AN ANALYSIS OF CONNECTED SPEECH SAMPLES OF APHASIC AND NORMAL SPEAKERS

Kathryn M. Yorkston and David R. Beukelman

University of Washington, Seattle

The purpose of this study was to develop a technique for quantifying connected speech samples of aphasic individuals as they recover from the moderate to the mild range of severity. Verbal picture descriptions elicited from mild and moderate aphasic speakers were audio-recorded and compared to those elicited from normal adult and geriatric speakers. These samples were compared on the basis of a measure of amount of information conveyed (content units) and two measures of efficiency including speaking rate (syllables per minute) and rate at which information was conveyed (content units per minute). Results indicated an inverse relationship between severity of aphasia and amount of information conveyed. However, mild and high-moderate aphasic speakers tended to communicate as much information as normal speakers. Both measures of efficiency differentiated groups of mild and high-moderate aphasics from normal speakers as well as differentiating low-moderate from mild aphasic speakers. Use of this quantification system which takes into account both amount of information and efficiency of communication is illustrated with data obtained from a recovering aphasic speaker.

The quantification of verbal output performance during the recovery of moderately and mildly aphasic individuals presents a unique clinical problem. Tasks used to sample verbal output on the majority of standard aphasia tests include object/picture naming, sentence completion, word or sentence imitation, and description of object function. Highly structured tasks, such as these, have had wide clinical use because they elicit responses reliably, are easy to score reliably, and taken as a group, provide a good index of severity for aphasic verbal impairment. However, many of the tasks that lend themselves to either plus/minus of multidimensional (Porch, 1971) scoring are so simple that mildly aphasic individuals may be indistinguishable from normal speakers, even though the aphasic speaker may have communication problems of considerable social and vocational consequence.

One alternative to the sampling of verbal performance with a series of highly structured, relatively simple tasks is the use of a rating scale system. Rating scale systems such as the ones developed for the Boston Diagnostic Aphasia Examination (Goodglass and Kaplan, 1972) allow analysis of relatively complex speech samples. The Rating Scale Profile of Speech Characteristics covers those features most elusive to quantification, which are: melodic line; phrase

length; articulatory agility; grammatical form; paraphasia in running speech; and word-finding. A judge listens to a sample of free conversation and oral description of a standard picture and rates each feature on a seven-point scale. The Profile of Speech Characteristics is used in conjunction with a five-point Aphasia Severity Rating Scale to supplement objective scores obtained from the more highly structured tasks included in the remainder of the test. Analysis of a running speech sample is critical to the assessment of mildly aphasic speakers. However, subjective rating scale systems, although useful descriptively, do not provide the objective quantification needed to monitor subtle changes in performance.

Besides rating scales, another method that is helpful for quantifying the speech of moderate and mildly aphasic speakers is the method reported by Howes (1973). He gathered extensive language samples (5000 words on continuous conversational speech) from aphasic speakers and measured a variety of statistical characteristics including distribution of word frequencies, rate at which words were emitted, and distribution of word length. Although the emphasis of Howes' work was to compare the mathematical equations derived from aphasic language samples with those equations known to apply to normal language, some of his results seem directly relevant to the clinical measurement of aphasic speech. Results indicated that the rate at which aphasics speak was relatively constant and varied only 2 to 3% from day to day and topic to topic. Howes suggested that rate of speech may be used as a valid index of a speaker's overall disturbance.

In summary, a procedure for sampling and quantifying running speech would be an important addition to an evaluation battery for moderate and mildly aphasic speakers who may demonstrate socially and vocationally significant communication deficits, yet have only minimal difficulty on some standard aphasia tests. The purpose of this study was to develop an objective, reliable, and clinically useful method of sampling and analyzing verbal output of moderate and mildly aphasic speakers and to determine if the measures obtained were sensitive to severity of aphasia and to differences between mildly aphasic and normal speakers. The clinical application for monitoring recovery as speakers move into mild range of severity will be discussed and illustrated with a brief case study.

