1286

Somatosensory basis of speech production

Stéphanie Tremblay*, Douglas M. Shiller* & David J. Ostry*†

** Department of Psychology, McGill University, Montreal, Quebec H3A 1B1, Canada*
† Haskins Laboratories, New Haven, Connecticut 06511-6695, USA

The hypothesis that speech goals are defined acoustically and maintained by auditory feedback is a central idea in speech production research[1–6]. An alternative proposal is that speech production is organized in terms of control signals that subserve movements and associated vocal-tract configurations[7–9]. Indeed, the capacity for intelligible speech by deaf speakers suggests that somatosensory inputs related to movement play a role in speech production—but studies that might have documented a somatosensory component have been equivocal. For example, mechanical perturbations that have altered somatosensory feedback have simultaneously altered acoustics[10–14]. Hence, any adaptation observed under these conditions may have been a consequence of acoustic change. Here we show that somatosensory information on its own is fundamental to the achievement of speech movements. This demonstration involves a dissociation of somatosensory and auditory feedback during speech production. Over time, subjects correct for the effects of a complex mechanical load that alters jaw movements (and hence somatosensory feedback), but which has no measurable or perceptible effect on acoustic output. The findings indicate that the positions of speech articulators and associated somatosensory inputs constitute a goal of speech movements that is wholly separate from the sounds produced.

We have adapted a technique used in studies of limb motor control to apply velocity dependent mechanical perturbations to the jaw. The perturbation was designed to be of sufficient strength to alter systematically the motion path of the jaw, and hence somatosensory feedback, without affecting the associated acoustic output. As in work on limb movement, adaptation to an artificial mechanical force field indicates the adjustment of control signals to take account of loads on the basis of sensory input[15–19].

We have altered somatosensory feedback in three different tasks—one involving normal vocalized speech, another during 'silent speech' (speech without vocalization), and a third that involves a non-speech jaw movement that is matched in amplitude and duration to that observed in speech. In the vocalized speech condition, subjects were required to repeatedly produce the utterance *siat* (pronounced 'see-at') at a subject-chosen rate and volume. This condition was tested to assess the extent to which adaptation to a somatosensory perturbation might occur in the presence of unaltered acoustic feedback. The silent speech condition explicitly removed auditory feedback, and hence examined the ability of subjects to adapt in the total absence of auditory input. Subjects in this group were asked to articulate the utterance *siat* without producing any sound. The non-speech condition addressed the issue of whether adaptation would occur in a cyclical jaw movement task that is matched only in amplitude and duration to that observed in speech. There was no reference whatsoever to speech production in the description of the task for the non-speech group.

A robotic device was connected to the mandibular teeth, and was used to deliver mechanical perturbations to the jaw (Fig. 1a). Sagittal plane forces were applied along a horizontal axis (parallel to the occlusal plane), in the direction of jaw protrusion. The forces were proportional to the instantaneous vertical velocity of the jaw (measured at the incisors) such that the magnitude of the perturbation increased with the velocity of movement (Fig. 1b). Performance was quantified for each subject by measuring on a trial-by-trial basis the maximum horizontal distance between the movement path under force-field conditions and the average movement path with the field off (null field). A decrease in the horizontal distance over trials reflects sensorimotor adaptation in which the effect of the force field is reduced.

Analyses of kinematic data revealed a systematic pattern of force-field adaptation in speech production. Figure 2 illustrates movements for individual subjects in the vocalized speech condition (Fig. 2a), the silent speech condition (Fig. 2b) and the non-speech condition (Fig. 2c). A baseline phase of 20 null-field trials provided a reference movement path under unperturbed conditions (black lines). As shown in blue, the jaw path deviated in the direction of protrusion with the introduction of the force field. Following training, adaptation (shown in red) was observed in the vocalized speech and the silent speech conditions, but not in the non-speech condition. The green paths illustrate an after-effect in which the jaw was retracted in comparison to the baseline following the unexpected removal of the force field (vocalized speech and silent speech conditions).

