REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (*Leave blank*)

2. REPORT DATE

3. REPORT TYPE AND DATES COVERED
FINAL 01 Nov 92 TO 31 Oct 96

4. TITLE AND SUBTITLE
PSYCHOPHYSICS OF COMPLEX AUDITORY AND SPEECH STIMULI

5. FUNDING NUMBERS
F49620-93-1-0033
61102F
2313/BS

6. AUTHOR(S)
Dr Richard E. Pastore

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
State University of New York
Dept of Psychology
P.O. Box 6000
Binghamton NY 13902-6000

AFOSR-TR-97
0150

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
AFOSR/NL
110 Duncan Ave Room B115
Bolling AFB DC 20332-8080

Dr John F. Tangney

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

19970602 120

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release;
distribution unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT (*Maximum 200 words*)

The supported research provides a careful examination of the many different, interrelated factors, processes, and constructs important to the perception by humans of complex acoustic signals, including speech and music. Traditional, solid psychophysical procedures were employed to systematically investigate perceptual interaction, grouping, and streaming as a function of physical and perceptual properties of stimuli. Models of stimulus interaction are being developed from research with simpler stimuli and tested with more complex stimuli, including speech. In addition, several cross-validated scaling measures (e.g., speeded classification, rating of goodness, similarity) and procedures were used to determine the multidimensional perceptual space for highly learned categories (e.g., place contrasts for speech), identifying the critical underlying dimensions, the function of each dimensions for every category, and the nature of interactions among dimensions. Results also were used to develop and evaluate prototype, exemplar, and threshold models for the underlying categorization process. The research provides a comprehensive picture of lower and higher level factors and processes which result in the perception of classes of complex auditory stimuli, including speech and music. In health, industry, and human factors, the

14. SUBJECT TERMS

15. NUMBER OF PAGES

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT (U)

18. SECURITY CLASSIFICATION OF THIS PAGE (U)

19. SECURITY CLASSIFICATION OF ABSTRACT (U)

20. LIMITATION OF ABSTRACT (U)

DTIC QUALITY INSPECTED 1

CONT: F49620-93-1-0033
SUNY-BINGHAMTON
Dr Richard E. Pastore

evaluation of problems and the development of appropriate approaches to treatment are limited by the accuracy of our understanding of the basic, underlying processes. Therefore, the improved understanding of perceptual processes for auditory and speech stimuli which result from this research has significant implications for scientific and practical advances in all of these fields.

1. Cover Page

Final Report

Air Force Office of Scientific Research
F49620-93-1-0033

Psychophysics of Complex Auditory and Speech Stimuli

November 1, 1993 - October 31, 1996

Principle Investigator:	**Richard E. Pastore**
Institution:	**State University of New York Binghamton, NY 13902-6000**

TABLE of CONTENTS

2. OBJECTIVES

The supported research provides a careful examination of the many different, interrelated factors, processes, and constructs important to the perception by humans of complex acoustic signals, including speech and music. Traditional, solid psychophysical procedures were employed to systematically investigate perceptual interaction, grouping, and streaming as a function of physical and perceptual properties of stimuli. Models of stimulus interaction are being developed from research with simpler stimuli and tested with more complex stimuli, including speech. In addition, several cross-validated scaling measures (e.g., speeded classification, rating of goodness, similarity) and procedures were used to determine the multidimensional perceptual space for highly learned categories (e.g., place contrasts for speech), identifying the critical underlying dimensions, the function of each dimension for every category, and the nature of interactions among dimensions. Results also were used to develop and evaluate prototype, exemplar, and threshold models for the underlying categorization process. The research provides a comprehensive picture of lower and higher level factors and processes which result in the perception of classes of complex auditory stimuli, including speech and music. In health, industry, and human factors, the evaluation of problems and the development of appropriate approaches to treatment are limited by the accuracy of our understanding of the basic, underlying processes. Therefore, the improved understanding of perceptual processes for auditory and speech stimuli which result from this research has significant implications for scientific and practical advances in all of these fields.

3. SUMMARY OF COMPLETED RESEARCH

A. MULTIDIMENSIONAL STRUCTURE OF PHONEME CATEGORIES

It is well accepted that the cues for speech categories are complex, with no single variable, or range of variable values, serving as an invariant, or even relatively consistent, cue. Yet, with only a few notable exceptions, most investigations of all aspects of speech perception over the last three decades have followed a long-standing procedure of studying perception by evaluating labeling, and occasionally discrimination, as a function of variation along a single physical dimension. A typical set of results is summarized in Figure 1, where the abscissa follows typical speech research by designating stimulus number [which here represents equal physical changes in the third formant (F3) onset frequency], and the ordinate is percent labeling of /d/. The two curves represent the results for different values of the second formant (F2) onset frequency; comparison across the two curves provides the typical evaluation of interaction between cues. (The results are taken from a small subset of our results, described below, for the /u/ vowel context without an initial release burst). One can conclude that a distinction or contrast between /d/ and /g/ (the alternative category) can be defined along the F3 onset frequency continuum, and that the value of F2 onset "trades" or interacts with (can alter the boundary location defined along) F3 onset variable. Thus, F3 and F3 onset frequencies are cues for the /d/ - /g/ contrast, and these two cues "trade" with each other. It should be obvious that these results provide a very limited perspective on the importance of either of the variables or the nature of their interaction (with the variable implicitly assumed to contribute equally to both labeling categories studied). In addition to such labeling studies, possible "perceptual" cues have been identified by analyzing the physical (spectral and temporal) properties of naturally produced stimuli, with some limited level of perceptual validation using a labeling task.

Using these very basic types of approaches, possible cues for voiced stop consonants varying in placement of the articulators prior to the onset of the consonant (thus, varying in place of articulation) were identified in systematic studies beginning in the 1950s (e.g., Liberman, Delattre, Cooper, & Gerstman, 1954; Delattre, Liberman, & Cooper, 1955; Halle, Hughes, & Radley, 1957). This early research identified F2, F3, and release burst as possible cues for perceptual categories contrasted in place of articulation. Later research further specified the complex nature of the stimulus features which might cue place categories (e.g., Fant, 1972; Cole & Scott, 1974). Somewhat more recent studies analyzing large sets of naturally produced stimuli and evaluating classification of complex sets of synthetic CV syllables, identified possible category-specific features in the gross dynamic spectral changes at consonantal release, with the release burst also possibly contributing to classification (e.g., Stevens & Blumstein, 1978, Zue, 1977). The formant transitions for velar consonants (e.g., /g/) tend to exhibit a prominent middle frequency spectral peak; alveolar (e.g., /d/) and labial consonants (e.g., /b/) exhibit a diffuse onset spectra, with the former rising and the latter falling in frequency, and with release bursts tending to enhance these spectral cues

(Ohde & Stevens, 1983). Later research tended to confirm the correlation between gross spectral onset shape, but raised questions about whether this information served as a primary critical feature for the place categories (e.g., Stevens & Blumstein, 1978, 1981; Blumstein & Stevens, 1979, 1980; Blumstein, Isaacs, & Mertus, 1982; Kewley-Port, 1982, 1983; Kewley-Port, Pisoni, & Studdert-Kennedy, 1983). Our research evaluates the role of the release burst and as well as dynamic changes in onset for CV syllables.

Overview of New Multidimensional, Multiple Measure Approaches

In recent years, a very few laboratories, including ours, have been working to advance our knowledge of categorical processes, and to expand the repertoire of effective research tools, by using a number of behavioral measures to carefully evaluate the nature of speech categories in a multidimensional perceptual space. The Perceptual Magnet findings of Kuhl and colleagues are probably the best known of these efforts; Kuhl used goodness, discrimination, and labeling measures to evaluate perception in different regions of a perceptual space defined in two (formant or resonant frequency) dimensions for vowels (e.g., Kuhl, 1991) and consonants (Iverson & Kuhl, 1995). The basic finding is that perceptual distance is reduced around the category prototype, thus the metaphor of a perceptual magnet. Another approach to studying categories is represented by Joanne Miller who also has used both selective adaptation magnitude (e.g., Volaitis & Miller, 1992) and goodness ratings (e.g., Hodgson & Miller, 1996) to map perception along broad ranges of physically important dimensions, as well as simple stimulus interactions (trading relation). Several researchers (e.g., Kingston & Macmillan, 1995; Macmillan, Braida, & Goldberg, 1987; Uchanski, Miller, Reed, & Braida, 1992) are employing important new approaches strongly grounded in Signal Detection Theory (SDT). Finally, Li and Pastore (1992) used goodness and similarity ratings, as well as speeded classification to evaluate prototype versus exemplar models of speech categories.

Our grant supported efforts overlap, to varying degrees, with each of these and other recent innovative approaches to studying auditory perception. Some of our work involved the perception of musical chords. Thus, Acker, Pastore, and Hall (1995) employed goodness ratings and accuracy measures to evaluate the possibility of perceptual magnet effects for musical chords; our finding of a perceptual anchor effect (opposite to a magnet effect) for musical chords provides a very important contrast to the perceptual magnet findings reported for speech by the Kuhl laboratory. Acker and Pastore (1996) then used an accuracy version of the Garner paradigm to investigate the nature of dimensional interaction for musical chords; this accuracy paradigm is less rigorously tied to SDT modeling, but also is more general than that developed by Kingston & Macmillan (1996) and less so than proposed by Ashby (1992). Acker & Pastore (under revision) also has evaluated the role of experience in the development of musical chord category. (This research is described in more detail later in this report).

Current Major Study

The major research effort under the AFOSR grant was a multi-year effort which investigated the multidimensional perceptual space for initial stop consonants (/b/, /d/, and /g/) in each of a number of vowel contexts (/a/, /ae/, /i/, /o/, and /u/). Stop consonants cannot exist in the absence of an accompanying vowel, and previous labeling research has indicated that each possible cue may play somewhat different (and largely unspecified) roles in the presence of different vowels. For each vowel, the consonant-vowel (CV) syllables was varied in a factorial manner across the three known possible cues: nature of release burst and the onset transitions to F2 and F3. For each vowel, we evaluated (within subjects) open-ended labeling [1] (or classification), goodness ratings (for each speech category), and pair-wise similarity ratings. The results of the classification and category goodness ratings are used to generate mappings of perception onto the space defined by the three physical dimensions (F2 and F3 onset frequencies and onset burst type). Similarity ratings were obtained from all possible pairings of s subset of the stimuli, with these ratings analyzed with Multidimensional Scaling (MDS) procedures to generate representations of perceptual spaces. Only those physical parameters (or combinations of parameters) which have psychological relevance will be represented in the MDS solution as perceptual dimensions; it is thus necessary to map the physical dimensions onto the perceptual dimensions. These physical dimensions are then

---

[1] The labeling (classification) is open-ended in the sense that all three consonant categories (/b/, /d/, and /g/) are allowed as responses, as well as a category for "none of the above" or "other". Using this fourth category, subjects could indicate stimuli which either belonged to none of the designated categories under consideration, or was sufficiently ambiguous that no clear category label could be applied to that stimulus.

mapped back onto the multidimensional perceptual space determined by the multidimensional scaling analyses of the similarity ratings. These processes allow us to determine which physical parameters are utilized in differentially categorizing the three consonants, as well as which of these parameters (or combinations of these parameters) are the most salient for distinguishing among the consonants.

Methods

Subjects: Each of the five experiments used a within subject design with a minimum of eight subjects completing all of the conditions (classification, goodness rating for each target consonant, similarity scaling). Subjects, who differed across experiments, were recruited from the university community using advertising signs and were paid for their time and effort. All reported normal hearing and American English to be their native language.

Stimuli: The stimuli were three formant CV syllables produced with a Klatt synthesizer program (CSRE 3.0 or 4.2). The original stimulus parameters were based upon a literature survey, reflecting those typically used in speech studies investigating initial voiced stop consonants varying in place of articulation. All stimuli were digitized (12-bit, 10 kHz sample rate) and were low pass filtered at 5 kHz. Stimulus parameters were varied systematically across the F2 and F3 onset frequencies producing a set of 27 to 30 stimuli, with the limitation that the F2 and F3 onset frequencies could not be closer together than the bandwidth of these formants. Stimulus sets were generated for the vowels /a/, /æ/, /i/, /o/, and /u/. In terms of placement of the articulators in production, the vowels /a/ and /u/ are both central, /æ/ is the most central of typical front vowels. [2] The vowel /i/ also is front, while /u/ is a high back vowel. In generating each stimulus set, considerable effort was made to make sure that the team working of the synthesis felt that the set included very good examples of each of the three target consonants (/b/, /d/, /g/).

Two additional versions of each stimulus then were created by adding an initial burst of noise corresponding to the release burst typically found at the onset or release of initial alveolar and velar stops (in labial stops, the release burst typically is weak or absent; Zue, 1976). Initial efforts used the synthesizer program to add the noise. When the resulting stimuli did not sound reasonable, we tried extracting release bursts from natural utterances, but adding these bursts to our stimulus set also produced stimuli which were heard as the CV syllable with a burst of noise occurring somewhere within the stimulus. We finally resorted to a brief (15 msec) burst of bandpass gaussian noise (2/3 octave) centered on the F2 (Low noise) and F3 (High noise) region. Adding the noise (followed by a 15 msec silent interval) resulted in sets of 87 to 96 stimuli for each vowel. For each set, pilot conditions were run with naïve subjects to insure that the set included reasonable examples of each of the three consonant categories. For several vowels, these pilot conditions resulted in either additional refinements of the stimuli, or even starting over with a new synthesis. In each experiment (defined by a given vowel), the full stimulus set was used for the labeling and goodness rating tasks.

