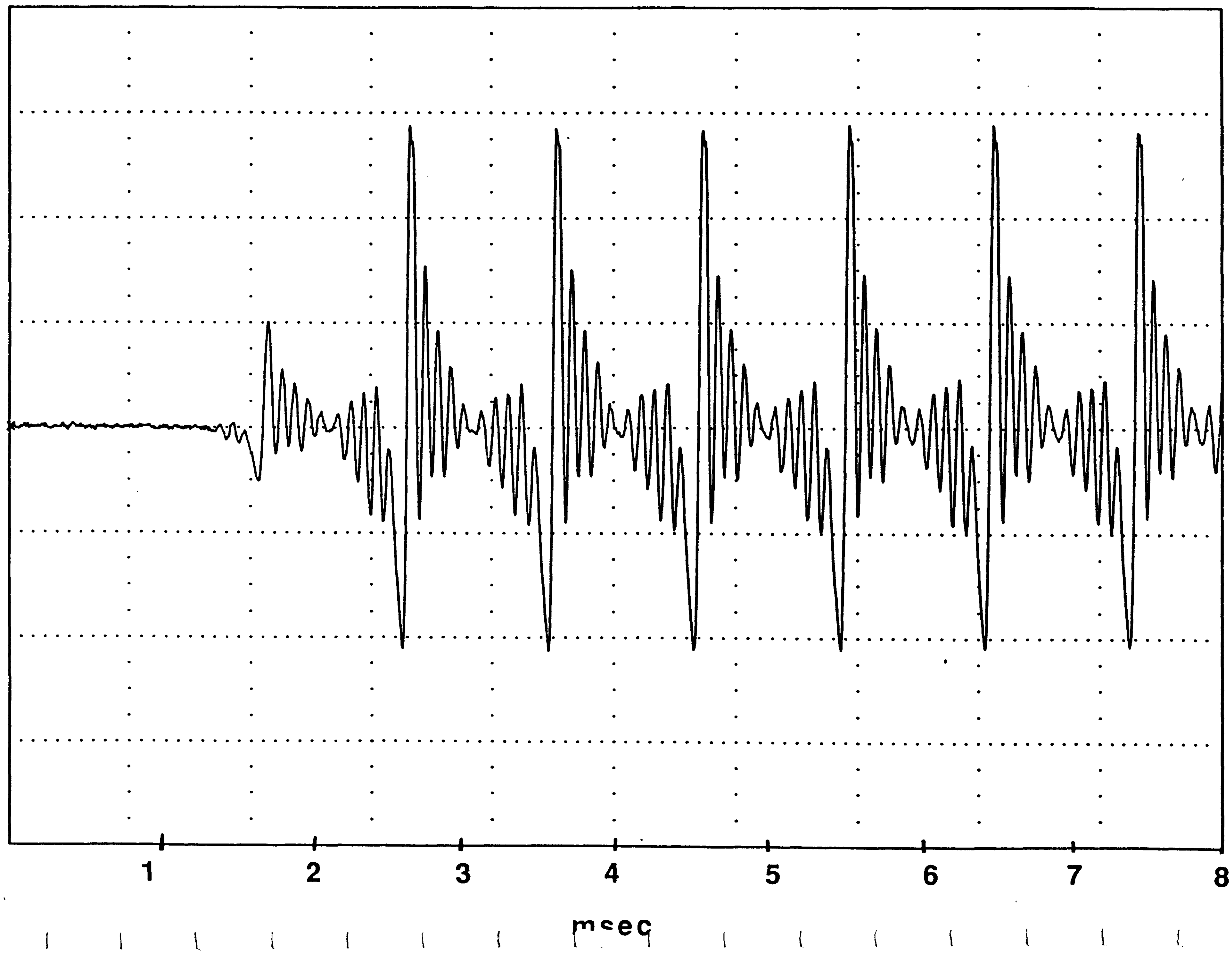
(a)

