Speech motor behavior can be altered through reinforcement learning

Benjamin Parrell[1,2,3]

[1] Department of Communication Sciences and Disorders, University of Wisconsin–Madison

[2] Waisman Center, University of Wisconsin–Madison

[3] Department of Linguistics and Cognitive Science, University of Delaware

Corresponding author: Benjamin Parrell, bparrell@wisc.edu

This work supported by a grant from the University of Delaware Research Foundation.

The authors have no competing interests to declare.

Abstract:

Reinforcement learning, the ability to change motor behavior based on external reward, has been suggested to play a critical role in early stages of speech motor development and is widely used in clinical rehabilitation for speech motor disorders. However, no current evidence exists that demonstrates the capability of reinforcement to drive changes in human speech behavior. Speech provides a unique test of the universality of reinforcement learning across motor domains: speech is a complex, high-dimensional motor task whose goals do not specify a task to be performed in the environment but ultimately must be self-generated by each speaker such that they are understood by those around them. Reinforcement learning may thus be more difficult for speech, given its high-dimensional and redundant motor system, while speech may also be particularly responsive to reinforcement given the ultimate goal is typically reliant on such feedback from our interlocutors. Across four experiments, we establish whether reinforcement learning alone is sufficient to drive changes in speech behavior and parametrically test two features known to affect reinforcement learning in reaching: how informative the reinforcement signal is as well as the availability of sensory feedback about the outcomes of one's motor behavior. We show that reinforcement learning can alter speech behavior and that more informative reward signals lead to greater learning. Contrary to results from upper limb control, masking feedback about movement outcomes has no effect on speech learning. Our results suggest reinforcement learning is active in speech but may operate differently than in other motor domains.

Introduction

When we are speaking with someone, we are usually understood without any problems. However, sometimes this seemingly effortless communication breaks down, whether due to a noisy environment, problems in communication technology, a distracted listener, speaking with someone from another part of the country or world, or myriad other reasons. In these situations, we need to change our speech to be better understood, but we may have limited or no information about why we were not understood or how to change our speech to maximize intelligibility. In these cases, we may try out different pronunciations of a word until we receive positive feedback from the listener that they understood what we were saying. This type of trial-by-trial learning driven by external feedback is typically often referred to as *reinforcement learning* (sometimes, as model-free learning).

Reinforcement learning has been studied extensively in upper limb control (e.g., Cashaback et al., 2017; Galea et al., 2015; Izawa & Shadmehr, 2011; Nikooyan & Ahmed, 2015; Therrien et al., 2016; Wu et al., 2014) and, to a smaller extent, in gait (Hasson et al., 2015). To date, it is essentially unknown to what extent reinforcement learning is active in speech production. Speech provides a unique test system to evaluate the universality of reinforcement learning across motor domains for two reasons. First, speech is a uniquely complex motor behavior, relying on coordination of close to roughly muscles between the respiratory, phonatory, and articulatory systems that requires complex control of both skeletal joints and the tongue, a muscular hydrostat. Second, speech is unique among human motor behaviors in that the targets for movements are internally generated rather than being defined in the environment. Ultimately, the goal in speech production is to be understood, and each speaker must come to define their own the motor task goals to accomplish this task.

Reinforcement learning in speech may be critical both during developmental speech acquisition and for treatment of motor disorders. Developmentally, reinforcement learning has been suggested to play a critical role in the first stages of speech acquisition (Howard & Messum, 2011, 2014; Messum & Howard, 2012; Warlaumont, 2014; Warlaumont et al., 2013; Warlaumont & Finnegan, 2016). In these models, the first words that infants produce are vocalizations produced with essentially random movements of the speech articulators. The productions that are recognized and positively reinforced by an external caregiver are more likely to be repeated. Over time, this reinforcement and repetition leads to consolidation of the motor plans that produce words that closely match the words in the language the infant is learning. In terms of motor rehabilitation, reinforcement also forms part of existing standards of care for motor speech disorders, typically combined with explicit instruction about how to produce a particular sound or set of sounds (Ballard et al., 2000; Duffy, 2013).

Despite the practical importance of reward learning in existing rehabilitation paradigms and its potential theoretical importance in human speech development, reinforcement learning in speech has received relatively little attention. The vast majority of studies on mechanisms of motor learning in speech has focused on sensory-error based learning (e.g., Daliri & Dittman, 2019; Houde & Jordan, 1998; Lametti et al., 2012, 2018; MacDonald et al., 2011; Mitsuya et al., 2015; Purcell & Munhall, 2006; Shiller et al., 2009; Villacorta et al., 2007). In this type of learning, differences between predicted sensory feedback and perceived reafferent feedback about movement outcomes lead to sensory prediction errors, which are used to update internal models and/or control systems to adapt behavior to oppose the perturbation. While sensorimotor learning can also drive changes in speech behavior, these changes are relatively short-lived in both speech and other motor domains compared to the longer-term impact of reinforcement learning (Krakauer, 2015; Roemmich & Bastian, 2018) and the two mechanisms rely on different neural substrates (Krakauer, 2015).

Perhaps because of its prominent role in speech rehabilitation, reinforcement learning has received some attention for clinical applications in speech. However, research to date, has almost universal focused on how the frequency of reinforcement affects learning, with mixed results (Adams et al., 2002; Adams & Page, 2000; Bislick et al., 2012, 2013; Hula et al., 2008; Katz et al., 2010; Steinhauer & Grayhack, 2000). While these studies have focused on the role of feedback frequency, they have not demonstrated clearly how reinforcement learning operates in speech. First, these studies mostly provided highly informative feedback about performance outcomes, giving participants either explicit instruction of *how* to improve their performance or highly informative feedback about their performance such as the difference between produced duration and a duration target (often referred to as "knowledge of performance" and "knowledge of results" (Schmidt & Lee, 2011)). In non-speech domains, this type of explicit information is known to aid learning during training but often decreases retention (Hasson et al., 2015; Schmidt & Lee, 2011). Second, these studies provided explicit instruction about the desired outcome. Although this is typical in clinical settings (Ballard et al., 2000), how explicit instruction interacts with other types of motor learning in unclear (Boyd & Winstein, 2004), and may in fact detrimentally affect learning in some cases (Green & Flowers, 1991; Shea et al., 2001). Critically, reinforcement learning in limb control is possible without explicit instruction (Galea et al., 2015; Izawa & Shadmehr, 2011; Nikooyan & Ahmed, 2015), suggesting it relies on a separate neural system.

The aim of the current study is to establish to what extent reinforcement learning is able to shape speech motor behavior. In addition to establishing the capability of the speech sensorimotor system to learn purely from reinforcement signals, we additionally explore two aspects of reinforcement learning that may affect its effectiveness in speech. First, learning is more likely in reaching tasks when the reward signal contains some information about the desired outcome compared to uninformative signals that relay only success or failure, particularly for motor tasks involving multi-dimensional control (Kooij & Overvliet, 2016; Manley et al., 2014). Second, there is evidence that the availability of sensory feedback may interfere with reinforcement learning. Cashaback et al. (2017) designed a task where participants learned to alter their reach location either through reinforcement alone when visual feedback was withheld or though sensory errors driven by visual feedback. Critically, they used a non-uniform distribution of perturbations such that the two learning mechanisms differed in the magnitude of compensation. When both visual feedback and reinforcement were combined, pitting the two learning systems against each other, learning was identical to the visual feedback alone. This result suggests the availability of sensory feedback may interfere with reinforcement learning in some cases.

We parametrically explore these two factors (information content of the reward signal and availability of sensory feedback) in a set of four studies on speech reinforcement learning where the factors are crossed in a 2 x 2 design. The basic goal, across all experiments, is to induce a change in the first vowel formant (F1) of the vowel /ɛ/ (as in *head*). Vowel formant are the characteristic resonances of the vocal tract, are closely tied to movements within of the lips, tongue, and jaw, and are typically used to characterize vowels in speech. Notably, a similar change in vowel formants is frequently the target of sensorimotor learning studies in speech. Thus, this paradigm will allow for comparison of our results with previous work in this area. To establish the ability of reinforcement learning to drive changes in speech behavior, we examine change in F1 in each study separately. To examine the effect of reward signal information content and sensory feedback availability, we compare results across all four studies.

Methods

Participants:

All participants were recruited from courses in the Linguistics and Cognitive Sciences department at the University of Delaware and were compensated with extra credit in those courses. No participant reported any history of speech or hearing problems. Experiments 1, and 2, and 4 had 20 participants each (Exp 1: 19 female/1 male; Exp 2: 20 female/0 male; Exp 4: 14 female/6 male). Experiment 3 had 21 participants (16 female/5 male). The experimental protocol was approved by Institutional Review Boards at the University of Delaware and the University of Wisconsin–Madison.