METHOD

Speakers

Aphasic Speakers. The 50 aphasic speakers (41 men and nine women) who participated in this study were native speakers of English judged during their neurological evaluations to have unilateral left hemisphere lesions. The authors reviewed audio-recorded speech samples and excluded all those speakers with more than minimal dysarthria. Ages ranged from 20 to 75 years with a mean age of 45 years. All speakers were at least one month post onset and had

verbal scores on the Porch Index of Communicative Ability (PICA) (1967) at or above the 50th percentile. All aphasic speakers were classified according to percentiles on the verbal subtests of the PICA into one mild and two moderate severity groups as follows:

Mild Group: 81st to 99th percentile (N = 17).
High-Moderate Group: 66th to 80th percentile (N = 16).
Low-Moderate Group: 50th to 65th percentile (N = 17).

Normal Speakers. Seventy-eight normal adult, native speakers of English (25 males and 53 females) were included in this study, none of whom reported a history of neurological deficit. Because of the potential relationship between age and efficiency of communication, normal speakers were classified into two groups according to age:

Normal Adult Group: Forty-eight adults ranging in age from 19 to 49 years with a mean age of 31 years were included. Selection was not made on the basis of education and speakers represented a cross section of occupations including orderlies, secretaries, nurses, physicians, physical therapists, and vocational counselors.

Normal Geriatric Group: Thirty geriatric adults ranging in age from 58 to 93 years with a mean age of 73 years were included. These speakers were drawn from senior citizen volunteer groups, retirement homes and patients hospitalized for nonneurological problems.

Speech Samples

Speech samples were elicited through a picture description task (*Cookie Theft* picture, Boston Diagnostic Aphasia Examination). Pilot work (Yorkston and Beukelman, 1977) indicated that this picture depicts sufficient action so that normal speakers were able to talk about the picture for 30 to 45 seconds. Furthermore, the content of speech produced was relatively predictable. Predictability was essential, as the analysis involves comparing the content of speech samples produced by different speakers. Pilot work also revealed that predictability of content could not be obtained in conversational speech samples. Instructions given to speakers were as follows: *Tell everything you see happening in this picture.* No verbal repeats or cues were given. Recording was stopped after 30 seconds of silence or after speakers indicated they were finished. No other restrictions were placed on the speakers.

Analysis of Speech Samples

Each speech sample was timed, transcribed verbatim, and analyzed using a method developed in pilot work (Yorkston and Beukelman, 1977). All of the samples were analyzed by a single judge. However, for reliability measures, this judge was joined by a panel of four speech-language pathologists who analyzed a series of randomly selected samples. The following measures were obtained from each sample.

Time. Each sample was timed beginning with the first utterance after instructions and ending when the speaker indicated that the task was completed

or after a 30-second period of silence elapsed. The 30-second period of silence was not included in the total sample time.

Syllables. To obtain a measure of speaking rate, the total number of syllables contained in each speech sample was tallied directly from the audio tape. All syllables, regardless of intelligibility, were counted. Counting was begun with the first utterance after instruction and was terminated on completion of the task. Interjudge reliability was assessed on a series of 10 normal and aphasic speech samples. All syllable counts made by the judges fell within a range of a ± 5%. When the same samples were rescored one week later, mean difference between test-retest syllable counts was 3.5%.

Content Units. The measure of amount of information conveyed was termed the *content unit.* Because speech samples potentially can be subdivided into such a large number of bits of information, the decision was made to compile a list of content units mentioned at least once as normal speakers described the *Cookie Theft* picture (Appendix). A content unit was defined as a grouping of information that was always expressed as a unit by normal speakers. For example, the six words and phrases in italics are the content units in the following sentence: The *little boy* is *on the stool* and *reaching up* for a *cookie*
1 2 3 4 5
and he's going *to fall over.* The words *on the stool* were considered one content
6
unit because in normal speech sample *on* did not appear without *the stool.* The words *little* and *boy* were considered different *content units* because many of the normal samples contained the concept *boy* without the concept *little.* Credit was given for words that were similar in meaning but not identical to the listed words. For example, the words *lady* and *woman* were considered synonomous as were the terms *ignoring, daydreaming* and *not paying attention.* To eliminate credit for redundant information each content unit was counted only one time. The total number of content units communicated in each speech sample was tallied by reviewing written transcripts and identifying which of the listed content units it contained. Interjudge reliability was assessed by having four speech-language pathologists judge a series of 10 normal and aphasic speech samples selected randomly. Judges were within ± 1 content unit 95% of the time. Test-retest agreement was 100%.

