ABSTRACT
A theoretical overview is presented. Phonemic information is transmitted by actions of neuro-muscular synergisms, which are organized to achieve articulatory and acoustic goals. Acoustic goals are achieved with the use of an internal model of relations between articulatory commands and the sound output. Auditory feedback is used to acquire and maintain the model, and to make situation-dependent adjustments in "postural" parameters underlying average sound level, rate and F0, which influence clarity and intelligibility. Supporting evidence concerning acoustic goals consists of findings of articulatory-to-acoustic motor equivalence. The hypothesized use of auditory feedback is illustrated by studies of patients who have experienced changes in hearing status.

Nous présentons un schéma-cadre théorique. L’information phonémique est transmise par le biais de synergies neuro-musculaires qui sont organisées par rapport à des objectifs articulatoires et acoustiques. La réalisation des objectifs acoustiques met en jeu un modèle interne des relations articulatoires-acoustiques. L’information auditive afférente est utilisée pour l’acquisition et la réactualisation de ce modèle, ainsi que pour permettre, selon les conditions de production, des ajustements de paramètres “posturaux” agissant sur l'intensité moyenne du signal, sur le débit et sur F0, qui modifient la clarté et l'intelligibilité de la parole. A l’appui de l’existence d’objectifs acoustiques nous présentons des données suggérant une équivalence motrice articulato-acoustique. Notre hypothèse sur l’utilisation du feedback auditif est illustrée par des études sur des patients qui ont vécu une évolution de leurs capacités auditives.

THEORETICAL OVERVIEW
We hypothesize that phonemic information in speech production is transmitted by the actions of neuro-muscular synergisms. The control of the synergisms is organized to achieve articulatory and acoustic goals. If speech motor control is based partly on acoustic goals, how is it accomplished? It is very unlikely that auditory feedback is used closed loop, moment-to-moment for the control of individual movements. The processing times for such functionality would probably be about the same as the duration of some brief movements, and many movements for vowels begin during relatively quiet or silent consonant intervals, when little is available in the way of audible cues about movement trajectories.

Thus, segmental speech movements toward acoustic goals are probably preprogrammed, based on “predictive simulation” with the use of an internal model of relations between articulatory commands and acoustic results [1]. Auditory feedback is used to help acquire and then maintain this internal model, which becomes resistant to change as speech motor patterns mature. Auditory feedback also has a secondary important role: to monitor changes in the acoustic environment. Such monitoring is used to help assure adequate intelligibility, which is accomplished by making adjustments in "postural settings" of articulatory mechanisms that underlie average sound level (SPL), speaking rate, F0, degree of clarity, and inflection of F0 and SPL.

SUPPORTING EVIDENCE
Acoustic goals
Preliminary evidence for the use of acoustic goals in speech motor programming comes from findings on “articulatory-to-acoustic motor equivalence” in production of the vowel /u/. The term “motor equivalence” refers to the observation that in multiple tries, the same goal is reached in more than one way. It is possible to produce approximately the same acoustic result with somewhat different area functions, at least for some sounds. We have tested the hypothesis that in multiple repetitions of the vowel /u/, there could be a trading relation between lip rounding and tongue-body raising, as evidenced by negative correlations across multiple repetitions [2].
Figure 1

Figure 1 shows locations of movement transducer coils on the tongue body (TB), upper lips (UL), lower lips (LL) and lower incisors (LI) for multiple repetitions of /u/ produced in a carrier phrase by a speaker of American English.

We have found negative correlations between values of tongue-body height and upper-lip protrusion in such data from five of seven speakers tested, which supports the hypothesis that there is motor equivalence in the transformation between the (articulatory) area function and the acoustic transfer function. In addition, in one of two subjects tested so far, we found preliminary indications of motor equivalence for /l/ and /r/, and for all three sounds, motor equivalence findings were confined to subsets that were acoustically less "prototypical", that is, near boundaries shared by other sounds in acoustic space.

We conclude tentatively that there is articulatory-to-acoustic motor equivalence, which reflects a control strategy used by articulatory synergisms to help keep acoustic variation within perceptually-acceptable regions in acoustic space. Since the use of auditory feedback for this purpose is unlikely, the control must rely on predictive simulation by using feedback from an internal (to the CNS) model of relations between articulatory commands and the sound output [3, 4, 5].

Auditory Feedback

Some characteristics of the hypothesized internal model can be illustrated with evidence from patients who have experienced a change in hearing status after having acquired speech.

