Temporal Control in Apraxia of Speech: An Acoustic Investigation of Token-to-Token Variability

Samuel A.K. Seddoh, Donald A. Robin, Carlin Hageman, Hyun-Sub Sim, Jerald B. Moon, and John W. Folkins

Since its inception in the late 1960s, the concept of apraxia of speech (AOS) has been associated with considerable controversy. Most of the debate has focused on whether the pathogenesis is motoric or linguistic (e.g., Rosenbek, Kent, and LaPointe, 1984). Although in the past disagreement existed (Martin, 1974), current acoustic, physiological, and perceptual data suggest that the disorder is primarily related to a disturbance in motor control during speech production as well as during nonspeech movements of the articulators (e.g., McNeil and Kent, 1990; Robin, 1992).

Among the most salient behavioral characteristics suggesting a motoric interpretation in apraxic speakers is articulatory mistiming. Temporal parameters of speech, such as those reflected in measures of laryngeal and supralaryngeal gestures, are abnormal in apraxic speakers (Collins, Rosenbek, and Wertz, 1983; Freeman, Sands, and Harris, 1978; Itoh, Sasanuma, Tatsumi, and Kobayashi, 1979; Kent and Rosenbek, 1983). For example, in contrast to the observation for normal speakers, Itoh et al. (1979) found overlapping ranges of voice onset time (VOT) for voiced and voiceless consonants in apraxic subjects. Kent and Rosenbek (1983) reported longer vowel durations for apraxic speakers compared to normal speakers.

Performance variability has also been studied in AOS (e.g., McNeil, Weismer, Adams, and Mulligan, 1990; Mc Coch, Darley, and Noll, 1982). Most of these studies focused on errors related to articulatory events for adjacent segments. For example, acoustic and perceptual studies have reported variable errors related to parameters such as point of articulation (e.g., in the form of substitutions, distortions, etc.) and VOT for stop consonants (Kent
and Rosenbek, 1983; Itoh et al., 1979; Odell, McNeil, Rosenbek, and Hunter, 1991; Weismer and Liss, 1991; see also Hardcastle, 1987, for data on electropalatography). Physiological studies have reported large variability in temporal coordination of the articulators (Itoh, Sasanuma, and Ushijima, 1979).

Comparatively little research has been conducted to determine whether apraxic speakers exhibit variability in the control of temporal parameters (Kent and McNeil, 1987; Weismer and Liss, 1991). Although in general, increased variability in speech production might occur from different sources (Nolan, 1982; Sharkey and Folkins, 1985; Smith and Kenney, 1994), many investigators have extrapolated from acoustic and physiological data to suggest that greater than normal temporal variability may be an indication of instability in the speech motor control system (Disimoni, 1974a, 1974b; Janssen and Wieneke, 1987; Kent and Forner, 1980; Tingley and Allen, 1975; Wieneke and Janssen, 1987). Increased variability in temporal parameters in AOS might therefore provide further support to the motoric interpretation of the disorder.

Most studies represent apraxic behavior by reporting only group data. Given that apraxic speech performance is variable both within and across subjects (Kent and Rosenbek, 1983; Square-Storer and Apeldoorn, 1991), information based on differences across subject groups alone may provide only partial clues regarding the mechanism underlying the disorder. Individuals with motor or structural speech disorders may use different strategies to compensate for their limitations (Folkins 1985). Thus, different subjects may exhibit different characteristics depending on the type of compensatory strategy employed. Characteristics may also differ between subjects qualitatively and/or quantitatively.

Accordingly, the purpose of this study was to examine token-to-token variability in the timing of various parameters at the intrasegmental and word levels, within individual subjects with AOS, and across normal and AOS subject groups.

