Administering The Token Test *Senza Intonazioni Particolari*: Does It Make Any Difference?

James R. Solomon
Arnold E. Aronson

Mayo Clinic
Rochester, Minnesota

For a diagnostic test that enjoys such clinical currency as the Token Test (TT), it is remarkable that instructions for administering it are so few and so incomplete. Then again, perhaps it is in part *because* of its apparent simplicity that the test is so popular. Certainly its freedom of administration has something to do with the existence of so many forms of the test (see Berry, 1976; Pettit and Heath, 1975).

One of the newest forms of the TT is that by DeRenzi and Faglioni (1975 [1976]). In their article, in which they introduce an abbreviated form of the test, the authors say (1975:257) that the commands should be read 'senza intonazioni particolari' (literally, 'without particular intonations'). This echoes the injunction against delivering the commands with 'any special prosodic emphasis' that is familiar from DeRenzi and Vignolo's original (1962) statement of the test.

It is assumed that these phrases refer to the use of distinctive intonation and stress patterns, such as occur in the use of contrastive stress, a prosodic feature of Italian, English, and many other languages. Contrastive stress can be used for several purposes. One is to highlight a difference between what is being said at the moment and either what has already been said or what is about to be said. Sometimes the contrast is apparent in the words used:

(1) No, I didn't say three *months*; I said three *days*.

In this example the two stressed words have a semantic similarity. In other instances the relationship may be largely phonological, as between two bisyllabic or polysyllabic words that share a common sequence of segments, and the use of contrastive stress to highlight the elements that are different may cause the usual syllabic stress in the second word to be shifted to another syllable:

(2) The first thing they heard about the situation was a sketchy report on television; then they received a tele*gram* that confirmed their worst fears.

At other times the contrast may not be expressed on the 'surface' of lexical items; instead it may be inherent in the context of the discourse, and the contrast may extend over a considerable amount of time. For example after a lengthy report on a meeting of ministers of the Oil Producing and Exporting Countries, in which the word *oil* is stated only at the outset and the word energy is not used at all, a newscaster begins the next item by saying:

(3) In *this* country today, *another* form of energy was in the news; at a conference of the Association of *Coal* Manufacturers...

The nature of the commands in the TT offers repeated opportunities to use contrastive stress. For example, in Part I of the test each succeeding command follows rather quickly upon the patient's response to the previous command; consequently, it is natural to read them in this manner:

			A	B	Type of shift
	M	M	H	F	
(4a)	Touch	the	green	circle.	--
	M	M	F	L L	
(4b)	Touch	the	white	circle.	A
	M	M	L	F	
(4c)	Touch	the	white	square.	B
	M	M	H-R	F	
(4d)	Touch	the	blue	circle.	AB

The letters M, L, and H indicate syllables produced with relatively mid-, low-, and high-level pitches or tones within the speaker's range. The letter F indicates a pitch that begins high and then falls to the lowest frequency of the sentence. The combination H-R refers to a pitch that is initially high, drops to a mid-level, and then rises again, although not necessarily to the level of the preceding high pitch.[1] The primary stress in each sentence, indicated by the underlining, occurs only in words produced with a falling pitch.[2] The letters A and B above the words for colors and shapes, respectively, refer to the type of 'shift' in focus or emphasis that is signalled by the position of the primary stress and the falling pitch in each sentence after the first; that is, sentence (4a) provides the context in which contrastive stress may operate in sentence (4b), sentence (4b) provides the context for (4c), and so on. These sentences exhaust the possibilities of contrastive stress in Part I of the test, corresponding to a shift in color (A), shape (B), or both color and shape (AB).

In Part IV of the test the situation vis-à-vis stress is somewhat more complex, mirroring the greater linguistic complexity of the commands generally. The use of contrastive stress in this part can signal semantic shifts between sentences (as in Part I) and also within them. To prevent the discussion from becoming too complex, only the latter situation will be considered; the following examples exhaust the possibilities:

			A	B	C						Type of shift
	M	M	M	M	R	M	M	F	L	L	
(5a)	Touch	the	small	yellow	circle	and	the	large	white	circle.	A
	M	M	M	M	R	M	M	M	F	L	
(5b)	Touch	the	large	yellow	square	and	the	large	blue	square.	B
	M	M	M	M	R	M	M	M	M	F	
(5c)	Touch	the	large	white	circle	and	the	large	white	square.	C
	M	M	M	M	R	M	M	H-R	F	L	
(5d)	Touch	the	small	yellow	circle	and	the	large	green	circle.	AB
	M	M	M	M	R	M	M	M	H-R	F	
(5e)	Touch	the	small	red	square	and	the	small	green	circle.	BC
	M	M	M	M	R	M	M	H-R	L	F	
(5f)	Touch	the	large	blue	circle	and	the	small	blue	square.	AC
	M	M	M	M	R	M	M	H-R	H-R	F	
(5g)	Touch	the	small	green	square	and	the	large	yellow	circle.	ABC

The letter R above the first noun in each sentence signifies a rising intonation, which is realized as a rise from the preceding mid-level pitch to a higher pitch, but not as high as the high pitch(es) in the second noun phrase. In these conjugate sentences the context is established by the first noun phrase, and contrastive stress may be used to highlight those elements—size (A), color(B), or shape (C)—in the second noun phrase that are different from the first. As may be seen, there are seven types of 'shift' possible in Part IV-type TT commands.

