

Specificity of Speech Motor Learning

Stéphanie Tremblay,¹ Guillaume Houle,¹ and David J. Ostry^{1,2}

¹Department of Psychology, McGill University, Montreal, Québec, Canada H3A 1B1, and ²Haskins Laboratories, New Haven, Connecticut 06511

The idea that the brain controls movement using a neural representation of limb dynamics has been a dominant hypothesis in motor control research for well over a decade. Speech movements offer an unusual opportunity to test this proposal by means of an examination of transfer of learning between utterances that are to varying degrees matched on kinematics. If speech learning results in a generalizable dynamics representation, then, at the least, learning should transfer when similar movements are embedded in phonetically distinct utterances. We tested this idea using three different pairs of training and transfer utterances that substantially overlap kinematically. We find that, with these stimuli, speech learning is highly contextually sensitive and fails to transfer even to utterances that involve very similar movements. Speech learning appears to be extremely local, and the specificity of learning is incompatible with the idea that speech control involves a generalized dynamics representation.

Key words: dynamics; generalization; adaptation; jaw; motor learning; speech

Introduction

As a child learns to talk and to produce the sounds of his native language, he must master some of the smallest and most temporally precise movements that humans are capable of making. What is it that the child learns that permits such precise control of articulation, and what enables speech learning to generalize? One possibility is that learning to talk involves the acquisition of a neural representation of vocal tract dynamics that can be applied broadly when a child encounters novel utterances. Indeed, the idea that a neural dynamics representation forms the basis of generalization of learning has been proposed in the context of other movements (Shadmehr and Mussa-Ivaldi, 1994; Wolpert et al., 1995; Flanagan and Wing, 1997). If speech learning involves the acquisition of a broadly generalizable dynamics representation, then, at the least, learning should transfer when similar movements are embedded in a different task. Speech motor control offers a unique window on transfer of learning because it is possible to assess transfer between phonetically distinct utterances that are matched in varying degrees in terms of their kinematics. For the three pairs of utterances that we tested here, motor learning is highly specific, and learning fails to transfer even to utterances that involve similar movements. The failure to observe any transfer of learning between movements that are closely matched in terms of their kinematics is consistent with the idea that speech learning is extremely local and is incompatible with the notion that speech motor function involves a generalized dynamics representation.

In work to date on dynamics learning, evidence for generalization is, for the most part, limited to situations in which there is a close match between movements in training and transfer conditions. Generalization is found to be consistently present when movement direction or joint rotations are similar in training and transfer conditions (see below). However, there is otherwise relatively little indication that learning generalizes broadly, and indeed the general rule seems to be that dynamics learning is spatially local. The capacity for broad generalization seems to depend on multiple instances of local learning and a process of interpolation between individual examples (Atkeson, 1989; Gandolfo et al., 1996; Ghahramani and Wolpert, 1997; Malfait et al. 2005; Mattar and Ostry, 2007a,b).

Instances of generalization of dynamics learning have been described for interlimb movements, typically when movements, equivalent to those in the training condition, are repeated with the contralateral limb (Dizio and Lackner, 1995; Criscimagna-Hemminger et al., 2003; Malfait and Ostry, 2004; Wang and Sainburg, 2004). Generalization for dynamics learning is also observed to movements that differ in amplitude and duration, but this is in the context of movements in a single direction (Goodbody and Wolpert, 1998). Similarly, generalization is observed when training and transfer movements involve the same joint rotations or joint torques and are in effect the same movements, in a different part of the workspace (Shadmehr and Mussa-Ivaldi, 1994; Shadmehr and Moussavi 2000; Malfait and Ostry, 2002; Hwang et al., 2003). Examples of more extensive generalization are found in trial-to-trial effects (Thoroughman and Shadmehr, 2000; Donchin et al., 2003; Thoroughman and Taylor, 2005), in visuomotor learning (Krakauer et al., 2000) and in the generation of movement sequences (Conditt et al., 1997). These latter cases are reviewed in Discussion.

In dynamics learning, there is considerable evidence that, when training and transfer movements differ to any substantial degree, and certainly when they differ in direction, there is little generalization of learning, suggesting, in effect, that motor learning is local. Thus, movements that are learned in one part of the

Received Jan. 29, 2007; revised Jan. 8, 2008; accepted Jan. 9, 2008.

This work was supported by the National Institute on Deafness and Other Communication Disorders Grant DC-04669, Natural Sciences and Engineering Research Council, Canada, and Fonds Québécois de la Recherche sur la Nature et les Technologies, Québec.

Correspondence should be addressed to Dr. David J. Ostry, Department of Psychology, McGill University, 1205 Dr. Penfield Avenue, Montreal, Québec, Canada H3A 1B1. E-mail: ostry@motion.psych.mcgill.ca.

DOI:10.1523/JNEUROSCI.4196-07.2008

Copyright © 2008 Society for Neuroscience 0270-6474/08/282426-09\$15.00/0

workspace can be repeated accurately only within a limited spatial range about the original training direction (Sainburg et al., 1999; Malfait et al., 2005). Grip-force modulation shows limited transfer to directions other than those on which subjects were specifically trained (Witney and Wolpert, 2003). There is, in addition, a rapid decrease in the magnitude of force-field aftereffects with increasing distance from the training direction such that relatively little transfer of learning is observed beyond $\pm 45^\circ$ (Gandolfo et al., 1996). Moreover, unlike visuomotor learning, the shape of the generalization gradient in dynamics learning is not modifiable even when subjects are permitted to extensively explore the workspace over several hundred trials (Mattar and Ostry, 2007b).

Additional examples of the specificity of learning are provided by demonstrations of context sensitivity. For example, subjects can simultaneously learn opposite force patterns if they are appropriately distinguished by contextual cues (Gandolfo et al., 1996; Osu et al., 2004; Waincott et al., 2005). Similarly, a movement that is learned in one context may fail to transfer to a novel context even when the movement itself and the force environment are unchanged (Cothros et al., 2006; Nozaki et al., 2006). There is similar evidence of specificity in speech motor learning. Subjects have been observed to adapt to a novel dynamic environment during vocalized speech and silent speech but do not adapt during nonspeech movements that are matched for amplitude and duration (Tremblay et al., 2003). Here we describe an additional instance of specificity in learning, which, although based on a small set of utterances, is nevertheless difficult to reconcile with the idea that speech learning is mediated by any kind of broadly generalizable representation of dynamics. We show that, for these utterances, when training and transfer materials are equated to varying degrees on kinematics, in no case is there transfer from one to another.

Subjects in these experiments learned to talk in an altered dynamic environment in which a robot applied mechanical loads to the jaw in proportion to jaw opening or closing velocity. After training with one utterance, subjects were asked to produce a test utterance to assess transfer of learning. The training and test utterances were matched on a continuum that extended from equating jaw movements alone to creating a match across movements of the jaw and tongue. If dynamics learning in speech production generalizes from training utterances to transfer utterances, performance in the transfer test should be comparable with performance at the end of training.

