The Use of Visual and Tactile Sensory Aids in Speech Production Training: A Preliminary Report

INTRODUCTION

Early work on the development of sensory aids focused primarily on instrumentation problems, e.g., the reliable detection of one or more acoustic parameters from the speech signal and their presentation in an effective tactile or visual format. During the past decade, as technology has become more sophisticated, emphasis has shifted from questions of instrumentation to concern with the evaluation of specific devices. Levitt, Pickett, and Houde, 1980 (1) provides a comprehensive review in this area.

The reception of acoustic cues for fundamental frequency (F0) and the monitoring of F0 (also called voice pitch) in speech production is a problem for hearing-impaired persons that has received considerable attention in sensory aid design and evaluation. Work on F0 forms part of every training program for the hearing impaired (2,3,4,5,6). Although considerable time and effort is often expended, development of good phonatory skills (e.g., speech produced with fundamental frequency appropriate to age and sex, and with intonation contours that convey a declarative or an interrogative sentence, etc.) can be a particularly frustrating task for both teacher and student. The difficulty may be due, in part, to the invisible positioning of the larynx that makes use of visual cues impossible; the hearing-impaired person may also find it difficult to perceive the auditory cues for control and variation in fundamental frequency. While such factors contribute to the difficulty in remediating phonatory disorders, some of these obstacles to successful perception and production can be surmounted by the use of sensory aids that provide a display of fundamental frequency.

Research such as that described by Levitt et al. (1) has examined sensory aids as speech reception devices (often as an adjunct to lipreading) but the potential of such devices in speech production training remains to be explored fully. Indeed, while speech reception aids (e.g., the hearing aid) often have usefulness in speech production training, the converse is not necessarily true. Furthermore, in evaluations of the effectiveness of speech training aids, results are often based on single-subject experiments. Subjects receiving speech training with either a visual or tactile sensory aid frequently demonstrate improvement in speech skills when compared to a control group. But in some instances, neither the subjects nor the control group may have received intensive speech training prior to the experiment. The result is that groups may have shown progress merely as a consequence of receiving training, and the

---

*Nancy S. McGarr, Ph.D.
Janet Head, M.A.
Meryl Friedman, M.S.*
Anne Marie Behrman, B.A.*
Karen Youdelman, Ed. D.*

*Center for Research in Speech and Hearing Sciences
City University of New York
33 West 42nd Street
New York, N.Y. 10036

†Lexington School for the Deaf
Jackson Heights, New York

---

*aThis study was supported by PHS Grant No. PO1-NS17764 from the National Institute of Neurological and Communication Disorders and Stroke (NINCDS).
effectiveness of a sensory aid per se could not be established reliably. Again, see Levitt et al. (1) for a discussion of these experimental issues; note in particular chapter VII. Unfortunately, there have been few studies evaluating sensory aids in a comprehensive and ongoing speech training program.

In this paper, we report preliminary data from the first year of a long-term study designed to examine the use of visual or tactile sensory aids in speech production training. The population studied is a group of hearing-impaired children at the Lexington School for the Deaf who have received speech training as part of the school's curriculum for a minimum of 2 years prior to participating in this project. Two questions are addressed:

1. Do hearing-impaired children who receive speech training that incorporates sensory aids as part of the curriculum show greater improvement in speech skills than a control group who receive training without sensory aids?

2. Is a tactile sensory aid, or a visual sensory aid, the more effective in remediating certain specific phonatory problems? The phonatory problems considered were (i) average voice pitch too high, and (ii) inappropriate pitch contours (e.g., monotonous or, alternatively, excessively variable). These specific problems were selected because they are among the most common forms of phonatory dysfunction in hearing-impaired speakers: see Osberger and McGarr, 1982 [7].