Reading TT stimuli in the fashion previously suggested above (that is, with 'special prosodic emphasis') is probably what DeRenzi, Vignolo, and Faglioni want us to avoid in administering the test. However, in neither the 1962 or 1975 presentations of the test nor in other studies by these authors and their colleagues is there further explanation of what these words mean or why this factor is considered important. In fact, as we observed several experienced clinicians read the test, we have noticed stress-intonation variations in the delivery of the commands among examiners, among readings by a single examiner on different occasions, and even within a single reading. The possibility exists that this variability may affect performance to such an extent that positive or negative changes in a patient's auditory comprehension as measured by this test, in either the presence or absence of therapy, may be obscured. This possibility provided the original impetus for this study, and its goal is to determine whether administering the TT with or without 'special prosodic emphasis' makes a significant difference in patient performance.

Method

Subjects

Ten aphasic and nine normal subjects participated in the study. All were native speakers of American English. None of the normal subjects had previous experience with the TT, and the aphasic subjects either were unfamiliar with the TT or had not been tested with it recently. All subjects were required to pass a screening battery for the TT, consisting of correctly identifying (either *ab initio* or, if necessary, after a brief training session) tokens for the five colors, two shapes, and two sizes used in the test.

The normal subjects, all women, were clerical personnel at the Mayo Clinic, except for one who was a medical student. During a preliminary interview these subjects gave no evidence of any history of neurologic disease that could cause or be related to a language problem, or any remote or recent history of a language, emotional, or intellectual disorder that might affect performance on the TT. The age range of these subjects was 19 to 25 years, with a mean of 21.3 years (SD = 2.3), and they had between 12 and 20 years of education, with an average of 13.2 years (SD = 2.6).

The aphasic subjects, seven men and three women, ranged in age from 34 to 79 years, with a mean of 59.4 years (SD = 14.3) and had between 7 and 16 years of education, with a mean of 10.9 years (SD = 3.0) (Table 1). These subjects had been identified as demonstrating aphasic language behavior by a consultant in speech pathology at the Mayo Clinic following a diagnostic examination. None of the aphasic subjects gave evidence of life-long mental retardation, senile or presenile dementia, psychiatric disorders, or a language or perceptual deficit limited to one modality. All had lesions confined to the left hemisphere, as determined by a consultant in neurology at the Mayo Clinic using clinical and laboratory data derived from standard measures, such as reflex and muscle testing, sensory testing, electroencephalography, angiography, brain scan, and computerized axial tomography. All aphasic subjects had been judged able to tolerate a testing session of approximately 1½ hours. In addition to performing satisfactorily on the screening measure previously described, the aphasic subjects had to achieve a score of at least one item correct on Part I and at least one item incorrect on Part IV of a 'standard' administration of the TT.[3]

Table 1. Characteristics of aphasic subjects.

Subj no.	Sex and age (yr)	Education (yr)	Time after onset (mo)	Cause	Token Test scores (no. correct): Parts I-IV (40 items)	Part V (21 items)	Total (61 items)
1	M, 63	12	71	Trauma	24	11	35
2	F, 75	16	½	Vascular	35	4	39
3	M, 69	14	12	Vascular	31	3	34
4	F, 79	8	½	Vascular	17	2	19
5	M, 49	8	¼	Tumor	28	14	42
6	M, 34	12	1	Vascular	13	3	16
7	M, 64	7	?	Tumor	32	6	38
8	M, 61	12	?	Tumor	28	5	33
9	F, 41	12	60	Vascular	34	12	46
10	M, 59	8	¼	Trauma	8	4	12

Conditions

After the initial screening procedure, a 61-item version of the TT (that of Boller and Vignolo, 1966, less item 11 of Part V and with squares substituted for rectangles throughout) was administered to each subject. In addition to providing data with which to determine further participation in the study by the aphasic subjects, this administration supplied baseline data on performance in Parts I and IV for comparison with performance in other conditions.

The style employed in delivering these stimuli was a rising intonation at the end of each sentence or each noun phrase. Normal rhythm and normal pauses (specifically at the end of the first noun phrase in Part IV-type commands) were maintained. This manner of delivery is considered one possible interpretation of DeRenzi and Vignolo's injunction against using special prosodic emphasis; it avoids introducing the cue of contrastive stress and at the same time is not an unfamiliar stress-intonation pattern for speakers of English.[4] This style—rising intonation with each command and no use of contrastive stress—constituted the control condition for the study.