Materials and Methods

Subjects and test utterances. Twenty-two subjects were assigned to one of three experiments in which tongue and jaw movements were matched to varying degrees between training and transfer utterances. In experiment 1, six subjects were trained to produce the utterance /sɪæs/ (“see-ass”) while the robot applied forces to the jaw. In *sɪæs*, the initial and final sibilant consonants impose precise jaw start and end points for each lowering/raising cycle. The transition from a high vowel (/i/) to a low vowel (/æ/) provides for large-amplitude movements. The associated high movement velocities result in forces that produce substantial deflection of the jaw. After adaptation, the subjects were tested for transfer of learning with the utterance /suæs/ (“sue-ass”). *Sɪæs* and *suæs* were selected because they are matched on the basis of jaw movements, whereas the tongue movements differ. In experiment 2, four subjects were trained with the utterance *sɪæs* and four others with the utterance /sæ/ [“sæ (as in sass)–ees (as in lease)”). All eight subjects were tested for transfer of learning with the utterance /sɪæis/ (“see-æ-ees”), such that part of the training and the test utterances, either the opening (*sɪæ*) or the closing (*æis*), required similar tongue and jaw movement, whereas the other half

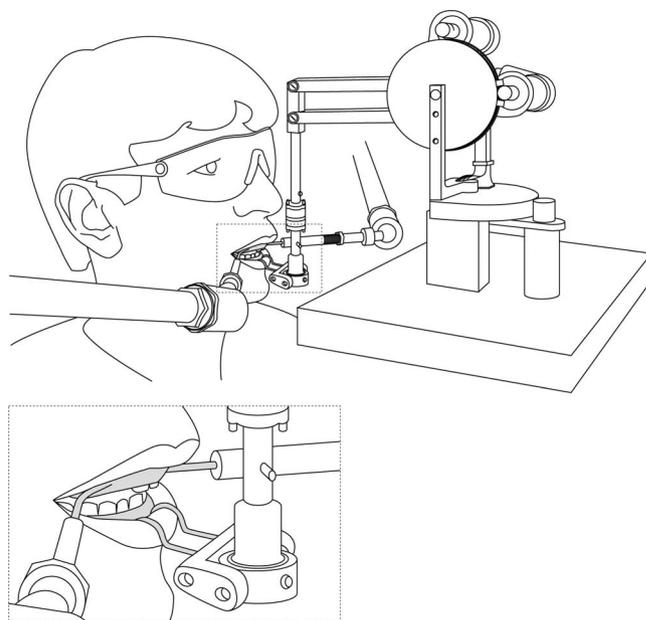


Figure 1. Schematic of the experimental setup. The subject is connected to a robotic manipulator by a dental appliance that is coupled to a rotary connector. The head is restrained during testing by connecting a similar dental appliance on the maxillary teeth to a set of articulated metal arms.

only shared jaw movements. In experiment 3, eight subjects were trained and tested with the utterance *sɪæs*; half of the subjects produced the utterance silently during the training phase and aloud during the test phase, whereas for the other half the order was reversed, that is, they trained with vocalized speech and were tested for transfer with silent speech. In this study, both tongue and jaw movements were similar in the training and the test phases.

Before testing, subjects in each study were given oral examples of the training and test utterances. They were asked to produce each utterance at a normal speech rate and volume and to place equal stress on each vowel. Subjects practiced each utterance with feedback before testing began. During testing, the experimenter monitored a real-time display of movement amplitude and velocity and provided verbal feedback to help subjects keep these constant. The purpose of the feedback was to equate movements in the training and transfer conditions.

Experimental manipulation. The motor task to be learned in the training phase was the same as the one described in a previous study (Tremblay et al., 2003). It involved adaptation of the orofacial apparatus to an artificial mechanical force field, delivered to the jaw by a robotic device (Phantom 1.0; Sensable Technologies, Woburn, MA). Specifically, the field acted to displace the jaw in the protrusion direction, along an axis orthogonal to jaw opening and closing movements, with a magnitude proportional to the instantaneous velocity of the jaw measured at the incisors. The force field was applied in the midsagittal plane. 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: $f = Bv$, where $B = (0, 20; 0, 0)$ N/s/m $^{-1}$. Peak forces ranged from 4 to 5 N.

The robot was coupled to the jaw by means of an acrylic and metal dental appliance that was glued with a dental adhesive to the buccal surface of the mandibular teeth (Fig. 1). The appliance was attached to a magnesium and titanium rotary connector that allowed unrestricted movement of the jaw in all degrees of freedom. The head was immobilized by connecting a second dental appliance that was on the maxillary teeth to a rigid metal frame that consisted of two articulated metal arms that were locked in place during data collection.

The testing was performed in blocks of trials in which subjects repeated an utterance 15 times at a normal rate and volume. Subjects were not told about the presence or the characteristics of the load. The experiment began with four blocks produced in the absence of load: two blocks

of the training utterance alternated with two of the test utterance. The first two blocks were for familiarization, whereas the last two comprised the baseline movements that were used in subsequent analyses. Then, immediately before the training phase of the experiment, a single block of the transfer utterance was collected with the field on to enable a comparison of the movement path of the test utterance before and after training. The main body of the experiment was performed under force-field conditions. This consisted of a training phase of 450 repetitions of the training utterance and of a test phase of 15 repetitions of the transfer utterance. Tests for possible aftereffects were performed immediately afterward. To test for aftereffects related to the training utterance, subjects produced five additional blocks of training trials (75 repetitions of the training utterance) with the force field on, followed by a one additional block in which the load was unexpectedly removed. Tests for possible transfer of aftereffects related to the transfer utterance were performed by having subjects produce another five blocks of training trials (force field on), followed by a one block of transfer utterances in which the load was absent.

Signal processing and statistical analysis. The kinematic data were recorded at 1 kHz and digitally low-pass filtered at 12 Hz. Performance was quantified for each subject by measuring on a trial-by-trial basis the maximum perpendicular distance between each movement path under force-field conditions and the average baseline path for the same utterance. Measures of perpendicular distance were then normalized for each subject separately, again on a trial-by-trial basis, with respect to the mean perpendicular distance in the first block of trials on initial exposure to the force field. The normalization was performed separately for training and test utterances. That is, deviation measures for the training utterance were divided by the mean initial exposure values for the first block of the training utterance, and deviation measures for the test utterance were divided by mean initial exposure values for the first block of the test utterance. The normalization accounted for differences in movement curvature between training and test materials.

The acoustical data were analog low-pass filtered at 5 kHz and digitally recorded at 10 kHz. A formant tracking algorithm coded in Matlab (MathWorks, Natick, MA) was used to obtain spectral measures during the voiced portion of each utterance. A 30 ms window associated with steady-state production of each initial and final vowel was manually selected to obtain estimates of first and second formant frequencies.

Statistical analyses of the kinematic data were conducted using repeated-measures ANOVA for both subjects who were successful in learning the task and separately for subjects who failed to show adaptation. The analyses compared the distance from baseline at the start of the training (averaged over the first five blocks), at the end of training (last five blocks), and for trials involving the transfer utterance also before and after training. Tests for transfer of learning and for aftereffects were based on the first five utterances in each condition (similar results were obtained using three to five trials). Bonferroni's corrected pairwise comparisons were performed to test for differences between individual means. Statistical analyses of first and second formant frequencies were conducted over the same phases of each experiment, also using repeated-measures ANOVA and Bonferroni's corrected pairwise tests.

Matching training and transfer utterances. We assessed the similarity of movements of training and transfer utterances in a number of ways. First, for each of the three experiments, we compared jaw movements that had been recorded by the robot under null field conditions, at the end of training (with the force field on), and in tests of transfer of learning (also with the force field on). The analysis assessed differences in initial jaw position, jaw movement amplitude, and maximum velocity of movements. We assessed differences using repeated-measures ANOVA followed by Bonferroni's corrected pairwise comparisons.

We also compared the kinematics of the movements that we used in the transfer task with jaw kinematics during force-field adaptation. The aim of this analysis was to assess the degree to which the jaw positions and velocities that were encountered in the transfer task had been previously experienced during training. We computed overlap as the proportion of the transfer utterance area (mean \pm 2 SD) that subjects experienced during training movements (mean \pm 2 SD). We used all trials from the transfer utterance block (force field on) and all training trials (force field

on) to conduct these analyses. We repeated this analysis for vertical and horizontal jaw positions, and vertical and horizontal velocity for each of the three experiments on a per-subject basis.