The third task was similarity rating between pairs of stimuli. In this task, each stimulus must be presented with every other one, including itself in each sequential order. If we used a full set of 90 stimuli, we would have to run 8,100 (90^2) trials to obtain one stimulus rating, representing approximately 18-20 hours of running time per subject. We therefore samples a subset of 9 or 10 stimuli defined by F2 and F3 onset frequencies (thus, 27 to 30 stimuli when considering the factorial combination of the three release burst conditions), allowing us to collect four rating responses per subject for each pair; all possible pairing once per session over four separate sessions. The F2 and F3 values of the stimuli were selected to include clear, strong examples of each of the three consonant categories, some weak or ambiguous examples, and a distribution across the F2 and F3 onset frequency values. The results of the similarity scaling were submitted to a Kruskal Multidimensional Scaling program (which maintains ordinal relationships)with the Euclidean metric. Optimum solutions all vowels were either in two or three dimensions, although the dimensions were not always consistent across vowels. Furthermore, the dimensions seldom simply reflected each of the three physical dimensions varied.

---

[2] The references to front, back, central and high back vowels are descriptive terms used to distinguish the placement of the articulators (with consequences for the resulting resonance frequencies) during pronunciation of the vowel.

Procedure: Subjects were run either alone or in pairs in commercial sound chambers. Stimuli were presented binaurally over Sennheiser HD450 headphones. In each experiment (defined by the vowel used in the CV syllable), subjects first listened to all stimuli to allow the subjects to become familiar with the complete set. Subjects then ran the classification task where they had to label each stimulus as "b", "d", "g", or "other". There were a minimum of ten repetitions of each stimulus for each subject. Subjects then ran three goodness rating tasks, where all stimuli were presented 10 times per subject for each task. In a given rating task, subjects used a 7 point rating scale to indicate the goodness of the stimulus as a member of a specified category, with 1 indicating very poor and 7 indicating excellent. The three tasks differed in terms of the consonant category being rated (/b/, /d/, and /g/), with the order of running counterbalanced across subjects. In the final task subjects used another 7 point rating scale to indicate the similarity between the pair of stimuli presented on each trial. The stimuli were a subset of the original stimuli (see stimulus section above). In this similarity rating task, subjects first listed to all pairs to provide a basis for judging the range of similarity present in the set. Data were collected for a minimum of 4 repetitions of each pair for each subject. In all three tasks, subjects were given a brief break at least once every 15 minutes. The tasks were distributed across a number of sessions distributed across several months.

Experiment 1. Results for /u/ Vowel Context

The results for the /u/ vowel are presented in Figures 2 and 3. The upper three sets of results in Figure 2 present the classification results (ordinate of each graph is percent labeling) as a function of F2-onset frequency each component row of graphs), F3-onset frequency (abscissa of each graph) and release burst type (three columns of graphs). The results report the proportion of labels ("b" "d" "g" or other, all color and pattern coded), with the sum for each stimulus (set of four bars) summing to 100. The goodness rating results are plotted in an analogous fashion in the lower set of panels. Since category goodness was rated for each of the three major phoneme categories (/b/, /d/, /g/), there is no restriction on the sum of the three ratings values for any stimulus.

The Labeling results for the No (release) Burst condition (upper left sets of panels in Figure 2) clearly indicate that that /bu/ is heard when the F2 onset transition is rising (red bars in lower two rows of results). When the F2 transition is falling (upper two rows of results), it is the F3 transition which determines the perceived category. Specifically, a rising F3 transition (and thus a prominent middle frequency spectral peak at onset) results in /gu/ and a falling /F3/ (thus, diffuse falling onset spectra) results in /du/. When F2 is flat, F3 differentiates between /bu/ and /du/. In essence, the F2 transition differentiates /bu/ from *non-bu* consonants, whereas the F3 transition differentiates the *non-bu* category in terms of specific consonants. We see a similar pattern of results in the goodness ratings (bar graphs in the lower left panel of Figure 2), thus the F2 transition is not equally a cue for all three phoneme categories. When a low frequency release burst is added to the stimuli (middle panel), perception is shifted toward /gu/, or away from /du/; for the flat F2 stimuli (F2 = 1400 Hz), where the stimuli where were weak /du/ in the absence of any release burst now are perceived as /bu/. Since we also see an overall increase in the goodness of /gu/ for all stimuli, we suspect that the low frequency release burst is providing evidence for /gu/ and against /du/. Finally, substituting the high frequency release burst for the low burst (right panels) clearly shift perception toward /du/ (yellow bars) at all values of F2 onset frequency. This shift in perception is seen both in the classification and the goodness rating results. Thus, for this vowel context, the high frequency release burst is providing strong evidence for /du/.

The 3-dimensional MDS solution, plotted in Figure 3, accounts for 96 percent of the variance. The two sets of figures (each in two parts) in Figure 3 present two different types of coding of the stimuli to display the three dimensional solutions (dimensions 1 versus 2 on left, dimensions 2 versus 3 on right, of each pair of figures) to the Multi-Dimensional Scaling (MDS) of similarity between pairs of stimuli (a subset of 30 of the 87 stimuli represented in Figure 2).) In each panel, the solid line represents dimensional grouping based upon the specific coding; the broken lines represent either separation based upon the coding found in the other graph or a consistent, but logically impossible, breakdown.[3] The lower pair of graphs code the nature and direction of F2 and F3 onset

---

[3] In the lower half of the dimension 2 and 3 (of the MDS solution) burst- and labeling-coded graph a separation of stimuli can be seen, between /b/ and /d/ categories on the one hand and the /g/ category on the other. This separation is a continuation of that seen in the upper half of the figure, which the stimuli separate based on burst frequency (high frequency to the left and low frequency to the right). The separation of stimulus categories based upon burst type is logically impossible when the burst is absent (lower portion of figure), indicating that there must be some other basis for the distribution of stimuli along dimension 2.

transitions. The upper pair of figures plot the MDS results in terms of the nature of the release burst (color coding), the dominant labeling category (letter) and relative goodness of category membership (large upper case indicating high goodness; small lower case for lower goodness; two letters indicating approximate equal classification and goodness for the two categories; "?" indicating ambiguous). In the upper right panel (dimension 3 versus 2) indicates that dimension 3 captures the contrast between burst absent (black print, lower portion) from burst present (red and blue in upper portion). However, dimension 3 does not differentiate among the perceptual categories. Dimension 2 does provide some separation of phoneme categories, specifically the pairing of /b/ and /d/ from the pairing of /b/ and /g/. When the burst is present, dimension 2 reflects the nature of the burst. Since the separation by classification category along dimension 2 is also found where no release burst is present, the nature of the burst must be only part of the story. In the upper left panel (plotting the MDS solution for dimension 2 versus 1) we see a relatively clear separation of the three phoneme categories. Dimension 2 again reflects a separation by nature of the burst, but with the no burst stimuli mixed across this separation. The nature of the burst seems to be irrelevant to the primary dimension of the perceptual space.

The nature of dimension 1, and the missing information about the nature of dimension 2, becomes more obvious in the lower set of panels which code the same MDS stimulus space in terms of F2 and F3 onset transitions. The nature of the F2 onset transition is coded in a manner consistent with the rainbow (or circle) of colors; red and yellow are rising, etc. (see legend). The shape of the symbol indicates the nature of the F3 onset transition (rising, flat, or falling). Keeping in mind that an MDS solution can be legitimately rotated (we have not done so), it is clear that dimension 1 reflects the nature of the F2 onset transition (as indicated on the figure), while dimension 2 reflects a combination of release burst and F3 transition. This overall pattern of results is quite consistent with the classification and goodness rating results in Figure 2. The results indicate a complex interaction of the three known (but not fully understood) cues for phoneme classification (e.g., Stevens & Blumstein, 1978; Kewley-Port, 1981; Kewley-Port & Luce, 1984;, Kewley-Port, Pisoni, & Studdert-Kennedy, 1983).

Figure 1, which illustrates typical labeling and trading relations findings for phoneme investigation, is actually derived from the upper two rows of bar graphs in the upper left panel in Figure 2, but with responses limited to /du/ and /gu/ (as is typical in speech research). In contrast to such a typical speech investigation which might map one behavioral measure onto one physical dimension (either holding the value of the other dimensions constant or, in a trading relation study, sampling only two values of one of the other dimensions, as in Fig. 1), the current research provides a very much more complete picture of perception.

Experiment 2. Results for /o/ Vowel Context

The classification and goodness results for the /o/ vowel are shown in Figure 4. In many ways the results are similar to those for /u/, but the pattern of differences are not quite as strong. In the absence of a release burst, a rising F2 transition results in /b/, and a falling F2 transition results in either /d/ for falling F3 transitions or /g/ for rising F3 transitions. Also, adding a low burst enhances perception of /g/ and adding a high frequency burst both enhances /d/ and diminishes /b/. Thus the overall pattern of results is similar to that for the /u/ vowel context, but the levels of category goodness are not as strong and the incidence of use of the "other" labeling category is higher than for any other vowel context investigated.

The MDS solution, shown in Figure 5, again provides a reasonable solution in three dimensions, accounting for 93 percent of the variability. As with /u/, dimension 3 separates burst present from burst absent, and dimension 2 provides some separation of the burst present stimuli into burst type (upper right panel of Figure 5). Dimension 2 also may reflect something about the F2 and F3 formant transitions (see lower right panel). Dimensions 1 and 2, together, seem to provide some separation between a combination of /b/ and /d/ from a combination of /b/ and /g/. (upper left panel of Fig. 5), with, at best, only a complex mapping of the F2 and F3 transition on to any of the dimensions (see lower set of panels in Figure 5).

Results for other Vowel Contexts

Presentation quality figures for the /a/, /æ/, and /i/ are still being developed, except as noted, are not contained in the following portion of this report. However, summaries of findings can be provided.

3. Results for /a/ Vowel Context

Figure 6 summarizes the classification and goodness results for the /a/ vowel context. The labeling results indicate that a rising F2 transition in the absence of a release burst results in the perception of /b/, with the stimuli all achieving moderate to high goodness (4 to 6). With a falling F2 transition, the stimuli with a falling F3 transition are consistently labeled as /d/, although with more moderate levels of goodness (3 to 5). With falling F2 and rising F3 transitions, classification reflects an approximately equal mixture of /d/ and /b/, reflecting only middle values of goodness for /d/ (3 - 4) and lower values of goodness for /g/ (2 - 3). When the F2 transition is flat, the dominant labeling response if for /b/ (reflecting goodness in the range of 2-3), and with the remaining responses distributed among the three alternative response categories (d, g, and other). The F2 transition thus again seems to differentiate /b/ from "other than /b/" stimuli, and the F3 transition seems to play a small role in defining perceptual category and categorical goodness for the other (/d/ and /g/) categories.

Adding a high frequency release burst results in consistent and significant increase in perceived goodness (5 - 6) and rates of classification (90 - 100%) for /d/ for falling F2 transitions, independent of the nature of the F3 transition. The high burst does not alter the strong perception of /b/ for rising transitions, but changes perception of flat F2 transition stimuli to the /d/ category (70-80 % labeling, goodness of 4-5). Adding a Low frequency release burst also does not alter the perception of /b/ for rising F2 transitions, but changes stimuli with falling F2 transitions to /g/.

The MDS solution is summarized in Figure 7 (with out the coding by perceived category and category goodness). The solution is similar in many ways to that found for /u/. A reasonable solution can be found in two dimensions (accounting for 98% of the variance), although the solution in 3 dimensions is easier to interpret. The primary dimension again reflects the nature of the F2 transition and (although not shown) provides a very good separation of the three consonant categories. Dimension 2, or dimensions 2 and 3, reflect properties of the release burst (in the 3-dimensional solution, the dimensions reflect the presence or absence of the release burst and, when present, the nature of the burst). Thus, the major difference between the /u/ and /a/ vowel contexts is that the F3 transition seems to not play a major role in differentiating /d/ and /g/ in the /a/ context.

Experiment 4. Results for /æ/ Vowel Context

Reasonable presentation formats for the /æ/ vowel results are still being developed. In the absence of any release burst, a rising F2 transition again results in the perception of /b/, and a falling F2 transition results in /g/. It is only for relatively flat F2 transitions that F3 plays any role in perception. When F3 is falling, the stimuli are perceived as /d/ with moderate goodness (3-5). When F3 is rising, the stimuli are somewhat ambiguous between /d/ and either /b/ (lower F2 onset) or /g/ (higher F2 onsets), with middle values of goodness for the alternative categories. Adding any release burst decreases the perceived goodness and the rate of classification for /b/ [although /b/ still remains the dominant category for rising F2 transitions; responses tend to be shifted to either /g/ (low frequency bursts) or to /d/ (high frequency bursts), and not to "other". Both release bursts also enhance the classification and goodness for /g/ when F2 is sharply falling. The major effects of low frequency burst and the high frequency burst can be seen only for rising (where /b/ is dominant) and flat F2 transitions, and these effects are quite small. Thus, there seems to be a different perceptual weighting of stimulus information for the three phoneme categories in the context of /æ/.

