Research Article

The Impact of Feedback Frequency on Performance in a Novel Speech Motor Learning Task

Mara Steinberg Lowe^a and Adam Buchwald^a

Purpose: This study investigated whether whole nonword accuracy, phoneme accuracy, and acoustic duration measures were influenced by the amount of feedback speakers without impairment received during a novel speech motor learning task.

Method: Thirty-two native English speakers completed a nonword production task across 3 time points: practice, short-term retention, and long-term retention. During practice, participants received knowledge of results feedback according to a randomly assigned schedule (100%, 50%, 20%, or 0%). Changes in nonword accuracy, phoneme accuracy, nonword duration, and initial-cluster duration were compared among feedback groups, sessions, and stimulus properties.

Results: All participants improved phoneme and whole nonword accuracy at short-term and long-term retention

time points. Participants also refined productions of nonwords, as indicated by a decrease in nonword duration across sessions. The 50% group exhibited the largest reduction in duration between practice and long-term retention for nonwords with native and nonnative clusters.

Conclusions: All speakers, regardless of feedback schedule, learned new speech motor behaviors quickly with a high degree of accuracy and refined their speech motor skills for perceptually accurate productions. Acoustic measurements may capture more subtle, subperceptual changes that may occur during speech motor learning.

Supplemental Materials: <https://doi.org/10.23641/asha.5116324>

This special issue contains selected papers from the March 2016 Conference on Motor Speech held in Newport Beach, CA.

Previous studies of motor learning in nonspeech domains (e.g., limb movements) have identified parameters for structuring practice and providing feedback that enhance motor skill acquisition (Schmidt & Lee, 2005). Broadly speaking, there is emerging evidence that incorporating these principles during speech motor learning tasks may be effective for promoting speech motor learning both in speakers without impairment and in individuals with motor speech disorders (e.g., Austermann Hula, Robin, Maas, Ballard, & Schmidt, 2008; Maas et al., 2008). However, the optimal approach to structuring practice and feedback in speech remains an active area of research (Bislick, Weir, Spencer, Kendall, & Yorkston, 2012; Maas et al., 2008; Wambaugh, Nessler, Cameron, & Mauszycki, 2013).

One parameter of interest in speech motor learning is feedback frequency (for a review, see Maas et al., 2008). Interventions for motor speech disorders, such as articulatory–kinematic or rate/rhythm approaches, do not always explicitly state instructions for feedback delivery (for an example of speech treatment recommendations for apraxia, see Wambaugh, Duffy, McNeil, Robin, & Rogers, 2006). Because little information is available on the optimal amount of feedback, high-frequency feedback is often provided (Ballard, Granier, & Robin, 2000). However, studies of motor learning from other domains (e.g., limb movements; Winstein & Schmidt, 1990) suggest that retention of newly learned motor skills is enhanced when feedback is not provided on every trial during practice and that the optimal amount of feedback needed may decrease as the learner becomes more proficient in performing a motor task (Wulf, Shea, & Matschiner, 1998).

The guidance hypothesis has been proposed to explain the seemingly counterintuitive finding that less knowledge of results feedback (e.g., whether performance was correct or incorrect; KR feedback) can result in enhanced retention of newly learned motor skills (Lee, White, & Carnahan, 1990; Salmoni, Schmidt, & Walter, 1984). The guidance hypothesis proposes that negative effects of high-frequency feedback occur for several reasons. First, the learner may become too dependent on frequent feedback (i.e., the feedback becomes

^aNew York University

Correspondence to Mara Steinberg Lowe: marasteinberglowe@nyu.edu

Editor: Yana Yunusova

Associate Editor: Ignatius Nip

Received May 26, 2016

Revision received September 27, 2016

Accepted November 17, 2016

https://doi.org/10.1044/2017_JSLHR-S-16-0207

Disclosure: The authors have declared that no competing interests existed at the time of publication.

part of the task), so that when it is removed at a later retention test, performance suffers. Second, KR feedback may interfere with information processing or reduce the need to perform memory retrieval processes, which are needed for learning and which hinder development of error-detection skills (Wulf & Shea, 2004). Third, KR feedback encourages the learner to correct even small errors that may occur from inherent variability present in the motor system. This may result in many “maladaptive short-term corrections” and failure to learn a stable and consistent behavior (Schmidt, 1991; Winstein & Schmidt, 1990; Wulf & Shea, 2004).

There is evidence that providing KR feedback at a reduced frequency (e.g., after only some trials) is advantageous for speech motor learning in speakers without impairment, and there is no evidence that it is disadvantageous (Bislick et al., 2012; Maas et al., 2008). Adams and Page (2000) compared the effects of high-frequency feedback (provided after every trial) with summary feedback provided after every fifth trial on the learning of a novel speech task in 20 healthy young adults. Although no significant differences in performance were found during acquisition, only the group that received summary feedback (20%) exhibited significantly lower error scores at a retention session 2 days later. This finding, where reduced feedback frequency resulted in better performance during a delayed retention test (also see Kim, LaPointe, & Stierwalt, 2012), is consistent with the limb motor learning literature (e.g., Winstein & Schmidt, 1990).

Similar results were found by Steinhauer and Grayhack (2000). Participants who received reduced KR feedback (50%) during a vowel nasalance task had better retention scores 24 hr after the initial practice session, as compared with participants who received 100% feedback. Participants who received reduced KR feedback (50%) also demonstrated better transfer skills when asked to nasalize a different vowel. Participants who received no feedback during practice (0%) had similar performance outcomes as the 50% group, suggesting that, at least for some speech motor learning tasks, providing feedback on every trial may be worse than not providing any feedback.

There is also evidence that reduced feedback frequency may improve relearning of speech targets (e.g., words, syllables, or sounds) relative to 100% feedback in individuals with motor speech impairments resulting from disorders such as Parkinson’s disease (Adams, Page, & Jog, 2002), childhood apraxia of speech (Maas, Butalla, & Farinella, 2012), and acquired apraxia of speech. For example, Austermann Hula et al. (2008) used a single-participant, alternating-treatment experimental design to evaluate how feedback frequency influenced relearning of speech skills in four people with acquired apraxia of speech. All participants simultaneously practiced two types of speech targets (described as “treatment conditions”), which differed in the manner of production (plosives vs. fricatives). One condition received high-frequency feedback (100%), whereas the other received low-frequency feedback (60%). Reduced feedback frequency resulted in enhanced performance as compared with high feedback frequency during retention and transfer for two of the four participants.

The studies described here and summarized in Table 1 compared feedback frequency under two conditions, high

feedback frequency (100% of trials) versus low feedback frequency (<100% of trials), with the exception of Steinhauer and Grayhack (2000), who included a third comparison condition that received no feedback. Overall, these studies demonstrate that speakers without impairment perform better at retention when provided with lower-frequency feedback during training and that the advantage for reduced-frequency feedback is more variable for children and adults with motor speech disorders. However, the different levels of reduced feedback were never directly compared with one another. Thus, the optimal level of feedback frequency to enhance speech motor learning, if one exists, remains unknown. In this study, we systematically investigated how different levels of feedback frequency (100%, 50%, 20%, 0%) affect acquisition and retention (short-term and long-term) of motor speech skills.

