

# FEEDBACK CONSIDERATIONS FOR SPEECH TRAINING SYSTEMS

*James J. Mahshie*

Gallaudet University  
Speech Communication Laboratory  
Department of Audiology and Speech-Language Pathology  
Washington, DC 20002

## ABSTRACT

The aim of this study was to explore the role of visual feedback of speech information on the development and use of specific speech production skills by a deaf learner. A young, deaf, female subject received training in which visual feedback of airflow was used to teach production of a voicing distinction. The results showed that provision of the feedback resulted in improvements in the segments being taught, and that providing the feedback during production resulted in improved performance during some phases of teaching, but not during others. The findings have implications on our understanding of speech learning that relies on the use of a visual surrogate for hearing.

## 1. INTRODUCTION

Feedback plays a very important role in speech learning (1). As children mature, they develop an internal model of articulatory/phonatory outcomes or aims. While the model may change as the child develops, it becomes at least a heuristic standard against which the child measures her own production attempts. Through feedback of their production attempts, children learn to adjust and modify their articulatory patterns until the utterance serves the intended function. Thus feedback is the mediator between desired outcomes and production.

Development of speech skills is vastly different for a deaf or hard of hearing individual than for a hearing speech-learner. Without adequate "built-in" feedback, the individual must increasingly rely on alternative sources of speech information. While amplification may permit improved access to some aspects of the produced signal, amplification alone may be inadequate in many cases for successful acquisition of spoken language.

Visual or tactile displays of speech information are often employed as a surrogate source of feedback for speech learning. In particular, a number of computer-based devices have been developed to provide deaf learners with visual feedback about their speech production attempts. While the role such feedback plays in learning is likely quite different from that of auditory feedback, details concerning the use of such surrogate feedback and its influence on speech learning has not been thoroughly explored.

The purpose of this investigation was to examine the role of a visual display of speech parameters (in the absence of audition) on both the ability to produce patterns that are being taught, and

on the overall learning of these patterns. More specifically, an experiment is described in which a deaf adult was taught to produce articulatory patterns through visual feedback about aerodynamic information that would lead to more accurate production of a voiced/voiceless consonant distinction (/p/ vs. /b/). The study was designed to address the following two questions: i) What is the time course of change in speech behavior associated with speech training using aerodynamic feedback? and ii) How effective is visual feedback for online monitoring of speech accuracy as training progresses?

## 2. PROCEDURES

### 2.1. Subject

The subject in this single subject experiment was a 21 year old, female, congenitally deaf, college student. She had a profound, bilateral hearing loss with a pure tone average threshold of 108 dB in her better ear. The subject reported not using amplification in the three years prior to the study, and amplification was not worn during the seven weeks of the training study.

Her speech was minimally intelligible and was characterized by numerous misarticulations, including production of glottal stops for /h/ and production of all plosives as voiced. Training was accordingly directed toward production of correct /h/ patterns and an appropriate consonant voicing distinction between /p/ and /b/. The /h/ segment was selected because the devoicing gesture required for accurate production of /h/ is similar to that required for all unvoiced segments. The bilabial consonants were chosen because their place of production is typically achievable by deaf learners due to the visibility of the articulators.

### 2.2. Instrumentation

Airflow feedback was obtained from a circumferentially vented pneumotachograph system (Glottal Enterprises Model MS100). The output of the device was low pass filtered at around 50 Hz to obtain average airflow, and then input to a laboratory computer (DEC PDP-11/34a). Custom designed software enabled display of airflow vs. time similar to the display on a storage oscilloscope. To provide the learner with a clear indication of the production target to be attempted, a target pattern based on a production by the teacher was drawn on the screen of the monitor with a marking pen.

The computer-based system was also used to acquire data associated with each training session with the subject. The system enabled digital storage of the airflow signal associated

with production of the stimuli words, and subsequent retrieval and measurement.

