The studies conducted examine both signal-dependent factors, and listener-dependent factors. The examinations of signal factors include experiments on perceptual degradation due to signal interruption at critical rates (approximately 4cps), and studies mapping the early levels of representation of speech. The data support the existence of two qualitatively different early processing stages; the first is relatively peripheral and subject to neural fatigue, while the second is central and subject to criterion shifts. The studies of listener-based factors include studies of perceptual restoration of deleted sounds (phonemes or musical notes), and studies of the perceptual effect of attentional allocation. The restoration experiments indicate similar architectures in the perceptual processing of speech and music.
The attentional investigations demonstrate rather fine-tuned attentional control under high-predictability conditions. Significant progress has been made in achieving the research objective of clarifying the properties of complex auditory pattern recognition.
The research project (AFOSR 86-0357) is a broad examination of the perception of complex auditory signals, particularly speech and music. The studies conducted this year examine both signal-dependent factors, and listener-dependent factors. The examinations of signal factors include experiments on perceptual degradation due to signal interruption at critical rates (approximately 4 cps), and studies mapping the early levels of representation of speech. The data support the existence of two qualitatively different early processing stages; the first is relatively peripheral and subject to neural fatigue, while the second is central and subject to criterion shifts. The studies of listener-based factors include studies of perceptual restoration of deleted sounds (phonemes or musical notes), and studies of the perceptual effect of attentional allocation. The restoration experiments indicate similar architectures in the perceptual processing of speech and music. The attentional investigations demonstrate rather fine-tuned attentional control under high-predictability conditions. Significant progress has been made in achieving the research objective of clarifying the properties of complex auditory pattern recognition.
II. Research Objectives

The objective of the research project is to delineate principles that underlie the perception of complex auditory patterns. The stimuli used are speech and musical patterns of varying complexity. A wide array of experimental procedures and analyses are used to try to determine properties that are true of the perception of complex auditory patterns across stimulus domains. In addition, we also are interested in discovering any principles that are domain specific (e.g., as "categorical perception" has traditionally been claimed to be a principle of perception specific to the speech domain). The various experimental investigations in the project may be broadly grouped into studies of signal-based factors, and studies of listener-based factors. The former group includes experiments that explore how properties of the input signal determine perception, while the latter group includes studies of how listeners' expectations influence perception/performance. The former group primarily focusses on early representations of the signal, and the latter includes higher-level factors (including, but not limited to, attentional influences). The long-term goal of the research is to understand both signal-based and listener-based factors, and their interaction in the perception of complex auditory patterns.
III. Progress toward research objectives

This has been a year of considerable progress: One paper appeared in the *Journal of Experimental Psychology: Human Perception and Performance*, one is in press at *Perception & Psychophysics*, two book chapters are in press, and three papers are under review at the journals; these seven papers are listed in the next section, and copies are included in the Appendices. Most of the material in these papers has been covered in earlier Progress Reports; this report will summarize portions of this work not previously covered, and will focus on several new lines of research currently being conducted in our laboratory.

The three papers currently under review each contain some research not summarized in previous Reports. The project on perceptual restoration of musical stimuli (DeWitt and Samuel; Appendix E) is a major research effort based on Lucinda DeWitt's dissertation. Our first annual report summarized the basic demonstration of the phenomenon (Experiment 1), the comparison of restoration of missing notes in familiar versus unfamiliar melodies (Experiment 2), and the stronger perceptual effects found with the well-established musical representations of scales (Experiment 4). Appendix E also includes an examination of the effect of structural properties, such as rhythm and tonality (Experiment 3), and a converging test of the effect of well-established musical representations, in this case, of chords (Experiment 5). The results of this research program support three conclusions: (1) High-level musical information, e.g., melodic familiarity, does not enhance perceptual restoration of missing notes; (2) Lower-level musical information, e.g., knowledge of a scale or chord in a particular musical key, does increase perceptual restoration; and (3) This pattern is the same as is found for speech, where (high-level) sentential information does not increase perceptual restoration, but (lower-level) lexical information does. The parallelism of the speech and music results suggests that perceptual restoration follows general principles that apply to complex auditory patterns.

