Speech of cochlear implant patients: A longitudinal study of vowel production

Joseph Perkell, Harlan Lane, Mario Svirsky, and Jane Webster
Massachusetts Institute of Technology, Research Laboratory of Electronics, Room 36-511, 50 Vassar Street, Cambridge, Massachusetts 02139

(Received 10 June 1991; accepted for publication 3 February 1992)

Acoustic parameters were measured for vowels spoken in /hVd/ context by four postlingually deafened recipients of multichannel (Ineraid) cochlear implants. Three of the subjects became totally deaf in adulthood after varying periods of partial hearing loss; the fourth became totally deaf at age four. The subjects received different degrees of perceptual benefit from the prosthesis. Recordings were made before, and at intervals following speech processor activation. The measured parameters included F1, F2, F0, SPL, duration, and amplitude difference between the first two harmonic peaks in the log magnitude spectrum (H1–H2). Numerous changes in parameter values were observed from pre- to post-implant, with differences among subjects. Many changes, but not all, were in the direction of normative data, and most changes were consistent with hypotheses about relations among the parameters. Some of the changes tended to enhance phonemic contrasts; others had the opposite effect. For three subjects, H1–H2 changed in a direction consistent with measurements of their average air flow when reading; that relation was more complex for the fourth subject. The results are interpreted with respect to: characteristics of the individual subjects, including vowel identification scores; mechanical interactions among glottal and supraglottal articulations; and hypotheses about the role of auditory feedback in the control of speech production. Almost all the observed differences could be attributed to changes in the average settings of speaking rate, F0 and SPL, which presumably can be perceived without the need for spectral place information. Some observed F2 realignment may be attributable to the reception of spectral cues.

PACS numbers: 43.70.Dn, 43.70.Fq, 43.70.Bk, 43.66.Ts

INTRODUCTION

This paper reports on vowel production in four patients before and after they received four-channel Ineraid cochlear implants (Youngblood and Robinson, 1988). This study is one of a series designed to explore the effects on speech production of restoring some hearing to postlingually deafened adults. The long-term goals of this work are (1) to contribute to the evaluation of cochlear prostheses and (2) to clarify the role of auditory feedback in the control of adult speech production. A previous study (Lane et al., 1991) reported on the speech breathing of three of the four subjects in the present experiment.

I. BACKGROUND

Lane and Webster (1991) review some studies of the speech of postlingually deafened speakers, including several on the effects of cochlear implants. They summarize as follows:

"Taken together, the studies ... implicate a role for audition in regulating, directly or indirectly, several speech properties; notably, voice quality, voicing, aspiration, pitch, intonation, stress, tempo, nasality and fricative and plosive articulation. However, there is wide variation in the reported speech anomalies introduced by profound deafening and in the consequences of reintroducing some hearing. The picture is clouded by the varying measures employed—ratings, transcriptions, and acoustic parameters; and, in some studies, by the experimenters' knowledge of their patients' auditory capacities and the concomitant use of speech therapy" (pp. 860–861).

Lane and Webster's own subjects were three postlingually deafened adults (different from the current subjects) who had not received speech therapy. In comparison to age- and gender-matched hearing controls, they were found to have more variable and higher segmental and suprasegmental F0, decreased speaking rate, reduced differentiation between palatal and alveolar voiceless fricatives, and reduced differentiation with respect to place of articulation in the spectra of voiced and voiceless plosive release bursts.

Lane et al. (1991) report on measures of speech breathing of three of the current subjects who read a paragraph during the same recording sessions used for this investigation. Using indirect volumetric measurements (made with an inductive plethysmograph), Lane et al. found that activa-
tion of each speaker’s cochlear prosthesis was followed by a significant change in average airflow during reading, which rose for the two subjects (called FA and FB in this report) with initially low airflow and fell for the one subject (MA) with a much higher average airflow pre-implant. These changes in average airflow were accompanied by corresponding changes in volume of air expended per syllable. It was suggested that the findings were due to changes in laryngeal valving as well as respiratory driving force.

A. Theoretical overview

On the one hand, speech deteriorates after prolonged profound deafening in adulthood, and the reintroduction of some self-hearing is followed by more normal speech respiration. On the other hand, most of the changes induced by deafening concern speech quality; the subject’s overall ability to speak, and usually to speak rather intelligibly is not usually undermined (Cowie and Douglas-Cowie, 1983; Goehl and Kaufman, 1984). Based in part on such background, we hypothesize that, in normally hearing adults, speech production mechanisms use nonauditory afferent information for their moment-to-moment function. The development and maturation of production mechanisms rely heavily on the use of auditory feedback; after maturation, however, the moment-to-moment function of production mechanisms is largely independent of auditory feedback (Lane and Tranel, 1971). Hearing can be considered to have two major roles in maintaining the communicative effectiveness of these mechanisms in adults. (1) Self-hearing helps to calibrate production mechanisms by monitoring relations between the speaker’s own articulations and his/her acoustic output. This calibration is performed in the face of numerous perturbations, ranging from adjustments in the speaker’s body posture to changing listener demands. (2) The speaker can also validate his acoustic output by observing the behavior of his listeners and by detecting discrepancies between his own speech and theirs. Thus validation is carried out with reference to the social environment.

Following the loss of hearing in adulthood, differential deterioration of speech parameters occurs at different rates, depending on such factors as mechanical constraints and the required precision of particular articulations, as well as the availability of alternative sources of calibrating information. In spite of the fact that the deteriorating speech may be intelligible under most circumstances, it will be characterized by anomalies. Some anomalies may be subtle, but very revealing about the role of auditory feedback (cf. Lane and Webster, 1991).

Reintroduction of “auditory” input with electrical stimulation of the eighth nerve (i.e., from activation of the pro-cessor of a cochlear implant) should result in changes in speech production. The nature of these changes will depend on many factors, including the type of information derived from the stimulation; how well the patient can make use of that information; the speaker’s articulatory adaptation to prolonged, profound deafness; differential dependence of speech characteristics on alternative feedback modalities in the deafened condition; and interdependencies among speech parameters.

B. Hypotheses

As described in detail below, we made longitudinal measurements on productions of nine different vowels in /hVd/ context, recorded before and after activation of the speech processor of a cochlear implant. Parameters measured were $F_1$, $F_2$, $F_0$, SPL, duration, and an indirect index of glottal aperture ($H_1-H_2$). We also had available longitudinal measures of average rate of airflow from paragraph readings during the same recording sessions (Lane et al., 1991) and measures of the subjects’ performance on a test of vowel identification (Rabinowicz et al., 1991). We developed several hypotheses about these parameters and how they would change over time, as a result of activation of a cochlear prosthesis.

Relations to normative absolute values and patterns of contrast. We expected to observe preactivation differences from normative values for some parameters. Depending on the parameter, degree of abnormality can be characterized in terms of absolute parameter values for individual vowels and groups of vowels pooled, as well as in terms of relative values for contrastive categories of vowels.

Relations between perceptual gains and changes in production. Following processor activation, we expected to observe normalization of absolute parameter values and enhancement of phonemic contrasts. The amount of normalization should be related to subject-specific differences in degree of pre-activation abnormality of the parameters concerned, and subsequent perceptual benefit from the prosthesis.

Interdependencies among production parameters. The measured parameters are not independent of one another. For example, there may be related changes in $F_0$ and SPL because both parameters can be influenced in part by the same glottal adjustments or by subglottal pressure. Or there might be a relation between duration and $F_1$ for low vowels, because at higher speaking rates there is less time for the tongue body to reach a maximally lowered position (cf. Lindblom, 1963).

Relative timing of changes in different parameters. Different parameters or groups of parameters could change at different rates following processor activation, depending on several factors such as: degree of parameter abnormality pre-implant, allowable variation in the parameter given phonemic contrasts, and degree of reliance on the new “auditory” input for parameter calibration (as opposed to continued reliance on alternative modalities such as tactile and proprioceptive feedback).

II. METHODS

Cochlear implant patients are not a homogeneous group. In the first place, many speech parameters of males and females have different baseline values. Furthermore, each cochlear-implant patient in our study (like the deaf subjects of Lane and Webster, 1991) had a distinct configuration of speech anomalies pre-activation—partly the result, no doubt, of their different histories of profound, prolonged deafness and their adaptation to it. Moreover, each of our patients had a unique processor tuning and configuration of electrodes along the length of the cochlea. Finally, vowel
productions may vary considerably and still be perceived as intended; thus, achieving precise vowel targets may be less critical to some speakers than others. We should expect then, that for most parameters, each subject will respond to stimulation differently. Consequently, we have adopted a single-subject longitudinal design which is replicated with four different subjects spanning both genders, a range of perceptual benefit from their prostheses, and diverse etiologies and ages at initial hearing loss. Findings are discussed separately and statistical tests are conducted separately for each subject. Despite the differences among patients, it proves possible to discern common patterns—for example, trends toward normal mean parameter values where their prestimulation values were abnormal.