In addition, we performed a control study involving six new subjects in which we used an electromagnetic motion-tracking device [ElectroMagnetic Articulograph (EMA) AG 500; Carstens Medizintechnik, Lenggern, Germany] to record and compare the trajectories of jaw and tongue movements in the absence of load for each of the training and transfer conditions that we tested. Subjects in this study produced each of the utterances used in the tests conducted with the robot. The study was performed in blocks of 10 trials. Subjects produced 30 trials of each utterance in total.

In the magnetometer control study, tongue and jaw movements were recorded in three dimensions using small (3–4 mm), light-weight receiver coils that were glued to the teeth and the tongue. Three coils on the maxilla were used to correct for head motion and defined a coordinate system with an origin at the tip of the maxillary incisors and a horizontal axis parallel to the occlusal plane. The coils on the maxilla were placed between the central incisors at the level of the gingiva and, at both the left and the right sides of the mouth, between the second premolar and the first molar tooth, also at the level of the gingiva. Speech movements were recorded using two coils on the mandible, one on the tongue dorsum and one on the tongue tip. The coils on the mandible were placed symmetrically at the left and the right between the lateral incisor and the canine (at the gingiva). The tongue tip coil was placed on the midline 1 cm back from the tip; the tongue dorsum marker was 3 cm posterior to the tongue tip marker. All movements were recorded at 200 Hz and low-pass filtered at 12 Hz.

Results

Transfer of speech learning

We selected our speech materials to provide a range of kinematic overlap between our training and test utterances. The goal was to match training and transfer movements in three ways, one involving jaw movements alone, one involving tongue and jaw movement in a restricted portion of an utterance, and one involving similar movement of the tongue and jaw throughout. In the jaw-matched condition (experiment 1), subjects were trained to produce the utterance *siæs* while the robot applied forces to the jaw. After adaptation, the subjects were tested for transfer of learning with the utterance *suæs*. *Siæs* and *suæs* were selected because they have similar jaw movement amplitudes, whereas the tongue movements differ. In the restricted-match condition (experiment 2), subjects were trained with either the utterance *siæs* or the utterance *sæis* and were then tested for transfer of learning with the utterance *siæis*. Subjects that were trained with *siæs* had tongue and jaw movements that overlapped those of the test utterance in the opening phase; subjects that trained with *sæis* were similar to the test utterance during closing. In the global-match condition (experiment 3), subjects were trained and tested with the utterance *siæs*; half of the subjects produced the utterance silently during the training phase and aloud during the test phase, whereas the other half trained with vocalized speech and were tested for transfer with silent speech. For these subjects, the goal was to have similar patterns of tongue and jaw movements in the training and the test phases.

Figure 2 displays records of typical jaw movements for the training and transfer utterances in each of the three experiments. *A* shows jaw lowering movements for training with *siæs* and transfer to *suæs* (jaw-match condition), *B* gives training with *siæs* and transfer to *siæis* (restricted-match), and *C* shows “*siæs* vocalized” and “*siæs* silent” (global-match) as training and transfer utterances. The left side of each panel is for the training utterance, and the right side shows the transfer utterance. The black lines give the baseline condition. The blue and red lines show perfor-

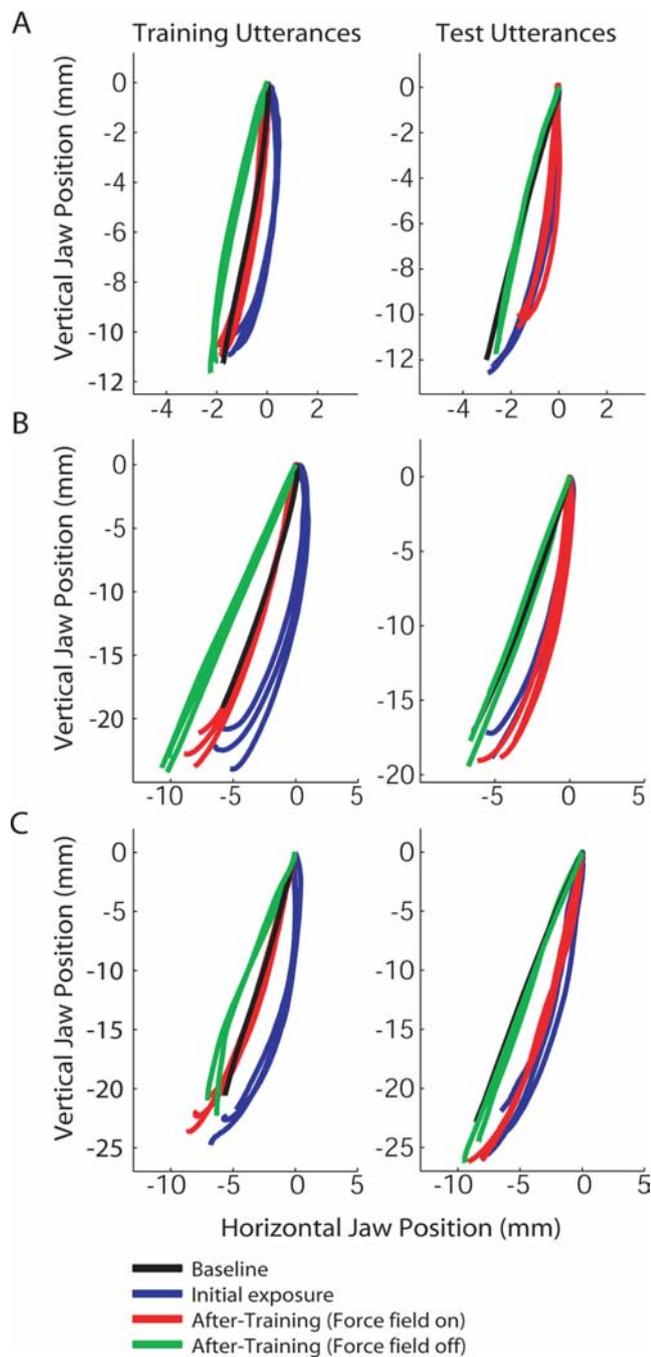


Figure 2. Jaw motion paths in the sagittal plane for individual subjects. The left column shows jaw movements for the training utterances; the right column shows movements for the test utterances. The black paths represent the baseline condition, the blue paths show individual movements during initial exposure to the force field, and the red lines correspond to individual motion paths after training. The green paths give jaw movements at the end of training, after unexpected removal of the field. **A**, Experiment 1. In the jaw-matched condition, adaptation to the force field observed at the end of training does not transfer to the test utterance. **B**, Experiment 2. In the restricted-match condition, training and test utterances had similar jaw and tongue lowering movements: adaptation was observed, but there was no transfer of learning. **C**, Experiment 3. Even when the training and test utterances were globally matched, the learning achieved during training does not transfer to the test phase of the experiment. A small aftereffect is observed for the training utterance (left) but not with the test utterance (right).

mance at the start and end of learning. The green lines show movements after training, after unexpected removal of load (aftereffect trials).