Speech Motor Learning in Speakers Without Impairment

Much of the evidence supporting the principles of motor learning within the limb motor domain has come from studies of people with intact motor systems, and then it has been applied to learning in populations with motor impairments (Schmidt & Lee, 2005). Likewise, studying how the structure of practice and feedback affects acquisition and retention of novel speech sequences in speakers without impairment can inform translation to use in clinical populations. Although using speakers without impairment allows for larger sample sizes (Kim et al., 2012), one major challenge is that speakers without impairment often learn new speech motor behaviors quickly and with a high degree of accuracy (e.g., Lisman & Sadagopan, 2013; Sadagopan & Smith, 2013). As a result, it can be difficult to observe potential effects of an experimental manipulation due to ceiling effects on performance. Therefore, it is necessary to use a task with stimuli that are sufficiently difficult to tax the speech motor system (Lisman & Sadagopan, 2013; Sasisekaran, Smith, Sadagopan, & Weber-Fox, 2010).

The nonword stimuli in this study were designed to be complex to avoid ceiling-level performance and to allow us to detect effects of feedback frequency condition on accuracy. However, they were also designed to contain properties of real English words (e.g., lexical stress patterns) to encourage speakers to use native-like speech production. All nonwords were four syllables long, because performance on nonword repetition tasks declines with increased length (Byrd, Valley, Anderson, & Sussman, 2012; Sadagopan & Smith, 2013). Onset consonant clusters were included in each syllable to increase complexity because they have been shown to influence accuracy and kinematic measures during a nonword repetition task, with reduced performance on nonwords that included more consonant clusters (Sasisekaran et al., 2010).

Half of the stimuli included phonotactically illegal (nonnative) clusters in word-initial position (e.g., /ft/). These were included because nonwords that include only allowable English phoneme combinations may not be sufficiently novel

Table 1. Summary of research assessing feedback frequency in speech motor tasks.

Authors (Year)	Feedback Groups (%)	Speech Task	Population	Summary of Results
Adams & Page (2000)	100 vs. 20	Slowed speech production	Adult speakers without impairment	<ul style="list-style-type: none"> • Better scores for 20% at retention
Steinhauer & Grayhack (2000)	100 vs. 50 vs. 0	Vowel nasalization	Adult speakers without impairment	<ul style="list-style-type: none"> • All had improved performance at retention • Better retention and transfer scores for 50% than 100% • 0% with similar performance at retention and transfer at 50%
Adams, Page, & Jog (2002)	100 vs. 20	Slowed speech production	Parkinson's disease	<ul style="list-style-type: none"> • Better retention scores for 20% at retention
Austermann Hula et al. (2008)	100 vs. 60	Articulatory–kinematic treatment approach	Apraxia of speech/aphasia	<ul style="list-style-type: none"> • 60% resulted in enhanced performance for two participants • 100% resulted in enhanced performance for one participant • No clear benefit for either frequency in one participant
Kim et al. (2012)	100 vs. 20	Production of novel (nonnative) phonetic utterances	Adult speakers without impairment	<ul style="list-style-type: none"> • 20% with better scores on intelligibility, naturalness, and precision during retention tests
Maas, Butalla, & Farinella (2012)	100 vs. 60	Articulatory–kinematic treatment approach	Childhood apraxia of speech	<ul style="list-style-type: none"> • Advantage for low-frequency feedback for two children • Small advantage for high-frequency feedback in one child • No clear improvement in either feedback condition for one child

and may be produced as automatically as real English words for people with robust speech motor systems (Lisman & Sadagopan, 2013). It is also possible that nonwords containing only legal English phoneme sequences may be produced by combining existing motor patterns, whereas nonnative phoneme sequences would not have an established motor pattern. Therefore, the production of novel (nonnative) sequences may require more explicit motor learning than phonotactically legal nonwords.

Present Investigation

The goal of this study was to determine if there is an optimal amount of feedback to provide during speech motor learning tasks to facilitate acquisition and retention of speech motor skills in speakers without impairment. Previous studies have typically only compared high (100%) and low (<100%) levels of feedback frequency during speech motor learning tasks (Adams & Page, 2000; Austermann Hula et al., 2008; Kim et al., 2012). This is the first study to compare different levels of low-frequency feedback (e.g., 50% vs. 20%) with each other and with high-frequency feedback and no-feedback conditions using a between-participants paradigm. These feedback levels were chosen because they have previously appeared in the literature (e.g., Adams & Page, 2000; Austermann Hula et al., 2008; Kim et al., 2012; Steinhauer & Grayhack, 2000), and they reflect a range of possible values. Because performance during acquisition can be an unreliable predictor of performance during retention, and true motor learning occurs only when practice leads to relatively permanent changes in the capability for movement

(Schmidt & Lee, 2005), performance during acquisition was compared with performance during two retention tests: short-term retention (30 min after practice) and long-term retention (2 days after practice). We compared whole nonword accuracy, phoneme accuracy, and two duration measures.

This study also explicitly compared motor learning during production of complex nonwords containing all native sequences to production of complex nonwords with nonnative sequences. Nonwords that include only allowable English phoneme combinations may be produced by combining existing motor patterns, whereas there may not be established motor patterns to support the production of nonnative phoneme sequences. Thus, accurate production of nonnative phoneme sequences may require the learning of a new motor skill, potentially providing a clearer test of the conditions that support speech motor learning.

The nonnative clusters included in this study consisted of /f + stop/ clusters (e.g., /fk/, /ft/) and /ʃ + stop/ clusters (e.g., /ʃk/, /ʃt/). These nonnative clusters were chosen because they were most likely to induce errors that could be easily perceived by a listener (i.e., the experimenter) for whom the clusters are nonnative (Davidson, 2006a, 2006b). This consideration was critical for providing accurate online feedback to participants. Using these clusters also allowed us to investigate differences in learning on the basis of similarity to native clusters. We predicted that participants may have a greater degree of familiarity and practice with /f/-initial clusters as compared with /f/-initial clusters, because English includes loanwords with /ʃ+stop/ clusters (e.g., *shtick*). Therefore, /f + stop/ clusters in the initial word position may be more novel and require more explicit learning.

Given the nature of the motor learning task, participants in all feedback groups were expected to demonstrate improved phoneme and whole nonword accuracy and shorter nonword durations at both short-term and long-term retention sessions relative to practice. However, on the basis of results from prior studies, we predicted that participants who received reduced feedback frequencies (i.e., 50% or 20%) or no feedback (0%) would demonstrate better performance on retention tests as compared with participants who received 100% feedback during the learning task. In addition, it was predicted that differences in performance between groups may not emerge until the long-term retention time point, on the basis of similar findings from studies in limb motor learning (e.g., Winstein & Schmidt, 1990). We did not have specific predictions about the level of low-frequency feedback that would provide the most benefit, because these conditions have not been directly compared before within the context of a complex nonword production task.

Method

Participants

Thirty-two healthy young adults (20 women; 12 men; mean age = 26;0 years; months, range = 19;8–37;8) were recruited from the New York University campus. Two additional participants were excluded; one did not complete the study, and the other did not meet the prepractice criterion (see below). All participants were native English speakers. Participants were excluded if they: were simultaneous bilinguals; had prior exposure to languages that included any of the nonnative clusters used in this study (e.g., Greek; see section on nonword stimuli); or if they reported a history of neurological, speech, or language problems and/or previous speech-language therapy. Two participants were sequentially bilingual in languages that do not allow the nonnative clusters described herein (L2 = Malay or Mandarin). Participants were also excluded if they had previous experience with phonetics or linguistics. All participants had unremarkable oral motor examinations and passed a pure-tone hearing screening at 500 Hz, 1000 Hz, 2000 Hz, and 4000 Hz with a criterion of 25 dB.

Participants were randomly assigned to one of four groups. The four groups differed in the amount of KR feedback they received during the practice session (see procedure section for further description): (a) eight participants received no feedback during practice (0%), (b) seven participants received feedback on every fifth production during practice (20%), (c) nine participants received feedback on every other production during practice (50%), and (d) eight participants received feedback on every production during practice (100%).