A. Implant characteristics, subjects, and speech materials

The Ineraid cochlear implant (Richards Medical Co.) consists of an implanted electrode array, a percutaneous pedestal and connector, and an external sound processor (Eddington, 1983; Youngblood and Robinson, 1988). The sound processor has an ear level microphone, a wideband automatic gain control, and a four-channel overlapping bandpass filter system with crossover frequencies of approximately 0.7, 1.4, and 2.3 kHz. The four analog filter outputs are delivered (via the percutaneous connector) individually to four monopolar intracochlear electrodes, with a common return electrode. The electrodes, spaced approximately 4 mm apart, were successfully positioned in all subjects by insertion into the scala tympani through the round window, with the first placed most apically, some 22 mm from the round window. Gain controls include user adjustments for input sensitivity and volume, and channel specific gains that are set (internally) for each subject.

Table I presents information concerning the subjects. Subject FA had a congenital monaural impairment but became totally deaf at age 33. Subject FB had normal hearing until age 21 and bilateral progressive hearing loss until age 40. Subject MA had a severe binaural hearing loss since birth; he became totally deaf four years prior to receiving his implant, i.e., at age 31. Subject MB became completely deaf at age 4, following meningitis. All four subjects wore hearing aids for extended periods during their lives. Consistent with reports that link age at hearing loss to intelligibility in adulthood (Cowie and Douglas-Cowie, 1983; Cowie et al., 1988), we noted informally that the speech of subjects MA and MB was somewhat difficult to understand while subjects FA and FB were quite intelligible; a formal evaluation has not been conducted, however.

All four subjects had pure tone average losses greater than 110 dB in each ear prior to implant; they derived no benefit from amplification and performed at chance levels on auditory tests of closed-set word recognition. The most recent post-stimulation values on a test of auditory four-vowel, forced-choice identification are shown in Table I. (Longitudinal values from this test are plotted in Fig. 6 and discussed below in Sec. III B 2.) Subjects FA, FB, and MA perform substantially above chance on this test; these three subjects have continued to use their prostheses regularly. Subject MB, however, received no measurable perceptual benefit from the implant. He gradually decreased his use of the prosthesis over the two years following processor activation, using it only at work toward the end of this period. After two years, he ceased using it altogether.

If there are abrupt changes in MB's speech parameters associated with activation of the processor, this would suggest that changes in speech production may be a more sensitive indicator of perceptual capabilities than speech discrimination tests, and that such changes may occur in response to relatively gross properties of the speech waveform, such as SPL. Thus MB is a type of control subject, in that systematic production changes in the absence of measured perceptual benefit provide terms of comparison for the interpretation of those changes in the more usual case of significantly enhanced speech discrimination following processor activation.

There were two pre-activation baseline recording sessions; they were separated by intervals of 10, 1, 4, and 9 weeks, respectively, for FA, FB, MA, and MB. Some pre-activation recordings were made before the implant surgery and others following it. Then the speech processor of the cochlear implant was activated. Post-activation recordings were made at intervals of approximately 0, 4, 12, 26, and 52 weeks. Subject FA made an additional recording at 85 weeks and MB at 85 and 141 weeks. The subjects did not receive explicit auditory training or speech therapy during the course of this research.

The speech material consisted of nine vowels spoken in the carrier phrase “It's a /hVd/ again”. The vowels and their corresponding presentation words were: /i/ (“heed”), /ɪ/ (“hid”), /ɛ/ (“head”), /æ/ (“had”), /ɑ/ (“had”), /ɛ/ (“hawed”), /u/ (“hood”), and /u/ (“who’d”). These utterances were arranged in a sequence read three times; other speech material was recited for approximately 20 min between each reading. The data also include average measures of speech respiration (ml/s) and of articulation rate (syllables/s) during three readings of a paragraph in the same recording sessions (see Lane et al., 1991, for details), and, finally, the vowel identification scores mentioned above.
B. Recording and calibration procedures

The subject was seated in a comfortable chair in a sound-attenuating room. A small electret microphone, attached with a stiff wire to a headband, was placed at a fixed distance of 20 cm from the subject's lips. The utterance materials were projected on a screen located several feet in front of the subject. For calibration of sound-pressure level, a sound source (electrolytix) was placed in front of the subject's lips, while an experimenter observed the sound pressure level (C scale) on a sound level meter held next to the electret microphone. The sound-pressure signal was recorded on an instrumentation recorder (either a Honeywell 5600E or a TEAC RD 111T PCM recorder).

C. Signal processing and data analysis

**Digitization and signal processing.** The recorded signal was low-pass filtered and digitized at 10 kHz. Digitization, signal processing and interactive data extraction were performed with procedures written in the MITSYN languages (Henke, 1989; Perkell et al., 1991) running on a Digital Equipment Corporation engineering workstation.

**Data extraction and analysis.** A command procedure, implemented with a MITSYN Command Language script, was developed to facilitate data extraction and calculation of a number of derived parameters. Its function is described in Perkell (1991).

For this report, we have performed graphical and statistical analyses of several variables. The frequency of F1 and F2, SPL, and H1-H2 were determined at the mid-vowel point, using a 51.2-ms window for calculating acoustic spectra and SPL. The parameter H1-H2 (the amplitude difference between the first two harmonics in the acoustic spectrum) is an indicator of the degree to which the glottal airflow waveform tends to be sinusoidal in shape. There is evidence that H1-H2 is correlated with the open quotient of the glottal waveform and perceived "breathiness" of the voice (cf. Ladefoged, 1981; Bickley, 1982; Klatt and Klatt, 1990). The measure of H1-H2 is invalid when the frequency of the second harmonic in the spectrum is within 150 Hz of the first formant (see Perkell, 1991); therefore, H1-H2 results are limited mainly to the nonhigh vowels. To calculate the SPL of each vowel token, the rms of the recorded, digitized sound-pressure signal was divided by the rms of the calibration tone and converted to dB.

Parameter changes from the two pre-activation sessions pooled to the last two post-activation sessions pooled were examined for each individual subject, using a two-way repeated measures analysis of variance, with activation and vowel type as main effects. Each cell contained all six repetitions of each presented vowel, even though not all productions were judged by the experimenters to be the same as the presented (stimulus) vowel. For each F-test of a significant change pre- to post-activation, there was 1 degree of freedom associated with the numerator, and 5 with the denominator, the between-tokens mean square. For some parameters, correlations were calculated to examine relations of each subject's data to normative mean values in the literature; the associated N was 9 (the number of vowels). Relations between pairs of parameters were also examined by correlating parameter values of individual tokens across sessions and by correlating values across vowels, within pre- and post-activation conditions. In addition, for each parameter, graphical analyses were made of longitudinal trends and of relations among the nine vowels pre- versus post-activation.

III. RESULTS AND DISCUSSION

All of the measured parameters changed with activation of the subjects' speech processors. In many respects, the results differ among the subjects, so to identify meaningful patterns in the data, we examine them in detail in the following sections on (1) characteristics of parameter changes; (2) the timing of those changes (longitudinal trends) and their relation to a measure of perceptual benefit; and (3) relations among the parameters.

A. Characteristics of parameter changes following activation of the processor

This section describes changes in parameter values, averaged across vowels, and for individual vowels, from the pre-activation condition (the two pre-activation sessions) to the post-activation condition (the last two post-activation sessions). The following observations concentrate on (a) the relation of values to normative data averaged across vowels, and/or (b) patterns of contrast of parameters among the nine vowels pro- versus post-activation. An asterisk below the difference value indicates a significant F (p < 0.05). For the first five parameters, product-moment correlations are shown pre- and post-activation (when significant at p < 0.05); they were computed by pairing off mean values for each vowel within each condition with normative values from the literature (Peterson and Barney, 1952, for F1, F2, and F0; Peterson and Lehiste, 1960, for duration; and Lehiste and Peterson, 1959, for SPL).