### 2.3. Stimuli

Two comparable lists of twenty monosyllabic and bisyllabic utterances were developed for testing and training. The composition of the stimulus set was as follows: twelve words contained /b/ and /p/ in initial and intervocalic contexts (3 words \* 2 contexts \* 2 segments = 12), three words contained /h/ in initial contexts and five words contained /h/ in intervocalic contexts. One set (List A) was used for training as well as within session testing (see description below) while the second set (List B) was presented only after completion of the entire training program.

### 2.4. Training Protocol

Training was approached in two overlapping phases. The first phase involved teaching the appropriate laryngeal gesture for an unvoiced segment through practice producing /h/-containing words. The second phase of training involved accurate coordination of the laryngeal and oral gesture required for production of the voicing distinction for /p/ vs. /b/.

The subject received visual feedback of airflow associated with her attempts at production. The experimenter provided additional instruction and directions, but the only feedback provided was through the visual display. As the subject demonstrated production success, feedback was delayed. To achieve this, practice words were attempted without watching the screen so that the information concerning the accuracy of the production was presented after completion of the utterance rather than during production.

### 2.5. General Procedures

The subject was seen a total of 12 times over a seven week period. Each meeting was comprised of three testing blocks and a training period. A typical session was comprised of:

**1. Testing: Before-Training with NO Feedback (BNF):** After a brief warmup, the subject produced the List A words while the average airflow signal was recorded. This provided information about the subject's internalized or self-monitored "learning" that could be carried-over from session-to-session.

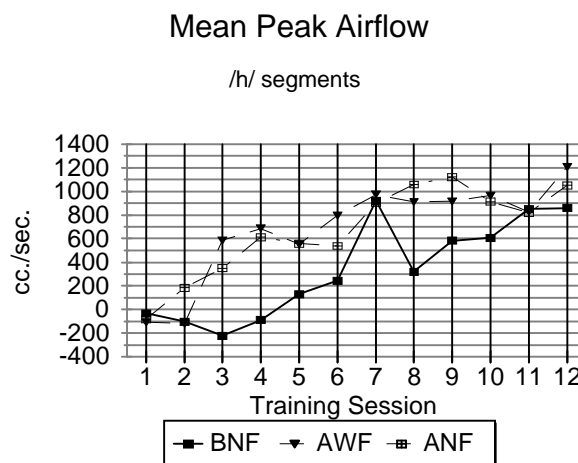
**2. Training period:** The training and practice portion of the session lasted for approximately 40 minutes. It involved production attempts in which the aerodynamic signal served as the primary source of feedback to the subject about how she was controlling the breath stream with her articulators. The List A stimuli set was used during this portion of the session.

**3. Testing - After-training WITH Feedback (AWF):** Following completion of the training portion of the session, the subject was again presented the List A stimuli list and was asked to produce the words while airflow signals were obtained. This time, the subject was allowed to see the display of her production attempt. The experimenter did not comment on the resultant patterns. This reflected the maximum usage of the feedback that might be expected from the subject.

**4. Testing - After-training with NO Feedback (ANF):** A second "after-training" testing block was also conducted in which the subject was again asked to produce the stimuli set while airflow signals were obtained. However, the subject was not provided access to the visual displays of her airflow patterns associated with these production attempts. This testing condition reflected the extent the subject was able to produce the pattern following an intensive feedback mediated training period without having to rely on the feedback.

### 2.6. Data Measurement

Measurements of peak airflow were made from printed copies of airflow waveforms associated with test productions from each session. There were thus 60 (20 stimuli x 3 conditions) traces measured for each subject for each session. Listener judgments of the productions from selected sessions were also obtained. The BNF data from three points in the training sequence (the first, seventh, and twelfth sessions) and the speaker's production of list B (the new stimulus list) were presented to a panel of ten judges. Tokens for each of the four sessions were randomized for each listener. Each judge was presented the tokens five times and using a four alternative forced choice paradigm<sup>1</sup>, judged the target consonant they heard.



**Figure 1:** Mean peak airflow associated with subject's production of medial and final /h/ during each of the training sessions. Data is shown for the Before No Feedback (BNF, After with Feedback (AWF) and the After with NO Feedback (ANF) test conditions.

