Effects of Temporal and Spectral Alterations of Speech Feedback on Speech Production by Persons With Hearing Loss

Dragana Barac-Cikoja, PhD

Objective: To evaluate the effects of delay and spectral alteration of speech feedback (SF) on the speaking rate and voice pitch in adult users of hearing aids (HAs) and cochlear implants (CIs).

Methods: Repeated-measure, completely crossed, 2-factor design. Spectral alterations were implemented by replacing SF with noise that was filtered into 1-, 2-, or 4-frequency bands and speech-modulated in real time. Delays varied from 25 to 200 milliseconds.

Subjects: Seven HA users with severe to profound hearing loss and 6 CI users were randomly recruited by public advertising. All were postlingually deafened adults with intelligible speech.

Results: The average speaking rate significantly decreased and rate variability significantly increased with increase in SF delay for both groups. Spectral alterations of SF reduced the effect on speaking rate in the HA group but not in the CI group. Spectral alterations did not significantly affect rate variability in the HA group but did so in the CI group. Average voice pitch increased significantly with increase in SF delay and with spectral alterations of SF in the HA group. No significant effects on average pitch of CI users were noted.

Conclusions: The 2 groups were affected differently by the delay and spectral alterations of SF. The differences possibly reflect greater spectral resolution ability in the case of CI users and greater audibility of bone-conducted SF (particularly in the low-frequency region) among the HA users.

Arch Otolaryngol Head Neck Surg. 2004;130:598-603

It has long been observed that prelingual deafness results in aberrant speech characteristics and that certain aspects of developed speech deteriorate with prolonged deafness.1-3 The apparent significance of hearing for the development and maintenance of speech production has been corroborated by complementary evidence from persons whose hearing has been partially restored via cochlear implants. Although postimplant speech changes vary substantially in kind and degree across patients, they are shown to be related to the age of the patients at implant and in particular to their age at onset of profound hearing loss.4,5

Separating the contribution of self-from other-generated speech in these processes is an unsettled theoretical issue and a difficult methodologic problem. If one thinks of speech acquisition via hearing as a process of discovering the systematic (language-determined) relationship between the sounds and the articulatory activity that can produce them, self-hearing has an essential role to play in tuning that mapping, while hearing others may serve primarily to establish linguistic significance (ie, the meaning of the sounds).6,7 Likewise, maintaining intelligible speech, both long-term and more immediately, with respect to specific communication circumstances requires, as much as any other skill, some information about the adequacy of the outcome. For instance, recent studies have shown that compensatory changes in individual sound production can be induced over time by systematically altering auditory feedback to indicate inaccurate articulation.8 To maintain speech audibility in different environments, it is clearly of great importance to monitor one’s own speech output relative to the particular noise conditions and make appropriate articulatory adjustments instantly.

More controversial, however, is the role of auditory feedback in the regulation of ongoing speech production. Although most researchers agree that auditory self-monitoring can influence some slowly changing aspects of speech production (prosodic organization of speech),9,10 it is less clear if it has any significance for directive control of articulation at the segmental level.11 Arguments against such a role in-
clude the fact that the speech of adventitiously deafened adults only slowly deteriorates at the segmental level; that a closed-loop feedback system would be too slow; particularly for consonant production; and that bone-conducted speech feedback (SF), because it is spectrally different from, but equally as loud as the air component, is analogous to a noise masker that reduces the overall feedback intelligibility (ie, the segmental information in SF).

The present study investigates the contribution of self-hearing to speech regulation in persons with hearing loss. It concentrates on adults who had developed speech before the onset of deafness but show changes in speech production due to prolonged deafness.

To investigate the specific contribution of self-hearing to speech regulation, researches have long used acoustic alterations of self-generated SF and analyses of their effects on various aspects of speech. The best-documented speech production response to feedback alterations involves speech delay. Although susceptibility to delayed auditory feedback (DAF) varies across individuals and may be affected by developmental factors as well as the person's speech fluency, the response commonly includes a decrease in speech rate, an increase in overall vocal intensity and pitch, and the emergence of dysfluencies. Such findings indicate the speaker's attention (not necessarily conscious) to auditory feedback and interference caused by asynchrony between the air-conducted auditory feedback and the speech production activity (with its accompanying somatosensory and bone-conducted auditory feedback).