Materials

Nonword Stimuli

Eight novel nonwords were designed for this study (see Table 2). Each nonword contained four syllables with consonant clusters in the onset position of each syllable.

Four nonwords had a phonotactically legal cluster (i.e., native) in the initial nonword position (e.g., /pl/, /br/), and four nonwords had a phonotactically illegal cluster (i.e., nonnative) in the initial nonword position (e.g., /ft/, /fk/). Half of the nonwords had a total of 18 phonemes; the other half had 19 phonemes (two- vs. three-phoneme cluster in the onset of the third syllable). All syllables had a consonant in the coda position.

The sum of the biphone probabilities, calculated for each nonword as a global measure of complexity (Vitevitch & Luce, 2004), was roughly equal for all nonwords within each category (native vs. nonnative initial cluster). For nonwords with nonnative consonant clusters, the biphone probability of the first pair of phonemes was zero because that cluster does not occur in the initial position of any English word. The nonword stimuli were produced by a female native English speaker who was a trained linguist. Stimuli were recorded, amplitude-normalized to 65 dB, and digitized using Praat software (Boersma & Weenink, 2014). Primary stress was placed on the first syllable, and secondary stress was placed on the third syllable.

Cognitive–Linguistic Testing

Participants completed a battery of tests to assess speech, language, and cognitive–linguistic skills. Nonword repetition ability has been correlated with measures of working memory and vocabulary size in both children and adults (Archibald & Gathercole, 2006; Edwards, Beckman, & Munson, 2004; Gathercole, 2006; Gathercole, Willis, Baddeley, & Emslie, 1994; Gupta, 2003). Therefore, each participant completed the Peabody Picture Vocabulary Test (PPVT-4; Form A; Dunn & Dunn, 2007) as a measure of receptive vocabulary to determine whether this correlated with individual differences in performance, as well as a digit span task (forward and backward) and the Corsi block span task (forward and backward) to assess verbal and nonverbal visuospatial working memory, respectively. Both forward and backward spans were collected because the forward span tests immediate recall and short-term memory, whereas the backward span is thought to target working memory (Burkholder-Juhász, Levi, Dillon, & Pisoni, 2007). The participants' articulation and phonological skills, phonological working memory, and perception of nonword sequences were also assessed using alternate motion rates and the Nonword Repetition Test (Dollaghan & Campbell, 1998; Lisman & Sadagopan, 2013), because underlying speech motor planning and execution abilities may influence performance on the nonword production task in this study (Sasisekaran et al., 2010). All participants performed within normal limits for their age on each of the tests listed, and there were no significant differences between groups. A summary of group performance on each of these tasks is provided in Supplemental Material S13.

Procedure

Data were collected over two sessions separated by 2 days. On the first day, after completing a language

Table 2. Nonword stimuli.

International Phonetic Alphabet	Orthography	Phonotactically Legal Onset	Nonword Duration (in ms)
plut.strɪf.brog.splæm	plootstrifbroagsplam	Yes	1,540
klop.sprɪn.truf.skwan	kloapsprintroofskwan	Yes	1,670
brit.sprɪl.strud.skɾɛm	breetssprilstroodskɾɛm	Yes	1,710
grul.stræb.sprot.splɛk	grootstrabsproatspleck	Yes	1,570
fkut.strɪg.drol.sklen	fkootstrigdroalsklen	No	1,650
ftik.strɛt.sprom.splad	fteekstretsproamsplad	No	1,690
fkop.strɛm.sprɪt.skwb	shkoapstremspreetskwib	No	1,850
ftud.strɪm.preb.skɾɛn	shtoodstrimpraybskɾɛn	No	1,800

background questionnaire, oral motor examination, and hearing screening, participants were familiarized with the accurate production of each nonword before practice (prepractice). Immediately following prepractice, participants completed the experimental practice session and received feedback from the experimenter according to a randomly assigned feedback frequency schedule. Two retention assessments were administered after practice: the first occurred approximately 30 min later (short-term retention), and the second occurred 2 days later (long-term retention). After the long-term retention, on the second session day, participants completed the Nonword Repetition Test (Dollaghan & Campbell, 1998), and diadochokinetic rates were collected. Each part of the procedure is described further in the following sections.

Prepractice

During the prepractice training, participants were instructed to repeat the target nonword after being presented with simultaneous auditory and visual (orthographic) models. The orthographic model was used to ensure accurate perception of the novel nonwords, in particular, the nonnative sequences, before practice. Stimuli were presented in a blocked fashion, and knowledge of results feedback (e.g., correct vs. incorrect) was provided on 100% of the trials for all participants. If the participant's production was inaccurate, specific feedback about the errors was provided (i.e., knowledge of performance feedback). Training ended when the participant accurately produced the nonword stimuli two times consecutively. If the participant was unable to meet this criterion after eight attempts, it was documented by the experimenter, and they moved on to the next nonword. All participants met the criterion on at least four of the eight stimuli nonwords. As mentioned earlier, one participant was excluded because she was not able to reach this criterion level. The criteria were met more often for nonwords having only native clusters than on nonwords having nonnative clusters. However, this pattern was consistent across all feedback groups.

Practice

Participants produced the eight nonword stimuli 10 times each in a randomized order during the practice session. The participants were instructed to repeat the nonword after hearing an auditory model and seeing an orthographic representation of the nonword simultaneously.

The orthographic model was provided to reduce short-term/working memory demands and to ensure proper speech perception of target nonwords, especially those with initial nonnative clusters. Participants were also instructed to try to be as fluent as possible without stopping during their production and to avoid repeating part or all of the nonword if they made an error. Stimuli were presented using E-Prime 2.0 Professional Software (Schneider, Eschman, & Zuccolotto, 2013). Auditory stimuli were presented over Altec Lansing speakers at a comfortable loudness level while a simultaneous orthographic model of the stimuli was presented in white writing on a black screen on a Dell desktop computer. Orthographic stimuli were presented for 3.5 s with an interstimulus interval of 5 s. The practice session lasted approximately 10 min. All productions were audio-recorded with a Shure SM10 head-mounted microphone connected to a Marantz PMD-660 digital recorder.

During the practice session, participants received feedback from the experimenter according to a randomly assigned feedback frequency schedule (100%, 50%, 20%, or 0%). For example, participants in the 100% group received feedback after every production, whereas participants in the 20% group only received feedback after every fifth production. Knowledge of results (KR) feedback was provided at three levels (e.g., correct, close, incorrect) approximately 2–3 s after the participants' response. KR feedback of "close" was provided if only one phoneme in the nonword was produced incorrectly or if all the phonemes were accurate but the nonword was produced disfluently. Knowledge of performance feedback (i.e., detailed information about how the nonword was produced) was not provided during the practice session, as previous work in the non-speech domain has shown that this type of feedback may interfere with retention of newly learned motor skills (e.g., Schmidt & Lee, 2005; for a review, see Maas et al., 2008). Participants completed cognitive–linguistic testing immediately after the practice session.

Retention Tests

During both short-term and long-term retention tests, participants repeated the same experimental protocol as the practice session (10 blocks of stimuli, with each block containing all eight stimuli in a random order); however, no feedback was provided. Participants were not refamiliarized with the stimuli before the long-term retention test.

Data Analysis

Transcription Analysis

Participants' productions were perceptually judged for accuracy and fluency online by the experimenter during data collection. Feedback was provided during the practice session on the basis of the online perceptual judgments. The utterances were scored again at a later time offline by the same experimenter. During the offline analysis, each nonword was transcribed from the audio recording, and accuracy was obtained by computing the percentage of phonemes correct and accuracy of the whole nonword (e.g., correct or incorrect). Online and offline judgments of whole nonword accuracy were in agreement 88.2% of the time.

