Applications and experiences of computer-based speech training

Öster, A-M.

journal: STL-QPSR
volume: 30
number: 4
year: 1989
pages: 037-044

http://www.speech.kth.se/qpsr
APPLICATIONS AND EXPERIENCES OF COMPUTER-BASED SPEECH TRAINING*

Anne-Marie Öster

Abstract

A microcomputer-based aid for speech training has been introduced on a trial basis in the Swedish schools for the deaf. The prototype has been used for some years to evaluate its applicability in the speech clinic of a bilingual school for prelingually deaf children. Results are reported from two training experiments and some general comments are given on this new type of visual technical aids.

INTRODUCTION

A new teaching situation has evolved in Swedish schools for the deaf. Sign-language is now the first language and Swedish is the second language and will mainly be learnt in its written form and by the aid of sign-language. Speech is only practised in the speech-clinic, about 20-40 minutes a week. To develop an intelligible speech on these conditions, efficient and individualized speech training methods have to be created. These must be based on written Swedish and the use of sign-language for instruction and explanation.

Imitation and self correction are important factors in speech learning. When a child does not have any residual hearing, other senses must replace the auditory feed-back that hearing children use when they learn to speak. The deaf child can establish oro-sensory-motor control of his speech movements through visual or tactile feed-back. Technical aids used in speech training can assist the teacher and help the child to follow three very important steps in speech learning: instruction, training, and generalization.

In speech training it is important to make the child aware of important dimensions of speech and distinctive contrasts between similar speech sounds. Distinctive features like, for example, voiced/voiceless are invisible to the child through speech-reading and consequently difficult to produce correctly. Many training aids have been developed where speech elements are shown as visual signals (Risberg, 1968). Computer-based speech training has capabilities to offer the children immediate and meaningful visual feed-back of distinctive contrasts and might make it easier for the therapist to instruct and explain what is wrong and what is correct and through motivational and frequent training establish a correct pronunciation.

The aim of this study was to get general experience from computer-based speech training and evaluate its efficiency. Two experiments were made with eight weeks of systematic training of two children with difficulties to produce some phonological contrasts in Swedish.

MATERIAL AND METHODS

Subjects

Two children were trained systematically with a microcomputer-based aid. One child (child 1) had difficulties in producing contrasts between phonologically long and short vowels, which is an important contrast in Swedish. It is realized as a difference in duration and for some pairs also with a difference in vowel quality. Furthermore, the following consonant has opposite

*This is a revised version of a paper presented at the European Conference on Speech Communication and Technology, Paris, September 1989.
quantity, i.e., long vowels are followed by short consonants and vice versa. The other child (child II) had difficulties to produce a distinction between voiced and voiceless velar stops.

Child I was 15 years old and attended the 8th grade of a school for the deaf in Stockholm. He had some residual hearing in the low frequencies. Child II was 13 years old and attended the 7th grade of the same school. Her hearing threshold was within the region of vibration. Both children were prelingually deaf.

**Training procedure**
The microcomputer-based aid used in this study was the IBM "SpeechViewer" (Öster, 1988). The system is made up of 12 modules for three areas of use: displays for basic awareness of selected dimensions of speech, such as pitch, loudness, timing and the presence or absence of voicing, skill-building voice-controlled games and exercises for pitch, timing and vowel production and speech patterning with graphical presentation of the speech signal for analysis and training of refined articulation and prosody.

In this study the module "Pitch and Loudness" of the speech patterning displays was used for the training. It is a graphical presentation of the speech signal where the voiced/voiceless contrast is clearly indicated by different colours. Voiced sounds are red and voiceless sounds are green. Discrimination between long and short vowels is visible through the differences in duration of the red colour that indicates voiced vowels. It was easy to explain to the children what was deviant in their production by comparing their speech with the speech of the therapist on a split screen display.

The children were trained about ten minutes each, two times a week during eight weeks. The therapist and the child worked systematically with the word-pairs of material A, which is described below in "Training materials".

**Training materials**
The children's speech was video recorded before and after the computer-training. The speech material of each child was identical at both recordings and was made up of minimal word pairs. The phonological contrast that differentiated the word pairs in the material of child I was vowel-length. The speech material of child II was made up of word pairs which contained the distinction between voiced and voiceless velar stops. The recorded material was composed of two parts (A+B). Material A contained the word pairs which were used in the training sessions. Material B was made up of word pairs which contained the same contrasts but was not used in the training sessions. The intention was to investigate the generalization effects on untrained words by assessing the production of phonological contrasts in material B before and after training. At the recording each word was presented in written form on separate cards. Short sentences in which the target words were included were also recorded to study the pronunciation of the words in isolation compared to the pronunciation of the words in running speech.