Each of the different speaker groups was compared using three measures generated from the speech sample: (1) content units indicated the amount of information conveyed; (2) syllables per minute served as an indicator of overall speaking rate; and (3) content units per minute were used as a measure of efficiency of communication or rate at which the speaker conveyed information.

RESULTS AND DISCUSSION

Amount of Information

The measure used in this study to quantify the amount of information con-

TABLE 1. Means ($\bar{X}$) and standard deviations (SD) of content units, syllables per minute and content units per minute for two normal adult and three aphasic severity groups for samples elicited with a picture description task.

Groups	*Content Units* $\bar{X}$	*(SD)*	*Syllables/Min.* $\bar{X}$	*(SD)*	*Content Units/Min.* $\bar{X}$	*(SD)*
Normal						
Adult	18.0	(4.7)	202.9	(40.2)	41.9	(13.2)
Geriatric	14.7	(3.6)	193.2	(39.8)	33.7	(13.5)
Aphasic						
Mild	16.4	(3.3)	120.8	(35.0)	18.7	(6.5)
High-Moderate	14.6	(4.6)	96.9	(28.3)	13.2	(7.0)
Low-Moderate	10.5	(2.5)	61.6	(19.1)	8.3	(4.2)

veyed was the total number of content units contained in the picture description sample. Means and standard deviations for this measure were computed for the three Aphasic Speaker Groups and the two Normal Speaker Groups. (Table 1). As illustrated in Figure 1A, there was an inverse relationship between severity of aphasia and mean content units produced in the speech samples. In other words, as severity decreased, the number of content units communicated increased to the point at which the mean content units pro-

