Achieving Speech Motor Goals: Feedforward and Feedback Control

Joseph S. Perkell

Speech Communication Group, Research Laboratory of Electronics
Massachusetts Institute of Technology,
50 Vassar St., Cambridge, MA 02139-4307, U.S.A.

perkell@speech.mit.edu

Abstract. *This paper summarizes three perturbation experiments that investigated feedforward and auditory feedback mechanisms in speech production. In the first experiment a sensorimotor adaptation apparatus shifted the first formant of speakers' auditory feedback of their own vowel productions without their being aware of the shift. Without realizing it, the subjects compensated by producing vowels with F1 shifted in the opposite direction; the amount of compensation varied among the subjects. Subjects' auditory acuity for vowels was also measured. Those with greater acuity compensated more than those with lesser acuity. In the second experiment, speakers' auditory feedback was mixed with increasing levels of masking noise. Produced sound levels and durations generally rose monotonically with noise to signal ratio (N/S); vowel contrasts increased at lower N/S levels and fell at higher ones. In the third study, six cochlear implant users pronounced multiple repetitions of /CVC#CVC/ utterances while their auditory feedback was blocked and restored between utterances, at intervals they could not predict. The feedback switches resulted in consistent, predictable changes in vowel durations, but not in segmental spectral measures. These results are compatible with the use of auditory goals and feedforward and feedback control as implemented in the DIVA model of speech motor planning.*

1. Introduction

This paper is concerned with the nature of motor programming goals for phonemic speech articulations and how feedforward and feedback mechanisms are used to produce the movements that achieve those goals. Recent evidence indicates that some phonemic goals are in the auditory domain. That evidence concerns articulatory-to-acoustic motor equivalence for the vowel /u/ (c.f. Perkell et al, 1993) and the semivowel /r/ (Guenther et al, 1999); changes in speech in response to changes in hearing (see Perkell et al., 2000, Vick et al, 2001); relations between speakers' auditory acuity and the degree of their produced vowel (Perkell et al., 2004a) and sibilant (Perkell et al., 2004b) contrasts; and sensorimotor adaptation (see Houde & Jordan, 1998). According to our interpretation, all of these results are compatible with the function of the DIVA

model of speech motor planning (see Guenther et al, 2006). DIVA is a neurocomputational model of relations among: cortical activity for producing speech sounds, the motor output, and the resulting sensory consequences. In the model, some phonemic goals are encoded in neural projections (mappings) from premotor cortex to sensory cortex, mappings that describe *regions in multidimensional auditory-temporal space*. The model has two control subsystems, a feedback subsystem and a feedforward subsystem. Feedback control employs error detection and correction to teach, refine and update the feedforward control mechanisms. As speech is acquired and becomes fluent, speech sounds, syllables and words become encoded as sequences of feedforward commands that no longer rely on auditory feedback.

To learn more about feedforward and auditory feedback control mechanisms in speech, investigators have conducted studies in which subjects' auditory feedback has been perturbed and their compensatory responses measured. Some of these studies used steady-state perturbations, such as blocking hearing with masking noise; others have used intermittent auditory perturbations that the subjects cannot anticipate. Unanticipated modifications of auditory feedback have revealed that mechanisms are available that can detect and correct production errors within about 100 to 150 ms from the onset of the perturbation (see Tourville et al., 2005). Therefore, if a movement lasts long enough, auditory errors can be corrected during the movement itself by closed-loop feedback. However, many articulatory movements in mature, fluent speech do not last long enough to be corrected by auditory feedback. It follows that fluent adult speech production is controlled almost entirely by feedforward mechanisms, as in the DIVA model (Guenther et al., 2006).

The current paper summarizes the results of three recent perturbation studies from our laboratory on auditory goals and feedback mechanisms that further test hypotheses based on the DIVA model. The studies are of: 1) sensorimotor adaptation in vowel production and the relation of the amount of speakers' adaptation to their auditory acuity, 2) the effects of different levels of masking noise on produced vowel contrasts and 3) the timing of changes in segmental and suprasegmental parameters in response to sudden, unanticipated switching of hearing on and off in cochlear implant users.

