Evidence for the Use of Somatosensory Feedback in Speech Production

Stéphanie Tremblay† and David J. Ostry†‡
† McGill University, Canada
‡ Haskins Laboratories, United States
E-mail: steff@motion.psych.mcgill.ca, ostry@motion.psych.mcgill.ca

ABSTRACT

This paper explores the role of somatosensory input in the achievement of speech targets. In a previous study [3], we showed that somatosensory input, independent of acoustics, plays a fundamental role in determining speech movements. This conclusion was based on the finding that subjects corrected for the effects of a velocity dependent mechanical load that significantly alters jaw kinematics — and hence somatosensory feedback — but has no effect on acoustic output. However, in work to date, the patterns of adaptation that we have observed have been specific to movements involving a vowel-to-vowel transition. To further investigate the latter outcome, the present study explores patterns of adaptation by manipulating the location of the vowel-to-vowel transition within the speech utterance. The present results are consistent with the findings of the previous study: adaptation to a mechanical load is only achieved in portions of speech movements that are associated with a vowel-to-vowel transition.

1. INTRODUCTION

Acoustical information has been shown to be fundamental to speech production in a variety of ways. In comparison, the potential role of somatosensory information — that typically accompanies auditory input — has been neglected experimentally, but may similarly play a role in the achievement of speech goals. The determination of the components of speech targets requires the ability to dissociate the contribution of various sensory modalities. This approach has been used to identify the importance of auditory input. For example, Jones and Munhall (2000) [1] altered the voice fundamental frequency (F0) feedback without modifying somatosensory feedback. They showed that subjects gradually adjusted their F0 in order to adjust the acoustic feedback they received to their normal unaltered pitch. Houde and Jordan (1998) [2] used a different method to perturb acoustics without manipulating somatosensory feedback: they presented subjects with shifted formant frequencies of their own voices. At the end of the training phase, subjects altered their production of speech utterances in order to compensate for the shift and to hear the expected sound. These studies underscore the idea that speech production is substantially dependent upon auditory information. On the other hand, the capacity for intelligible speech by deaf speakers suggests that somatosensory feedback on its own may contribute to the achievement of speech targets.

In a previous study [3], we have succeeded in manipulating somatosensory feedback independent of speech acoustics and provided direct evidence that somatosensory input is central to the achievement of speech targets. We 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 systematically alter the motion path of the jaw, and hence somatosensory feedback, without affecting the associated acoustic output (Appendix A).

Our results showed that when mechanical perturbations which have no effect on speech acoustics are delivered to the jaw, adaptation in the movement path can be observed on the basis of somatosensory feedback alone (Figure 1 Panels A & B). That is, even when the acoustic goal is achieved, subjects modify their motor commands in order to reach somatosensory targets. Adaptation was not observed in matched non-speech movements (Figure 1 Panels C) even following extended practice. Moreover, subjects compensated for the perturbation only in portions of the movement that involved a vowel-to-vowel transition (Figure 1 Panels A & B). The present study explores the extent to which sensorimotor adaptation to mechanical perturbations is specific to the phonetic composition of the utterance and in particular to position of vowel-to-vowel transitions.

2. METHODS

Subjects

Eleven subjects were randomly assigned to one of the three experimental conditions. The subjects had no known speech, hearing or motor problems. Also, none of them had dental implants, protheses, crowns or temporomandibular joint dysfunction.