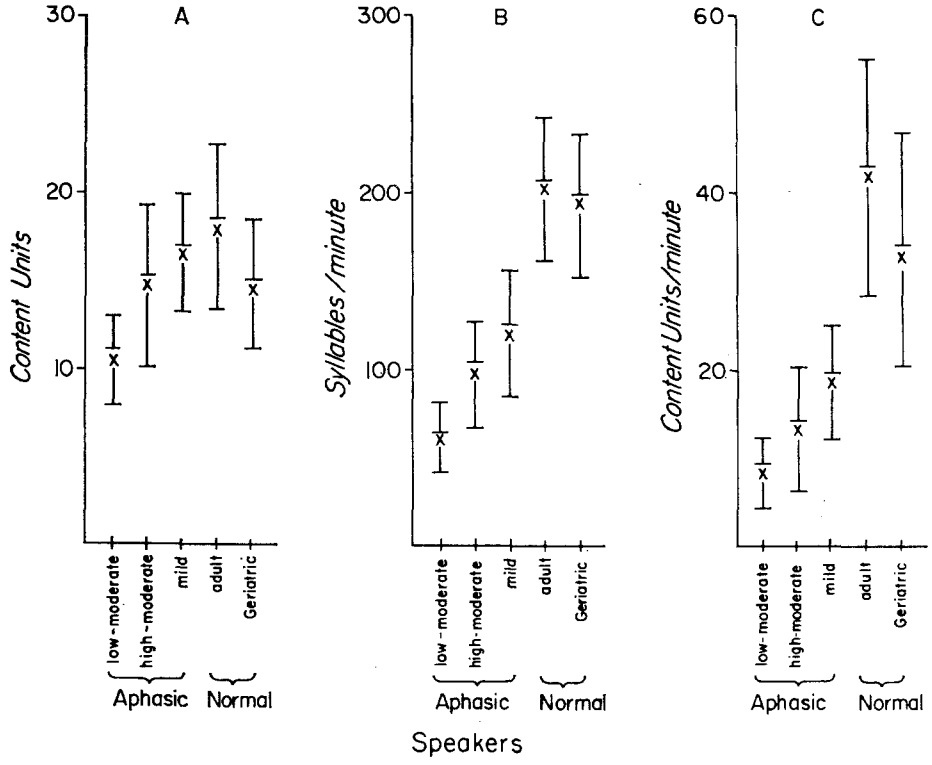

Figure 1. Means (X) and a one standard deviation range of content units, syllables per minute and content units per minute for Normal and Aphasic Speaker Groups.

duced by speakers in the Mild Aphasic Group was between the means produced by the Normal Adult and the Normal Geriatric Groups.

Analysis of variance revealed significant differences in number of content units communicated among the Speaker Groups ($df = 4,123$; $F = 12.1$, $p < 0.001$). Neuman Keuls post hoc testing (Kirk, 1968) indicated that the mean content units produced by the Low-Moderate Aphasic Group was significantly lower at the 0.01 level than those produced by any of the other groups. Differences between the Normal Speaker Groups, both Adult and Geriatric, and the Mild and High-Moderate Aphasic Groups were not significant. Taken together, these results suggest that a measure of content units communicated in a sample of connected speech may be useful in distinguishing between mild and moderate aphasic deficits. However, this measure is insensitive to the differences between mild aphasic and normal speakers. Because mildly aphasic speakers were able to communicate as many concepts as normals without time restrictions and without prompting or cueing, other measures must be used to distinguish between these groups.

Efficiency of Communication

Two measures that give an indication of efficiency of communication were calculated. These measures included speaking rate (syllables per minute) and rate at which content units were communicated (content units per minute). Means and standard deviations of each of these measures for the Normal Speaker Groups and the three Aphasic Groups were computed. Examination of Figure 1B reveals that mean syllables per minute was related inversely to severity of aphasia. However, the speaking rates of the Aphasic Groups were all slower than either of the Normal Speakers Groups. Mean speaking rates for Normal Adult, Normal Geriatric, Mild, High-Moderate and Low-Moderate Aphasics were 202.9, 193.2, 120.8, 96.9 and 61.6 syllables per minute respectively. Analysis of variance revealed significant differences among these groups ($df = 4,123$; $F = 72.2$, $p < 0.001$). Neuman-Keuls post hoc testing indicated that the two Normal Speaker Groups were not significantly different from one another but that mean speaking rates for both Normal Adult and Geriatric Groups were higher than any of the Aphasic Groups at the 0.01 level. Further, the mean speaking rates of Mild and High-Moderate Aphasic Groups were not significantly different from one another but both were significantly higher than the speaking rates of the Low-Moderate Aphasic Group at the 0.05 level.

The second measure of efficiency investigated in this study was content units per minute. Examination of Figure 1C reveals that there was an inverse relationship between this measure of efficiency and severity of aphasia. Further, none of the aphasic groups achieved a rate as rapid as normal speakers. Analysis of variance revealed significant differences among speaker groups ($df = 4,123$; $F = 44.87$, $p < 0.001$). Neuman Kuels post hoc testing indicated that the Normal Adult speakers produced significantly more content units per minute than did Normal Geriatric speakers (0.05 level). However, both normal

groups produced significantly more content units per minute than any of the Aphasic Groups (0.01 level). Significant differences were found between the Mild and Low-Moderate Aphasic Groups but neither of these Aphasic Groups were different from the High-Moderate Group.