General methods:

The experiments are designed to induce participants to alter the first vowel formant (F1) in the vowel /ɛ/ solely through external reinforcement. Participants wore a head-mounted microphone (AKG C520) that was used to record their speech, and wore closed-back, over-the-ear headphones (Beyerdynamic DT 770) that were used to play auditory reward signals and, in Experiments 3 and 4, to play speech-shaped noise designed to mask auditory feedback. Audio data was digitized using a Scarlett 2i2 USB audio interface and recorded with the Audapter program (Cai et al., 2008; Tourville et al., 2013) in MATLAB.

Each experiment has three phases: baseline, training, and washout (Figure 1, example for “head” shown). During all phases, participants read words out loud, one at a time, as they appear on a computer screen. Stimuli for the baseline, training, and washout phases were *head*, *bed*, and *dead* for all experiments. These stimuli contained the target vowel /ɛ/. Experiments 2, 3, and 4 additionally included the words *hid*, *bid*, *did* and *had*, *bad*, *dad* during the baseline phase only to measure F1 for the vowels /ɪ/ and /æ/, respectively. The order of the stimuli was randomized for each participant. Each word with /ɛ/ was repeated at least 20 times during the baseline phase of each experiment.

In order to provide real-time feedback based on participants' vowel formants, the target vowel for each trial was detected automatically as the part of the speech signal for that trial above a participant-specific amplitude threshold. Then, vowel formants were tracked using Praat (Boersma & Weenink, 2019). A single F1 value for that trial was then calculated as the average F1 within a 50ms window centered around the vowel midpoint. Using a small window ensured the F1 measurement was taken from the steady-state portion of the vowel even with a somewhat noisy estimate of vowel onset and offset. The participant-specific amplitude threshold used for vowel detection and Linear Predictive Coefficient order for formant tracking were set in a brief parameter setting session immediately prior to the main experiment.

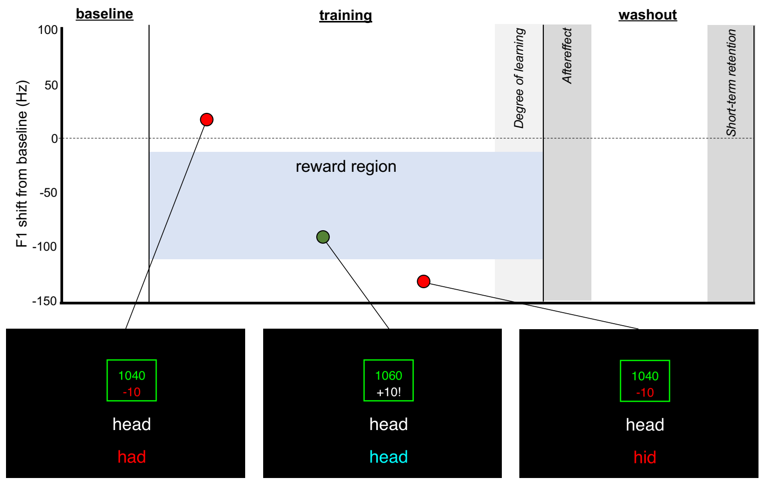


Figure 1: Schematic of general methods. Examples of trials with F1 above (left), in (center), and below (right) the target region are shown.

- <u>Baseline phase</u> (80-120 trials): Participants are told that they are training a computer program to recognize their particular voice. During this phase, the mean and standard deviation of F1 is measured. No reward or reinforcement signal was given during the baseline phase.
- <u>Training phase</u> (250-350 trials): Participants are told the computer program that was just trained will try to recognize the words they speak. Participants gain points when the computer recognizes the target word (+ in Fig 1) and lose points when it recognizes another word (**x** Figure 1). Rewards are presented visually and accompanied by auditory reward signals (chimes, spoken words) which vary by experiment. Participants are told that their goal is to gain points by being recognized correctly by the computer. Unknown to the participants, the computer recognizes words as correct only when the first vowel formant (F1) falls within a specific target region (blue shaded region in Fig 1). This target region is 100 Hz wide, and is defined relative to the participant's mean F1 for the vowel /ɛ/ produced during the baseline phase (10-110 Hz below the mean). The overlap of the reward region with participants baseline

productions was chosen to ensure that participants would receive positive reward on some productions without changing their baseline behavior, as large shifts that do not overlap with baseline production may be difficult to learn (Therrien et al., 2016). A positive reward (+10 points) was given if F1 falls within a target region defined relative to the participant's mean F1 in the baseline phase (10-110 Hz below the mean). Productions above this region are recognized as containing the vowel /æ/ (e.g., *had*); those below this region, the vowel /ɪ/ (e.g., *hid*). The direction of the target region shift relative to baseline values (positive or negative) was always negative; thus, participants needed to shift their production of /ɛ/ towards /ɪ/ to produce F1 in the target region. Participants started with 1000 points.

- <u>Washout phase</u> (100-150 trials): Participants are told that the game is over, and that they are to simply read the words as they appear. Participants do not receive any feedback or earn/lose points during the washout phase. The long washout period (100-150 trials, depending on the experiment) allows for testing short-term retention of learning. Notably, changes in speech behavior due to sensorimotor learning return to near baseline values within 30-50 trials (MacDonald et al., 2011; Parrell et al., 2017). The washout phase is used to assess both the degree of learning (*aftereffects*, measured during first 20 trials) and short-term retention (last 20 trials). No reward or reinforcement signal was given during the washout phase.

Each trial lasted 3 seconds. Feedback about performance, if shown, was displayed for an additional 2 seconds. There was a 0.5 second pause between each trial when no stimulus word was displayed.

<u>Experiment-specific methods:</u>

Experiment 1: The baseline phase consisted of 80 trials; the training phase, 350 trials; and the washout phase, 100 trials. During the training phase, when participants production fell within the reward region, a pleasant chime was played over the headphones. When the production fell above or below this region, a pre-recorded voice saying the "recognized" word was played. For example, when the stimulus was "head", "had" was played when the production was above the target region, while "hid" was played when the production fell below the target region.

Experiment 2: The baseline phase consisted of 120 trials; the training phase, 250 trials; and the washout phase, 150 trials. All acoustic reinforcement signals were based on each participants' own productions recorded during the baseline phase. For each word in the baseline phase, the production with median F1 was chosen to be played back to the participant during the training phase. In order to create a positive reinforcement signal that fell within the target region, F1 for the chosen productions of *head, bead,* and *dead* was shifted by -60 Hz using Audapter. This resulted in an F1 in the center of the reward zone for these words. During the training, when the production fell above or below the target region, the participant's recording of the "heard" word was played. For example, when the stimulus was "head", "had" was played when the production was above the target region, while "hid" was played when the production fell below the reward zone. When the production fell within the reward zone, the modified version of the "heard" word was played. For example, when the stimulus was "head", the participant's own production of "head" from the baseline phase, with F1 shifted by -60 Hz, was played.

Experiments 3 and 4: Experiments 3 and 4 were designed to mirror the reinforcement signals used in Experiments 1 and 2 with the addition of speech-shaped noise designed to mask participants' ability to hear their own speech. For Experiment 3, the baseline phase consisted of 120 trials; the training phase, 250 trials; and the washout phase, 150 trials. For Experiment 4, the baseline phase consisted of 90 trials; the training phase, 250 trials; and the washout phase, 100 trials. Stimuli with all vowels (/ɪ/, /ɛ/, and /ae/) were included in the baseline phase, where each stimulus word was repeated 10 times each. For both experiments, only the /ɛ/ stimuli were used after the baseline phase. Reinforcement signals were the same as those used in Experiment 1 (Experiment 3) and Experiment 2 (Experiment 4). The amplitude of the masking noise was modulated by the amplitude of the participant's speech using Audapter, with the noise played at a constant gain above the speech amplitude and calibrated to be roughly 80 dB when speaking at a normal volume (Figure 2). This allowed us to prevent participants from receiving auditory feedback about their speech, while largely avoiding potential Lombard affects associated with speaking in the presence of background noise. A summary of differences between experiments in shown in Table 1.

Table 1: Methodological differences between experiments.