**METHOD**

**Subjects**

Five patients with AOS, and five control subjects participated in the study. The apraxic subjects were part of a larger, on-going study in our laboratory aimed at understanding visuomotor tracking abilities of this population. Three of them were males and two were females. Their ages
ranged from 35 to 72 years. The normal speaking subjects were age and
gender matched with the AOS subjects. The AOS subjects fit selection
criteria provided by Kent and Rosenbek (1983). They had no evidence of
aphasia and they were not agrammatic. They exhibited normal perform-
ance on various neuropsychological tests related to verbal perform-
ance, memory, and vision. Tests conducted included the Wechsler
Adult Intelligence Scale, Revised (Weschler, 1981), for the evaluation of
verbal ability and Intelligent Quotient (IQ), the Weschler Memory Scale
(Russel, 1975), for the assessment of memory function in verbal and non-
verbal domains, the Rey Auditory-Verbal Learning Test (Rey, 1941), for ver-
bal memory, and the Complex Figure Delayed Recall Test (Osterrieth, 1944),
for visuospatial memory. Visual perception was tested using the Revised
Visual Retention Test (Benton, 1974), the Facial Recognition Test (Benton,
Hamsher, and Varney, 1983) and the Judgement of Line Orientation Test
(Benton et al., 1983). Constructional abilities were tested with the Complex Figure Test-Copy (Osterrieth, 1944). All neuropsychological tests
were obtained from and interpreted by a neuropsychologist. Language
was assessed using the Multilingual Aphasia Examination (Benton and
Hamsher, 1989) and the Token Test (Boller and Vignolo, 1966; De Renzi
and Vignolo, 1962). All subjects were right-handed and had a minimum
of 12 years of education. Detailed descriptive information on the sub-
jects is displayed in Tables 1 (demographic), 2 (language and neuropsy-
chological) and 3 (lesion). All brain-damaged subjects suffered a single,
left cerebral vascular accident.

Speech Materials

The stimuli included four target words: “pop,” “pea,” “Bob,” and “bee,”
embedded in the carrier phrase, “That’s a _____ a day.”

Table 1. Brain-Damaged Subjects Demographic Information

<table>
<thead>
<tr>
<th>Subject</th>
<th>Age</th>
<th>Gender</th>
<th>Handedness*</th>
<th>Education</th>
<th>Job History</th>
</tr>
</thead>
<tbody>
<tr>
<td>AOS1</td>
<td>35</td>
<td>Female</td>
<td>+100</td>
<td>12</td>
<td>Homemaker</td>
</tr>
<tr>
<td>AOS2</td>
<td>55</td>
<td>Male</td>
<td>+100</td>
<td>12</td>
<td>Farmer</td>
</tr>
<tr>
<td>AOS3</td>
<td>66</td>
<td>Male</td>
<td>+100</td>
<td>16</td>
<td>Electrical Engineer</td>
</tr>
<tr>
<td>AOS4</td>
<td>72</td>
<td>Female</td>
<td>+100</td>
<td>12</td>
<td>Homemaker</td>
</tr>
<tr>
<td>AOS5</td>
<td>68</td>
<td>Male</td>
<td>+100</td>
<td>16</td>
<td>Retired</td>
</tr>
</tbody>
</table>

*Note: Handedness was based on modified Oldfield Questionnaire (1971), +100 = full
right-handedness; Education = years.
Table 2. Brain-Damaged Subjects' Language and Neuropsychological Information (1 = Normal; 2 = Mild Impairment; 3 = Moderate Impairment; 4 = Severe Impairment)*

<table>
<thead>
<tr>
<th>Subject</th>
<th>Spontaneous Speech</th>
<th>Auditory Comprehension</th>
<th>Repetition</th>
<th>Naming</th>
<th>Reading Comprehension</th>
</tr>
</thead>
<tbody>
<tr>
<td>AOS1</td>
<td>Nonfluent</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>AOS2</td>
<td>Nonfluent</td>
<td>1</td>
<td>3</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>AOS3</td>
<td>Nonfluent</td>
<td>1</td>
<td>4</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>AOS4</td>
<td>Nonfluent</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>AOS5</td>
<td>Nonfluent</td>
<td>1</td>
<td>4</td>
<td>NA</td>
<td>1</td>
</tr>
</tbody>
</table>

*Note: Severity judgments were made by a speech-language pathologist (second author) and a neuropsychologist based on results from the Multilingual Aphasia Examination (Benton and Hamsher, 1989), clinical interview and chart review.
Table 3. Lesion Information (All brain-damaged subjects had a single left hemispheric stroke)

<table>
<thead>
<tr>
<th>Subject</th>
<th>Time Post Onset (yrs.)</th>
<th>Lesion*</th>
</tr>
</thead>
<tbody>
<tr>
<td>AOS1</td>
<td>8</td>
<td>Left inferior sector of precentral gyrus and posterior sectors of inferior parietal gyrus, superior temporal gyrus, and inferior parietal lobule</td>
</tr>
<tr>
<td>AOS2</td>
<td>7</td>
<td>Left inferior sector of precentral gyrus, posterior sector of inferior frontal gyrus and insula</td>
</tr>
<tr>
<td>AOS3</td>
<td>7</td>
<td>Left inferior sector of precentral gyrus and posterior sector of inferior frontal gyrus</td>
</tr>
<tr>
<td>AOS4</td>
<td>6</td>
<td>Left sensory motor cortex, posterior sector of inferior frontal gyrus, insula and basal ganglia</td>
</tr>
<tr>
<td>AOS5</td>
<td>13</td>
<td>Left basal ganglia and insula</td>
</tr>
</tbody>
</table>

*Note: Lesion data reported by Dr. Hanna Damasio using a standard plotting analysis system (Damasio and Damasio, 1989).

