The Influence of Three Types of Speech Production Upon Auditory Comprehension in Aphasia

Judith E. Treadwell
York County Counseling Services, Braintree, Massachusetts

Reg L. Warren
Braintree Hospital, Braintree, Massachusetts

Mary S. Wilson
Laureate Learning Systems, Braintree, Massachusetts

Computers are now being used to treat aphasic patients. The addition of an audible (spoken) component has enabled computer programs to talk to patients, thus increasing their potential use for treatment. For instance, Mills (1982 and 1983) has reported success in retraining auditory comprehension using artificial speech.

There are various methods of speech production available to the microcomputer user--live voice recorded, digitization, Linear-Predictive Coding or LPC, and phoneme synthesis. Each differs in its relative intelligibility and "naturalness." This study examines two forms of Linear-Predictive Coding or LPC--phoneme based LPC and what we have termed customized LPC.

These two forms of linear-predictive coding were produced by a Texas Instrument system which sampled a human speech input, and compressed, encoded, and stored the sample, which was then reproduced by an Echo II speech synthesizer. For phoneme-based LPC, the data base consisted of individual sounds concatenated into word strings. As a result the prosody, intonation and pitch of the original human speech signal were lost, resulting in a type of "robotic" speech. In contrast, the data base for customized speech consists of whole phrases, resulting in a more natural speech form. Customized speech sounds more natural, but phoneme-based speech requires considerably less memory, making it more desirable from a programming point of view. For example, a program with a 200-word vocabulary requires almost twice the memory for customized LPC as it does for phoneme-based LPC. Memory-efficient speech allows for greater flexibility in programming, improved graphics, and animation.

The purpose of this study was to compare the performance of a group of aphasic patients during a word categorization task across three conditions of speech presentation; live voice, customized, and phoneme-based speech. The dependent measures were accuracy of comprehension--that is, percent correct and speed of response in milliseconds.

METHOD

Subjects. Nine aphasic subjects, four male and five female, participated (Table 1). All used English as their first language, were premorbidly right-handed, had minimum speech discrimination of 88% bilaterally, adequate visual abilities, praxis, reading and categorization skills to participate. Comprehension scores ranged from 60-90% on the Boston Diagnostic Aphasia Examination (Goodglass and Kaplan, 1972).

Procedure. Experimental stimuli were 30 high-frequency nouns representing six categories; animals, body parts, clothing, foods, utensils, and vehicles. Each subject first listened to the question "Which one is a vehicle?" (presented in one of the three speech conditions), and then viewed three printed

Table 1. Description of subjects.

Subject	Gender/Age	Education	Lesion	Aphasia Type	MPO	BDAE (Aud. Comp.)
1	M-53	13	Parietal	Anomic	24	90%
2	F-67	12	Frontal/ Parietal	Brocas	1	90%
3	F-71	10	Parietal	Mixed	1	90%
4	F-67	12	Parietal/ Temporal	Anomic	4	60%
5	M-49	12	Internal Capsule	Anomic	15	75%
6	F-70	10	Frontal	Brocas	1.5	75%
7	M-54	16	Temporal	Anomic	18	75%
8	M-64	12	Frontal/ Parietal	Brocas	6	90%
9	F-63	12	Temporal/ Occipital	Anomic	5	60%

nouns on a monitor (a target word and two foils). The nouns were arranged vertically and were numbered "1," "2," or "3." The subject responded by moving his index finger from a fixed rest position to the corresponding number on the computer keyboard.

Prior to the experimental task, each subject was familiarized with both forms of LPC by listening to sentences identifying the category of a noun which was paired with a picture of the object and a printed sentence. During the experimental task, each subject was instructed to respond as accurately and as rapidly as possible. Appearance of the words and operation of the reaction time clock began with the end of the question. The clock stopped with the subject's response. Accuracy of response and reaction time (in milliseconds) were computed and later printed. Order of the speech conditions was counter-balanced across subjects. The sequence and position of each subordinate category (represented by the target nouns) was randomly assigned and distributed across the three key positions. Sound level for the two LPC conditions was 66-68 dB. Each session lasted approximately one hour.

RESULTS

Of the 810 responses, 705 (87%) were correct. A two-tailed t-test for related measures (Ferguson, 1966) compared accuracy across the three modes of speech presentation. These values are shown in Table 2. There were significant differences between the live voice and phoneme-based speech and between customized and phoneme-based speech. Accuracy of performance was not significantly different between live and customized conditions.