EXP	POSITIVE REWARD SOUND	MASKING NOISE	VOWELS IN BASELINE	BASELINE TRIALS	TRAINING TRIALS	WASHOUT TRIALS
EXP 1	chime	no	/ɛ/	80	350	100
EXP 2	resynthesized token from baseline with F1 in target region	no	/ɛ/, /ae/, and /ɪ/	120	250	150
EXP 3	chime	yes	/ɛ/, /ae/, and /ɪ/	120	250	150
EXP 4	resynthesized token from baseline with F1 in target region	yes	/ɛ/, /ae/, and /ɪ/	90	250	100

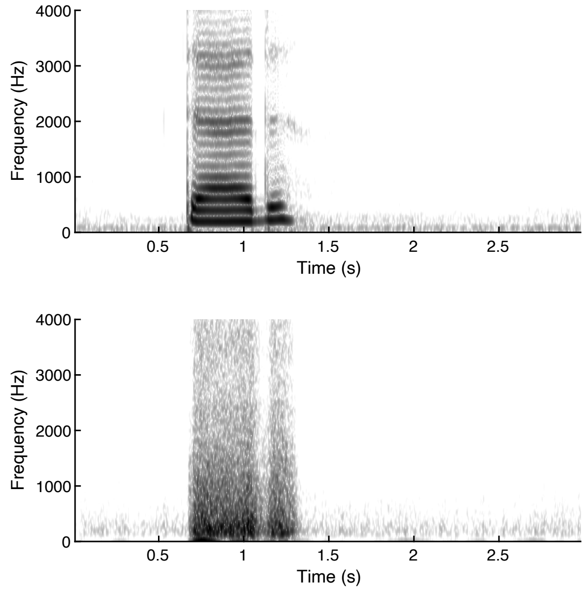


Figure 2: Spectrograms showing speech input (top) and amplitude-modulated masking noise (bottom) used in experiments 3 and 4. The amplitude-modulated noise served to mask auditory feedback while limiting Lombard effects associated with speaking in noise.

Post-participation survey:

Participants in Exp 2 and 3 were given a survey after they completed the experiment to assess whether they adopted any strategy and, if so, what that strategy was. Participants were also asked a set of questions regarding their level of engagement and attention during the experiment.

Data analysis:

The primary outcome for all experiments was the change in F1 for /ɛ/ from its baseline value. For each participant, all trials for a given participant were normalized to the mean F1 for words with /ɛ/ from the baseline phase. To measure learning, we took the mean of this normalized F1 over the last 20 trials of the training phase. Aftereffects were measured as the mean F1 during the first 20 trials of the washout phase, and short-term retention was measured as the mean during the last 20 trials of the washout phase. In order to test whether learning occurs, we used linear mixed-effects models using the *lme4* package (Bates et al., 2014) in *R* (R Core Team, 2013) with a fixed factor of phase (baseline, end of training, aftereffects, short-term retention) and random intercepts for participants (there were not enough observations to fit random slopes). Statistical significance was evaluated with the *lmerTest* package (Kuznetsova et al., 2017). Separate tests were conducted for each experiment. Post-hoc comparisons were conducted using the *emmeans* package (Lenth et al., 2020) with corrections for multiple comparisons.

On visual inspection of the data, it became clear that learning was not uniform—some participants clearly showed a change speech behavior that moved their F1 to the target region, while others showed no change (Figure 3A). To quantify these differences, we sorted participants into "learners" and "nonlearners" based on their behavior in the last 50 trials of the training phase. Participants whose F1 in these trials was significantly lower than baseline (towards the target), as assessed through a t-test with $\alpha = 0.05$, were classified as learners. All other participants were classified as non-learners. Classifying participants based on a metric of task success—i.e., participants who produced a significantly greater number of rewarded trials than would be expected given the standard deviation of their baseline production of words with /ɛ/—resulted in essentially the same classification pattern. Each method classified 2 participants as learners that were classified as non-learners by the other method. Pooling across all experiments, the distribution of learning is highly non-normal (Kolmogorov-Smirnov test: $D(81) = 0.67$, $p = 3 \times 10^{-32}$, Figure 3B). The figure shows learning as the change in F1 from baseline to the end of the training phase, expressed as a z-score based on baseline variability. When fitting the data with two Gaussian distributions, the two distributions have centers at -2.04 and -0.15, consistent with a group of learners who lowered their F1 and a group of non-learners who did not. We report the number of learners for each experiment and descriptive statistics for learners and non-learners. However, no inferential statistics are reported for either group since the division was done a posteriori based on the data.

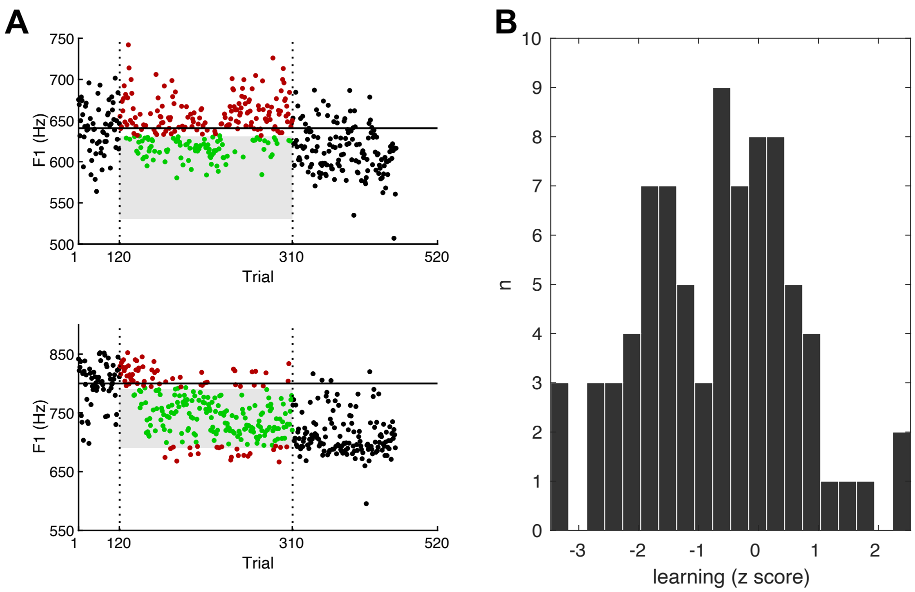


*Figure 3: **A**: Two example participants from experiment 3. The target region for receiving reward is shown in grey. Productions in the baseline and washout phases are shown as black circles. Productions during the training phase are shown as green circles if participants received a positive reward and as red circles if participants received a negative reward. The participant in the top panel shows no change in F1 frequency over the course of the experiment, while the participant in the bottom panel shows a clear shift in F1 frequency to the rewarded region that is maintained during washout. **B**: Distribution of learning for all participants across all four experiments. Learning shown as z-scored change in F1 from baseline values. The distribution is non-normal and has two peaks near -2 and 0.*

In addition to the individual experiment analyses, we conducted a series of meta-analyses across experiments. These analyses allowed us to test directly whether the different manipulations across experiments—the type of reward signal on positively rewarded trials and the presence of masking noise—affected the degree of learning. For these analyses, we conducted ANOVAs with reward signal and masking noise as fixed factors. Separate analyses were conducted for the training, aftereffects, and short-term retention measures of learning. We conducted separate analyses on both the full dataset as well as a dataset limited to only participants classified as learners. This second analysis allows us to determine whether potential differences between experiments are due to different degrees of learning or, conversely, to differences in the fraction of participants who learn without any difference in the magnitude of the change in participants who do learn. To further probe whether the proportion of learners varies across experiments, we conducted Chi-squared tests comparing the proportion of learners 1) across all experiments, 2) across experiments without masking noise (Exp 1 and 2) and with masking noise (Exp 3 and 4), and 3) across experiments with no implicit imitation target (Exp 1 and 3) and with an implicit imitation target (Exp 2 and 4).

A second goal of the meta-analysis was to further probe the potential mechanisms driving reward learning in speech. For this, we measured another set of speech parameters related to either overall variability or trial-to-trial corrections, both of which have been suggested to be related to reward in other motor domains (Dhawale et al., 2017; Wong & Shelhamer, 2011). We measured F1 variability during the baseline phase (taken only from words with /ɛ/), to test whether participants who are naturally more variable may learn better. Variability was measured in two ways: as the standard deviation of all /ɛ/ productions in the baseline phase as well as the average trial-to-trial change in these trials. We additionally measured the change in F1 standard deviation during the first 30 training trials (early learning) compared to baseline variability to assess whether learning is associated with increased exploration of the potential solution space. We also measured the F1 distance from /ɛ/ to /ɪ/ during the baseline phase (Exp 2-4 only), as participants who have a larger space between these vowels may be able to lower F1 for /ɛ/ without encroaching on /ɪ/. Lastly, we measured the average magnitude of the trial-to-trial change in F1 after trials with positive and negative reward. This allows us to assess how much participants change their production after a negative reward ("exploration") and whether participants maintain similar F1 values after positive reward ("exploitation"). Statistical tests were conducted by correlating these measures with the magnitude of learning at the end of the hold phase across participants. Results were very similar using either aftereffects or short-term retention measures.