The research reported in Appendix F (Pitt and Samuel) examines whether listeners make particular use of information in stressed syllables to decode incoming speech. A critical aspect of the methodology is the neutralization of the simple amplitude and duration advantages in such syllables; the goal is to determine if the perceptual process has developed procedures that focus on stressed syllables because they normally would serve as "islands of reliability" in the generally messy speech signal. If so, then we should observe processing advantages for syllables in stressed contexts, even though the syllables are acoustically no longer stressed. A set of three experiments supports the following conclusions: (1) When the signal provides robust cues to predict the occurrence of a stressed syllable, there is in fact better processing of the predicted-to-be-stressed syllable than if no such expectation exists, even in the absence of any acoustic basis for the advantage; and (2) Normal sentential context does not provide very strong predictions of stress. Thus, the system does seem to be geared to work on stressed syllables differentially, presumably because they can serve as reliable anchors for processing. The weakness of stress predictability in normal sentences suggests that under normal conditions, the system rapidly recognizes the occurrence of stressed syllable on the basis of its acoustic advantages (e.g., greater amplitude and duration), and then focuses processing on such syllables.

Most of the research on signal alternation effects (Appendix G) has been covered in previous reports. However, the final version of the paper includes a new experiment that is an initial attempt to examine a source-effect basis to the phenomenon (Experiment 5). The basic effect is a breakdown in perceptual...
processing when a signal is alternately presented to the right and left ears over headphones. At a critical alternation rate of about 4 cps, intelligibility is seriously impaired; slower and faster rates are much less detrimental. The phenomenon indicates that information presented to the two ears does not get integrated. The existing explanation for the effect invokes a separate sensory store for each ear to account for the lack of signal integration. However, no other research implicates monaurally-driven sensory storage. Experiment 5 is an initial study aimed at developing an alternative account in which the failure of integration is traced to processes that assign portions of the auditory input to different environmental sources. Because the signals in this paradigm are presented alternately to each ear, two different locations (180 degrees apart) are implied, and thus two (separate) sources. The initial test of this analysis presented in Experiment 5 did not produce evidence to support this account, but new experiments are planned that should provide more sophisticated tests.

Perceptual Restoration: Attentional Factors. As noted previously, in addition to the seven research projects included in the Appendices, several new projects have been conducted during the past year. One of these new studies examines possible attentional effects in the phonemic restoration illusion. As noted in the discussion of musical restoration, previous research has shown that listeners use pre-existing lexical knowledge to perceptually restore missing pieces of words. In a study conducted in our own laboratory some years ago (Samuel and Ressler, 1986), we examined a possible attentional component to the effect: Under normal listening conditions, people try to extract words from the signal, and have no incentive to analyze lower-level aspects. They may therefore have difficulty attending to the lower-level representations, and be unable to judge whether any low-level piece of the signal was present or not. In the earlier study, we used two techniques to help listeners focus attention on the lower level: training (with feedback), and attentional cues. We found that a thousand training trials did not have any noticeable effect, suggesting that restoration is a basic perceptual property. However, we did find one attentional cue condition that reliably helped subjects toward more veridical perception: If subjects were given both the identity of the impending test word, and the identity/location of the to-be-restored phoneme, they were better able to resist the illusion. Interestingly, either of these two cues individually was useless, suggesting that listeners had to direct attention to the appropriate low-level representation via the lexical representation.

We have finished two new training conditions, and are in the middle of two new tests of attentional cues. The new training conditions are intended to give subjects enough training to allow any improvements that might be possible. Thus, we have run subjects for 10,000 trials with feedback. In one condition, half of the stimuli are real words, and half are pseudowords (e.g., "pafis", or "garfiniay"). We were interested in whether lexical representations would develop for these new "words", and what effect this would have on performance. The results were very simple: Subjects performed terribly on both words and nonwords, and showed essentially no improvement through the six weeks of training. The second training study only used real words, to allow listeners to use a consistent strategy (if they could). We found that performance was somewhat better than in the mixed condition, but that there was still very little improvement in discrimination performance over the course of 10,000 training trials. Thus, the earlier conclusion, that restoration is a fundamental process, seems to be sound.