Offline transcription was completed by the first author, who is a trained speech-language pathologist and was blinded to feedback group and session. A second listener, a research assistant trained in phonetic transcription and blinded to feedback group and session, transcribed all productions from nine out of 32 (28%) participants. For phoneme accuracy, interrater agreement was computed for each participant and ranged from 96.3% to 98.7%, with an average of 97.8%. For whole nonword accuracy, interrater agreement ranged from 78.3% to 90.4%, with an average of 84.6%.

Acoustic Measures

Duration measures were collected only from perceptually accurate and fluent productions. Durations of the target nonwords and clusters were measured on the basis of waveforms and wide-band spectrograms in Praat (Boersma & Weenink, 2014).

The onset of nonwords with an initial stop was defined by the burst release. In nonwords with an initial fricative, the start of the nonword was determined to be at the first point of high-frequency spectral energy on the spectrogram and the onset of aperiodic noise in the waveform. For nonwords that contained a nasal stop in the coda position, the end of the nonword was defined by the disappearance of the nasal formant. The offset of the nonword was defined by the release burst in nonwords that ended with stops. Some tokens with no release burst were included because final stops are not always released in the speech of adults without impairment. In these cases, an approximation of the end of the stop consonant was made on the basis of the waveform and spectrogram and by listening to the audio recording.

Duration of the initial cluster was measured only for the nonwords with nonnative clusters. We analyzed initial cluster durations only for nonwords with nonnative clusters because we believed that the participants were less likely to have existing motor patterns for the phonotactically illegal clusters. This portion of the nonnative nonword was predicted to have the potential for the most changes in speech motor learning. All nonnative clusters were composed of an initial fricative followed by a stop consonant (e.g., /fk/, /ft/). The onset of the cluster was determined to be at the first point of high-frequency spectral energy on

the spectrogram and the onset of aperiodic noise in the waveform. The offset of the cluster was defined by the stop release burst.

Measurement of whole nonword duration for all tokens was completed by a research assistant trained in acoustic analysis and blinded to feedback group and session. The first author measured whole nonword duration for 22% (1,158/5,294) of accurate nonwords productions. The mean difference between coders was 2 ms, with a standard deviation of 55 ms. There was a significant strong correlation ($r = .92$, $p < .001$) between the first and second coders' measurements of whole nonword duration.

Measurement of cluster duration was completed by two research assistants who were trained in acoustic analysis and blinded to feedback group and session. One research assistant measured cluster duration for all accurate initial nonnative clusters, whereas the other took measurements for 28% (657/2,373) of the tokens. The mean difference between coders was 11 ms, with a standard deviation of 24 ms. There was a significant strong correlation ($r = .87$, $p < .001$) between the first and second coders' measurements of cluster duration.

Statistical Analysis

Statistical analysis was conducted using the *lmer* and *lmerTest* packages (Baayen, 2008; Baayen, Davidson, & Bates, 2008; Bates, Maechler, Bolker, & Walker, 2010; Kuznetsova, Brockhoff, & Christensen, 2013) in R (<http://www.r-project.org/>). There were four dependent measures analyzed in this study: (a) Whole nonword accuracy was calculated for each nonword and used as a dependent measure in a logistic mixed-effects model; (b) phoneme accuracy was calculated for each nonword, and a logit transformation was applied before it was used as a dependent measure in a linear mixed-effects model to make it an unbounded continuous variable. Proportions that are bound cannot be used because the model may predict values that are less than zero or greater than one; (c) whole nonword duration was calculated only for productions judged to be perceptually accurate and was used as a continuous dependent measure in a linear mixed-effects model; (d) initial cluster duration was calculated only for productions of nonnative nonwords judged to be perceptually accurate and was used as a continuous dependent measure in a linear mixed-effects model.

For models analyzing whole nonword accuracy and phoneme accuracy, fixed effects included Feedback group (0%, 20%, 50%, 100%), Session (practice, short-term retention, long-term retention), and Nateness of nonword-initial phoneme cluster (native, nonnative). To allow for comparisons between all groups, each model was run with reference levels for Feedback set at 100%, 50%, 20%, and 0%. Practice was used as the reference level for the Session fixed effect in all models so that comparisons could be made between short-term retention and practice or between long-term retention and practice for each dependent variable. Nateness of the initial phoneme cluster was included in the models because we hypothesized that participants may perform differently on nonwords with native and nonnative

clusters that varied in novelty and complexity. Native was the reference group for the Nateness fixed effect in all models. In addition to main effects, three two-way interactions (Feedback $\times$ Session; Feedback $\times$ Nateness; Session $\times$ Nateness) and one three-way interaction (Feedback $\times$ Session $\times$ Nateness) were included in the full models for the accuracy measures.

For models analyzing the duration measures, fixed effects included Feedback group (0%, 20%, 50%, 100%), Session (practice, short-term retention, long-term retention), and the Feedback $\times$ Session interaction. If a participant did not produce three accurate productions in practice, productions of that nonword were removed from all sessions and not included in the analysis. Nateness was not included as a fixed factor because there were inherent differences in the phonemes included in the clusters (i.e., native clusters consisted of stop + liquid combinations, whereas nonnative clusters consisted of fricative + stop combinations). In addition, cluster duration was not measured on nonwords with only native cluster. To allow for comparisons between all groups, each model was run with reference levels for Feedback set at 100%, 50%, 20%, and 0%. Practice was used as the reference level for the Session fixed effect in all models so that comparisons could be made between short-term retention and practice or between long-term retention and practice for each dependent variable.

For all models, log-likelihood comparisons were used to determine the interactions that significantly contributed to the final models chosen for each dependent variable at short- and long-retention time periods (Baayen, 2008; Baayen et al., 2008). Random intercepts for participant and nonword were also included in all models to account for nonword- and speaker-specific variation. Full model outputs are provided in Supplemental Materials S1, S2, S3, S4, S5, and S6.

The final logistic mixed-effects regression model for whole nonword accuracy included Feedback, Session, and Nateness as fixed effects and Participant and Nonword as random effects. All two-way interactions among Feedback, Session, and Nateness and the three-way interaction between all of the fixed effects were excluded from the final model because log-likelihood comparisons revealed that these interactions did not significantly contribute to the model fit (all $p > .05$).

The final linear mixed-effects regression model for phoneme accuracy included Feedback, Session, and Nateness as fixed effects and Participant and Nonword as random effects. Two-way interactions between Feedback and Session, Feedback and Nateness, and Session and Nateness were also included in the model. The three-way interaction among Feedback, Nateness, and Session was not included because log-likelihood comparisons revealed it did not significantly contribute to the model ($p > .05$).

For whole nonword duration, two separate linear mixed-effects regression models were fit: one for nonwords with only native clusters, and one for nonwords with an initial nonnative cluster. Both models included Feedback and Session as fixed effects and Participant and Nonword

as random effects. The two-way interaction between Feedback and Session was also included in both models.

Two linear mixed-effects regression models were fit for initial cluster duration: one for nonwords with /f/-initial nonnative clusters, and one for nonwords with /f/-initial nonnative clusters. Both models included Feedback and Session as fixed effects and Participant and Nonword as random effects. The two-way interaction between Feedback and Session was also included in both models.

Results

Transcription Accuracy

Whole Nonword Accuracy

Whole nonword accuracy data are presented by Session, Feedback group, and Nateness of the initial cluster in Table 3.

Feedback condition was not a significant predictor of whole nonword accuracy (all $p > .05$). However, nateness was a significant predictor of nonword accuracy ($\beta = -.737$, $z = -2.99$, $p = .003$, $d = 3.06$). Nonwords with only native clusters were produced more accurately ($M = 76.9\%$, $SD = 4.2\%$) than nonwords that contained nonnative clusters in the word-initial position ($M = 63.1\%$, $SD = 4.8\%$).

