Assessing Gestural Intelligibility of Normal and Aphasic Subjects

Charles R. Flowers
University of Washington, Seattle, Washington

Marilyn Wyse
The Mason Clinic, Seattle, Washington

In helping aphasic patients improve their ability to express themselves, a total communication approach is sometimes taken. An aphasic patient is encouraged to use all expressive modalities, and the criterion for task accuracy is based on the success of message transmission, including via gestures. The communication of some patients with moderate to severe oral verbal expressive disorders is enhanced when gestures are used in place of (or along with) speech attempts.

Helping aphasic patients improve gestural communication is probably not a top priority for most aphasia clinicians. This is understandable partly because gestural communication is less desirable as a communication system than other means of expression. It is less efficient, less diverse, and has more ambiguity compared with spoken, written or manual sign language communication. Secondly, it is often an unnatural means of communication when used in place of speech. It calls attention to itself and contributes to a communicator's bizarreness. Finally, some aphasic patients have severe gestural deficits and do not improve gestural ability with treatment. In spite of these problems with gestural communication, we see an occasional patient in therapy for gestural training. In order to assess the benefits of treatment with pre- and post-therapy testing, an overall index of gestural communication would be helpful.

From a functional communication standpoint, one overall measure of interest is intelligibility, that is, how well a person's gestures are understood. Yorkston and Beukelman (1982) found that clinicians could not estimate dysarthric persons' speech intelligibility with adequate reliability or validity when the clinicians had prior familiarity with sentences and words on the AIDS (Assessment of Intelligibility of Dysarthric Speech). In this study, we started to answer some similar questions about gestural intelligibility assessment that were addressed by Yorkston and Beukelman for dysarthric speech intelligibility. We were interested in the validity and interrater reliability of clinicians' judgments of gestural intelligibility. The test situation was one in which the clinicians were familiar with the items that testees were gesturing.

Our first step was to test normal persons' ability to gesture intelligibly to other normal persons. The purposes of this part of the study were (1) to get some insights into the factors that contribute to intelligibility, (2) to use this information to select items for a short test of gestural intelligibility, and (3) to investigate the range of normal persons' gestural ability.

Thirty normal couples volunteered for this part of the study. In each couple, one person participated as a gesturer and the other as a receiver. The persons were familiar to each other as friends or relatives. Mean age of the gesturers was 44 years with a range of 16 to 86, and the mean age of the receivers was 45 with a range of 16-79.

Two tests of gestural ability were administered to these couples. The first was a Draw-in-the-Air Test. Gesturers were shown 28 cards, each with a

symbol printed on it and a label for the symbol, for example, a question mark with the word "question mark" under it. Gesturers attempted to draw the items in the air with their nonpreferred hand. This was the left hand for 27 subjects and the right hand for three subjects. The receivers drew what they thought was being gestured and also wrote a label for it on an answer sheet. They did not know the items beforehand.

Intelligibility results for the 28 items are shown in Table 1. As shown, intelligibility for the items ranged from 40 to 100 percent. The mean percent correct for the group was 81 with a range of 54 to 100 percent and a standard deviation of 12.4.

Table 1. Percent intelligibility for 28 draw-in-the-air items (N = 30 couples). Items on the 10-item final test are underlined.

circle	100	square	90	letter C	77
figure 8	97	cross	87	heart	73
letter Z	97	triangle	87	number 3	63
letter B	97	letter M	87	daisy	63
question mark	97	quotation marks	83	star	60
letter L	93	exclamation point	83	number 9	53
dollar sign	93	checkmark	83	arrow	53
letter X	90	letter S	80	quarter moon	40
diamond shape	90	number sign	77		
letter W	90	number 5	77		

The second test involved pantomiming object usage. The gesturers saw 35 cards with the names of objects printed on them. Their task was to demonstrate to the receivers how the objects would be used, again with their nonpreferred hand. The receivers wrote down what they thought was the name of the object. Percent intelligibility results are shown in Table 2. The range of intelligibility for these items was 33 to 100 percent. The mean percent correct for the 30 couples was 74 with a range of 46 to 91 percent and a standard deviation of 15.9.

