Effects of Emphatic Stress and Speaking Mode on Articulatory Organization

H. Timothy Bunnell and James Polikoff

Speech Processing Laboratory, A.I. duPont Institute, Wilmington, DE 19899

An adult male talker recorded multiple repetitions of short nonsense sentences. Each sentence was of the form “A huC1 ate a C2uffle” where C1 and C2 were from the set {/b/, /d/, /g/}. Sentences containing all 9 combinations of C1-C2 pairings were recorded in several speaking modes: CLEAR, CONVERSATIONAL, STRESS1, and STRESS2. The latter two conditions entailed production of the sentences in conversational mode, but with emphatic stress on the syllable containing either the first (STRESS1) or the second (STRESS2) variable consonant. These materials were recorded at the University of Wisconsin Microbeam facility and include both acoustic data and tracings of the trajectories of flesh points on the tongue, lips, and jaw. Analyses of the acoustic and articulatory data suggest that local effects of stress on consonant articulation are similar to those of clear speech. The largest acoustic and articulatory differences were observed for consonants in syllable final position. The various significant articulatory effects of stress and speaking mode will be compared to perceptual measures of consonant intelligibility.

Telephone: (302) 651-6830, (H. T. Bunnell)

Overall Objectives:

1.) Examine differences in stop consonant articulation associated with two speaking modes (clear and conversational) and two levels of emphatic stress (stressed and non-stressed).

2.) Observe relationship between mode-related articulatory effects in production and stop consonant perception.

Articulatory Data Analysis

Method:

1.) Stimuli were short sentences of the form: “A hu<u>C</u> ate a <u>C</u>uffle” where the variable consonant was /b/, /d/, or /g/ in all nine combinations.

2.) Each of the nine sentence types was repeated several times in each of four speaking mode conditions: CLEAR; CONVERSATIONAL; STRESS_1 (hu<u>C</u> stressed); and STRESS_2 (<u>C</u>uffle stressed). In all, 108 different sentences provided approximately three repetitions of each sentence in each speaking mode condition.

3.) Audio (20 kHz sampling rate) and simultaneous X-ray microbeam pellet tracings were recorded digitally for each token.

4.) Pellets were placed at three tongue locations (tongue tip, tongue body, and tongue dorsum), the mid-line vermilion border of the upper and lower lips, and at the base of a mandible incisor.

5.) Digitized X and Y coordinates of pellet locations over time were filtered at 12 Hz to remove noise.

6.) Segments were located and marked for further analysis using a graphic display program.

7.) Scaled data from marked regions were averaged, coded for STOP consonant, speaking MODE, syllable POSITION, and XY coordinate.

8.) Velocities were estimated from filtered position data using the central difference method.

9.) Peak velocities leading into (CLOSING velocity) and out of (OPENING velocity) marked consonant closure regions were located automatically and corrected manually when needed.

10.) Data were coded as per position data with the addition of a GESTURE (opening or closing) factor.

Results:

1.) Many of the “classical” clear speech versus conversational speech differences were observed. E.g., Figure 1 shows rate difference, syllabification of final consonant, and apparent target consonant location differences.

2.) Figure 2 illustrates other consonant articulator position differences associated with different speaking modes.

3.) Velocity data also showed several significant differences associated with speaking mode. Figure 3.

Factor Analysis

Method:

1.) All usable pellet location data was equated for sampling rate and submitted for principle components factor analysis to determine components which specify the X-Y locations of all six pellets for each sampling epoch.

2.) The resulting principle component factors were used to reduce the dimensionality of articulatory data in additional analyses.

Results:

1.) Figure 4 illustrates the principal components determined from these data. The first four components account for about 86% of the variance in pellet locations.

2.) Figure 5 plots the locations of each consonant token (measured during the closure interval) in a two dimensional factor space.

Perception Experiment

Method:

1.) Stimuli adjusted for equal long-term RMS amplitude.

2.) Noise masking added to provide -4dB SNR re the long-term average RMS amplitude.

3.) Following one presentation of each sentence, listeners identified both of the nonsense target words by touching two of the six response alternatives (hub, hud, hug, buffle, duffle, guffle) displayed on a touch sensitive terminal screen.

4.) There were 7 volunteer listeners each of whom responded to stimuli in two 30 minute sessions. In all, each listener heard each sentence 7 times.

Results:

1.) The main effect of speaking mode was significant with CLEAR and STRESSED tokens identified correctly more often than CONVERSATIONAL and UNSTRESSED tokens. Figure 6.

2.) When RMS amplitude of the target syllable was used as a covariate, the effect of speaking mode was not significant. However, effects of speaking mode continued to interact significantly with other factors. Figure 7.

Articulatory - Perceptual Relations

Purpose:

1.) Preliminary look at relationships between articulatory factors and consonant intelligibility.

Method:

1.) The dependent measure was the arcsine transformed proportion of correct identification responses for each consonant token.

2.) Independent measures were the RMS amplitude of the vocalic portion of the target-bearing syllable and the average values of each of the first four pellet factors in the three regions shown in Figure 1. Thus, there were a total of 13 predictor variables that could enter the regression equation.

Results:

1.) Amplitude Effects - Obviously loudness is distributed differently over a sentence in different speaking modes. Stressed syllables tended to be louder than the same syllable in unstressed or conversational mode. Figure 8 shows how well target syllable amplitude alone predicts consonant identification. In this figure, RMS amplitude predicts about 22% of the variation in arcsine transformed consonant identification scores.

2.) Articulatory Effects - Independent of amplitude effects, articulatory differences also predict differences in consonant identification. Figure 9 plots predicted versus obtained identification scores (arcsine transformed) when RMS and additional articulatory factors are both included. In this figure, the multiple linear regression with 10 and 175 degrees of freedom accounts for approximately 52% of the variance. This model was applied across all consonants in all locations and speaking modes.

3.) Analysis of the terms participating in the best regression models suggests that assimilative rather than contrastive articulatory effects lead to improved identification of consonants in noise. That is, smaller differences in articulator positions between the consonant intervals and preceding or following vowels tend to be associated with better consonant identification.

Conclusions

I. In the Articulatory Domain:

1.) The present talker produced CLEAR and STRESSED syllables in generally similar ways that differed from those used in conversational speech or with unstressed syllables.

2.) Some effects, notably fronting gestures associated with /d/ production suggested a more stereotypic consonant target location for those productions.

3.) Factor analysis may provide a useful way of reducing the dimensionality of microbeam pellet data.

II. In the Perceptual Domain:

1.) The present talker was not uniformly successful at increasing the perceptual clarity of target consonants.

2.) For /d/ and /g/, intelligibility in syllable-final position was generally greater than in syllable-initial position.