There were pre- to post-activation differences in average parameter values that indicate overall changes in average formant values, SPL, F0 and speaking rate and reflect changes in the settings of the underlying physiological mechanisms that control them; we will refer to these as "postural" changes. Pre- to post-activation differences for individual vowels reveal changes in patterns of contrast of parameters such as duration and formant frequencies.

**Vowel space.** Table II shows that, averaged across all the vowels, F1 decreased pre- to post-activation in all subjects; the drop for FA was not statistically reliable, however. Average F2 decreased for FA and MB, and increased for FB and MA; the latter change was not statistically reliable.
Figure 1 shows average pre- versus post-activation formant values by vowel for F1 in the top half and F2 in the bottom half. Data from the two female subjects are plotted in the left two panels, and data from the two male subjects are plotted in the right two panels. The highest vowels are on the ends of each plot and the lowest vowels are in the middle, with front vowels on the left and back vowels on the right. The open circles represent the pre-activation average values (in Hz) for each vowel (from the two pre-activation recordings) and closed circles represent post-activation average values (from the last two recordings). Error bars indicate +/- one standard error about the mean; in some cases small error bars are obscured by the circles. Normative mean values from Peterson and Barney (1952, P&B) are indicated by small closed squares connected with dotted lines. (As in some of the following plots, parameter scales and ranges are consistent for the two females and for the two males, but different for the males versus the females.)

For the most part, changes for individual vowels in Fig. 1 show that the lower vowels accounted for more of the F1 decrease than the higher vowels, indicating a higher tongue position following processor activation, especially for the lower vowels. The overall F2 changes suggest some tongue backing for FA and MB and slight fronting for FB. The average F1 changes noted above were away from normative mean (P&B) values for MA and MB, with a marked compression of the F1 range compared to the P&B values for MA. MB's F1 range was compressed pre- and post-activation. The congruence of MB's F1 values for individual vowels with respect to the pattern of P&B means decreased somewhat following processor activation. For the other three subjects, the general patterning of F1 over vowel quality was altered very little from pre- to post-activation, showing little change in congruence with the P&B patterns (see r values in Table II).

The effect of processor activation on the second formant was different from its effect on F1. Three of the four speakers showed significant F2 changes, and these were toward P&B values. This F2 realignment took place mainly among the front vowels. For FA, the front vowels, incorrectly ordered on F2 initially, were correctly ordered post-activation and their average deviation from the P&B values had decreased 15%. For FB, the decrease was 56%. For MB an apparent increase in congruence with P&B values (see r values in Table II) is due to the large decrease in F2 for /a/; however, his pre-activation pronunciations of the stimulus word did not sound consistently like the target word "hod". For FA and FB, the back vowels changed little in the direction of the P&B values, in spite of considerable differences from those normative mean values in their pre-activation data.

The highlights of the formant changes are as follows: compression of F1 range and movement toward normative mean values of F2 in the front vowels for some subjects.
FIG. 1. Average pre- (open circles) versus post-activation (closed circles) formant values by vowel (for F1 in the top half and F2 in the bottom half), with normative values from Peterson and Barney (1952) indicated by small closed squares connected with dotted lines. The “highest” vowels are on the ends of the plots and the lowest vowels are in the middle, with “front” vowels on the left and “back” vowels on the right. Error bars indicate ± one standard error about the mean.

FIG. 2. Pre- and post-activation mean vowel nucleus durations in msec, by vowel, along with normative mean data from Peterson and Lehiste (1960). The symbols are used in the same way as in Fig. 1, but the vowels are ordered differently: the “lax” vowels, /i/, /e/, /ɛ/, and /u/ (cf. Chomsky and Halle, 1968) are in the left half of each panel.

Changes are consistent with articulation rate increases observed in paragraph readings from the same recording sessions: FA increased 7.5% (4.0 syl/s pre, 4.3 post); FB, 7% (4.3 pre, 4.6 post); MA, 26% (2.7 pre, 3.4 post); and MB, 10% (3.0 pre, 3.3 post). (All changes were significant at \( p < 0.01 \), except the first where \( p = 0.08 \); see Lane et al., 1991, for details.) Figure 2 shows pre- and post-activation mean vowel nucleus durations in ms, by vowel, along with normative mean data from Peterson and Lehiste (1960). The symbols are used in the same way as in Fig. 1, but the vowels are ordered differently: the “lax” vowels, /i/, /ɛ/, /ɛ/, and /u/ (cf. Chomsky and Halle, 1968) are in the left half of each panel. FA, FB, and MB had an exaggerated tense-lax durational contrast pre-activation and preserved that pattern in the course of decreasing durations post-activation; however, FB also compressed the overall range of durations somewhat in the direction of normative mean values. On the other hand, MA had a pre-activation pattern of durations that was poorly congruent with the normative pattern and did not have the same kind of well-defined tense-lax durational contrast. Post activation, while decreasing durations overall, he compressed the range to be considerably smaller than the normative one.

Inference from these results suggests that FA, FB, and MB may have been using exaggerated duration cues along with some formant information to signal the tense-lax distinction pre-activation and would not give it up post-activation. Perhaps their newly-acquired “hearing,” while providing some formant information, was not sufficient to enable
FIG. 3. Pre- and post-activation mean F0 by vowel, along with normative mean data from Peterson and Barney (1952). Symbols and vowel ordering are as in Fig. 1.

FIG. 4. Average pre- and post-activation SPL by vowel, along with normative mean data from Lehiste and Peterson (1959) plotted as in Fig. 2. For all four subjects, pre-activation, open vowels, had higher SPL values than more closed vowels, presumably due to more efficient radiation of sound from the open oral cavity, as in hearing speakers. Post-activation this pattern changed minimally for MB, somewhat for MA because of a decrease in SPL for the vowel /e/, and markedly for FA and FB. For FA, the post-activation pattern of openness-related SPL is the inverse of the normative pattern, and for FB post-activation, the vowels /æ/, /o/, and /u/ have levels that are in the same range as /i/ and /u/, about 5 dB lower than the levels for /I/, /œ/, /n/, and /u/. These changes away from the normative pattern observed for FA and FB are discussed to them to rely more on formant cues and less on duration to signal the tense-lax contrast.

Fundamental frequency (F0). Reference to Table II shows that average F0 decreased dramatically for FA, from a high pre-activation mean of 292 Hz, by 101 Hz. It did not change significantly for FB or from high pre-activation means of 155 for MA and 163 for MB. Figure 3 shows pre- and post-activation mean F0 by vowel, along with normative mean data from Peterson and Barney (1952). Symbols and vowel ordering are as in Fig. 1. The pre-activation data of subjects FA and FB have a height-related pattern of "intrinsic F0" across vowels which is congruent with the normative one but appears in the figure to be somewhat exaggerated. This result is consistent with Bush's (1979) finding of exaggerated intrinsic F0 relations and among the vowels of congenitally deaf children. The overall range of F0 is diminished post-activation for FA; the diminution is approximately proportional to the amount of the average F0 decrease, and the exaggeration of the intrinsic F0 pattern has diminished. The plot for FB shows some decrease in the exaggeration of the intrinsic F0 pattern. For MA and MB, the pre-activation pattern of F0 was also exaggerated in relation to the P&B data, but congruence with P&B values was somewhat less than for FA and FB (see r values, Table II). Post-activation, the exaggerated pre-activation range persisted for MB and (with exception of the vowel /u/) diminished somewhat for MA. MA and MB had significant linear correlations with P&B values pre-activation; those relations were not significant post-activation (see Table II). Thus, congruence with the P&B pattern diminished for MB and MA; in the figure, it appears that this decrease in congruence is mainly due to a paradoxical lowering of F0 for the vowel /u/.

To summarize: one subject with abnormally high F0 lowered it significantly post activation. All subjects had exaggerated intrinsic F0 differences among vowels pre-and post-implant, although visual inspection of the plots suggests a slight post-implant diminution of the exaggerated intrinsic F0 differences. The congruence with the normative pattern decreased for the two male subjects, post-activation. SPL. Averaged across all vowels, SPL decreased for each of the four subjects (Table II). Figure 4 shows average pre- and post-activation SPL by vowel, along with normative mean data from Lehiste and Peterson (1959) plotted as in Fig. 2. For all four subjects, pre-activation, open vowels, had higher SPL values than more closed vowels, presumably due to more efficient radiation of sound from the open oral cavity, as in hearing speakers. Post-activation this pattern changed minimally for MB, somewhat for MA because of a decrease in SPL for the vowel /e/, and markedly for FA and FB. For FA, the post-activation pattern of openness-related SPL is the inverse of the normative pattern, and for FB post-activation, the vowels /æ/, /o/, and /u/ have levels that are in the same range as /i/ and /u/, about 5 dB lower than the levels for /I/, /œ/, /n/, and /u/. These changes away from the normative pattern observed for FA and FB are discussed.
further below in relation to changes in $H1-H2$ (Sec. III C 2).