Session was also a significant predictor of whole nonword accuracy when comparing both short-term retention ($\beta = .325$, $z = 5.03$, $p < .001$, $d = -1.84$) and long-term retention ($\beta = .620$, $z = 9.31$, $p < .001$, $d = -3.33$) to practice. Whole nonword accuracy was greater at the short-term retention time period ($M = 70.6\%$, $SD = 3.0\%$) as compared with practice ($M = 64.5\%$, $SD = 3.6\%$). Also, nonword accuracy was greater at long-term retention ($M = 75.7\%$, $SD = 3.1\%$) than during practice ($M = 64.5\%$, $SD = 3.6\%$).

Phoneme Accuracy

Phoneme accuracy data are presented by Feedback group and Session in Figure 1 and in Supplemental Material S7. Phoneme accuracy data are presented by Nateness of the initial cluster and Session in Supplemental Material S8.

Feedback condition and Nateness were not significant predictors (all $p \geq .05$) of phoneme accuracy. However, there was a significant effect of Session, $F(2, 7604) = 40.97$, $p < .001$. Although participants were highly accurate across all sessions, they were more accurate at both short-term retention ($M = 97.8\%$, $SD = 5.5\%$, $d = -0.10$) and long-term retention ($M = 98.4\%$, $SD = 4.7\%$, $d = -0.21$) sessions as compared with practice ($M = 97.2\%$, $SD = 6.8\%$).

The interaction of Session and Feedback contributed significantly to the model's fit, $F(6, 7604) = 2.74$, $p = .012$ (see Figure 1). When comparing performance between practice and short-term retention, participants who received no feedback during practice had a larger improvement in phoneme accuracy across sessions than participants who received feedback on 100% of trials during practice ($\beta = -.093$, $t = -1.97$, $p = .048$; see Figure 1A). No significant

Table 3. Mean whole nonword accuracy by Feedback group, Session, and Nativeness.

Feedback Group	Whole Nonword Accuracy, Percent Nonwords Correct (<i>SD</i>)					
	Practice		Short-Term Retention		Long-Term Retention	
	Native	Nonnative	Native	Nonnative	Native	Nonnative
0	67.8 (20.2)	55.3 (17.3)	73.6 (19.4)	63.4 (16.9)	80.9 (17.2)	73.1 (15.0)
20	78.6 (15.3)	60.0 (23.4)	78.2 (18.4)	71.1 (14.9)	83.2 (8.1)	72.5 (12.2)
50	71.7 (13.5)	52.2 (19.3)	77.5 (18.4)	58.6 (17.1)	76.9 (17.1)	64.7 (19.9)
100	74.4 (19.6)	55.9 (18.5)	77.2 (20.6)	64.7 (24.3)	84.7 (13.1)	68.8 (23.6)
Total	72.9 (16.9)	55.6 (18.8)	76.6 (18.4)	64.1 (18.3)	81.3 (14.8)	69.5 (17.9)

differences were found when comparing performance of other feedback conditions across practice and short-term retention (all $p > .05$). When comparing performance between practice and long-term retention, participants who received no feedback during practice had a larger improvement in phoneme accuracy across sessions than participants who received feedback on 20% of trials ($\beta = -.133$, $t = -2.75$, $p = .006$) or on 50% of trials ($\beta = -.151$, $t = -3.32$, $p = .001$) during practice (see Figure 1B). Participants who received no feedback during practice also had a larger improvement in phoneme accuracy between practice and long-term retention sessions than participants in the 100% feedback group, but this difference did not reach statistical significance ($\beta = -.08$, $t = -1.71$, $p = .087$).

Session also significantly interacted with Nativeness, $F(2, 7604) = 5.47$, $p = .004$, indicating that the difference in accuracy between practice and short-term retention and the difference in accuracy between practice and long-term retention were greater for nonwords with nonnative clusters than those with native clusters. Although all nonword productions were highly accurate, this interaction is most likely due to a ceiling effect. Nonwords with native clusters did not have as much room to improve as nonnative clusters.

Although the main effects Feedback and Nativeness were not significant, the interaction between these variables was a significant predictor of phoneme accuracy, $F(3, 7604) = 9.81$, $p < .001$. Participants who received feedback on 50% of their productions during the practice

session performed significantly worse on nonwords that included nonnative clusters, across sessions, than all other groups. Again, it is unclear if this finding reflects a true underlying interaction, because performance on the nonwords with native clusters was close to ceiling during practice.

Acoustic Measures

Whole Nonword Duration

Whole nonword duration data are presented in Figures 2 (native nonwords) and 3 (nonnative nonwords). These data are also presented by Feedback group and Session in Supplemental Material S9 and by Nativeness of the initial cluster and Session in Supplemental Material S10.

For the nonwords with only native clusters, Feedback condition was not a significant predictor ($p > .05$) of nonword duration. There was a significant effect of Session, $F(2, 2784.3) = 76.60$, $p < .001$. There was no difference in duration of native nonwords between short-term retention ($M = 1,553$ ms, $SD = 128$ ms, $d = 0.04$) and practice ($M = 1,558$ ms, $SD = 130$ ms) sessions. However, participants produced nonwords that were shorter in the long-term retention session ($M = 1,516$ ms, $SD = 120$ ms, $d = 0.34$).

The effect of Session significantly interacted with Feedback group, $F(6, 2784.3) = 8.45$, $p < .001$; see Figure 2). There was a larger difference in duration between practice and short-term retention sessions for the groups that received feedback on 100% ($b = 25.26$, $t = 2.27$, $p = .024$),

Figure 1. Mean phoneme accuracy by session and feedback group (across nativeness). (A) Comparison between practice and short-term retention sessions. (B) Comparison between practice and long-term retention sessions.

Figure 2. Mean nonword duration by session and feedback group for nonwords with an initial native cluster. (A) Comparison between practice and short-term retention sessions. (B) Comparison between practice and long-term retention sessions.

50% ($b = 45.45$, $t = 4.14$, $p < .001$), or 20% ($b = 37.33$, $t = 3.29$, $p = .001$) of trials during practice as compared with the group that received no feedback during practice (0%; see Figure 2A). Between practice and long-term retention, there was a larger difference in duration for the group that received 50% feedback during practice compared with all other groups (0%: $b = 67.62$, $t = 6.25$, $p < .001$; 20%: $b = 48.08$, $t = 4.39$, $p < .001$; 100%: $b = 47.47$, $t = 4.43$, $p < .001$; see Figure 2B).

For the nonwords with an initial nonnative cluster, Feedback condition was not a significant predictor ($p \geq .05$) of nonword duration. However, there was a significant effect of Session, $F(2, 2070.5) = 131.29$, $p < .001$. Overall, participants produced shorter nonwords in both short-term retention ($M = 1,693$ ms, $SD = 133$ ms, $d = 0.17$) and long-term retention ($M = 1,656$ ms, $SD = 139$ ms, $d = 0.43$) sessions as compared with practice ($M = 1,715$ ms, $SD = 133$ ms).

The effect of session significantly interacted with Feedback group, $F(6, 2070.41) = 2.79$, $p = .011$; see Figure 3). There were no significant differences in duration among feedback groups between practice and short-term retention sessions (all $p \geq .05$; see Figure 3A). Between practice and long-term retention sessions (see Figure 3B), there was a larger difference in duration for the group that received 50% feedback during practice compared with

all other groups (0%: $b = 45.53$, $t = 3.18$, $p = .002$; 20%: $b = 39.03$, $t = 2.84$, $p = .005$; 100%: $b = 37.85$, $t = 2.91$, $p = .004$).