2. Sensorimotor Adaptation to Perturbations of Vowel Acoustics and its Relation to Perception

We have conducted a study that investigated feedforward control in 20 normal-hearing speakers (Villacorta et al., 2004, 2005). The speakers participated in a sensorimotor adaptation (SA) experiment in which they pronounced CVC words containing the vowel /ɛ/. They received auditory feedback of their productions in which a digital signal processing apparatus shifted the vowel F1 either up or down in nearly real-time (18 ms delay). This paradigm was based on that of Houde and Jordan (1998), but differed from it in that only F1 was shifted and the words were pronounced naturally, not whispered. Ten of the subjects received upward shifts and the other 10, downward shifts. The subjects were unaware of the shift, which was introduced gradually in a ramp phase, maintained in a full-shift phase and then removed in a post-test phase (Fig. 1A). The entire experiment lasted about 80 minutes. The plots in Fig. 1A show that the subjects partially compensated for the shifts over many trials by modifying their productions so that produced F1 moved in the direction opposite to the experimentally

introduced shift. Produced F2 did not change significantly. When the shift was removed, there was a period during which produced F1 gradually returned to baseline values. The amount of compensation varied across subjects.

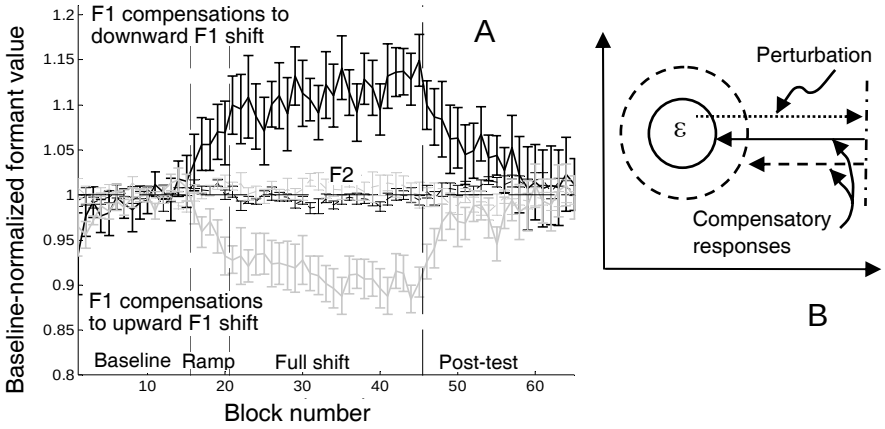

Figure 1. A: Compensatory responses to F1 shifts in normal-hearing subjects. Average values of subjects' baseline-normalized F1 and F2 (Fn_{Hz}/mean baseline Fn_{Hz}) vs. block number. Each block contains one repetition of each of 18 different words. The curves above baseline show the average of 10 subjects' productions in response to a downward shift of F1; the curves below baseline, the average of 10 subjects' responses to an upward F1 shift. The error bars show +/- one standard error around the mean. B: Schematic of goal regions and hypothetical compensatory responses for /ɛ/ for a high-acuity speaker (solid circle) and a low-acuity speaker (dashed circle). An F1 perturbation is indicated by the dotted arrow, and compensatory responses, by the solid and dashed arrows.

We were able to recall 13 of the subjects for tests in which their auditory acuity was measured in terms of just noticeable differences (JNDs) for natural-sounding /CɛC/ synthetic stimuli, based on their own productions. The amount of subjects' compensation in the SA experiment was related to the size of their vowel spaces and to their auditory acuity. When vowel space size was partialled out, there was a strong correlation (r = .8) between acuity and amount of compensation to F1 shift: speakers with better acuity tended to compensate more.