III. Articulatory and Perceptual Domains:

1.) Coarticulatory effects may be related to consonant intelligibility in noise with <u>greater</u> degrees of coarticulation associated with higher intelligibility.

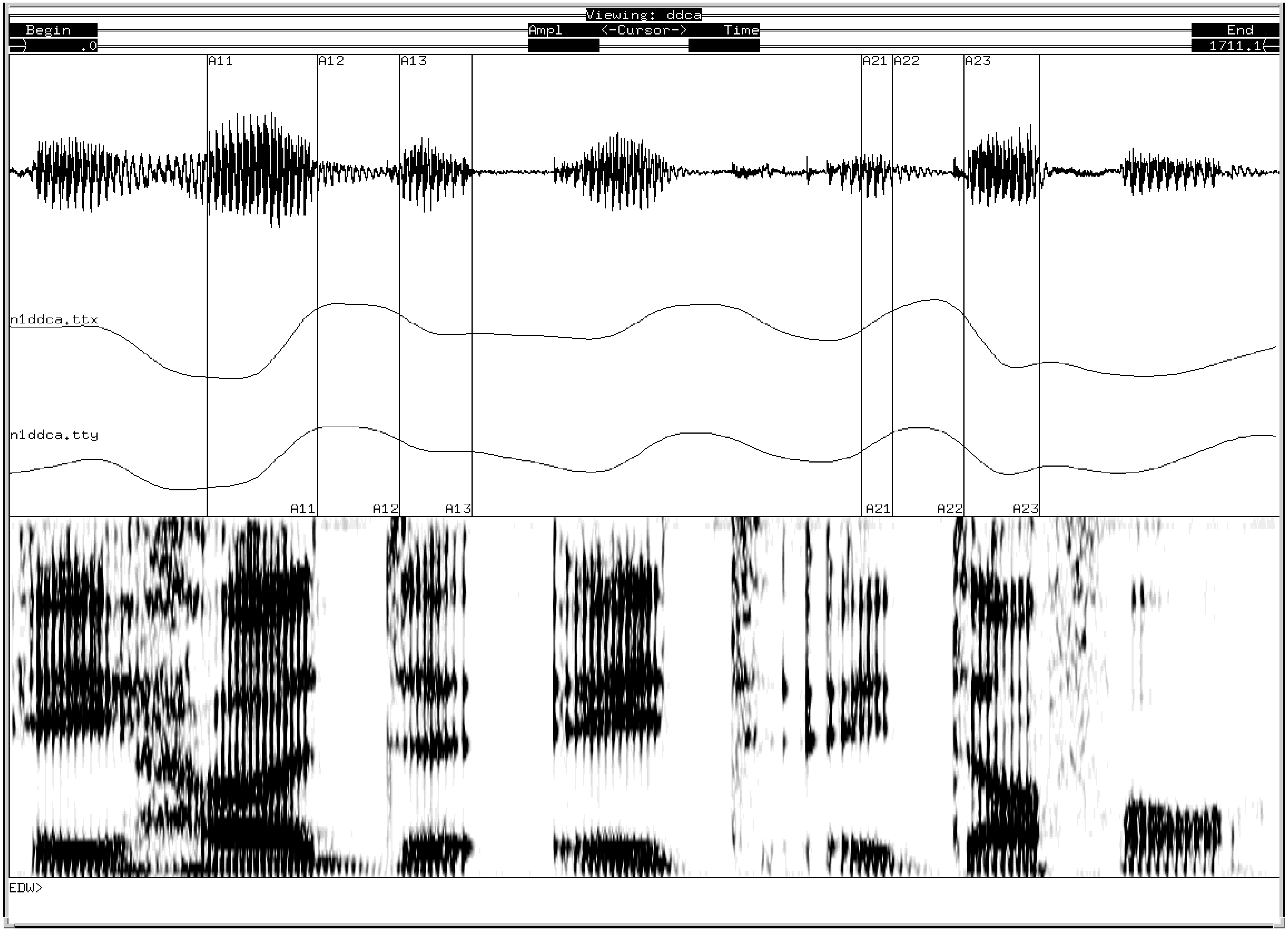
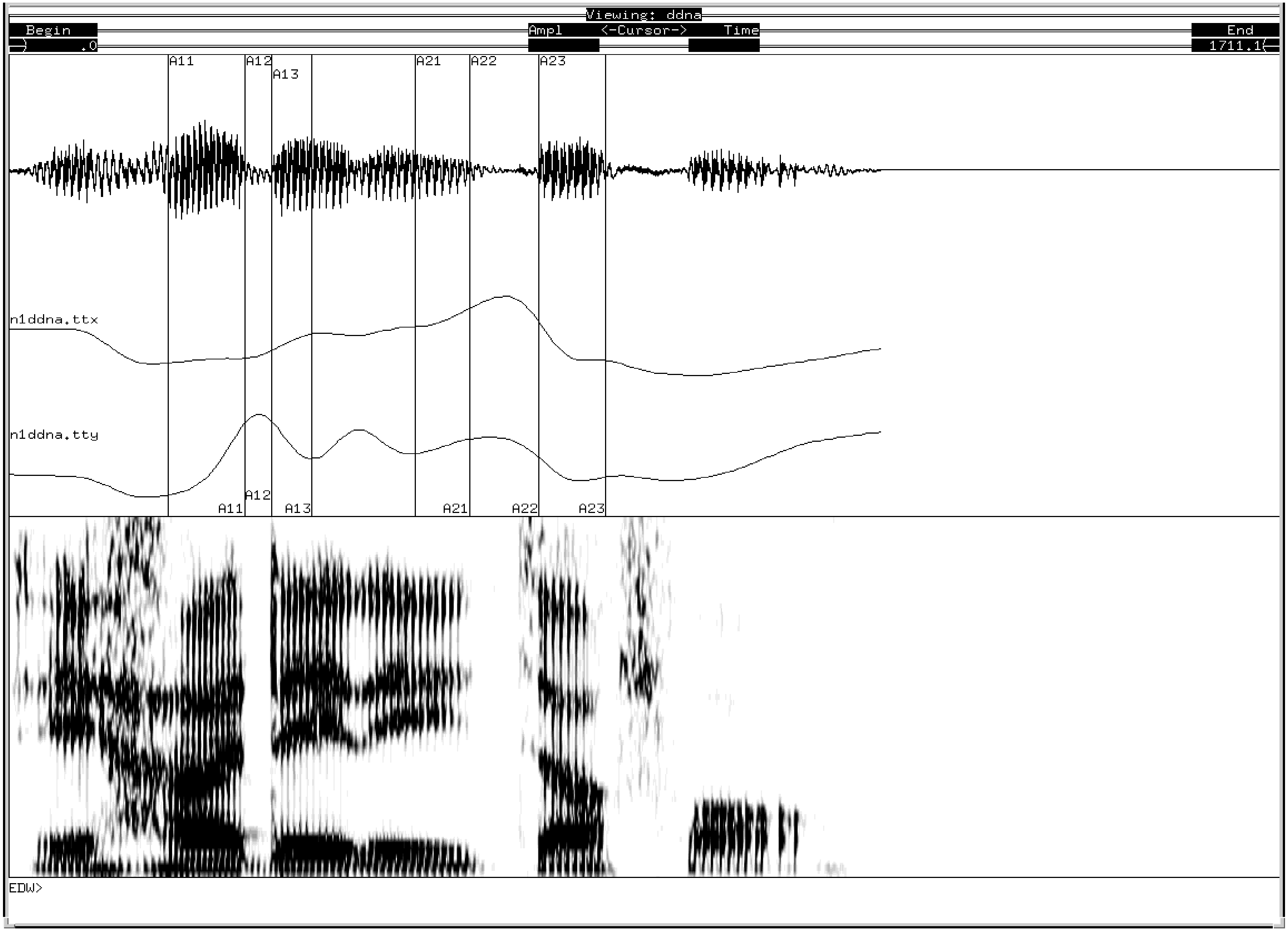