A two dimensional MDS solution captures the separation among the phoneme categories (accounting for 90% of the variance) and reflects the pattern of results from the labeling and goodness conditions. Dimension 1 reflects the nature of the F2 transition and the separation of /b/ from rising transitions, /g/ from falling transition, and a mixture of /d/ and relatively poor /g/ in the center. Dimension 2 captures a combination of F3 transition and burst type (low versus high or missing), separating /d/ from the other phoneme categories. Moving to a 3-dimensional solution provides a separation between the presence and absence of burst, but adds little to separating the classification of the stimuli.

Experiment 5. Results for /i/ Vowel Context

Past labeling studies have often found that the cues for place categories are quite different in the context of an /i/ vowel, and, to some extent, this was the case in our study. In the absence of a release burst, a rising F2 transition again leads to perception of /b/ with goodness ranging from good to very good (4-6, with 7 indicating maximum goodness). However, a flat or falling F2 transition leads to mixed classification results, with all categories rated very low in goodness. Thus, although the F2 transition differentiates /b/ from other types of percepts, the other percepts do not correspond to good phonemes. Adding a release burst of any kind resulted in a decrease in classification and goodness of /b/ and an increase in both measures for /g/, with the goodness rating for /g/ independent of whether the burst was high or low frequency. Adding the low frequency burst did not alter the goodness rating for /d. When the burst was high frequency, there was an even greater drop in perception of /b/, and enhanced goodness and classification for /d/; perception of /d/ now was consistently stronger than /g/ for all but the steepest rising and falling F2 transitions. Thus, there is some consistency across vowel contexts in that the high frequency release burst again providing information which is positive toward /d/ and negative toward /b/, the low frequency burst providing information which is positive toward /g/, and, unless stronger cues are present (e.g., from the release burst), rising F2 transitions providing evidence for /b/. However, in contrast to the other vowel contexts investigated, the low frequency burst diminishes perception of /b/, and falling F2 transitions alone are not adequate for the clear perception of phonemes other than /b/.

The MDS scaling procedure resulted in an adequate fit of the results in two dimensions (accounting for 91% of the variance), with both dimensions reflecting properties of the release bursts. All of the no burst stimuli are in two closely spaced groups, both high on dimension 2 and either central (all ambiguous percepts) or high (all /b/ percepts) on dimension 2. All low burst stimuli are in two closely spaced groups which are both low on dimension 2 and are either central (strong /g/ percepts) or somewhat higher than central (weak /g/ and /b/ percepts) on dimension 1. The high burst stimuli are relatively closely spaced in a region which is central to dimension 2 and low on dimension 1; there is some indication of grouping (but not really separation) of /d/ and /g/ which seem to reflect the distribution of /d/ and /g/ perception found in the labeling and classification results.

Concluding Remarks

It is clear from the patterns of results that there are some broad general principles in the perception of initial stop consonants varying in place of articulation, but with each different possible cues varying in importance and specific relevance depending upon vowel context. This basic notion is not new. However, the current results provide a significantly improved understanding of the complex nature and structure of perceptual phoneme categories. The mapping provided by this work also establishes a basis for other types of investigations which should allow for the identification of the nature of processes which underlie the perception of speech and other types of complex auditory stimuli (e.g., see below study of perceptual magnet).

References

Acker, Barbara E., & Pastore, Richard E (1996) Perceptual Integrality of musical chord components. **Perception & Psychophysics, 58**, 748-761.

Acker, Barbara E., & Pastore, Richard E (in preparation) Perceptual Integrality of musical chord components II: First and second inversion chords. **Perception & Psychophysics.**

Acker, Barbara E., Pastore, Richard E., & Hall, Michael D., (1995) Within-category discrimination of musical chords: Perceptual magnet or anchor? **Perception & Psychophysics, 57**, 863-874.

Ashby, F. Gregory. (1992). Multidimensional models of categorization. In F. G. Ashby (Ed.) **Multidimensional models of perception and cognition**, pp. 449-483. Erlbaum, Hillsdale, NJ.

Blumstein, Sheila E., Isaacs, Ellyn, & Mertus, John (1982). The role of the gross spectral shape as a perceptual cue to place of articulation in initial stop consonants. **Journal of the Acoustical Society of America, 72**(1). 43-50.

Blumstein, Sheila E. & Stevens, Kenneth N. (1979). Acoustic invariance in speech production: Evidence from measurements of the spectral characteristics of stop consonants. **Journal of the Acoustical Society of America, 66**, 1001-1017.

Blumstein, Sheila E. & Stevens, Kenneth N. (1980). Perceptual invariance and onset spectra for stop consonants in different vowel environments. **Journal of the Acoustical Society of America, 67**, 648-662.

Cole, Ronald A., & Scott, Brian (1974) Toward a theory of speech perception. **Psychological Review, 81**(4), 348-374.

Delattre, Pierre C., Liberman, Alvin M., & Cooper, Frank C. (1955). Acoustic loci and transitional cues for consonant. **Journal of The Acoustical Society of America, 27**, 769-773.

Fant, Gunnar (1960). **Acoustic Theory of Speech Production.** The Hague: Mouton.

Fant, Gunnar (1973). **Speech Sounds and Features.** Cambridge, MA: MIT Press.

Halle, Morris, Hughes, George W. & Radley, J. P. H. (1957). Acoustic properties of stop consonants. **Journal of the Acoustical Society of America, 29**, 107-116.

Hodgson, Phillip, & Miller, Joanne L. (1996). Internal Structure of phonetic categories: Evidence for within category trading relations. **Journal of the Acoustical Society of America**, 565-576.

Iverson, Paul., & Kuhl, Patricia K. (1995) Mapping the perceptual magnet effect for speech using signal detection theory and multidimensional scaling. **Journal of the Acoustical Society of America, 97**, 553-562.

Kewley-Port, Diane (1982). Measurement of formant transitions in naturally produced stop consonant-vowel syllables. **Journal of the Acoustical Society of America, 72**, 379-389.

Kewley-Port, Diane (1983). Time-varying features as correlates of place of articulation in stop consonants. **Journal of the Acoustical Society of America, 73**, 322-335.

Kewley-Port, Diane & Luce, Paul A. (1984). Time-varying features of initial stop consonants in auditory running spectra: A first report. **Perception & Psychophysics, 68**, 830-835.

Kewley-Port, Diane, Pisoni, David B., & Studdert-Kennedy (1983). Perception of static and dynamic acoustic cues to place of articulation of initial stop consonants. **Journal of The Acoustical Society of America, 73**, 1779-1793.

Kingston, John & Macmillan, Neil A. (1995). Integrality of nasalization and F1 in vowels in isolation before and after oral and nasal consonants: A detection-theoretic application of the Garner paradigm. **Journal of the Acoustical Society of America, 97(2)**, 1261-1285.

Kuhl, Patricia K. (1991) Human adults and human infants show a "perceptual magnet effect" for prototypes of speech categories, monkeys do not. **Perception & Psychophysics, 50**, 93-107.

Li, Xiao-Feng., & Pastore, Richard E. (1992) Evaluation of prototypes and exemplars for a phoneme place continuum. In. M.E.H. Schouten (Ed), **Audition, Speech and Language**, Berlin: Mouton-De Gruyter, 303-308.

Liberman, Alvin M., Delattre, Pierre C., Gerstman, Louis J. & Cooper, Frank S. (1956) Tempo of frequency change as a cue for distinguishing classes of speech sounds. **Journal of Experimental Psychology, 52**, 127-137.

Macmillan, Neil A., Braida, Louis D., & Goldberg, Rina F. (1987). Central and peripheral processes in the perception of speech and nonspeech sounds. In M.E.H. Schouten (Ed.). **The Psychophysics of Speech Perception.** Boston: Martinus Nijhoff, 28-46.

Ohde, Ralph N. & Stevens, Kenneth N. (1983). Effect of burst amplitude on the perception of stop consonant place of articulation. **Journal of the Acoustical Society of America**, 74(3), 706-714.

Stevens, Kenneth N. (1980). Acoustic correlates of some phonetic categories. **Journal of the Acoustical Society of America, 68**, 836-842.

Stevens, Kenneth N. (1981) Constraints imposed by the auditory system on the properties used to classify speech sounds: Data from phonology, acoustics, and psychoacoustics. In T. Myers, J.Laver, & J. Anderson (eds.). **The Cognitive Representation of Speech.** Amsterdam: North-Holland. 61-74.

Stevens, Kenneth N., & Blumstein, Sheila E. (1978). Invariant cues for place of articulation in stop consonants. **Journal of the Acoustical Society of America, 64**(5), 1358-1368.

Stevens, Kenneth N., & Blumstein, Sheila E. (1981). The search for invariant acoustic correlates of phonetic features. In P. D. Eimas and J. L. Miller, (Eds.). **Perspectives in the Study of Speech**, (Erlbaum, Hillsdale, N.J.).

Stevens, Kenneth N., Blumstein, Sheila E., & Glicksman, Laura. (1992). Acoustic and perceptual characteristics of voicing in fricatives and fricative clusters. **Journal of the Acoustical Society of America, 91**(5), 2979-3000.

Uchanski, Rosalie M., Miller, Kathleen M., Reed, Charlotte M., & Braida, Louis D. . (1992). Effects of Token Variability on Vowel Identification. In M.E.H Schouten (Ed). **The Auditory Processing of Speech: From Sounds to Words.** NY: Mouton de Gruyter. 291-302.

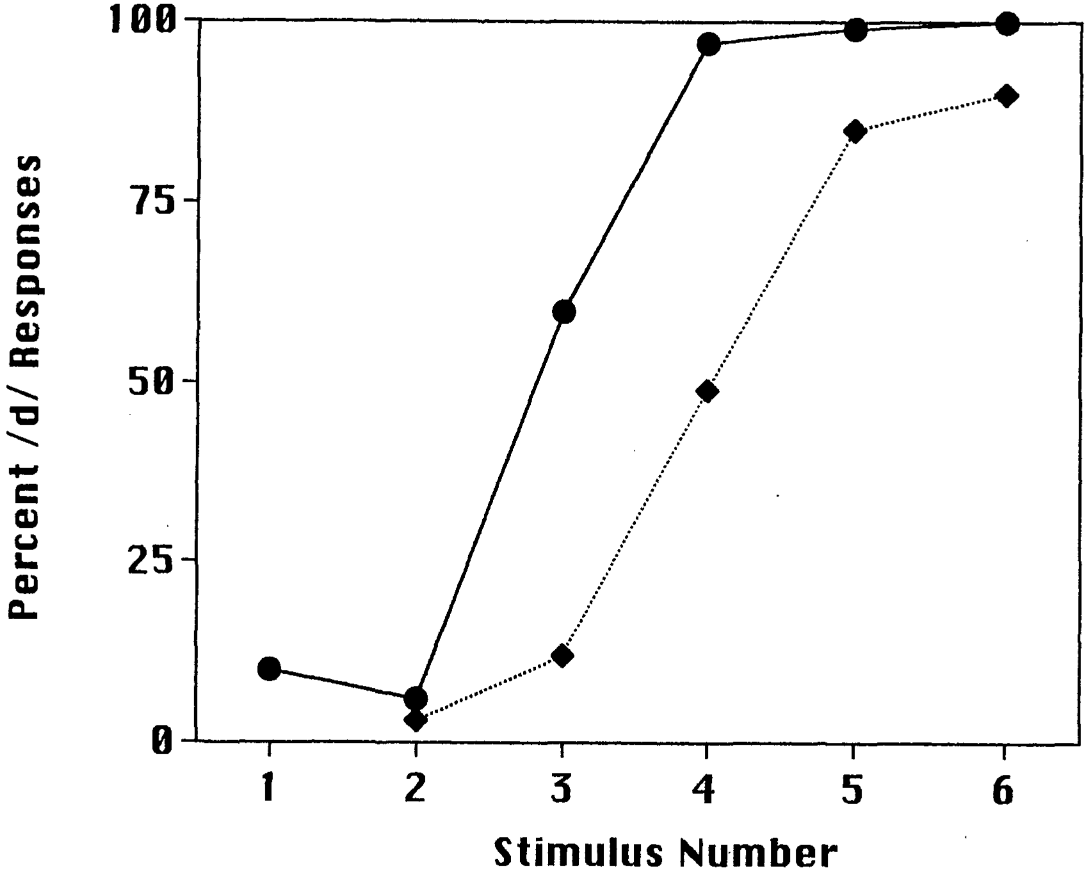
Volaitis, Lydia E. & Miller, Joanne L. (1992). Phonetic prototypes: Influence of place of articulation and speaking rate on the internal structure of voicing categories. **Journal of the Acoustical Society of America, 92**(2) 723-735.

Walley, Amanda C. & Carrell, Thomas D. (1983). Onset spectra and formant transitions in the adult's and child's perception of place of articulation in stop consonants. **Journal of the Acoustical Society of America, 73,** 1011-1022.

Zue, Victor (1976) Acoustic characteristics of stop consonants: A controlled study. Unpublished Doctoral Distribution, Massachusetts Institute of Technology (Reproduced by Indiana University Linguistics Club).