It is worth noting that FA's SPL decrease of 14.1 dB was accompanied by an $F0$ decrease of 101 Hz (35%), while MB's similar decrease in SPL of 11.7 dB was accompanied by a notably smaller $F0$ decrease, 21 Hz (13%). We hypothesize that MB's relatively small decrease in $F0$ is mainly a by-product of reducing SPL via a large decrease in subglottal pressure, whereas FB's large $F0$ decrease was probably the combined effect of a change in subglottal pressure and active control of vocal fold tension.

In summary, SPL decreased post-activation for all subjects. The normal height-related pattern deteriorated markedly for the two female subjects.

$H1-H2$. Values of $H1-H2$ derived from Klatt and Klatt's (1990) data on normally hearing speakers average $-4.1$ dB for 10 females (range: $-7.6$ to 1.1 dB) and $-9.8$ dB for 6 males (range: $-11.4$ to $-6.3$ dB). (Our computation of $H1-H2$ is from spectra with 6 dB per octave of preemphasis, and Klatt and Klatt add 10 to all their $H1-H2$ values; therefore, to compare $H1-H2$ data, we subtracted 16 dB from Klatt and Klatt's values.) Holmberg et al. (in press) found somewhat similar (but larger) ranges for groups of 15 female and 15 male speakers. Our subjects showed some differences of $H1-H2$ values among individual vowels; for this reason and because of different utterance materials, comparisons with other data must be made with caution. Reference to Table II shows that, pre- and post-activation, the $H1-H2$ mean was essentially within the Klatt and Klatt range for FA and FB, and it was above the Klatt and Klatt range for MA and MB. Pre- to post-activation, values increased for FA, FB, and MB and they decreased for MA. These results are discussed further below, in relation to changes in other parameters. The distributions of $H1-H2$ over the vowel set pre- and post-activation are not plotted here, since the measurable set was severely reduced by the restriction that $H2$ cannot be within 150 Hz of $F1$ (see Sec. II C).

Summary of parameter changes. Table III summarizes the major pre- to post-activation changes by subject (each pair of columns) and parameter (row). The left-hand column within each pair shows average parameter change, and the right-hand column indicates change in pattern of contrast between classes of vowels or congruence with respect to the normative pattern. A "+$" indicates an increase and a "$-$", a decrease. Parentheses enclosing an entry indicate that the effect was expressed weakly. Post-activation, FA and FB enhanced congruence of $F2$ with respect to the normative pattern for front vowels, retained an exaggerated long-short contrast and decreased congruence of SPL with respect to the pattern of normative mean values. Most of MA and MB's changes were: away from normative mean values, in the direction of decreasing contrast or congruence with normative patterns; or, an abnormal pre-activation pattern persisted. Thus, there were pre-to post-activation changes in average values for duration (rate), SPL and $F0$; and in cross-vowel patterning for $F1$, $F2$, duration, $F0$, and SPL. Some of these changes were toward normative mean values for FA and FB, but not for MA and MB.

### B. Longitudinal trends

In this section, temporal trends are examined in plots of parameter values averaged across vowels within each session, primarily to determine whether the parameters fall into different groups, based on the timing of changes. In addition, trends in the subjects' vowel identification scores on a perceptual test are examined for relations to changes in produc-

<table>
<thead>
<tr>
<th>Parameter</th>
<th>FA pattern</th>
<th>FB pattern</th>
<th>MA pattern</th>
<th>MB pattern</th>
</tr>
</thead>
<tbody>
<tr>
<td>Avg</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$F1$ range</td>
<td></td>
<td></td>
<td></td>
<td>below normal, pre and post diminished,</td>
</tr>
<tr>
<td>$F2$ congruence with normal</td>
<td>$+$</td>
<td>$+$</td>
<td>(+)</td>
<td></td>
</tr>
<tr>
<td>Dur tense-lax/long-short distinction</td>
<td>exaggerated</td>
<td>pre and post</td>
<td>pre and post</td>
<td>pre and post</td>
</tr>
<tr>
<td>$F0$ Congruence with normal pattern of &quot;intrinsic $F0$&quot;</td>
<td>(exaggerated pre, less so post)</td>
<td>(exaggerated pre, less so post)</td>
<td>exaggerated pre, less so post</td>
<td></td>
</tr>
<tr>
<td>SPL Congruence with normal pattern of &quot;openness&quot;</td>
<td>(congruent pre and post)</td>
<td>(congruent pre and post)</td>
<td>(congruent pre and post)</td>
<td></td>
</tr>
<tr>
<td>$H1-H2$</td>
<td>$+$</td>
<td>+</td>
<td>$-$</td>
<td>$+$</td>
</tr>
</tbody>
</table>
Figure 5 shows parameter trends by subject (column) and parameter (row). In the first six rows, each panel shows session means (averaged across vowels and repetitions) versus time re activation of the speech processor (in weeks). From top to bottom, the rows correspond to: $F_1$ (Hz—row 1), $F_2$ (Hz—row 2), duration (ms—row 3), $F_0$ (Hz—row 4), SPL (dB—row 5), and $H1-H2$ (dB—row 6). Each point in rows 1–6 represents the mean value from three repetitions of nine vowels. The vertical bars in rows 3–6 represent ±1 standard error around the mean. The asterisks denote significant pre- to post-activation differences ($p < 0.05$) according to a repeated measures ANOVA. Row 7 shows mean air flow data (mL/s) measured with inductive plethysmography from readings of a passage made during the same recording sessions. In each panel, the vertical line indicates the time of processor activation.

Figure 5 shows parameter trends by subject (column) and parameter (row). Separate plots of parameter values versus time for each individual vowel (not shown) evidenced trends that were (with some exceptions) similar across vowels; the averaged parameter trends were generally
representative of those for individual vowels within subjects. The vertical lines indicate the time of processor activation. In the first six rows, each panel shows session means (averaged across vowels and repetitions) versus time relative to activation of the speech processor (in weeks). From top to bottom, the rows correspond to: F1 (Hz—row 1), F2 (Hz—row 2), duration (ms—row 3), F0 (Hz—row 4), SPL (dB—row 5), and H1-H2 (dB—row 6). The vertical bars in rows 3–6 represent ±1 standard error around the mean. Error bars are not shown for average F1 and F2 because of the inherently large differences of these parameters across vowels. Error bars are present for SPL, but some of them are small enough to be obscured by the closed circles. The asterisks denote significant changes from pre- to post-activation conditions.

Row 7 shows mean air flow data (ml/s) derived from lung volume versus time, recorded from readings of a passage made during the same recording sessions (Lane et al., 1991). Lung volumes were measured with an inductive plethysmograph that transduced the cross-sectional areas of the speaker’s chest and abdomen. Isovolume maneuvers at tidal-end respiration level and inflation of a bag of known volume allowed summation and calibration of the plethysmograph signals. Lung volumes at the initiation and termination of the speakers’ expiratory limbs were derived from tracings of calibrated lung volume displayed by computer. The difference between the two, divided by the duration of the limb, yielded an estimate of mean airflow in ml/s during the limb. The changes in flow shown in row 7 were in the direction of normal for FA and FB; this could not be determined for MA and MB (see Lane et al., 1991, for details).

1. Trends in vowel acoustic parameters and average air flow

In order to compare the timing of parameter changes with one another, we used the following procedure to determine the time when each parameter arrived within a heuristically defined “post-activation range,” viz, when it attained a value in the range between the last two post-activation session means. For each subject, when an acoustic parameter changed pre- to post-activation (p < 0.05), we generated a plot of mean values versus time corresponding to a panel in Fig. 5, and fit the data with a smooth curve such that the influence of neighboring points decreases exponentially with distance (McLain, 1974). We considered the parameter to have changed when its smoothed curve entered the post-activation range. (Admittedly, the degree of smoothing can influence these determinations.) Since MB showed no perceptual benefit, no significant change in two of the acoustic parameters and great variability pre-activation in two others (SPL and H1-H2), his trends have not been included in this analysis. Table IV shows the results of these determinations. The rows correspond to the parameters. For each subject, there are two columns; the left column lists the approximate time of change (in weeks), and the right column lists a classification, “S” for slowly changing, “R” for rapidly changing. As it turned out, each subject's data were sufficiently bimodal to allow this categorizing unequivocally.

