Proceedings of the 15th International Congress of Phonetic Sciences. Barcelona, Spain. August 3–9, 2003. Universitat Autònoma de Barcelona.

1306

Sensorimotor adaptation to auditory perturbations during speech: Acoustic and kinematic experiments

Ludo Max[†‡], Marie E. Wallace[†‡], and Irena Vincent[†]

† University of Connecticut, Storrs, CT, USA

‡ Haskins Laboratories, New Haven, CT, USA

E-mail: ludo.max@uconn.edu

ABSTRACT

We investigated sensorimotor adaptation to auditory feedback perturbations during speech production. In Study I, formant or fundamental frequency (F_0) feedback was manipulated during sustained vowels. When F_0 feedback was altered, group data showed upward F_0 adjustments regardless of the feedback shift direction. When formant feedback was altered, group data showed opposing adjustments in both the first and second formant. Sensorimotor adaptation was present at vowel onset, and subjects showed aftereffects when the auditory perturbation was removed. For Studies II (acoustics) and III (kinematics), subjects produced monosyllabic words while upward or downward formant feedback shifts were applied. Acoustic results replicated those for sustained vowels. Kinematic analyses of jaw and tongue positions and displacements indicated motor-equivalent adaptation in the overall gestures rather than individual articulators. Findings are highly consistent with recent data on limb movements and suggest continuous updating of forward and/or inverse internal models of the articulation-to-acoustics transformations in the vocal tract.

1. INTRODUCTION

A large body of empirical data and several theoretical models suggest that speech movements are planned in terms of auditory/acoustic goals [1-4]. However, given that delays in the auditory feedback are too long to contribute to corrections of ongoing movements, a crucial role for the auditory system may be related to the acquisition, consolidation, and updating of an internal model of the vocal tract. Internal models of the various motor systems and, when applicable, the environment have been widely proposed as a solution for the selection of accurate motor commands by the central nervous system (CNS). The transformation from efferent signals to movements is complex due to the time-varying influence of several variables that depend on neural and muscular physiological factors, the current state of the system, and biomechanics. Therefore, planning movements on the basis of desired movement consequences requires access to an inverse internal model that consists of a representation of the dynamic mapping between central efferent signals and the sensory consequences of the resulting movements.

Evidence in support of this perspective comes primarily from two lines of research. First, kinematic and kinetic characteristics of limb movements are consistent with a control scheme in which the effector system's dynamics and external loads are taken into account during movement planning [5-7]. Hence, a representation of the effects of dynamics and load on the multiple transformations from efferent neural signals to movement consequences must have been available to the CNS during the planning stage. Second, subjects adjust movement planning in the presence of externally manipulated consequences such as those resulting from altered visual information or system dynamics [8-11]. Taking into account that subjects are able to achieve desired movement consequences with modified central commands, that the adjusted movements show correct anticipation of the manipulated sensory environment, and that aftereffects are observed in the form of continued but unnecessary compensation immediately after feedback is restored to normal, it is indeed likely that the CNS has access to a continually updated representation of the mapping between efferent and afferent signals.

If the CNS relies on such models to generate motor commands, then understanding the nature, level of detail, and use of these models appears particularly important for speech production. In speech, the goal is to generate an acoustic signal that is intelligible to a listener, and, thus, at some level movements must be planned in terms of sequences of acoustic targets. In addition, no visual feedback is available, auditory feedback delays are too long, articulatory movements are extremely fast, and the task involves coordinating different subsystems (pulmonary, laryngeal, orofacial) that affect each other's activity through aerodynamic and biomechanical interactions.

To investigate internal models of the multiple transformations from motor commands to acoustic output in speech, we have initiated a program of research focusing on the ability to learn altered command-to-output mappings. Here, we present and integrate the results from a series of three experiments investigating sensorimotor adaptation to formant perturbations in the auditory feedback signal. Whereas a previous study by others had examined articulatory adaptation to manipulated auditory feedback during whispering [12], these studies from our laboratory focused on speakers' articulatory compensations to formant shifts in the auditory feedback during typical productions of sustained vowels and monosyllabic words.