Viewing: ddca
Begin
Ampl
<-Cursor->
Time
End
.0
1711.1
A11
A12
A13
A21
A22
A23
n1ddca.ttx
n1ddca.tty
A11
A12
A13
A21
A22
A23
EDW>

Figure 1a



Viewing: ddna
Begin
Ampl
<-Cursor->
Time
End
.0
1711.1
A11
A12
A13
A21
A22
A23
n1ddna.ttx
n1ddna.tty
EDW>

Figure 1b

Pellet Locations During Maximum Closure

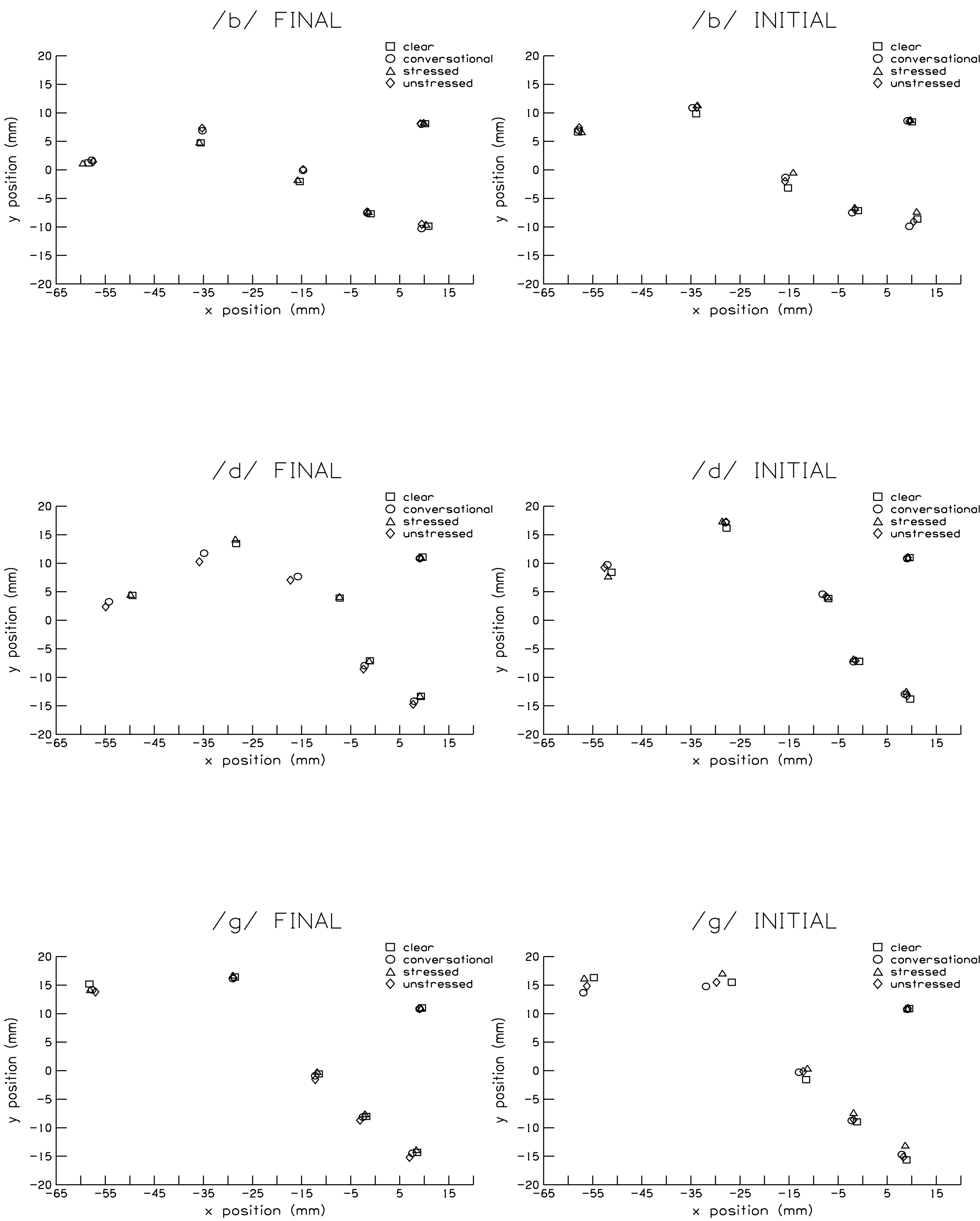


Figure 2

Displacement X Velocity

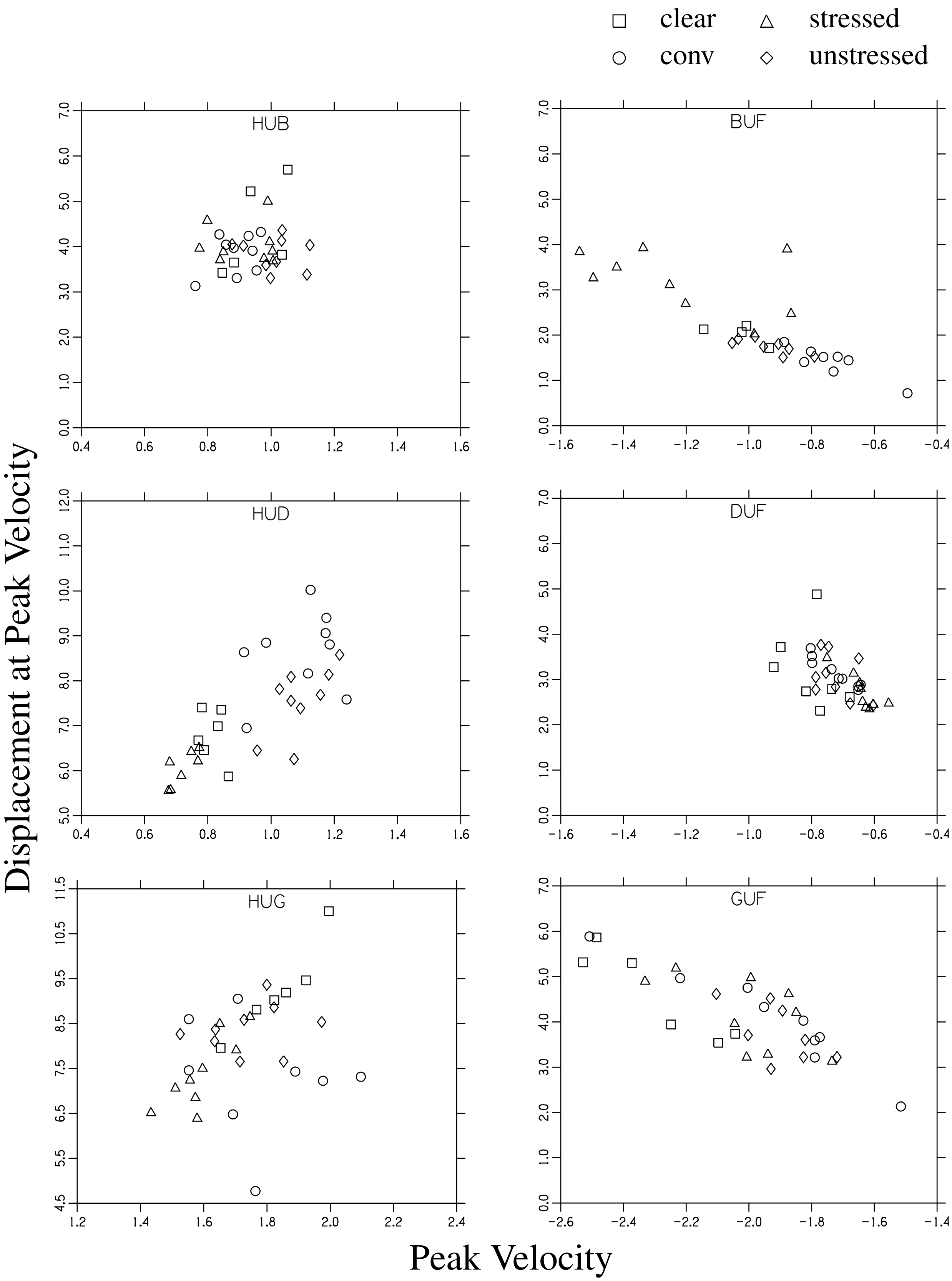


Figure 3

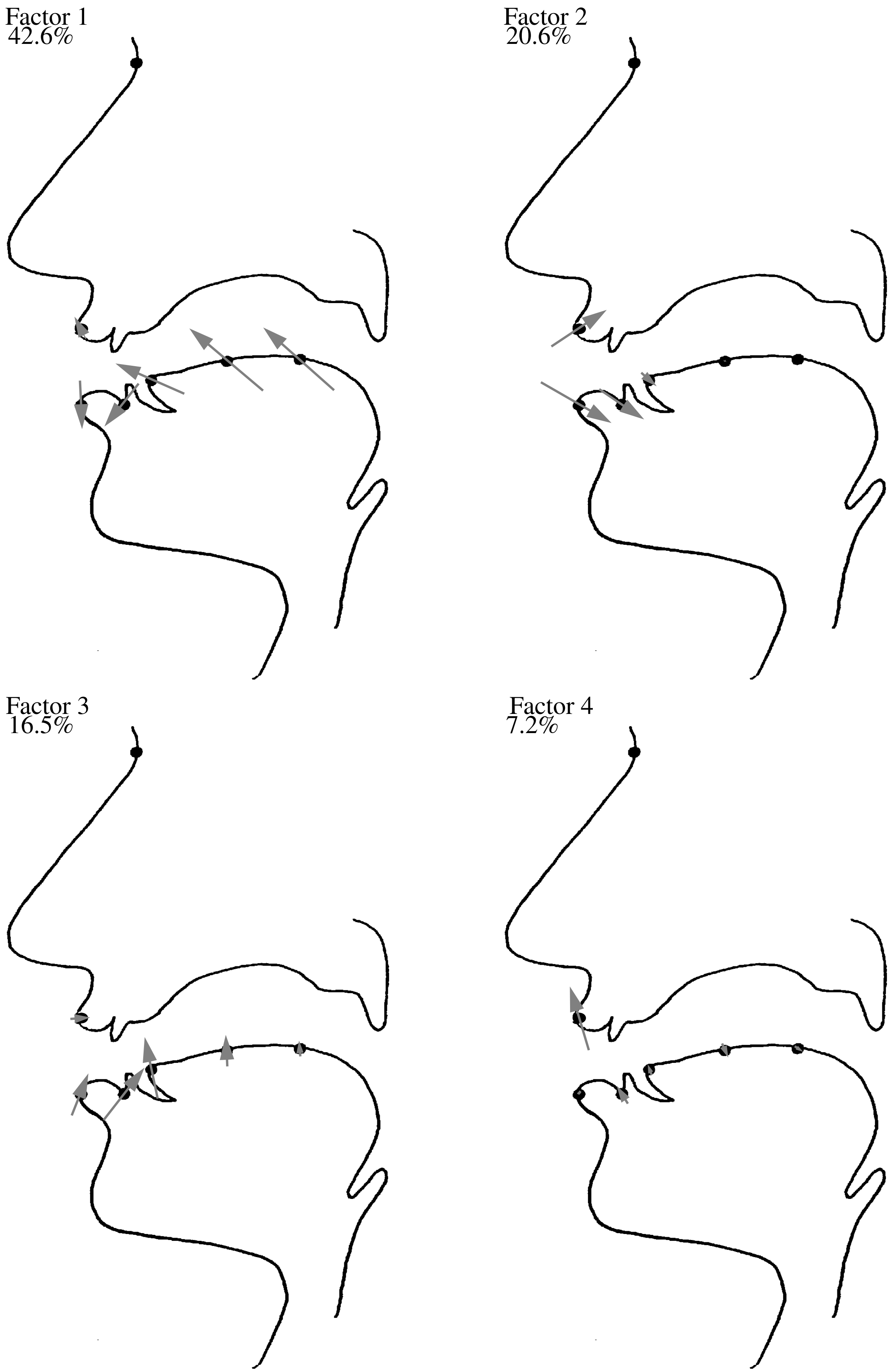