Figure 1 presents percent correct for all nine subjects across three speech conditions. Subjects responded most accurately during live voice condition, the group average being 94% with this condition having the least between-subject variability. Customized condition elicited slightly less accurate performance (91.4%), with variability between subjects increasing

Table 2. Values for two-tailed t-test for accuracy across three speech conditions.

	Live	Customized	Phoneme-based
Live	---	1.956	4.757*
Customized		---	4.860*
Phoneme-based			---

*significant < .01

slightly. Phoneme-based condition was most difficult for the subjects, with considerable variability between subjects. Individual performance for six of the nine subjects parallelled group performance across conditions. Two subjects performed equally well in live and customized conditions while one subject was most accurate in customized condition.

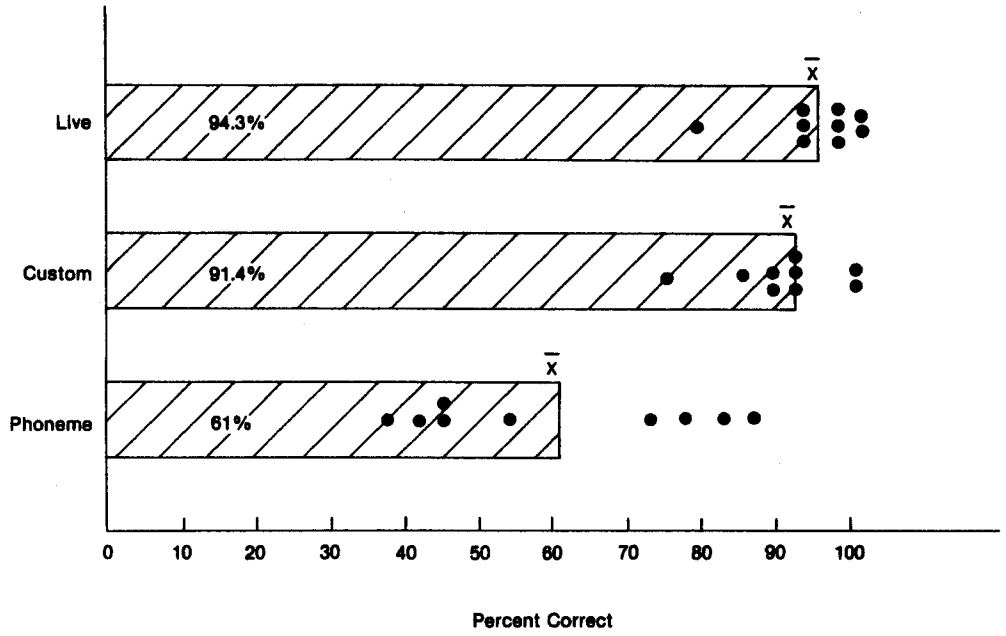

Figure 1. Percent correct across three speech conditions for nine subjects.

Response Time. For each condition, the analysis was limited to correct responses falling within 1.5 standard deviations of the mean response time per subject. This eliminated 8.7% of subjects' correct responses and removed the influence of slow but accurate or inaccurate guesses upon the response time analysis. A simple RT measure of subjects' recognition of the numbers 1, 2, and 3 on the monitor followed by the depression of the appropriate key indicated little effect of key position. Table 3 shows that speed of response was slower in phoneme-based condition than in live-voice condition. Reaction times were slower in phoneme-based than in customized condition. Reaction times did not differ significantly between live-voice condition and customized condition.

Table 3. Values for two-tailed t-test for response time across three speech conditions.

	Live	Customized	Phoneme-based
Live	---	.8568	3.689**
Customized		---	2.776*
Phoneme-based			---

** significant < .01
* significant < .05

Figure 2 presents average response times across the three speech conditions. Mean response time was fastest in live-voice condition. Mean RT for customized condition was only 134 milliseconds slower while phoneme-based condition elicited the slowest mean RT. Variability between subjects was considerable in all three conditions, with scores approximately equally dispersed about the mean. Individual performances of six subjects were similar to the group pattern of between-condition performance while three subjects responded fastest to customized speech.