For the experimental portion of the study four additional styles of delivery were employed:

1. Condition I was a monotone-monostress reading, in which normal rhythm was maintained. The purpose of this condition was to test the hypothesis that removal of the suprasegmental cue of not just stress (as in the control condition) but intonation as well would produce significant differences in speed and accuracy of performance when compared with performance in the control condition.

2. Condition II was a reading that employed contrastive stress, realized in the manner set forth in the introductory remarks. Lehiste (1976) has pointed out that although fundamental frequency is the most potent cue in the perception of stress, the parameters of duration and intensity are also important. Accordingly, in reading the stimuli in the contrastive stress condition, no attempt was made to vary fundamental frequency only, while holding the other parameters constant. Thus when words were singled out to receive contrastive stress, those words (or in bisyllabic words, the syllables in those words that normally receive lexical stress) were produced with a combination of higher fundamental frequency, greater duration, and greater intensity than the corresponding words (or lexically stressed syllables in words) occupying syntactically parallel positions in the immediately preceding noun phrase. The purpose of this condition was to test the hypothesis that the use of the suprasegmental cues of stress and intonation would produce significant differences in speed and accuracy of performance when compared with performance in the control condition (stress but not intonation removed) and condition I (both stress and intonation removed).

3. Condition III was identical to condition II except that the final word of each command was omitted. The purpose of this condition was to test the hypothesis that contrastive stress and the sentence intonation pattern in which it is embedded constitute a semantic cue of such strength that in a succession of TT commands native speakers of English can predict a missing final word and execute the command correctly significantly more often than by chance. Stimuli for this condition were read in the same manner as for condition II, except for the omission of the final word.

4. Condition IV was identical to condition II, including the manner of realizing contrastive stress, except that in each command the word selected to receive contrastive stress was semantically inappropriate. An element of inappropriateness was present also in the control condition, in the sense that naturally occurring opportunities to use contrastive stress were not acted on. In condition IV, however, contrastive stress was employed, but it was placed on a word that normally would not have received such stress. For example, in Part I-type commands, if the preceding command was 'Touch the green circle' and the next command was 'Touch the white circle', it would normally be realized as

M M F L
(6a) Touch the white circle.

With a deliberate attempt to mislead the listener, however, it would be produced as

M M L F
(6b) Touch the white circle.

Similarly, in Part IV-type commands a sentence might be realized in condition IV as

M M M M R M M M F L
(7a) Touch the small red circle and the large red circle

or

M M M M R M M M M F
(7b) Touch the small red circle and the large red circle

rather than as the usual

M M M M R M M F L L
(7c) Touch the small red circle and the large red circle

The purpose of this condition was to test the hypothesis that contrastive stress is a sufficiently powerful semantic cue that, even if subjects hear the segmental information correctly, the incorrect suprasegmental information contained in misplaced contrastive stress can interfere with performance to the extent that subjects' speed and accuracy of response in this condition will differ significantly from the same in the control condition and condition II (correct contrastive stress).

Experimental Stimuli

The experimental stimuli were contained in four lists, each consisting first of 13 commands modeled on those in Part I of the TT and then of 12 modeled on Part IV. Since in Part I-type stimuli contrastive stress operates only between (and not within) commands, the first command in each of the four lists was a dummy item, to which each subject responded without knowing that his response would not be scored. (This item was always read with the same rising intonation used in the control condition.) Thus there was a possibility of 24 scorable items in each list and a total of 96 in the four lists. Individual stimuli were selected by use of a table of random numbers and a set of restrictions that assured equal or almost equal representation, both within and across lists, of each experimental condition, each shift type, and each token color, size, and shape.

The four lists of stimuli and the 61-item TT were read by the first author and recorded with an Ampex reel-to-reel tape recorder, model AG600. An assistant using a stopwatch gave hand signals to begin reading the next command in each part, with a goal of a 7 second interval between stimuli; the resulting mean interval between the Part I-type stimuli on all four experimental lists was 6.9 seconds (range = 6.2-7.4; SD = 0.3); for the Part IV-type stimuli, 7.2 seconds (range = 6.8-7.8; SD = 0.3); and for the 61 TT items read in the control condition, 7.3 seconds (range = 6.8-8.0; SD = 0.3). The TT and the four sets of experimental stimuli so recorded were then copied onto five separate cassette tapes by means of a Craig cassette tape recorder, model 2619.

Procedures

After the initial screening test, each subject was first given the full TT and then the four lists of experimental stimuli. A Latin-square design was employed to establish four different orders of presentation for the four lists, with every fifth subject in each group receiving the same order. Before being presented each set of experimental stimuli (that is, after the TT and each subsequent list), the subject had a rest period of approximately four minutes, the nature of which (conversation, silence, and so forth) was determined by the subject.