The blue paths show that initial exposure to the force field had

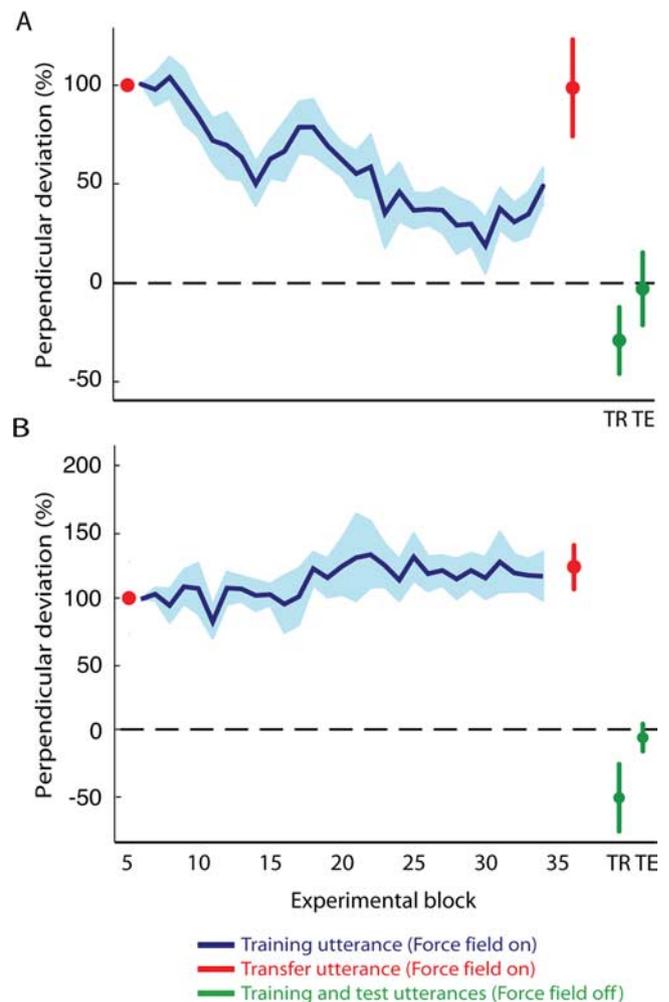


Figure 3. Learning curves showing training and transfer trials for subjects who showed adaptation (**A**) and those who did not (**B**). Mean values (\pm SE) are shown on a between-subjects basis over a series of 15 trial blocks. Deviation measures are normalized to the deviation observed at the initial introduction of load. Each point on the learning curves corresponds to average values of the normalized maximum perpendicular deviation of all 15 movements of a given block. The separate red data points at the beginning and end of each curve give performance for the transfer utterance before and after training. The green data points labeled TR and TE show performance for the training and test utterances after unexpected removal of load (\pm SE). The points without error bars at the left of each panel give normalized deviation scores for initial exposure to transfer and test utterances, respectively. In the test for transfer of learning at the end of training in **A**, it can be seen that the test utterance is significantly more deviated than the training utterance.

similar effects on the jaw movement path for both the training and transfer utterances. The red paths demonstrate that subjects adapted to the perturbation in each of the three experiments (left panels) but that there was no transfer of learning (right panels), regardless of the degree to which movements in the training and the transfer utterances were matched. The green lines show that unexpected removal of load after training produces a slight aftereffect for the training utterance (left) and no aftereffect for the transfer utterance (right).

Performance was quantified for each subject separately by measuring, on a trial-by-trial basis, the maximum perpendicular distance between each movement path under force-field conditions and the average baseline path for the same utterance as measured in the absence of load. Figure 3 shows these deviation or curvature measures relative to the perpendicular distance on initial exposure to the force field at the beginning of training (see

Materials and Methods). Figure 3A gives the learning curve (\pm SE) pooled over the three experiments (for the 13 subjects of 22 who showed adaptation). Figure 3B shows the curve for nine subjects who failed to adapt. The separate red data points at the two ends of each curve show performance for the transfer utterance on initial exposure to the force field (left side) and in the test for transfer of learning at the end of the experiment (right). The green data points (\pm SE) labeled TR for training and TE for test show performance after unexpected removal of load (aftereffects).

We assessed differences in normalized movement curvature (perpendicular distance) using repeated-measures ANOVA and Bonferroni's corrected pairwise comparisons. Consistent with the patterns seen in Figure 3A, an overall ANOVA found reliable differences in normalized curvature over the course of testing ($F_{(7,70)} = 32.70$; $p < 0.001$). Mean curvature was similar in the three experiments ($F_{(2,10)} = 0.67$; $p > 0.5$), and there was no statistical interaction ($F_{(14,70)} = 0.51$; $p > 0.5$); that is, the patterns of curvature over the course of training and test measures were similar for the three different sets of utterances. For the subjects in Figure 3B who failed to adapt, there were no statistical differences in curvature between the start and end of training ($p > 0.5$) nor between the end of training and tests for transfer of learning ($p > 0.5$). Moreover, performance on the transfer test was no different for subjects who adapted to the load and those who failed to adapt (between-subjects t test, $t_{(18)} = 1.057$; $p > 0.30$).

Bonferroni's comparisons showed that, for the subjects in Figure 3A, there was adaptation to the force field but no transfer of learning. Specifically, on initial exposure to the force field, the jaw path was displaced relative to the baseline ($p < 0.001$). By the end of training, jaw movements for the training utterance were reliably less deviated than at the beginning ($p < 0.01$). After unexpected removal of the load, the jaw trajectory was deviated in a direction opposite to the force field. The resulting curvature was marginally negative, that is, curved in the opposite direction (95% confidence interval, -0.67 to -0.03). However, the observed value was not different than null-field curvature based on a Bonferroni's pairwise comparison ($p > 0.5$). Three results suggest that, despite reliable adaptation, there was no transfer of learning: jaw movements in the transfer test were significantly more deviated than jaw movements at the end of the training phase ($p < 0.05$), jaw deviation for the transfer test utterance was no different after training than before ($p > 0.4$), and there was no aftereffect for the test utterance; after unexpected removal of load, jaw movements for the transfer utterance did not differ from the baseline path ($p > 0.5$; 95% confidence interval for test utterance aftereffect, -0.44 to 0.20). We also verified that performance at the end of each of the retraining blocks, that is, before aftereffect testing, was no different than adapted performance at the end of the first training phase ($F_{(2,20)} = 1.39$; $p > 0.25$). Thus, the learning that occurred for the training utterance did not transfer to the test utterance.

A series of acoustical analyses was performed to determine the degree to which speech acoustical patterns were altered by the presence of the force field or in response to its unexpected removal. We separately assessed changes in first and second formant frequencies (F1 and F2) at both the initial vowel and at the final vowel in *siæs* and *sæis*. The tests were performed using repeated-measures ANOVA and also using ANOVA on a per-subject basis. Figure 4 shows an example of the raw acoustical signal during the production of *siæs* under null-field conditions and the associated speech spectrogram (Fig. 4A). The spectro-

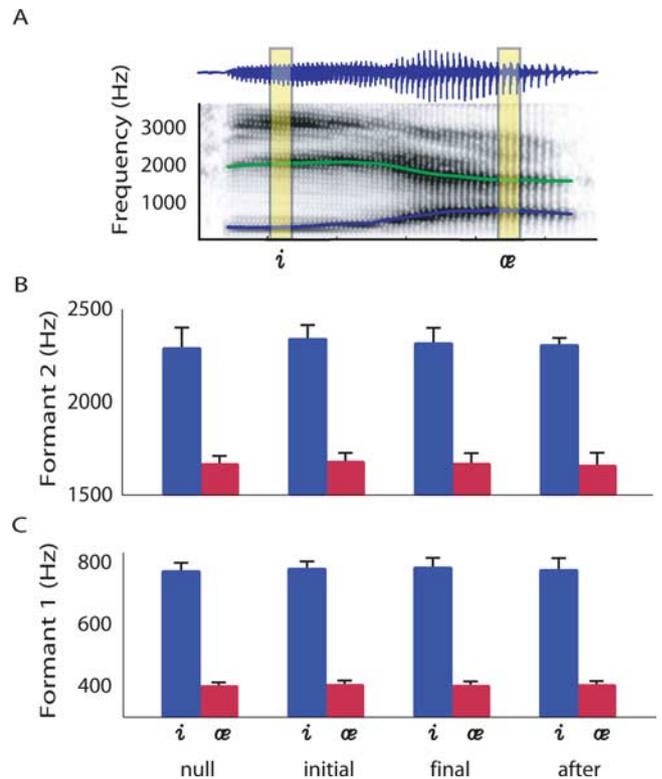


Figure 4. Acoustical analysis of first and second formant frequencies over the course of training. **A**, An example of the raw acoustical signal during the production of *siæs* under null-field conditions. The associated speech spectrogram and first and second formant tracks for the utterance are shown. The highlighted vertical panels from the steady-state production of the initial and final vowels were used to obtain estimates of first and second formant frequencies. **B**, Mean first formant frequencies (\pm SE) for subjects who showed statistically reliable adaptation with the utterance *siæs* in the vocalized speech condition. **C**, Mean second formant frequencies (\pm SE) for the same subjects.