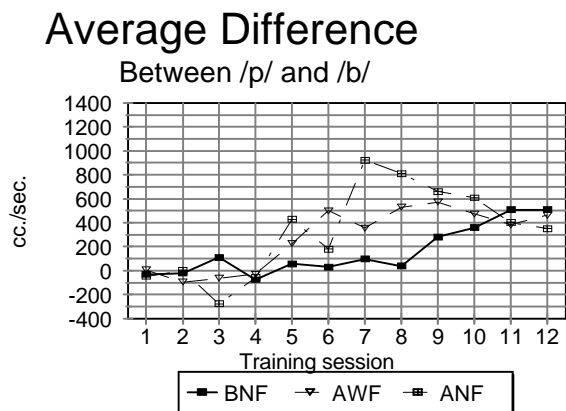
## 3. RESULTS

### 3.1. Aerodynamic results

Aerodynamic results are shown for /h/ production in Figure 1. Beginning in the second session the subject showed increased

<sup>1</sup>The listeners were given 4 alternatives, /p/, /b/, /h/ and the glottal stop /?/.

peak airflow associated with production of /h/ segments in the AWF and ANF conditions. However, there was little evidence of increased peak airflow in the BNF condition until much later in the training sequence. These findings suggest that: i) The use of aerodynamic feedback in conjunction with training corresponded with a rapid rise in her ability to produce increased airflow rates during /h/ production following the training period regardless of the presence or absence of the feedback during these attempts; and ii) Practice with the visual display appeared to have a cumulative affect on the overall learning of the task (as reflected in the BNF test condition). While learning appeared facilitated by the use of visual feedback, the subject did not appear to depend on the displayed information for online monitoring of her productions.



**Figure 2:** Difference between average Mean Peak Airflow for /p/ productions and average for /b/ productions. As measure varies from zero, it reflects distinction in production of the two voicing categories.

Figure 2 shows the difference in the peak average airflow between /p/ and /b/ segments. This measure reflects the distinction in the aerodynamic patterns associated with voiced vs. unvoiced plosives.

During the first three training sessions the subject worked exclusively on production of a laryngeal devoicing gesture. There was little indication during these sessions of any difference in peak airflow between the intended voiced and voiceless plosive segments. In session four, training was initiated on the plosive segments. While no distinction was observed in the peak airflow for /p/ vs. /b/ for that session, the subject did begin to demonstrate a distinction during the next session (session five). In sessions five through ten, the subject began demonstrating a dramatic increase in the peak airflow difference measure in the ANF condition. Interestingly, she tended to produce very large peak airflow patterns in the ANF condition in sessions five, seven and eight. These large peaks likely reflected the lack of fine control that might be expected as the learner was becoming aware of control of the laryngeal mechanism for this task. With continued practice, the subject began to produce more acceptable

patterns until by session ten and eleven, the subject showed fairly consistent distinction between /p/ and /b/ in the ANF condition.

In contrast, the subject began much earlier to show greater control in the laryngeal pattern for voiceless production during the AWF testing condition. Evidence for this was the production of a consistent and non-exaggerated distinction in peak airflow between /p/ and /b/. The consistent and controlled distinction was evident in the AWF condition beginning in session five. This suggests that the subject was able to use the aerodynamic feedback effectively in monitoring her productions.

As noted earlier, the BNF testing-condition is considered the best reflection of how “internalized” the motor pattern has become, since it reflects the learner’s performance without recent experience of visual feedback related to her production patterns. The subject showed little evidence of production of a distinction in the BNF condition until session nine, when a clear distinction was produced in the peak airflow rates between /p/ and /b/ segments. This distinction increased in the next two sessions, and stabilized at around 500 cc./sec. during the final session.