The stimuli were presented in a quiet room on the same model Craig cassette recorder on which they had been copied; the machine was AC-powered, and the loudness was set at a comfortable range for the subject. The recorder was placed on the far side of a table from the subject, with the tokens lying between; the examiner (the first author) sat to the subject's right. Forward movement of the tape was controlled by the examiner by means of the on-off switch of the recorder microphone; the tape was stopped whenever it appeared that the subject would not be able to respond during the 7 second interstimulus interval.

The tokens were made of pressed wood; the large circles were 50 mm in diameter, the large squares were 50 mm on each side, and the small circles and squares, 25 mm. They were arranged before the subject in a random pattern, but the same pattern was used for each subject. Except for Part V of the TT, subjects responded to all stimuli by tapping the tokens with the end of a metal pen held in the hand preferred by the subject. Subjects were uniformly instructed to wait until the command was read in its entirety and then to move as quickly as possible. They were reminded regularly of the need to respond quickly, and every time they began to move prematurely, they were reinstructed in the importance of waiting to respond.

Responses as observed by the examiner were entered on a score sheet for later evaluation. The sound of the subject's striking the tokens with the pen, as well as the sound of the stimuli emanating from the cassette recorder, was detected by an RCA microphone, model BK-6B, lying on the table near the cassette recorder and was recorded on the same model Ampex recorder on which the stimuli had originally been recorded. The Ampex machine was located behind the examiner and was controlled by him; the tape speed was 3-3/4 ips.

Scoring

Subject performance was evaluated in three ways. The first was a simple pass-fail score based on the examiner's observation of each response. The same observation provided the input for the second method of evaluation, a weighted system of scoring based on the probabilities of the subject's selecting the correct elements of the stimulus (color, shape, and, where appropriate, size) by chance. Each Part I-type stimulus was thus worth 7 points: 5 for choosing the correct color out of the five and 2 for the correct one of the two shapes or two sizes. Each Part IV-type stimulus was worth 20 points: 2 for each correct size, 5 for each correct color, 2 for each correct shape, and 2 for selecting the two tokens in the order in which they had been presented.

The third method of evaluating performance was the determination of latency of response, in terms of reaction time (RT). The measurement of RT is a classic technique of experimental psychology (see Woodworth and Schlossberg, 1954, ch. 2); recently it has been explained by Welford (1977: 463) in this manner:

> The reaction time, that is the period elapsing between the appearance of a signal and the beginning of a responding movement, is commonly regarded as a measure of the time taken by central processes. Obviously, it includes some time taken by peripheral processes as well, but...this is usually short compared with that taken by perception, translation from perception to action, and the shaping and initiation of the responding movement. The central processes which take place during reaction time may be of greater or less complexity, and the length of the reaction time is commonly regarded as a measure of this.

In RT studies the stimulus is typically of short duration, for example, a briefly flashing light); thus the beginning point for the measurement of the response latency is, as described by Welford, the appearance of the stimulus. In the current study, the stimuli are of considerably greater duration; for that reason, the onset of the RT period was taken as the termination of the stimulus, after which subjects were allowed to respond.

RTs were computed from the end of the stimulus to the sound of the subject's tapping the pen against the token (for Part IV-type stimuli, the second token). This was accomplished by playing back, on the Ampex recorder, the tape that was recorded during the testing session, the signal being amplified through an Ampex amplifier-loudspeaker, model AA620. From the amplifier the signal was fed into a Sanborn Viso strip-chart recorder, model 100. Gain (that is, the amplitude of the tracing) on the strip-chart recorder (SCR) was controlled primarily with the volume control of the amplifier-loudspeaker; by this means it was usually a simple matter to correlate the audio signal with deflections of the SCR stylus and thus to identify with confidence the end-point of the stimulus and the end of the response. The speed at which the paper traveled through the SCR (25 mm/s) allowed for estimating RT within 0.01 s; this was done by visual examination of the tracing on the SCR. This method of recording and analyzing RT data is essentially the same as that used by Saxman and Fay (1970).

The results obtained from the three methods of scoring were, when appropriate, submitted to statistical analysis using a two-tailed Student's t test. The minimal level of significance adopted was a $p < 0.05$.

Results

All subjects, both normal and aphasic, passed the screening test requiring simple identification of colors, shapes, and sizes. Some aphasic subjects initially demonstrated signs of agnosia for shape or color, but after a short training session each made reliable correct responses.

Pass-fail scoring (Table 2)

For the pass-fail scoring, the percent correct figures for the control condition were based on the number of correct responses each subject made to the ten items in Part I and the ten items in Part IV of the TT. For the four experimental conditions, the figures for each condition represent a pooling of subject performance on the three occurrences of that condition in each of the four lists (this applies to all discussion of results by experimental conditions to follow).

For the simpler commands, all normal subjects performed perfectly on Part I of the TT and on the experimental stimuli in conditions I, II and IV. For the same sets of stimuli, the aphasic subjects showed small, statistically nonsignificant differences in scores. For condition III, however, in which commands were delivered with contrastive stress but with the last word omitted, both groups of subjects showed a considerable decrement in performance; when compared with performance in each of the other conditions, the scores for condition III were significantly lower ($p<0.01$-0.001), for both groups. Interestingly, for condition III the aphasic subjects as a group obtained higher scores than did the normal subjects (a mean correct score of 49% vs. 34%); this difference was not statistically significant, however.