Results

All experiments had the same structure (Figure 1). In all phases, participants spoke one word per trial our loud (*head*, *bed*, or *dead,* all containing the same /ɛ/vowel). First, participants completed a baseline phase to measure a participant-specific mean F1 values for the vowel /ɛ/. No reinforcement was given during this phase. Participants were told this phase was being used to train the computer to recognize their speech. The baseline phase was followed by a training phase where participants were instructed that the computer would attempt to recognize the word they spoke, and were instructed to try to get the computer to recognize them correctly. In the training phase, the computer recognized the "correct" word if participants produced the vowel /ɛ/ with an F1 value 10-110 Hz below their baseline mean. Positive reward was given by earning points (+10), visual feedback of the correctly recognized word, and an auditory reward. In experiments 1 and 3, auditory reward was a pleasant chime. In experiments 2 and 4, auditory reward was a token of each participant's own speech from the baseline phase with F1 for the vowel /ɛ/ shifted by -60 Hz to the middle of the reward region. Negative reward was given by losing points (-10), visual feedback of the incorrectly recognized word, and the an audio recording of the incorrectly recognized word. Learning was measured as the change in F1 from baseline at the end (last 20 trials) of the training phase. Following training, participants competed a washout where no reward was given. The washout phase was used to examine immediate aftereffects of learning (first 20 trials) as well as short-term retention of learning (trials 80-100). Experiments 1 and 2 had no masking noise. In Experiments 3 and 4, speech-shaped noise was played over headphones to mask participants' ability to hear their own speech. Results for each experiment are first presented individually. All descriptive statistics show mean and standard error. Data for all experiments is shown in Figure 4.

<u>*Experiment 1:*</u>

Experiment 1 had no masking noise and used a chime as auditory feedback for positive reward. At the group level, participants showed a very slight change in F1 values towards the target region by the end of the training phase (-2.7±5.8 Hz), which persisted into the aftereffects (-3.5±6.2 Hz) and retention (-5.9±7.4 Hz) measures. However, this change was not significant ($F(3,57) = 0.37$, $p = 0.78$). Despite the lack of an overall effect, 6/20 participants showed significant learning at an individual level, producing a change in their F1 relative to baseline values by -30.0 ± 4.4 Hz at the end of training. This change persisted into both the aftereffects (-29.1 ± 9.1 Hz) and retention (-39.9 ± 9.8 Hz) phases.

<u>*Experiment 2:*</u>

Experiment 2 had no masking noise and used a resynthesized token of each participant's own speech, with F1 shifted to the middle of the target region as auditory feedback for positive reward. Participants produced a significant change from baseline after training ($F(3,57) = 6.4$, $p < 0.001$). Across all participants, F1 was lower than baseline ($p < 0.001$) at the end of the training phase (-29.9 ± 5.8 Hz), in the aftereffects (-23.6 ± 6.1 Hz), and in retention (-23.8 ± 6.8 Hz). These phases did not differ from each other (all $p > 0.97$). At the individual level, 17/20 participants exhibited significant learning. When considering only these participants, learning was greater than for the whole group (training: -37.9 ± 4.4 Hz; aftereffects: -30.4 ± 5.8 Hz; retention: -30.0 ± 6.8 Hz).

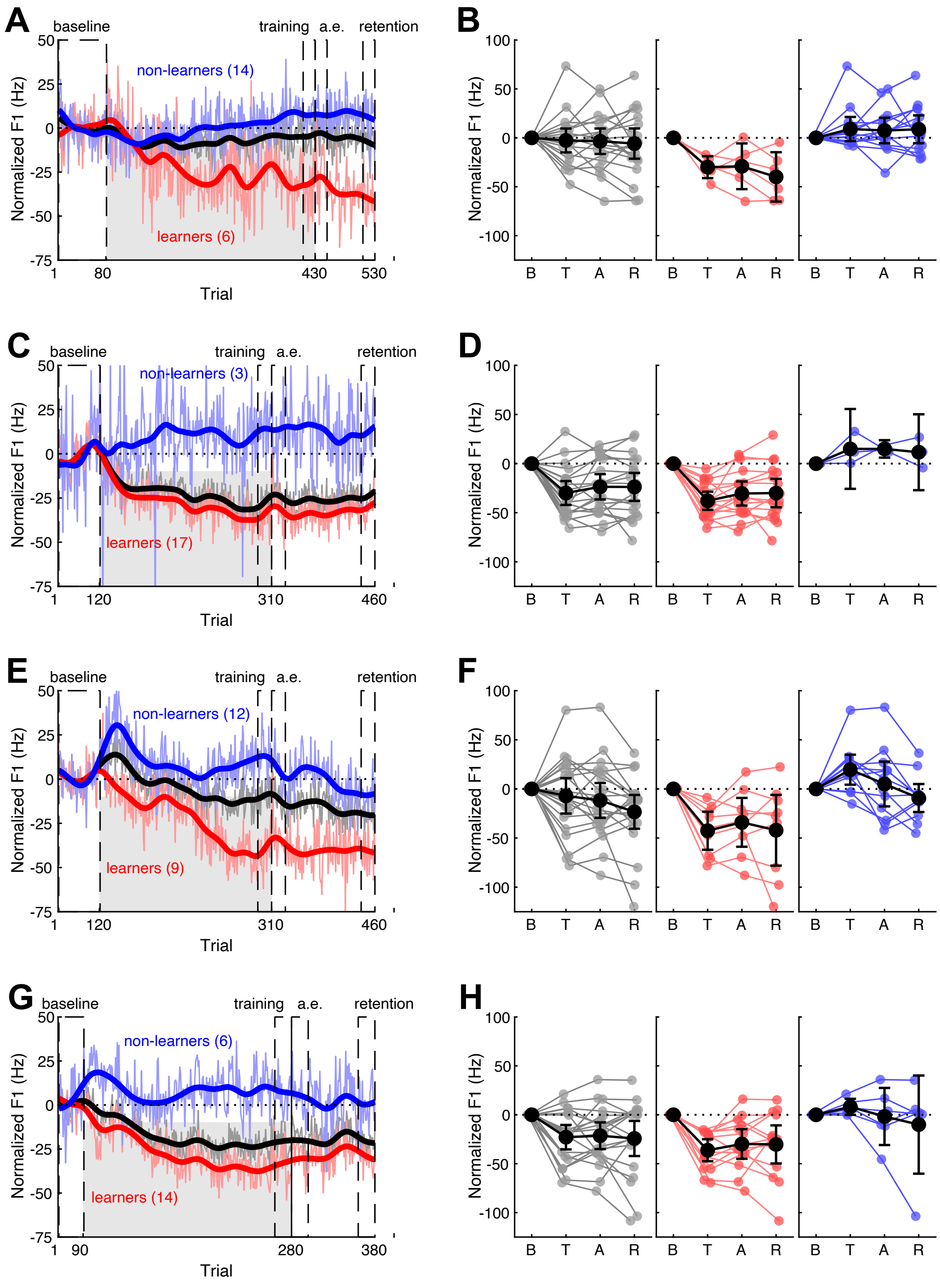
<u>*Experiment 3:*</u>

Experiment 3 had masking noise that blocked participants' perception of their own speech and used a chime as auditory feedback for positive reward. Participants did change their F1 from baseline, as reflected by a main effect of phase in the statistical model ($F(3,60) = 3.5$, $p = 0.02$). F1 was lower than baseline in all phases (training: -7.1 ± 8.6 Hz; aftereffects: -11.7 ± 8.5 Hz; retention: -23.3 ± 8.3 Hz). However, only the retention phase was significantly different from baseline ($p = 0.01$, other $p > 0.41$). The retention phase was not significantly different from either the training ($p = 0.14$) or aftereffects measures ($p = 0.41$). 9/20 participants exhibited significant learning, producing much larger changes in F1 than the group overall (training: -42.6 ± 8.4 Hz; aftereffects: -34.0 ± 10.8 Hz; retention: -42.1 ± 15.6 Hz).

<u>*Experiment 4:*</u>

Experiment 1 had masking noise that blocked participants' perception of their own speech and used a resynthesized token of each participant's own speech, with F1 shifted to the middle of the target region as auditory feedback for positive reward. Across all participants, F1 was reduced, relative to baseline, in the training (-22.9 ± 6.0 Hz), aftereffects (-21.4 ± 6.5 Hz), and retention (-24.2 ± 8.6 Hz) measures. These values were significantly lower than baseline ($F(3,57) = 12.4$, p

< 0.0001, all individual measures $p < 0.001$). There were no differences between the three phases (all $p > 0.63$). 14/20 participants showed learning at an individual level (training: -26.2 ± 5.2 Hz; aftereffects: -34.0 ± 10.8 Hz; retention: -42.1 ± 15.6 Hz).