Multiple Samples and Reliability

If the technique described for sampling and quantifying connected speech samples is to be used for clinical management of high-level aphasic individuals, the potential for learning that may result from multiple sampling must be explored. Two samples were elicited from 10 aphasic speakers whose communication deficits and etiology met the criteria described earlier in this report. The samples were taken on the same day with at least a half-hour interval between them. T-test comparisons between the means of content units, syllables per minute and content units per minute obtained from the sample and those obtained from the second sample indicated no significant differences between samples for any measures. These data support the contention that familiarity with the picture and task does not improve performance significantly and suggests that learning did not affect the scores obtained.

Also of interest when making clinical use of this technique is test-retest reliability. Pearson product-moment correlations between first and second samples for content units, syllables per minute and content units per minute were 0.94, 0.96 and 0.94, respectively. These correlations suggest a close relationship between speech samples obtained on the same day.

CLINICAL IMPLICATIONS

From the comparisons of connected speech samples of mild and moderate aphasic and normal speakers, several points emerge that are relevant to assessment of verbal output and the clinical management of aphasic individuals. An inverse relationship exists between severity of aphasia and number of content units conveyed. However, mild and high-moderate aphasic speakers do not differ from normal speakers, both adult or geriatric, in the amount of information they are able to convey. Therefore, as an aphasic speaker moves from the moderate into the mild range of severity, other measures are needed to monitor progress. Because measures of efficiency of communication distinguish normal speakers from the mildly aphasic speakers, these measures are potentially useful in monitoring verbal output of the high-level aphasic speakers. Both of the efficiency measures investigated in this report, syllables per minute and content units per minute, essentially produced the same pattern of results. Namely, scores produced by normal speakers were significantly higher than scores produced by any of the aphasic speakers groups and aphasic scores were related inversely to severity of aphasia.

Content and efficiency data derived from a picture description task are not intended to replace traditional measures of verbal output. Rather, these mea-

sures quantify connected speech and thus extend the task difficulty into a range appropriate for the assessment of moderate and mildly aphasic speakers. The following example illustrates how the progress of a 30-year-old man, JB, was monitored as he moved from the moderate into the mild severity range. JB had suffered a cerebral hemorrhage that resulted in moderately severe fluent aphasia. During the initial recovery period, spontaneous speech was fluent but marked by numerous word finding problems and phonemic-articulatory breakdowns.

From the third through the seventh month post-onset, changes in verbal output were monitored with the verbal subtests of the PICA and measures obtained from the picture description task. Figure 2 illustrates the pattern of

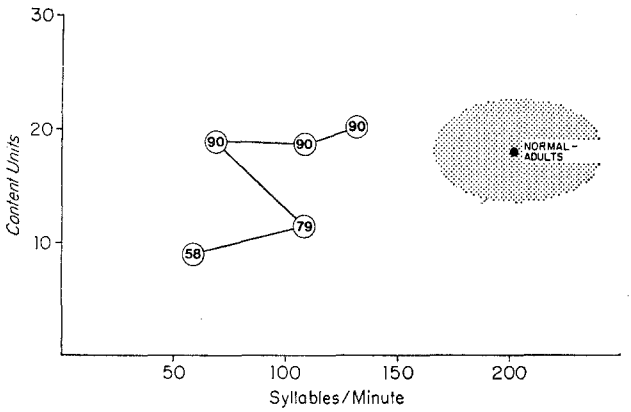

Figure 2. Content units and syllables per minute obtained from speech samples elicited at monthly intervals during the course of recovery for JB. Also illustrated are the means and a one standard deviation range for Normal Adult speakers. Encircled numbers are PICA verbal percentiles at time when speech samples were elicited.

JB's recovery and contains a plot of content units against syllables per minute over a five-month recovery period. The encircled numbers on JB's recovery graph are PICA verbal percentile scores obtained at the time the picture description samples were elicited. As a point of reference, the mean and standard deviation data for the Normal Adult Speaker Group were also plotted. Examination of the figure reveals that as JB's verbal scores on the PICA moved from the 58th to the 90th percentile, his performance on the verbal picture description sample also improved. In fact, by the third sample (fifth month post onset), the amount of information, content units, fell within the normal range. However, communication efficiency was considerably below the normal range with a speaking rate of 78 syllables per minute as opposed to approximately 200 syllables per minute for normal speakers. Ninetieth percentile scores on the verbal subtests of the PICA represent performance that is characterized by an occasional response delay or phoneme distortion. The relatively infrequent articulatory distortions experienced by JB persisted and prevented higher scores. However, in the presence of stable PICA scores the num-

ber of syllables per minute on the picture description task continued to increase.