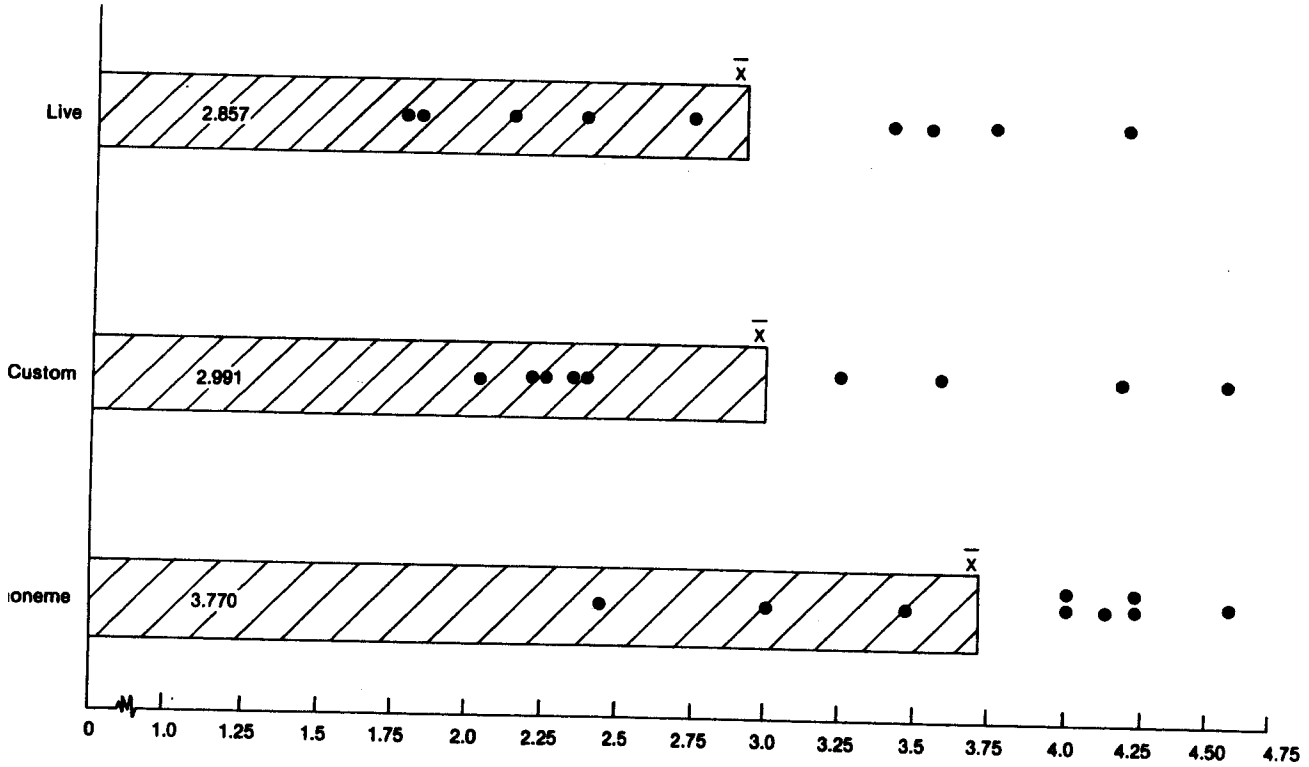

Figure 2. Response time (in sec.) across three speech conditions for nine subjects.

The use of multiple-t comparisons (three for each of the two dependent variables) increases the probability of committing a Type 1 error. Accordingly, the reader may wish to select a more conservative alpha level. Dunn (1961) suggests that the preselected alpha level (i.e., .05) be divided by the number

of comparisons per dependent variable: .05 ÷ 3 = .0166. Using the adjusted alpha level, the comparison between response times in the customized versus phoneme-based conditions (Table 3, t=2.776) is not significant.

Conclusions. Typically, software containing audible components would not utilize live speech since one objective in using microcomputer programs is to free clinicians' time. However, the most important question concerning computer generated speech is whether it is of sufficient intelligibility to be used as a substitute for "live" speech. In this study, live speech was used as a standard against which two forms of Linear Predictive coding were contrasted. The results suggest that accuracy and speed of comprehension in the customized form resembled performance under live conditions in a word categorization task. The significantly slower and less accurate performance in the phoneme-based condition suggests that this form of LPC may compound already existing comprehension deficits in an aphasic patient. Therefore, we conclude that LPC phoneme-based speech may not be appropriate for use with aphasic patients.