For the more complex stimuli, the normal subjects again performed perfectly (or nearly so) in the control condition (Part IV of the TT) and on the Part IV-type stimuli in experimental conditions I, II, and IV. As might be expected, the aphasic subjects' scores in the same conditions were significantly lower than both the normal subjects' scores for the same stimuli ($p<0.001$ in all cases) and their own scores for the same conditions with the

		Pass-fail scoring			Weighted scoring		
		No. of items	$\overline{X}$ (% correct)	Range (% correct)	No. of points	$\overline{X}$ (% of points)	Range (% of points)
Part I-type commands							
Aphasic subjects (N = 10)	Control	10	84	20–100	70	93	59–100
	Cond. I	12	89	58–100	84	96	88–100
	Cond. II	12	88	50–100	84	95	69–100
	Cond. III	12	49	25—83	84	83	71—95
	Cond. IV	12	86	33–100	84	93	68–100
Normal subjects (N = 9)	Control	10	100	(100)	70	100	(100)
	Cond. I	12	100	(100)	84	100	(100)
	Cond. II	12	100	(100)	84	100	(100)
	Cond. III	12	34	17—67	84	81	76—90
	Cond. IV	12	100	(100)	84	100	(100)
Part IV-type commands							
Aphasic subjects (N = 10)	Control	10	29	0—60	200	83	69—93
	Cond. I	12	43	8—75	240	86	70—98
	Cond. II	12	51	8—92	240	89	76—97
	Cond. III	12	25	8—52	240	83	77—95
	Cond. IV	12	41	0—83	240	84	68—98
Normal subjects (N = 9)	Control	10	100	(100)	200	100	(100)
	Cond. I	12	98	92–100	240	99	98–100
	Cond. II	12	99	92–100	240	99	98–100
	Cond. III	12	48	25—75	240	95	93—98
	Cond. IV	12	100	(100)	240	100	(100)

Table 2. Performance of Aphasic And Normal Subjects With Part I- And Part IV-Type Commands: Analysis By Pass-Fail And Weighted Scoring Methods.

simpler (Part I-type) commands ($p<0.001$ in all cases). Furthermore, the superiority of performance that the aphasic subjects demonstrated over the normal in condition III with the Part I-type commands was reversed with the Part IV-type commands (25% for the aphasic vs 48% for the normal subjects), and this difference was statistically significant ($p<0.02$).

Weighted scoring (Table 2)

With this scoring method points were awarded not only for correct selection of individual elements of the token but also, with Part IV-type commands, for selection of the two tokens in the order in which they were mentioned in the command. However, no points were awarded for correct performance of the action (touching). The results of testing justify the latter two decisions. No subject responded to the command 'touch' by doing anything other than that; thus assigning points in the weighted scoring to the verb would not have enhanced any differences in performance. All normal subjects responded to the Part IV-type stimuli in terms of order of mention, whereas all aphasic subjects but one committed errors of reversal of order. More than a third of the responses of two aphasic subjects to these items involved reversals. More than two-thirds of such errors in all the aphasic subjects who committed them were obvious during testing; the remainder appeared during the analysis of results.

The differences in performance between the aphasic and normal groups were not as dramatic with weighted scoring as with pass-fail scoring. As mentioned previously, for each of the ten items in Part I of the TT and the twelve Part I-type stimuli, 7 points were possible, for a total of 70 points for Part I and 84 for each of the four experimental conditions. In the control condition and conditions I, II, and IV, normal subjects' perfect scores remained unchanged, and the aphasic subjects' scores improved between 7 and 9% (percent ot total points vs. percent correct in pass-fail scoring). A dramatic change occurred in condition III: the superiority (a 15% difference in mean number of items correct) that the aphasic subjects demonstrated over the normal, when analyzed in terms of pass-fail scoring, was reduced to only 2% with weighted scoring and remained statistically non-significant.

For the Part IV-type commands the normal subjects' performance in all but condition III remained virtually unchanged from the pass-fail analysis. For the aphasic subjects, however, weighted scoring produced an increase in their scores in these conditions of between 39 and 54%. For condition III, the aphasic subjects increased from an average of 25% correct with pass-fail scoring to an average of 83% of the total possible points in weighted scoring. For the same condition the normal subjects increased their mean scores from 48 to 95%, maintaining their superiority in performance over the aphasic subjects for these truncated stimuli delivered with contrastive stress, but by much less (a superiority of only 12%, vs. the previous 23%). However, largely because the range of normal subjects' scores with weighted scoring was much less than with pass-fail scoring (93 to 98% of total points vs. 25 to 75% correct), this approximately 50% smaller percentage difference in scores attained a somewhat higher degree of statistical significance ($p<0.01$) than did the difference in pass-fail scoring ($p<0.02$).