What underlies this correlation between acuity and compensation? Figure 1B schematizes how two speakers differing in acuity, and therefore in the sizes of their auditory goal regions for the vowel /ɛ/, might respond to a perturbation of F1. We postulate that the high-acuity speaker has a smaller goal region. The perturbation of F1 is indicated by a dotted arrow pointing to the right, and the shifted value of F1, by a vertical broken line. This high-acuity speaker, in response to the shift in F1, will produce a greater compensatory response (middle arrow) than the speaker with lesser acuity. That is because the high-acuity speaker continues to compensate until the F1 of his or her auditory feedback (which includes the shift) moves into the goal region. The distance between the shifted value of F1 (vertical line) and the edge of the goal region is greater for the high-acuity speaker. This interpretation is consistent with our finding that speakers' vowel and sibilant contrast distances were related to their auditory acuity for those sounds (Perkell et al., 2004a, b). In the DIVA model, auditory feedback

provides closed-loop corrections of current motor commands, accompanied by modifications of feedforward commands for subsequent movements. The gradual return to baseline after removal of the shift indicates that there was a temporary modification of feedforward commands during the shift that persisted for a while without the shift (Villacorta et al., 2004, 2005).

3. Effects of Masking Noise on Vowel and Sibilant Contrasts in Normal-hearing Speakers and Postlingually Deafened Cochlear Implant Users

We have investigated other aspects of the role of auditory feedback in speech production by examining speakers' produced vowel contrasts under different noise-to-signal (N/S) levels (Perkell et al., 2005, submitted). Seven postlingually deafened cochlear implant users and six normal-hearing controls pronounced multiple repetitions of utterances containing the vowels /i/, /u/ /ɛ/ and /æ/ while hearing their speech mixed with noise at seven equally spaced levels between their thresholds of detection and discomfort (plus a baseline condition with no added noise). Measures were made of average vowel spacing (AVS in mels – average of all possible inter-vowel distances in $F1_{mel}$ x $F2_{mel}$ space), SPL (dB) and duration (ms). The implant users were recorded at two time samples, one-month and one-year post-implant. Overall, the controls had the highest AVS, while the implant users had the lowest AVS at one month and intermediate values at one year.

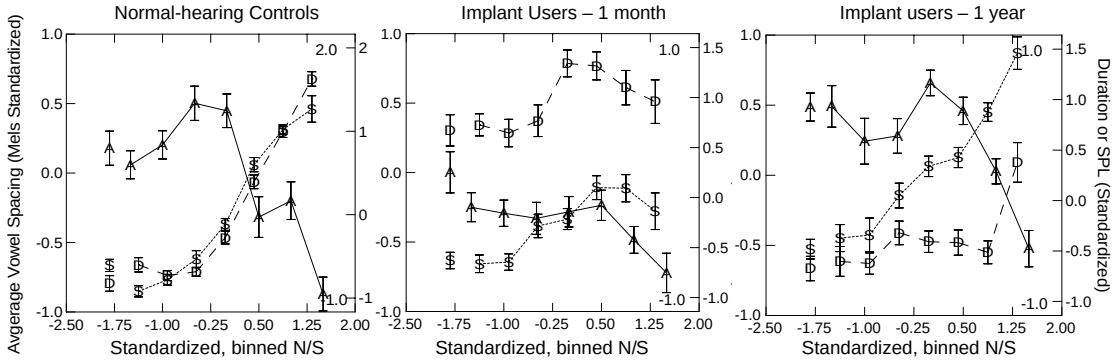

Figure 2. AVS (A), Duration (D) and SPL (S) vs. noise-to-signal ratio (N/S) (all as standard scores) averaged across 6 controls (left), 7 implant users at 1-month post-implant (middle) and the implant users at 1-year post-implant (right). AVS scores are on the left vertical axis; SPL, on the right vertical axis; the ranges of duration scores are shown by numbers in the upper and lower right-hand corners. Scores from a baseline condition are shown separately at the left of each plot (N/S is arbitrary). Error bars: standard error about the mean.