gram shows the first and second formant tracks obtained from the voiced portion of the utterance. The highlighted vertical bars show the intervals used to obtain the first and second formant frequencies. Figure 4, B and C, shows mean formant frequencies (\pm SE across subjects) at different phases of the experiment for the seven subjects that trained with repetitions of *siæs* in the vocalized speech condition and also showed statistically reliable adaptation. As can be seen from a visual inspection of the figure, formant frequencies were similar throughout the experiment. Indeed, there were no systematic statistical effects related to the force field. Mean formant frequencies were no different for null-field trials, the first five and the last five training blocks and the first five utterances after unexpected removal of the force-field ($F_{(3,21)} = 0.36$; $p > 0.5$). As expected, values for F2 were greater than for F1, and steady-state values for /i/ were different than for /æ/ ($F_{(1,7)} = 77.70$; $p < 0.001$). For /i/, mean F1 and F2 values (\pm SE) were 396 ± 14 and 2265 ± 86 Hz, respectively. For /æ/, mean values were 751 ± 33 and 1635 ± 56 Hz, respectively.

We conducted ANOVA on the acoustical data on a per-subject basis for both subjects who showed adaptation and those who did not. These analyses compared null-field formant frequencies (from the second block of null-field trials), formant frequencies at the start of the training (averaged over the first five blocks), and formants at the end of training (last five blocks). Aftereffect trials were not included in the within-subjects tests because of the small number of trials involved. Significant differences in formant frequencies between these three phases of the

Table 1. Jaw movement kinematics for training and test utterances

	Null TR	Null TE	TR end	TE
<i>siaes</i> – <i>suæs</i>				
X	0.00 ± 0.00	–0.03 ± 0.09	0.29 ± 0.58	0.21 ± 0.69
Y	0.00 ± 0.00	–0.27 ± 0.26	–1.50 ± 2.23	–2.37 ± 3.81
Amp	16.40 ± 3.61	15.13 ± 3.82	15.61 ± 3.94	13.81 ± 3.53
Vel	161.49 ± 33.69	163.06 ± 34.46	148.80 ± 34.51	149.33 ± 40.10
<i>siaes</i> silent – <i>siaes</i> vocalized				
X	0.00 ± 0.00	0.30 ± 0.20	–0.72 ± 0.43	–0.74 ± 0.43
Y	0.00 ± 0.00	0.20 ± 0.28	0.25 ± 0.52	0.53 ± 0.66
Amp	23.53 ± 1.64	23.48 ± 0.96	24.73 ± 0.92	23.92 ± 1.50
Vel	220.57 ± 17.38	220.54 ± 14.79	222.46 ± 18.67	225.10 ± 16.80
<i>siaes</i> – <i>siaeis</i>				
X	0.00 ± 0.00	0.30 ± 0.48	1.30 ± 0.43	0.66 ± 0.34
Y	0.00 ± 0.00	0.02 ± 0.14	–1.70 ± 1.32	–2.13 ± 0.36
Amp	22.24 ± 2.20	23.78 ± 3.74	21.15 ± 2.29	21.61 ± 3.24
Vel	213.03 ± 30.54	219.89 ± 39.69	199.48 ± 26.52	202.47 ± 29.88

Mean values across subjects (\pm SE) are given for initial horizontal position (X; in millimeters), initial vertical position (Y; in millimeters), movement amplitude (Amp; in millimeters), and maximum tangential velocity (Vel; in millimeters per second). Values are given for null-field trials for training utterances (Null TR) and test utterances (Null TE), for training utterances at the end of training (TR End), and for test utterances in tests for transfer of training with the force field on (TE). Initial horizontal and vertical positions are expressed relative to values obtained under null-field conditions with training utterances.

experiment were obtained in some cases using Tukey's *post hoc* tests (at either $p < 0.05$ or $p < 0.01$), but there was no systematic pattern of acoustical change associated with the introduction of load or over the course of adaptation to the force field. For each subject, four possible acoustical effects were examined, that is, possible effects on F1 and F2 at the initial vowel and on F1 and F2 at the final vowel. In total, for the subjects who showed adaptation, we obtained reliable increases in either F1 or F2 frequency with the introduction of load in six cases, reliable decreases in four cases, and no change in 22. For the same subjects, a comparison of formant frequencies at the start and end of training showed 13 reliable increases in F1 or F2 frequency, 8 reliable decreases, and 11 cases in which there was no significant acoustical change over the course of training.

A comparable examination of possible within-subject acoustical effects in subjects who did not show adaptation produced a similarly variable pattern. The introduction of load resulted in a reliable increase in F1 or F2 frequency in two cases, a reliable decrease in one, and no change in nine cases. Over the course of training, these same subjects showed a reliable increase in F1 or F2 frequency in two cases, reliable decreases in four, and no change in six.

Similarity of training and transfer utterances

We used repeated-measures ANOVA and Bonferroni's corrected pairwise comparisons to assess differences recorded by the robot in initial jaw position, movement amplitude, and maximum velocity in each of the three experiments. We examined differences in training and transfer utterances under null-field conditions, at the end of the training phase and in the test for transfer of learning. The mean values for these variables over the course of the experiment are given in Table 1.

ANOVA indicated no reliable differences in initial jaw position over the various phases of the experiment ($F_{(3,30)} = 0.44$, $p > 0.5$; $F_{(3,30)} = 0.77$, $p > 0.5$, for horizontal and vertical position, respectively). There was no interaction for either initial horizontal position or initial vertical position between the phases of testing and the three different sets of training and tests utterances ($F_{(6,30)} = 2.26$, $p > 0.05$; $F_{(6,30)} = 0.37$, $p > 0.5$, for horizontal and vertical position, respectively).

There was likewise no dependence of jaw movement amplitude or jaw maximum velocity on the specific phase of the exper-

iment ($F_{(3,30)} = 0.42$, $p > 0.5$; $F_{(3,30)} = 0.35$, $p > 0.5$, for amplitude and velocity, respectively). Movement amplitudes but not velocities differed marginally between experiments ($F_{(2,10)} = 3.23$, $p = 0.08$; $F_{(2,10)} = 1.95$, $p > 0.2$), with tests involving *siaes* and *suæs* having somewhat smaller amplitudes than the other two manipulations. There was no statistical interaction; the kinematic patterns in the different phases of testing were similar for the different test utterances ($F_{(6,30)} = 0.70$, $p > 0.5$; $F_{(6,30)} = 0.18$, $p > 0.5$, for amplitude and maximum velocity, respectively).