 ISBN 1-876346-48-5

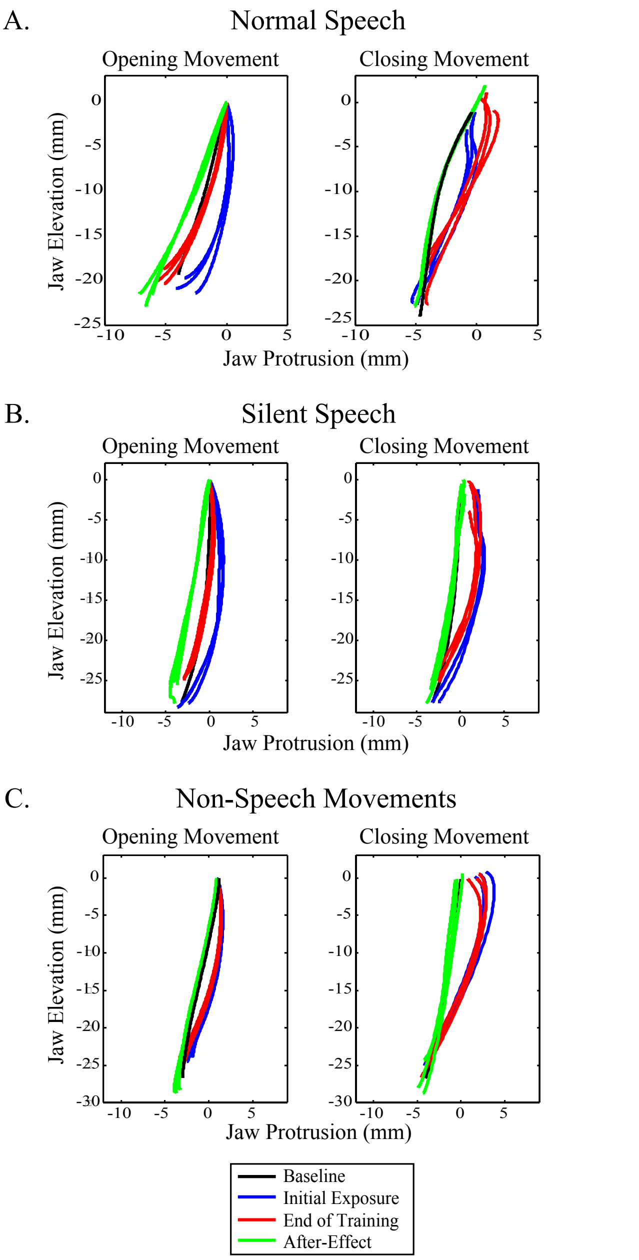

Figure 1: Sagittal plane jaw motion paths during the baseline condition (shown in black), initial exposure to the force-field (blue), at the end of training (red), and following unexpected removal of the field (green). The figure shows individual trials for single subjects. Left hand panels show the opening phase of the movements, whereas the right hand panels show the closing phase. Adaptation to the force-field is represented by both a restoration of the jaw path from first exposure to the force field to baseline movements under null field, and a subsequent motion dependent after-effect in the opposite direction. **A.** During vocalized speech, adaptation to the force-field is observed in the opening phase of the movement, but not in the closing phase. **B.** During silent speech, the pattern of adaptation observed in vocalized speech is unaltered by removal of acoustic feedback: adaptation is observed in the opening phase only. **C.** Matched non-speech movements show no adaptation in any phase of the movement.

Material and Perturbations

A robotic device (Sensable Technologies Phantom 1.0) was connected to the mandibular teeth and was used to deliver mechanical perturbations to the jaw (Figure 2). 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 — attached to the maxillary teeth — to a rigid metal frame, the head restraint.

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 (Figure 2B). The force vector, **f**, produced by the manipulandum depended on the velocity vector of the jaw at the incisors, **v**, according to the following linear equation: **f = B . |v|**, where **B** is a constant matrix representing viscosity in N.sec.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} relates force in the horizontal direction to horizontal velocity and B_{xy} relates horizontal force to vertical velocity. Peak forces ranged from 4 – 5 N.

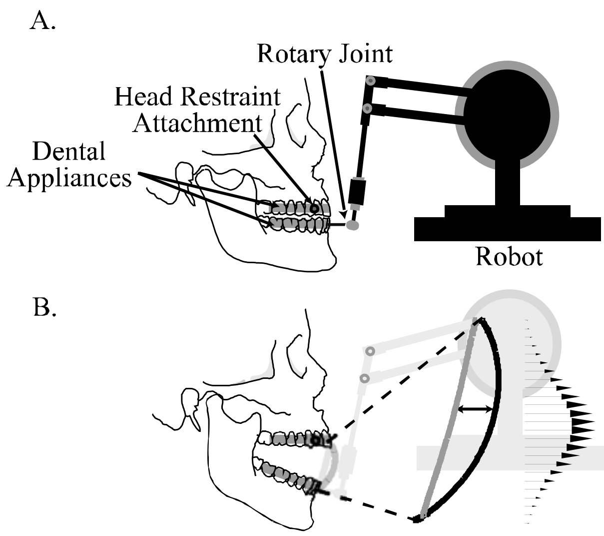

Figure 2: **A.** Schematic showing subject attached to the robotic device. **B.** Jaw opening movement with the force-field off (grey) and upon initial exposure to the field (black). Vectors depict the magnitude and direction of force applied by the robot over the course of the movement. The double headed-arrow shows the maximum horizontal deviation between null-field and force-field movements which served as a performance index.

ISBN 1-876346-48-5 © 2003 UAB

Tasks and Procedure

The four subjects of the first group were asked to repeatedly produce an utterance in which there was a vowel-to-vowel transition in the opening portion of the speech movement, namely “s-i-æ-s”. The second group (4 other subjects) produced an utterance that had a vowel-to-vowel transition in the closing portion of the speech movement, that is, “s-æ-i-s”. The last group (3 subjects) was trained with an utterance that involved the open vowel /æ/ but no vowel-to-vowel transition (“s-æ-s”). Every subject was required to produce the assigned utterance at his preferred rate and volume. The experimenter monitored a real-time display of movement parameters and provided verbal feedback when amplitude, duration or volume deviated from their initial values by more than ~20%.