Table 2. Percent intelligibility for 35 Pantomime Object Usage items (N = 30 couples). Items on the 10-item final test are underlined.

telephone	100	salt shaker	83	fan	70
comb	97	basketball	80	key	67
toothbrush	97	hypodermic needle	80	paintbrush	67
mirror	97	dice	77	flashlight	63
lipstick	93	bubble stick	77	telescope	63
razor	93	bubble gum	77	teacup	63
needle/thread	93	yo yo	73	saw	63
barbells	90	pitcher	73	facial tissue	63
hat	90	drinking glass	70	trumpet	57
bowling ball	90	thermometer	70	hammer	50
fly swatter	87	dart	70	spatula	33
camera	87	frisbee	70		

To prepare for the next phase of the study we shortened each test to 10 items. We wanted a test in which intelligibility was determined primarily by the gesturer's proficiency and not by the receiver's comprehension ability. We did not use items which appeared to be unintelligible for reasons other than gestural ability. These included those that were ambiguous to the receivers even when gestured well, because the movement was too complex or did not clearly indicate the item. Examples were daisy and quarter moon. We also eliminated items for which a proficient gesture could be perceived as some other item, such as cross which could look like a t or a plus sign. Other items were not used because the gesture looked like a different item when reversed from left to right. For example, a backward 3 looks like E. The 10 items selected for the Draw-in-the-Air Test are indicated in Table 1. The mean number of items correct on these items by the normal couples was 9.3 with a range of 7 to 10 and a standard deviation of 1.0.

In shortening the Pantomime Object Usage Test to 10 items, we chose from among items which had at least 85 percent intelligibility, except for thermometer which was 70 percent intelligible. We chose that item because it appeared particularly sensitive to gestural completeness. The final 10 items are indicated in Table 2. The mean number correct on these items by the normal couples was 9.1 with a range of 7 to 10 and a standard deviation of 1.0.

In the next phase of the study we were interested in clinicians' judgments of intelligibility on the two tests. Eight aphasic subjects representing a range of aphasia type and severity and two normal subjects were videotaped as they performed the test items. Before having speech and language pathologists score the tests, we established the level of intelligibility of the 10 taped gesturers on each item. The videotape of each gesturer was shown to 10 freshman and sophomore university students enrolled in speech and hearing sciences classes. The students had not had a course in aphasia. Each gesturer was viewed by a different set of students, for a total of 100 student judges. It was necessary to have a set of students view only one gesturer because the measure of interest was intelligibility to persons who were unfamiliar with the gesturers and items.

As the students watched the videotapes, they wrote down what they thought was being gestured. They were encouraged, but not forced, to guess. Responses were scored as correct or incorrect. A taped gesturer's intelligibility on each test was the total number of correct answers by the 10 students who viewed that gesturer. The variability of the students' answers was great, suggesting that gestural intelligibility of persons with deficits is influenced in large part by the comprehension ability of the receiver. Some persons are better guessers than others. Of the 2000 answers (2 tests x 10 gesturers x 10 items x 10 students), 29 percent were intelligible to all 10 students, 12 percent were intelligible to none, while the remaining 59 percent were intelligible to some but not all. These results suggested to us a method of scoring intelligibility. Testing the validity and reliability of the scoring method was the final objective of this study.

The intelligibility measures generated by the students were compared with scores by 10 certified speech and language clinicians, each with at least two years experience with aphasic patients. Before viewing the tapes, the clinicians were made familiar with the test items and with the scoring procedure. Total time for viewing the tapes was 1.5 hours, during which there were three short breaks.

The clinicians scored each item of all 10 gesturers on both tests as either 0, 1 or 2. A score of 2 was assigned when they judged that all persons viewing a gesture would comprehend it, a 1 when they thought that only some

persons would comprehend the gesture, and 0 when they judged that the gesture would be intelligible to no one.

The total score of the clinicians on an item could range from 0 to 20, since each item was worth 2 points and there were 10 scorers. These scores were correlated with the intelligibility measures established in the earlier study by the student judges who were unfamiliar with the items. Recall that these scores could range from 0 to 10. The Spearman rank order correlations between the clinicians' scores and intelligibility on the draw and pantomime tests were .85 and .82 respectively. In Table 3, items with scores of 0, 1 or 2 are compared with intelligibility for the Draw-in-the-Air Test. For the 300 items that were 100 percent intelligible to the students, 268 (89 percent) were scored as 2. Items that were 80-90 percent intelligible tended to be rated as 1 or 2. Items that were 30-70 percent intelligible clearly should have been rated as 1, that is, intelligible to some but not all persons. Clinicians in error tended to overestimate intelligibility on these items--that is, assign a 2 instead of a 1.

Table 3. Total number of 2, 1 and 0 test scores according to percent intelligibility for the Draw-in-the-Air Test.

Percent Intelligibility	Test Scores		
	2	1	0
100	268	26	6
80 - 90	221	52	7
30 - 70	89	101	30
10 - 20	22	32	66
0	2	11	67

Similar results are shown for the Pantomime Object Usage Test (Table 4.) On this test it appears that items that had moderate intelligibility (30-70 percent) were more likely to be underestimated than overestimated by the clinicians. This was opposite to the trend for the other test. Seventy-six, or 35 percent, of the moderately intelligible items were judged to be intelligible to no one when in fact they clearly were intelligible to some persons. For the 150 items with 0 intelligibility, the clinicians' scores were overestimations on 2/3 of them in contrast to the other test on which 16 percent of the 0 intelligible items were overestimated. Generally, the scores on the first test appeared more valid estimations of intelligibility than on the Pantomime Object Usage Test.