When used with an Apple system, peripheral devices and the software necessary to generate LPC speech cost between $120 - $200. Although the number of currently available programs using LPC speech are limited, the demand for it is expected to rise as the quality of LPC increases and the price of LPC decreases. As a matter of fact, since the speech used in this study was produced (in 1983) editing programs have been developed which improve the final speech product. In 1983 Mills compared synthesized, digitized, and cassette-tape forms of artificial speech in a picture identification task by aphasic patients. Error rates were highest for synthesized speech, less for digitized speech, and least for cassette-tape reproduction. Put another way, when artificial speech more closely resembled human speech, comprehension improved. Actually, the synthesized speech used by Mills was a form of LPC phoneme-based speech. Considering that our study found no significant difference in performance between live speech and a phrase-based form of LPC that is customized, the next step would be to compare digitized speech to LPC customized speech during the same comprehension task. Further study also needs to investigate the effects of familiarization and training with all forms of artificial speech.

REFERENCES

Dunn, O.J. Multiple comparisons among means. Journal of American Statistical Association, 56, 52-64, 1961.

Ferguson, G. Statistical Analysis in Psychology and Education. New York: McGraw-Hill, 1966.

Goodglass, H., Kaplan, E. Boston Diagnostic Aphasia Examination. Philadelphia: Lea and Febiger, 1972.

Mills, R. Microcomputerized auditory comprehension training. In R.H. Brookshire (Ed.), Clinical Aphasiology: Conference Proceedings, 1982. Minneapolis, MN: BRK Publishers, 1982.

Mills, R. The talking Apple - comparisons of three microcomputerized speech production methods. Paper presented to the American Speech-Language-Hearing Association Convention, Cincinnati, 1983.

DISCUSSION

Q: What sort of instructions did you give your subjects relative to the time/accuracy issue?

A: They were told, first, to respond as accurately as possible and then as quickly as possible.

Q: So you did emphasize accuracy. What do you think would have happened if you had taken the time element out of it?
A: I think we would still have found differences between the live-voice and phoneme-based and probably between the customized and the phoneme-based speech on the accuracy measure. We wanted to put the response time in because we thought that if there were differences between custom and live-voice it might show in response time, which is more sensitive.

Q: Were these pictures or written words that the subjects saw?
A: Pictures were used originally for familiarizing subjects. During the task itself only printed words were presented.

Q: I think in your conclusion you implied that live voice and customized were equivalent, not different? You base that I think on your nonsignificant t-values, is that right? You can't legitimately use nonsignificant differences to say that conditions are the same. Statistical tests are built to find differences and you can't confirm the null hypothesis.
A: That's a good point. We're aware of that and we thought we were careful in the conclusions to indicate only that they were not different.

Q: I may have missed this but I was a little surprised at how fast the computerized versions were presented. Did you control for live voice as far as presentation time?
A: Yes, when we gave the live-voice condition we had it set so that as with the other presentations it was timed so that when there was a "beep" we would ask the question.

Q: How long were the possible choices presented?
A: They stayed up on the screen until the patient responded. In live-voice condition it wasn't possible to perfectly time the end of the question with the onset of the stimuli and therefore in that condition the clock did start with the onset of the three response conditions, but the live voice did not always stop immediately at that point. It certainly did in the two computerized versions of speech.

Q: Did you look at your data to see if the patients responded differently after becoming accustomed to the speech conditions?
A: We didn't really look at the effect. Familiarization was not a question under study and it is an issue that needs to be looked at. Each patient received about 10 trials of each speech condition just prior to starting the experimental task in that condition. My feeling is, in the phoneme-based condition, typically the subject either understood it or didn't.

Q: I have a question about the customizing procedures. You said that was done with the package that was $120 - $200?
A: No, if you get an Echo you get phoneme-based speech. You have to buy a tool that Texas Instruments produces in order to encode the customized speech. So you can't have the phrase-based speech by buying the $200 package.

Q: Did that take a lot of time to prepare those stimuli?
A: Yes, it's very time-consuming. I didn't do it myself, but the people who did it can average maybe five minutes on a single word.

A: I think you should consider looking at some of the synthesized speech packages that are commercially available. With very little time and expense some of the speech synthesis programs sound great.
A: Commercially available synthesis is based on phoneme concatenation which is also the case with LPC-phoneme condition, and that was the condition where patients had more difficulty.