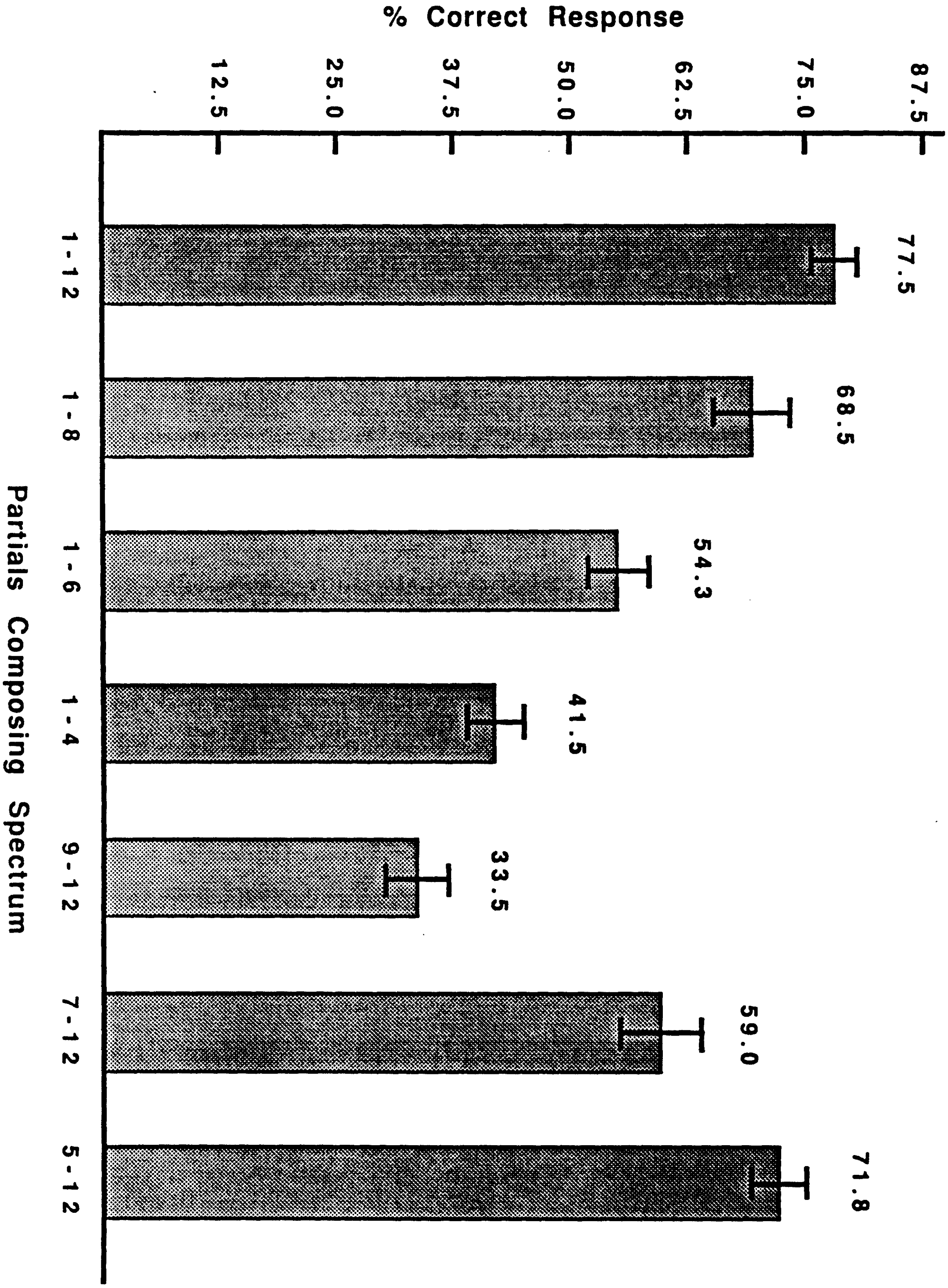


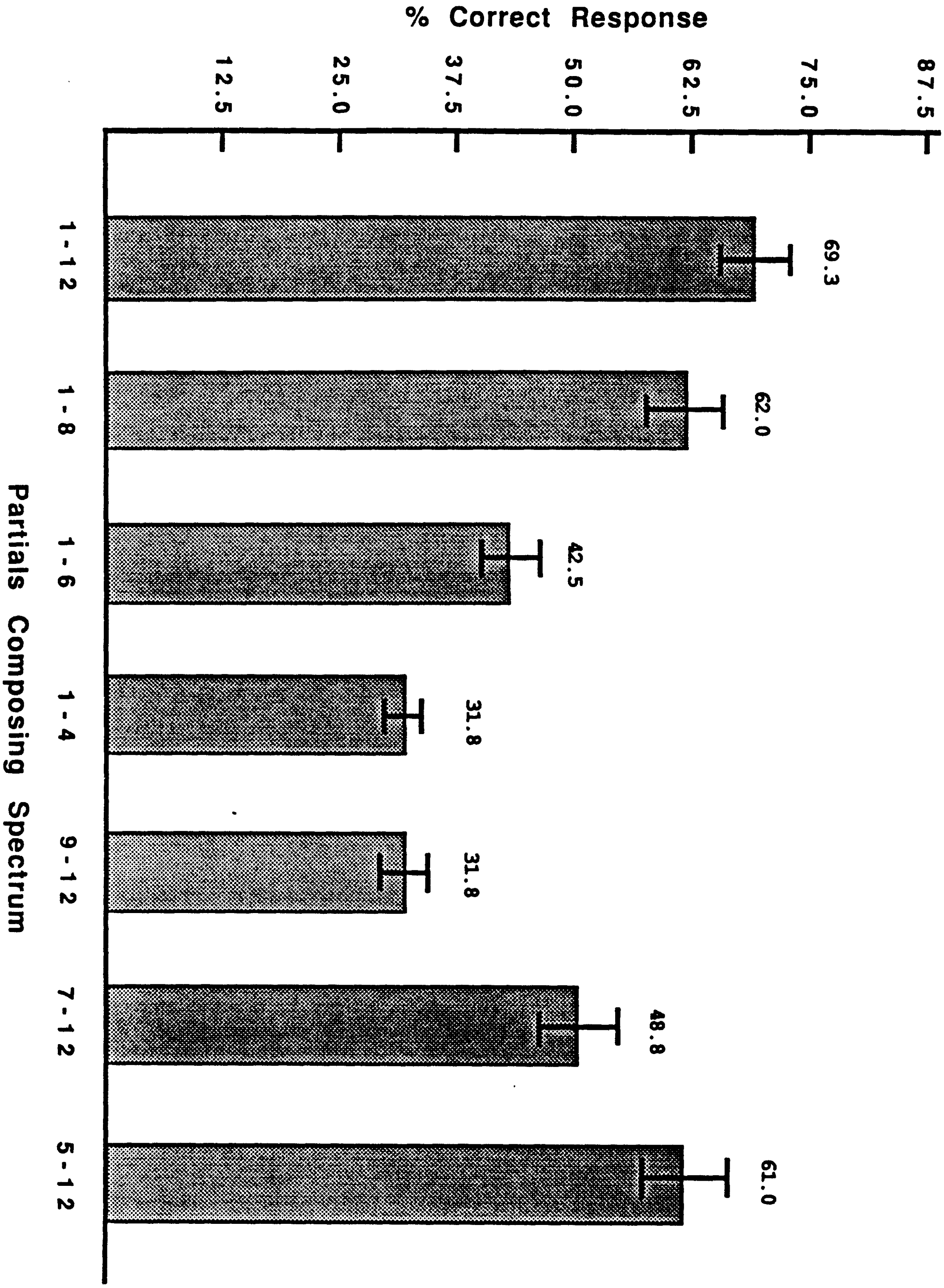
(b)


8
16
24
32
40
48
56
64
eq


8
16
24
32
40
48
56
64
msec


1
2
3
4
5
6
7
8
msec


% Correct Response
87.5
75.0
62.5
50.0
37.5
25.0
12.5
77.5
68.5
54.3
41.5
33.5
59.0
71.8
1-12
1-8
1-6
1-4
9-12
7-12
5-12
Partials Composing Spectrum


% Correct Response
12.5
25.0
37.5
50.0
62.5
75.0
87.5
Partials Composing Spectrum
1-12
1-8
1-6
1-4
9-12
7-12
5-12
69.3
62.0
42.5
31.8
31.8
48.8
61.0

Ta

Individual Subjects

Partials Composing Spectrum	1	2	3	4	5	6	7	8	9	10	
1-12	92.5	82.5	77.5	75.0	67.5	87.5	62.5	85.0	70.0	75.0	77.5
1-8	52.5	85.0	85.0	72.5	42.5	70.0	62.5	85.0	55.0	75.0	68.5
1-6	55.0	42.5	75.0	60.0	35.0	60.0	50.0	62.5	47.5	55.0	54.3
1-4	55.0	32.5	50.0	45.0	25.0	57.5	45.0	40.0	32.5	32.5	41.5
9-12	25.0	35.0	55.0	32.5	32.5	30.0	32.5	20.0	47.5	25.0	33.5
7-12	70.0	67.5	72.5	70.0	40.0	77.5	62.5	25.0	52.5	52.5	59.0
5-12	77.5	72.5	80.0	77.5	65.0	85.0	57.5	82.5	57.5	62.5	71.8
	61.1	59.6	70.7	61.8	43.9	66.8	53.2	57.1	51.8	53.9	