Perceptual Analysis

Speech samples of the pathological subjects were perceptually rated by six experienced full-time speech-language pathologists in terms of overall speech defectiveness (OSD), overall articulatory imprecision (OAI), and intelligibility. A 10-point equal-appearing interval scale with 1 indicating normal and 10 severely defective utterance, was used in the ratings.

Reliability

Measurement of the temporal variables was completed by two individuals who were experienced in making acoustic measures. To determine intraobserver reliability, both individuals reanalyzed 10% of their data. Interobserver reliability was determined by having each individual reanalyze 10% of the other individual’s data. Agreements for both procedures were 99% and 98%, respectively.

Intrajudge reliability for the perceptual judgments involved a repeat of 10% of the samples from all subjects. Results showed a 99% agreement. To determine the interjudge reliability, Pearson Product Moment correlations were computed. Average correlations were found to be 0.84 for OSD, 0.88 for OAI, and 0.81 for intelligibility.
Procedure

Subjects were asked to repeat each target word embedded in the carrier phrase 10 times. The number of repetitions elicited was limited to 10 because of time constraints and patient availability. Subjects spoke into a microphone at a fixed mouth-to-microphone distance of 15 cm. The utterances were recorded on a digital tape recorder and were later low-pass filtered at 11 kHz and digitized at a 20 kHz sampling rate. Only perceptually accurate tokens, as defined by broad transcription, were included in the study.

Token-to-token variability was assessed using the standard deviations\(^1\) for each subject's 10 repetitions. Standard deviations were calculated for the following five durational measures: stop-gap duration (SGD), i.e., the period of silence preceding the initial consonant of each target word; voice onset time (VOT); second formant transition duration (F2D); steady-state vowel duration (VD); and the total duration (TWD) of the target words, defined as the duration of the initial consonant and the following vowel, i.e., CV syllables only. Temporal measures were made from spectrographic displays of each target word as illustrated in Figure 1.

RESULTS

Figures 2 through 6 show the distributions of the raw data for each subject for SGD (Figure 2), VOT (Figure 3), F2D (Figure 4), VD (Figure 5), and TWD (Figure 6). A total of 40 tokens (4 target words × 10 repetitions)

---

\(^1\) Following McNeill, Weismer, Adams, and Mulligan (1990), we chose to analyze our data using standard deviation (SD) instead of the more frequently used coefficient of variation. Unlike the use of SD, the expression of variability with coefficient of variation assumes that a relationship exists between mean and variance in the data in question. This assumption implies that any change in variance corresponds to a proportional change in the square of the mean. Our data showed no consistent pattern of relationship between mean durations and their SDs, which is in consonance with Day and Fisher's (1937) observation that proportional relationship between mean and variance is by no means a logical necessity for every distribution. As McNeill and colleagues (1990) have argued, any justification for "an assumed relationship between mean and variance characteristics of a distribution should be based on explicit knowledge about the form of that relationship" (p. 265). Besides, it has also been demonstrated that results may become reversed (sometimes counter intuitively) when variability is expressed as a ratio (SD/mean, i.e., coefficient of variation) (Colsher, Cooper, and Graff-Radford, 1987).

We need to mention, however, that our choice of SD from among other alternative measures of variability in speech (Munhall, 1989), was by no means intended to suggest that this method is the best way to deal with this issue. Sharkey and Folkins (1985), for example, noted that the use of logarithm of variance has the advantage of eliminating "any effects on the measure of variability due to differences in distribution means" (p. 10). More research is needed to determine how best to measure variability.
Figure 1. A spectrograph of “That’s a pop a day,” produced by one of the normal control subjects, indicating the five temporal parameters measured: Stop-Gap Duration (SGD); Voice Onset Time (VOT); Second Formant Transition Duration (F2D); Steady-State Vowel Duration (VD); Total Duration of Target Word (TWD).