Figure 3 gives mean values of maximum horizontal deviation in each of these conditions on a per-subject basis. It can be seen that the force field had a similar initial effect for subjects in all experimental conditions: compared to the baseline, movements at the start of training in the force field deviated significantly in the protrusion direction ($P < 0.001$ for all subjects). Although there

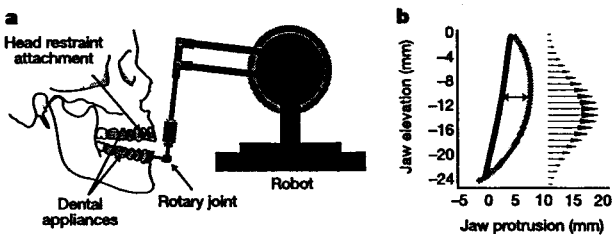

Figure 1 Experimental set-up and representative data. **a**, Diagram showing subject attached to the robotic device. **b**, Jaw opening movement with the force field off (black) and on initial exposure to the field (grey). Vectors depict the magnitude and direction of force applied by the robot over the course of the movement. The double-headed arrow shows the maximum horizontal deviation between null-field and force-field movements that served as a performance index.

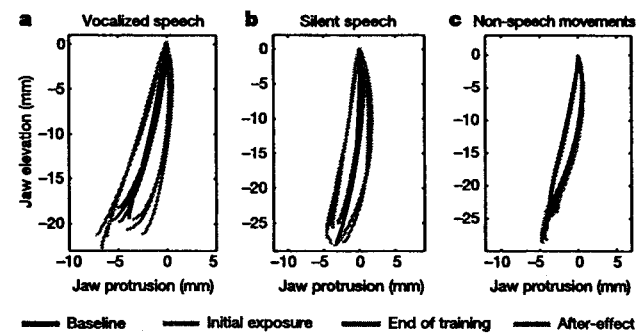

Figure 2 Sagittal plane jaw motion paths. Data were acquired during the baseline condition (black trace), on initial exposure to the force field (blue), at the end of training (red), and following unexpected removal of the field (green). The figure shows individual trials for single subjects. **a**, During vocalized speech, adaptation to the force field and a subsequent after-effect are observed. **b**, During silent speech, the pattern of adaptation and after-effect observed in vocalized speech are unaltered by removal of acoustic feedback. **c**, Matched non-speech movements show neither adaptation nor an after-effect.

were individual differences in the initial deviation caused by the field, there was no overall difference in the magnitude of the perturbation between the groups ($P > 0.05$). Somatosensory feedback was thus initially altered in a comparable way in each of the conditions tested. At the end of training (shown in red), adaptation was observed as indicated by a significant reduction in deviation from baseline for all but two subjects (S7 and S8) in the vocalized speech group ($P < 0.01$), and for all subjects in the silent speech group ($P < 0.001$) (Fig. 3a, b). In contrast, adaptation was not observed in the non-speech group: for two subjects, movements at the end of training did not differ reliably from those at the beginning ($P > 0.05$), for the other two subjects there was an increase in the deviation (Fig. 3c, left side). A motion-dependent after-effect (shown in green) following the unexpected removal of the force field illustrates that adjustments to offset the effects of the force field were generally as large in magnitude as the initial deviation for the vocalized speech and the silent speech groups (Fig. 3a, b). None of the four subjects in the non-speech condition showed an after-effect in the direction of retraction relative to baseline (Fig. 3c).

Two subjects in the non-speech group were tested over an extended period of time to see whether the absence of adaptation in that condition arose simply as a consequence of limited practice. Both were tested under force-field conditions for four days in a row (2,100 trials in total for each subject). The right side of Fig. 3c shows results of the final day of training, labelled 2B and 3B. Neither of the subjects showed an adaptation or an after-effect. This indicates that matching jaw movements on duration and amplitude alone is insufficient for the achievement of the adaptation observed in the speech groups.

As a further control, we tested a variant of the non-speech condition in which subjects were required to produce discrete jaw lowering movements from an initial position with the mouth closed to a remembered target location. We reasoned that discrete jaw lowering movements might have absolute spatial goals more comparable to those observed in speech. The opening movements were equal in amplitude and duration to the opening movements in the vocalized speech condition. Subjects were instructed to briefly hold the target position and then to return slowly to the starting position. Subjects were trained in the same force field as in the previous conditions. A pattern similar to that observed in cyclical non-speech movements was obtained (Fig. 3d). For all subjects, initial exposure to the force field resulted in a reliable deviation from baseline conditions ($P < 0.01$). Training did not produce a reduction in the deviation ($P > 0.05$). The sudden removal of the force field did not result in an after-effect comparable to that observed in Fig. 3a, b ($P > 0.05$).