To account for inter-subject differences in data values and ranges when averaging across subjects, each subject's data were first converted to standard scores. As shown in Fig. 2, the controls' average vowel duration and SPL rose monotonically with increasing N/S. Their average vowel spacing rose initially; then it fell. Implant users' durations and SPL generally increased with increasing N/S, but less regularly. Their AVS at one month was flat at low N/S, and fell at high N/S. Their AVS function at one

year rose at low N/S and then fell, resembling the controls more than at one month; however the increase in AVS still was less than the controls' at low N/S.

The shapes of the AVS functions are interpreted as being the product of two underlying behaviors with increasing levels of noise: 1) Speakers seek to maintain clarity by increasing contrast distance. 2) Speakers also tend to minimize effort, which predominates more and more as masking increasingly prevents them from perceiving their produced contrasts and using auditory feedback to help maintain them. All three AVS samples (controls, implant users at one month and at one year) hypothetically may be characterized by an underlying function that incorporates the effects of clarity, economy of effort and masking. We hypothesize that the differences among the three observed functions depends on the maximum degree of contrast that the speakers from the different samples are able to produce. Presumably, this maximum degree of contrast depends on the state of the speakers' feedback and feedforward control subsystems. Implant users have lower baseline levels of contrast and are less able to increase contrasts in noise than normal-hearing speakers (Perkell et al., 2005, submitted).

4. Time Course of Speech Changes in Response to Short-term Changes in Hearing Status

In this study, the timing of changes in segmental and suprasegmental speech parameters was investigated in six cochlear implant users by switching their implant microphones off and on, a number of times in a single experimental session. The subjects produced multiple repetitions of the /dV_1n#SV_2d/ utterances, *Don shad*, *Don sad*, *Dun shed*, and *Dun said,* in quasi-random order. Thus, there were two vowel contrasts, /ɑ/-/ʌ/ in the first word position and /æ/-/ɛ/ in the second, and the sibilant contrast /s/-/ʃ/. The changes between hearing and non-hearing states were introduced under computer control by a voice-activated switch at V_1 onset; the number of utterances between switches was varied to minimize subject anticipation of the switches. Measures of the suprasegmental parameters of SPL, duration (reflecting speaking rate) and F0 were made from the vowels, and segmental contrast distances were measured for the vowels and sibilants. Changes in parameter values were computed by averaging data from multiple tokens, lined up with respect to the switch. The changes were calculated between averaged pre-switch values and values from the first, second and third utterances following the switches. Each subject's data were converted to standard scores and the results were averaged across speakers. ANOVA and post-hoc t-tests were calculated to examine pre- to post-switch changes. Because of space limitations, only changes in contrast distances and vowel durations are discussed here.

As shown in Table 1, contrast measures for the vowels and sibilants did not exhibit changes that were maintained consistently during the three post-switch utterances. On the other hand, vowel durations increased during the vowel in which hearing was blocked (V1*, utterance 1 – shaded column) and they decreased for the second vowel in utterance 1 when hearing was restored. The changed duration values were maintained consistently until the time of the next switch in hearing state. We speculate that the duration decrease with hearing restored did not take place until the following syllable (V2, utterance 1) because neural processing and muscle activation delays made it impossible to truncate motor commands already issued for producing V1*.

Parameter	Type	Switch	Post-switch Utterance								
			1			2			3		
			V1*	S	V2	V1	S	V2	V1	S	V2
Contrast Distance	Segmental	Block	0	–	0	0	0	0	0	0	0
		Restore	0	0	0	0	0	+	0	0	0
Duration	Suprasegmental	Block	+		+	+		+	+		+
		Restore	0		–	–		–	–		–

Table 1. Summary of the direction and statistical significance of parameter changes when feedback was blocked and restored in a group of six cochlear implant users. The changes are between pre-switch utterances and the first, second and third utterances following the hearing switch, which was made within the first 20 ms of V1* (shaded column). V1 = the vowel in the first word; S = the sibilant at the beginning of the second word; V2 = the vowel in the second word. + = significant increase; - = significant decrease; 0 = no significant change.