Given its well-documented robust effects on speech production, a delayed feedback paradigm was chosen to probe the present subjects' reliance on their auditory feedback during speech production. The question asked was how susceptible to feedback delay are persons whose hearing is severely impaired (hearing aid [HA] group) or altered by the use of a cochlear implant (CI group). The hypothesis was that the effect of DAF would depend on the degree of auditory access to speech information in the feedback. That is, limited access to spectral details in SF for persons with severe to profound hearing loss was expected to reduce and possibly eliminate any temporal and pitch changes that a delay might impose on speech. In successful CI users, on the other hand, spectral resolution is significantly better, and their susceptibility to DAF was consequently expected to parallel the response in normal-hearing individuals.

Another goal of this study was to assess speech production effects when feedback intelligibility was experimentally manipulated. Specifically, the availability of spectral details in the feedback (and consequently its intelligibility) can be reduced by applying spectral smearing, a specific acoustic manipulation that has been effectivley used in speech perception studies. Spectral smearing is accomplished by filtering a speech signal into an experimentally manipulated number of frequency bands, extracting an amplitude contour for each band, and amplitude-modulating similarly filtered noise bands with the respective contours. As a result, speech is replaced by a signal the intelligibility of which is directly related to the number of noise bands that compose it. In the present study, this acoustic manipulation, spectrally altered feedback (SAF), has been applied in real time to each subject's speech and fed back to him or her.

This reduction in the signal's spectral details not only severely affects segmental identity by reducing or eliminating segmental information, but the information on speech intonation contour is absent, and the fine temporal structure of the running speech is progressively less specified. Lack of research with this particular feedback manipulation limits the predictions regarding its effects on speech production. However, based on the general understanding of the regulatory function of auditory feedback, it was expected that the effects would primarily manifest at the suprasegmental level (ie, in terms of temporal and pitch characteristics). Furthermore, it was expected that the imposed reduction in spectral information would affect the feedback's regulatory function and therefore reduce the speaker's susceptibility to feedback delay. Clearly, this would be the case only if the spectral information in the speech signal were accessible to the speaker to some degree. Thus, it was expected that between-group differences might be observed in the interaction between the 2 factors (temporal and spectral feedback alterations).

A note regarding the parameters of speech production that were selected for the analyses in this study: The decision to focus on the suprasegmental level was based on models of speech production regulation outlined earlier. However, given that spectral smearing changes the speech signal primarily at the segmental level, it is possible that such feedback would have an effect on articulatory precision at the segmental level. While such analyses are needed, they are beyond the scope of this report.

METHODS

SUBJECTS

Seven HA users and 5 CI users participated in this study. Data for 1 additional CI user were eliminated from the analyses owing to partial loss of her speech recordings caused by technical problems. Listeners in the HA group had severe to profound hearing loss in the better (tested) ear using a 3-frequency average. In the CI group, 2 subjects used a Clarion Hi Focus 1.2 model, 1 subject used a Clarion Platinum model, 1 used a Nucleus Esprit 24 (Cochlear Corporation, Sylmar, Calif), and 1 used a Nucleus Sprint. All devices had been implanted 9 to 18 months before the testing, and all patients had postlingual deafness. For all subjects, the preferred communication mode was aural/oral, and all had intelligible speech. Average age in the HA group was 23 years (range, 19-31 years) and in the CI group, 39 years (range, 21-54 years). Subjects were recruited through public advertising and consented to participate after reading a written description of the study. The project was approved by the Gallaudet University institutional review board, and subjects were paid for their participation.