To verify that the perturbations produced by the force field did not systematically alter the speech acoustics, the first and second formant frequencies were examined during the transition between *i* and *a* in the test utterance *siat*. Figure 4a gives an example of the acoustic spectrogram for a single utterance. The spectrogram depicts F1 and F2 frequencies during the voiced portion of the utterance. The white rectangle highlights the transition between vowels. Figure 4b shows transitions in formant frequencies between vowels (time-normalized) at different phases of the experiment for a single subject. It may be seen that the formant frequency transitions were similar throughout the experiment. Statistical tests were carried out across subjects in order to compare the acoustic signals in the baseline, the first and the last 20% of training, and in the final null-field trials in which the load to the jaw had been unexpectedly removed. Differences in average F1 and F2 frequency were assessed at the midpoint of the vowel transition as well as at its start and end

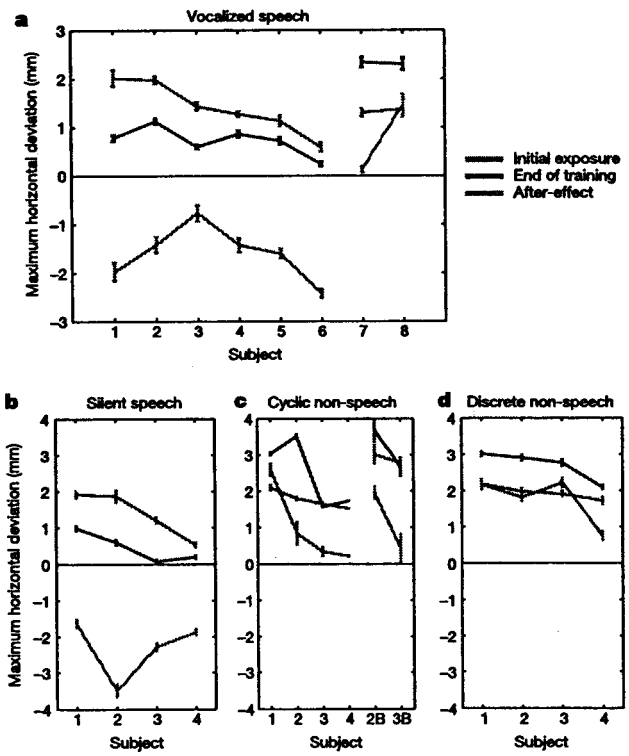

Figure 3 Jaw position during force-field adaptation shown on a per-subject basis. Average values of maximum horizontal deviation are presented for initial exposure to the force field (blue), at the end of the training (red), and following the unexpected removal of the field (green). Connecting lines are provided for visualization purposes only, and do not imply correlations across subjects. **a**, In vocalized speech, six out of eight subjects showed adaptation (reduction of deviation and an after-effect). **b**, All subjects in the silent speech condition showed adaptation. **c**, None of the subjects in the cyclical non-speech condition showed either a reduction in deviation or an after-effect (S1–S4). S2 and S3 were tested for an extended period of time and showed no adaptation (2B and 3B). **d**, Subjects in the discrete non-speech condition showed a pattern similar to those in the cyclical non-speech condition.