The experiment began with a familiarization phase with the field off (null field) in which subjects produced 30 repetitions of the utterance that was to be subsequently tested in the experiment. A baseline phase of 20 further null field repetitions provided a reference movement path under unperturbed conditions. This was followed by a field-on training phase of 525 repetitions after which the force field was unexpectedly removed and 30 further repetitions were collected. These final utterance productions under null field conditions assessed the possible presence of movement after-effects. Performance was quantified for each subject by measuring on a repetition-by-repetition basis the maximum horizontal distance between the movement path under force-field conditions and the average baseline path. Deviation in the horizontal direction was used because it corresponded to the direction of the applied perturbation.

3. RESULTS

Figure 3 illustrates examples of movements generated from the production of the three utterances; *sias* is illustrated in the top panels, *sais* is found in the middle panels, and the bottom panels show the movements produced by *sas*. The left-hand side panels show the opening phase of the movements, whereas the right-hand side panels show the closing phase of the movements.

The black lines depict the average movement paths produced under null field condition before training (baseline). The jaw movements at initial exposure to the force-field are represented in blue. Compared to the baseline, movements at the start of training in the force-field deviated significantly in the protrusion direction for both opening and closing movements. There was no difference in the magnitude of the perturbation between the groups. Somatosensory feedback was thus initially altered in a comparable way in each of the three groups. At the end of training (shown in red), adaptation was achieved if the deviation from baseline was significantly reduced. In the top panels, it can be noted that the group which produced the utterance *sias* adapted to the force-field in the opening phase of the movements (s-i-æ), but not in the closing phase (æ-s). The opposite pattern was observed in the group that produced the utterance *sais*: subjects did not adapt in the opening phase of the movement (s-æ), but significantly reduced their movement deviation in the closing phase (æ-i-s). In the group that repeated the utterance *sas*, adaptation was not observed in the opening or the closing phase of the movements.

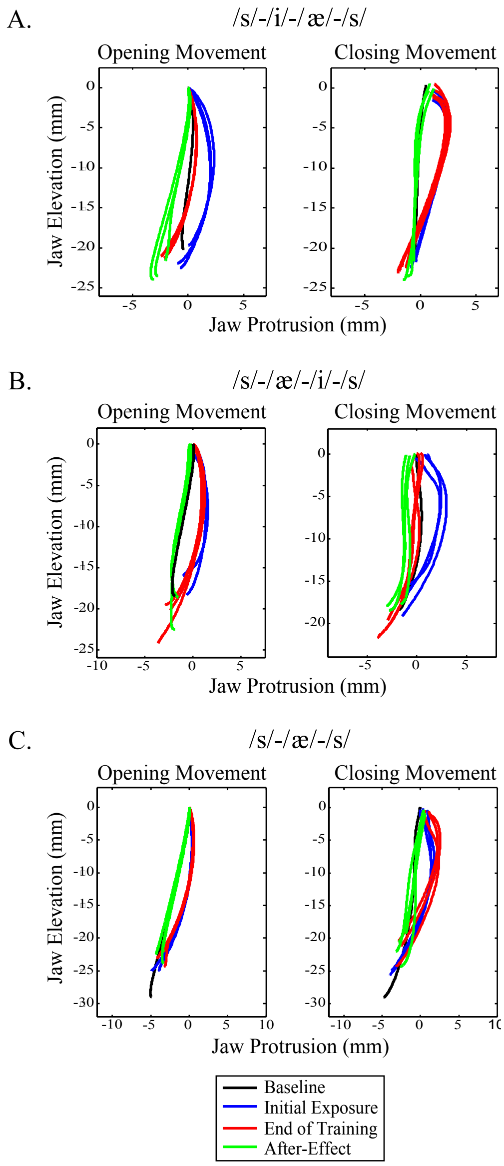

Figure 3: Sagittal plane jaw motion paths during the baseline condition (black), initial exposure to the force-field (blue), at the end of training (red), and following unexpected removal of the field (green). The figure shows individual trials for single subjects. Left hand panels show the opening phase of the movements, whereas the right hand panels show the closing phase. **A.** During the production of *sias*, adaptation to the force-field and a subsequent after-effect are observed in the opening phase of the movement, but not in the closing phase. **B.** During the production of *sais*, the pattern of adaptation and after-effect are observed in the closing phase of the movement only. **C.** Force field training with *sas* did not produce adaptation in either the opening or the closing phase of the movement.