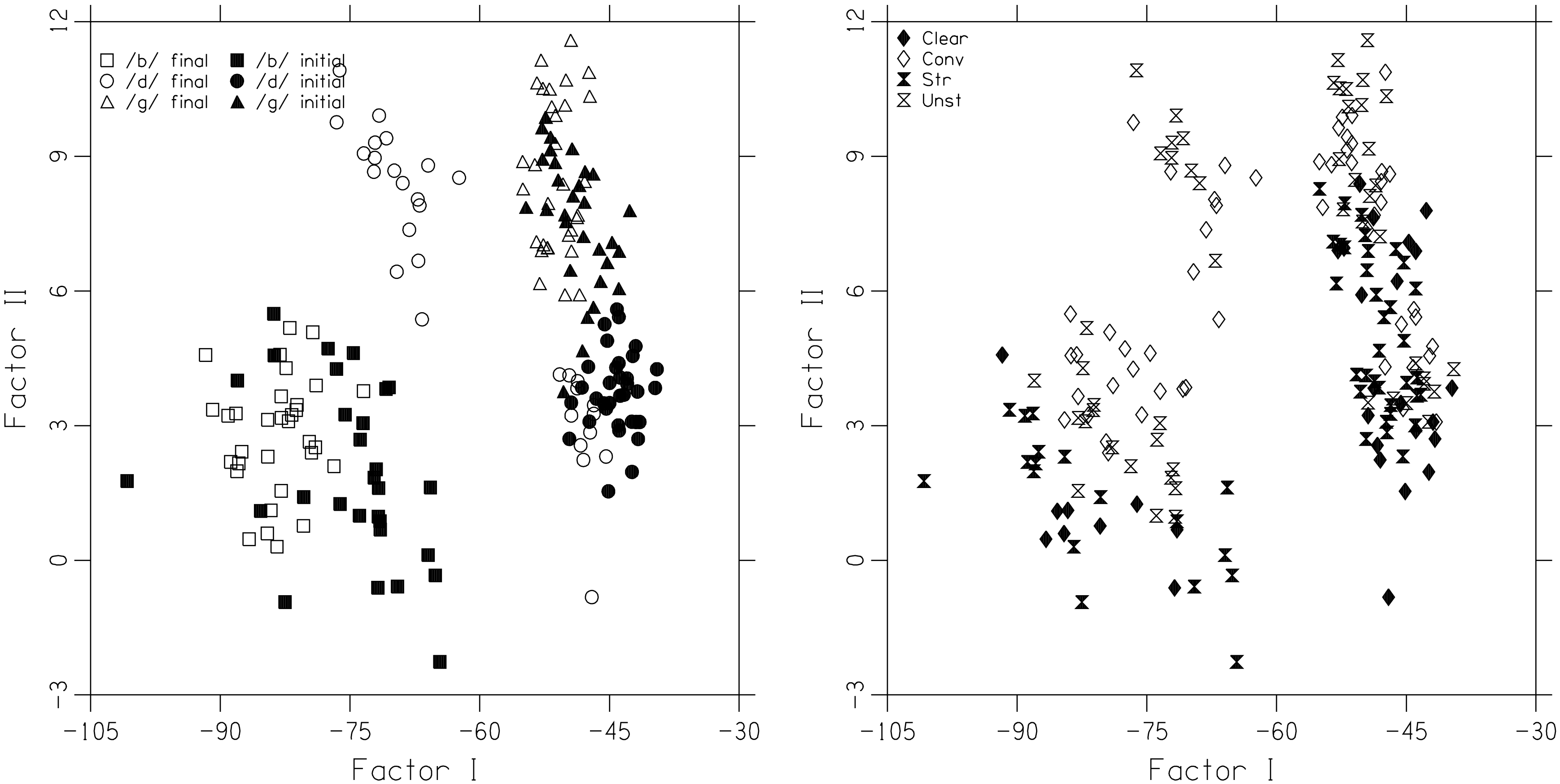


Figure 4



/b/ final
/d/ final
/g/ final
/b/ initial
/d/ initial
/g/ initial
Factor II
Factor I
12
9
6
3
0
−3
−105
−90
−75
−60
−45
−30
Clear
Conv
Str
Unst
Factor II
Factor I