METHODS AND PROCEDURES

Subjects

The test group consists of 20 hearing-impaired children (mean pure tone average 92 dB ISO) who were enrolled in the Lexington School for the Deaf and who were receiving speech training as part of the regular school curriculum prior to participating in this experimental program. All of the children were aided binaurally. None had other known handicapping conditions. The subjects ranged in age from 8 to 12 years old (n = 6) and 15 to 19 years old (n = 14). Subjects in the intervening ages were not included, in order to avoid phonatory instability that frequently accompanies puberty. The subjects, whose speech production was characterized as either “inappropriate in fundamental frequency” (generally, this meant too high) or “inappropriate in modulation of the pitch contour” (perceptually flat and monotonous), were selected from the speech department records in consultation with the supervisor of the Lexington School speech program. The children’s speech production skills were subsequently assessed independently (as described below) in order to corroborate these records.

The test children were divided into three groups for experimental purposes.

The Auditory Group (n = 6) consisted of children of 8 to 12 years who received the experimental speech program using only auditory input, usually the child’s own hearing aid or an auditory trainer. No visual or tactile sensory aid was used with this group, in keeping with the school’s policy of developing maximum use of residual hearing.

The Visual Group (n = 5) comprising older (15 years) students, received the experimental speech program using auditory input, as with the Auditory Group, and also a visual sensory aid, the Visipitch (Kay-Elemetrics Corp.). Figure 1 shows that the Visipitch was interfaced with an Apple II computer to provide a visual display of the fundamental frequency contour produced by the child. The contour could be stored by the teacher and later retrieved for comparative analysis with other productions. Perceptual judgements of the child’s attempt at each trial could also be stored, and learning curves were generated at the end of the program to assess each child’s progress. This is described by Friedman, in press (8). The computer also served as a data collection device for the other two groups so that similar analyses could be made.

The Tactile Group (n = 9), also older students, received the experimental speech training using auditory input and also a multichannel tactile sensory aid developed for the project; this aid was described by Boothroyd, in press (9). Figure 2 shows the tactile device worn by one of the students. Briefly, this sensory aid extracts fundamental frequency from either a laryngeal accelerometer or a microphone and presents F0 to the child tactually along a series of eight solenoids. As shown in this photograph, high frequency would be displayed near the top of the sensory aid, i.e., close to the wrist; lower frequencies are coded near the bottom of the array. The device was worn by the children only during the speech tutoring sessions for the purposes of this experiment.

Test subjects were placed on either of two curricula depending on their phonatory problem. The philosophy for either curriculum is quite similar; the children progress through steps ranging from vowels in isolation, to consonant and vowel combinations, to multisyllables varying in durational patterns, to words, phrases, sentences, and spontaneous speech. Speech training is accomplished using only those phonemes that are in a child’s
phonetic repertoire as determined by the Phonetic Level Speech Evaluation, of Ling, 1976 (6), administered at the beginning of the school year. Teaching consists of three tasks: (i) auditory training (without use of any sensory aid except the hearing aid), (ii) imitation of the teacher's productions (with a sensory aid) and (iii) production of the stimulus on demand (again, with the sensory aid). While each curriculum is hierarchical in progression, from simple to complex activities and also in level of difficulty, the child and teacher can be working on several steps simultaneously. For example, a child may need additional practice on production of voiceless consonants but at the same time be working on a more advanced step such as multisyllabic strings of voiced consonants and low vowels.

Twenty children, matched for age, sex, and hearing loss as well as phonatory problem, formed the control group. These children received their speech training as part of the Lexington School curriculum, which would include working on segmental and suprasegmental production as well as focusing on each child's particular phonatory problem. While the curriculum for the experimental groups was slightly different from that of the controls, both training regimens incorporated many common teaching principles and techniques. Children in the test group substituted the experimental speech program for the school speech tutoring program. Approximately 9 months of speech training, corresponding to the school calendar year, were given to each group.
Evaluation Measures

A test battery of speech production measures was administered to both the test and control groups prior to speech training and again at the end of the training period (hereafter, pre- and post-training respectively). In all cases, evaluations were conducted without the use of a sensory aid except the hearing aid.