ISBN 1-876346-48-5

2. METHOD

Study I. Subjects were eight male adults (21-33 years of age) with normal voice and speech. Normal hearing thresholds were confirmed with an audiological screening. Subjects were unaware of the purpose of the study.

Formant or F_0 feedback was manipulated in real time (delay approximately 20 ms) with a digital signal processor (Boss VT-1, Roland) during sustained productions of front (/ɛ/), central (/ʌ/), and back (/ɔ/) vowels. Auditory feedback was delivered through insert earphones with all instrumentation calibrated such that a 1 KHz sine wave with an intensity of 70 dB SPL at the microphone resulted in an output intensity of 72 dB SPL in the left and right earphones. Productions were elicited in randomized order by displaying monosyllabic words (*pet, bus, law*) with the target vowels on a computer monitor and instructing subjects to produce and sustain only the vowel for approximately 1.5 seconds.

Subjects produced 90 vowels (30 trials of each vowel) in each of 9 conditions. In 8 conditions, the first 45 vowels were produced while either F_0 or the formants were shifted up or down (0.2 and 2.0 semitones (ST) up or down for F_0 and 1.8 and 4.5 ST up or down for the formants). The remaining condition was a control condition in which no frequency shifts were applied. Subjects' speech intensity was kept approximately constant by means of intensity feedback displayed on the computer monitor.

Using the PRAAT acoustic analysis software [13], F_0 and the first (F1) and second (F2) formant frequencies were automatically extracted at various locations throughout each vowel. To reflect anticipatory adaptation, data reported here were obtained 100 ms into the vowel for F_0 and 10 ms into the vowel for F1 and F2.

Study II. Subjects were three female and two male adults (22-44 years of age) with normal voice and speech. All subjects passed a hearing screening, were unaware of the purpose of the study, and had not participated in Study I.

Procedures, instrumentation, and analyses were identical to those in Study I except for the fact that subjects produced the words *tech, tuck,* and *talk* and F1 and F2 were measured at 50% into the vowel because (a) the anticipatory nature of the adaptation had already been established in Study I, and (b) the word-initial consonant-related gesture limited the possibility of adaptation from the very onset of phonation.

Study III. Both acoustic and kinematic data have been collected from four female and four male young adults with no diagnosed communication or neurological disorders. At the time of this writing, initial results are available from three male adults (19-21 years of age). Subjects were again unaware of the purpose of the study and had not participated in any of the previous studies. They produced the same monosyllabic words as used in Study II while relatively small and relatively large upward and downward formant shifts were gradually introduced in the auditory feedback using a real time (latency 10 ms) digital signal processor (VoiceOne, TC Helicon) under computer control. Based on acoustic measures (TF32 software [14]) of the original and processed speech signal of one subject, the average shift across F1 and F2 that was introduced by the processor was 1.0 ST in the small shift up condition, -0.4 ST in the small shift down condition, 2.5 ST in the large shift up condition, and -3.0 ST in the large shift down condition.

Throughout each condition, subjects produced the target words at a rate of 18 words per minute while receiving auditory feedback through insert earphones. Each of the four experimental conditions consisted of two blocks, a formant-manipulated block and a non-manipulated block. The formant-manipulated block lasted for 8.5 minutes with the shift being introduced gradually over the first 5 minutes and the full shift being maintained for the next 3.5 minutes. In a control condition, no formant shift was applied during either the 8.5 or the 3.5 minute blocks.

Movements of the lips, jaw, and tongue were transduced with a two-dimensional electromagnetic midsagittal articulograph (EMA; Carstens AG200). Receiver coils were attached at the vermillion border of the upper and lower lip, just below the lower incisors, and on the tongue. For the tongue, three coils were positioned 1 to 1.5 cm apart with the most anterior coil approximately 1 cm from the tongue tip. Reference receiver coils were attached at the nasion and just above the upper incisors.