Figure 5a

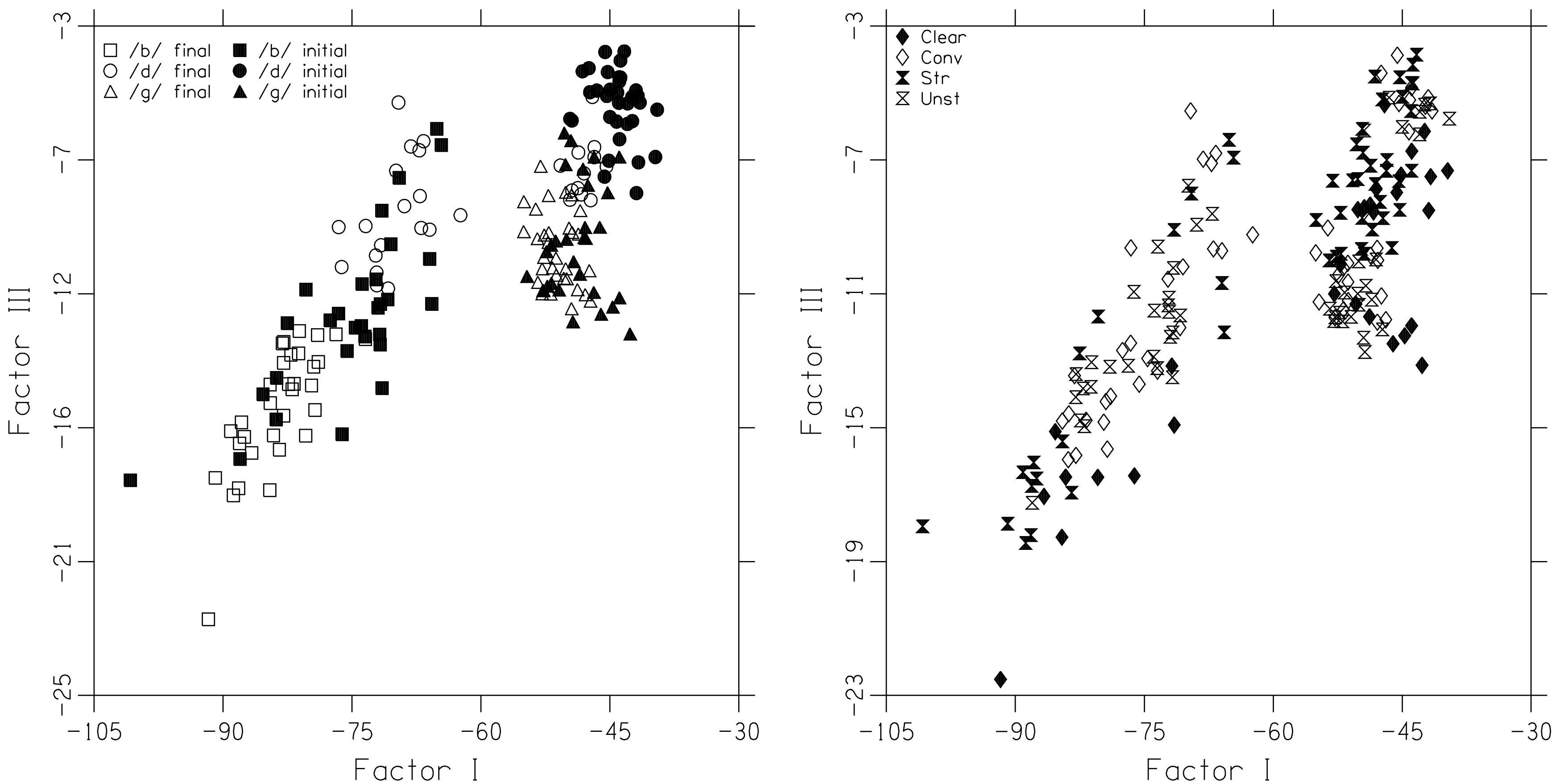

/b/ final
/d/ final
/g/ final
/b/ initial
/d/ initial
/g/ initial
Factor III
Factor I
-3
-7
-12
-16
-21
-25
-105
-90
-75
-60
-45
-30
Clear
Conv
Str
Unst
Factor III
Factor I
-3
-7
-11
-15
-19
-23
-105
-90
-75
-60
-45
-30

Figure 5b

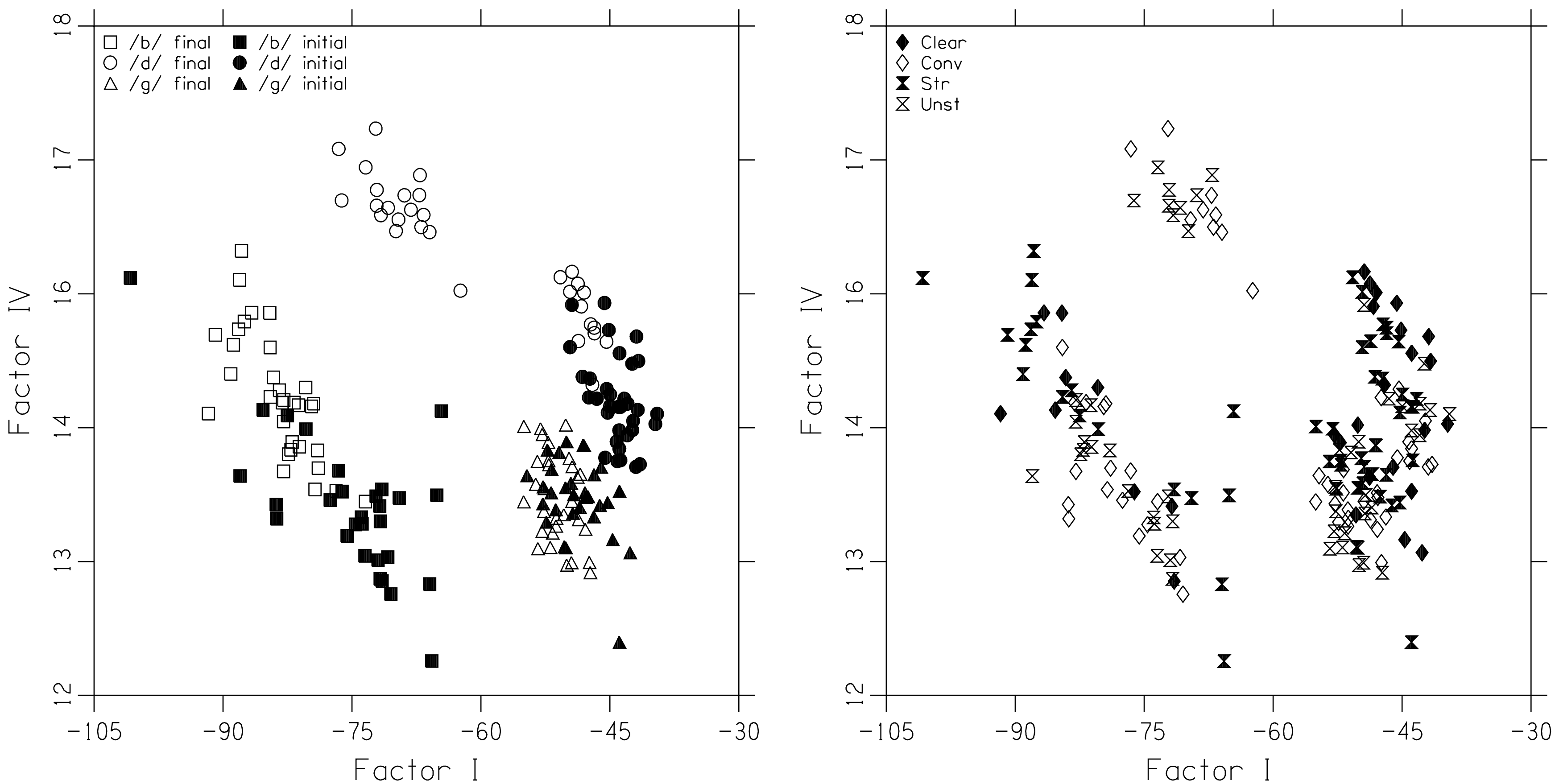

/b/ final
/d/ final
/g/ final
/b/ initial
/d/ initial
/g/ initial
Factor IV
Factor I
Clear
Conv
Str
Unst
Factor IV
Factor I

figure 5c

Insert mode.plt (figure 6) here.