Initial Cluster Duration

Initial cluster duration data are presented by Feedback group and Session for nonwords with /f/-initial nonnative clusters in Supplemental Material S11 and for nonwords with /t/-initial nonnative clusters in Supplemental Material S12.

For the nonwords with /f/-initial nonnative clusters, Feedback condition was not a significant predictor ($p > .05$) of cluster duration. However, there was a significant effect of Session, $F(2, 1256.7) = 26.0$, $p < .001$. Although participants produced nonwords that had similar lengths between practice ($M = 243$ ms, $SD = 43$ ms) and short-term retention ($M = 247$ ms, $SD = 44$ ms) sessions, productions at the long-term retention session were shorter ($M = 234$ ms, $SD = 38$ ms, $d = -0.30$). The effect of session significantly interacted with Feedback group, $F(6, 1256.7) = 3.51$, $p = .001$; see Figure 4). Between the practice and short-term retention sessions, the group that received feedback on 50% of trials in practice had a significant increase in cluster duration relative to the groups that received feedback on 100% ($b = -13.29$, $t = -2.33$, $p = .020$) and 0% ($b = -14.45$, $t = -2.46$, $p = .014$) of trials in practice. The group that received feedback on 20% of trials in practice also had a

Figure 3. Mean nonword duration by session and feedback group for nonwords with an initial nonnative cluster. (A) Comparison between practice and short-term retention sessions. (B) Comparison between practice and long-term retention sessions.

Figure 4. Mean cluster duration by session and feedback group. (A) /f/-initial clusters. (B) /f/-initial clusters.

significant increase in cluster duration relative to the groups that received feedback on 0% ($b = -12.43$, $t = -2.10$, $p = .036$) of trials in practice. Between practice and long-term retention sessions, the group that received feedback on 20% of trials during practice had a larger decrease in cluster duration than the group that received 100% feedback ($b = 11.27$, $t = 1.96$, $p = .050$). There were no other significant differences in duration between feedback groups across sessions (all $p > .05$).

For the nonwords with /f/-initial nonnative clusters, Feedback condition and Session were not significant predictors ($p > .05$) of cluster duration. The interaction Feedback condition $\times$ Session was also not significant ($p > .05$; see Figure 4). Table 4 is provided as an overview of all significant findings across dependent measures for the main effect of Session and the interaction Feedback $\times$ Session.

Discussion

This study examined how feedback frequency affected the acquisition and retention of newly learned speech motor skills in speakers without impairment. Perceptual and acoustic measures of complex nonword productions were

compared among groups who received two different levels of low-frequency feedback (e.g., 50% vs. 20%), high-frequency feedback (100%), or no feedback (0%) to determine the amount that was most beneficial for speech motor learning. Performance was compared across sessions (practice vs. short-term retention; practice vs. long-term retention) and stimulus properties (i.e., nativeness of word-initial consonant cluster).

As predicted, participants in all groups demonstrated improved phoneme and whole nonword accuracy at short-term and long-term retention sessions and shorter nonword productions at the long-term retention session. However, participants who received no feedback during practice improved the most in phoneme accuracy between practice and long-term retention sessions. Participants who received feedback on 50% of their productions during practice demonstrated the largest reduction in nonword duration between practice and long-term retention sessions. In sum, these findings suggest that participants continue to refine their speech motor skills for perceptually accurate productions. In the remainder of this section, we discuss the implications of our findings with respect to accuracy, acoustic changes, and nonnative cluster production.

Table 4. Summary of significant results.

	Session	Feedback $\times$ Session	
		STR vs. P	LTR vs. P
Accuracy ^a			
Whole nonword	LTR, STR > P	—	—
Phoneme	LTR, STR > P	0 > 100	0 > 20, 50
Nonword duration ^b			
Native nonwords	LTR < P	100, 50, 20 < 0	50 < 0, 20, 100
Nonnative nonwords	LTR, STR < P	—	50 < 0, 20, 100
Initial cluster duration ^b			
/f/-initial clusters	LTR < P	100, 0 < 50 0 < 20	20 < 100
/f/-initial clusters	ns	ns	ns

Note. STR = short-term retention; LTR = long-term retention; P = practice; ns = not significant.

^aA greater-than sign indicates more accurate performance. ^bA less-than sign indicates shorter (faster) durations.

Impact of Feedback Frequency on Speech Motor Learning

On the basis of results from prior studies of feedback frequency and speech motor learning (e.g., Adams & Page, 2000; Kim et al., 2012), it was predicted that participants who received reduced feedback frequencies (i.e., 50% or 20%) or no feedback (0%) would demonstrate better performance on retention tests as compared with participants who received 100% feedback during the learning task. However, results from this experiment revealed that all participants, regardless of the frequency of feedback received during practice, demonstrated similar degrees of improvement at short-term and long-term retention tests when measured by whole nonword accuracy. Small differences in performance were noted between feedback groups when analyzed by phoneme accuracy, but the results must be interpreted cautiously in the presence of a possible ceiling effect.

In terms of the acoustic measurements, all participants produced nonwords that were shorter in the long-term retention session compared with practice. However, the group that received feedback on 50% of their productions in practice had the largest reduction in duration between baseline and long-term retention for nonwords with native and non-native clusters. Accurate productions with shorter durations were thought to represent a more stable motor pattern and possibly more efficient control of speech movements (Lisman & Sadagopan, 2013).

There are several factors that may explain differences between the findings of the present work and those of past research. First, the null results may be related to the specific task used in this study. It has been proposed that there is an optimal amount of frequency to benefit motor learning and that this amount may vary for different tasks (Sparrow, 1995). Feedback frequency has never been studied using a complex nonword production task such as the one in this study and, although unlikely, it remains possible that none of the frequencies tested represents the optimal amount of feedback for this task.

Second, the relationship between feedback frequency and the other practice conditions (particularly practice amount and target complexity) may have affected the results of this study. Motor learning studies have demonstrated that large amounts of practice with complex targets are beneficial for retention of newly learned motor skills (Maas et al., 2008). However, it has been shown that feedback frequency interacts with the number of practice trials and target complexity (Kim et al., 2012; Wulf et al., 1998). Kim et al. (2012) demonstrated that speech motor learning with reduced feedback frequency was enhanced when the number of practice trials increased. In addition, Wulf et al. (1998) suggested that high-frequency feedback may be beneficial for learning complex motor skills until a certain level of expertise is obtained. For this study, each stimulus was produced 10 times during practice and an additional 10 times in each retention session. This reflected a smaller number of repetitions, particularly during practice, than in some

other studies of speech motor learning (Adams & Page, 2000; Kim et al., 2012; Steinhauer & Grayhack, 2000). It is possible that the large number of repetitions with complex targets may obscure the more subtle effects of feedback manipulations. It is not known how these differences affected the results obtained and reported here, highlighting the need for additional research to determine the optimal parameters and combinations of the practice and feedback variables.

Last, despite our efforts to generate complex nonword stimuli, our participants demonstrated a high degree of phoneme accuracy, close to ceiling level, even during practice. Therefore, it is possible that any effect of feedback frequency on phoneme accuracy would have been impossible to observe in accuracy measures. In addition, perceptual measures of whole nonword accuracy may have been too gross to capture the small changes in motor performance across sessions. Perceptual measures can be used because of their clinical relevance and ecological validity; however, they may not capture more subtle (subperceptual) changes that may occur during speech motor learning (Austermann Hula et al., 2008). By measuring the stability of productions, we may be able to observe smaller changes in performance and provide further information about how speech motor learning is affected by changes in practice and feedback conditions. For example, Lisman and Sadagopan (2013) used kinematic measures to supplement behavioral (accuracy) measures when assessing the impact of internal versus external focus during different speech motor learning tasks. They found differences in movement trajectories and durations between conditions even when accuracy was close to ceiling level. The present study analyzed acoustic duration measurements, instead of kinematic measurements, to supplement perceptual accuracy measures in order to better assess speech motor learning.