Table 4. Total number of 2, 1 and 0 test scores according to percent intelligibility for the Pantomime Object Usage Test.

Percent Intelligibility	Test Scores		
	2	1	0
100	221	48	1
80 - 90	90	78	12
30 - 70	34	110	76
10 - 20	13	83	84
0	4	56	90

agreement among the clinicians' scores. For the Draw-in-the-Air Test, pertinent data are presented in Table 5. For each taped gesturer, the table shows the range of total scores of the 10 clinicians and how many scores agreed within 1, 2 and 4 points. The smallest ranges of scores were three, four and four points for Subjects 5, 7 and 8 respectively. Subjects 5 and 8 were the normal subjects, and Subject 7 was the most severe aphasic subject. For the remaining aphasic subjects the range of scores across items was from 6 to 10 test points (30 to 50 percent).

Table 5. Agreement among 10 speech pathologists' scores on the Draw-in-the-Air Test for the 10 gesturers.

		Number of Scores in Agreement		
Gesturer	Range of Scores	Within 1 point	Within 2 points	Within 4 points
1	13 - 19	7	8	9
2	11 - 19	6	8	8
3	5 - 15	4	5	9
4	13 - 19	3	6	7
5	17 - 20	9	9	10
6	13 - 20	5	7	8
7	0 - 4	5	7	10
8	16 - 20	9	9	10
9	4 - 12	3	5	7
10	9 - 18	4	5	8

Similar results are shown in Table 6 for the Pantomime Object Usage Test. On this test, scores for the normal subjects were not in any better agreement than for most of the aphasic subjects. The best agreement was for aphasic Subject 3 with a 20 percent range in scores, and the poorest was for Subject 7, with a 55 percent range. Agreement was slightly poorer on this test than on the Draw-in-the-Air Test.

Table 6. Agreement among 10 speech pathologists' scores on the Pantomime Object Usage Test for the 10 gestures.

		Number of Scores in Agreement		
Gesturer	Range of Scores	Within 1 point	Within 2 points	Within 4 points
1	8 - 14	5	6	8
2	10 - 15	4	6	9
3	9 - 13	5	7	10
4	8 - 16	4	4	8
5	15 - 20	7	8	9
6	8 - 13	4	6	9
7	0 - 11	3	4	5
8	7 - 16	5	7	9
9	9 - 15	4	6	9
10	7 - 13	5	8	9

Conclusions.

Normal couples' gestural intelligibility is highly variable. It is not safe to make assumptions about normal gestural intelligibility for an item without testing it.

Receivers naive to test items are highly variable in their ability to guess the gesture a testee is attempting. Clinicians may desire to measure gestural intelligibility by having a patient gesture to another person who is naive about test items, as on the AIDS. Clinicians would be wise to obtain several measures and take the average or highest rating as a patients' intelligibility index.

The 2, 1, 0 scoring system appears to have some validity, but it needs improvement. The 10 clinicians who scored the gestures were not given any information about what constituted 2, 1 and 0 scores beyond scoring category definitions. Additional guidelines, examples and practice perhaps would improve scoring accuracy. These should be item specific, since it is our impression that items differ in what a gesturer needs to do to be intelligible. On some items, it seems clear that a score of 7 or 12 on the PICA (Porch Index of Communicative Ability, Porch, 1981) would be intelligible, while on other items more complete gestures would be required.

Interclinician scoring agreement also was not adequate. It appears that clinicians cannot judge gestural intelligibility reliably when they have prior knowledge of the items.

We have begun to analyze descriptive data from this study to determine which factors contribute to intelligibility. Some are variables other than gestural proficiency. As we tested the normal couples in the first study, it appeared that there was an age effect. We compared scores of couples who were over 60 to those under 40, and there was a significant difference. But as we thought about why that might have occurred, it was not clear at all that it was a true age effect. Some items may have favored the young group. For "camera," for example, some older gesturers pantomimed a box camera by holding it at waist length. This was not as intelligible as gesturing a camera at the face. Some older subjects appeared not to know what a frisbee was. The older subjects also appeared to guess less often, with more failures to respond. Some older subjects admitted inability to recall the names of some items, for example, exclamation point and quotation marks. The point is that in testing gestural intelligibility of aphasic patients, a clinician should attempt to control for testees' familiarity with items and other variables extraneous to the patients' gestural proficiency.