Latency of response (Table 3)

For analysis by latency of response (reaction time), data from only nine aphasic subjects were used, whereas in the pass-fail and weighted scoring

		Latency											
		Correct responses				Incorrect responses				Combined responses			
		No. of Ss	No. of Rs	$\bar{X}$ (sec)	Range (sec)	No. of Ss	No. of Rs	$\bar{X}$ (sec)	Range (sec)	No. of Ss	No. of Rs	$\bar{X}$ (sec)	Range (sec)
		Part I-type commands											
Aphasic subjects	Control	9	79	1.52	0.95–2.42	3	10	1.23	0.35–1.39	9	89	1.49	0.95–2.42
	Cond. I	9	92	1.70	0.83–3.29	3	8	2.38	1.61–2.71	9	100	1.76	0.83–3.06
	Cond. II	9	93	1.66	0.77–2.36	2	7	2.16	1.89–2.28	9	100	1.69	0.77–2.36
	Cond. III	9	51	2.35	1.77–3.64	9	46	2.39	1.42–2.73	9	97	2.37	1.84–2.98
	Cond. IV	9	87	1.68	0.83–3.16	3	3	1.80	1.67–2.09	9	98	1.69	0.83–2.92
Normal subjects	Control	9	86	0.34	0.22–0.52	0	...	...	...	9	86	0.34	0.22–0.52
	Cond. I	9	104	0.41	0.23–0.73	0	...	...	...	9	104	0.41	0.23–0.73
	Cond. II	9	105	0.42	0.29–0.59	0	...	...	...	9	105	0.42	0.29–0.59
	Cond. III	9	37	1.04	0.73–1.28	9	67	0.96	0.67–1.57	9	104	0.99	0.69–1.36
	Cond. IV	9	100	0.45	0.29–0.95	0	...	...	...	9	100	0.45	0.29–0.59
		Part IV-type commands											
Aphasic subjects	Control	8	27	5.95	2.12–20.85	8	58	5.95	2.26–11.41	9	75	5.95	2.22–13.05
	Cond. I	9	43	3.78	1.69– 7.87	8	55	6.64	1.53–12.41	9	98	5.38	1.64–11.86
	Cond. II	9	41	5.18	1.94– 8.19	7	53	6.03	1.74–12.82	9	94	5.66	1.84–11.12
	Cond. III	9	29	5.00	2.71– 8.29	9	66	5.59	2.58–10.95	9	95	5.41	2.66– 9.62
	Cond. IV	7	42	4.16	2.36– 9.47	9	52	6.48	2.24–13.36	9	94	5.45	2.29–13.14
Normal subjects	Control	9	88	0.78	0.59– 1.39	0	...	...	...	9	88	0.78	0.59– 1.39
	Cond. I	9	105	0.76	0.50– 1.16	2	2	2.57	0.96– 4.18	9	107	0.80	0.50– 4.18
	Cond. II	9	106	0.86	0.67– 1.15	1	1	0.82	(0.82)	9	107	0.86	0.67– 1.15
	Cond. III	9	51	1.48	0.93– 2.05	9	54	1.23	0.94– 1.55	9	105	1.35	0.94– 1.89
	Cond. IV	9	107	0.84	0.67– 1.08	0	...	...	...	9	107	0.84	0.67– 1.08

Table 3. Performance of aphasic and normal subjects with Part I and Part IV type commands: analysis by latency of response (reaction time) method.

the results of ten subjects were used. One aphasic subject frequently began to respond before the stimulus was completed (the beginning point for the RT calculation), despite repeated reminders; his RTs were consequently not taken into account.

The RT data are presented in categories of correct responses, incorrect responses, and the combination of the two. The breakdown into correct and incorrect RTs was because differences in RT for those two types of response have been recognized in the literature. In the current study, only the aphasic subjects' responses to the Part IV-type stimuli permit investigation of this question.[5] Paired dependent $\underline{t}$ tests for the mean RTs of these subjects' correct and incorrect responses suggested significantly different results only for condition IV ($\underline{t}_d = 5.127$; $\underline{df} = 6$; $\underline{p} < 0.01$); the differences between correct and incorrect responses for the other four conditions were not significant.

When combined (correct plus incorrect) responses were analyzed, only the difference in mean RT between the control condition and condition II (contrastive stress, complete sentences) for the normal subjects when responding to Part I-type stimuli was statistically significant, and only at a borderline level ($\underline{p} < 0.05$). All other comparisons of mean RTs for the control condition with those of conditions I, II, and IV, as well as comparisons among these three experimental conditions, for both Part I- and Part IV-type commands and for both the normal and aphasic subjects, showed no significant differences. When RTs for condition III were compared with any of the other four conditions, they were significantly slower for the normal subjects for both types of commands ($\underline{p} < 0.001$ in all cases) and for the aphasic subjects for the simpler stimuli ($\underline{p} < 0.005$-0.001). For Part IV-type stimuli, however, comparisons of the aphasic subjects' mean RTs for condition III with the other conditions showed no significant differences.