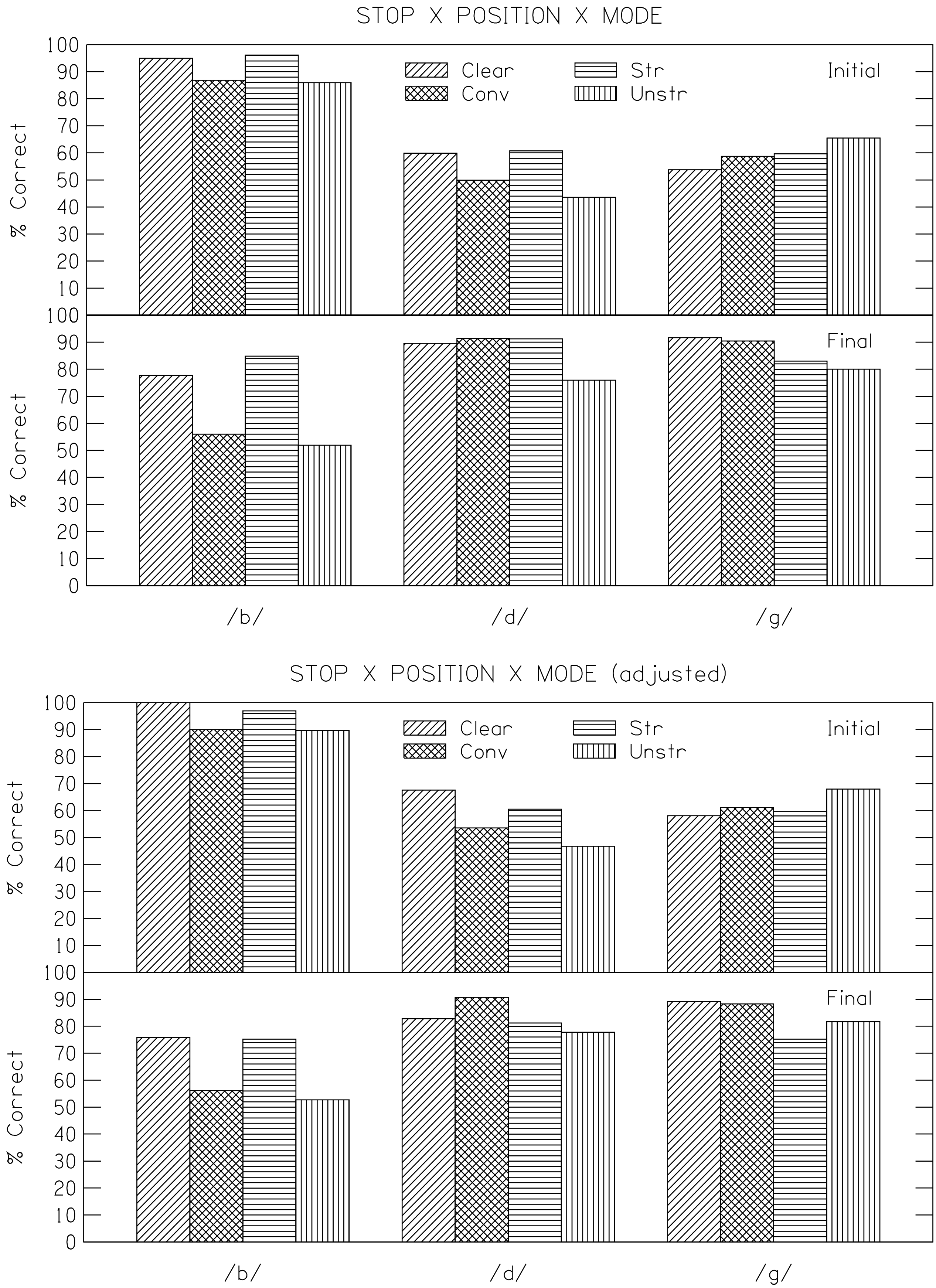

STOP X POSITION X MODE
Clear
Conv
Str
Unstr
Initial
Final
% Correct
100
90
80
70
60
50
40
30
20
10
0
/b/
/d/
/g/
STOP X POSITION X MODE (adjusted)