Kinematic data were filtered, corrected for head/helmet movement, and rotated and shifted into a coordinate system in which the x-axis lies within the individual subject's occlusal plane and the y-axis is normal to the occlusal plane and intersects it at the tip of the upper incisors. For the vowel-related opening gesture, each receiver coil's displacement, position at the time of movement onset and offset, movement duration, and peak velocity were extracted using custom-developed MATLAB routines.

3. RESULTS

Study I. When F_0 feedback was altered, group data showed upward adjustments in F_0 regardless of the direction of the feedback shift. When formant feedback was altered, however, group data showed opposing adjustments (i.e., in the opposite direction of the shift) in both F1 and F2. Figure 1 shows these opposing adjustments for productions of the vowel /ɔ/ in the four formant-shifted conditions (note that the x-axis in this and all following graphs indicates the total number of trials produced at a given point into the condition, even though each graph shows the trials for only one of the three vowels). Thus, adaptation was observed for the articulatory but not the phonatory system. Articulatory sensorimotor adaptation (a) was measurable at vowel onset (indicating *anticipatory* rather than reactive adjustments), (b) occurred after only a few trials, (c) occurred for all three vowels but was largest for /ɔ/ and smallest for /ɛ/, and (d) was associated with aftereffects in approximately the first ten trials after the manipulation was removed.

ISBN 1-876346-48-5 © 2003 UAB

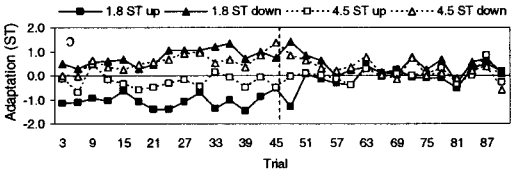

Figure 1: Difference (in ST) in F2 for productions of /ɔ/ in four experimental conditions vs a control condition. The dashed vertical line marks the point when normal feedback was restored. Data averaged across 8 subjects.

Study II. The results for productions of consonant-vowel-consonant words in Study II closely matched those obtained for sustained vowels in Study I. As illustrated in Figure 2, specific compensation (opposing adjustments) for the applied formant shifts (a) was seen for all three words, although, again, adaptation was largest for /ɔ/ and smallest for /ɛ/, (b) appeared early in the conditions after only a few trials, (c) was similar in magnitude to that seen for sustained vowels, and (d) was much smaller or even absent for the larger formant shifts as compared with the smaller ones.

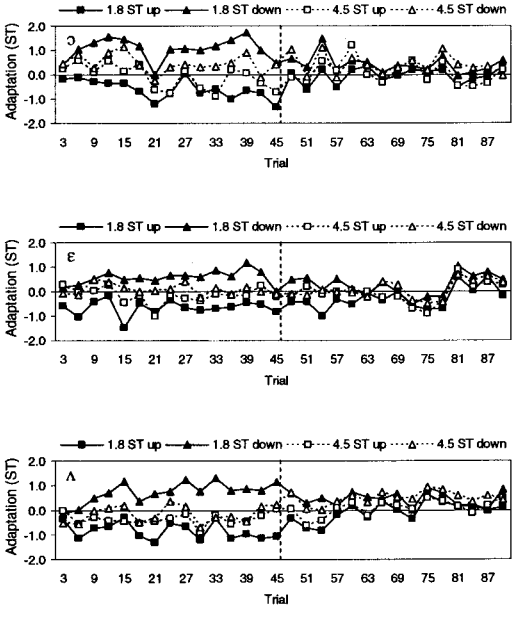

Figure 2: Difference (in ST) in F2 for productions of /ɔ/ (top panel), /ɛ/ (middle panel), and /ʌ/ (bottom panel), in four experimental conditions vs a control condition. Vowels were produced in monosyllabic words. Data averaged across 5 subjects.