Why were there no consistent changes in sound contrasts when hearing state was switched? According to the DIVA model, *short-latency contrast changes* should occur when auditory feedback is *modified* (as in the SA experiment described above), but *not* when it is simply *blocked or restored*. The current results indicate that the mechanism regulating duration is at least partly under closed-loop control since changing the availability of auditory feedback resulted in changes in vowel durations. These differences in changes between segmental contrasts and a suprasegmental parameter (rate, as measured by durations) are consistent with previous findings, which also indicate that the two types of parameters are controlled differently (Svirsky et al., 1992; Perkell et al., 2000; Denny et al., submitted).

5. Summary and Discussion

The three studies described above have shown the following results. Speakers compensate for shifts in the first formant in their auditory feedback of vowels they are producing, and the amount of compensation is related to their acuity for small differences in the vowel spectra. This finding is consistent with earlier results and the idea that speakers with higher acuity will acquire auditory goal regions that are smaller and spaced further apart than speakers with lower acuity (Perkell et al, 2004a, b). When speakers' auditory feedback of their vowel productions is mixed with different levels of noise, they will increase contrasts at lower N/S levels as long as they can presumably perceive the contrasts, but as contrast perception is overwhelmed at higher N/S levels, contrasts decrease, hypothetically due to a predominating influence of economy of effort (see Lindblom, 1990). When hearing is blocked and restored unexpectedly, there are rapid changes in vowel durations that are compatible with closed-loop feedback control; however, segmental contrasts do not change consistently over relatively short intervals between switches in hearing state.

As mentioned in the Introduction, feedback control of segmental parameters involves the detection and correction of mismatches between expected and actual sensory consequences of speech articulation. Experimentally induced, unexpected modifications of auditory feedback can elicit observable rapid responses that seem to be closed-loop

(Tourville et al., 2005). Under real-world circumstances auditory disparities between intended and produced speech sounds tend to occur or be maintained over long time spans, e.g., as a consequence of vocal-tract growth or the insertion of dentures. Therefore, the primary role of auditory feedback control of segmental contrasts is to provide corrections that are incorporated into feedforward commands (as demonstrated in the laboratory in SA experiments). On the other hand, externally induced changes in acoustic transmission conditions, such as the occurrence of loud noises, occur often and call for rapid responses for the maintenance of intelligibility. It follows that unexpectedly blocking and restoring auditory feedback engages a different feedback control mechanism from the one that helps to acquire and maintain segmental contrasts.

All of the results described above are compatible with the function of the DIVA model of speech motor planning – in the way it employs auditory goal regions, economy of effort and feedforward and feedback control mechanisms. Since the DIVA model is formulated in terms of patterns of cortical connectivity and activity, it can also be tested with brain imaging experiments (see Guenther et al., 2006). When imaging studies and behavioral studies are used in combination to test the same DIVA-based hypotheses, they provide a valuable means of quantifying relations among phonemic specifications, brain activity, articulatory movements and the speech sound output.

Acknowledgements

The work from our laboratory that is described in this paper was done in collaboration with a number of people, including Satrajit Ghosh, Frank Guenther, Harlan Lane, Melanie Matthies, Mark Tiede, Majid Zandipour, Margaret Denny, Jennell Vick and Virgilio Villacorta. Support was from grants R01-DC001925 and R01-DC003007 from the National Institute on Deafness and Other Communication Disorders, N.I.H.