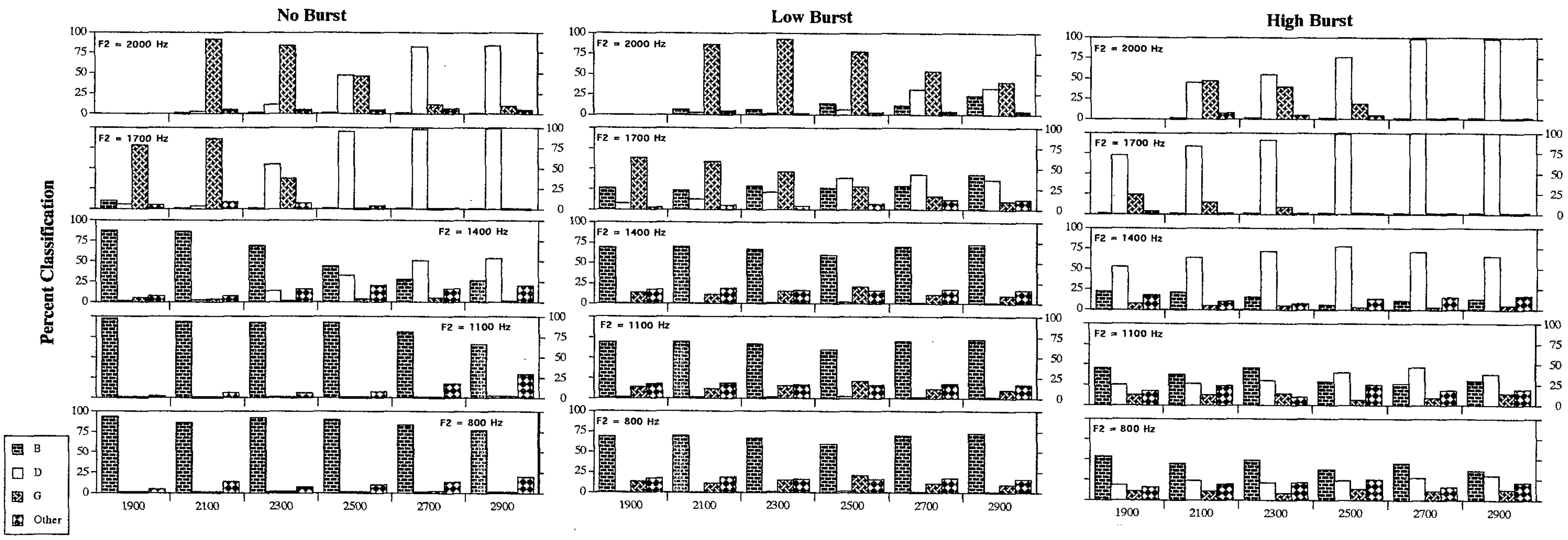
{Reports on the component experiments in this study have been presented at three different meetings of the Acoustical Society of America. The results are now being developed into a major manuscript which, once published, will be provided to AFOSR }.

Typical Speech Classification Study Results

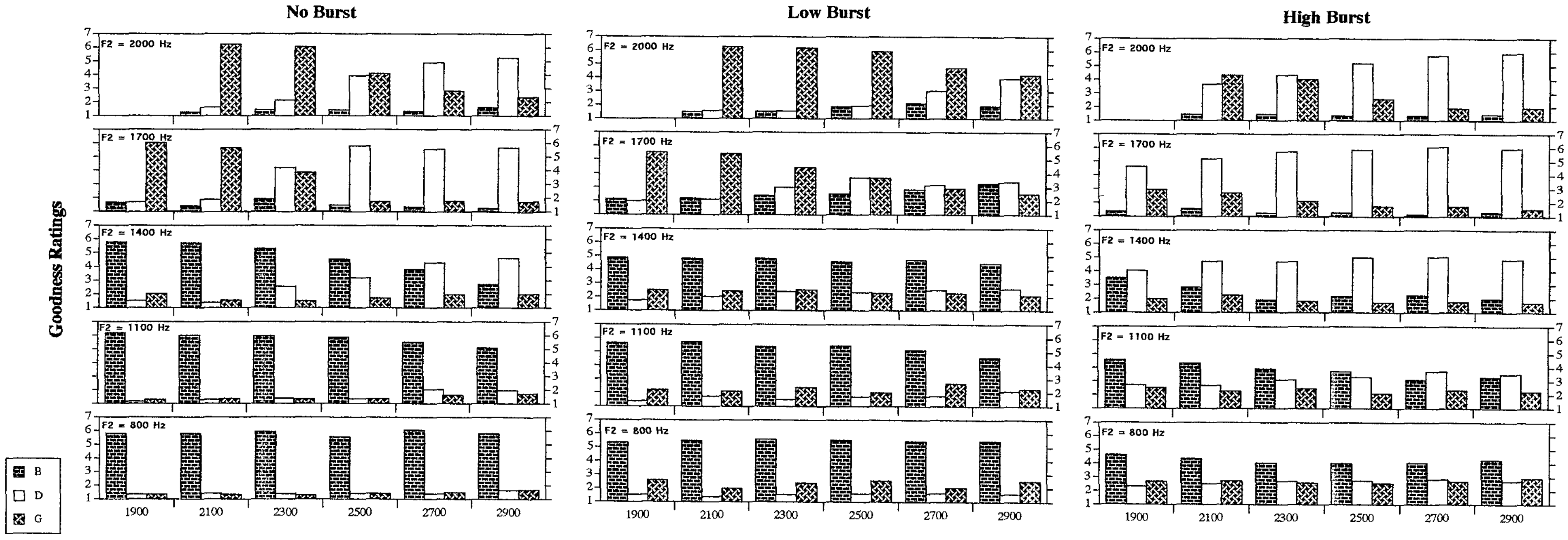


Data from No Burst /u/ vowel condition with stimulus number corresponding to F3 onset frequency. Two curves are for F2 onset frequencies of 1700 (circles) and 2000 (diamonds) Hz.

Labeling /u/ Vowel

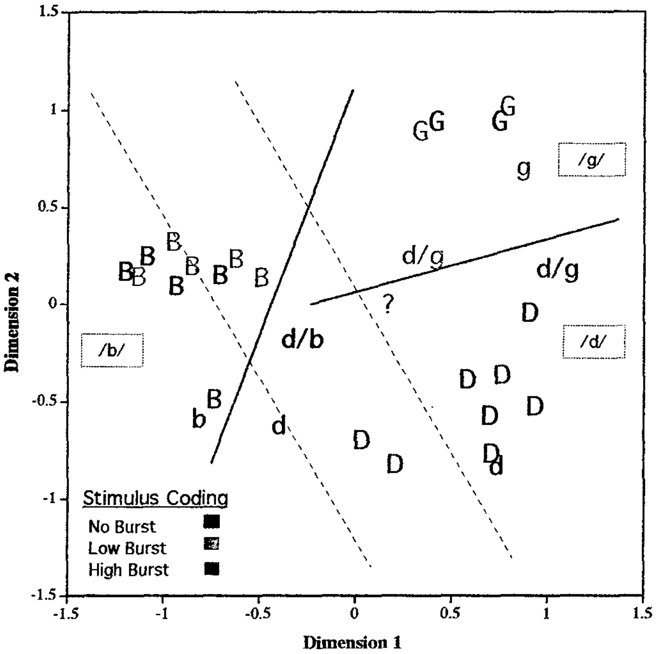


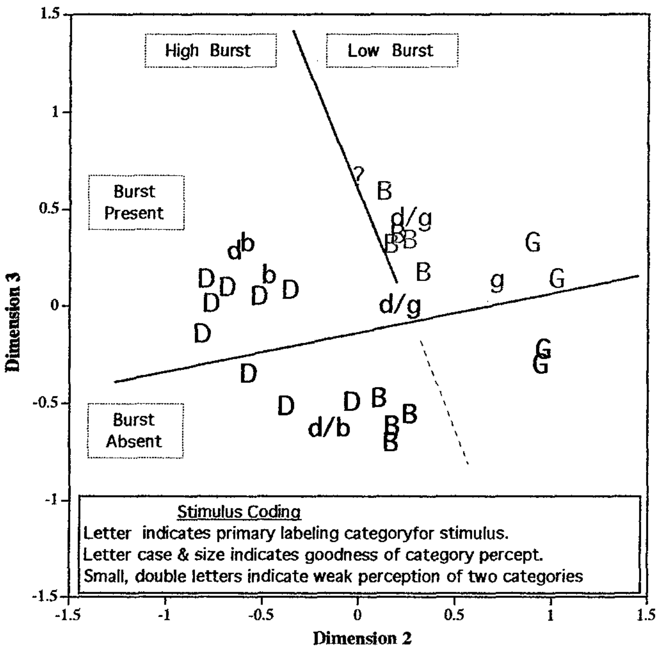
Goodness /u/ Vowel



F3 Onset Frequency (in Hz)

/u/ Vowel Similarity MDS - Coded for Burst, Labeling & Goodness





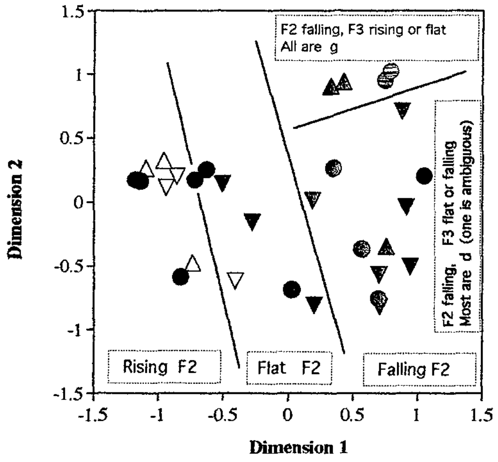
Dimension 1:
Separates B from other (D and G).

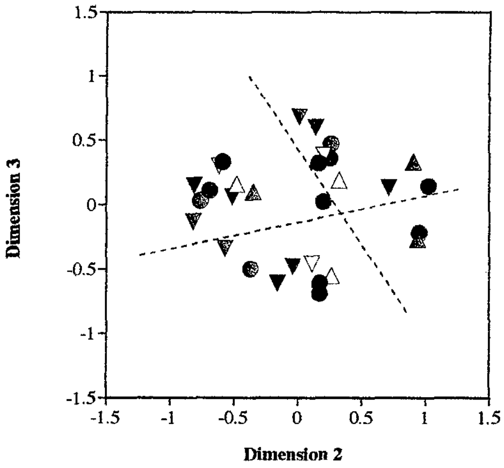
Dimension 2:
Separates other into D and G.

Dimension 2:
Separates on Low vs. High.
Separates d from g; No consequences for B.

Dimension 3 :
Codes Burst Present versus Absent.
No consequences for classification.

/u/ Vowel Similarity MDS - F2 & F3 Onset Frequency Coding





Quick Overview

F 2	Transition	F 3
■ □	Rising	Δ
■	Flat	O
▨ ■	Falling	∇

Detailed Coding {F2 / F3 Onset Freq.}

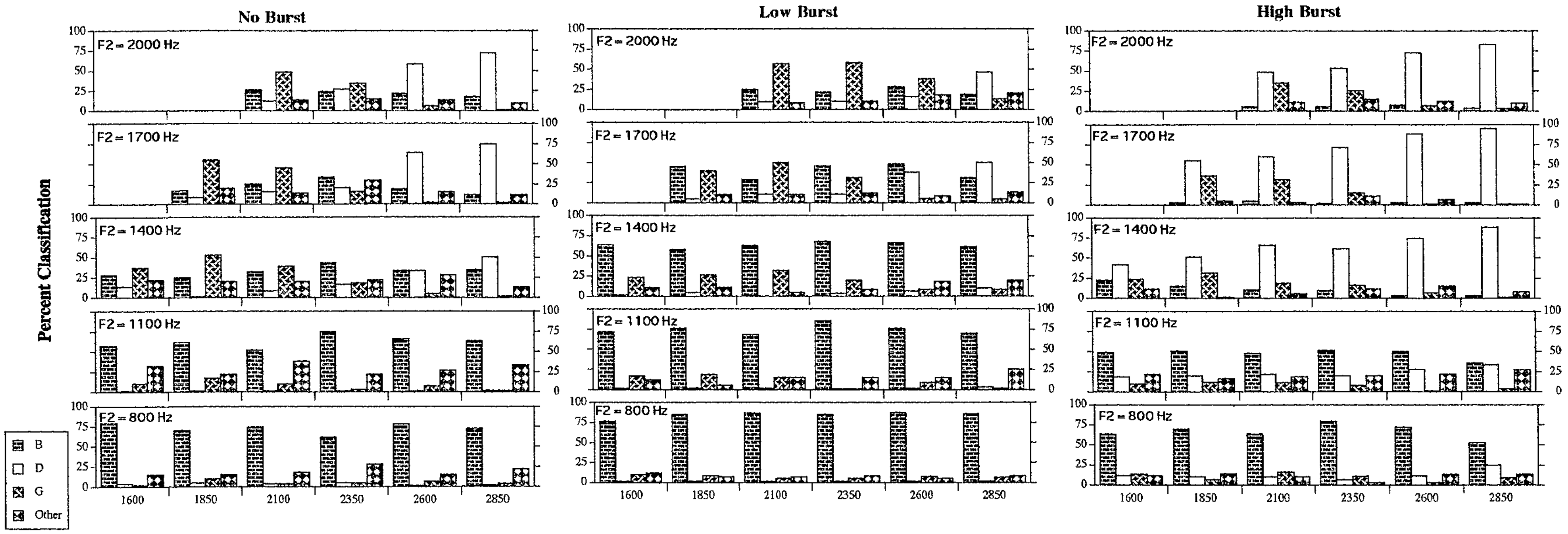
Key: F2 - F3 Onset Frequency			
●	800 - 2300	▲	1700 - 2100
△	1100 - 1900	◍	1700 - 2500
▽	1100 - 2700	▼	1700 - 2900
●	1400 - 2300	●	2000 - 2300
▼	1400 - 2700	▼	2000 - 2700

Dimension 1:
Clearly coded primarily by F2 onset frequency or transition.