Study III. Given the unusual speaking conditions created by the articulograph "helmet" and the receiver coils attached to the articulators, a limited set of acoustic analyses was first conducted to determine whether or not subjects showed auditory-based adaptation as did the subjects in Studies I and II. Using the TF32 software, F1 and F2 were extracted at a time point 50% into the vowel, and the difference in F1 and F2 between each experimental condition and the control condition was computed to obtain data that are directly comparable with those discussed and displayed above for Studies I and II. Figure 3 illustrates the results for one representative subject. The data show clear evidence of sensorimotor adaptation consisting of a compensatory adjustment of the acoustic output in the opposite direction of the shift.

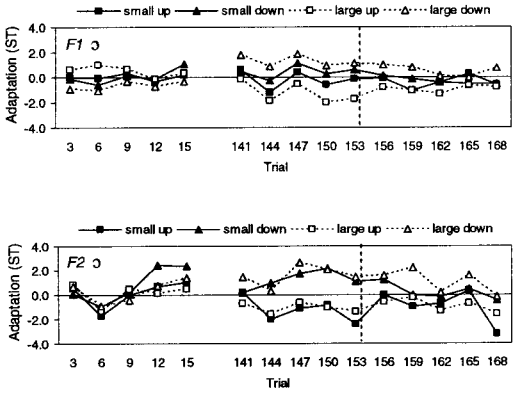

Figure 3: Difference (in ST) in F1 (top panel) and F2 (bottom panel) for productions of /ɔ/ in four experimental conditions vs a control condition. Vowels were produced in monosyllabic words. Data for one individual subject.

Interestingly, despite the remarkable adaptation in the acoustic output, no consistent adjustments were obvious in the articulatory kinematics. That is, although actual tongue and jaw positions and displacements may differ from condition to condition, (a) the kinematic measures included here did not show adjustments that occurred *in parallel with* the experimental manipulations, and (b) the consistent acoustic compensation in opposite directions for upward and downward formant shifts was not accompanied by *opposing* adjustments in tongue or jaw positions for upward vs downward shifts. Figure 4 contains data from the same subject whose acoustic data are in Figure 3. Using the same trials of the same word, it shows the y-axis coordinate of the jaw receiver at the point of maximum displacement, and both the y- and x-axis coordinate of the middle tongue receiver at the point of maximum y-axis displacement for the control, small upward shift, and small downward shift conditions. Although the latter two conditions resulted in strong acoustic adaptation for F2, no opposing adjustments or other trends can be identified in the kinematic measures.

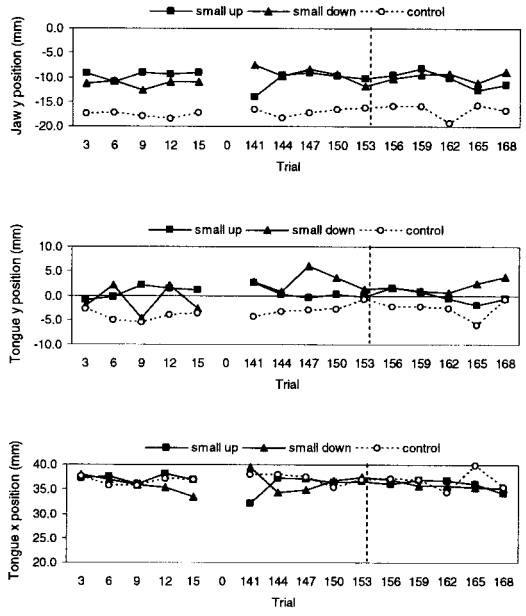

Figure 4: Jaw y-axis and tongue y- and x-axis coordinates for /ɔ/ opening movements in the small up and down shift conditions (which show strong F2 acoustic adaptation). The same trials are shown as in Figure 3. Data from one subject.