were produced by each subject for each variable. With the exception of apraxic speakers number 3 (AOS3) and number 5 (AOS5), whose performances were comparable to those of normal speakers, all of the AOS subjects exhibited greater range in their tokens for SGD (Figure 2) than did the normal control subjects. The AOS subject with the most variable performance on this measure was apraxic speaker number 4 (AOS4). The same subject produced the greatest range in the tokens for VOT (Figure 3), although in general the performances of all the AOS subjects on this variable were very similar to those of their normal counterparts. Two of the AOS subjects, apraxic speaker number 2 (AOS2) and AOS4, exhibited a relatively large range of tokens for F2D in comparison to the performance of the normal control subjects (Figure 4). The tokens for the other AOS subjects were similar to the normal control speakers. Interestingly, among all subjects (i.e., both normal control speakers and AOS), AOS3 showed the least variable tokens for this measure (F2D) as well as for VD (Figure 5) and TWD (Figure 6). For VD, the tokens for AOS4 and AOS5
Figure 1. A spectrograph of "That's a pop a day," produced by one of the normal control subjects, indicating the five temporal parameters measured: Stop-Gap Duration (SGD); Voice Onset Time (VOT); Second Formant Transition Duration (F2D); Steady-State Vowel Duration (VD); Total Duration of Target Word (TWD).

were produced by each subject for each variable. With the exception of apraxic speakers number 3 (AOS3) and number 5 (AOS5), whose performances were comparable to those of normal speakers, all of the AOS subjects exhibited greater range in their tokens for SGD (Figure 2) than did the normal control subjects. The AOS subject with the most variable performance on this measure was apraxic speaker number 4 (AOS4). The same subject produced the greatest range in the tokens for VOT (Figure 3), although in general the performances of all the AOS subjects on this variable were very similar to those of their normal counterparts. Two of the AOS subjects, apraxic speaker number 2 (AOS2) and AOS4, exhibited a relatively large range of tokens for F2D in comparison to the performance of the normal control subjects (Figure 4). The tokens for the other AOS subjects were similar to the normal control speakers. Interestingly, among all subjects (i.e., both normal control speakers and AOS), AOS3 showed the least variable tokens for this measure (F2D) as well as for VD (Figure 5) and TWD (Figure 6). For VD, the tokens for AOS4 and AOS5
Figure 2. Distribution of the raw data (tokens) for individual subjects for Stop-Gap Duration (SGD). The total number of tokens for each subject was 40 (4 target words x 10 repetitions).

were of the largest range. Apraxic speaker number 1 (AOS1) and AOS2 also exhibited larger than normal range in their tokens for VD. For TWD (Figure 6) all the AOS subjects except AOS3 exhibited large ranges in their tokens, compared to the performance of the normal control subjects. The largest range was exhibited by AOS4.

Thus, in general, there was a greater range in performance for the AOS subjects compared to that of the normal control speakers, particularly for SGD (Figure 2), VD (Figure 5), and TWD (Figure 6). For these
Figure 3. Distribution of the raw data (tokens) for individual subjects for Voice Onset Time (VOT). The total number of tokens for each subject was 40 (4 target words × 10 repetitions).

measures, many of the tokens for the AOS subjects fell outside the range exhibited by the normal control speakers, although many tokens also fell well within the control subjects' range. VOT (Figure 3) and F2D (Figure 4) measures, however, showed less appreciable differences between the tokens for the AOS subjects and those for the normal control speakers.

To determine whether the performance of the AOS subjects differed significantly from that of the normal control subjects, group comparisons of the token-to-token standard deviations (SDs) for each variable were completed.
Figure 4. Distribution of the raw data (tokens) for individual subjects for Second Formant Transition Duration (F2D). The total number of tokens for each subject was 40 (4 target words × 10 repetitions).

T tests were performed on the SDs for each measure. Table 4 shows the individual SDs for each subject for each variable as well as the means for groups. The results of the T tests revealed that the AOS subjects' performance differed significantly from that of the normal control speakers for SGD (p = .0037), VD (p = .0232), and TWD (p = .0026). The performance of the AOS subjects did not differ significantly for VOT (p = .7508) and F2D (p = .6080).
Figure 5. Distribution of the raw data (tokens) for individual subjects for Steady-State Vowel Duration (VD). The total number of tokens for each subject was 40 (4 target words × 10 repetitions).