Discussion

The results of this study are essentially negative: the presence or absence of the prosodic features of stress and intonation do not seem to produce any consistent differences in the performance of either aphasic or normal subjects on TT-type tasks. This finding suggests that DeRenzi and Vignolo's (1962) and DeRenzi and Faglioni's (1975) concern about the use of special prosodic emphasis in presenting TT stimuli is unfounded and perhaps that aphasic patients do not profit, in terms of improved auditory comprehension, from manipulation of linguistic stress.

However, the negative outcome of this study may be due to its design. With the Part I-type commands, the presence of an equal number of sentences in which contrastive stress was used correctly and in a misleading manner induced in some subjects a feeling of uncertainty; they remarked that they quickly made a decision not to rely on or pay attention to either cue. It was also the opinion of some subjects that the inclusion of condition III, with its omission of final words, introduced an element of game-playing that detracted from the task. The feeling that there was something anomalous about condition III is supported by the RT data. With the exception of responses to the Part IV-type stimuli by the aphasic subjects (to be discussed), all mean RTs for condition III were significantly slower than those for any other condition, whereas the overwhelming finding was that there was no significant difference in performance in terms of RT with any of the other conditions. These reactions and intuitions suggest that a less ambitious study, using only complete sentences as stimuli and only monotone-monostress and contrastive stress as experimental conditions, might produce a different outcome.

The results may not be without interest, however. The virtual absence of any significant differences in the performance of the normal subjects in all conditions but condition III suggests that prosodic features are fragile elements in language comprehension. Normal subjects may profit from suprasegmental cues only when they are presented in a 'subliminal' fashion; once a normal subject suspects that such cues are being manipulated, especially in a misleading way, he will ignore them and operate on the basis of segmental information.

Aphasic patients may not be as distrustful of alterations of prosodic features as normal subjects appear to be . This is suggested by the greater success of the aphasic subjects, compared with the normal subjects, in completing the truncated contrastive stress sentences (condition III) when the stimuli were relatively simple (Part I-type). Where the normal subjects as a group seemed to have been constructing false hypotheses, the aphasic subjects seemed to have been profiting from the information provided by the presence of contrastive stress. Nevertheless, the aphasic subjects' rate of success was only 49%, that is, no better than chance.

It is also instructive to note that the superiority enjoyed by the aphasic subjects as a group was noted only in the pass-fail scoring. When their performance was evaluated by means of weighted scores and reaction times, the advantage disappeared. This suggests, once again, that pass-fail scoring is a crude index of performance and that other techniques, such as weighted scoring and latency of response, are more sensitive measures.

The superiority of the aphasic subjects in condition III with Part I-type stimuli also disappeared in responding to Part IV-type stimuli. Regardless of the method of evaluation, the aphasic subjects did not do as well as the normal subjects in condition III for the more complex stimuli. This, in turn, suggests that the primary source of difficulty for the aphasic subjects was the segmental rather than the prosodic information. It also supports the view of some clinicians that the part of the TT most sensitive to aphasic disorders is not Part V but Part IV. A corollary to this view is that the difficulty of this part for aphasic subjects obscures any benefit they may derive from contrastive stress.

Another source of sensitivity in Part IV is the use of two noun phrases in each command and the consequent possibility that a subject may reverse their order. It may be noted that in terms of symbolic logic the phrases 'small red circle and large blue square' and 'large blue square and small red circle' are equivalent; reversal of the elements of a conjunction does not affect its truth value. It should also be noted that in the instructions given at the outset of testing, subjects were not cautioned about responding in a particular order (for example, the order of mention). Nevertheless, no normal subject made a reversal, whereas all aphasic subjects but one reversed the order of the tokens in at least one command, and more than a third of all these subjects' responses to Part IV-type stimuli involved reversals. This finding suggests that reversal is an abnormal behavior for the TT and is an indication of aphasic behavior. One should also note that such potentially valuable diagnostic information may be lost without a detailed system of scoring.

Finally, the RT data are worthy of note themselves. A long series of studies by many different investigators (reviewed by Phillon, 1973; also see Pillon and Lhermitte, 1974) has shown that the RTs of brain-damaged subjects are significantly slower than those of normal subjects, regardless of the task and the modality in which the stimulus is presented. The RT data in this study agree with these findings: with the exception of one response (an incorrect response) by one normal subject (out of over 1000 scorable RT responses made by normal subjects in the study), the RT for which was more than three standard deviations above (that is, below) the mean of her responses for that list and that condition—with that sole exception, the fastest response made by any aphasic subject in a particular condition and with a particular type of command was always slower than the slowest response made by a normal subject for that condition and stimulus type. However, such a finding fails to distinguish aphasic patients from the general population of brain-damaged patients. It is hoped that with modifications in design to make the study more linguistically sensitive, a clear distinction in RTs between aphasic and non-aphasic brain-damaged subjects may emerge, and that the role of prosodic features of language in auditory comprehension may be understood more clearly.