APPARATUS

The subjects’ speech was recorded with a miniature microphone (Sennheiser MKE2; Sennheiser Electronic GmbH & Co, Wedemark, Germany) placed at their nontest ear. The microphone output was fed to a mixer (Mackie 1202 VLZPRO; Mackie...
The wideband speech signal (digitized into 16 bits at a sampling rate of 44100 Hz) and KYMA-generated white noise were fed (in parallel) through a filter bank with IIR bandpass filters (filter slopes at least 36 dB/octave). The number of filters (1, 2, or 4) was experimentally controlled. Filter cutoff frequencies were 300 Hz to 5500 Hz for a single-band condition (N1); 300 Hz to 1284 Hz and 1284 Hz to 5500 Hz for the 2-band condition (N2); and 300 Hz to 621 Hz, 621 Hz to 1285 Hz, 1285 Hz to 2658 Hz, and 2658 Hz to 5500 Hz, for the 4-band condition (N4). For each frequency band, signal amplitude was averaged over a 4-millisecond window and was used to modulate noise of the same bandwidth. Multiple band signals were combined, and the level of the new signal was adjusted to have the same average sound level as the input signal from the microphone. The processing introduced a delay between the input to and the output from the KYMA, which was kept fixed at 4 milliseconds and applied across all experimental conditions (although in the case of unaltered feedback, it was not technically necessary).

PROCEDURES

Subjects read a 6-sentence passage (Rainbow Passage) under each of 22 experimental conditions, including 4 spectral and 5 temporal manipulations of the auditory feedback in a completely crossed design. Four levels of spectral manipulation (SAF) consisted of SAF in addition to the 3 noise conditions (N1, N2, and N4). All feedback signals (speech or noise) were high-pass filtered at the cutoff frequency of 300 Hz. The 5 levels of temporal manipulation (DAF) included delays of 25, 50, 100, and 200 milliseconds and a no-delay condition. Presentation level was controlled via audiometer and was selected by the subject based on personal comfort level at the onset of the experiment. Subjects were tested individually in a sound-treated booth.

ACOUSTIC ANALYSES

Digitized speech recordings were used to compute duration and fundamental frequency (F0) contours for each sentence. Duration was measured from the initiation of the utterance’s energy to its termination based on the waveform and spectrogram displays. F0 contours consisted of F0 estimates made over 512 samples of data based on an autocorrelation function (WinPitch 1.8; Pitch Instruments Inc, Toronto, Ontario).

RESULTS

TEMPORAL ORGANIZATION OF SPEECH

The speech rate was calculated based on the average syllable duration across 6 sentences. Figure 1 shows changes in average speaking rate (syllables per second) as a function of the length of the delay and the spectral content of the feedback for the 2 groups of subjects. It is evident that subjects in both groups responded to an increasing delay in the speech condition (SF) by reducing their speech rate. However, the 2 groups differed in how the spectral smear of the feedback affected their response to delay. For the HA group, the effect of DAF on speech rate systematically declined with the reduction of spectral information in the feedback. For the CI group, no such interaction was observed: feedback delays affected speaking rate equally regardless of the degree of its spectral smear. Results of the repeated-measures design analysis of variance (ANOVA) with each individual subject’s average speaking rate as the dependent variable confirm these observations. The effect of DAF was found to be significant for both groups (F12,48=3.91, P < .001), not for the CI group (F12,48=1.74, P = .09). In-
individual subjects’ data were consistent with the group trends. An apparent increase in speaking rate at the 200-millisecond delay was not found statistically significant for either group.

Given that each speech sample comprised an uninterrupted sequence of 6 sentences, it was possible to assess the stability of speech rhythm as a function of the different feedback manipulations. The dependent measure for these analyses was the standard deviation (SD) calculated on syllabic duration values for each experimental condition and each subject. Figure 2 shows changes in SD as a function of the magnitude of delay and the degree of spectral alteration for the 2 groups of subjects. Inspection of the graph reveals that for the HA group (Figure 2A), increasing delay resulted in an overall increase in SD for all SAF conditions. In contrast, for the CI group (Figure 2B), the effect of delay varied as a function of SAF: For the SF, there was a marked increase in SD for the delays of 100 milliseconds or more, while no clear trends were evident for the 3 noise conditions. While repeated-measures ANOVA for the HA group proved a significant effect of DAF (F4,24 = 5.70, P < .002), neither SAF nor DAF×SAF was found significant (F3,18 = 1.63, P = .22 and F12,72 = 1.05, P = .42, respectively). Results for the CI group revealed a significant DAF×SAF in addition to a significant main effect of DAF (DAF, F4,10 = 3.99, P = .02; SAF, F3,12 = 1.48, P = .27; DAF×SAF, F12,48 = 2.25, P = .02).