The data reported in this paper are taken from results of the Fundamental Speech Skills Test, described by Levitt, Rotunno, Sullivan, and Head in 1979 (10). Briefly, this test assesses the production of suprasegmentals, including listeners' ratings of the overall appropriateness of fundamental frequency during the production of spontaneous speech samples, global ratings of the appropriateness of the intonation contours in spontaneous speech samples, rating of the presence of a terminal fall in the production of phrases and simple declarative sentences, as well other test items assessing such abilities as breath-stream management for vowel and consonant production, phoneme-related pitch breaks, etc.

The children's productions were tape-recorded. Two teachers who were experienced in evaluating the speech of the deaf (and who knew the children) rated the tapes at the end of the year. The tapes were randomized so that, while the rater knew the name of the child and the test material, she did not know whether the sample was pre- or post-training.

RESULTS

The results are divided into two parts, differences between test and control groups, and differences between the Auditory, Visual, and Tactile groups. In comparing test and control groups, the data were averaged across the Auditory, Visual, and Tactile groups. This was done because of the small number of children in each subgroup. In evaluating the effect of the various types of sensory aid, the data were averaged over the two phonatory-problem groups. It should be noted that children with average-pitch problems typically also had inappropriate intonation contours. For this reason, children with average-pitch problems were evaluated with respect to both average pitch and intonation contour. The children within the inappropriate intonation group were evaluated with respect to overall intonation contour and, in addition, with respect to terminal fall, which is an important specific aspect of the intonation contour.

Evaluations took the form of ratings by two teachers who were experienced in evaluating the speech of the deaf. An analysis of variance showed no significant differences between the two raters in their evaluations.

Test versus Control Groups

Average pitch problems—Figure 3 shows the percent of judgements, pre- or post-training, for test and control subjects (n = 11 in each group) who were characterized as having pitch that was judged to be inappropriately high for the child's age and sex. The plots are collapsed across the age of the subjects and, for the test group, also across the sensory aid.

The raters judged speech productions that had been obtained from a description of a picture stimulus, and judged also a sample of conversation—the picture description and conversation sample are two subtests of the FSST (11 children x 2 listeners x 2 stimuli = 44 judgements per group).

Analysis of variance on the data averaged over stimuli showed no statistically significant difference between raters and no rater interactions with the other main effects—subjects and time of evaluation. The test subjects and the control subjects received very similar ratings pre-training: 45 percent of the total judgements were judged to be in an acceptable pitch range. Test children showed a significant increase in the percent of acceptable judgements post-training (an increase of approximately 40 percent). This increase was due primarily to a decrease in the percent of judgements rated as "too high."

The control group also showed progress, although the magnitude of their improvement was not as large: the percent of acceptable judgements of overall pitch for the control group increased by only 10 percent. As noted above, the speech training sessions for the test children concentrated heavily on the remediation of inappropriately high pitch, while the children in the control group worked not only on this problem but also on other aspects of speech production.

Figure 4 shows judgements obtained for the same group of children on an additional characteristic of their speech, one that was not taught specifically in their speech sessions—the "overall appropriateness of the intonation contour." Listeners were asked to judge if the intonation contours in the abovementioned speech samples were "adequate," "monotonous," or "excessively variable." An overall judgement was encouraged from the listeners rather than attention to specific phrases or sentences.

Test children with "average pitch too high" who
**INAPPROPRIATE AVERAGE PITCH:**

**PITCH RATINGS**

**FIGURE 3**
Percent of judgements of the appropriateness of average pitch on two speech samples (see text). Subjects were characterized as having "inappropriately high fundamental frequency." Listeners rated the samples: too low (LOW), acceptable (ACCP.), or too high (HIGH) in F0. Data are pre- and post-training.