An analysis of jaw movement kinematics in training and transfer conditions showed that a large proportion of the states that subjects encountered in the transfer test had already been experienced during force-field training. Figure 5 shows an example from a single subject in which

vertical jaw position and velocity are plotted as a function of normalized time for the training utterance *siaes* (mean \pm 2 SD over all training trials) and the transfer utterance *suæs* (mean \pm 2 SD for the transfer block, field on). In this example, *siaes* trajectories (training) are indicated by solid lines, *suæs* trajectories (transfer) are indicated by dashed lines, and the overlap is shown in gray. The overlap is calculated as the proportion of the transfer utterance area (based on all trials in the transfer utterance block with the force field on) that subjects experienced during training movements (based on all training trials in the force field). We assessed differences in the proportion of overlap in jaw position and velocity using repeated-measures ANOVA. We found that the jaw trajectory overlap was comparable in all three experiments ($F_{(2,10)} = 0.43$; $p > 0.5$), but the overlap measures themselves differed reliably ($F_{(3,36)} = 7.08$; $p < 0.01$). Bonferroni's corrected comparisons found the overlap between training and transfer movements along a horizontal axis to be less than the overlap between training and transfer movements in the vertical direction (Fig. 5C) ($p < 0.04$). The average percentage overlap in training and transfer for the first experiment (utterances *siaes* and *suæs*) was 62.4 for horizontal jaw position, 92.8 for vertical jaw position, 76.6 for horizontal jaw velocity, and 87.2 for vertical jaw velocity. The overlap percentages for the test with *siaes* and *siaeis* were 68.0, 72.0, 80.8, and 73.0 for horizontal and vertical position and velocity, respectively. Values for *siaes* vocalized and silent were 68.2, 84.4, 79.4 and 84.4, respectively. The substantial overlap particularly in the case of jaw velocities suggests that subjects had experienced, in the context of the initial training, a large proportion of the states needed for good performance in the transfer test conditions.

The similarity of training and transfer utterances was also assessed in a control study in which tongue and jaw movements were recorded using an electromagnetic motion tracking device. Subjects repeated each of the test utterances used in the main experiments (*siaes* vocalized, *siaes* silent, *suæs*, and *siaeis*) in separate blocks of trials. Using repeated-measures ANOVA and Bonferroni's corrected pairwise comparisons, we tested for differences in the amplitude, velocity, and duration of movement. Tests were conducted separately for horizontal and vertical components of movement of the two receiver coils on the tongue and a single coil on the jaw. Mean values for these variables (\pm SE) are given in Table 2.

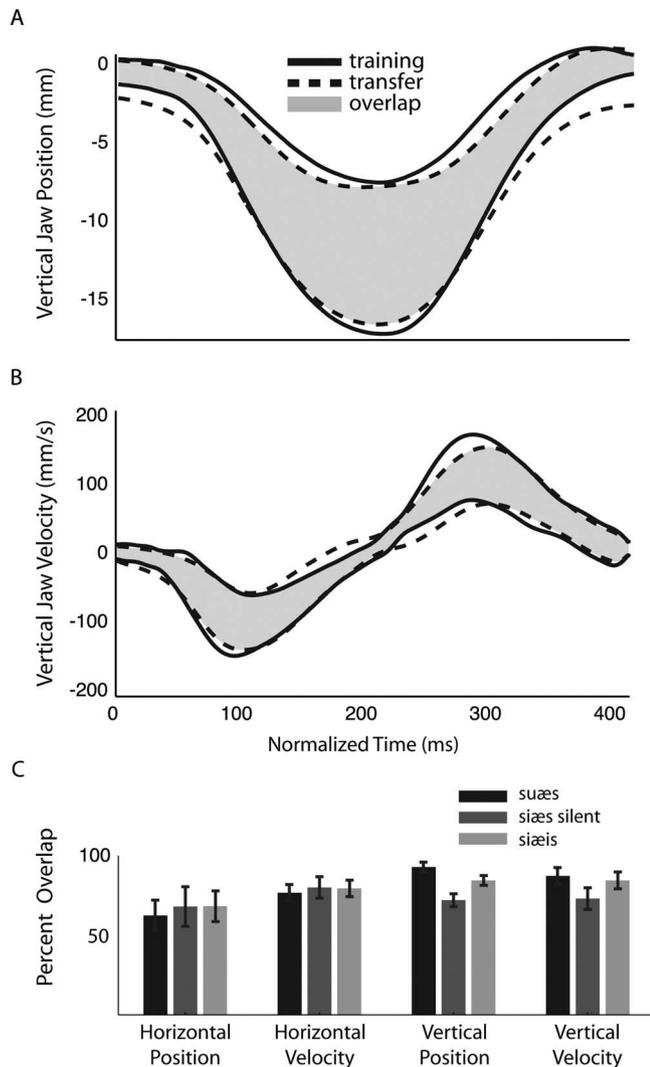


Figure 5. Analysis of kinematic overlap between training and transfer utterances. **A**, Mean vertical jaw position (± 2 SD) for all training trials (field on, utterance *siæs*) and for the transfer block (field on, utterance *suæs*). In this example, training utterance trajectories are indicated by solid lines, transfer utterance trajectories are indicated by dashed lines, and the overlap is shown in gray. The computed overlap is the percentage of the transfer utterance area that is shared by the training utterance. **B**, Mean vertical jaw velocity (± 2 SD) under the same conditions. In this example, overlap is 0.87 and 0.86, for vertical position (**A**) and vertical velocity (**B**), respectively. **C**, Mean overlap (\pm SE) between training and transfer utterances (averaged over subjects and experiments). Percentage overlap is shown between horizontal and vertical jaw position and velocity for *siæs* and each of the different utterances used in tests of transfer of training.

We observed no reliable differences in horizontal movement amplitude for any of the four test utterances ($F_{(3,15)} = 1.18$; $p > 0.3$). However, horizontal amplitudes differed reliably for the different receiver coils ($F_{(2,10)} = 5.51$; $p < 0.02$) such that larger horizontal movement amplitudes were observed for both coils on the tongue. For vertical amplitude, the receiver coils on the tongue and the jaw had different amplitudes for different utterances ($F_{(6,30)} = 3.56$; $p < 0.01$). Greater amplitudes at the tongue blade than at the incisor in *siæs* were the source of the interaction ($p < 0.01$). There was no evidence that individual receiver coils had different vertical amplitudes for different utterances ($p > 0.2$ in all cases).

Maximum horizontal velocities differed for the three receiver coils ($F_{(2,10)} = 10.19$; $p < 0.01$), with reliably greater velocities

observed for the two tongue coils on the tongue ($p < 0.05$ in both cases). However, there was no evidence that horizontal velocity differed for the four test utterances ($F_{(3,15)} = 1.87$; $p > 0.15$).

Maximum vertical velocity was found to differ both for different utterances ($F_{(3,15)} = 20.64$; $p < 0.01$) and for different receiver coils ($F_{(2,10)} = 5.35$; $p < 0.03$). There was also a reliable statistical interaction ($F_{(6,30)} = 7.10$; $p < 0.01$) such that the vertical velocities of both tongue markers were higher for *siæs* in vocalized speech condition than in the silent speech condition ($p < 0.01$ in each case). In contrast, jaw marker vertical velocities were similar for the four utterances ($p > 0.5$ in all cases).

Movement durations likewise differed for different utterances ($F_{(3,15)} = 3.89$; $p < 0.03$). However, they were comparable for different receiver coils ($F_{(2,10)} = 0.26$; $p > 0.5$), and there was no statistical interaction ($F_{(6,30)} = 0.97$; $p > 0.4$). Durations were greatest for *siæs* under silent speech conditions and least for *suæs*. However, none of the individual pairwise tests for differences in duration reached significance by Bonferroni's corrected comparisons ($p > 0.3$ in all cases).