**ANALYSES**
Narrow phonetic transcriptions were made of the video recordings using the symbols of the International Phonetic Alphabet (IPA). For this application we felt a need for additional dia-critical marks. Some of those which Bush, Edwards, Luckau, Stoel, Macken, & Petersen (1973) and Roug, Landberg, & Lundberg (1987) have developed for the transcription of babbling and phonetic development in early infancy were used. Spectrograms were also made to objectively confirm the perceptual results. A Kay Elemetrics Sonagraph, model 5500 was used.
RESULTS

Training of long/short vowels (child I)

Fig. 1 shows the vowel transcriptions made of the video recorded speech of child I before and after training of long and short vowels. The symbol /:/ is used to indicate that the preceding phoneme is long. The transcriptions are made on both isolated words and words in sentences. The figure shows all the variants in quality and quantity that the child produced at the two recordings. The target vowels are given by their conventional orthographic representation. For å and ö, there exist two open allophones in Swedish, which are used before r.

<table>
<thead>
<tr>
<th>ORTHOGRAPHY</th>
<th>IPA</th>
<th>BEFORE TRAINING</th>
<th>AFTER TRAINING</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>i</td>
<td>i:</td>
<td>i: i</td>
<td>i: i</td>
<td>8</td>
</tr>
<tr>
<td>I</td>
<td>ι:</td>
<td>ι: ι</td>
<td>ι: ι</td>
<td>8</td>
</tr>
<tr>
<td>y</td>
<td>y:</td>
<td>ο: ο</td>
<td>ο: ο</td>
<td>2</td>
</tr>
<tr>
<td>Y</td>
<td>ο: ο</td>
<td>ο: ο</td>
<td>ο: ο</td>
<td>2</td>
</tr>
<tr>
<td>u</td>
<td>u:</td>
<td>ο: ο</td>
<td>ο: ο</td>
<td>2</td>
</tr>
<tr>
<td>θ</td>
<td>ο:</td>
<td>ο: ο</td>
<td>ο: ο</td>
<td>2</td>
</tr>
<tr>
<td>e</td>
<td>e:</td>
<td>ο: ο</td>
<td>ο: ο</td>
<td>4</td>
</tr>
<tr>
<td>Ε</td>
<td>ο:</td>
<td>ο: ο</td>
<td>ο: ο</td>
<td>4</td>
</tr>
<tr>
<td>ö</td>
<td>φ:</td>
<td>φ: ο</td>
<td>φ: ο</td>
<td>2</td>
</tr>
<tr>
<td>ö</td>
<td>φ:</td>
<td>φ: ο</td>
<td>φ: ο</td>
<td>2</td>
</tr>
<tr>
<td>οε</td>
<td>οέ</td>
<td>οέ οέ</td>
<td>οέ οέ</td>
<td>2</td>
</tr>
<tr>
<td>οέ</td>
<td>οέ</td>
<td>οέ οέ</td>
<td>οέ οέ</td>
<td>2</td>
</tr>
<tr>
<td>0</td>
<td>ο:</td>
<td>ο: ο</td>
<td>ο: ο</td>
<td>2</td>
</tr>
<tr>
<td>ο</td>
<td>ο:</td>
<td>ο: ο</td>
<td>ο: ο</td>
<td>5</td>
</tr>
<tr>
<td>å</td>
<td>ο:</td>
<td>ο: ο</td>
<td>ο: ο</td>
<td>5</td>
</tr>
</tbody>
</table>

Fig. 1.  Phonetic analysis of long and short vowels before and after training. n = number of observations in each reading (child I).

It is obvious that the child did not control the vowel quality in some of the cases, e.g., ο, y, and η, but it was not the intention to train the vowel quality in this study. However, in the case of short η, the pronunciation improved after the durational training and in some other cases, the pronunciation became more stable, as for å and ö.
Fig. 2 shows the assessment of the child's productions of the correct quantity of Swedish long and short vowels before and after training. The result shows, although a limited amount of data, that the child controlled the contrast between one vowel, o, before training. All his productions of the vowels d and u were long before training. After training, he learnt to produce the short versions correctly. Short vowels improved more than long vowels.

Fig. 2. Subjective evaluation of correct quantity of Swedish long and short vowels before and after training (child I). The vowels are represented in orthographic symbols.

An improvement in producing durational contrasts between vowels with the help of computer training was found in all vowels except for o which was correct from the beginning, and y, where this contrast was not established. The child produced the same differences in duration between long and short Swedish vowels when he read the words in isolation as when he read the words in running speech. Furthermore, the improvements found in material B, after training were also found in material B, which indicates a carry-over effect.

Training of voiced/voiceless velar stops (child II)
In Fig. 3, the subjective evaluation of the child's productions of the voicing contrast of velar stops in medial and final positions is shown. Before training, the child produced a devoiced, aspirated velar stop for both k and g in medial position. Sometimes k was realized as a stop with unspecified friction. After training, he separated the phonemes by producing a voicing contrast. Also in final positions, before and after training, k was realized as a devoiced stop.
with aspiration or in some cases with unspecified friction, while \( g \) always was omitted. After training, it is obvious that she learnt the voicing contrast between velar stops in final as well as in medial positions. The IPA-symbol [\( g \)] was preferred to represent the child's pronunciation of \( k \) instead of the symbol [\( k \)]. The reason for this was owing to the fact that the expiration pressure was perceived as too low to be a correct production of \( k \).