Figure 2 shows spectral data from sibilants produced by three cochlear implant patients (1-3) who became profoundly deaf postlingually, and after a number of years of profound deafness acquired some hearing from multi-channel cochlear implants [6]. The figure shows average values of spectral median (on the left) and symmetry (on the right) for the consonants /s/ (unfilled circles) and /f/ (filled triangles) at three times with respect to receipt of the implant: pre-implant, just post implant, and six months later. Two of the patients (1 and 3) showed a good distinction between the two consonants pre-implant, even after a number of years of profound deafness. Data from two additional patients, not shown, are similar. On the other hand, Patient 2 had reversed values of the two measures pre-implant, consistent with an auditory impression that the consonants were quite distorted. She was able to correct her pronunciation with the use of the auditory feedback from her implant, but only after months of practice. The pre-implant integrity of the sibilants in four of five patients and the fact that correction took a number of months in Patient 2 indicate that the control mechanism is quite resistant to change, even in the face of profound changes in auditory feedback.

The nature of the control of sibilant production is further illustrated by data from a bilateral acoustic neuroma (NF2) patient who had tumor-removal surgery that severed her remaining eighth nerve [7]. At the
time of surgery, she received an auditory brainstem implant, which ended up providing her with auditory envelope but not spectral cues. Figure 3 plots spectral median versus week from the onset of hearing loss (OHL) for /s/ and /ʃ/. From before OHL and continuing for over 70 weeks post-OHL, she maintained a good contrast between the two sounds. At week 72, she had surgery to anastomose her left hypoglossal nerve to the facial nerve, in an attempt to restore some facial function that had also been lost at the time of tumor removal surgery. The anastomosis denervated some tongue muscles on the left side, producing a tongue weakness that effectively altered a property of the "plant". Without auditory feedback to help maintain the control mechanism, the acoustic contrast between the two sounds then gradually collapsed. (From additional measures, it was observed that her VOT—a phonemic mechanism—did not change longitudinally, in spite of changes in the related postural parameters of average vowel SPL, F0 and a spectral indicator of average glottal abduction.)

Figure 4 shows data that illustrate the relation between phonemic and postural parameters and a somewhat surprising aspect of phonemic control. The figure plots spectral median for /ʃ/ (phonemic) and SPL for a following vowel (postural) versus time during an experiment with another cochlear implant patient, in which his speech processor was on at the beginning and end of the experiment (500 to 1000 sec. and 2000 to 2500 sec.) and off in the middle of the experiment (1000 to 2000 sec.) [8]. When his processor was turned off, vowel SPL increased abruptly; it remained high until the final on condition, when it dropped abruptly. On the other hand, the sibilant spectral median did not change abruptly at the on-to-off transition. During the off condition, the median gradually drifted upward (toward /s/ values) and then at the off-to-on transition, dropped abruptly. According to our theoretical overview, this drift of a phonemic parameter over the course of less than an hour was unexpected. Further insight is gained from examining the underlying articulation, as illustrated in the next figure.

Figure 5 shows the horizontal position of a movement transducer coil on the speaker's tongue blade versus time for repetitions of /ʃ/ spanning the off-to-on transition. The gradual forward movement during the off condition presumably is responsible for the upward drift of the spectral median. The first /ʃ/ produced after the speech processor is turned on follows the trend from the preceding off condition; however, by the next repetition, the subject seems to have detected the acoustic distortion and pulls his tongue back by 2 mm, in an apparent overcompensation for the detected error. Observation of this speaker's dental cast revealed that his palatal vault was the narrowest of the approximately 100 palates we have examined. His extremely narrow palatal vault may have made the consonant acoustics especially sensitive to small articulatory changes. Without the
help of auditory feedback for maintenance purposes (possibly in combination with the presence of the movement transducer coil on his tongue blade), the control parameters degraded slightly and caused the unexpectedly rapid articulatory and acoustic drift.

CONCLUSIONS

The data cited above, along with observations by a number of other researchers, lead us to the following tentative conclusions. Depending on the type of goal (acoustic or articulatory) and the available feedback mechanisms, an optimal combination of peripheral (auditory and somatosensory) and internal feedback is employed in the control of synergisms that transmit phonemic information. Parameter values of the controller and an internal model are acquired and then maintained with the help of auditory feedback. Auditory feedback is also used to monitor acoustic transmission conditions to help make situation-dependent, rapid adjustments in postural parameters underlying average sound level, F0 and rate— for adequate clarity and intelligibility. Phonemic mechanisms are usually robust, but this may depend on individual factors (such as amount of experience or anatomy) that determine the difficulty of parameter maintenance under adverse conditions, such as hearing loss or perturbations to articulation. Since much of the supporting data is variable and based on a few individual subjects, results from additional subjects are needed to make our conclusions less tentative.

ACKNOWLEDGMENTS

This work was supported by grants DC01925, DC01291, DC02525 and DC00361 from the National Institute on Deafness and Other Communication Disorders, National Institutes of Health. We are also grateful for helpful input from Pascal Perrier.

REFERENCES