Figure 7

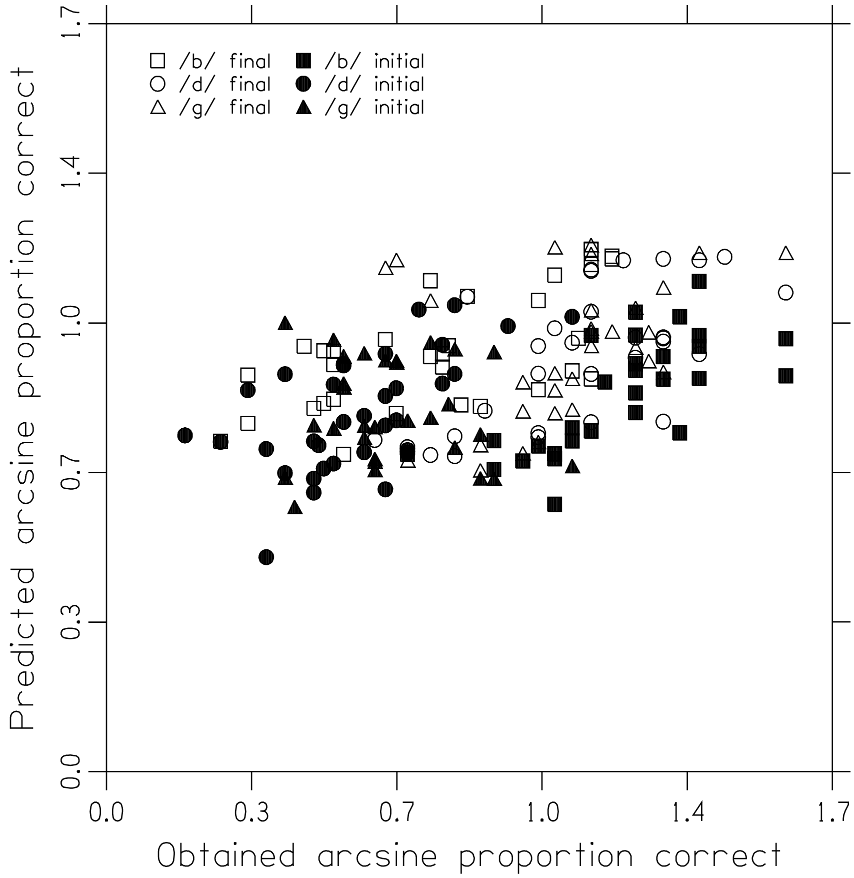


Figure 8

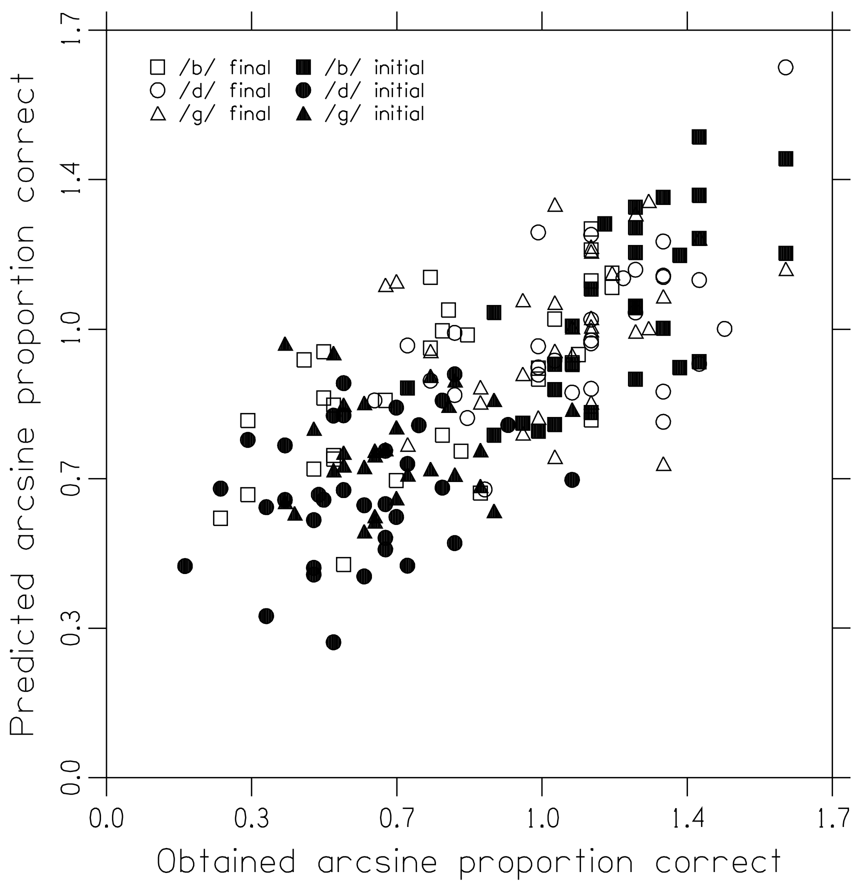
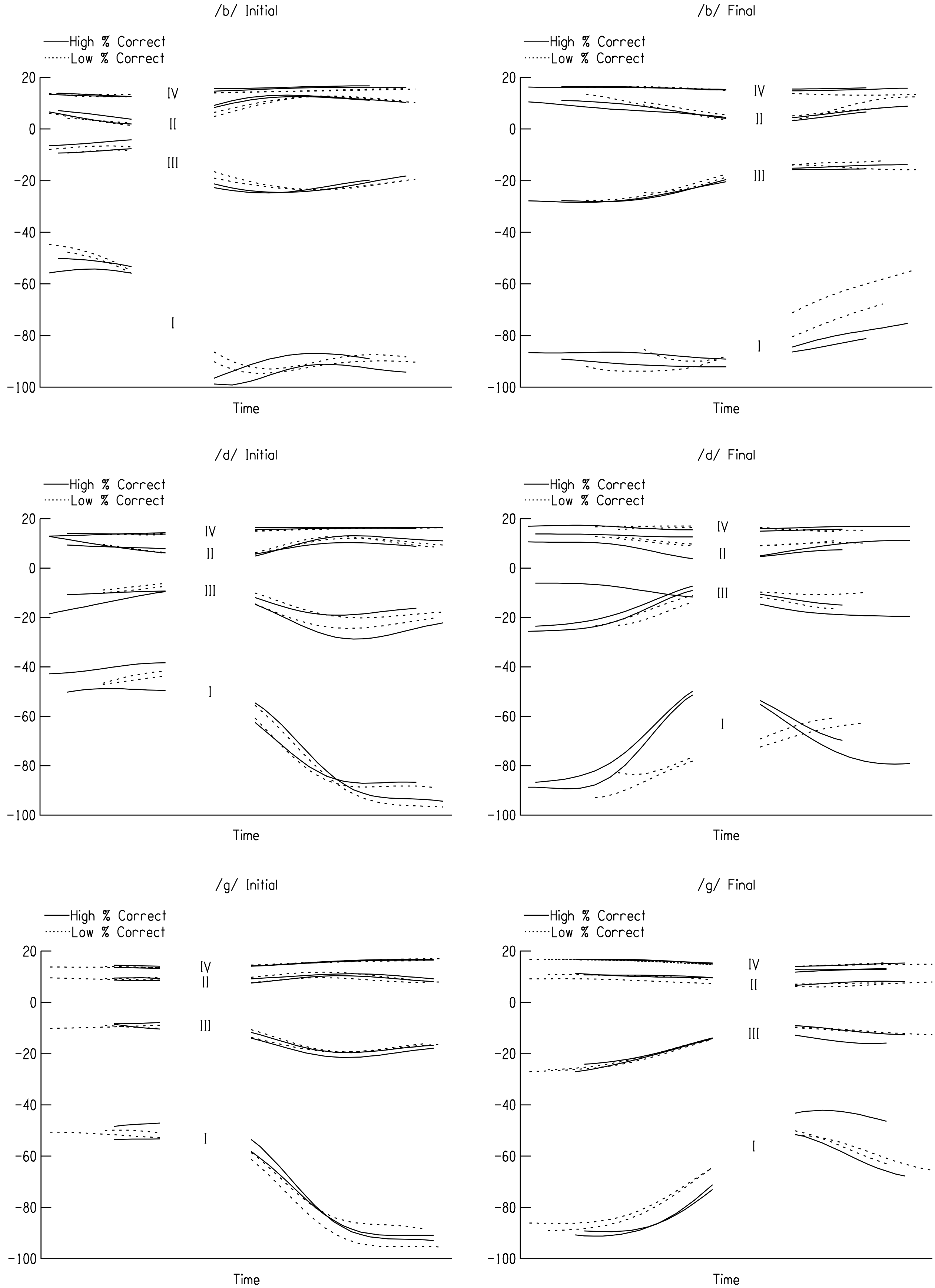


Figure 9



/b/ Initial
High % Correct
Low % Correct
20
0
-20
-40
-60
-80
-100
IV
II
III
I
Time
/b/ Final
High % Correct
Low % Correct
IV
II
III
I
Time
/d/ Initial
High % Correct
Low % Correct
IV
II
III
I
Time
/d/ Final
High % Correct
Low % Correct
IV
II
III
I
Time
/g/ Initial
High % Correct
Low % Correct
IV
II
III
I
Time
/g/ Final
High % Correct
Low % Correct
IV
II
III
I
Time

Figure 10