<table>
<thead>
<tr>
<th></th>
<th>BEFORE TRAINING</th>
<th>AFTER TRAINING</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Medial</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>/( k )/</td>
<td>[( g^h )]</td>
<td>[( g^h )]</td>
</tr>
<tr>
<td>/( g )/</td>
<td>[( g^h )]</td>
<td>[( g )]</td>
</tr>
<tr>
<td><strong>Final</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>/( k )/</td>
<td>[( g^h )]</td>
<td>[( g^h )]</td>
</tr>
<tr>
<td>/( g )/</td>
<td>[-]</td>
<td>[( g^a )]</td>
</tr>
</tbody>
</table>

Fig. 3. Phonetic analysis of velar stops in medial and final position before and after training (child II).

In Figs. 4 and 5, spectrograms of the child's utterances before and after training of the word pairs *make-mage* (husband-stomach), representing medial positions, and *tack-tagg* (thanks-thorn), representing final positions, are shown. The spectrograms confirm the perceptual evaluation that the child improved the contrast of voicing with the help of the computer training. Before training, \( k \) and \( g \) are both produced without voicing in medial position. In final position, \( g \) is omitted. After the training, \( g \) is pronounced as a voiced velar stop in both medial and final positions. This child also produced the differences in voicing between velar stops when she read the words in isolation as well as when she read the words in running speech. As for child I, the improvements found in material A after training were also found in material B, which indicates a carry-over effect.

**DISCUSSION**

The work with the SpeechViewer shows that computer-aided speech training may well be used as a valuable expansion of traditional speech training of prelingually deaf children, as it offers the children an immediate, meaningful, and motivational feed-back of distinctive contrasts. By means of the speech patterning module, it is easy to show and instruct the child about the different distinctive features, which underlie phonological contrasts in language. Training with minimal word pairs, which only differ with respect to one opposition, has shown to be an appropriate and efficient training strategy.

A very important observation is the difference in feed-back provided with computer-based speech training compared to traditional speech training. There are different types of feed-back to benefit from in order to replace the auditory feed-back in speech training. In traditional speech training, the feed-back is given to the child by the therapist after the production and hence, the feed-back process is subjective, delayed, and verbal. In computer-based speech training the child gets an objective, immediate, and nonverbal feed-back. This type of feed-
back is more efficient and makes the information clear and meaningful to the child and hence also makes the training motivating.

![Spectrogram](image)

Fig. 4. Spectrograms of the word-pair magel/make before and after training (child II).
Another very important advantage with computer-based speech training is that it gives an objective evaluation of the child's speech. The speech therapist has a difficult role to encourage and motivate the child at the same time as she/he must criticize and evaluate the child's attempts. The "SpeechViewer" and other visual speech training aids give the therapist and the child better possibilities to cooperate.

**BEFORE TRAINING**

![Spectrograms of the word-pair tack/tagg before training](image)

**AFTER TRAINING**

![Spectrograms of the word-pair tack/tagg after training](image)

*Fig. 5. Spectrograms of the word-pair tack/tagg before and after training (child II).*
The system also gives the speech therapist increased flexibility in training since the task levels and performance criteria offer many choices. Much of traditional training is combined in one piece of equipment, built to display only one or a few speech parameters or features which make the selection of the most suitable training easy. Furthermore, efforts and improvements in training can easily be registered and documented since the information on the screen can be printed out.

There are, however, some possible drawbacks with computer-based speech training that are worth consideration. The costs of the systems are high and in some respects, a new teaching technique must be developed. There is always a risk that the training will be adjusted to the technical aid and not to the child's needs.

CONCLUSION
Improvements in producing durational contrasts between long and short vowels and voicing contrasts between velar stops in medial and final word positions were found after some training with a computer-based aid. One explanation of the positive training result is the very instructive and pedagogical effect that such systems can provide in the form of a direct, meaningful, motivational, and visible feedback. If a technical aid gets a child to understand what is wrong and what is correct in his production, it will be a powerful tool in the speech clinic. Computer-based speech training has shown to be useful and it is very likely that speech training methods in a bilingual school for the deaf ought to be based on this technique. It is a good complement in the speech training situation but must be used with restriction and in moderation.

ACKNOWLEDGMENTS
This study is a result of a valuable collaboration with the speech therapists Astri Wennerberg and Görel Delin of the Manilla School, Stockholm, which are responsible for the training of the children. The work has been financed by the Swedish Board for Technical Development and the Swedish Board of Education.

References