We are currently examining the effects of two new attentional cues. The earlier research had suggested that listeners' attempts to focus attention on a particular phoneme were obligatorily directed through lexical representations. The new
conditions are designed to show whether a sublexical route is possible: the syllable. In one condition, subjects are given a visual cue that provides syllabic specification. For example, if the potentially deleted phoneme is the final "t" in "candidate", a subject would see ".//date". If the previous results were due to implicitly cueing the critical syllable (by giving the phoneme in a word), then this condition should improve discrimination as much as the word-plus-phoneme cue. The second cuing condition being tested gives subjects the same syllabic information as the first, plus an indication of the critical phoneme ("//date"). If subjects use a syllabic representation of speech, then this condition could allow them to direct attention to the phonemic detail through the syllable. Again, this mechanism could have produced the observed advantage found for word-plus-phoneme cues. These conditions thus provide a test of whether syllabic representations, rather than lexical ones, are fundamental. Either outcome (syllable or word as basic) would be of interest.

Sublexical Attentional Effects. We are currently conducting another set of experiments, using a very different methodology, that is concerned with how attention can be allocated at different levels of analysis. These experiments use the phoneme monitoring technique in which subjects are told to respond to the occurrence of a specified phoneme (e.g., /b/); reaction time is the dependent measure. We have introduced the use of cost/benefit analyses to this procedure. Virtually all previous studies using phoneme monitoring have only included word-initial target phonemes. In the handful of studies using other target positions, targets were equally probable in the possible positions (word-initial or word-medial). In our current work, the critical test words are bisyllabic, and have a consonant-vowel-consonant-consonant-vowel/consonant (cvc-cvc) structure. In a baseline condition, target phonemes are equiprobable in the four possible consonant positions. In each of the four experimental conditions, a very high percentage of the targets occur in only one of the four consonant positions. There are occasional targets in low-probability positions. The cost/benefit analyses examine whether those targets yield slower detection reaction time relative to the baseline (costs), and whether targets in the high-probability position are detected faster than in the baseline case (benefits).

This methodology provides a very neat measure of the precision with which attention can be focussed. For example, if subjects attend at a syllabic level in this task, then targets in the same syllable should behave similarly: Targets in second position (cvc-cvc) should accrue similar benefits as those in first position (Cvc-cvc) in the high-probability first-position condition because they are in the same syllable. If subjects work with CV or VC codes, (called "diphones" or "demi-syllables") rather than syllabic ones, then third position consonants could benefit in the first-position weighted case, because they share CV structure (Cvc-Cvc). The finest-tuned case possible would be one in which only the high-probability position itself showed benefits. This pattern would support an allocation process that can operate at a phonemic level.

The results of our first experiment in this line of research indicated a rather fine-tuned allocation of attention. Figure 1 presents reaction time difference scores for the four experimental conditions relative to the baseline. As is clear in the Figure, targets in all four experimental conditions showed reliable benefits for the high-probability location. Interestingly, these benefits were generally accompanied by reliable costs for targets in other positions. For example, while targets in (high probability) first position were detected about 115 ms faster than in the baseline condition, (unexpected) second-positions targets in this case were approximately 125 ms slower than baseline. The results thus imply (1) an ability
to focus attention in accord with the experimental probability structure, and (2) a relatively fixed overall level of performance; gains in the expected case are accompanied by costs in unexpected locations. The shape of the cost/benefit profile across positions provides us with an estimate of the "envelope" of attentional focus.

We have run a second experiment in this project to determine if subjects use lexical representations to help them in this monitoring task. Once subjects hear the first syllable of a word, they may generate expectancies as to the possible identity of subsequent phonemes in the word, based on their knowledge of the language. If so, monitoring times to targets occurring in later positions (second syllable) may be facilitated by the use of such knowledge-based expectations. We tested this hypothesis in the second experiment by using nonword stimuli created by changing at least two consonants in each of the real words used in Experiment 1. Such nonword stimuli do not allow listeners to take advantage of lexical information because they have no lexical representations. If listeners were using such representations for the words in Experiment 1, monitoring times should be differentially slower to second-syllable targets in the nonwords of Experiment 2.