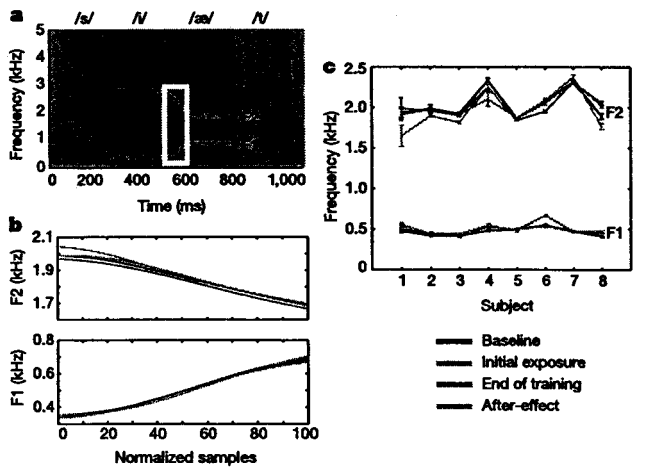

Figure 4 Acoustic data. **a**, Spectrogram of the acoustic signal during a single repetition of the utterance *siat*. The dark bands show the frequency composition of acoustical energy. F1 and F2 formant tracks are indicated by yellow lines. The white rectangle indicates the region of acoustical transition between vowels *i* and *a*. **b**, First and second formant frequencies during the transition from *i* to *a*: baseline (black), initial exposure (blue), end of training (red), and following unexpected removal of the field (green). Formant frequency trajectories for a single subject are shown. The curves give average values for individual blocks. **c**, No systematic acoustic effect is observed when F1 and F2 frequencies are examined on a per-subject basis.

(25% of the peak velocity of the transition). These points were chosen to ensure that the velocity of the vertical movement was sufficiently high to induce a significant mechanical perturbation due to the field. No significant difference was found between the four phases of the experiment at the three measurement points in F1 ($P > 0.05$) or in F2 ($P > 0.05$). Thus, the perturbation did not measurably alter the acoustic signals.

Figure 4c shows the pattern of acoustic effects for each subject separately. The individual data points give the formant frequencies at the peak velocity of the transition between vowels. It may be seen that there is no systematic pattern of acoustic changes associated with the force field. Significant differences in formant frequencies between the four phases of the experiment were observed in a small number of subjects (subject 6 for F1 frequency, and subjects 6 and 8 for F2, $P < 0.01$). In all but one case (F2 frequency for subject 8), values for baseline and initial exposure to the force field were comparable. An examination of formant frequencies on a per-subject basis was also undertaken at the start and end of the acoustic transition between vowels. As in the case of the peak velocity measure, no systematic acoustic effect was observed.

An additional acoustic analysis was carried out to examine the possibility of an acoustic effect on the very first trial in which the force field was applied. No difference was observed in F1 frequency ($P > 0.05$) or in F2 frequency ($P > 0.05$) between the first force-field trial and the remaining three experimental phases (baseline, end of training and after effect). These results further support the idea that the mechanical perturbation created by the force field was not large enough to alter measurably the acoustic signals.

Possible perceptible differences in the acoustics due to the presence of the force field were assessed using a perceptual discrimination task adapted from ref. 20. The results of these perceptual tests showed no ability of listeners to distinguish utterances produced in the force-field condition from those in the baseline condition. The average proportion of correct identification of utterances produced in the force-field condition was 0.54 (99% confidence interval, CI, ± 0.064). These tests were repeated by having the subjects of the vocalized speech condition make perceptual discrimination judgments on their own utterances. For these subjects, the average proportion correct was 0.49 (99% CI $\pm$ 0.063). There was thus no perceptible auditory cue that would distinguish utterances produced in force-field and null-field conditions.

In summary, we have examined the role of somatosensory information in speech production by applying velocity-dependent mechanical loads to the jaw that altered the movement path, and hence somatosensory feedback, without affecting the speech acoustics. We found that when speech acoustics were unaltered, or even absent altogether, subjects nevertheless adapted to the perturbation such that the motion path of the jaw approached that observed in the absence of load. The observation that adaptation in jaw movements occurs when speech acoustics are unaltered by the perturbation indicates that somatosensory information on its own is a principal component of the speech target. Changes to somatosensory input result in modifications to speech movements that restore the normal path of the jaw even when the acoustic goal is achieved. The similarity of movements following adaptation to those observed in the absence of load suggests that a precise pattern of somatosensory feedback related to the entire course of the movement is used to update the control signals underlying jaw motion in speech. Moreover, the non-speech conditions demonstrate that adaptation is not an inevitable consequence of training in a force field, that is, it is neither due to reflexes nor a consequence of the active force-generating properties of jaw muscles. Our findings indicate that in speech, a somatosensory goal is pursued independent of the acoustics. □

Methods

Experimental procedures

Mechanical perturbations were delivered to the jaw using a robotic device (Phantom 1.0, Sensable Technologies). The coupling between the robot and the jaw involved (1) an acrylic and metal dental appliance that was glued to the buccal surface of the teeth, and (2) a magnesium and titanium rotary connector that permitted motion of the jaw in all six translational and rotational degrees of freedom. The head was immobilized by connecting a second dental appliance, which was attached to the maxillary teeth, to a rigid metal frame that consisted of two articulated metal arms, one on each side of the head, that were locked in place during data collection.