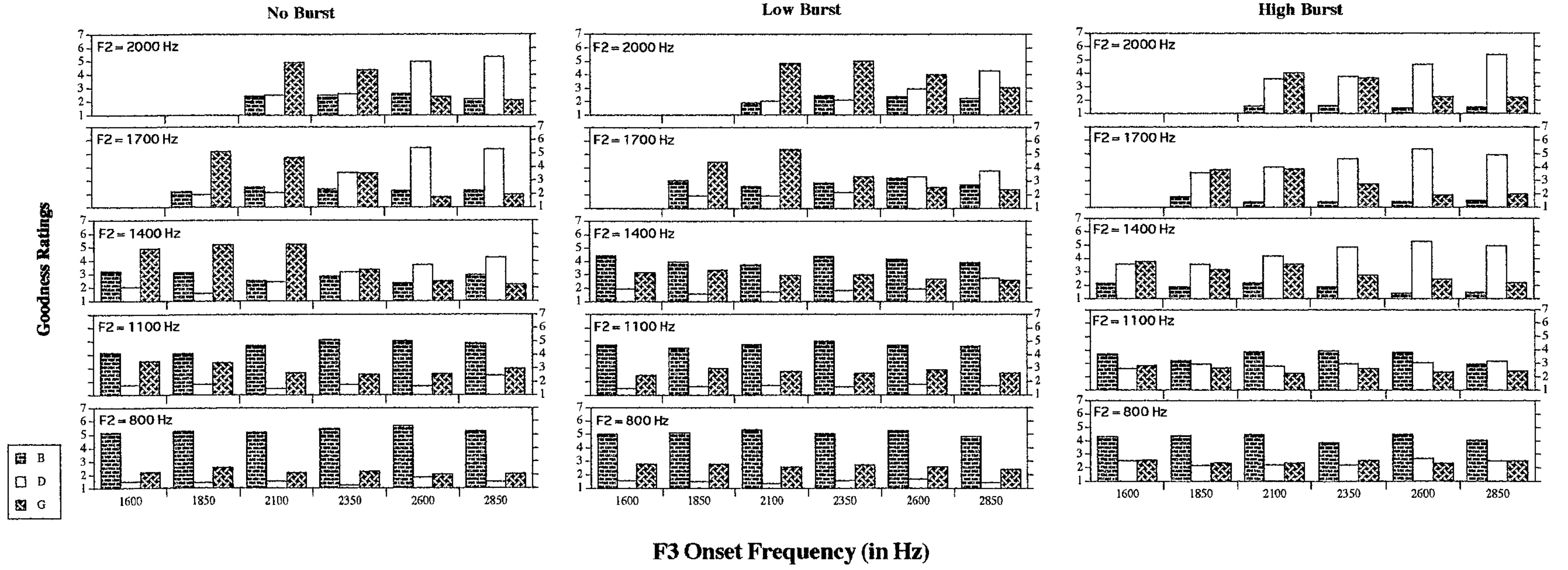
Dimension 2:
Combination of cues: Diverging F2 & F3, plus Burst type (see Coding for Burst)

Dimension 3:
Codes Burst Present versus Absent (see Coding for Burst).

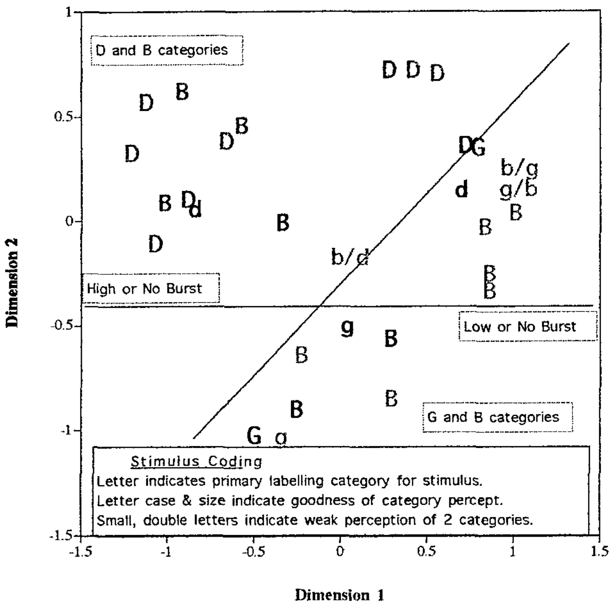
Labeling /o/ Vowel

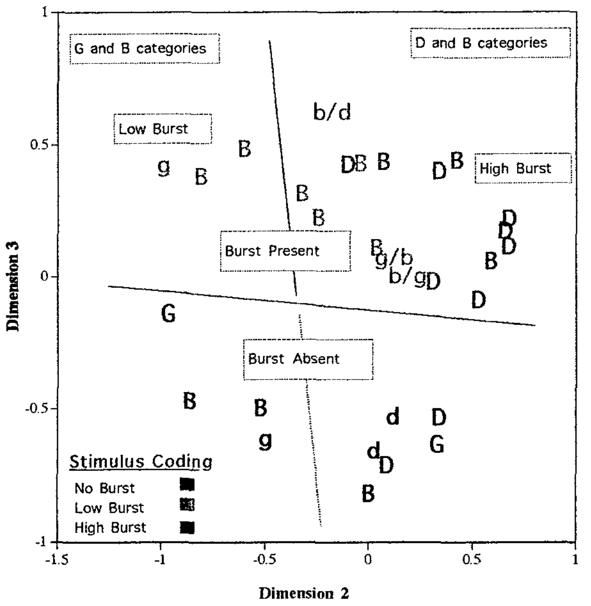


Goodness /o/ Vowel



/o/ Vowel Similarity MDS - Coded for Burst, Labeling & Goodness





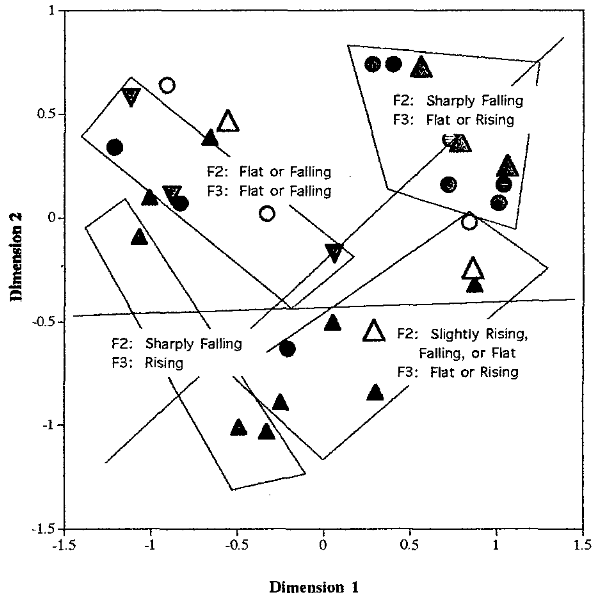
Dimension 1:
Separates Low from High Burst (with rotation, cf. Dim 1 vs. 3).

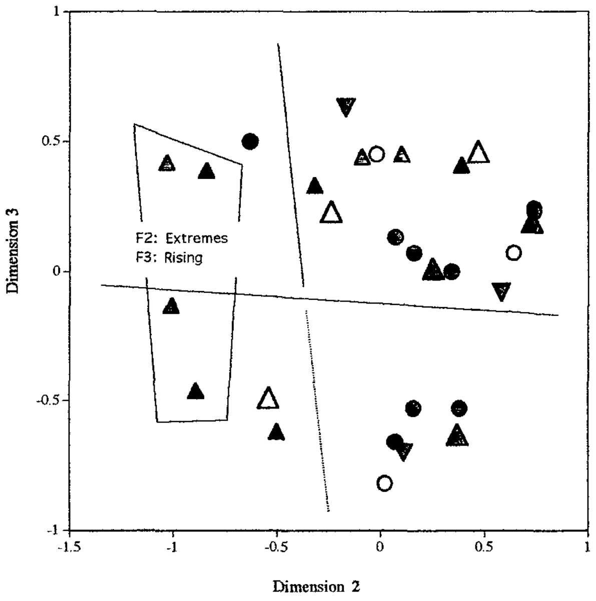
Dimension2:
Separates D from G.

Dimension 2:
Separates D from G (both when Burst Present and Absent).

Dimension 3:
Burst Present versus Absent. No consequences for classification.

/o/ Vowel Similarity MDS - F2 & F3 Onset Frequency Coding





Quick Overview

F 2	Transition	F 3
■□	Rising	Δ
■	Flat	O
▨▨	Falling	∇

Detailed Coding [F2 / F3 Onset Freq.]

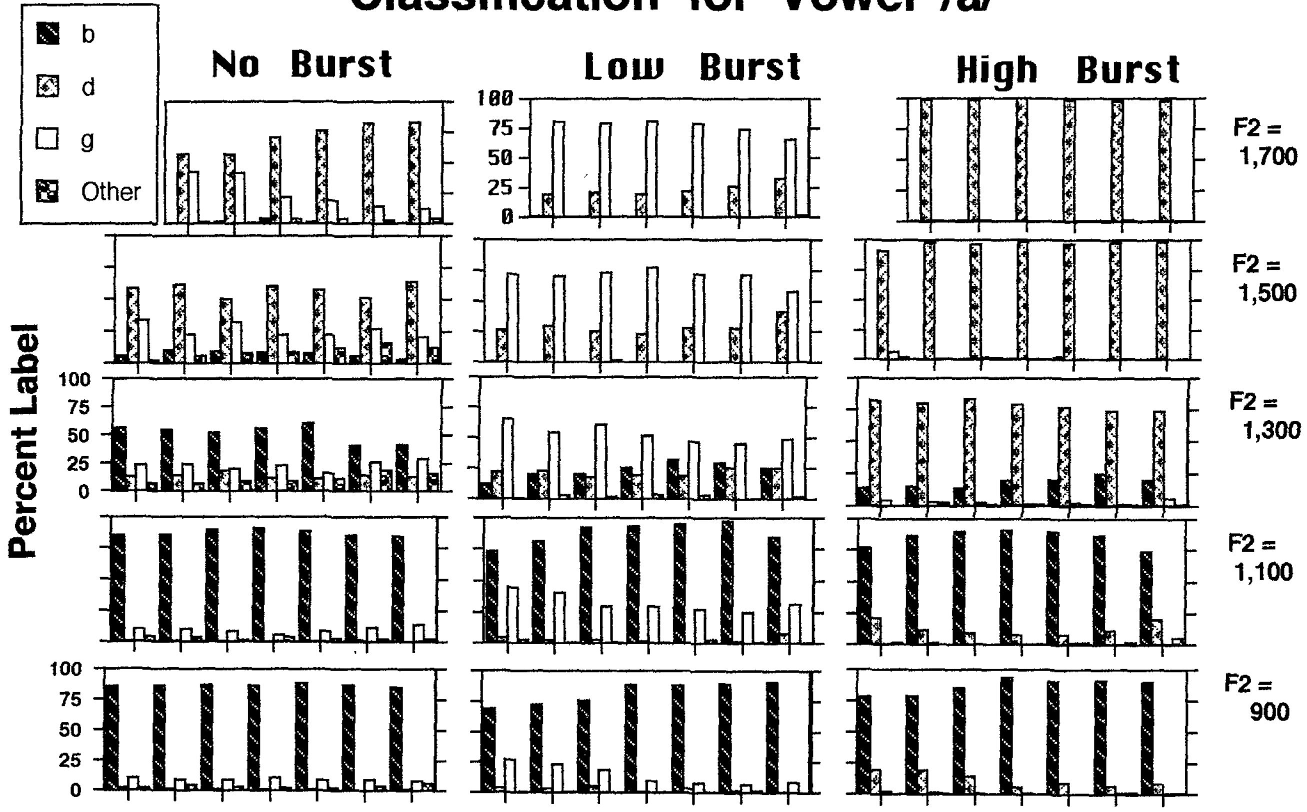
Key: F2 - F3 Onset Frequency			
▲	800 - 2100	▲	1700 - 1850
△	1100 - 1600	●	1700 - 2350
○	1100 - 2600	▼	1700 - 2850
▲	1400 - 2100	▲	2000 - 2100
●	1400 - 2600	●	2000 - 2600

Dimension1:
Codes mainly separation between Low & High Burst (See coding for Burst).