380 *Figure 4: Change in F1 for all experiments. Experiments 1-4 are shown in order from the top down.* ***A***, ***C***, ***E***, ***G***: *mean F1 value over the course of the experiment for all participants (black), learners (red) and non-learners (blue). Raw trial averages (thin lines) as well as a smoothed running average over 10 trials (thick lines) are shown.* ***B, D, F, H***: *F1 values in the baseline (B), end of training (T), aftereffects (A), and short-term retention (R) phases for experiments 1 (**B**), 2, (**D**), 3, (**F**), and 4 (**H**). From right to left, data is shown for all participants(back), learners (red), and non-learners (blue).*

Differences between experiments:

All the presented meta-analyses comparing results from different experiments measured learning as the change in F1 from baseline to the end of the training phase. Analyses using the aftereffects produced essentially the same results. Analyses using retention showed no differences between experiments based on either the presence of masking noise or the type of auditory reward signal. In terms of overall change in F1, there was a significant effect of positive reward signal ($F(1,77) = 10.2$, $p < 0.01$), such that the change was greater in Experiments 2 and 4, where the reward signal was a token of each participant's own speech with a shifted F1 value, than in Experiments 1 and 3, where the reward signal was a chime. Contrary to our initial hypothesis, masking noise had no effect on F1 change ($F(1,77) = 0.04$, $p = 0.85$), nor was there any interaction between the presence of masking noise and the reward signal ($F(1,77) = 0.7$, $p < 0.40$).

However, the effect of reward signal was not significant when examining only participants classified as learners ($F(1,42) = 0.004$, $p = 0.95$). Neither masking, nor the interaction between masking and reward signal were significant in this group (both $p > 0.25$). This result suggests that the difference in the magnitude of F1 change between experiments with different reward signals was likely driven by differences in the proportion of learners, rather than in the degree to which participants changed F1 if they did learn. A set of chi-squared tests on the proportion of learners in each experiment supports this idea. There was an overall difference in the proportion of learners between all experiments (χ^2 (3, $N = 81$) = 14.5, $p = 0.001$). This was largely driven by a difference between experiments with different reward signals (χ^2 (3, $N = 81$) = 12.2, $p < 0.001$). There was no difference in the proportion of learners based on masking noise (χ^2 (3, $N = 81$) = 1.0, $p = 0.32$).

Across experiments, the magnitude of F1 change at the end of the training phase was not well predicted by variability. Neither baseline variability, change in variability from the baseline to the training phase, nor distance between /ɛ/ and /ɪ/ in the baseline phase predicted learning (Table 2).The exceptions are the amount of F1 change after receiving positive and negative reinforcement during the training phase. The best predictor of learning was the trial-to-trial change in F1 after receiving positive reward. Participants who produced smaller changes in these trials learned more ($R^2 = 0.26$, $p < 0.0001$). Additionally, increased learning was associated with participants who produced larger trial-to-trial F1 changes after receiving negative reward, though the magnitude of this effect was relatively modest ($R^2 = 0.05$, $p = 0.03$). Results for all factors are shown in Figure 5.

Based on the significant relationship between change after positive reward and learning, we considered whether the difference in overall learning magnitude (driven by the proportion of learners) between experiments with informative and non-informative reward signals could be related to differences in the degree to which participants shifted their productions after positive reward. For example, participants may be less likely to shift their production after they hear a word with the “correct” F1. However, we found no evidence that the magnitude of shift after positive reward differed between studies with different reward signals ($F(1,77) = 2.4$, $p = 0.13$) or based on the presence of masking noise ($F(1,77) = 0.3$, $p = 0.59$). There was similarly no significant interaction between the two factors ($F(1,77) = 0.003$, $p = 0.96$).

We additionally examined whether variability in the baseline phase or early in the training phase affected the percentage of trials that were produced with F1 in the target region. Recall that the target region ranged from 10 to 110 Hz below each participant’s baseline mean. This was chosen to ensure that all participants received reward on some trials without needing to change their baseline F1 values. Indeed, baseline variability, as measured by the standard deviation of F1, ranged from 13-56 Hz. Variability in the first 50 trials of the training phase ranged from 12-127 Hz. Even at the small end of this range, we would expect participants to receive positive reward on at least 20% of trials. In our data, all participants received at least some positive reward for trials with F1 within the target region during the training phase, as expected (1.2%-94.4% of trials, across participants). There was a small but significant relationship between baseline variability and percentage of trials produced with F1 in the target region across the training phase ($R^2 = 0.03$, $p = 0.04$). However, there was no relationship between variability in the training phase itself and percentage of trials with F1 in the target region ($R^2 = 0.002$, $p = 0.028$). Together, these results suggest little relationship between variability and percentage of rewarded trials.

Table 2: Correlations between the change in F1 from the baseline to the end of the hold phase and various potential predictors of learning

MEASURE	R^2	***P***
Baseline standard deviation	0.01	0.16
Baseline trial-to-trial change	-0.01	0.56
Increase in standard error from baseline to training	-0.01	0.78
/ɛ/ - /ɪ/ distance	-0.02	0.78
F1 change after negative reward	0.05	0.02*
F1 change after positive reward	0.28	< 0.0001*

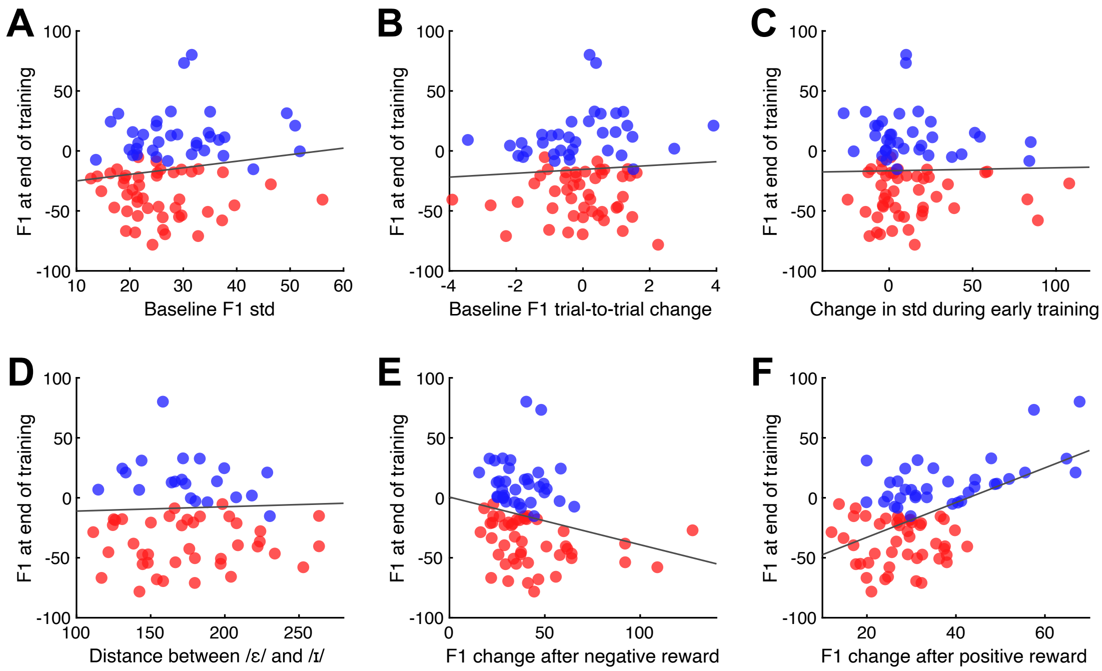


Figure 5: Potential factors associated with learning, defined as the magnitude of F1 change from baseline at the end of the training phase. In all panels, learners are shown in red and non-learners in blue. A solid black line indicated the regression model. ***A:*** *Baseline standard deviation of F1.* ***B****: trial-to-trial change in F1 in the baseline phase.* ***C:*** *change in standard deviation of F1 from the baseline phase to the training phase.* ***D****: F1 distance between /ɛ/ and /ɪ/ in the training phase.* ***E****: trial-to-trial change in F1 after receiving negative reward in the training phase.* ***F****: trial-to-trial change in F1 after receiving positive reward in the training phase.*

<u>*Strategy use and engagement:*</u>

Strategy use was assessed in a follow-up survey after experiments 2 and 3. Participants were asked the question "Did you develop any techniques or strategies during the task? If so, what was that strategy?". In Experiment 2, 16/20 participants reported using a strategy. Only 4 of these strategies related to changing the quality of the vowel, which was required to perform the task successfully. Despite the presence of a highly informative auditory reward signal for positive reward (a token of the participants' own speech with F1 shifted to the middle of the target region), only 2/20 participants reported imitating the reward signal (both of these participants were classified as learners. In Experiment 3, 19/21 participants reported using a strategy. Of these, only 2 were plausibly related to changing vowel quality. Positive reward was accompanied by a chime in this experiment, so participants could not imitate the reward signal.