Examination of Table IV leads to the following observations. A rapid F1 change accompanied a rapid duration change for MA, and a slow F1 change accompanied a slow duration change for FB. Mean F2 change was rapid (FA and FB). SPL changed slowly (FA, FB, and MA). F0 changed significantly only for FB; that change was relatively slow. In all three cases, H1-H2 and flow change together: slow and increasing for FA and FB, rapid and decreasing for MA.

Several additional trends are noteworthy in Fig. 5. Except for MA, whose SPL change was the smallest, the trends for SPL and H1-H2 approximately mirrored one another: As SPL dropped, H1-H2 increased. The trends for SPL and F0 were similar to one another for all four subjects, regardless of whether change was significant. Thus, if MA's relatively small SPL change is excluded, there was a general tendency for SPL, H1-H2, flow and F0 to group together.

For MB, the relative time courses of H1-H2 (Fig. 5, row 6) and average air flow (Fig. 5, row 7) evidence a relation which differs from the other three subjects. In the first post-activation session at seven weeks, mean H1-H2 had increased by about 3 dB, indicating a more spread glottis, while paradoxically, average airflow dropped by about 100 ml/s to a very low value of 64 ml/s. [The lowest value in a study by Horii and Cooke (1978) of normally hearing male subjects was 143 ml/s.] MB's longitudinal trends must be interpreted with caution, however, since 3 of the 7 parameters show the trends between the two pre-activation sessions.

Since MB's relation between flow and H1-H2 appeared to be paradoxical, we questioned the validity of comparing his vowel parameters from /hVd/ citation pronunciations with flow measures from a reading passage. We checked the correspondence between vowel parameters and flow measures for MB by extracting acoustic parameters from ten vowel nuclei selected from each of the same three readings of the Rainbow Passage used for the flow measurements (N = 30). The trends in acoustic parameters were nearly identical to those in Fig. 5 (last column) for MB. Somewhat exceptional were the formant frequencies (probably due a different balance of vowel type and context effects between the two samples) and the overall magnitude of durations (due to differences in context). This result helps to validate our approach of comparing the /hVd/ vowel parameters with average flow values from the reading passage for all

<table>
<thead>
<tr>
<th>Parameter</th>
<th>FA</th>
<th>FB</th>
<th>MA</th>
</tr>
</thead>
<tbody>
<tr>
<td>F1</td>
<td>n.s.</td>
<td>29</td>
<td>S</td>
</tr>
<tr>
<td>F2</td>
<td>5</td>
<td>R</td>
<td>7</td>
</tr>
<tr>
<td>Dur</td>
<td>n.s.</td>
<td>24</td>
<td>S</td>
</tr>
<tr>
<td>F0</td>
<td>18</td>
<td>S</td>
<td>n.s.</td>
</tr>
<tr>
<td>SPL</td>
<td>22</td>
<td>S</td>
<td>29</td>
</tr>
<tr>
<td>H1-H2</td>
<td>36</td>
<td>S</td>
<td>23</td>
</tr>
<tr>
<td>Flow</td>
<td>58</td>
<td>S</td>
<td>16</td>
</tr>
</tbody>
</table>

TABLE IV. Times (in weeks) of change, when parameters reached a final “post-activation range.” The rows correspond to the parameters. For each subject, the left column lists the approximate time of change, and the right column lists a classification, “S” for slowly changing, “R” for rapidly changing. The abbreviation n.s. indicates that the change was not significant based on the ANOVA.
subjects, and it strongly reinforces the observation that MB's relation between flow and $H_{1}-H_{2}$ was different from the other three subjects. We will speculate below that the apparently paradoxical flow decrease with a spread glottis was due to a large drop in subglottal air pressure, which is consistent with the low flow reading (and the $F_0$ decrease).

2. Comparison with vowel identification measures

The first row of Fig. 6 reproduces the longitudinal plots for $F_2$ shown in Fig. 5 to facilitate comparison with trends in accuracy of vowel identification following activation of the processor.

Row 2 of Fig. 6 shows the results of a four-vowel, forced-choice test of vowel identification (Boothroyd, 1985), which was administered at intervals following processor activation. The vertical axes in row 2 represent percent correct; chance performance is 25%.

Subject FB had the highest identification scores and evidenced some slight improvement between the first score (at 3 weeks post-activation) and subsequent ones (beginning at 26 weeks). MB consistently performed at or below chance throughout the testing. MA showed some slight fluctuation over time, in the range of 55% to 60% correct. (An additional score was obtained from him beyond the 78 week limit of his plot; that value was 60% correct, close to the value of 63% at 55 weeks.) FA began performing at near chance level immediately post-activation, and over the course of 30 weeks improved her performance to a level of better than 45% correct, which she maintained for the remainder of the tests.

3. Summary of longitudinal observations

Because of the limited number of subjects, the variability in their trends, and because the classification of variables into rapidly and slowly changing groups depends on the graphical methods used, the following generalizations are tentative. $H_{1}-H_{2}$ and flow change together and there is a tendency for SPL and $F_0$ to follow their time course. SPL and $H_{1}-H_{2}$ trends are complementary. Duration and $F_1$ also covary. Where vowel identification improved substantially, it was accompanied by a decrease in $F_2$, reflecting realignment of the vowel space toward normative values.

C. Relations among parameters and underlying mechanisms

In the previous two sections, we observed changes pre- to post-activation in averages of the measured parameters; changes in patterns of contrast among vowels; and differences in the timing of changes for groups of interrelated parameters. When processor activation causes a speaker to make physiological adjustments to the speech production mechanism, the consequences for acoustic speech parameters are complex because one active articulatory change (e.g., shorter syllables) often entrains others (e.g., less open vowels). If the speaker simultaneously seeks to maximize phonemic contrasts that she or he can now discriminate aurally (e.g., short versus long vowels), the implementation of this phonemic goal will be superimposed on the other active
and passive articulatory changes. In this section, we undertake to relate the patterns of covariation in parameters observed with our speakers to a model of mechanical and acoustical linkages among articulators and cavities.

We looked at two different kinds of parameter relations—within condition and longitudinal. To the extent that individual token values of parameters covary among vowels within a condition, that covariation may suggest a mediating functional mechanism, or that there are additional mediating parameters. To explore such parameter covariation, we performed pairwise correlations on parameter values. To explore how evolving postural changes of the underlying mechanisms interact with one another, we can reasonably hypothesize some mediating mechanism.

We hypothesize that there are two sets of mechanisms.

### Table V. Correlations between pairs of acoustic parameters.

Each row corresponds to a parameter pair for which patterns of significant correlations were found among the subjects. There are three columns for each subject. The first column indicates the results of the within-condition pre- and post-activation correlations. The second and third columns indicate the results of the within-condition pre- and post-activation correlations. Each within-condition correlation was calculated from mean values for the vowels, averaged over six tokens of each vowel. The numbers in the cells are values of significant product-moment correlations. An empty cell indicates that there was no significant correlation.

<table>
<thead>
<tr>
<th>Parameter pair</th>
<th>FA</th>
<th>FB</th>
<th>MA</th>
<th>MB</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) DUR, F1</td>
<td>0.38 0.70</td>
<td>0.32</td>
<td>0.25</td>
<td>0.17</td>
</tr>
<tr>
<td>(2) F1, F0</td>
<td>-0.28 -0.90</td>
<td>-0.86 -0.52</td>
<td>-0.95 -0.81</td>
<td>-0.44 -0.78</td>
</tr>
<tr>
<td>(3) DUR, F0</td>
<td>-0.85 -0.62</td>
<td>-0.38</td>
<td>-0.61</td>
<td>-0.46</td>
</tr>
<tr>
<td>(4) H1-H2, SPL</td>
<td>-0.36</td>
<td>-0.82</td>
<td>-0.80</td>
<td>-0.91</td>
</tr>
<tr>
<td>(5) F1, SPL</td>
<td>0.84</td>
<td>-0.87</td>
<td>0.23</td>
<td>0.65</td>
</tr>
<tr>
<td>(6) H1-H2, FLOW</td>
<td>0.52</td>
<td>-0.68</td>
<td>0.81</td>
<td>0.53</td>
</tr>
<tr>
<td>(7) SPL, FLOW</td>
<td>-0.64</td>
<td>-0.68</td>
<td>0.81</td>
<td>0.53</td>
</tr>
</tbody>
</table>

![FIG. 7. A schematic diagram which illustrates hypothesized functional mechanisms and their effects on relations among the measured parameters.](http://example.com/figure7.png)

---

that account for many of the relations shown in Table V. Figure 7 is a schematic diagram which illustrates those mechanisms and their interactions with the measured parameters. The purpose of the diagram is to help interpret relations among the average and contrastive parameter changes made by the subjects, possible underlying mechanisms, and the correlation results.