Subject Means

T

Individual Subjects

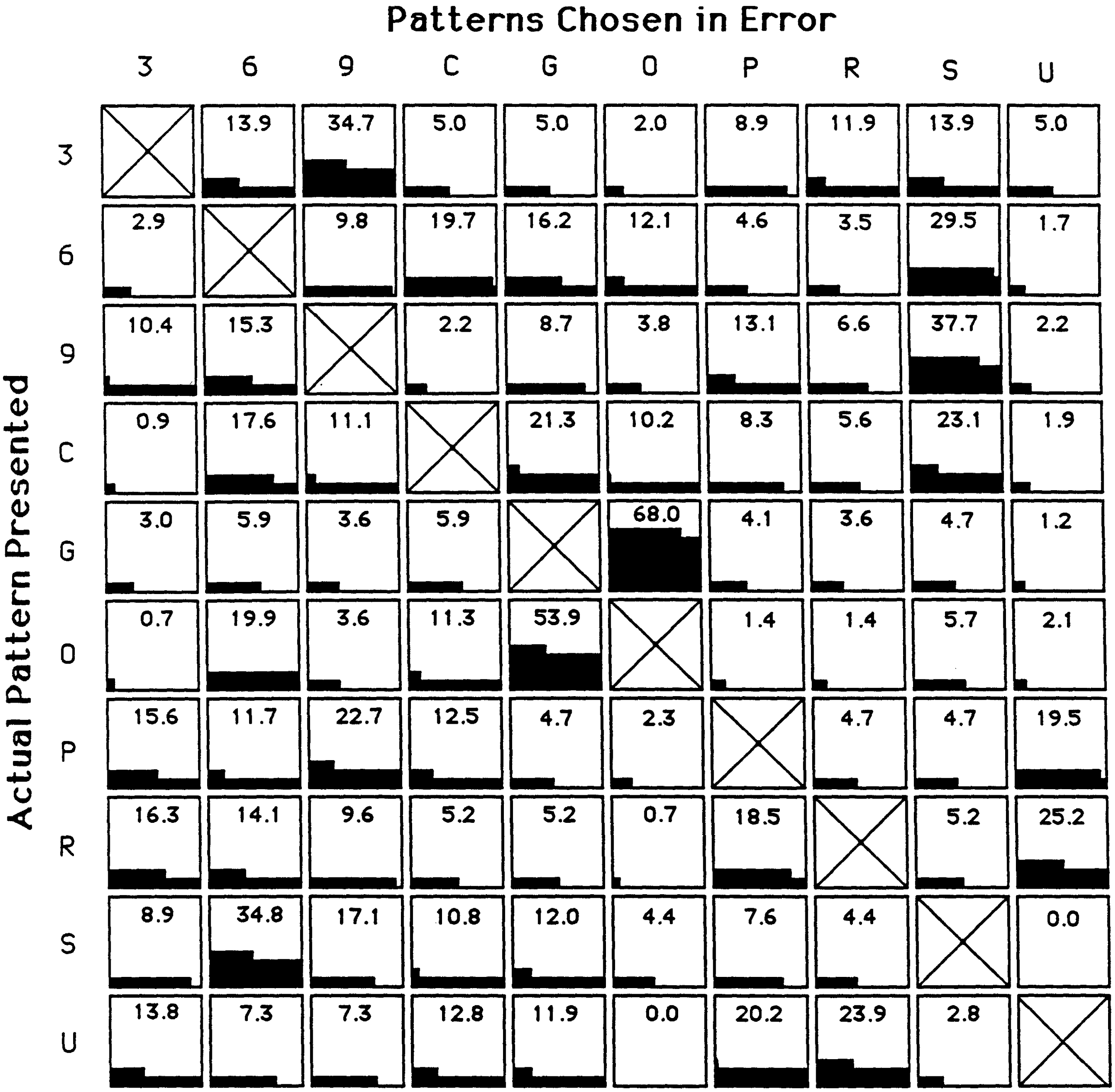
Partials Composing Spectrum	1	2	3	4	5	6	7	8	9	10	
1-12	65.0	72.5	65.0	85.0	47.5	80.0	90.0	65.0	52.5	70.0	69.3
1-8	55.0	52.5	60.0	77.5	47.5	80.0	82.5	60.0	45.0	60.0	62.0
1-6	30.0	32.5	50.0	45.0	40.0	70.0	32.5	47.5	37.5	40.0	42.5
1-4	35.0	27.5	32.5	30.0	27.5	45.0	32.5	42.5	27.5	17.5	31.8
9-12	25.0	37.5	35.0	42.5	30.0	47.5	35.0	17.5	27.5	20.0	31.8
7-12	60.0	62.5	32.5	67.5	42.5	62.5	70.0	22.5	42.5	25.0	48.8
5-12	55.0	77.5	55.0	82.5	40.0	75.0	75.0	52.5	42.5	55.0	61.0
	46.4	51.8	47.1	61.4	39.3	65.8	59.6	43.9	39.3	41.1	