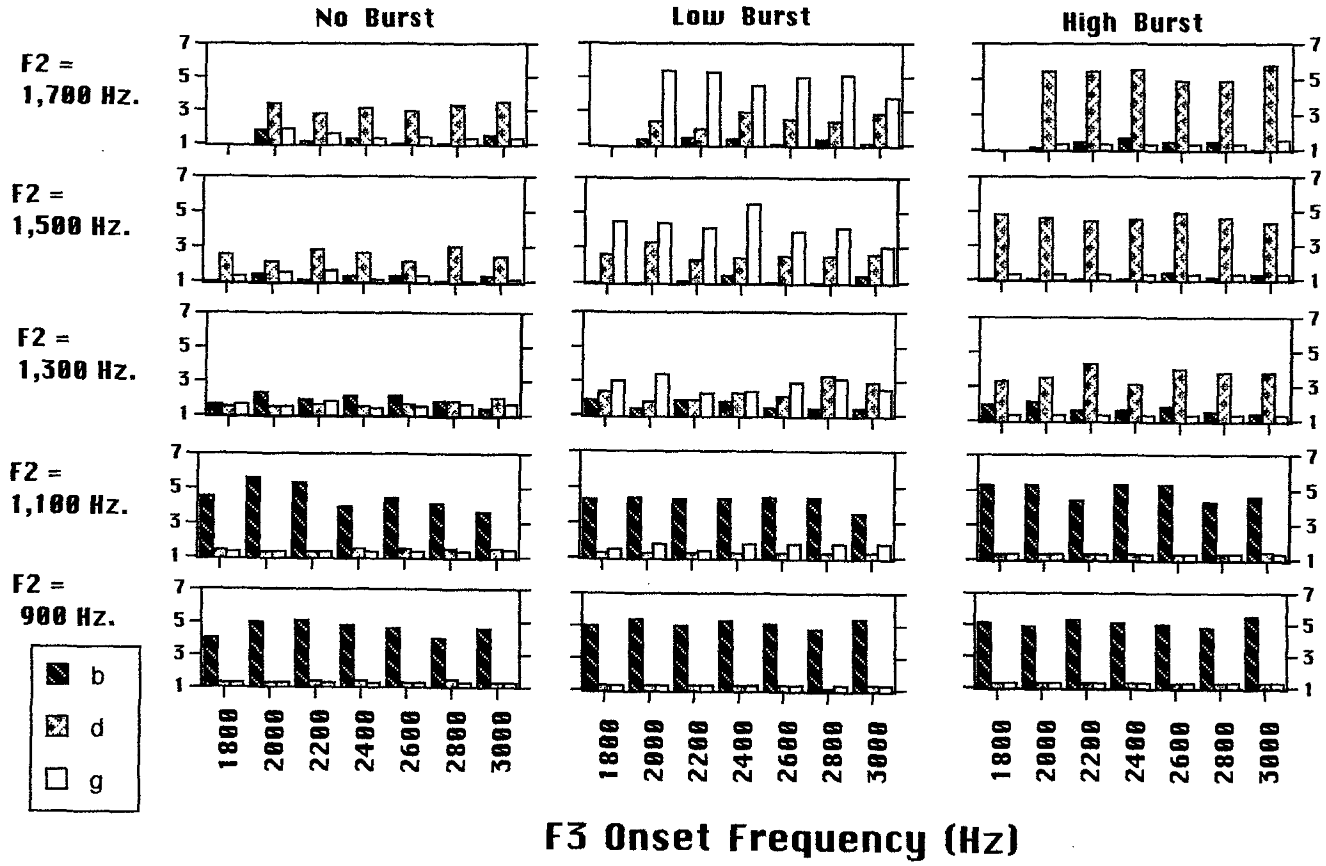
Dimension2:
Codes separation between D & G categories (See coding for Labelling). Also, extreme coordinates loosely correlated with extreme values of F2.

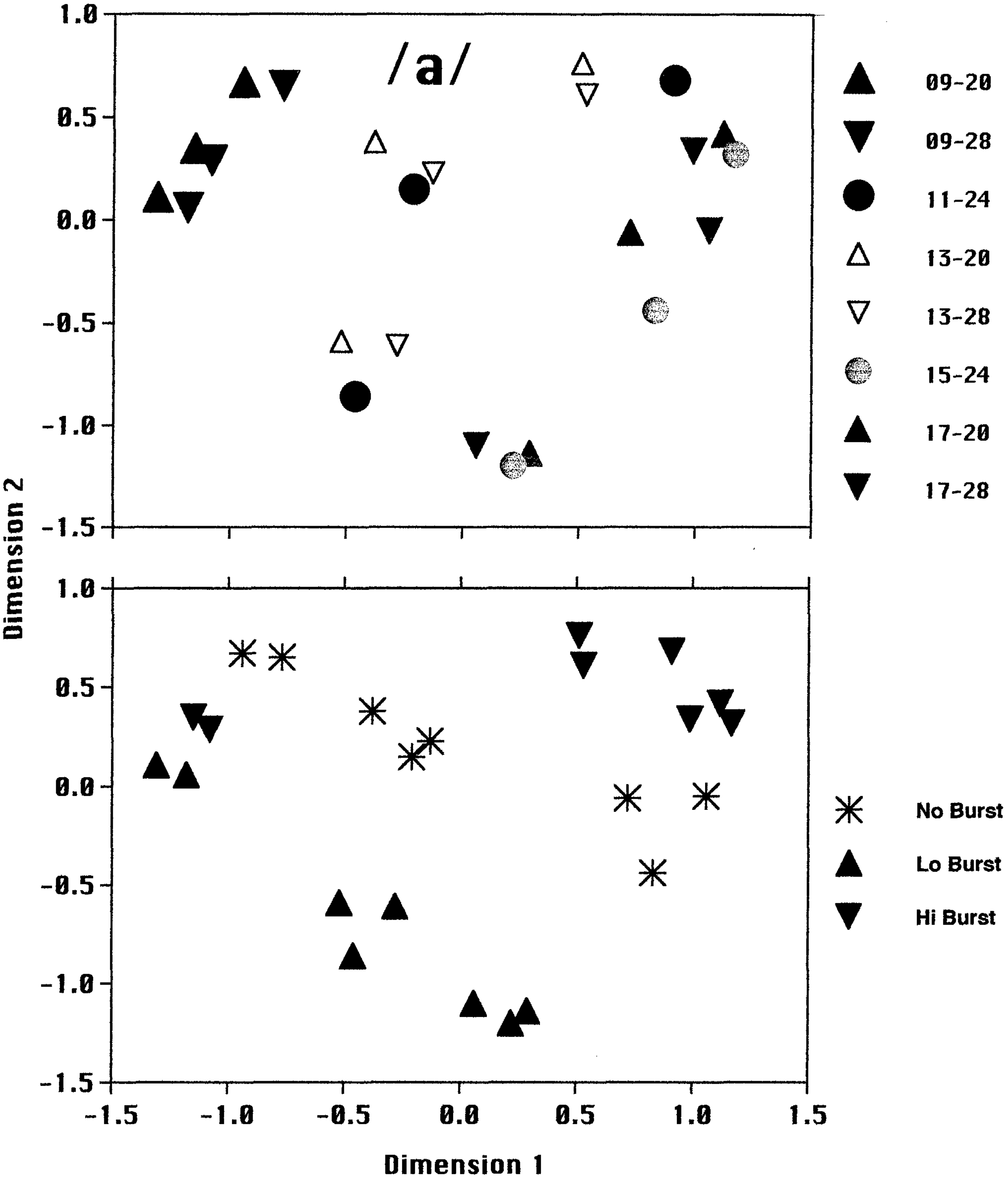
Dimension 3:
Codes Burst Present vs. Absent (see coding for Burst).

Classification for Vowel /a/



Goodness Ratings for Vowel /a/




/a/
▲ 09-20
▼ 09-28
● 11-24
△ 13-20
▽ 13-28
● 15-24
▲ 17-20
▼ 17-28
✳ No Burst
▲ Lo Burst
▼ Hi Burst
Dimension 2
Dimension 1
1.0
0.5
0.0
-0.5
-1.0
-1.5
-1.5
-1.0
-0.5
0.0
0.5
1.0
1.5

B. MULTIDIMENSIONAL STRUCTURE OF OTHER PERCEPTUAL CATEGORIES

1. PROTOTYPE FUNCTION IN MUSICAL CHORDS

Specification of the internal structure and organization of auditory perceptual categories, especially for speech sounds, has recently generated a considerable theoretical and empirical research. One important finding is that category prototypes reduce discrimination for stimuli nearby in the perceptual space (e.g., Kuhl, 1991; Iverson & Kuhl, 1995). This result also occurs in young preverbal infants who have had only passive exposure to their native language (Kuhl, 1991). Several studies in this laboratory have explored the function of musical chord prototypes - another natural, but nonspeech category. Our first study (Acker, Pastore, & Hall, 1995) evaluated musical chord category structure for musicians who had extensive formal musical training. Two sets of major chords were constructed; a "prototype (P)" set centered around an in-tune (Equal Tempered) chord and a "nonprototype (NP)" set centered around an out-of-tune chord. Each listener consistently rated one chord the highest in the P set, indicating the presence of a prototype (though the precise stimulus varied slightly across subjects), but with ratings systematically declining for stimuli around the prototype. ratings for all stimuli in the NP set were low, indicating the absence of a prototype, although stimuli closest to the prototype received somewhat higher ratings, thus indicating the influence of the prototype. Discrimination results were in contrast to the speech work; compared to the NP context, discrimination was better in the P context, with the chord prototype enhancing, not impairing, discrimination. These results show that non-speech categories also posses internal structure, but that category representations may function differently from those of speech.

References

Acker, Barbara E., Pastore, Richard E., & Hall, Michael D., (1995) Within-category discrimination of musical chords: Perceptual magnet or anchor? **Perception & Psychophysics**, **57**, 863-874.

Iverson, Paul., & Kuhl, Patricia K. (1995) Mapping the perceptual magnet effect for speech using signal detection theory and multidimensional scaling. **Journal of the Acoustical Society of America**, **97**, 553-562.

Kuhl, Patricia K. (1991) Human adults and human infants show a "perceptual magnet effect" for prototypes of speech categories, monkeys do not. **Perception & Psychophysics**, **50**, 93-107.

{A copy of the Acker, Pastore, & Hall (1995) is attached}.

2. ROLE OF EXPERIENCE / TRAINING IN DEVELOPMENT OF AUDITORY CATEGORIES

Influences of experience on the development of musical chord categories was investigated in a subsequent study based upon the same stimulus set (Acker & Pastore, under review). Separate groups of nonmusicians completed the goodness rating and discrimination tasks described above. Rating results in the P stimulus set indicated only very rough differentiation of goodness, with no one chord receiving a high rating. These results probably indicate the absence of a strong prototype for the C-major chord. Stimuli in the NP set received uniformly low ratings from the nonmusicians, with discrimination performance equivalent for the P and NP sets; these goodness ratings and discrimination results from the nonmusicians indicated a lack of category structure. The discrimination results for nonmusicians are in sharp contrast to nearly equivalent results for musicians in two studies; discrimination was not only significantly better for the P stimuli, but for the NP stimuli was no better than that for the nonmusicians. Thus, musical training improved perception of clearly tuned stimuli, but had little effect on perception of other stimuli. These results also are in contrast to the speech work with infants, where only passive exposure to the native language apparently is sufficient for the formation of strong speech sound categories. Language is a pervasive and integral part of human experience and it is probably impossible to find even young infants who have had no language exposure. Music, while somewhat perceptually pervasive (e.g., radios, Muzak), is not something that a large percentage of the population performs (or produces) and has extensive knowledge about. Thus, future work with nonmusicians and musical categories will be able to more easily determine what is required for the actual development of musical categories.

References

Acker, B.E., & Pastore, R.E. (under review) Musicians show an "anchor effect" for a major chord category, non-musicians do not. **Perception & Psychophysics.**

{A copy of the Acker & Pastore (under review) can is attached }.

3. INTEGRALITY OR SEPARABILITY OF AUDITORY FEATURES

Acker & Pastore (1996) used an accuracy version of the Garner paradigm to evaluate the perceptual integrality or separability of notes (frequencies) in root position major chord. This study demonstrated that the E and G notes in a root position C-major chord are perceived in an asymmetrically integral fashion, with subjects unable to respond separately to the notes in the chord, but with E, the frequency distinguishing between major and minor chords, contributing more to perception. Although these results stand on their own, there is an inherent confound which limits conclusions about the cause of the asymmetry. Specifically, in a root position C-major chord, the E note not only differentiates the major from minor chord, but also is lower in spectral position than the G. A subsequent study (Acker & Pastore, in preparation) manipulating the spectral position (highest, middle, or lowest tone) of the location of the E note, determined that subjects can best attend to the lowest frequency, which had the least potential for masking from the other notes. This last study demonstrated that a basic perceptual phenomena (masking) is more influential than a cognitive factor (distinguishing note) in processing individual chord components. It also provided a replication of our original perceptual anchor effect for chords (Acker, Pastore, & Hall, 1995); performance was much better for in-tune chords than for out-of-tune chords.

References

Acker, Barbara E., & Pastore, Richard E (1996) Perceptual Integrality of musical chord components. **Perception & Psychophysics**, **58**, 748-761.

Acker, Barbara E., Pastore, Richard E., & Hall, Michael D., (1995) Within-category discrimination of musical chords: Perceptual magnet or anchor? **Perception & Psychophysics**, **57**, 863-874.