The force vector, f, produced by the robot depended on the velocity vector of the jaw at the incisors, v, according to the following linear equation: $\mathbf{f} = B\cdot|\mathbf{v}|$, where B is a constant matrix representing viscosity in N s m^{-1}. Specifically, we used a force field defined by:

$$B = \begin{bmatrix} B_{xx} & B_{xy} \\ B_{yx} & B_{yy} \end{bmatrix} = \begin{bmatrix} 0 & 20 \\ 0 & 0 \end{bmatrix}$$

where B_{xx} represents force in the horizontal direction in proportion to horizontal velocity, and B_{xy} represents horizontal force in proportion to vertical velocity. Peak forces ranged from 4 to 5 N.

Twenty subjects were divided into two groups, a voicing group (8 subjects) and a non-voicing group (12 subjects). The non-voicing group was further separated into three conditions: a silent speech condition (4 subjects), a non-speech cyclical movement condition (4 subjects), and a non-speech discrete movement condition (4 subjects). Subjects in all conditions were instructed to keep movement amplitude and duration constant. The experimenter monitored a real-time display of movement parameters, and provided verbal feedback when amplitude or duration deviated from their initial values by more than ~20%. Subjects in the voicing group were also instructed to keep volume constant by monitoring values on a sound pressure level meter.

The experiment began with a familiarization phase with the field off (null field), in which subjects produced 30 repetitions (trials) of the utterance *siat* or the non-speech movement that was to be subsequently tested in the experiment. A baseline phase of 20 further null-field trials provided a reference movement path under unperturbed conditions. This was followed by a field-on training phase of 525 trials, after which the force field was unexpectedly removed and 30 further trials were collected. These final trials under null-field conditions assessed the possible presence of movement after-effects. Data analyses focused on jaw lowering alone, as this phase of movement was associated with potential acoustic effects in the vowel-to-vowel transition (*i-a*) due to the perturbation.

Subject performance was quantified by measuring maximum horizontal distance between perturbed movements and the average baseline path. The analyses were repeated using the horizontal distance between the perturbed movement at maximum lowering velocity and the baseline. The two yielded similar results. Statistical analyses using analysis of variance (ANOVA) were conducted for each subject separately. The analyses compared the maximum deviation from baseline movements in the first 20% of the training trials, the last 20% of the training trials, and the null-field trials following training (after-effect). Pair-wise comparisons of means were carried out using Tukey's method, where appropriate.

Acoustical analyses

The acoustic signal associated with the production of utterances in the vocalized speech condition was analogue low-pass-filtered at 10 kHz and sampled at 22.5 kHz. Acoustic formant tracking was carried out on the transition between the vowels *i* and *a*. For the purpose of subsequent analyses, values for the first and second formants (F1 and F2, respectively) were taken at 25% of the peak velocity of each formant transition (at the beginning and the end of the transition) and also at the point of peak velocity. Note that the acoustics associated with the vowel-to-vowel transitions were of particular interest as the force field depended on movement velocity and hence would have had the greatest effect during the transition between vowels. Statistical tests of potential acoustic effects due to the force field were conducted on a per-subject basis for the first and second formant frequencies separately using ANOVA.

Perceptual task

Six listeners were presented with randomly selected individual utterances recorded from null-field and initial force-field trials. Each test sequence comprised four utterances. The sequences were either of the form AABA or ABAA, where A represents an utterance randomly chosen from null-field trials in the baseline condition and B is an utterance randomly selected from initial force-field trials. The listeners were told that they would hear utterances that were recorded in two different conditions, and were required to indicate whether the second or the third differed from the other three. Listeners were presented with three blocks of 50 such sequences. Each block contained utterances selected from a single speaker.

The test was repeated using four subjects in the vocalized speech condition who were required to make perceptual judgments on their own utterances. In this case, two blocks of 50 sequences composed entirely of the listeners' own productions were used as stimuli. All other aspects of the procedure were identical to those described above.