To further examine this interaction, tests of simple effects of each factor were conducted. A significant difference among DAF means was found only for the speech condition (F4,16 = 13.02, P < .001). The subsequent t tests (on differences between means for the adjacent levels of DAF) showed that a significant increase in SD occurred only between 50 and 100 milliseconds (t4 = −5.33, P < .01). On the other hand, the analysis of simple effects of SAF at each DAF level revealed no significant differences.

In summary, while the HA group showed a tendency to speak at a more variable rate as feedback delay increased, the CI group showed a differential response to delay as a function of the spectral content of the feedback. For the SF, variability in the production rate rose significantly when the feedback was considerably delayed (>50 milliseconds). For the spectrally smeared feedback, a delay did not add significantly to the speech rate variability.

**VOICE PITCH**

Two dependent measures were derived from the individual F0 values that made up sentential contour: their mean and SD, representing average pitch and its variability, respectively. Figure 3 shows changes in average F0 (measured in hertz) as a function of the size of delay and the spectral content of the feedback for the HA group. One can see that the average pitch was the lowest for the SF with either no delay or a relatively short delay (25 milliseconds) and rose considerably for delays of 50 milliseconds or longer. The spectral smearing resulted in higher pitch in general, and the changes due to delay were relatively small. Repeated-measures ANOVA on average F0 for the HA group showed a significant main effect of DAF and DAF×SAF (F4,24 = 3.30, P = .03 and F12,72 = 2.50, P = .01, respectively). A significant simple effect of DAF was found for the speech condition (F4,10 = 3.84, P = .02) but not for any of the 3 noise conditions. Thus, significant interaction reflects a differential effect of delay on average pitch for speech vs SAF.
The effect of feedback manipulations on voice pitch proved significant for the HA group but was less clear in the case of CI users. Specifically, HA subjects significantly raised their voice pitch when the feedback was spectrally altered and when unaltered speech was substantially delayed (≥50 milliseconds). The increase in pitch was comparable across different feedback manipulations. As a group, CI users showed no significant effect on voice pitch under any of the feedback manipulations. However, this finding should be considered with caution given the apparent individual differences in the direction of pitch changes under altered feedback. Future research should provide more data to evaluate this issue.

Two arguments may illuminate these findings. In normal-hearing adults, the response to delayed feedback most commonly includes pitch increase and is interpreted as an indication of an increase in the overall speech effort induced by difficult listening conditions. Findings from the HA group are consistent with this trend and perhaps warrant the same interpretation.

Second, subjects in the HA group might have received significant information about their voice pitch through bone conduction. Given the amount of their hearing loss in the low-frequency region where F0 changes occur, pitch information was potentially audible to them. In contrast, CI users have no access to such signal, and the low-frequency information was filtered out from their feedback. The absence of such information could account for the inconsistencies in their response to various feedback alterations.

For the CI group, on the other hand, repeated-measures ANOVA on average pitch data showed no significant main effects or interaction. Inspection of Figure 4 suggests a possible reason. Figure 4A shows subjects’ data for the speech condition with delay as a parameter, and Figure 4B illustrates the no-delay condition with spectral content as a parameter. One can see that the subjects were split in terms of the general direction of their pitch changes (increase vs decrease) in response to feedback alterations. For instance, while 3 subjects showed a tendency to increase their voice pitch in response to increasing delay or spectral smearing, 1 clearly showed the opposite trend. As a result, changes that may have been due to the experimental manipulations were cancelled out. Data from this study do not allow statistical tests of this possibility.

The effect of feedback manipulations on the stability of voice pitch was assessed using the SD calculated on F0 values (SD$_{F0}$) from the sentential pitch contour. Analysis of variance revealed no significant effect on either the HA or the CI subjects. Careful inspection of individual subjects’ plots of SD$_{F0}$ as a function of DAF and SAF did not provide additional insights into possible trends. Although it is possible that temporal and spectral alterations of the feedback have no systematic influence on the stability of voice pitch during speech production, it is also possible that SD$_{F0}$ was not a sensitive measure of this function.