By the seventh month post-onset, JB returned on a part-time basis to the business he had previously managed. Despite the excellent progress JB had made since the onset of aphasia, he found that returning to work was frustrating. He commented frequently that he was unable to handle the communication demands of his job, although he seemed to communicate functionally in conversational settings. This feeling of reduced ability to communicate rapidly was expressed by many of the mildly aphasic individuals interviewed by Rolnick and Hoops (1960). Moss (1972), in discussing his recovery from aphasia, said that despite reassurances from those around him that he was doing well, the communication process seemed very slow, laborious, and at times inaccurate.

The analysis system described in this report allowed the authors to document for JB that the amount of information he communicated was within normal limits and at the same time to confirm his frustrating reduction in communication efficiency. This information was also shared with JB's family and business associates, so they could understand better his frustration with his communication performance.

ACKNOWLEDGMENT

This study was supported in part by RSA Grant #16-P-56818. Pilot work for the present study was presented at the Seventh Annual Clinical Aphasiology Conference, May 1977 and appears in the Conference Proceedings. The authors wish to acknowledge the assistance of Kitty Turner, Pat Mitsuda, Pat Waugh and Robert C. Marshall. Reprint requests should be sent to Kathryn M. Yorkston, Speech Pathology Services, Department of Rehabilitation Medicine RJ-30, University of Washington, Seattle, Washington, 98195.

REFERENCES

GOODGLASS, H., and KAPLAN, E., *Boston Diagnostic Aphasia Examination,* Philadelphia: Lea and Febiger (1972).

HOWES, D., Some experimental investigations of language in aphasia. In H. GOODGLASS and S. BLUMSTEIN, (Eds.), *Psycholinguistics and Aphasia.* Baltimore: The Johns Hopkins University Press (1973).

KIRK, R. E., *Experimental Design: Procedures for the Behavioral Sciences.* Belmont, California: Brooks/Cole Publishing Co. (1968).

MOSS, C. S., *Recovery with Aphasia.* Urbana: University of Illinois Press (1972).

PORCH, B. E., Multidimensional scoring in Aphasia testing. *J. Speech Hearing Res.,* **14,** 776–792 (1971).

PORCH, B., *The Porch Index of Communicative Ability.* Palo Alto: Consulting Psychologist Press (1967).

ROLNICK, M., and HOOPS, H. R., Aphasia as seen by the Aphasic. *J. Speech Hearing Dis.,* **34,** 48–53 (1969).

YORKSTON, K., and BEUKELMAN, D., A system for quantifying verbal output of high-level aphasics. *Proceedings of the Seventh Annual Clinical Aphasiology Conference,* Minneapolis: BRK 175–180 (1977).

Received April 30, 1979.
Accepted July 30, 1979.

APPENDIX

A. Content units used in analyzing speech samples describing the *Cookie Theft* picture

Two
children
little
boy
brother
standing
on stool
wobbling
(off balance)
3-legged
falling over
on the floor
hurt himself
reaching up
taking (stealing)
cookies
for himself
for his sister
from the jar
on the high shelf
in the cupboard
with the open door
handing to sister
little
girl
sister
standing
by boy
reaching up
asking for cookie
has finger to mouth
saying *shhh* (keeping
him quiet)
trying to help
(not trying to help)
laughing
mother
woman (lady)
children behind her
standing
by sink
washing (doing)
dishes
drying
faucet on
full blast
ignoring
(daydreaming)
water
overflowing
onto floor
feet getting wet
dirty dishes left
puddle
in the kitchen (indoors)
general statement
about disaster
lawn
sidewalk
house next door
open window
curtains