Thus, with one exception, we found no differences across the utterances that formed training/transfer pairs in terms of the movement amplitudes, maximum velocities, and durations of movement of the receiver coils. The one exception was that vertical velocities of tongue markers were reliably higher for *siæs* vocalized than for *siæs* silent. In evaluating the findings of the EMA recordings, it should be kept in mind that this control study was conducted without attaching subjects to the robot. Patterns of movement may be different with the robot in place.

Discussion

The instances of adaptation that we have observed in the experiments described above are in accord with the idea that the nervous system corrects for dynamics in speech production. However, this correction on its own does not mean that adaptation is based on a generalizable dynamics representation. To assess the extent to which there is generalization of speech learning, we manipulated the phonetic content of utterances so as to vary the similarity of movements in training and transfer conditions. Although we examined a reasonably small set of utterances, we find that, in each case, speech learning fails to generalize regardless of the extent to which utterances are matched in terms of their kinematics. The present evidence is compatible with the idea that dynamics learning is primarily local or instance based. Generalization of learning presumably arises on the basis of interpolation between instances of local learning (Atkeson, 1989; Gandolfo et al., 1996; Ghahramani and Wolpert, 1997; Malfait et al. 2005; Mattar and Ostry, 2007a,b). Generalization in speech learning would presumably be evident after training with multiple utterances that could be combined to form the elements of the transfer set.

The varied phonetic conditions of speech can be compared with context cues that have been the focus of previous work in motor learning. In limb dynamics studies, it has been shown that a learned movement that is acquired in one context may fail to transfer to a different context even when the movement and the force environment remain unchanged (Cothros et al., 2006; Nozaki et al., 2006). Moreover, subjects can adapt to multiple environments based on a change in context (Gandolfo et al., 1996; Osu et al., 2004; Waincott et al., 2005). The present study shows that context cues, including different phonetic conditions, are highly selective. There is little transfer of speech learning from one context to another. The specificity of the learning argues

Table 2. Tongue and jaw movement parameters as recorded with EMA

	<i>siaes</i> vocalized	<i>siaes</i> silent	<i>suas</i>	<i>siaeis</i>
Mandibular incisor				
X	1.34 ± 0.25	1.54 ± 0.25	1.29 ± 0.38	1.28 ± 0.28
Y	9.43 ± 1.72	10.56 ± 1.28	8.30 ± 1.31	9.59 ± 1.62
VX	14.91 ± 2.87	13.31 ± 1.66	14.64 ± 3.90	15.81 ± 3.29
VY	62.40 ± 13.66	56.80 ± 7.58	63.37 ± 10.14	73.87 ± 15.34
Dur	335.77 ± 21.52	398.39 ± 21.84	276.14 ± 5.21	290.85 ± 17.91
Tongue blade				
X	2.66 ± 0.85	2.54 ± 0.46	4.12 ± 1.52	2.58 ± 0.56
Y	11.16 ± 1.75	9.03 ± 1.79	9.49 ± 1.75	11.49 ± 1.85
VX	25.87 ± 6.30	20.80 ± 3.18	36.74 ± 11.77	28.07 ± 4.57
VY	100.95 ± 15.22	60.94 ± 10.53	79.70 ± 9.26	110.83 ± 14.76
Dur	313.79 ± 24.24	336.25 ± 26.56	287.39 ± 29.41	295.38 ± 34.55
Tongue dorsum				
X	3.17 ± 1.10	3.34 ± 0.94	3.77 ± 0.87	2.92 ± 0.69
Y	9.19 ± 1.50	7.96 ± 2.00	9.05 ± 1.25	10.04 ± 1.24
VX	31.33 ± 7.87	28.49 ± 6.16	37.39 ± 8.79	34.07 ± 6.40
VY	101.43 ± 13.40	62.95 ± 10.87	81.35 ± 11.35	107.39 ± 13.47
Dur	306.55 ± 26.49	373.97 ± 46.99	297.88 ± 31.12	318.02 ± 20.09

Mean values are shown across subjects (\pm SE) for horizontal amplitude (X; in millimeters), vertical amplitude (Y; in millimeters), horizontal velocity (VX; in millimeters), vertical velocity (VY; in millimeters), and opening movement duration (Dur; in milliseconds). Values are given for markers attached to the mandibular incisor, tongue blade, and tongue dorsum in each of the training or test utterances.

against a model-based explanation of generalization of speech learning.

There are a number of instances in which generalization of dynamics learning has been previously reported. The majority of these cases are situations in which training and transfer movements are closely related. For example, generalization across limbs is observed in situations in which movements of the contralateral limb are closely matched to those of the training limb (Dizio and Lackner, 1995; Criscimagna-Hemminger et al., 2003; Wang and Sainburg, 2004). Interlimb generalization may also reflect cognitive factors, in that transfer is eliminated when loads are introduced gradually such that there is no kinematic error and no conscious awareness of load (Malfait and Ostry, 2004). Transfer of learning has also been reported across workspace locations (Shadmehr and Mussa-Ivaldi, 1994; Shadmehr and Moussavi, 2000; Malfait et al., 2002; Hwang et al., 2003; Mah and Mussa-Ivaldi, 2003). However, this kind of transfer appears to be limited to cases in which training and transfer movements involve the same joint rotations and are thus in effect the same movements in a different part of the workspace. Generalization over differences in amplitude and duration has likewise been observed, but it is in the context of movements in a single direction (Goodbody and Wolpert, 1998).

There are a small number of examples of which we are aware in which dynamics learning generalizes more broadly. The most extensively documented example is that errors in force-field learning that occur on a trial-by-trial basis affect the immediately after movement over a wide range of movement directions (Thoroughman and Shadmehr, 2000; Donchin et al., 2003; Thoroughman and Taylor, 2005). It is not clear how these trial-to-trial effects are reflected in the final level of adaptation that follows a longer period of training. Indeed, a rather different pattern is observed when learning occurs over several hundreds of movements. In this case, there is no evidence of generalization whatsoever beyond $\pm 45^\circ$ from the training direction (Mattar and Ostry, 2007b).

In the case of visuomotor learning, adaptation to altered visuomotor gain generalizes broadly and the shape of the generalization gradient is modifiable (Krakauer et al., 2000). Krakauer et al. (2006) report another rather specific example of generaliza-

tion. They show that learning generalizes from the shoulder to wrist but not vice versa. One example, that may reflect true generalization, is described by Hwang et al. (2003) who report generalization across workspace locations. The most important aspect of this dataset is that the magnitude of the aftereffect changes systematically over the workspace and may indicate the capacity for extrapolation. The presence in the Hwang et al. data of transfer across workspace locations on its own is not unexpected because similar joint rotations are involved in the training and transfer configurations. In one additional example, generalization of dynamics learning is observed in the generation of movement sequences. Specifically, adaptation is observed to generalize across movements in the same region of the workspace in cases in which the components of a movement sequence have been reordered (Conditt et al., 1997).

Speech learning data are different in a number of ways than patterns of learning observed in limb movement. We found that only $\sim 60\%$ of subjects adapted to the new dynamics. This adaptation rate is typical of that observed for force-field adaptation in speech production tasks (Tremblay et al., 2003; Nasir and Ostry, 2006) and also for adaptation to altered formant frequency in speech (Houde and Jordan, 2002; Purcell and Munhall, 2006). We also found that adaptation was incomplete. In the present study, subjects show a reduction in movement curvature with training that averages $\sim 50\%$. In speech learning tasks involving acoustics, adaptation to altered formant frequencies results in even smaller effects, with adaptation correcting for $\sim 25\text{--}30\%$ of the experimentally induced formant shift (Purcell and Munhall, 2006). It would be of interest to know whether these failures are related. It may be that some speakers are more reliant on auditory information and others more on proprioception. Individuals who fail to adapt in the present situation may be more sensitive to auditory than to somatosensory feedback and vice versa. The incomplete adaptation observed in both kinds of speech learning tasks may indicate that targets in tasks such as these are less precise.