The data were in accord with this view: Reaction times to target phonemes in final position of the nonwords were significantly slower than detection times for the corresponding targets in real words. A similar (though not statistically significant) effect was found for monitoring times in the third consonant position (cvc-Cvc). In addition, the cost/benefit analysis for the nonwords showed a very similar pattern to the pattern found for words: At all four expected locations, targets were detected faster than in the equiprobable baseline (see Figure 1). These benefits were accompanied by reliable costs. There were some interesting differences in the pattern of costs and benefits between the word and nonword cases, possibly indicating differences in how listeners can allocate attention to words versus nonwords. We plan further experiments to pursue this, and to use this apparently sensitive technique to explore additional issues involving possible sublexical representations.

Properties of Two Early Sublexical Levels of Analysis. We have argued (Kat and Samuel, 1984; Samuel, 1986; Samuel and Newport, 1979) that there are two qualitatively different levels of analysis involved in early speech processing. The first level is essentially a "neural spectrogram", while the second is a more abstract representation. The first level has been called "simple acoustic", and it is thought to be relatively peripheral (monaurally-driven). The second level (called "complex acoustic") is hypothesized to include central (binaurally-driven) units, with significant recoding of the signal. These units can be used for phonetic coding, but are not exclusively linguistic; other complex sounds (e.g., a plucked violin string) are represented at this level as well.

Previous research has demonstrated that the identification shifts induced by selective adaptation are accompanied by reliable reaction time changes. For example, if a /ba/-/da/ test series is used, reaction times to identify tokens at the /b/ end of the series are higher after /ba/ adaptation than after /da/ adaptation; at the /da/ end, the reverse is true. Our current research looks for (1) absolute changes in reaction time and (2) differential effects of "acoustic"
factors and more abstract ("phonetic") factors. In all conditions, the test series was an 8-step /ba/-/da/ continuum.

Two conditions were designed to test the hypothesis that adaptation causes changes in the function of first-level units primarily through fatigue, while second-level units are instead subject to criterion shifts. The reaction time prediction was that fatigue should reduce efficiency of a processing structure, resulting in slower responses; criterion shifts, with no change in efficiency, should not affect reaction times.

The structure of the baseline identification test (an uninterrupted series of syllables to identify) is different enough from the adaptation phase (test syllables interspersed among blocks of adaptor presentations) to make comparisons of absolute reaction times risky. Figure 2 provides a comparison of labelling and reaction times in three cases: (a) adaptation with /a/ versus baseline identification (without any adaptation) [left panels]; (b) adaptation with /ba/ versus adaptation with /a/ [center]; (c) adaptation with /da/ versus adaptation with /a/ [right].

In principle, the /a/ adaptation condition serves as a better baseline than the baseline identification test, since responses are collected under the same timing conditions. Because repeated presentation of /a/ should have no effect on identification of /ba/ or /da/, we would expect no shift in responses, and as the figure shows, none is found. In contrast, the endpoints of the continuum cause large shifts in labelling. The reaction time functions (plotted against reaction times in the /a/ condition) also show robust adaptation effects: Responses are significantly slower to stimuli in the adapted category. This one-sided lifting of the RT function is what should occur if the adaptation procedure fatigues structures responsible for identifying /ba/ or /da/, making the operation of such structures less efficient.

A second experiment in this series was designed to test whether the reaction time effects do in fact reflect fatigue of processors. The key to this test is the use of two different types of adaptors, one type that should match the test series acoustically (and thus induce fatigue), and one type that should not.