Notes

[1]It should be noted that this discussion of stress, in terms of differences in pitch, is a linguistic description, not an articulatory or acoustic one.

[2]This description assumes that part of the movement from high to low pitch occurs in the second syllable. It is important to keep in mind that considerable variation exists among speakers in the phonetic realization of stress. Thus another possible description of the mechanism of stress in a bisyllabic word such as <u>circle</u>, in which the first syllable is the more prominent, is that the drop in fundamental frequency occurs almost entirely during that syllable; if so, this would be represented as $\overset{F}{\text{cir}}\overset{L}{\text{cle}}$.

[3]On the basis of this criterion, one subject who achieved a perfect score of 10 on Part IV was excluded from the study.

[4]A final rising intonation is not necessarily an aberrant pattern for imperative sentences in English; indeed, in some geographical and social dialects and subdialects this pattern is frequently employed in statements as well.

[5]The gross disparity between the number of correct vs. incorrect responses for the aphasics to Part I-type commands and for the normal subjects to both types makes statistical comparisons between them unreliable.

Acknowledgement

This study was supported in part by the Mayo Foundation and in part by Public Health Service Grant NS-05446-13 from the National Institute of Neurological and Communicative Disorders and Stroke. The continual guidance offered by Raymond J. Kent, Ph.D., of the University of Wisconsin, Madison, contributed greatly to the strengths of the study. The authors are grateful also to Mr. Irvin R. Zimmerman of the Section of Engineering, Mayo Clinic, for his valuable assistance in developing the reaction time recording technique, and to William M. O'Fallon, Ph.D., of the Section of Medical Research Statistics, Mayo Clinic, for his suggestions regarding the statistical treatment of the data. The first author's current address is Speech Pathology, Department of Neurological Sciences, Marshfield Clinic, Marshfield, Wisconsin 54449.

REFERENCES

Berry, W.R. Testing auditory comprehension in aphasia: clinical alternative to the Token Test. In R.H. Brookshire(Ed.), Clinical Aphasiology, Proceedings of the Conference, 1976, 43-63. Minneapolis: BRK Publishers (1976).

Blough, D.S., and Lipsett, L.P. The discriminative control of behavior. In J.W. Kling and L.A. Riggs (Eds.), Woodward & Schlossberg's Experimental Psychology, 3rd ed., 743-92. New York: Wiley (1971).

Boller, F., and Vignolo, L.A. Latent sensory aphasia in hemisphere-damaged patients: an experimental study with the Token Test. Brain 89:815-830 (1966).

DeRenzi, E., and Faglioni, P. L'esame dei disturbi afasici di comprensione orale mediante una versione abbreviata del test dei gettoni [Verbal comprehension in aphasic and in normal subjects with a shortened version of the Token Test]. Rivista di Patologia Nervosa e Mentale 96:252-269 (1975[1976]).

DeRenzi, E., and Vignolo, L.A. The Token Test: a sensitive test to detect receptive disturbances in aphasics. Brain 86:815-30 (1966).

Fry, D.B. Experiments in the perception of stress. Language and Speech 1:126-152 (1958).

Laming, D.R.J. Information Theory of Choice-Reaction Times. London: Academic Press (1968).

Lehiste, I. Suprasegmental features of speech. In N.J. Lass (Ed.), Contemporary Issues in Experimental Phonetics, 225-230. (Perspectives in Neurolinguistics and Psycholinguistics.) New York: Academic Press (1976).

Pettit, J.M., and Heath, J.H. A comparison among four variations of the Token est as administered to a group of adult aphasics. Paper presented at the American Speech and Hearing Association Convention, Washington, D.C. (Nov. 1976).

Pillon, B. L'apport de la methode des temps de reaction dans l'etude des performances des malades atteints de lesions cerebrales. Annee Psychologique 73:261-272 (1973).

Pillon, B., and Lhermitte, F. Designation et denomination a differents rhythmes chez patients atteints de lesions cerebrales. Neuropsychologia 12:55-63 (1974).

Saxman, J.H., and Fay, W.H. Latency of echoic verbal responses by three-year-old children. Journal of Speech and Hearing Research 13:232-238 (1970).

Welford, A.T. Motor performance. In J.E. Birren and K.W. Schaie (Eds.), Handbook of the Psychology of Aging, 450-496 New York: Van Nostrand Reinhold (1977).

Woodowrth, R.S., and Schlossberg, H. Experimental Psychology. Revised ed. New York: Henry Holt (1954).

DISCUSSION

Q. Is there a correlation between Token Test percentiles and the PICA?

A. We attempted to gather PICA data on these subjects but it was not available I would like to make that correlation but I have not as yet. I do not know of any.