Acker, B.E. & Pastore, R.E. Integrality of frequency components in first and second inversion major chords. **Perception & Psychophysics.**

{A copy of Acker & Pastore (1996) is attached. The Acker & Pastore (in preparation) manuscript is several weeks away from completion and will be provided then.}

4. CONTEXTUAL FACTORS IN THE TRACKING OF AUDITORY SEQUENCES

Recent work presented at two conferences (International Conference on Music Cognition and Perception, Acoustical Society of America) investigated context complexity on target detection in longer, more complex sequences of auditory stimuli. Listeners learned a short melody (the target) which was subsequently embedded in three line musical pieces. Two different musical contexts were created; one where the other two lines of music were harmonically static and identical to the melody in rhythmic features, and one where the other lines were more harmonically and rhythmically complex. On each trial, the presented piece contained a one note error. The musically trained subjects had to indicate if the error occurred in the pre-learned melody or in the other two musical voices. Performance generally was better when the melody was in the more complex pieces. Thus, the distinctive features of the non-melodic voices in the complex context aided in segregation of the target (the melody). Continuing research is manipulating the target by making it more distinctive (i.e. in a different instrument timbre than the other musical voices) and less distinctive (i.e. presenting the musical pieces in random timbres). The goal of the latter is to evaluate the influence of a perceptual manipulation (i.e. timbre) on higher-level representations (the pre-learned melody). Whereas these ideas are being explored with musical stimuli, the basic findings have generally applicable.

References

Acker, Barbara E., & Pastore, Richard E (1996). Melody perception in homophonic and polyphonic contexts. **Proceedings of the Fourth International Conference of Music Perception and Cognition**, Montreal, Canada: McGill University.

[A copy of Acker & Pastore (1996) is us attached.]

C. PERCEPTUAL MAGNET EFFECTS FOR CV SYLLABLES, A MULTIDIMENSIONAL APPROACH

In an attempt to demonstrate the generality of the finding of a perceptual magnet effect (described above) found for vowels, Iverson & Kuhl (1995) investigated the effects of category goodness on the perception of the American English CV contrast between /ra/ and /la/ categories. In the original vowel study, perceptual distances were found to be reduced around the best exemplars of a category relative to poor exemplars of that category, where this pattern of results is characterized using the metaphor of a perceptual magnet (Kuhl, 1991). The findings of the original vowel study have not always been replicated, and there have been assertions that the findings may simply be a different demonstration of the category boundary effect (enhanced discrimination in the region of the category boundary) studied in the 1960s and 70s. The Iverson and Kuhl CV study used perceptual identification (classification) and category goodness ratings to determine the best and worst exemplars within the /ra/ and /la/ categories as well as to determine the location of the boundary between categories. A multidimensional scaling (MDS) analysis then demonstrated results consistent with a perceptual magnet effect for the /ra/-/la/ categories. However, the use of only a small range of stimuli largely concentrated in the region of the category boundary again leaves open the very real possibility that the results reflect no more than the classic finding of enhanced discrimination (and thus perceptual distance) across the category boundary.

Past work in our lab with musical stimuli (C Major chord triads) has shown the opposite pattern of results, termed the perceptual anchor effect, where perceptual distances are greater the best exemplars of a category, and reduced around poor exemplars (see above). The study described here moves back to the speech domain, evaluating the basic pattern of findings of Iverson & Kuhl (1995). We started the experiment described here by synthesizing a set of stimuli based upon the parameters provided by Iverson and Kuhl. Because we found that the stimulus set did not contain strong examples of both categories, we decided to use a different set of CV stimuli. The stimuli for the current study were developed from those used in our multidimensional analysis of phoneme categories in the context of the vowel /u/. Because of the extensive data we had collected, we knew the locations of the category boundaries in multidimensional space and could extend the range of stimuli beyond the best category exemplars in a direction away from the category boundary. We followed a procedure similar to that used by Iverson and Kuhl, evaluating goodness ratings, paired discrimination, and similarity for stimuli within and across a /bu/-/du/ and /bu/-/gu/ contrasts, but with stimulus differences which were smaller than that used in our original study.

All stimuli were all 300 ms in length, without release bursts at onset, and varied in F2 and F3 formant onset frequencies. The two stimulus sets were based upon phonetic identification and category goodness rating results obtained previously (Pastore et al., 1996). We first conducted a phonetic identification task, in which subjects labeled which syllable (/bu/, /du/, or /gu/, or none of the above) a given stimulus sounded most like. All four were provided in order to ensure that the stimuli in each subset were members of only one of the two consonant categories comprising that stimulus set (so that there would be only one category boundary within that set of stimuli). Next, a category goodness rating task was administered, in which subjects were asked to rate on a 5-point scale (5 being an excellent exemplar of that category) how good an exemplar of a specific category each of the stimuli were. For each stimulus set, subjects were asked to rate, in separate experimental sessions, how good each stimulus was as a member of each of the two categories comprising that set. For example, for the /bu/-/du/ set, subjects rated in separate blocks of trials each stimulus as a member of the /bu/ category and as a member of the /du/ category. The third task used similarity ratings in which subjects were presented with a pair of stimuli, randomly selected from all the possible pairs of stimuli within a set, and asked to judge how similar, on a scale from 1 to 7, the stimuli were (7 being a perfect match). In the final task, subjects were presented with an AXB discrimination task, in which 3 stimuli were presented together, with either the first two (AX) or last two (XB) stimuli being identical, and the task was determining which stimulus (A or B) was the same as the middle stimulus (in pilot work we found that a same different task was very difficult for our subjects and tended to elicit strong response biases). In the discrimination task, there were two separate phases. In the first phase, the stimuli were two steps apart on the F2 onset frequency. In the second phase, they were two steps apart on the F3 onset frequency. This task was used to generate an alternate set of measures to the similarity ratings to determine the effect on perceptual distances between stimuli. Specifically, perceptual distance between two stimuli should be directly proportional to their similarity and inversely proportional to the ability to discriminate the two. The data collection phase of this research has only recently been completed and we still are in the process of analyzing the results.

References

Kuhl, Patricia K. (1991) Human adults and human infants show a "perceptual magnet effect" for prototypes of speech categories, monkeys do not. **Perception & Psychophysics**, **50**, 93-107.

Iverson, Paul., & Kuhl, Patricia K. (1995) Mapping the perceptual magnet effect for speech using signal detection theory and multidimensional scaling. **Journal of the Acoustical Society of America, 97**, 553-562.

Pastore, R.E. Liberto, J.W., & Crawley, E. J. (1996). Mapping multidimensional perceptual consonant space for place contrasts. **Journal of the Acoustical Society of America, 100(No. 4 Pt. 2)**, 2694 [Abstract].

{A manuscript will be prepared for submission later this year. Once published, copies of this manuscript will be provided to AFOSR.}

D. NATURE AND BASIS FOR SPECIFIC PERCEPTUAL CATEGORY TYPE

One long term project in our laboratory had investigated the nature and probable basis for some limitation on auditory perceptual processing which may well have significant implications for understanding aspects of a number of different types of percept, including phonemes contrasted in voicing. Initial position stop consonants with a common place of articulation can be contrasted in manner of articulation (thus, for labial stops, the phonemes /b/, /p/, and /m/ are voiced, voiceless, and nasal). Voicing contrasts (voiced versus voiceless) differ primarily along the production continuum of voice onset time (VOT) which maps on to a complex set of physical and perceptual dimensions. For labial stop, stops of American English are perceived as voiceless only when voicing onset is delayed by more than approximately 24 msec. The category boundary for alveolar (/d/ versus /t/) and velar (/g/ versus /k/) stops typically have longer category boundaries defined along the VOT continuum. Voicing contrasts are perceived categorically and VOT trades with several stimulus parameters, such as syllable duration and intensity of aspiration noise. The location of the voicing boundary (or boundaries) also differs considerably across languages.

The original idea that there may be an auditory basis for the perception of voicing contrasts stems from Hirsh (1959), and two of the earliest demonstrations of categorical perception for nonspeech stimuli (Miller, Wier, Pastore, Kelly, and Dooling, 1976; Pisoni, 1977) are based upon Hirsh's research. It is a combination of (1) misconceptions of Hirsh's findings, (2) some new research findings, and (3) a reasonable conjecture of the nature of the limitations underlying the basic phenomena which motivated our current research. Hirsh (1959) reported that there is a threshold of approximately 2 msec. for being able to detect an asynchrony in the onset of a pair of auditory stimuli and a threshold of approximately 20 msec. for being able to identify the order of onset of the stimuli. This difference of approximately 10 dB in the thresholds for detection and recognition is quite common throughout the auditory perception literature (e.g., detection versus recognition thresholds for speech in a masking noise). Hirsh conjectured that the detection threshold may have a sensory basis, but that the order threshold was probably perceptual in nature. It is the latter, perceptual temporal order threshold (hereafter, TOT) which has been conjectured to be a possible auditory basis for the perception of voicing contrasts. One misconception often found in the literature addressing temporal order and VOT is that there is only one threshold (at approximately 20 msec.) which is sensory in origin. Thus, many studies ask subjects to make a simultaneity (simultaneous-successive) judgment when studying TOT. The second misconception is the belief that Hirsh (1959) found that TOT was independent of the stimulus parameters investigated, and thus constant; any finding of a variation in TOT threshold therefore is attributed to other processes. Although many of his condition yielded threshold estimates in the 15-20 msec range (with stimuli spaced every 10 msec around onset synchrony), Hirsh found some indication that the psychometric functions may be different when the stimuli were close in frequency, when one of the stimuli was noise, and when stimuli had gradual rise times. Some of our later research provided clear demonstrations that TOT thresholds are longer when stimuli have dynamic frequency onsets and/or graduate rise times, and that TOT is a direct function of total stimulus duration. In addition, a number of studies (including our own) have demonstrated that when subjects are given extensive training with a limited set of stimuli, TOT values can be reduced to relatively small onset differences. Finally, recent work by Sinex and McDonald (1988; Sinex, McDonald, & Mott, 1991) indicated that there is a change (increase) in the synchrony of firing in auditory neurons for speech stimuli when onset asynchrony (VOT) reaches 20 to 40 msec., with this relatively peripheral interaction conjectured as possibly serving as a cue for voicing contrast and possibly TOT.

There is what may be a relatively simple explanation for TOT which also may be applicable to at least some of the different category boundaries defined along VOT continua. Very early work on the perception of sounds of varying duration demonstrated that very brief sounds (10 ms or less) are perceived as clicks, with perception moving to tone-pips (clicks with a crude pitch-like quality) as duration is increased, with pitch perceived for stimuli longer than approximately 30 msec. More recent work has indicated that pitch discrimination continues to improve with increasing duration up to approximately 100 msec. (several very recent publications by William Hartmann, as well as some older work by Brian Moore and Charles Watson, all in JASA, address these issues). These perceptual findings are consistent with the physical properties of stimuli, where the effective bandwidth of signals are inversely proportional to duration. In a typical temporal order identification task, and probably for many voicing contrasts, the listener must make a judgment of the nature of the stimulus with the earlier onset based solely upon that portion of the stimulus which occurs before the onset of the second. If the initial stimulus is a tone and it lead the second by approximately 10 msec, the listener can tell that there was an onset asynchrony, but after only 10 msec, the bandwidth of the earlier stimulus is too broad to make a reasonable judgment of its nature. According to this conceptualization, TOT reflects a limit on the quality of information necessary to perform the required recognition task. The term quality of information certainly reflects the functional bandwidth of the isolated portion of the initial stimulus which, in turn, is a function of duration or onset asynchrony. Starting from this conceptualization, it is relatively straightforward to conjecture that longer onset differences will be required for stimuli which are closer together in frequency, where one or both stimuli is broad band, or where the onsets of the stimuli are dynamically changing in frequency or amplitude. Likewise, shorter onset differences will be required where the listeners are given extensive practice with a specific set of stimuli which do not vary other than in order of onset. Finally, the findings reported by Sinex may well reflect the relationship just described; after 20 to 40 msec, the initial stimulus may have become sufficiently narrow in bandwidth to result in some firing synchrony before the second stimulus is added.

Our research provides an indirect test of these conjectures. We used two tones which were fairly close together in frequency and were long (1,000 msec), and which thus should result in relatively long values for TOT. Three different basic conditions were run, all with stimuli varying in which stimulus had the earlier onset and the amount of the onset difference. In one condition, the two tones were presented to the same ear, with the subjects required to indicate which (high or low pitch) had the earlier onset. TOT values here serve as a baseline for the other conditions. In the other two conditions the tones were presented to separate ears; these conditions thus were dichotic. In one task the subjects had to again identify which pitch had the earlier onset; judgments still had to be made on the basis of the spectral information present prior to the onset of the second (independent of ear). The values of TOT for this pitch dichotic condition should be, and were found to be, equivalent to those for the single ear condition. In the other dichotic condition subjects were asked to indicate which ear received the earlier onset (independent of pitch). In the dichotic ear condition the judgment of order could be made on the basis of where, rather than what, had the earlier onset., and the values of TOT should be, and were, significantly shorter than the pitch conditions. Finally, when ear and pitch are correlated, subjects should make responses based upon the better information, and performance was equal to that found for the dichotic ear condition.

References

Hirsh, I.J. (1959) Auditory perception of temporal order. **Journal of the Acoustical Society of America, 31,** 759-767.

Miller, J.D., Wier, C.C., Pastore, R,.E., Kelly, W.J., & Dooling, R.J. (1976) Discrimination and labeling of noise-buzz sequences with varying noise-lead times: An example of categorical perception. **Journal of the Acoustical Society of America, 60,** 410-417.