Individual participant responses are reported in the Appendix.

Participants in these studies were also asked to rate how engaging they found the task. Specifically, they were asked to rate their agreement with the statements "I was motivated to perform well in this task" and "I was motivated by the points I was earning" on a scale from 0 (disagree) to 100 (agree). The median overall motivation was 95 (mean: 84.5, 9 participants reported "yes" instead of reporting a number). The median motivation related to the points was 100 (mean: 83.5, 8 participants reported "yes" and 1 participant reported "no" instead of reporting a number).

Discussion

In a set of four experiments, we examined whether positive and negative reinforcement alone could cause participants to change their speech production in the absence of any explicit instruction. Specifically, we examined whether participants could learn to lower the first formant of the vowel /ε/, analogous to a widely-demonstrated change that can be induced through sensorimotor adaptation. We tested two additional aspect of reinforcement learning. First, we examined the effects of the auditory signal given for positive reward, comparing an arbitrary sound (a chime) with a potentially-informative sound (a resynthesized version of each participant's own speech, with F1 shifted to the center of the target region). We hypothesized that the more informative reward signal would lead to a larger magnitude of learning. Second, we examined the effect of masking auditory feedback of participants' speech would affect learning. Based on previous work in reaching showing that visual feedback of hand position reduces the effectiveness of reinforcement learning to change reach angle, we hypothesized that learning would be reduced when auditory feedback was available, as shifting F1 in this case would conflict with participants internal targets for speech.

Our results showed that reinforcement can indeed drive participants to learn to shift their vowel production even in the absence of any explicit instruction. While we observed learning in some participants in all experiments, the average magnitude of learning was greater in experiments with informative reward signals. This increase in average learning, however, was driven by a greater proportion of participants who were able to learn to shift their F1 towards the target region. When examining only participants who exhibited learning, the magnitude of learning was similar across studies. Thus, it seems that an informative reinforcement signal makes learning more likely, but does not affect the magnitude of learning.

Perhaps surprisingly, this effect does not seem to be driven by explicit imitation of the informative reinforcement signal. In Experiment 2, 17/20 participants were classified as learners. However, only 2/20 reported imitating the reinforcement signal. These results suggest that the benefit of an informative reward signal does not come from allowing for explicit imitation, but rather serves as an implicit guide to achieve success. One possibility is that participants are implicitly imitating the reward signal, without being consciously aware. This is similar to the concept of phonetic convergence or accommodation, where speakers adjust their own productions to align with speech that they hear even over very short time scales (e.g., Babel, 2010; Fowler et al., 2003; Goldinger, 1998; Pardo, 2006, 2013; Pickering & Garrod, 2013). Alternatively, the resynthesized speech reward signal may give participants implicit information about the dimension along which speech must be altered to achieve success, which may be important for reinforcement learning in high-dimensional motor systems (Manley et al., 2014). These results suggest that providing informative feedback may help reinforcement learning without the need to instruct participants to explicitly imitate the feedback. This finding has important clinical implications, as explicit instruction about how to change motor behaviors may reduce the retention of learning after training generally (Green & Flowers, 1991; Hasson et al., 2015; Shea et al., 2001; Winstein & Schmidt, 1990), and in some neurological disorders (Boyd & Winstein, 2004, 2006; Masters et al., 2004). Interestingly, the resynthesized speech feedback condition is somewhat similar to the “reformulations” of infant speech typically made by caregivers, where they repeat the word they perceive the infant to have intended with a more adult-like pronunciation (Howard & Messum, 2011). The current results showing that feedback with implicit production targets increase learning suggests that such reformulations may in fact facilitate infant speech learning even in the absence of any attempts to imitate or match adult-like speech (c.f. Guenther, 2016)

Contrary to our second hypothesis, we found no evidence that masking auditory feedback of participants’ speech affected either the magnitude or the probability of learning. This is contrary to previously demonstrated results in reaching. In these tasks, participants are presented with a visual target, and must learn to alter the angle or location of their reach away from the target to receive reward. Providing visual feedback about the position of the hand in these task seems to bias the system to weight sensory errors over reinforcement feedback, such that the effect of reinforcement on learning is eliminated (Cashaback et al., 2017). Here, we found no such effect for speech when auditory feedback is available. This may result from an important difference in how speech and reaching targets are defined. Targets in laboratory reaching tasks are externally defined (e.g., move your hand to the circle on the screen). However, movement targets in speech are defined internally by each participant. Thus, when participants change their F1 in response to reinforcement feedback, they may be simultaneously altering the intended target of their speech, eliminating any potential conflict between the sensory and reinforcement learning systems. As stated above, speech targets are relatively flexible even at short time scales, which provides some support for this idea. More broadly, these results suggest that the interaction between sensory error-based learning and reinforcement learning is complex and potentially reliant on whether movement targets are defined externally in the environment or internally.

Our data suggest that the primary factor driving learning is the magnitude of the change in F1 after trials that receive positive reward during the training phase. Participants who change F1 less after positive reward learn more, suggesting they are more capable of "exploiting" the correct behavior to receive reward. There was also a significant, but small, relationship between learning and the magnitude of F1 change after negative reward such that participants who have a greater change in F1 after negative reward learn more. This is consistent with the idea that reinforcement learning is accomplished through an exploration of the solution space. However, this seems to play a minor role in learning in these experiments. Somewhat surprisingly, learning was not related to production variability in the baseline phase or to the change in variability from the baseline to the hold phase. It may have been expected that participants who were more variable were more likely to receive positive reward and thus, to learn more readily (Dhawale et al., 2017) or that higher variability in the dimension of control that must be changed would itself facilitate learning (Wu et al., 2014); however, this seems to not be the case here. The lack of an effect between learning and variability has also been reported for some reaching tasks (Cashaback et al., 2017).

Lastly, we found that the changes in F1 caused by reinforcement learning were maintained through the washout phase, up to 150 trials after reinforcement was removed. This is substantially longer than changes in formant values caused by sensorimotor adaptation are retained; in this case, speakers return to producing formant values near to their baseline within 30 trials. This has at least two important implications. First, from a theoretical side, it suggest that reinforcement learning caused participants to shift their production goals to the target region. Without anything to push them back to their pre-training targets, they maintained these goals after reinforcement was removed. Second, from a clinical view, this suggest that reinforcement learning has the potential to cause long-lasting changes in speech production, potentially even after a relatively short training session. This suggest reinforcement learning is a likely powerful clinical tool for speech rehabilitation, consistent with previous suggestions in limb control (Roemmich & Bastian, 2018).

In sum, our results suggest that reinforcement learning is an active process in speech motor control and that it can cause changes in behavior even in the absence of explicit instruction. Reinforcement learning is not affected by the availability of auditory feedback and is retained after reinforcement is removed. Together, this suggests that reinforcement operates by causing a shift in the intended movement target. Notably, this shift is at least largely implicit, as few participants reported using any explicit strategies related to changing vowel quality. These results suggest altering behavior through reinforcement is possible even in complex, high-dimensional motor tasks such as speech production. These results suggest reinforcement is a plausible mechanism for early speech development, consistent with recent computation models (Howard & Messum, 2011; Messum & Howard, 2012; Warlaumont, 2014; Warlaumont et al., 2013; Warlaumont & Finnegan, 2016). Moreover, they suggest reinforcement may be a powerful clinical tool for speech rehabilitation, even without explicit instruction or detailed “knowledge of performance/results” feedback provided about errors. However, potential differences between speech and other motor domains, such as the effects of sensory masking, suggest reinforcement learning should be further studied to maximize its effectiveness in rehabilitative paradigms.

References

Adams, S., & Page, A. (2000). Effects of Selected Practice and Feedback Variables on Speech Motor Learning. *Journal of Medical Speech·Language Pathology*, *8*(4), 215–220.

Adams, S., Page, A., & Jog, M. (2002). Summary Feedback Schedules and Speech Motor Learning in Parkinson's Disease. *Journal of Medical Speech-Language Pathology*, *10*(4), 215–220.