**INAPPROPRIATE AVERAGE PITCH:**

**INTONATION CONTOUR RATINGS**

**FIGURE 4**
Percent of judgements on the appropriateness of the intonation contour. Subjects are as in Figure 3. Listeners rated the intonation contours of the speech samples: monotonous (MONO.), adequate (ADEQ.), or variable (VAR.). Data are pre- and post-training.

**INAPPROPRIATE INTONATION CONTOURS:**

**CONTOUR RATINGS**

**FIGURE 5**
Percent of judgements on the appropriateness of the intonation contour. Ratings are as in Figure 4. Subjects were those whose intonation contours were characterized as "inappropriate."
in addition were rated as “excessively variable in the control of intonation contours,” showed a striking decrease in variability as fundamental frequency became more appropriate. Nearly 30 percent of the judgements were rated as “adequate” after training and about 15 percent were rated as “monotonous.” The control group showed a pattern similar to the test group although, again, the magnitude of the improvement was not as great. While inappropriately high fundamental frequency may be among the most salient perceptual characteristics for listeners to deaf speech, there are obviously also other phonatory problems that may co-exist. These less salient problems may improve as the primary problem is corrected. We also noted that, as variability in production decreased, some productions became more monotonous as the children acquired adequate phonatory control.

**Inappropriate pitch contours**—Nine children in the test groups and nine in the control groups were categorized as having pitch contours that were “flat and monotonous”. Figure 5 shows the percent of judgements in which listeners rated production as “adequate,” “monotonous,” or “excessively variable” for the two speech samples described above (9 children x 2 listeners x 2 stimuli = 36 judgements per group). Pre-training, none of the test children were rated as having “adequate” intonation contours; more of their productions were judged “excessively variable” (54 percent) than “monotonous” (46 percent). The control group likewise evidenced more variability (71 percent) than monotony (29 percent) in intonation-contour production. (Anecdotally, it was noted that many of these children also had phoneme-related pitch problems which might have influenced the judgements of excessive variability.) Both groups showed an improvement in percent of productions judged to be “adequate” (an increase of 28 percent for the test group and 7 percent for the control group). These judgements reflect a general impression of the child’s ability to produce a variety of intonation contours in the speech samples, including a terminal fall for a declarative sentence, a terminal rise for interrogatives, and peak stress for emphasis. These are complex clues, and we are currently investigating the perception of intonation in speech produced by this group: see Rubin-Spitz and McGarr, 1985 (11). Furthermore, learning curves obtained for the test children showed that acquiring phonatory control for producing a terminal fall in a declarative sentence was an exceedingly difficult task for many subjects.

This result is also illustrated in Figure 6, which shows judgements of terminal fall in 15 phrase and sentence items on the FSST (9 children x 2 listeners x 15 items = 270 judgements per group). Listeners were asked to judge whether the F0 contour at the end of the phrase or sentence fell, was flat, or increased. As before, there was no statistically significant difference between raters. For the test group, the judgements were nearly 20 percent higher for post-training testing, although there was still considerable room for improvement. It is interesting that a number of children (57 percent of the test children, 35 percent of the controls) were initially judged to produce a terminal fall in a controlled stimulus environment (a phrase or simple sentence) but that the overall impression of their speech (as

![Figure 6](image-url)
produced in less structured tasks such as the picture description or conversation sample) did not reflect that terminal-fall skill.

The Effect of the Sensory Aid:

In order to assess the effectiveness of the sensory aid in remediation of phonatory problems, an analysis was done of the 3 test groups: Auditory, Visual, and Tactile. As noted earlier, data were collapsed across phonatory problems because of the small number of subjects in this pilot study. Figure 7 shows the change in percent of judgements of average pitch, pre- to post-training. Ideally, one would expect that as a result of training, the percent of "acceptable" judgements of average pitch would increase, and that the percent of judgements awarded for pitch too low or too high would decrease. That pattern was noted for the Auditory group.