As in previous studies of speech learning, aftereffects after the removal of load were small. In the study by Tremblay et al. (2003), aftereffect magnitudes were, in many cases, less than the deflection produced by the initial exposure to load. Indeed, Nasir and Ostry (2006) provide empirical evidence that, for lateral loads, the modest aftereffects after speech learning result from adaptation achieved at least in part through impedance control. Adaptation in speech learning is presumably based on a combination of increased stiffness along with learned compensatory adjustments. Indeed, although not well documented in the literature, the presence of increased stiffness, long after kinematic and electromyographic measures of learning have reached asymptotic values is characteristic of not only speech production but learning in arm movement as well (Thoroughman and Shadmehr, 1999).

The duration of movement in these experiments was typically ~ 600 ms, with movement more or less equally divided between opening and closing phases (Table 2). The individual phases of movement were thus long enough for feedback corrections, and

indeed on-line correction may well play a role in the adaptation observed here. An examination of Figure 2 indicates that there are corrective adjustments to jaw paths that are present from the very start of the movement. This suggests that adaptation that is observed here depends on a substantial feedforward adjustment even if there are feedback-based changes as well.

In summary, dynamics learning in speech production appears to be highly context dependent. For the utterances tested here, learning fails to generalize when similar tongue and jaw movements are embedded in different words or different production contexts. The findings point to the specificity of speech motor learning and are inconsistent with the idea that transfer of speech learning is based on a generalizable dynamics representation.

References

- Atkeson CG (1989) Learning arm kinematics and dynamics. *Annu Rev Neurosci* 12:157–183.
- Conditt MA, Gandolfo F, Mussa-Ivaldi FA (1997) The motor system does not learn the dynamics of the arm by rote memorization of past experience. *J Neurophysiol* 78:554–560.
- Cothros N, Wong JD, Gribble PL (2006) Are there distinct neural representations of object and limb dynamics? *Exp Brain Res* 173:689–697.
- Criscimagna-Hemminger SE, Donchin O, Gazzaniga MS, Shadmehr R (2003) Learned dynamics of reaching movements generalize from dominant to nondominant arm. *J Neurophysiol* 89:168–176.
- Dizio P, Lackner JR (1995) Motor adaptation to Coriolis force perturbations of reaching movements: endpoint but not trajectory adaptation transfers to the nonexposed arm. *J Neurophysiol* 74:1787–1792.
- Donchin O, Francis JT, Shadmehr R (2003) Quantifying generalization from trial-by-trial behavior of adaptive systems that learn with basis functions: theory and experiments in human motor control. *J Neurosci* 23:9032–9045.
- Flanagan JR, Wing AM (1997) The role of internal models in motion planning and control: evidence from grip force adjustments during movements of hand-held loads. *J Neurosci* 17:1519–1528.
- Gandolfo F, Mussa-Ivaldi FA, Bizzi E (1996) Motor learning by field approximation. *Proc Natl Acad Sci USA* 93:3843–3846.
- Ghahramani Z, Wolpert DM (1997) Modular decomposition in visuomotor learning. *Nature* 386:392–395.
- Goodbody SJ, Wolpert DM (1998) Temporal and amplitude generalization in motor learning. *J Neurophysiol* 79:1825–1838.
- Houde JF, Jordan MI (2002) Sensorimotor adaptation of speech. I. Compensation and adaptation. *J Speech Lang Hear Res* 45:295–310.
- Hwang EJ, Donchin O, Smith MA, Shadmehr R (2003) A gain-field encoding of limb position and velocity in the internal model of arm dynamics. *PLoS Biol* 1:209–220.
- Krakauer JW, Pine ZM, Ghilardi M-F, Ghez C (2000) Learning of visuomotor transformations for vectorial planning of reaching trajectories. *J Neurosci* 20:8916–8924.
- Krakauer JW, Mazzoni P, Ghazizadeh A, Ravindran R, Shadmehr R (2006) Generalization of motor learning depends on the history of prior action. *PLoS Biol* 4:e316.
- Mah CD, Mussa-Ivaldi FA (2003) Generalization of object manipulation skills learned without limb motion. *J Neurosci* 23:4821–4825.
- Malfait N, Ostry DJ (2004) Is interlimb transfer of force-field adaptation a cognitive response to the sudden introduction of load? *J Neurosci* 24:8084–8089.
- Malfait N, Shiller DM, Ostry DJ (2002) Transfer of motor learning across arm configurations. *J Neurosci* 22:9656–9660.
- Malfait N, Gribble PL, Ostry DJ (2005) Generalization of motor learning based on multiple field exposures and local adaptation. *J Neurophysiol* 93:3327–3338.
- Mattar AAG, Ostry DJ (2007a) Neural averaging in motor learning. *J Neurophysiol* 97:220–228.
- Mattar AAG, Ostry DJ (2007b) Modifiability of generalization in dynamics learning. *J Neurophysiol* 98:3321–3329.
- Nasir SM, Ostry DJ (2006) Somatosensory precision in speech production. *Curr Biol* 16:1918–1923.
- Nozaki D, Kurtzer I, Scott SH (2006) Limited transfer of learning between unimanual and bimanual skills within the same limb. *Nat Neurosci* 9:1364–1366.
- Osu R, Hirai S, Yoshioka T, Kawato M (2004) Random presentation enables subjects to adapt to two opposing forces on the hand. *Nat Neurosci* 7:111–112.
- Purcell DW, Munhall KG (2006) Compensation following real-time manipulation of formants in isolated vowels. *J Acoust Soc Am* 119:2288–2297.
- Sainburg RL, Ghez C, Kalakanis D (1999) Intersegmental dynamics are controlled by sequential anticipatory, error correction, and postural mechanisms. *J Neurophysiol* 81:1045–1056.
- Shadmehr R, Moussavi ZM (2000) Spatial generalization from learning dynamics of reaching movements. *J Neurosci* 20:7807–7815.
- Shadmehr R, Mussa-Ivaldi FA (1994) Adaptive representation of dynamics during learning of a motor task. *J Neurosci* 14:3208–3224.
- Thoroughman KA, Shadmehr R (1999) Electromyographic correlates of learning an internal model of reaching movements. *J Neurosci* 19:8573–8588.
- Thoroughman KA, Shadmehr R (2000) Learning of action through adaptive combination of motor primitives. *Nature* 407:742–747.
- Thoroughman KA, Taylor JA (2005) Rapid reshaping of human motor generalization. *J Neurosci* 25:8948–8953.
- Tremblay S, Shiller DM, Ostry DJ (2003) Somatosensory basis of speech production. *Nature* 423:866–869.
- Wainscott SK, Donchin O, Shadmehr R (2005) Internal models and contextual cues: encoding serial order and direction of movement. *J Neurophysiol* 93:786–800.
- Wang J, Sainburg RL (2004) Interlimb transfer of novel inertial dynamics is asymmetrical. *J Neurophysiol* 92:349–360.
- Witney AG, Wolpert DM (2003) Spatial representation of predictive motor learning. *J Neurophysiol* 89:1837–1843.
- Wolpert DM, Ghahramani Z, Jordan MI (1995) An internal model for sensorimotor integration. *Science* 269:1880–1882.