Babel, M. (2010). Dialect divergence and convergence in New Zealand English. *Lang. Soc.*, *39*(04), 437–456. https://doi.org/10.1017/s0047404510000400

Ballard, K. J., Granier, J. P., & Robin, D. A. (2000). Understanding the nature of apraxia of speech: Theory, analysis, and treatment. *Aphasiology*, *14*(10), 969–995. https://doi.org/10.1080/02687030050156575

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). lme4: Linear mixed-effects models using Eigen and S4. *R Package Version*, *1.1-12*(7).

Bislick, L. P., Weir, P. C., & Spencer, K. A. (2013). Investigation of Feedback Schedules on Speech Motor Learning in Individuals with Apraxia of Speech. *Journal of Medical Speech· Language Pathology*, *20*(4), 18–23.

Bislick, L. P., Weir, P. C., Spencer, K., Kendall, D., & Yorkston, K. M. (2012). Do principles of motor learning enhance retention and transfer of speech skills? A systematic review. *Aphasiology*, *26*(5), 709–728. https://doi.org/10.1080/02687038.2012.676888

Boersma, P., & Weenink, D. (2019). *Praat: Doing phonetics by computer.* (Version 6.0.47)

[Computer software]. ttp://www.praat.org/

Boyd, L. A., & Winstein, C. J. (2004). Providing explicit information disrupts implicit motor learning after basal ganglia stroke. *Learn Mem*, *11*(4), 388–396. https://doi.org/10.1101/lm.80104

Boyd, L. A., & Winstein, C. J. (2006). Explicit Information Interferes with Implicit Motor Learning of Both Continuous and Discrete Movement Tasks After Stroke. *Journal of Neurologic Physical Therapy*, *30*(2), 46–57. https://doi.org/10.1097/01.NPT.0000282566.48050.9b

Cai, S., Boucek, M., Ghosh, S., Guenther, F. H., & Perkell, J. (2008). A System for Online Dynamic Perturbation of Formant Trajectories and Results from Perturbations of the Mandarin Triphthong /iau/. *Proceedings of the 8th International Seminar on Speech Production*, 65–68.

Cashaback, J. G. A., McGregor, H. R., Mohatarem, A., & Gribble, P. L. (2017). Dissociating error-based and reinforcement-based loss functions during sensorimotor learning. *PLoS Comput Biol*, *13*(7), e1005623. https://doi.org/10.1371/journal.pcbi.1005623

Daliri, A., & Dittman, J. (2019). Successful auditory motor adaptation requires task-relevant auditory errors. *Journal of Neurophysiology*, *122*(2), 552–562. https://doi.org/10.1152/jn.00662.2018

Dhawale, A. K., Smith, M. A., & Ölveczky, B. P. (2017). The Role of Variability in Motor Learning. *Annu Rev Neurosci*, *40*, 479–498. https://doi.org/10.1146/annurev-neuro-072116-031548

Duffy, J. R. (2013). *Motor speech disorders: Substrates, differential diagnosis, and management* (Third edition). Elsevier.

Fowler, C. A., Brown, J. M., Sabadini, L., & Weihing, J. (2003). Rapid access to speech gestures in perception: Evidence from choice and simple response time tasks. *Journal of Memory and Language*, *49*(3), 396–413.

Galea, J. M., Mallia, E., Rothwell, J., & Diedrichsen, J. (2015). The dissociable effects of punishment and reward on motor learning. *Nat Neurosci*, *18*(4), 597–602. https://doi.org/10.1038/nn.3956

Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review*, *105*(2), 251–279.

Green, T. D., & Flowers, J. H. (1991). Implicit versus Explicit Learning Processes in a Probabilistic, Continuous Fine-Motor Catching Task. *Journal of Motor Behavior*, *23*(4), 293–300. https://doi.org/10.1080/00222895.1991.9942040

Guenther, F. H. (2016). *Neural control of speech*. The MIT Press.

Hasson, C. J., Manczurowsky, J., & Yen, S.-C. (2015). A reinforcement learning approach to gait training improves retention. *Frontiers in Human Neuroscience*, *9*. https://doi.org/10.3389/fnhum.2015.00459

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

Howard, I. S., & Messum, P. (2011). Modeling the development of pronunciation in infant speech acquisition. *Motor Control*, *15*(1), 85–117.

Howard, I. S., & Messum, P. (2014). Learning to pronounce first words in three languages: An investigation of caregiver and infant behavior using a computational model of an infant. *PLoS One*, *9*(10), e110334. https://doi.org/10.1371/journal.pone.0110334

Hula, S. N. A., Robin, D. A., Maas, E., Ballard, K. J., & Schmidt, R. A. (2008). Effects of feedback frequency and timing on acquisition, retention, and transfer of speech skills in acquired apraxia of speech. *Journal of Speech, Language & Hearing Research*, *51*(5), 1088–1113.

Izawa, J., & Shadmehr, R. (2011). Learning from sensory and reward prediction errors during motor adaptation. *PLOS Computational Biology*, *7*(3), e1002012. https://doi.org/10.1371/journal.pcbi.1002012

Katz, W. F., McNeil, M. R., & Garst, D. M. (2010). Treating apraxia of speech (AOS) with EMA-supplied visual augmented feedback. *Aphasiology*, *24*(6–8), 826–837. https://doi.org/10.1080/02687030903518176

Kooij, K. van der, & Overvliet, K. E. (2016). Rewarding imperfect motor performance reduces adaptive changes. *Experimental Brain Research*, *234*(6), 1441–1450. https://doi.org/10.1007/s00221-015-4540-1

Krakauer, J. W. (2015). The applicability of motor learning to neurorehabilitation. In V. Dietz & N. Ward (Eds.), *Oxford Textbook of Neurorehabilitation* (pp. 55–64). Oxford University Press. https://doi.org/10.1093/med/9780199673711.003.0007

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software*, *82*(13). https://doi.org/10.18637/jss.v082.i13

Lametti, D. R., Nasir, S. M., & Ostry, D. J. (2012). Sensory preference in speech production revealed by simultaneous alteration of auditory and somatosensory feedback. *Journal of Neuroscience*, *32*(27), 9351–9358. https://doi.org/10.1523/JNEUROSCI.0404-12.2012

Lametti, D. R., Smith, H. J., Watkins, K. E., & Shiller, D. M. (2018). Robust Sensorimotor Learning during Variable Sentence-Level Speech. *Current Biology*, *28*(19), 3106-

3113.e2. https://doi.org/10.1016/j.cub.2018.07.030

Lenth, R., Singmann, H., Love, J., Buerkner, P., & Herve, M. (2020). *emmeans: Estimated Marginal Means, aka Least-Squares Means*.

MacDonald, E. N., Purcell, D. W., & Munhall, K. G. (2011). Probing the independence of formant control using altered auditory feedback. *J Acoust Soc Am*, *129*(2), 955–965. https://doi.org/10.1121/1.3531932

Manley, H., Dayan, P., & Diedrichsen, J. (2014). When money is not enough: Awareness, success, and variability in motor learning. *PLoS One*, *9*(1), e86580. https://doi.org/10.1371/journal.pone.0086580

Masters, R. S. W., MacMahon, K. M. A., & Pall, H. S. (2004). Implicit Motor Learning in Parkinson's Disease. *Rehabilitation Psychology*, *49*(1), 79–82. https://doi.org/10.1037/0090-5550.49.1.79

Messum, P., & Howard, I. S. (2012). Speech development: Toddlers don't mind getting it wrong. *Curr Biol*, *22*(5), R160-1. https://doi.org/10.1016/j.cub.2012.01.032

Mitsuya, T., MacDonald, E. N., Munhall, K. G., & Purcell, D. W. (2015). Formant compensation for auditory feedback with English vowels. *The Journal of the Acoustical Society of America*, *138*(1), 413–424.

Nikooyan, A. A., & Ahmed, A. A. (2015). Reward feedback accelerates motor learning. *J Neurophysiol*, *113*(2), 633–646. https://doi.org/10.1152/jn.00032.2014

Pardo, J. S. (2006). On phonetic convergence during conversational interaction. *J Acoust Soc Am*, *119*(4), 2382–2393.

Pardo, J. S. (2013). Measuring phonetic convergence in speech production. *Front Psychol*, *4*, 559. https://doi.org/10.3389/fpsyg.2013.00559

Parrell, B., Agnew, Z., Nagarajan, S., Houde, J. F., & Ivry, R. B. (2017). Impaired Feedforward Control and Enhanced Feedback Control of Speech in Patients with Cerebellar Degeneration. *Journal of Neuroscience*, *37*(38), 9249–9258. https://doi.org/10.1523/JNEUROSCI.3363-16.2017

Pickering, M. J., & Garrod, S. (2013). An integrated theory of language production and comprehension. *Behav Brain Sci*, *36*(4), 329–347. https://doi.org/10.1017/S0140525X12001495

Purcell, D. W., & Munhall, K. G. (2006). Adaptive control of vowel formant frequency: Evidence from real-time formant manipulation. *Journal of Acoustical Society of America*, *120*(2), 966–977.