Received 24 March; accepted 10 April 2003; doi:10.1038/nature01710.

1. Guenther, F. H., Hampson, M. & Johnson, D. A theoretical investigation of reference frames for the planning of speech movements. *Psychol. Rev.* 105, 611–633 (1998).
2. Guenther, F. H. *et al.* Articulatory tradeoffs reduce acoustic variability during American English /r/ production. *J. Acoust. Soc. Am.* 105, 2854–2865 (1999).

3. Houde, J. F. & Jordan, M. I. Sensorimotor adaptation in speech production. *Science* **279**, 1213–1216 (1998).
4. Jones, J. A. & Munhall, K. G. Perceptual calibration of F0 production: Evidence from feedback perturbation. *J. Acoust. Soc. Am.* **108**, 1246–1251 (2000).
5. Perkell, J. S., Matthies, M. L., Svirsky, M. A. & Jordan, M. I. Trading relations between tongue-body raising and lip rounding in production of the vowel /u/: A pilot "motor equivalence" study. *J. Acoust. Soc. Am.* **93**, 2948–2961 (1993).
6. Perkell, J. S. & Nelson, W. L. Variability in production of the vowels /i/ and /a/. *J. Acoust. Soc. Am.* **77**, 1889–1895 (1985).
7. Browman, C. P. & Goldstein, L. Articulatory phonology: An overview. *Phonetica* **49**, 155–180 (1992).
8. Guenther, F. H. Speech sound acquisition, coarticulation and rate effects in a neural network model of speech production. *Psychol. Rev.* **102**, 594–621 (1995).
9. Saltzman, E. L. & Munhall, K. G. A dynamical approach to gestural patterning in speech production. *Ecol. Psychol.* **1**, 333–382 (1989).
10. Hamlet, S. L. & Stone, M. L. Compensatory vowel characteristics resulting from the presence of different types of experimental dental protheses. *J. Phonet.* **4**, 199–218 (1976).
11. Lindblom, B., Lubker, J. & Gay, T. Formant frequencies of some fixed-mandible vowels and a model of speech motor programming by predictive simulation. *J. Phonet.* **7**, 147–161 (1979).
12. McFarland, D., Baum, S. R. & Chabot, C. Speech compensation to structural modifications of the oral cavity. *J. Acoust. Soc. Am.* **100**, 1093–1104 (1996).
13. McFarland, D. & Baum, S. R. Incomplete compensation to articulatory perturbation. *J. Acoust. Soc. Am.* **97**, 1865–1873 (1995).
14. Savariaux, C., Perrier, P. & Orliaguet, J. P. Compensation strategies for the perturbation of the rounded vowel [u] using a lip tube: A study of the control of space in speech production. *J. Acoust. Soc. Am.* **98**, 2428–2442 (1995).
15. Gandolfo, F., Mussa-Ivaldi, F. A. & Bizzi, I. Motor learning by field approximation. *Proc. Natl Acad. Sci. USA* **93**, 3843–3846 (1996).
16. Goodbody, S. J. & Wolpert, D. M. Temporal and amplitude generalization in motor learning. *J. Neurophysiol.* **79**, 1825–1838 (1998).
17. Lackner, J. R. & Dizio, P. Rapid adaptation to coriolis force perturbations of arm trajectory. *J. Neurophysiol.* **72**, 299–313 (1994).
18. Shadmehr, R. & Mussa-Ivaldi, F. A. Adaptive representation of dynamics during learning of a motor task. *J. Neurosci.* **14**, 3208–3224 (1994).
19. Thoroughman, K. A. & Shadmehr, R. Learning of action through adaptive combination of motor primitives. *Nature* **407**, 742–747 (2000).
20. Bernstein, L. R. & Trahiotis, C. Detection of interaural delay in high-frequency noise. *J. Acoust. Soc. Am.* **71**, 147–152 (1982).

Acknowledgements We thank G. Houle, M. Tiede and C. Dolan for technical support. This work was supported by the National Institute on Deafness and Other Communication Disorders, the Natural Sciences and Engineering Research Council of Canada, and Le Fonds pour La Formation de Chercheurs et l'Aide à la Recherche, Quebec.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to D.J.O. (ostry@motion.psych.mcgill.ca).