Subject Means

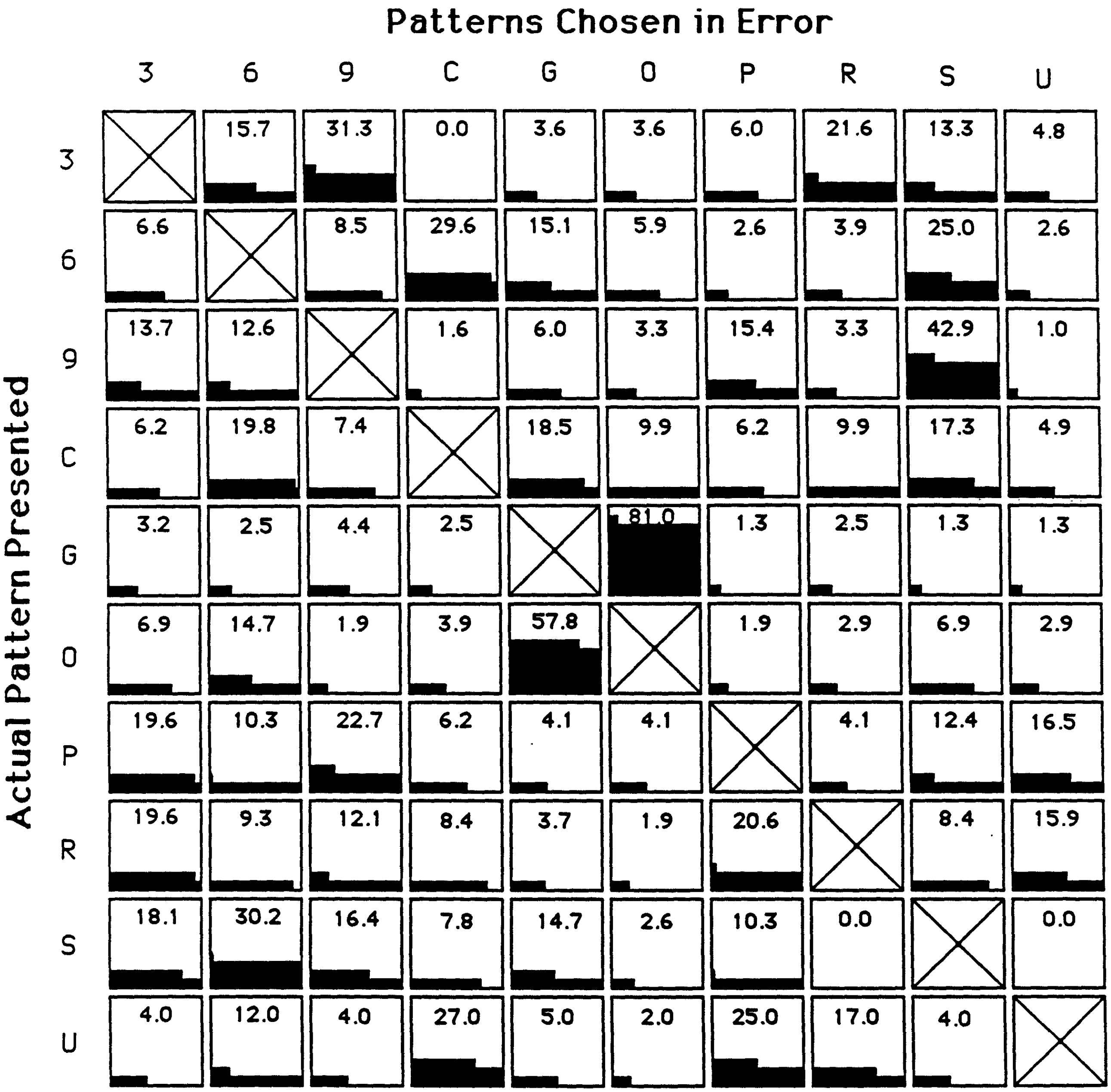
Partials Composing Spectrum	3	6	9	C	G	O	P	R	S	U	Spectrum Means
1-12	87.5	65.0	65.0	92.5	62.5	80.0	80.0	77.5	75.0	90.0	77.5
1-8	92.5	60.0	37.5	92.5	35.0	90.0	67.5	57.5	62.5	90.0	68.5
1-6	87.5	57.5	35.0	60.0	27.5	62.5	45.0	50.0	65.0	52.5	54.3
1-4	57.5	22.5	15.0	47.5	37.5	50.0	47.5	52.5	47.5	37.5	41.5
9-12	35.0	20.0	20.0	40.0	32.5	30.0	57.5	47.5	25.0	27.5	33.5
7-12	65.0	47.5	42.5	77.5	57.5	52.5	77.5	57.5	50.0	62.5	59.0
5-12	77.5	55.0	50.0	92.5	62.5	72.5	80.0	80.0	70.0	77.5	71.8
Pattern Means	71.8	46.8	37.9	71.8	45.0	62.5	65.0	60.4	56.4	62.5	