R Core Team. (2013). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. http://www.R-project.org/

Roemmich, R. T., & Bastian, A. J. (2018). Closing the Loop: From Motor Neuroscience to Neurorehabilitation. *Annual Review of Neuroscience*, *41*(1), 415–429. https://doi.org/10.1146/annurev-neuro-080317-062245

Schmidt, R. A., & Lee, T. D. (2011). *Motor control and learning: A behavioral emphasis* (5th ed). Human Kinetics.

Shea, C. H., Wulf, G., Whitacre, C. A., & Park, J. H. (2001). Surfing the implicit wave. *The Quarterly Journal of Experimental Psychology Section A*, *54*(3), 841–862. https://doi.org/10.1080/713755993

Shiller, D. M., Sato, M., Gracco, V. L., & Baum, S. R. (2009). Perceptual recalibration of speech sounds following speech motor learning. *Journal of Acoustical Society of America*, *125*(2), 1103–1113. https://doi.org/10.1121/1.3058638

Steinhauer, K., & Grayhack, J. P. (2000). The role of knowledge of results in performance and learning of a voice motor task. *Journal of Voice*, *14*(2), 137–145. https://doi.org/10.1016/S0892-1997(00)80020-X

Therrien, A. S., Wolpert, D. M., & Bastian, A. J. (2016). Effective reinforcement learning following cerebellar damage requires a balance between exploration and motor noise. *Brain*. https://doi.org/10.1093/brain/awv329

Tourville, J. A., Cai, S., & Guenther, F. (2013). Exploring auditory-motor interactions in normal and disordered speech. *Proceedings of Meetings on Acoustics*, *19*, 060180. https://doi.org/10.1121/1.4800684

Villacorta, V. M., Perkell, J. S., & Guenther, F. H. (2007). Sensorimotor adaptation to feedback perturbations of vowel acoustics and its relation to perception. *Journal of Acoustical Society of America*, *122*(4), 2306–2319. https://doi.org/10.1121/1.2773966

Warlaumont, A. S. (2014). Reinforcement-modulated self-organizaiton in infant motor speech learning. In J. Mayor & P. Gomez (Eds.), *Computational Models of Cognitive Processes* (pp. 112–124). World Scientific.

Warlaumont, A. S., & Finnegan, M. K. (2016). Learning to Produce Syllabic Speech Sounds via Reward-Modulated Neural Plasticity. *PLoS One*, *11*(1), e0145096. https://doi.org/10.1371/journal.pone.0145096

Warlaumont, A. S., Westermann, G., Buder, E. H., & Oller, D. K. (2013). Prespeech motor learning in a neural network using reinforcement. *Neural Netw*, *38*, 64–75. https://doi.org/10.1016/j.neunet.2012.11.012

Winstein, C. J., & Schmidt, R. A. (1990). Reduced frequency of knowledge of results enhances motor skill learning. *Journal of Experimental Psychology: Learning, Memory, and*

Cognition, *16*(4), 677–691. https://doi.org/10.1037/0278-7393.16.4.677

Wong, A. L., & Shelhamer, M. (2011). Exploring the Fundamental Dynamics of Error-Based Motor Learning Using a Stationary Predictive-Saccade Task. *PLoS ONE*, *6*(9), e25225. https://doi.org/10.1371/journal.pone.0025225

Wu, H. G., Miyamoto, Y. R., Gonzalez Castro, L. N., Ölveczky, B. P., & Smith, M. A. (2014). Temporal structure of motor variability is dynamically regulated and predicts motor learning ability. *Nat Neurosci*, *17*(2), 312–321. https://doi.org/10.1038/nn.3616

Appendix

Participant responses to post-experiment survey about strategy use.

Exp	**Used a strategy**	**Strategy related to changing vowel quality**	**Imitated feedback**	**Strategy**
Exp 2	✓		✓	listening to correct words said by the computer and trying to mimic them
	✓		✓	If I got a word correct, I tried to copy the way the playback of myself said the word
	✓	✓		Yes. For "bed," I tried to make more of a schwa sound, therefore rounding my lips more than usual and keeping my mouth more closed. Usually, when I say "bed," I open my mouth wider. For the words "head" and "dead," I tried to not let my pitch go up and tried to keep my pitch level throughout the vowel sound. Again, I tried to keep my mouth more closed than usual and tried to keep the vowel sound consistent. If I said the words with a more open mouth like I usually do, the computer registered them as "had" or "dad."
				not really i just tried to not say the four letter words too fast
	✓			At times I would look at my reflection and read the word while looking at my reflection to keep myself entertained and from dozing off. Another strategy I used was counting on my fingers to see how many words I would say.
				I did not
	✓			I tried to be as articulate as possible when I said the words
	✓	✓		To adjust my pronunciation.
	✓			I realized that the d on each of the words was short.
	✓			I tried to pronounce the words very clearly

	✓	✓		Yes. If I got the word wrong and had points deducted then I tried to change how I pronounced the word a little bit. I usually tried focusing on the middle part of the word and changed how I pronounced that.
	✓			Speaking more clearly by focusing on one word at a time
	✓			The words dead and head were similar to say, found if you kept A silent it gave you the points
	✓			I TRIED TO PRONOUNCE THE WORDS WITH A STEADY TONE AND EMPHASIS ON THE LAST PHONEME WHICH GOT ME POINTS AND THEN FOR THE RETRAIN SECTION I TRIED TO CHANGE THE WAY I PRONOUNCED THEM TO MATCH UP WITH THE SECOND TASK
	✓			yes for head i looked at the e so i remembered to pronounce it correctly.
				No, I did not
	✓			to fully stretch out my e's
				not really-i thought that emphasizing certain parts of the word helped at times.
	✓	✓		CHANGING THE WAY I PRONOUNCED WORDS IN ORDER TO GET POINTS
	✓			During the testing phase, if I spoke lower I was more likely to get the answer right and get the points.
Exp 3	✓		n/a	I tried to keep my voice low and pronunciate each word
			n/a	no.
	✓		n/a	I tried to read the words very clearly and with diction.
	✓		n/a	i noticed when i articulated the /d/ at the end of the sentence i gained points
			n/a	No
	✓		n/a	i tried to say the word not like how the person thought i was saying it on the screen
	✓		n/a	If I got one right where the computer gave me 10 points (rare) I would try to not move at all and hope the next word was the same or similar in order to get another one right.
	✓	✓	n/a	I tried to do a 'short e' sound as much as possible, as my words kept getting confused with 'a' sounds. I tried to change my sound so that the computer would recognize it, without it seeming forced.
	✓		n/a	yes, when I got a word correct I tried to repeat the next word in the exact same way by positioning/moving my mouth the same way
	✓		n/a	Yes- I repeated words a certain way once I finally noticed how I was expected to say them to earn points.
	✓		n/a	I tried to annunciate my E's more
			n/a	No, just kept going for it
	✓		n/a	I started to stress the vowels making them longer in order for the machine to approve them
	✓		n/a	I noticed that the computer thought I used "ad" endings a lot more than I did so I would try and pronounce the "e" sounds more in words that this occurred. This did not always work.
	✓		n/a	How to speak clearly so words are apparent

	✓		n/a	For words with an "e" sound like "bed" or "dead" or "head" I had to prolong the "e" sound for the computer to understand. For too short of a word it would think I said an "a" sound like "bad" or "dad" or "had". I could tell if I was saying the word long enough by listening to the static in the headphones. It needed to be a certain length of static before I knew I should end the word with the next consonant. Also, I wasn't staring at the screen. I knew the next word was up when the static started to play in the headphones, and then I read it from the screen. When I was originally staring at the screen I could predict what word was coming next, or at least my brain was trying to, and then I felt like I had a harder time saying the next word because I already thought I knew what the word should be. So I stopped looking and only looked up when I heard static.
	✓	✓	n/a	I'm not sure, but maybe pronouncing words slightly different to see if changes would make the word correct. If it was correct, then using that change when the same word came up again.
	✓		n/a	I pronounced "head" with less emphasis on the "ea" part.
	✓		n/a	I tried to enunciate my vowels
	✓		n/a	I started to say the words head, bed, dead quicker
	✓		n/a	I contorted my mouth and diaphragm in ways I did not think possible in order to enunciate the words. My main strategy was to try and hit the first syllable as hard as possible.

738