The left panels of Figure 3 show the labelling and reaction time results for a /pa/ adaptation condition plotted against those from a /ta/ adaptation condition. As noted above, the test series was /ba/-/da/. The /pa/ adaptor shares place of articulation with /ba/, and produces an attenuated, but reliable shift in the same direction as /ta/; /ta/ matches /da/ in place, and produces a comparably reduced
but reliable effect. Some of the earliest adaptation studies (e.g. Eimas and Corbit, 1973) used such "cross-series" adaptors, and argued that the observed shift reflected effects at an abstract level of representation, since the acoustic match (of /pa/ to /ba/, for example) is not very good. In terms of the two-level model just discussed, the /pa/ and /ta/ adaptors should affect the second, abstract level of analysis. This level was hypothesized to be subject to criterion shifts, and as the reaction time data show, there was absolutely no differential slowing of responses with cross-series adaptation. Thus, the data are exactly as predicted.

The right panels of Figure 3 show the labelling and reaction time results for two adaptors derived from the endpoint /ba/ and /da/ syllables. These adaptors only included second and third formant energy; F1 was eliminated in the synthesis procedure. These adaptors contain most of the acoustic cues for /ba/ or for /da/, but the absence of F1 severely reduces their phonetic quality. These adaptors were expected to affect the simple acoustic level (because of the acoustic match), but to have little effect on the more abstract level (because of the virtual elimination of phonetic quality). As the Figure shows, these adaptors produced labelling shifts of exactly the same magnitude as the /pa/ and /ta/ adaptors. In contrast to /pa/ and /ta/, these adaptors produced significant differential reaction times. This results is predicted by the view that these adaptors should affect the simple acoustic level, and that this level is subject to response-slowing fatigue effects.

Our reaction time measures of selective adaptation thus appear to delineate the early levels of speech processing. Results for the single endpoint adaptors are consistent with a model in which such adaptors produce fatigue of low-level acoustic analyzers; these adaptors also presumably affect more abstract levels of analysis (via criterion shifts). The final set of data were particularly diagnostic: Adaptors sharing abstract properties with the test series produced adaptation shifts of the same magnitude as those produced by acoustically-matched but phonetically-degraded ones. However, only the latter ones caused a change in the time it took subjects to identify syllables. These results provide detailed support for the two-level model, and help to delineate the properties of units at each level of analysis.

References


IV. List of Publications

Papers Appearing During this Period


Papers in Press


Papers Submitted for Publication


V. Personnel

Principal Investigator: Arthur G. Samuel, Associate Professor of Psychology at Yale University. Ph.D. from University of California, San Diego, 1979.

Associate in Research: Donna Kat. B.A. in Psychology from University of California, San Diego, 1979. Ms. Kat has been first author on one journal article and second author on two. She has been first author on two Acoustical Society presentations.

Graduate Student: Lucinda DeWitt. Ph.D. to be awarded by Yale University, May 1989. Title: The Role of Knowledge-based Expectations in Music Perception. Dr. DeWitt is now an Assistant Professor of Psychology at Depauw University.

Graduate Student: Mark Pitt. Currently finishing a dissertation entitled The Allocation of Attention During the Processing of Speech. The dissertation will be finished this summer, and Mr. Pitt will assume a Postdoctoral position in our lab in the fall.
VI. List of Interactions (coupling activities)

Conference Presentations

During this reporting period, we have made five presentations at four conferences. These are summarized here:


Other Interactions:

In addition to the conferences listed above, the PI has been heavily involved in the review process for both granting agencies and journals. This includes grant reviews for AFOSR and NSF, and manuscript reviews for:

- Behavioral Research Methods and Instrumentation
- Cognition
- Cognitive Psychology
- Journal of the Acoustical Society of America
- Journal of Experimental Psychology: Human Perception and Performance
- Language and Speech
- Memory and Cognition
- Perception & Psychophysics
- Quarterly Journal of Experimental Psychology

During this reporting period, the PI was asked (and agreed) to join the Editorial Board of three journals: Cognition, Memory and Cognition, and the Journal of Experimental Psychology: Human Perception and Performance. These professional activities produce a great deal of interaction with other researchers.
Fig 1

Expected-Location Condition

- cyc-cyc

- cvc-cvc

- cvc-cvc

- cvc-cvc

Consonant-Location Condition

c1 c2 c3 c4
Figure 3

Stimulus Number vs. Reaction Time (RT)