Patterns

Partials Composing Spectrum	3	6	9	C	G	O	P	R	S	U	Spectrum Means
1-12	92.5	45.0	57.5	82.5	60.0	57.5	80.0	80.0	52.5	85.0	69.3
1-8	82.5	67.5	32.5	67.5	32.5	77.5	47.5	62.5	70.0	80.0	62.0
1-6	75.0	40.0	15.0	65.0	10.0	52.5	32.5	25.0	57.5	52.5	42.5
1-4	50.0	35.0	20.0	45.0	17.5	35.0	25.0	20.0	30.0	40.0	31.8
9-12	20.0	15.0	22.5	37.5	42.5	22.5	55.0	42.5	12.5	47.5	31.8
7-12	52.5	35.0	40.0	50.0	65.0	32.5	67.5	75.0	27.5	42.5	48.8
5-12	70.0	37.5	45.0	77.5	52.5	67.5	65.0	57.5	55.0	82.5	61.0
Pattern Means	63.2	39.3	33.2	60.7	40.0	49.3	53.2	51.8	43.6	61.4	



	3	6	9	C	G	O	P	R	S	U
3	☒	13.9	34.7	5.0	5.0	2.0	8.9	11.9	13.9	5.0
6	2.9	☒	9.8	19.7	16.2	12.1	4.6	3.5	29.5	1.7
9	10.4	15.3	☒	2.2	8.7	3.8	13.1	6.6	37.7	2.2
C	0.9	17.6	11.1	☒	21.3	10.2	8.3	5.6	23.1	1.9
G	3.0	5.9	3.6	5.9	☒	68.0	4.1	3.6	4.7	1.2
O	0.7	19.9	3.6	11.3	53.9	☒	1.4	1.4	5.7	2.1
P	15.6	11.7	22.7	12.5	4.7	2.3	☒	4.7	4.7	19.5
R	16.3	14.1	9.6	5.2	5.2	0.7	18.5	☒	5.2	25.2
S	8.9	34.8	17.1	10.8	12.0	4.4	7.6	4.4	☒	0.0
U	13.8	7.3	7.3	12.8	11.9	0.0	20.2	23.9	2.8	☒



Patterns Chosen in Error

Actual Pattern Presented

	3	6	9	C	G	O	P	R	S	U
3	☒	15.7	31.3	0.0	3.6	3.6	6.0	21.6	13.3	4.8
6	6.6	☒	8.5	29.6	15.1	5.9	2.6	3.9	25.0	2.6
9	13.7	12.6	☒	1.6	6.0	3.3	15.4	3.3	42.9	1.0
C	6.2	19.8	7.4	☒	18.5	9.9	6.2	9.9	17.3	4.9
G	3.2	2.5	4.4	2.5	☒	81.0	1.3	2.5	1.3	1.3
O	6.9	14.7	1.9	3.9	57.8	☒	1.9	2.9	6.9	2.9
P	19.6	10.3	22.7	6.2	4.1	4.1	☒	4.1	12.4	16.5
R	19.6	9.3	12.1	8.4	3.7	1.9	20.6	☒	8.4	15.9
S	18.1	30.2	16.4	7.8	14.7	2.6	10.3	0.0	☒	0.0
U	4.0	12.0	4.0	27.0	5.0	2.0	25.0	17.0	4.0	☒