4. DISCUSSION

This series of experiments demonstrates that adult subjects can re-learn the mapping between central motor commands to the muscle systems in the vocal tract and sensory consequences of the resulting movements. In particular, this work shows that speakers compensate for external manipulations of the formant frequencies in the auditory feedback channel by adjusting their acoustic output in such a way that their formant frequencies change in the opposite direction of the experimental shift—thereby in effect minimizing or canceling the experienced discrepancy between anticipated and actual sensory consequences.

This main finding is highly consistent with recent studies of limb sensorimotor control, and suggests continuous refinements and adjustments of the internal representations of efferent-afferent mappings relevant for the planning and organization of orofacial movements for speech. Specifically, these data suggest continuous updating of forward and/or inverse internal models of the articulation-to-acoustics transformations in the vocal tract. In terms of the proposed goals for speech movement planning, it may be of theoretical importance that these initial analyses revealed strong adaptation in the acoustic output, apparently without accompanying major changes in the positioning of individual articulators.

REFERENCES

[1] F.H. Guenther, M. Hampson, and D. Johnson, "A theoretical investigation of reference frames for the planning of speech movements", *Psychol Rev*, **105**, pp. 611-33, 1998.

[2] P. Ladefoged, J. DeClerk, M. Lindau, and G. Papnun, "An auditory-motor theory of speech production", *Working Papers in Phonetics*, **22**, pp. 48-76, 1972.

[3] H. Lane, M. Matthies, J. Perkell, J. Vick, and M. Zandipour, "The effects of changes in hearing status in cochlear implant users on the acoustic vowel space and CV coarticulation", *J Speech Lang Hear Res*, **44**, pp. 552-63, 2001.

[4] J. Perkell, F. Guenther, H. Lane, M. Matthies, P. Perrier, J. Vick, R. Wilhelms-Tricario, and M. Zandipour, "A theory of speech motor control and supporting data from speakers with normal hearing and profound hearing loss", *J Phonetics*, **28**, pp. 233-72, 2000.

[5] J.R. Flanagan and S. Lolley, "The inertial anisotropy of the arm is accurately predicted during movement planning", *J Neurosci*, **21**, pp. 1361-9, 2001.

[6] M. Suzuki, D.M. Shiller, P.L. Gribble, and D.J. Ostry, "Relationship between cocontraction, movement kinematics and phasic muscle activity in single-joint arm movement", *Exp Brain Res*, **140**, pp. 171-81, 2001.

[7] P.N. Sabes, M.I. Jordan, and D.M. Wolpert, "The role of inertial sensitivity in motor planning", *J Neurosci*, **18**, pp. 5948-5957, 1998.

[8] J.R. Flanagan and A.K. Rao, "Trajectory adaptation to a nonlinear visuomotor transformation: evidence of motion planning in visually perceived space", *J Neurophysiol*, **74**, pp. 2174-8, 1995.

[9] R. Shadmehr and F.A. Mussa-Ivaldi, "Adaptive representation of dynamics during learning of a motor task", *J Neurosci*, **14**, pp. 3208-24, 1994.

[10] K.A. Thoroughman and R. Shadmehr, "Electromyographic correlates of learning an internal model of reaching movements", *J Neurosci*, **19**, pp. 8573-88, 1999.

[11] D.M. Wolpert, Z. Gahramani, and M.I. Jordan, "Are arm trajectories planned in kinematic or dynamic coordinates? An adaptation study", *Exp Brain Res*, **103**, pp. 460-70, 1995.

[12] J. Houde and M. Jordan, "Sensorimotor adaptation in speech production", *Science*, **279**, pp. 1213-16, 1998.

[13] P. Boersma and D. Weenink, "Praat". Computer software, 2002.

[14] P. Milenkovic, "TF32". Computer software, 2002.

ISBN 1-876346-48-5