The Effect of Nativeness of Stimuli on Speech Motor Performance and Learning

Nonnative phoneme sequences within nonwords were of interest in this study because it was thought that they may provide a novel and complex speech context that would not be produced using established motor patterns. Therefore, production of nonwords that contained nonnative clusters may provide a speech context in which differences in motor learning on the basis of feedback frequency could be assessed. Nonwords with only native phoneme clusters were produced more accurately (phonemes and whole nonwords) than nonwords with an initial nonnative cluster by all participants in each session. However, this difference was statistically significant only for whole nonword accuracy. The difference in phoneme accuracy between nonwords with native clusters and nonwords with an initial nonnative cluster may not have reached statistical significance because performance for all nonwords was close to ceiling level.

The effect of nativeness interacted with the feedback condition for phoneme accuracy. Participants who received

feedback on 50% of their productions during practice performed significantly worse on nonwords that included nonnative clusters, across sessions, than those who received feedback on 100%, 20%, or 0% of their productions during practice. It is unclear if these findings reflect a true underlying interaction because all groups demonstrated near-ceiling performance on phoneme accuracy for nonwords with native clusters.

An interaction between Nateness and Session was also found for phoneme accuracy when comparing performance during practice and both retention sessions. The number of phonemes correct increased more for nonwords with nonnative clusters between practice and either short-term or long-term retention sessions than for nonwords with only native clusters. Once again, the presence of a possible ceiling effect complicates the interpretation of this result.

Cluster duration was only measured in nonwords with nonnative clusters, because that is where we predicted that more explicit learning would be needed. All participants, regardless of the feedback group, produced nonwords with /f+stop/ initial clusters that were shorter in long-term retention compared with practice. However, no changes in duration were observed across sessions for /f+stop/ initial clusters. Participants may have a greater degree of familiarity and practice with /f/-initial clusters as compared with /f/-initial clusters, because English includes loanwords with /f+stop/ clusters (e.g., *shtick*). Therefore, changes in acoustic duration of the /f+stop/ initial clusters may respond to motor practice and feedback more like the phonotactically legal clusters than the /f+stop/ initial clusters. This is consistent with work by Davidson (2006a), which found that English speakers are more accurate in producing nonnative clusters that have more features in common with phonotactically legal clusters. In addition, the feedback provided in this study was based on the accuracy of the whole nonword, not specifically the accuracy of the cluster. It is possible that feedback specific to the target motor behavior (e.g., nonnative cluster production) may be more beneficial for learning more novel or complex motor skills until a certain level of familiarity or expertise is obtained.

Methodological Considerations

The aim of the study reported here was to determine whether there were systematic differences in performance on a motor speech task in speakers without impairment on the basis of the amount of feedback they received. Although there were some significant findings on various measures, there was not clear evidence in favor of any one of the feedback levels. We note that the absence of evidence for a difference is not equal to evidence of the absence of differences, and in this section, we outline some methodological details in this study that may have masked an underlying difference if one exists.

First, we used a between-participants paradigm to explore how feedback frequency affected speech motor learning. This was done in large part because of practical

constraints, as having each individual participate in each of four feedback conditions would have been extremely time-consuming and may have led to other issues of interpretation (see below). Unfortunately, we obtained evidence of interparticipant variability, as there were some group differences obtained during practice despite random group assignment. Although we note that these differences were not statistically significant, it is hard to determine if these smaller numerical differences reflected substantive differences between the individuals in each group that might mask other differences. While it may be ideal for subsequent studies to investigate this question using a within-participant paradigm, this has already been shown to be pragmatically and logistically difficult (e.g., Austermann Hula et al., 2008) because it is hard to equate stimulus complexity across conditions and to control for possible generalization effects.

An additional consideration was the role of feedback in guiding the participants' performance. KR feedback informed participants about the accuracy of their productions during practice, and it was provided because previous literature suggests that it results in better retention of newly learned motor skills (e.g., Schmidt & Lee, 2005). However, it remains possible that participants were able to judge the accuracy of their production by comparing their own production to the auditory model. Therefore, KR feedback may not provide any additional information about the correctness of their production above and beyond what the participants gather independently, so differences in the frequency of such feedback may be uninformative. On this point, we have assumed that the complexity of the nonwords, including some with nonnative consonant clusters, made the comparison with the model a bit more difficult and that the feedback did provide additional information. In addition, participants were instructed to be as fluent and accurate as possible in their productions, but they were not provided with specific instructions regarding the speed of their productions. This may have resulted in variability in the nonword durations because some participants could have focused on increasing the speed of their productions, whereas other participants placed greater emphasis on producing their productions accurately.

Last, it is possible that one of the other parameters related to the structure of practice and feedback obscured any effect we may have found. In order to test the single parameter of feedback frequency, it was necessary for us to hold the other parameters constant. As mentioned above, one concern regarding the structure of practice may be related to the number of productions for target words during practice. In the practice portion of this study, participants produced eight nonword stimuli 10 times each, whereas other studies have had participants produce a single target speech motor skill between 40 and 100 times (Adams & Page, 2000; Kim et al., 2012; Steinhauer & Grayhack, 2000). Although the participants completed 80 productions during the practice session, it is possible that they were not able to create a stable motor pattern for each nonword with this amount of repetitions, and that resulted in

variable performance across groups during the retention sessions.

Taken together, the issues addressed here highlight the large number of methodological considerations in a study designed to identify optimal practice and feedback parameters. We hope that future researchers will benefit from considering the complexity of these issues.

Conclusion

The current study systematically investigated how the amount of feedback provided during a complex nonword production task affects both acquisition and retention of speech motor skills in speakers without impairment. All participants, regardless of how much feedback they received during practice, demonstrated improved phoneme and whole nonword accuracy at short-term and long-term retention sessions. Participants also continued to refine their productions of perceptually accurate nonwords, as noted by a decrease in nonword duration across sessions. However, participants who received no feedback during practice made the most improvement in phoneme accuracy at the long-term retention session as compared with participants in all other feedback groups, whereas participants who received feedback on 50% of their productions in practice had the largest reduction in nonword duration between practice and long-term retention sessions. Reduction in cluster duration was observed across sessions in nonwords with /f+stop/ initial clusters but not in nonwords with /t+stop/ initial clusters. Additional research is needed to replicate and expand on these findings because this is the first study to report changes in acoustic duration measurements during a speech motor learning task, and no consistent measurable differences were found between feedback conditions.

Acknowledgments

The research and preparation of this article was supported by grants from the New York University–University Research Challenge Fund and National Institutes of Health Grant K01DC014298 to Adam Buchwald. The authors would like to thank Maria Grigos, Tara McAllister Byun, and Susannah Levi for their input on the study design and analyses; Kelly Karpus, Holly Jane Wilde Calhoun, Jonathon Boyd, and Sarah Kastner-Ziemann for help with data analysis; and Stacey Rimikis for help with statistical analyses.