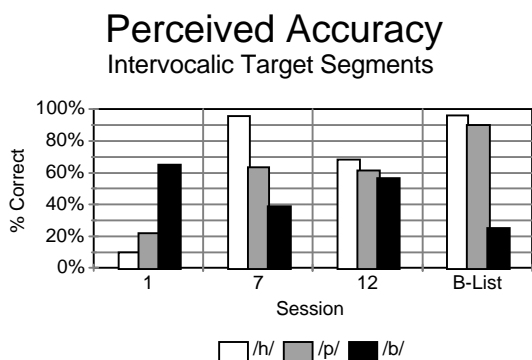
Together, these findings suggest that the subject was able to achieve the articulatory patterns needed to produce a distinction between /p/ and /b/ almost immediately upon initiation of instruction using visual feedback of airflow. Evidence of acquisition of appropriate production patterns was reflected in the after-training test data, with greater precision observed when feedback was provided than when it was withheld. Evidence for internalization of learning, however, was not observed until later in the training, as reflected in the BNF test condition. By the end of the twelve training sessions, the subject showed evidence of producing a distinction in average airflow between /p/ and /b/, and this distinction was evident in the data obtained both before and after the training portion of the session. This convergence of the data for the three testing conditions during the final few sessions suggests that a degree of motor learning has been accomplished that no longer requires online monitoring through the visual display provided by the training system.

Following conclusion of training, the subject was asked to produce the second or List B utterances. The results showed a consistent tendency to produce the /h/ segments with peak airflow reflecting use of an appropriate laryngeal opening-closing gesture. In addition, the subject also produced the /p/ segments with an average peak airflow that was 311 cc/sec greater than was used for production of the /b/ segments. This again suggests use of appropriate laryngeal behavior for production of the voicing distinction being trained.

### 3.2. Perceptual results

Figure 3 shows the perceived accuracy of the target segments in the intervocalic context for the first, seventh, and twelfth training sessions, as well as for the List B stimuli. All data are from the BNF condition and thus reflect the extent of internalized learning that has occurred at these points in the training sequence. The results suggest that while there was improvement in production

of the /h/ and /p/ production accuracy, the subject's production of /b/ segment voicing actually decreased. This is likely the result of overgeneralizing the production pattern.



**Figure 3:** Perceived accuracy of /h/, /p/ and /b/ productions at three points during the training sequence, and for production of the B-list words.

#### 4. IMPLICATIONS

This study was initiated to examine the role that external, aerodynamically-based visual-feedback has on both the learning of a speech skill, and on the maintenance outside of training of patterns being taught. The results suggest that such feedback can play an important role for both learning and production early in the process of learning a skill, but that with sufficient practice and experience, such external feedback becomes less necessary for accurate production.

A distinction has been made in the motor learning literature between knowledge of results (KR) feedback and knowledge of performance (KP) feedback (1,2). KR refers to augmented feedback related to the nature of the result produced, and typically reflects the correctness or incorrectness of the outcome of a behavior. In contrast, KP is augmented feedback about the production patterns actually used. While clear that KR is an essential component of learning a motor behavior (such as speech) the role of KP is not well understood. Intuitively, it makes sense that when speech is approached from a specific task perspective, that the learner would benefit from KP.

The present study suggests that KP can be a useful element of learning for a deaf individual learning to speak, but that once the behavior has stabilized, the moment to moment information provided by KP may not be necessary. It is likely that following attainment of consistent patterns on trained productions where the presence or absence of feedback or proximate practice opportunity appears to have little benefit, that the such feedback is likely less important. It is probable that the learner would benefit more from KP at this stage of learning, particularly as training focuses on automating the behavior (3) or generalizing it. Further study of the potential role of KP and its relation to KR is needed.

#### 5. REFERENCES

1. Mahshie, J. . The use of sensory aids for teaching speech to children who are deaf. In Spens, K.-E., and Plant, G. (Eds.) Speech Communication and Profound Deafness. London: Whurr Publishers, Ltd., 1995.
2. Schmidt, R. Motor Control and Learning. Champaign, IL: Human Kinetics Publishers, Inc., 1988.
3. Ling, D. Foundations of Spoken Language for Hearing-Impaired Children. Washington: A.G.Bell Association for the Deaf, 1989.