Computer-based Speech Therapeutic Intervention

H. T. Bunnell, R. Walter, J. Polikoff, J. McNicholas

Alfred I. duPont Hospital for Children, Nemours Children's Clinic, Wilmington, DE

ABSTRACT

A computer-based speech training system for young children with /r/ articulation delay was evaluated. Treatment subjects trained with /r/; controls trained with /t/. Probe words recorded before and after each of 18 training sessions over 6 weeks were scored by SLPs. Some treatment subjects, but no controls, showed improved articulation.

INTRODUCTION

The present work is intended to address the expanding demand for speech therapeutic services by providing software for articulation training. The Speech Training, Assessment, and Remediation (STAR) system would assist clinical professionals in assessing and monitoring the progress of therapeutic intervention, augment their efforts in highly repetitive articulation drill and training, and assist in record keeping and reporting. Ultimately, we expect that children will be able to use the software on a home computer, interacting with the software via an animated computer character. Since the system will be constantly eliciting speech from a child and measuring the speech produced, it will be capable of extensive record keeping and report generation, further assisting clinical staff in their duties.

We recently completed and are analyzing results from an initial evaluation of a prototype of this system. This computer-based training was included as an adjunct to the participants' traditional speech therapy sessions with a Speech Language Pathologist. Subjects' productions of a set of probe words were recorded before and after each computer training session in order to assess carryover from one training session to the next and to track improvement. In the following, we describe briefly the evaluation study and its results.

METHOD

Subjects:

- 21 children (4 to 7 years old) recruited
- Diagnosed with articulation delays related to syllable-initial /r/
- Not receiving therapy
- Due to drop-outs, final group was 18 Ss:
  - 6 females & 12 males
  - 10 in "treatment" group & 8 in the control group
  - Age ranged from 56 to 94 months (mean 77.5, S.D., 10.3)

Procedure:

- 6-week study
- 1/2 hour SLP intervention per week (all Ss)
- 3 computer training sessions per week (probe – train – probe)
- Treatment Ss trained on /r/ - /w/ contrasts
- Control Ss trained on /t/ - /k/ contrasts

Apparatus:

The speech training system was an extension of the system described by Bunnell, Yarrington, and Polikoff (2000).[1] It comprised a game-like computer interface that presented a cartoon character with which children interacted verbally. Figures 1 through 3 illustrate the three activity screens associated with the game which is set in a spaceship.

- A 'bridge' scene (Figure 1) is used as the introductory screen.
- A briefing room (Figure 2) is the setting for descriptions of speech segments, the aliens associated with the speech segments, and movies of a Speech-Language Pathologist describing how each segment should be produced.
- The communications room (Figure 3) is the scene where most of the actual speech training activity occurs.

[1]Bunnell, H.T., Yarrington, D., and Polikoff, J. (2000). STAR: Articulation Training for Young Children. Proceedings of the Sixth International Conference on Spoken Language Processing. Beijing, China. 4: 85-88.

Figure 1. Opening Screen of Speech training game.

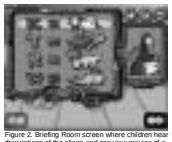

Figure 2. Briefing Room screen where children hear descriptions of the aliens and can view movies of a speech language pathologist describing the articulation of each sound.

Figure 3. Communications room in speech training game.

The speech-training component of the game guided children through a sequence of increasingly more difficult levels, based on the words associated with each level. Progression through the levels used up-down staircase logic (Wetherill and Levitt, 1965).[2] Rewards for performance were given frequently in the form of on-screen entertainment. Certificates of achievement were presented as further rewards when a child completed each of several stages of the game.

On each training trial, the child was asked to say a word—ostensibly to help cartoon characters learn to pronounce it—which was recorded by the computer and passed to a speech recognition engine for evaluation. The recognition engine evaluated the received utterance against a small active lexicon and returned a measure of confidence that the utterance was the word that was requested. If the returned confidence measure was sufficiently high the utterance was considered correct and reinforcing feedback was given verbally and in terms of a gauge that rose in level toward full scale. If the utterance was judged incorrect, feedback such as "you'll do better next time" was given instead and the visual feedback from the displayed gauge dropped.

Before and after each training session, another program was run to probe the child's progress using a set of 36 words that sampled a variety of segments of interest in a variety of syllabic and phonetic contexts. These recordings comprise the dataset used to assess progress in speech training.