The findings of this study have justified the use of an altered feedback paradigm to investigate the function of auditory feedback in speech regulation. Significantly slower speech and a more variable speaking rate were observed in response to delaying the SF even when feedback audibility was substantially limited by the degree of hearing loss (severe to profound in the HA group) and when the asynchronous air-conducted signal was not competing with the bone-conducted speech (in the case of CI users). For the HA group, however, such effects were significantly reduced with the loss of speech information that was introduced by spectral smearing. Taken together, these findings suggest that the speakers were attending to their auditory feedback, and when the speech information was not sufficiently discernible owing to the acoustic degradations, asynchrony of production and auditory feedback was of no consequence to their speech production. In contrast, the CI users’ response to feedback delay was not significantly changed by spectral alterations, indicating that even the most spectrally degraded feedback provided some speech information for these speakers. This conclusion is perhaps further corroborated by the finding that the regularity of speaking rate significantly decreased under SAF (taking such variability to indicate an increasing difficulty in detecting speech information in the feedback).

In contrast, CI users have no access to such signal, and the low-frequency information was not sufficiently discernible owing to the acoustic degradations, asynchrony of production and auditory feedback was of no consequence to their speech production. Two arguments may illuminate these findings. In normal-hearing adults, the response to delayed feedback most commonly includes pitch increase and is interpreted as an indication of an increase in the overall speech effort induced by difficult listening conditions.

Two arguments may illuminate these findings. In normal-hearing adults, the response to delayed feedback most commonly includes pitch increase and is interpreted as an indication of an increase in the overall speech effort induced by difficult listening conditions.
In summary, this study focused on speakers who already had an established mapping between the sounds and the articulatory activity that can produce them. It sought to discover how experimentally imposed changes in self-generated speech sounds may affect articulation. The results suggest that the effects depend on the degree to which this mapping has been disturbed. At this time, little is known about the regulatory role of auditory feedback in the speech-acquisition stage. It is not possible to predict if attention to auditory feedback would equally depend on the quality of its content and how discernible the speech information is when the perception-production correspondence is still developing. However, acoustic alterations of speech feedback are a promising tool for investigating such questions.

In conclusion, temporal and spectral alterations of speech feedback influenced the speaking rate and voice pitch differentially for the 2 groups of subjects. The results indicated that the impact of auditory feedback on speech production depended on the amount and kind of speech information this feedback provided to the speaker. The method proved useful in probing the regulation of speech production based on auditory feedback. Future work should examine the effects on other parameters of speech production, including those pertaining to segmental articulation. Also warranted would be research focusing on developmental changes in the manner and the degree to which auditory feedback is used for speech regulation.

Submitted for publication September 2, 2003; final revision received November 20, 2003; accepted November 26, 2003.

This work was supported by the Priority Area Grant from the Gallaudet Research Institute.

This study was presented at the Ninth Symposium on Cochlear Implants in Children; April 25, 2003; Washington, DC. Preliminary results were also reported at the 141st meeting of the Acoustical Society of America; June 8, 2001; Chicago, Ill.; and at the Conference on Implantable Auditory Prostheses; August 20, 2001; Asilomar, Calif.

I thank Chizuko Tamaki, BS, Nancy McIntosh, AuD, and Lannie Thomas, BS, for their help in collecting and analyzing data. I am also thankful to Peggy Nelson, PhD, for helpful discussions of various aspects of this study, and Paula Tucker, EdS, and Robert C. Johnson, MA, for help in preparation of the manuscript.

Corresponding author and reprints: Dragana Barac-Cikoja, PhD, Gallaudet Research Institute, HMB N205F, Gallaudet University, 800 Florida Ave NE, Washington, DC 20002 (e-mail: dragana.barac-cikoja@gallaudet.edu).

REFERENCES

1. Levitt H, McGarr NS, Geffner D. Development of language and communication skills in hearing impaired children. ASHA Monogr. 1987;26:1-158.