In these tests, the gesturers were performing for the examiner and the camera. They were not attempting to communicate. It was our impression that the normal couples in the first study gestured in a qualitatively different way than the taped gesturers. Most of the normal couples were quite serious about the task and appeared to want to communicate the message, while the aphasic and two taped normal subjects appeared just to be going through the motions. Testing and treatment of gestural communication should be arranged so that the aphasic person is attempting to communicate a gesture, and not simply performing it to command. The Pantomime Referential Abilities Test in the New England Pantomime Tests by Duffy and Duffy (1984) is a step in that direction.

REFERENCES

Duffy, R.J. and Duffy, J.R. New England Pantomime Tests. Tigard, OR: CC Publications, 1984.

Porch, B.E. Porch Index of Communicative Ability. Palo Alto, CA: Consulting Psychologists Press, 1981.
Yorkston, K. and Beukelman, D. Assessment of Intelligibility of Dysarthric Speech. Tigard, OR: CC Publications, 1982.

DISCUSSION

Q: What was your rationale for including the drawing of figures test?
A: We thought it would be more sensitive to some patients' gestural impairment than the pantomime test. The pantomime test is similar to the traditional tests of praxis. The draw-in-the-air test requires different kinds of movements, finger movements that are two-dimensional in space with more fine, as opposed to gross, movements. Also, some patients are able to communicate to some degree by indicating the size or shape of objects or by drawing an actual symbol in the air. So, there is an interest in these items from a communication standpoint.

Q: Did you find that normal subjects tried to draw the pantomime objects in the air?
A: No, but then their instructions clearly indicated that they were to demonstrate how the objects would be used. Their task was to communicate the object by demonstrating use, and not by some other kind of gesture. I do recall one subject who had difficulty with thermometer, and he finally resorted to drawing its shape.

Q: Did you make any observations about body part as object?
A: Yes, we did. We noted when that occurred in the first study with the normal couples, and it was not uncommon. The normal gesturers did use body part as object on some of the items. I wonder if that behavior can be considered abnormal.

Q: Yes, we have some data on normal young and old subjects. Depending upon the item, they will not infrequently use body part as object. I am skeptical of using that as an index of brain injury.
A: I agree. There were some of our items where using a body part as object helped considerably to communicate the object in pantomime. I wonder about the validity of Goodglass and Kaplan's statements on that issue.

Q: I'm looking forward to your investigating the localization of gestural impairment in brain damaged patients. The literature suggests that high parietal, left hemisphere lesions produce gestural deficits. I wonder if you have made any observations about localization?
A: No, we have not studied that ourselves. This study was confined to measurement issues and variability of normal performance. The study grew out of a clinical problem. We were seeing some patients for gestural training, and we were looking for ways to measure gestural communication improvement.

Q: I wonder whether recognition or pantomime ability was affected by the sex of the gesturer, like a male putting on lipstick as opposed to a woman doing that.
A: We anticipated a problem there, but did not find it. Males were not reluctant to use the items more associated with females, and vice versa. A few of the females did gesture a razor by movements across the legs, and I believe that was less intelligible than movement across the face. Nevertheless, in testing, it might be desirable to eliminate items with a sex bias.

Q: I would think that, even if two gesturers were doing the same thing, it might be more intelligible if, for example, a male gestured an item that ordinarily is used by males than if a female were gesturing. Did you allow facial gestures?

A: Yes. That's another factor that makes the second test different from the first. Acting ability, including facial expression, seems to help performance on the pantomime test, and not on the draw-in-the-air test.

Q: We are grappling with this same problem in Toronto in our neuropraxis research program. We are studying the three modalities, and what we have tried to get away from is a subjective analysis of all of the behaviors. Instead, we have moved towards a 9-point error system in which we look at the location in which the gesture was produced, the posture of the hand, and whether or not the action was correct. We hope to make behavioral inferences from that as to what the concept is within the person. We are also moving to a zero, one, two error system after we do the second level of analysis.

A: I think assessing gestural proficiency requires quite a different approach than testing for gestural intelligibility. Our procedures were designed more for testing intelligibility. The items could perhaps also be used for testing aspects of proficiency. In assessment of gestures, a component analysis is just as crucial as obtaining an overall measure. Your work should help us know which components to analyze, and perhaps which aspects are important for intelligibility. Some aspects of deficient gesturing may help indicate something about the way the patient is attempting the task and the nature of brain injury, but not be an important contributor to intelligibility. It would be important to know what is deficient about a person's gestures to serve as a basis for treatment.

ACKNOWLEDGMENT

The assistance of Amy Clark, Marsha Adams, Winona Lau, Sheila Hallauer, Margaret Redenbaugh, Tracey Yonick, Elaine Peizer, Joan Jaeger, Kathryn Yorkston, Joyce Hedges, Vicki Hammen, and Kristine Keough is gratefully acknowledged.