[2]Wetherill, G.B. and Levitt, H. (1965). "Sequential estimation of points on a psychometric function." British Journal of Mathematical and Statistical Psychology 18: 1-10.

Data Analysis:

- 18 probe words containing /r/ and/or /er/ in mixed contexts
- 2070 words in all sampled at intake (session 0) and training sessions 3, 6, 9, 12, 15, and 18
- 17 practicing SLPs each classified ~485 words as correct/incorrect
- One SLP classified half the total set as correct/incorrect
- Each word token rated by 4 or 5 SLPs

RESULTS

Data for the 2070 probe words were expressed in terms of the proportion of "correct" responses each word received from the 4 or 5 raters who heard it.

There were differences among the probe words such that some words were, overall, less frequently rated "correct" by the judges. Figure 4 shows the proportion of "correct" judgments recorded for each of the 18 probe words averaging over subjects (the 2 /r/ segments in *graveyard* and *rooster* are each considered separately). Words with syllable-final /r/ were generally more difficult for these children than were words with syllable-initial /r/. The /r/ in the /gr/ cluster of graveyard was classified correct more frequently than any other /r/ segment.

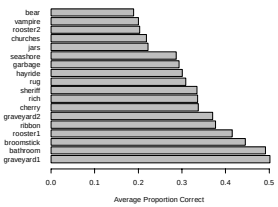

Figure 4. Average proportion of "correct" classifications by judges for each of the 18 probe words analyzed.

Of greater interest is whether subjects in the treatment group showed more improvement than subjects in the control group as measured by their production of the probe words. In fact, three subjects in the treatment group reached a criterion level of greater than 50% correct /r/ production in all contexts for the probe words. The remaining treatment subjects and all control subjects maintained approximately stable performance throughout the study. These data are shown in Figure 5.

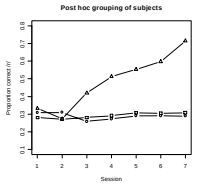

Figure 5. Three post hoc groupings of subjects. Average proportion correct /r/ responses per session for N=3 subjects in the experimental treatment group who acquired /r/ during the study (triangles), contrasted with averaged data from N=7 subjects in the treatment group who did not acquire /r/ (squares), and averaged data from N=8 subjects in the control group (circles).

In addition to monitoring the progress of children as measured by their productions of probe stimuli, we have examined several measures of performance on the training tokens as well. One measure that we have found useful is illustrated in Figure 6, which shows—for each child—the maximum level the child was able to sustain at least 60% correct productions (as measured by the ASR system). Subjects 3, 6, and 16 were the three subjects whose data were plotted separately in Figure 4. This figure suggests that a fourth subject (s02) may also have been having some success despite not showing improvement with the probe stimuli.

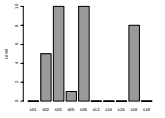

Figure 6. Highest level at which each child in the experimental group was able to maintain a performance level of 60% correct productions.

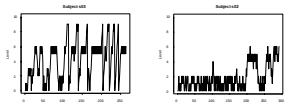

Figure 7. Performance level of two subjects throughout the 6-weeks of the experiment. The panels chart the level transitions for each child.

To further examine this, Figure 7 shows series-by-series performance for s02 (and for comparison s03) throughout the experiment. The plot for s03 is characteristic of a child who is succeeding with this task by running the game up to the highest level frequently, especially in later sessions. This figure also suggests that s02 was beginning to reach higher levels of difficulty in sessions late in the session.

DISCUSSION

Although this study lasted only six weeks and involved only one half-hour traditional therapy session per week, three subjects in the treatment group made substantial improvement in /r/ production, and a fourth subject may have been starting to acquire /r/ toward the end of the protocol.

In addition to showing potential for efficacy as a speech training aid, we would emphasize the potential value of the data obtained by computer aided speech training systems. In particular, it is noteworthy that:

- Detailed records of activity and progress are available to help a supervising SLP (e.g., Figure 7).
- Broad classification of performance in terms of the maximum level at which a certain level of performance can be maintained provides useful summaries of performance (e.g., Figure 6).
- Large amounts of speech data are recorded during the training and provide specific examples that a supervising SLP can review to inform clinical decisions.
- With appropriate consent/assent, data collected by the system can be used in large scale acoustic analysis studies to examine fine differences among children with speech delays and to search for acoustic markers associated with readiness to acquire specific segments.

This study was supported by Nemours Biomedical Research.