The left column in Fig. 7 (at levels A, C, and D) indicates average acoustic parameters which the subjects appear to have changed in an overall purposeful way across sessions (in different combinations, depending on the subject). Arrows point from those changes toward implementing physiological mechanisms. The top-most parameter, speaking rate, was increased by three of the subjects; speaking rate obviously interacts with durations. The subjects also made average changes in the other two parameters in the left column, F0 and SPL. (SPL was presumably changed mainly with modifications of subglottal pressure and glottal aperture, as indicated in the figure.) The postural changes underlying these average parameter settings can dominate the behavior of mechanisms shown in the center column, depending on how strongly the change is expressed.

The text in each box in the center column indicates, for each physiological mechanism, an associated acoustic parameter (in italics), and any associated contrast (in brackets). The active expression of any of the contrasts (between classes or groups of vowels) can influence the behavior of a mechanism. The arrows connecting pairs of boxes imply an influence of a mechanism on the one below (or next to) it. A "+" sign next to an arrow indicates that the influence causes an increase. (The parentheses enclosing two of the signs and the dashed line connecting the two boxes at levels D and E indicate that the influence works in opposition to the change described within the target box.) Thus the center column describes two hypothetical sets of (physically mediated) relations among physiological and acoustic parameters (rows A–C and D–E, respectively).

In the right-hand column of Fig. 7 are listed parameter pairs with signs of the correlations that are predicted by the relations shown in the middle column. Lines connect pairs of mechanisms (boxes) with their respective predicted correlations. (Two additional parameter pairs, "H1–H2, flow" and "SPL, flow," are shown at the bottom of the figure.) Numbers in parentheses below the predictions correspond to rows in Table V.

The diagram in Fig. 7 reflects the strong hypothesis that almost all of the observed parameter changes following processor activation in this research are due to changes in the postural settings of physiological mechanisms regulating speaking rate, F0 and SPL. Those three parameters should be readily perceived by cochlear implant patients in their own speech and in the speech of others without resorting to spectral place information. According to the diagram, changes in glottal aperture are made mostly to assist in implementing SPL changes, and not to adjust voice quality in response to perceived spectral imbalance (i.e., an abnormal degree of breathiness). Changes in F1 are a consequence of rate changes; since those changes compress the F1 range, they are unlikely to have been made actively in response to perceived F1 abnormality. We hypothesize that the only spectrally specific changes we have observed are in F2, as evidenced mainly by FA, to a lesser extent by FB and perhaps by MA. Those changes (and small tongue posture changes underlying average F2 differences) are not represented in the diagram. We interpret the retention of the exaggerated tense–lax durational contrast as evidence that the spectral information being delivered by the prosthesis did not provide sufficient vowel space differentiation for the subjects to abandon the exaggerated length contrast in their vowel productions.

In the remainder of this section, the hypotheses represented in Fig. 7 are explained in detail, and the results shown in Table V are interpreted with respect to those hypotheses.

1. Relations among duration, F1 and F0

Inference from the data suggests that there are mechanically mediated relations among duration, F1, and F0 which can affect intrinsic F0 and durational differences among vowels. Those relations can be influenced, in turn, by mechanisms underlying changes in speaking rate and average F0, and by expression of the contrast between long and short vowels.

a. Duration and F1. An increase in speaking rate, (level A in Fig. 7) should induce a decrease in vowel nucleus durations. Our findings in this study of decreased vowel durations are consistent with other results showing relatively low speaking rates for deaf speakers (Lane and Webster, 1991; Plant, 1984; Leder et al., 1987) and rate increases after stimulation with a single-channel implant (Plant and Oster, 1986).

Decreased durations should allow less time for tongue body lowering for low vowels, leading to an average increase in their tongue height and a decrease in their values of F1 (level B, Fig. 7) (cf. Lindblom, 1963). This mechanism implies a prediction of a positive longitudinal correlation between duration and F1 (row 1 in Table V), and there were significant, weak longitudinal correlations between duration and F1 for all four subjects. The expression of the long–short contrast (in which each length category, long and short, includes both high and low vowels) should reduce the positive correlation between duration and F1 within sessions, but not between them. With one exception, there were no significant within-session correlations of duration and F1. The exception was a pre-activation positive correlation for FA. Her duration data in Fig. 3 show that, pre-activation, the vowels /ı/ and /u/, with low F1, group abnormally with the short vowels, thus establishing the positive correlation of F1 and duration.

b. F1 and F0. An increase in tongue height may lead to increased tension on the vocal folds because of anterior "tongue pull," i.e., contraction of the posterior genioglossus which is transmitted to the thyroid cartilage, with a resulting increase in F0 (cf. Honda, 1983—level C in Fig. 7). This mechanism could result in negative longitudinal and within-condition correlations between F1 and F0. Indeed, significant negative correlations were found within all pre- and post-activation conditions as well as across sessions for all four subjects, with the single exception of the post-activation...
condition for MA (Table V, row 2). For all four subjects, the correlation was slightly weaker post-activation than pre-activation, and it was weaker for the males than for the females. (The finding that MA did not have a significant post-activation correlation of F1 and F0 may be due to his marked post-activation compression of the range of F1—see above).

These within-condition results, along with our observations about the height-related pattern of intrinsic F0, suggest the following account. Without auditory input, the observed exaggerated pattern of intrinsic F0 is due largely to tongue pull, which is greater for high vowels than low vowels (cf. Honda, 1983). Following processor activation, a number of changes took place (discussed further below); one effect of those changes was to reduce exaggerated intrinsic F0 differences among vowels, i.e., to interfere somewhat with the tongue-pull mechanism. This explanation does not preclude the possibility of some active regulation to moderate F0 differences among the vowels: In order to enhance the perception of height distinctions, speakers may raise F0 for high vowels to minimize the distance between F0 and F1 (Trannmuller, 1981) with active use of the cricothyroid muscles (cf. Vilkman et al., 1989). Thus active control when hearing is available may be superimposed on the mechanical infrastructure of the tongue-pull mechanism.

Taking F1 as an indirect index of tongue pull, we interpret the correlation coefficients in row 2, Table V as showing that the effect on F0 of varying tongue pull across sessions is weaker than the effect of varying tongue pull across vowels within a session. In the longitudinal correlations, the changes in average F1 were induced by rate changes; in the within-condition correlations, by differences in vowel height.

c. Duration and F0 (Levels A-C in Fig. 7; row 3 in Table V). The data of Peterson and Lehiste (1960) in Fig. 3 show that low, more-open vowels are longer than high vowels. The combination of the above hypotheses about the relation between duration and tongue height (levels A–B) and tongue height and F0 (levels B–C) leads to a prediction of a negative correlation between duration and F0: in contrast to high vowels, the production of low, more open, longer vowels should exert less tension on the vocal folds (via less anterior pull on the thyroid cartilage) and a resulting lower F0. Because this effect spans two mechanisms, it should be expressed weakly. FA, FB, and MA indeed had negative post-activation correlations of duration and F0, and FA had a negative pre-activation correlation as well. FB and MA also had weak negative longitudinal correlations between F0 and duration, reinforcing the notion that tongue-pull may have an influence on average F0. FA’s lack of a negative longitudinal correlation may be related to her large, purposeful, long-vowel pre-supralaryngeal pull on the thyroid cartilage) and a resulting lower F0. Because this effect spans two mechanisms, it should be expressed weakly. FA, FB, and MA indeed had negative post-activation correlations of duration and F0, and FA had a negative pre-activation correlation as well. FB and MA also had weak negative longitudinal correlations between F0 and duration, reinforcing the notion that tongue-pull may have an influence on average F0. FA’s lack of a negative longitudinal correlation may be related to her large, purposeful, long-vowel pull on the thyroid cartilage) and a resulting lower F0. Because this effect spans two mechanisms, it should be expressed weakly. FA, FB, and MA indeed had negative post-activation correlations of duration and F0, and FA had a negative pre-activation correlation as well. FB and MA also had weak negative longitudinal correlations between F0 and duration, reinforcing the notion that tongue-pull may have an influence on average F0. FA’s lack of a negative longitudinal correlation may be related to her large, purposeful, long-vowel pull on the thyroid cartilage) and a resulting lower F0. Because this effect spans two mechanisms, it should be expressed weakly. FA, FB, and MA indeed had negative post-activation correlations of duration and F0, and FA had a negative pre-activation correlation as well. FB and MA also had weak negative longitudinal correlations between F0 and duration, reinforcing the notion that tongue-pull may have an influence on average F0. FA’s lack of a negative longitudinal correlation may be related to her large, purposeful, long-vowel pull on the thyroid cartilage) and a resulting lower F0. Because this effect spans two mechanisms, it should be expressed weakly. FA, FB, and MA indeed had negative post-activation correlations of duration and F0, and FB also had a post-activation negative correlation between H1–H2 and SPL. These findings lead us to speculate that the mechanically mediated relation between vowel height and F0 (above) may also include some effect on SPL: increased tension on the vocal folds may tend to adduct them somewhat, increasing SPL slightly. The correlation between H1–H2 and SPL diminishes when F0 is actively controlled, as appears to be the case for FA who showed significant reduction in F0 and nonsignificant within-condition correlations of H1–H2 and SPL. FB and MA, on the other hand, did not change F0 pre- to post-activation; FB shows the expected strong negative post-activation correlation (although MA does not).