Pastore, R E. & Farrington, S.M., (1996) Measuring the Difference Limen for Identification of Order of Onset for Complex Auditory Stimuli. **Perception & Psychophysics, 58,** 510-526.

Pastore, R.E., Layer, J.K., Morris, C.B., and Logan, R.J. (1988). Temporal order identification for tone/noise stimuli with onset transitions, **Perception & Psychophysics, 44,** 257-271.

Pisoni, D.B. (1977). Identification and discrimination of relative onset time of two-component tones: Implications for voicing perception in stops. **Journal of the Acoustical Society of America, 61,** 1352-1361.

Sinex, D.G., & McDonald, L.P. (1988) Average discharge rate representation of voice-onset time in the chinchilla auditory nerve. **Journal of the Acoustical Society of America, 83**, 1817-1827.

Sinex, D.G., McDonald, L.P., & Mott, J.B. (1991) Neural correlates of nonmonotonic temporal acuity for voice onset time. **Journal of the Acoustical Society of America, 90**, 2441-2449.

{A manuscript describing these results (Crawley, Pastore, and Hinds) has been through an initial publication review and requires only relatively minor revision. A copy of this manuscript is attached. A poster describing these results was presented at the Spring 1996 meeting of the Acoustical Society.]

4. AFFILIATED PERSONNEL:

A. Faculty:

Richard E. Pastore, Ph.D., Project Director

B. Current Graduate Students who Worked on Project

Barbara E. Acker, BM, MA, MM
Edward Crawley, MA
James Liberto, BS
Lisa Gunther, BA, MA

C. Former Graduate Students who Worked on Project

Michael D. Hall, Ph.D.	Research Associate, U. Washington, Seattle, Washington
X.F. Sheldon Li, Ph.D.	Research Scientist, AT&T Bell Laboratories, NJ
Jennifer Cho, MA	Human Factors Specialist, Applications & Technology., Inc., Arlington, VA
Wenyi Huang MA	IBM Contractor, Charlotte, North Carolina
Shannon Farrington (MA pending)	IBM, San Jose, California
Sajni Jassal, MA	

D. Current Undergraduate Students who Worked on Project

Shawn Weil (on leave at Oxford University)
James Rao
Melissa A. Vindigni
Tina M. Proust
Kelly Streger
Trisha J. Widmer
Stephen T. Brush

E. Former Undergraduate Students who Worked on Project

Denise Rotavera	(10/93 to 5/94)	Now in graduate school
Laura Peyser	(10/93 to 5/95)	Now in graduate school
Ellen Hoffman	(10/93 to 5/94)	
Tracy Samuel	(8/94 to 5/96)	Now in graduate school
Linda Choi	(8/94 to 1/96)	Now in graduate school
Kevin Hinds	(1/95 to 5/96)	Now in graduate school
Elizabeth Stocum	(10/94 to 8/95)	

5. SUMMARY OF GRANT PUBLICATIONS

A. **PUBLISHED MANUSCRIPTS:** (* indicates that copy is attached)

* Acker, Barbara E., & Pastore, Richard E (1996) Perceptual Integrality of musical chord components. **Perception & Psychophysics**, **58**, 748-761.

* Acker, Barbara E., & Pastore, Richard E (1996). Melody perception in homophonic and polyphonic contexts. **Proceedings of the Fourth International Conference of Music Perception and Cognition**, Montreal, Canada: McGill University.

* Acker, Barbara E., Pastore, Richard E., & Hall, Michael D., (1995) Within-category discrimination of musical chords: Perceptual magnet or anchor? **Perception & Psychophysics**, **57**, 863-874.

* Li, Xaio-Feng., & Pastore, Richard E. (1995) Perceptual Constancy of a Global Spectral Property: Spectral Slope Discrimination. **Journal of the Acoustical Society of America**, **98**, 1956-1968.

* Pastore, R.E., & Farrington, S.M. (1996). Measuring the Difference Limen for Identification of Order of Onset for Complex Auditory Stimuli. **Perception & Psychophysics**, 58(4), 510 - 526.

Pastore, R. E. (in press). Some modern speech phenomena may be less than current beliefs. In J. Charles-Luce, P. Luce and J. R. Sawusch (Eds.), **Theories in Spoken Language: Perception, Production, and Development**. Norwood. NJ: Ablex.

B. **Manuscripts under revision:** (all require relatively simple revisions)

* Acker, B.E., & Pastore, R.E. (under review) Musicians show an "anchor effect" for a major chord category, non-musicians do not. **Perception & Psychophysics.**

* Crawley, E.J., Pastore, R.E., & Hinds, K.J. (under review) Auditory Temporal Order Thresholds for Dichotic Listening Conditions. **Perception & Psychophysics.**

* Hall, M.D., Pastore, R.E., Acker, B.E., & Huang, W. (under review) Evidence for auditory feature integration with spatially distributed items. **Perception & Psychophysics.**

C. **MANUSCRIPTS STILL IN PREPARATION:** (based upon completed Research)

Acker, B.E. & Pastore, R.E. Integrality of frequency components in first and second inversion major chords. **Perception & Psychophysics.**

Farrington, S.D. & Pastore, R.E. Perceiving Source Characteristics from Complex Natural Sounds: Walker Identification. **Journal of Experimental Psychology: Human Perception & Performance.**

Hall, M.D., & Pastore, R.E. Effects of stimulus complexity on the perceptual organization of musical tones. **Perception & Psychophysics.**

Liberto, J.W., Pastore, R.E, Huang, W., & Hall, M.D. The Octave Illusion: Exploring Dichotic Pitch Perception. **Perception & Psychophysics.**

Liberto, J. & Pastore, R.E. A Multidimensional Evaluation of the Perceptual Magnet for Consonants Contrasted in Place of Articulation. **Journal of the Acoustical Society of America**

6. INTERACTIONS & TRANSACTIONS

A. MEETING PRESENTATIONS: (Acoustical Society presentations cited in terms of published abstracts)

Acker, B. E. (1996) Compositional style, frequency height, and harmonic influences on melody perception. **Journal of the Acoustical Society of America, 100(No. 4 Pt. 2)**, 2844 [Abstract]

Acker, B.E., & Pastore, R.E. (1997) Effects of timbre manipulations on melody perception. Talk to be presented at the 133rd meeting of the **Acoustical Society of America**. State College, PA. (June 16 - 20, 1997).

Acker, B.E. & Pastore, R.E. (1996). Melody perception in homophonic and polyphonic contexts. **Fourth International Conference of Music Perception and Cognition**, Montreal, Canada, August, 1996.

Acker, B.E., & Pastore, R.E. (1996) Integrality of first inversion C-major chord components. **Journal of the Acoustical Society of America, 99**, 2481 [Abstract].

Acker, B.E., & Pastore, R.E. (1996) Directed attention and perception of frequency changes. **Journal of the Acoustical Society of America, 99**, 2482 [Abstract].

Acker, B.E., & Pastore, R.E. (1995) Major chord prototypes are based on just temperament. **American Psychological Society**, NY, NY, July 1, 1995.

Acker, B.E., & Pastore, R.E. (1995). Discrimination of musical chord components. **Journal of the Acoustical Society of America, 97**, 3391 [Abstract].

Acker, B.E., & Pastore, R.E. (1995). The role of experience in the development of category structures. **Journal of the Acoustical Society of America, 97**, 3241-2 [Abstract].

Acker, B.E., Pastore, R.E., & Hall, M.D., (1994) Within-category discrimination of musical chords: Perceptual magnet or anchor? **Journal of the Acoustical Society of America, 95**, 2937 [Abstract]

Cho, J.L., Hall, M.D., & Pastore, R.E. (1993) Stimulus properties critical to normalization of instrument timbre. **Journal of the Acoustical Society of America, 93**, 2402. [Abstract]

Crawley, E.J., Acker, B.E., & Pastore, R.E. (1997) . Ability to detect changes in musical pieces is a function of musical experience and musical context. Talk to be presented at the 133rd meeting of the **Acoustical Society of America**. State College, PA. (June 16 - 20, 1997).

Crawley, E.J. & Pastore, R.E. (1996) Dichotic temporal order thresholds. **Journal of the Acoustical Society of America**, 99, 2598 [Abstract].

Farrington, S.M. & Pastore, R.E. (1995) Auditory temporal order identification: A discrimination analysis. **American Psychological Society**, NY, NY, July 1, 1995.

Farrington, S.M. & Pastore, R.E. (May, 1996). Perceiving source characteristics from complex sounds. **Journal of the Acoustical Society of America, 99(No. 4 Pt. 2)**, 2598 [Abstract].

Hall, M.D., & Pastore, R.E. (1993). An Auditory Analog to Feature Integration. **Psychonomic Society**, Washington, D.C., Nov., 1993 [Poster Presentation].

Hall, M.D., & Pastore, R.E. (1995). Defining features of steady-state timbres. **Journal of the Acoustical Society of America, 97**, 3275 [Abstract].

Huang, W., Hall, M.D., & Pastore, R.E. (1993) An illusion based on dichotic fusion of harmonically related tones. **Journal of the Acoustical Society of America, 93**, 2316 [Abstract]

Li, XF, Pastore, R.E., & Cho, J., (1993) An exploration of phoneme structure and models of classification for place of articulation. **Journal of the Acoustical Society of America, 93**, 2390 [Abstract]

Pastore, R.E. (1993) Implicit assumptions in modeling higher level auditory processes. **Journal of the Acoustical Society of America, 93**, 2307 [Abstract of *Invited presentation.*]

Pastore, R.E. & Crawley, E. J. (1996). Dichotic temporal order thresholds. **Journal of the Acoustical Society of America, 99(No. 4 Pt. 2)**, 2598 [Abstract].

Pastore, R.E. Liberto, J.W., & Crawley, E. J. (1996). Mapping multidimensional perceptual consonant space for place contrasts. **Journal of the Acoustical Society of America, 100(No. 4 Pt. 2)**, 2694 [Abstract].

Pastore, R.E. & Farrington, S. (1995). Auditory temporal order identification: A discrimination analysis. **Meeting of the American Psychological Society,** New York City, July 1, 1995. [Poster presentation].

Pastore, R.E., Farrington, S.M., & Acker, B.E. (1994) Exploration of the phonetic structure of cues for place of articulation. **Journal of the Acoustical Society of America, 95**, 2976 [Abstract]

B. CONSULTATIVE & ADVISORY FUNCTIONS

Richard E. Pastore
Consulting Editor, **Perception & Psychophysics**
Extramural Personnel Reviewer: Tenure, Promotion to Associate Professor, Promotion to Full Professor (Institutions named upon request)
Ad hoc reviewer for peer review journals: **Journal of the Acoustical Society of America** (Psychological Acoustics, Speech Communication), **Perception & Psychophysics, Psychological Science.**

Barbara E. Acker
Ad hoc reviewer for **Perception & Psychophysics**

C. TRANSACTIONS

Richard Pastore & Barbara Acker, Co-chairs. **Speech and music: Exchange of ideas, methods, and findings.** Special session organized for the 131st meeting of the Acoustical Society of America, May, 1996. This session, recommended by members of an Acoustical Society Technical Committee, was motivated by reports of completed research projects.

7. NEW DISCOVERIES, INVENTIONS, PATENT DISCLOSURES

No inventions or patent disclosures.

8. HONORS AND AWARDS

A. Richard E. Pastore
1. Past year
   Consulting Editor, **Perception & Psychophysics.**

2. Lifetime
   Fellow, American Psychological Association, Division 3.
   Fellow, American Psychological Society.

B. Barbara E. Acker
Nominated by Psychology Department for University 1994 Excellence Award
Nominated by Psychology Department for 1997 Dissertation Year Fellowship Award

SUMMARY of ATTACHED MANUSCRIPTS

Acker, Barbara E., & Pastore, Richard E (1996) Perceptual Integrality of musical chord components. **Perception & Psychophysics, 58**, 748-761.

Acker, Barbara E., & Pastore, Richard E (1996). Melody perception in homophonic and polyphonic contexts. **Proceedings of the Fourth International Conference of Music Perception and Cognition**, Montreal, Canada: McGill University.

Acker, B.E., & Pastore, R.E. (under review) Musicians show an "anchor effect" for a major chord category, non-musicians do not. **Perception & Psychophysics.**

Acker, Barbara E., Pastore, Richard E., & Hall, Michael D., (1995) Within-category discrimination of musical chords: Perceptual magnet or anchor? **Perception & Psychophysics, 57**, 863-874.

Crawley, E.J., Pastore, R.E., & Hinds, K.J. (under review) Auditory Temporal Order Thresholds for Dichotic Listening Conditions. **Perception & Psychophysics.**

Hall, M.D., Pastore, R.E., Acker, B.E., & Huang, W. (under review) Evidence for auditory feature integration with spatially distributed items. **Perception & Psychophysics.**

Li, Xaio-Feng., & Pastore, Richard E. (1995) Perceptual Constancy of a Global Spectral Property: Spectral Slope Discrimination. **Journal of the Acoustical Society of America, 98**, 1956-1968.

Pastore, R.E., & Farrington, S.M. (1996). Measuring the Difference Limen for Identification of Order of Onset for Complex Auditory Stimuli. **Perception & Psychophysics**, 58(4), 510 - 526.