References

Denny, M., Perkell, J.S., Lane, H., Matthies, M.L., Tiede, M., Zandipour, M., Vick, J. & Burton, E. (submitted) Time course of speech changes in response to short-term changes in hearing state. *Journal of the Acoustical Society of America*

Guenther, F.H., Espy-Wilson, C., Boyce, S.E., Matthies, M.L., Zandipour, M. & Perkell, J.S. (1999). Articulatory tradeoffs reduce acoustic variability during American English /r/ production. *Journal of the Acoustical Society of America*, *105*, 2854-2865.

Guenther, F.H., Ghosh, S.S., & Tourville, J.A. (2006). Neural modeling and imaging of the cortical interactions underlying syllable production. *Brain and Language, 96*, 280-301.

Houde, J.F. & Jordan, M.I. (1998). Sensorimotor adaptation in speech production. *Science, 279*, 1213-1216.

Lindblom, B.E.F. (1990). Explaining phonetic variation: A sketch of the H&H theory. In W.J. Hardcastle & A. Marchal (Eds.), *Speech Production and Speech Modeling* (pp. 403-439). Netherlands: Kluwer Academic Publishers.

Perkell J.S., Denny, M., Lane, H., Guenther F.H., Matthies, M.L., Tiede, M., Vick, J., Zandipour, M. & Burton, E. (submitted) Effects of masking noise on vowel and sibilant contrasts in normal-hearing speakers and postlingually deafened cochlear implant users, *Journal of the Acoustical Society of America.*

Perkell, J.S., Guenther, F.H., Lane, H., Matthies, M.L., Perrier, P., Vick, J., Wilhelms-Tricarico, R. and Zandipour, M. (2000). A theory of speech motor control and supporting data from speakers with normal hearing and with profound hearing loss, *Journal of Phonetics, 28*, 233-272.

Perkell J.S., Guenther F.H., Lane, H., Matthies, M.L., Stockmann, E., Tiede, M. & Zandipour, M. (2004a). The distinctness of speakers' productions of vowel contrasts is related to their discrimination of the contrasts, *Journal of the Acoustical Society of America 116*, 2338-44.

Perkell, J.S., Matthies, M.L., Svirsky, M.A. & Jordan, M.I. (1993). Trading relations between tongue-body raising and lip rounding in production of the vowel /u/: A pilot motor equivalence study, *Journal of the Acoustical Society of America 93*, 2948-2961.

Perkell J.S., Matthies, M.L., Tiede, M., Lane, H., Zandipour, M., Marrone, N., Stockmann, E. & Guenther, F.H. (2004b).The distinctness of speakers' /s/-/ʃ/ contrast is related to their auditory discrimination and use of an articulatory saturation effect, *Journal of Speech, Language and Hearing Research 47*, 1259-69.

Svirsky, M.A., Lane, H., Perkell, J.S., & Wozniak, J. (1992). Effects of short-term auditory deprivation on speech production in adult cochlear implant users. *Journal of the Acoustical Society of America, 92*, 1284-1300.

Tourville, J.A., Guenther, F.H., Ghosh, S.S., Reilly, K.J., Bohland, J.W., & Nieto-Castanon, A. (2005). Effects of acoustic and articulatory perturbation on cortical activity during speech production. In 11th Annual Meeting of the Organization for Human Brain Mapping (pp. S49).

Vick, J., Lane, H., Perkell, J.S., Matthies, M.L., Gould, J., & Zandipour, M. (2001). Speech perception, production and intelligibility improvements in vowel-pair contrasts in adults who receive cochlear implants. *Journal of Speech, Language and Hearing Research 44*, 1257-68.

Villacorta, V., Perkell, J. S., & Guenther, F. H. (2004). Sensorimotor adaptation to acoustic perturbations in vowel formants. *Journal of the Acoustical Society of America 115*, 2430 (A).

Villacorta, V., Perkell, J.S., & Guenther, F.H. (2005). Relations between speech sensorimotor adaptation and perceptual acuity. *Journal of the Acoustical Society of America 117*, 2618-2619 (A).