References

- Adams, S. G., & Page, A. D. (2000). Effects of selected practice and feedback variables on speech motor learning. *Journal of Medical Speech-Language Pathology*, 8(4), 215–220.
- Adams, S. G., Page, A. D., & Jog, M. (2002). Summary feedback schedules and speech motor learning in Parkinson's disease. *Journal of Medical Speech-Language Pathology*, 10(4), 215–220.
- Archibald, L. M., & Gathercole, S. E. (2006). Nonword repetition: A comparison of tests. *Journal of Speech, Language, and Hearing Research*, 49(5), 970–983.
- Austermann Hula, S., 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, and Hearing Research*, 51, 1088–1113.
- Baayen, R. H. (2008). *Analyzing linguistic data: A practical introduction to statistics*. Cambridge, United Kingdom: Cambridge University Press.
- Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, 59(4), 390–412. <https://doi.org/10.1016/j.jml.2007.12.005>
- 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.
- Bates, D., Maechler, M., Bolker, B., & Walker, S. (2010). Linear mixed-effects models using Eigen and S4 (Version 1.1). Retrieved from <http://lme4.r-forge.r-project.org/>
- 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. (2014). Praat: Doing phonetics by computer (Version 5.4.03). Retrieved from <http://www.praat.org/>
- Burkholder-Juhász, R. A., Levi, S. V., Dillon, C. M., & Pisoni, D. B. (2007). Nonword repetition with spectrally reduced speech: Some developmental and clinical findings from pediatric cochlear implantation. *Journal of Deaf Studies and Deaf Education*, 12(4), 472–485. <https://doi.org/10.1093/deafed/enm031>
- Byrd, C. T., Valley, M., Anderson, J. D., & Sussman, H. (2012). Nonword repetition and phoneme elision in adults who do and do not stutter. *Journal of Fluency Disorders*, 37(3), 188–201. <https://doi.org/10.1016/j.jfludis.2012.03.003>
- Davidson, L. (2006a). Phonology, phonetics, or frequency: Influences on the production of non-native sequences. *Journal of Phonetics*, 34(1), 104–137. <https://doi.org/10.1016/j.wocn.2005.03.004>
- Davidson, L. (2006b). Phonotactics and articulatory coordination interact in phonology: Evidence from nonnative production. *Cognitive Science*, 30(5), 837–862. https://doi.org/10.1207/s15516709cog0000_73
- Dollaghan, C., & Campbell, T. F. (1998). Nonword repetition and child language impairment. *Journal of Speech, Language, and Hearing Research*, 41, 1136–1146.
- Dunn, L. M., & Dunn, D. M. (2007). *Peabody Picture Vocabulary Test—Fourth Edition*. Minneapolis, MN: Pearson Assessments.
- Edwards, J., Beckman, M. E., & Munson, B. (2004). The interaction between vocabulary size and phonotactic probability effects on children's production accuracy and fluency in nonword repetition. *Journal of Speech, Language, and Hearing Research*, 47(2), 421–436.
- Gathercole, S. E. (2006). Nonword repetition and word learning: The nature of the relationship. *Applied Psycholinguistics*, 27, 513–543.
- Gathercole, S. E., Willis, C. S., Baddeley, A. D., & Emslie, H. (1994). The Children's Test of Nonword Repetition: A test of phonological working memory. *Memory*, 2(2), 103–127. <https://doi.org/10.1080/09658219408258940>
- Gupta, P. (2003). Examining the relationship between word learning, nonword repetition, and immediate serial recall in adults. *Quarterly Journal of Experimental Psychology A: Human*

- Experimental Psychology*, 56(7), 1213–1236. <https://doi.org/10.1080/02724980343000071>
- Kim, I., LaPointe, L. L., & Stierwalt, J. A.** (2012). The effect of feedback and practice on acquisition of novel speech behaviors. *American Journal of Speech-Language Pathology*, 21, 89–100. [https://doi.org/10.1044/1058-0360\(2011/09-0082](https://doi.org/10.1044/1058-0360(2011/09-0082)
- Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H.** (2013). lmerTest: Tests for random and fixed effects for linear mixed effect models (lmer objects of lme4 package) (Version 2.0). Retrieved from <http://lmerTest.r-forge.r-project.org/>
- Lee, T. D., White, M. A., & Carnahan, H.** (1990). On the role of knowledge of results in motor learning. *Journal of Motor Behavior*, 22(2), 191–208. <https://doi.org/10.1080/00222895.1990.10735510>
- Lisman, A. L., & Sadagopan, N.** (2013). Focus of attention and speech motor performance. *Journal of Communication Disorders*, 46(3), 281–293. <https://doi.org/10.1016/j.jcomdis.2013.02.002>
- Maas, E., Butalla, C. E., & Farinella, K. A.** (2012). Feedback frequency in treatment of childhood apraxia of speech. *American Journal of Speech-Language Pathology*, 21, 239–257. [https://doi.org/10.1044/1058-0360\(2012/11-0119](https://doi.org/10.1044/1058-0360(2012/11-0119)
- Maas, E., Robin, D. A., Austermann Hula, S. N., Freedman, S. E., Wulf, G., Ballard, K. J., & Schmidt, R. A.** (2008). Principles of motor learning in treatment of motor speech disorders. *American Journal of Speech-Language Pathology*, 17, 277–298.
- Sadagopan, N., & Smith, A.** (2013). Age differences in speech motor performance on a novel speech task. *Journal of Speech, Language, and Hearing Research*, 56, 1552–1566. [https://doi.org/10.1044/1092-4388\(2013/12-0293](https://doi.org/10.1044/1092-4388(2013/12-0293)
- Salmoni, A. W., Schmidt, R. A., & Walter, C. B.** (1984). Knowledge of results and motor learning: A review and critical reappraisal. *Psychological Bulletin*, 95(3), 355–386.
- Sasisekaran, J., Smith, A., Sadagopan, N., & Weber-Fox, C.** (2010). Nonword repetition in children and adults: Effects on movement coordination. *Developmental Science*, 13(3), 521–532.
- Schmidt, R. A.** (1991). Frequent augmented feedback can degrade learning: Evidence and interpretation. In G. E. Stelmach, & J. Requin (Eds.), *Tutorials in motor neuroscience* (pp. 55–75). Dordrecht, The Netherlands: Kluwer Academic Publishers.
- Schmidt, R. A., & Lee, T.** (2005). *Motor control and learning: A behavioral emphasis* (4th ed.). Champaign, IL: Human Kinetics.
- Schneider, W., Eschman, A., & Zuccolotto, A.** (2013). *E-prime user's guide*. Pittsburgh, PA: Psychology Software Tools, Inc.
- Sparrow, W. A.** (1995). Acquisition and retention effects of reduced relative frequency of knowledge of results. *Australian Journal of Psychology*, 47(2), 97–104.
- 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.
- Vitevitch, M. S., & Luce, P. A.** (2004). A web-based interface to calculate phonotactic probability for words and nonwords in English. *Behavioral Research, Methods, Instruments & Computers*, 36(3), 481–487.
- Wambaugh, J. L., Duffy, J. R., McNeil, M. R., Robin, D. A., & Rogers, M. A.** (2006). Treatment guidelines for acquired apraxia of speech: Treatment descriptions and recommendations. *Journal of Medical Speech-Language Pathology*, 14(2), xxxv.
- Wambaugh, J., Nessler, C., Cameron, R., & Mauszycki, S.** (2013). Treatment for acquired apraxia of speech: Examination of treatment intensity and practice schedule. *American Journal of Speech-Language Pathology*, 22, 84–102.
- 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.
- Wulf, G., & Shea, C. H.** (2004). Understanding the role of augmented feedback: The good, the bad, and the ugly. In M. A. Williams, & N. J. Hodges (Eds.), *Skill acquisition in sport: Research theory and practice* (pp. 121–144). London, United Kingdom: Routledge.
- Wulf, G., Shea, C. H., & Matschiner, S.** (1998). Frequent feedback enhances complex motor skill learning. *Journal of Motor Behavior*, 30(2), 180–192. <https://doi.org/10.1080/00222899809601335>