The evidence that decreases in SPL were associated with increases in glottal aperture suggests an explanation of the inversion of the normal relation between vowels “openness” and SPL that occurred with FA and FB post-activation (Sec. III A, above). Recall that with the processor in use, FA’s more open vowels had about the same or lower SPL than /i/ and /u/; FB showed this effect, but selectively, for /æ/ and /æ/, and /ı/ only (see Fig. 4). Furthermore, FA and FB both had post-activation increases in H1–H2 and average air flow, consistent with an increase in glottal aperture. SPL values for vowels are usually dominated by the amplitude of the F1 prominence in the spectrum. Increased glottal acoustic resistance (from a larger glottal aperture) generally would result in an increase in the bandwidth of F1 and a reduction in its amplitude. Due to a source-filter interaction effect, these changes would be greater for the more open vowels than for /i/ or /u/ (Nord et al., 1984). For FB, there may have been additional F1 attenuation from substantial coupling to a subglottal (tracheal) resonance which occurs who showed the most extreme contrast between long and short vowel durations, which he alone had both pre- and post-activation (see Fig. 2 above).

2. Relations involving H1–H2, SPL, and flow

a. H1–H2 and SPL (Table V, row 4). A spread glottis (indicated by higher values of H1–H2) results in a more sinusoidal glottal airflow waveform with a spectrum that has decreased energy at high frequencies and increased formant bandwidths. There may be increased coupling to subglottal resonances, with an accompanying damping of F1 (Fig. 7, level E). Thus we expect a negative correlation between SPL and H1–H2. There were negative longitudinal correlations of H1–H2 and SPL for all four subjects. FA, FB, and MB presumably increased glottal aperture, along with decreasing subglottal pressure, in order to achieve their significant SPL decrements pre- to post-activation. MA’s longitudinal correlation between these two parameters was the weakest among the four subjects. In comparison to the other three, his net SPL decrement was much smaller and he was the only one to decrease airflow pre- to post-activation.

A within-condition negative correlation of H1–H2 and SPL suggests that cross-vowel SPL differences might be mediated in part by differences in glottal aperture (which may be in turn influenced by height-related tongue pull, as implied by the chain of mechanisms in Fig. 7). FB and MA each had a negative pre-activation correlation and FB also had a post-activation negative correlation between H1–H2 and SPL. These findings lead us to speculate that the mechanically mediated relation between vowel height and F0 (above) may also include some effect on SPL: increased tension on the vocal folds may tend to adduct them somewhat, increasing SPL slightly. The correlation between H1–H2 and SPL diminishes when F0 is actively controlled, as appears to be the case for FA who showed significant reduction in F0 and nonsignificant within-condition correlations of H1–H2 and SPL. FB and MA, on the other hand, did not change F0 pre- to post-activation; FB shows the expected strong negative post-activation correlation (although MA does not).
in the same frequency region as F1 for her low vowels. Klatt and Klatt (1990) report a median value of 750 Hz for the first tracheal resonance of five females, with one having a value of 800 Hz.

Figure 8 shows LPC and DFT spectra for pre- and post-activation tokens of the vowels /α/ and /i/ for subject FB. The spectra for /α/ show a considerable post-activation decrease in the prominence of the F1 peak in relation to the amplitudes of higher formants; such an increase in damping is not so apparent for /i/. Examination of FB's post-activation F1 values in Fig. 1 reveals that /α/, /α/, and /α/ alone have values in the region around 800 Hz. Thus the selective attenuation of these three vowels may be mainly from coupling to the first tracheal resonance which acts like an antiresonance in the vocal-tract transfer function at about 800 Hz. On the other hand, comparable spectra for FA did not reveal a relative reduction in the amplitude of F1 for /α/, and her post-activation F1 values for /α/ were in the region of 1 kHz. We speculate that FA's post-activation inversion of the normal relation between vowel "openness" and SPL was due predominantly to F1 damping from increased glottal acoustic resistance.

b. F1 and SPL (levels B-E in Fig. 7; row 5 in Table V). The relation between vowel height and vowel SPL is well known and is illustrated with the data of Lehiste and Peterson (1959) in Fig. 3. The lower the vowel, the greater the SPL, which implies a positive correlation between F1 and SPL. This expected outcome was observed in both longitudinal and within-condition correlations. Both male subjects had positive pre- and post-activation correlations between F1 and SPL, as did the females pre-activation. On the other hand, FA had a negative post-activation correlation and FB had no significant correlation post-activation. These results are consistent with the observation made above that post-activation, both female subjects produced the more open vowels with intensity levels that were abnormally low in relation to levels for the high vowels; whereas the male subjects had more "normal" patterns both pre- and post-activation.
3. Summary and implications of relations among the parameters

The correlation results support the overview schematized in Fig. 7. There is a set of mechanically mediated effects among vowel durations, tongue height, and tension on the vocal folds; and a set of relations among subglottal pressure, glottal aperture, airflow, and coupling to subglottal resonances. The former set of effects can influence longitudinal and cross-vowel relations among duration, $F_1$, $F_0$; the latter can influence relations between $H_1$--$H_2$ and SPL. The two sets may be linked, but more work is needed to explore such a linkage. Relations between pairs of parameters can also be influenced by segment-specific control over vowel length (to enhance the tense--lax contrast). We speculate that the observed diminution of exaggerated intrinsic $F_0$ differences was due to interference with the underlying tongue-pull mechanism by direct control over the laryngeal configuration in changing the postural setting of $F_0$ and/or SPL. Implant-induced changes in the postural settings underlying speaking rate, average $F_0$ and SPL can also influence other relations among the parameters, not only longitudinally, but across vowels, within conditions.

IV. DISCUSSION

There were pre-activation abnormalities of average parameter values and patterns of contrast. For several reasons (including a paucity of directly comparable normative data and variation in the normative data that can be cited), it is difficult to quantify the degree of abnormality of the measured parameters; however, a number of observations support the hypothesis that there were abnormalities in the pre-implant speech of the subjects. FB had low average airflow, low $H_1$--$H_2$ and exaggerated patterns of intrinsic $F_0$ and tense--lax durational differences. FB was similar to FB pre-activation, with the addition of an abnormally high average $F_0$, an irregular pattern of $F_2$ for the front vowels and a somewhat aberrant cross-vowel pattern of durations. MA had an abnormally high average $F_0$, somewhat irregular patterns of intrinsic $F_0$ and $F_2$, and an aberrant pattern of durations. MB had a high average $F_0$, a compressed range of $F_1$, a compressed range and somewhat irregular pattern of $F_2$, and exaggerated patterns of intrinsic $F_0$ and tense-lax durational contrast. The $F_0$ values for MA and MB are consistent with the report of Leder et al. (1987) that postlingually deafened men with profound sensorineural losses bilaterally had a higher fundamental frequency when reading than normally hearing age-matched men. Although formal tests of intelligibility have not yet been conducted, we observed that the speech of our four subjects sounded abnormal; the two women were nevertheless quite intelligible, while the two men, with earlier hearing losses, were much less intelligible.