DISCUSSION

The results of the present study showed that token-to-token variability is greater for apraxic speakers for the control of the temporal parameters studied relative to a normal age- and gender-matched control group. This finding is similar to previous investigations that have demonstrated that, as a group, AOS patients exhibit greater variability during speech
Figure 6. Distribution of the raw data (tokens) for individual subjects for Total Duration of Target Word (TWD). The total number of tokens for each subject was 40 (4 target words × 10 repetitions).

production than normal individuals (Kent and McNeil, 1987; McNeil, Liss, Tseng, and Kent, 1990; McNeil, Weismer, Adams, and Mulligan, 1990; Mlcoch, Darley, and Noll, 1982). Compared to the performance of the normal control subjects, many of our apraxic patients exhibited greater token-to-token variability in the control of stop-gap duration (SGD) (Figure 2), steady-state vowel duration (VD) (Figure 5), and total target word duration (TWD) (Figure 6), but not voice onset time (VOT) (Figure 3) and second formant transition duration (F2D) (Figure 4). Interestingly, all of the apraxic speakers also produced tokens that fell within the range
Table 4. Standard Deviations for Individual Subjects and the Mean Standard Deviations for Each Temporal Variable

<table>
<thead>
<tr>
<th>Measures</th>
<th>Normal Control Subjects</th>
<th>AOS Subjects</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>SGD</td>
<td>13</td>
<td>13</td>
</tr>
<tr>
<td>VOT</td>
<td>28</td>
<td>23</td>
</tr>
<tr>
<td>F2D</td>
<td>24</td>
<td>19</td>
</tr>
<tr>
<td>VD</td>
<td>36</td>
<td>30</td>
</tr>
<tr>
<td>TWD</td>
<td>32</td>
<td>29</td>
</tr>
</tbody>
</table>

Note: SGD = Stop-Gap Duration; VOT = Voice Onset Time; F2D = Second Formant Transition Duration; VD = Steady-State Vowel Duration; TWD = Total Duration of Target Word.
of normal controls, and one AOS subject (AOS3) exhibited a relatively normal range for all variables. It should be noted that the speaker with the least token-to-token variability (AOS3) was the least impaired speaker, and the one with the greatest variability (AOS4) was the most severely impaired speaker. The severity rating was based on the perceptual judgment of speech, i.e., overall speech, defectiveness, overall articulatory imprecision, and intelligibility.

Different interpretations of variability are possible (Sharkey and Folkins, 1985). Increased variability in speech may be an indication that the speaker is unable to reach intended motor goals. Several acoustic investigations of token-to-token variability in children have also shown that variability within the parameters studied decreases with increases in age (DiSimoni, 1974a, 1974b; Kent and Forner, 1980; Tingley and Allen, 1975). These investigators have suggested that greater variability might point to the involvement of the speech motor control system. Similarly, Janssen and Wieneke (1987) and Wieneke and Janssen (1987) noted that greater than normal token-to-token temporal variability in stutterers may be an indication of instability in the speech motor control system in this population.

One possibility is that speech production, as a motor task, involves a functional operating range within which a speaker must perform. The abnormal variability exhibited by the AOS subjects may be an indication that these patients operate more often outside this functional range than within it. Therefore they may be more likely to make errors.

Greater token-to-token variability may also be a reflection of the motor system's flexibility. Folkins (1985) suggested that a speaker may use different alternatives within an existing motor rule system to attain a perceptual goal. This implies that, following changes that affect the speech production system, speakers may compensate for their limitations in different ways and to varying degrees, perhaps depending on the level of their limitations or the demands of the stimuli. Our apraxic speakers were able to compensate for their limitations during the experimental task. Even though they showed greater variability than the normal control speakers, they still produced perceptually adequate tokens. For example, AOS4 who had the greatest token-to-token variability, and AOS3 whose variability was close to normal, both produced perceptually adequate tokens. Presumably AOS3 compensated less than AOS4.

Finally, the present findings suggest that knowledge of variability in AOS might be clinically useful. Speakers with large variability and good compensation may have a better prognosis for recovery and/or may be better candidates for treatment than those who do not compensate well for their impairment. Further research on variability and compensation is warranted.
ACKNOWLEDGMENTS

We wish to thank John C. Rosenbek, Malcolm R. McNeil, Joseph R. Duffy, Robert T. Wertz, and Patrick J. Doyle for generating a stimulating discussion following the presentation of an earlier version of this paper at the 24th annual Clinical Aphasiology Conference at Traverse City, Michigan. We also thank the anonymous reviewers of the manuscript for their careful scrutiny and considerations. Finally, we wish to express our gratitude to Dr. Hannah Damasio for providing neuroanatomic information, and to Dr. Daniel Tranel for providing neuropsychological data on the brain-damaged patients.

This research was supported by NIH (NIDCD) grant no. DC90076 and NIH (NINDS) grant no. NS19632.

REFERENCES