ISBN 1-876346-48-5 © 2003 UAB

Following the unexpected removal of the field after the training, an after-effect (shown in green) in which the jaw was retracted in comparison to the baseline was observed wherever adaptation has been achieved: that is, in the opening phase of the *sias* movements and in the closing phase of the *sais* movements. In the third group (*sas*) as well as in the closing phase of *sias* and the opening phase of *sais,* the corresponding "after-effect" movements overlapped those in the baseline condition. The motion dependent after-effect seen in the portions of movements in which the subjects compensated for the perturbations reflects the changes in the motor commands to the jaw that enabled the "adapted" performance in the presence of the force field. These results suggest that the achievement of somatosensory targets in speech is particularly relevant in portions of utterances associated with a vowel-to-vowel transition.

4. CONCLUSIONS

The results obtained in this paper are consistent with the outcome of a previous study: subjects can adapt to a velocity dependent force field that alters somatosensory feedback, but not acoustic feedback, during the production of speech utterances. Adaptation indicates that the nervous system takes account of dynamics while planning speech movements. The presence of a motion dependent after-effect shows that subjects adjusted their motor commands to cancel out the effect of the field by applying a force, which is equal but opposite to the one delivered by the robot. Overall, this suggests that a somatosensory goal, independent of the acoustic goal, is pursued in speech production. The second finding is that subjects adapt to the force field only in the portion of the speech movement that held the vowel-to-vowel transition. The results highlight the importance of precision of movement during vowel-to-vowel transitions. Extensions to this procedure may offer a means to explore the relative precision requirements of somatosensory feedback during speech.

5. APPENDIX A

Acoustic analyses were performed to verify that the perturbations produced by the force-field did not result in alterations to the speech acoustics that could explain the observed adaptation in vowel-to-vowel transitions. Thus far, the first and second formant frequencies have been examined during the vowel-to-vowel transitions between i and æ of the utterance *siat*. Figure 4 shows time-normalized vowel-to-vowel transitions at different phases of the experiment. It may be seen that the formant frequency transitions were similar throughout the experiment. No significant difference was found between the four phases of the experiment in F1 or in F2 frequencies. Thus, adaptation observed in the vowel-to-vowel transition of the speech movements cannot be explained by the presence of altered acoustic feedback.

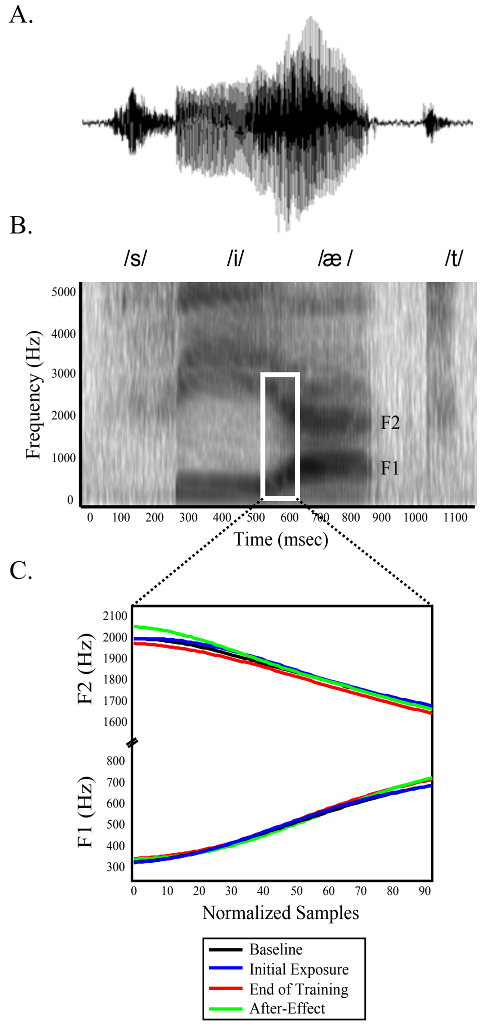

Figure 4: A. Raw acoustic signal during a single repetition of the utterance *siat*. **B.** Spectrogram of the signal shown in panel A. The dark bands show the frequency composition of acoustical energy. The white rectangle indicates the region of acoustical transition between vowels i and a. **C.** 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.

REFERENCES

[1] J.A. Jones and K.G. Munhall, "Perceptual Calibration of ú0 Production: Evidence from Feedback Perturbation", *J. Acous. Soc. Am.*, vol. 108, pp. 1246-1251, 2000.

[2] J.F. Houde and M.I. Jordan, "Sensorimotor Adaptation in Speech Production," *Science*, vol. 279, pp. 1213-1216, 1998.

[3] S. Tremblay, D.M. Shiller and D.J. Ostry, "Somatosensory Basis of Speech Production," *Nature,* in press.