The Visual group also showed an increase in the percent of "acceptable" judgements as well as an increase in the percent of judgements of "too low." It may be that in attempting to control for an inappropriately high fundamental frequency, some children overcompensate and explore the low end of their pitch range. However, this result must be interpreted with caution, because children with both inappropriately high F0 and inappropriate pitch contours were included in this analysis.

For the Tactile group, the results were perplexing. The percent of judgements for inappropriately high F0 increased from pre- to post-training, while those for "too-low" or "acceptable" productions actually decreased. Again, this may reflect the combination of phonatory problems within the Tactile group and also the possibility that the Tactile sensory aid may be less appropriate for conveying more static aspects of pitch. Again, we interpret these trends cautiously because of the small number of subjects in each group.

Figure 8 shows the change in percentage of judgements of overall intonation from pre- to post-training for the three test groups. Ratings are of the "overall acceptability" of the intonation contour. Ideally, one would expect that as the result of training the percentage of acceptable productions would increase and the percentage of monotonous and excessively variable judgements would decrease. However as noted in Figure 4, a number of children seemed to become more monotonous in production as they tried to achieve a correct contour. A second pattern, an intermediate stage in remedial training, may be what we are seeing.

The Tactile group showed the preferred pattern—an increase in the percentage of "adquate" judgements and a decrease in productions judged as monotonous or excessively variable. The multi-channel tactile sensory aid may be particularly well suited for conveying dynamic changes in F0 such as are required for overall intonation contour production.

The Auditory and Visual groups likewise showed significant change in performance from pre- to post-training, although the patterns of change showed an increase in the number of productions judged as monotonous—the secondary pattern. We emphasize again that we do not consider this to be a serious error, and anecdotally report that some children who have continued on the project now show more adequate variation of intonation contours and less monotonity.

Figure 9 shows the pattern of changes from pre- to post-training for items related to terminal fall. Ideally, one would anticipate an increase in terminal fall, and a decrease in judgements of terminal rise or relatively flat productions. Again, that pattern of ideal change is noted for the Tactile group and also for the Auditory group. The Visual group also makes some progress, but it is not as dramatic.

DISCUSSION

To summarize briefly, the test children showed greater progress than the controls in remediation of either inappropriately high fundamental frequency or inappropriate pitch contours. However, these results reflect a combination of parameters of the test children's speech training, including in some cases a visual or tactile sensory aid and also the experimental computer-based curriculum. Although that curriculum incorporated many of the features of the Lexington school curriculum for remediating phonatory problems, it may be sufficiently distinct to account for some of the differences between test and control groups. We are presently evaluating the contribution of the curriculum alone in effecting change, as well as examining more closely the trends noted for the different sensory aids. Nonetheless these preliminary results are encouraging. Children who received the experimental curriculum without sensory aid intervention (Auditory group) made significant progress when compared to their controls. While all of these children were younger, sensory aid intervention would have been provided if progress arrested. This was not necessary, although it remains to be seen if this group might have progressed faster and further in their skills with the use of sensory aids.

Older subjects who received training with either
the tactile or visual sensory aid also made significant progress. These children could be characterized as having more impervious phonatory problems owing to their age. Within this group, children who received speech training with a visual aid and who manifested phonatory problems of inappropriately high average pitch seemed to show advantages over subjects with similar problems who were using the tactile device. However, children with inappropriate intonation contours who used the tactile device seemed to show some advantage over subjects with similar problems who were using a visual display. This trend should be interpreted with caution due to the between-children differences and the small numbers in each group. We are presently analyzing data from additional subjects. Such trends make intuitive sense however, since a multichannel device may be better suited to conveying more dynamic changes in fundamental frequency than to presenting more static pitch levels.

Acknowledgements

The authors acknowledge the cooperation of the faculty, staff and students of the Lexington School for the Deaf.

REFERENCES