These observations give the impression that the relatively poor intelligibility of the speech of the male subjects was reflected in a higher degree of abnormality of the measured parameters. This possible difference in degree of abnormality may be due to age at onset of a linguistically limiting degree of hearing loss (see Cowie and Douglas-Cowie, 1983; Cowie et al., 1988), although additional measures on a larger number of subjects would be needed to support such a suggestion.

Our data only enable us to begin exploring relations between perceptual gains and changes in production, and the initial picture that emerges is complex. The changes made in average duration (decreased in all subjects, significantly in two of the four), $F_0$ (decreased in FA) and SPL (decreased in all subjects) suggest that the subjects were attempting to correct "discrepancies" that they detected between these characteristics of their own speech and the speech of those around them. It is noteworthy that MB changed all three of these parameters toward normative mean values, in spite of receiving no benefit, as measured by vowel identification tests. These three parameters should be easily perceived with relatively undifferentiated input from a cochlear prosthesis, without the use of spectral place information. Following the reasoning in the previous section, we speculate that changes in average values of $F_1$ follow from durational changes. Changes in $H_1$--$H_2$ and flow were most likely made in the course of implementing the observed changes in $F_0$ and SPL.

Such changes were least pronounced for MA. His decrease in average duration served to compress the range of an initially aberrant pattern of durations among vowels, and he did not change average $F_0$ in spite of a relatively high pre-activation value. The good perceptual benefit MA received from his implant might have been responsible for his slight $F_2$ normalization; however, most characteristics of his vowel production did not seem to improve and might have deteriorated somewhat. The same conclusion can be drawn about MB's speech production changes, although such a result is not surprising in view of his lack of perceptual benefit and eventual nonuse of the implant.

For subject FA, an information transfer analysis showed good results for reception of $F_2$ information: over her three most recent tests, the average information received was 75.2% for $F_2$ (Rabinowitz et al., 1991). The measure was obtained as follows: Eight vowels uttered by one male speaker three times each were presented on 720 trials in all for identification without feedback. The confusion matrix was collapsed to a single binary contrast—high versus low $F_2$ (the range was from 0.92 to 1.9 kHz; note that the crossover frequency between channels 2 and 3 of the implant was 1.4 kHz). An information transmission of 75% corresponds to an error rate for confusions across the two groups of (four) high- and low-$F_2$ vowels of less than 5%. This finding reinforces our observation that there was a relation between perceptual gain from the implant and the normalization of $F_2$ values for FA. FA's realignment of $F_2$ provides the most convincing evidence in this study of specific changes in speech production that might be related to gains in perception from the use of spectral place information and that could also enhance vowel contrasts. Like FA, the subjects of Svirsky and Tobey (1991) produced /I/ and /e/ more distinctly from one another when their (22 channel) speech processors were activated. Since her pre-implant speech was noticeably loud and harsh, FA's post-activation decrements in average $F_0$ and SPL seemed to enhance the subjective quality of her voice as well. Her $F_0$ decrease is consistent
with the results of several studies that find normalization of $F_0$ following electrical stimulation of the cochlea (Fourcin et al., 1983; Ball and Ison, 1984; Ball and Faulkner, 1989; Engleman et al., 1981; Oster, 1987; Plant and Oster, 1986).

Like MA and MB, the subject of Tartter et al. (1989) was deafened in childhood. Following activation of her multichannel implant at age sixteen, she had lower $F_1$, shorter stressed syllables and lower $F_0$, all consistent with changes observed in our subjects and the view schematized in Fig. 7. She also had a paradoxical reduction in her acoustic vowel space, consistent with the differences among our subjects in suggesting that age at deafening is an important factor in governing speech production responses to stimulation from a cochlear implant.

Thus we speculate that production gains are governed at least as much by prior linguistic experience as by perceptual gains. One interpretation of our results is that the four subjects were using their newly acquired auditory feedback to recalibrate pre-existing but somewhat degraded internal models of the relation between their production routines and the resulting acoustics. In the case of MA and MB, the model was never fully developed in the first place, so their post-activation speech could not readily improve with respect to normative mean values. For them, post-activation normalization of their speech patterns might have been somewhat like learning a second language in adulthood with relatively impoverished (or for MB, virtually useless) auditory feedback. The fact that MB’s changed speech patterns persisted after a year of nonuse of the prosthesis supports the idea that such motor patterns (including relatively unproductive ones) are largely independent of auditory feedback for their moment-to-moment function.

A major finding of this study is the large degree of interdependency among production parameters. Vowel production changes are likely to be influenced not only by perceptual gains and age at deafening, but also by: (1) mechanically mediated relations among durations, $F_1$ and $F_0$; (2) aerodynamically and acoustically mediated relations among flow, SPL and $H_1$--$H_2$; (3) parameter interactions with postural changes underlying speaking rate, average $F_0$ and SPL; and (4) strategies for enhancing or maintaining specific contrasts, e.g., exaggerating durational differences between tense and lax vowels.

We also found that there could be differences in the relative timing of changes in different parameters. There appears to be separate control over changes in respiration and laryngeal postural mechanisms on one hand, and duration and formants on the other. The relative rates of change of those two groups of parameters can differ among subjects, but our data are still too sparse to provide meaningful information about the factors that may govern those rates.

V. CONCLUSIONS

There were trends toward normative patterns among the two female subjects, who were deafened in adulthood, but generally not among the two males, who were deafened in childhood. The two female subjects have less pressed voices ($H_1$--$H_2$ is higher); their flow is more normal; their formant spaces have better structure along the $F_2$ dimension; their rate is up and their level down. To some extent their more normal speech may be due to their newfound ability to hear others and to compare the (distorted) sound of their own speech with the (distorted) information about the speech of others, allowing them to validate articulatory routines so that their speech sounds more “right” to them.

Speakers respond differently to deafening, so a common tendency toward normalization requires them to respond differently to processor activation. The women had pressed voices, so more airflow and breathiness is a trend toward normal; MA had much higher rates of flow and breathiness, which fell with processor activation.

The male subjects’ early deafening and MB’s lack of perceptual benefit complicate the picture since the pre-deafening routines they are trying to recalibrate may be faulty in the first place, and MB’s post-activation stimulation was apparently of little use to him.

There are mechanical interactions among adjustments, and the adjustments may not always be accomplished in the best ways. Thus higher rate—a trend toward normal—may lead to compression of the vowel space—a trend away from normal. MA decreased segment duration but failed to make additional adjustments to restore duration contrasts between tense and lax vowels.

The general pattern of relationships among speech parameter values proved to be consistent with a model of speech production in cochlear implant patients which has three interacting components. First, patients discriminate relatively gross properties of the speech waveform such as rate, $F_0$ and SPL and make postural adjustments regulating the corresponding speech parameters. Second, those adjustments operate on a mechanism with complex interdependencies among its components, with the result that further speech changes (such as formant values and breathiness) are induced. Third, speakers may make adjustments in articulation in order to convey phonemic contrasts. These three forces acting in concert—postural adjustments, physical linkages, and phonemic adjustments—determine the empirical covariation of speech parameters during vowel production.

ACKNOWLEDGMENTS

We are grateful to Dr. William Rabinowitz, Research Laboratory of Electronics, Massachusetts Institute of Technology and Dr. Donald Eddington, Massachusetts Eye and Ear Infirmary, for providing data and analyses from perceptual tests, to Dr. Kenneth Stevens, Research Laboratory of Electronics, Massachusetts Institute of Technology and Dr. T. V. Ananthapadmanabha, Voice and Speech Systems, Malleswaram, Bangalore, India for many helpful suggestions, and to Dr. Vivien Tartter, City University of New York, Dr. Steven Leder, Veteran’s Administration Medical Center, West Haven, Conn., and Dr. James Mahshie, Galaudet University, for constructive reviews. This work was supported by N.I.D.C.D. Grant No. DC00361 to the Massachusetts Eye and Ear Infirmary and the Massachusetts Institute of Technology, Dr. Joseph B. Nadol, Principal Investigator.
To facilitate subject identification in this paper, they are called FA, FB, MA, and MB (for the two females and the two males). For cross reference to other publications from the Massachusetts Eye and Ear Infirmary that include data from these subjects, they are S09, S15, S11, and S08, respectively.

To obtain a signal from which a form of “open quotient” of the glottal vibration pattern could be derived, two electrodes of an electrogastrograph (EGG—Kay Elemetrics, Inc.) were placed on either side of the larynx and secured with an elastic band. However, a variable signal-to-noise ratio in the EGG signal prevented our obtaining very many valid measures of an EGG-based open quotient, especially from the female subjects. Therefore, open quotient is not discussed further here.


