Speaker compensation for local perturbation of fricative acoustic feedback^{a)}

Elizabeth D. Casserly^{b)}

Department of Psychological & Brain Sciences, Speech Research Laboratory, Indiana University, 1101 East 10th Street, Bloomington, Indiana 47405

(Received 17 June 2010; revised 14 January 2011; accepted 14 January 2011)

Feedback perturbation studies of speech acoustics have revealed a great deal about how speakers monitor and control their productions of segmental (e.g., formant frequencies) and non-segmental (e.g., pitch) linguistic elements. The majority of previous work, however, overlooks the role of acoustic feedback in consonant production and makes use of acoustic manipulations that effect either entire utterances or the entire acoustic signal, rather than more temporally and phonetically restricted alterations. This study, therefore, seeks to expand the feedback perturbation literature by examining perturbation of consonant acoustics that is applied in a time-restricted and phonetically specific manner. The spectral center of the alveopalatal fricative [ʃ] produced in vowel-fricative-vowel nonwords was incrementally raised until it reached the potential for [s]-like frequencies, but the characteristics of high-frequency energy outside the target fricative remained unaltered. An “offline,” more widely accessible signal processing method was developed to perform this manipulation. The local feedback perturbation resulted in changes to speakers’ fricative production that were more variable, idiosyncratic, and restricted than the compensation seen in more global acoustic manipulations reported in the literature. Implications and interpretations of the results, as well as future directions for research based on the findings, are discussed. © 2011 Acoustical Society of America.

[DOI: 10.1121/1.3552883]

PACS number(s): 43.70.Mn, 43.71.Es, 43.70.Bk [AL]

Pages: 2181–2190

I. INTRODUCTION

Research examining speakers’ responses to altered linguistic feedback has proven to be a rich field of inquiry. Speech elements such as the configuration of the vocal tract (Fowler and Turvey, 1980; Baum and McFarland, 1997), the relative position of articulators like the jaw, lips, and tongue tip (Kelso *et al.*, 1984), and the acoustics available to speakers (Elman, 1981) have all been altered both unexpectedly and in a predictable pattern in order to observe the strategies and temporal structure of language users’ responses. The role of acoustic feedback during speech production has been a particular focus in the last twenty years, investigating how speakers monitor their speech acoustics in real-time, how that feedback is used in articulatory control, and whether linguistic contrastivity plays a role in the feedback correction loop (Elman, 1981; Burnett *et al.*, 1997; Burnett *et al.*, 1998; Houde and Jordan, 1998; Jones and Munhall, 2000; Donath *et al.*, 2002; Houde and Jordan, 2002; Jones and Munhall, 2002; Xu *et al.*, 2004; Sivasankar *et al.*, 2005; Purcell and Munhall, 2006b; Chen *et al.*, 2007; Villacorta *et al.*, 2007; Munall *et al.*, 2009; Shiller *et al.*, 2009).

These studies have made great strides in increasing understanding of speakers’ compensation behavior under conditions of altered acoustic feedback, but two issues have been

largely overlooked. First, most of these experiments use either fundamental frequency (f0) or vowel formants as targets, leaving the role of acoustic feedback in consonant articulation in question, and second, perturbations to targeted speech acoustics are typically temporally or phonetically global, applying to entire utterances and entire acoustic signals rather than specific linguistic units and their phonetic signatures. An additional concern is the difficulty in conducting experiments of this type from an engineering perspective—isolation and manipulation of particular aspects of speech acoustics in real-time, with sufficiently little delay in transmission to a speaker’s ears, is not a trivial task and may prevent wide application of the feedback perturbation method to both a variety of researchers and a variety of linguistic targets.

This investigation aims to meet these gaps in the feedback perturbation literature by exploring speakers’ responses to altered consonant acoustics, particularly the spectral characteristics of the anterior fricative [ʃ], and by modifying the fricative acoustics in a context-specific way, leaving the high-frequency acoustics elsewhere in the utterance intact. In addition, a more accessible, “offline” method is used to do so.

II. PREVIOUS ACOUSTIC FEEDBACK PERTURBATION STUDIES

The majority of previous real-time acoustic feedback perturbation experiments have focused on altering the acoustics corresponding to vowel quality or pitch on-line, during speakers’ relatively natural production of speech. Although f0 and formant frequency targets differ in their phonetics

^{a)}Portions of this work were presented in “Phonological representation of contextual variants: New evidence from sensorimotor adaptation,” 12th Conference on Laboratory Phonology, Albuquerque, NM, July 2010.

^{b)}Author to whom correspondence should be addressed. Electronic mail: casserly@indiana.edu

and contrastive role in English, results from the domains are largely parallel.

The earliest studies investigating the role of acoustic feedback in speech focused on the perception and production of pitch (Elman, 1981). Elman (1981), in his groundbreaking work, found that speakers who hear their pitch artificially shifted to a fundamental frequency other than that which they intended to produce, alter their productions to compensate for the feedback perturbation. Subjects in these experiments typically do not realize that a shift has been made, and they respond systematically to the perturbation, usually by changing the pitches they produce to counter the effect of the manipulation. That is, if the f_0 of their speech is raised, speakers will produce lower f_0 's in response and vice versa. The basic phenomenon has since been replicated with varying task and manipulation parameters (e.g., Burnett *et al.*, 1997; Burnett *et al.*, 1998; Jones and Munhall, 2000; Donath *et al.*, 2002; Sivasankar *et al.*, 2005) and even implicated as a treatment method for stuttering (Howell *et al.*, 1987; Kalinski *et al.*, 1993).

The pitch-shift paradigm was extended to transformations of vowel formant frequencies by Houde and Jordan (1998, 2002), who systematically shifted F1 and F2 frequencies and observed dramatic changes in speakers' vowel production as a result. Specifically, they found that subjects compensated for increasing shifts in formant frequency, often producing radically different vowels (outside the formant-shifting manipulation) than they intended. Despite this, speakers were naïve to the acoustic manipulation and never became aware of the changes in what they were hearing or producing.

Compensation for altered vowel acoustics has now been explored in detail. Compensation is observed both in whispered and regularly voiced speech (Purcell and Munhall, 2006b), and the compensatory effects disappear slowly after perturbation ends (Houde and Jordan, 1998, 2002; Purcell and Munhall, 2006b; Villacorta *et al.*, 2007; Munhall *et al.*, 2009; Shiller *et al.*, 2009). Villacorta *et al.* (2007) related individual differences in the degree of vocalic compensation to differences in perceptual sensitivity to vocalic contrasts and found generalization of compensation effects across unaltered words containing the same vowel (see also Houde and Jordan, 2002) and (to varying degrees) words containing un-manipulated vowels (Villacorta *et al.*, 2007). Compensation was also recently shown to be outside conscious control, occurring to equal degrees in naïve subjects and those explicitly warned of the manipulation and told to ignore it (Munhall *et al.*, 2009). Finally, Purcell and Munhall (2006a) have also demonstrated that compensation occurs both in response to the typical gradual, between-token perturbation of segmental acoustics and in response to rapid, within-token shifts of formant frequency, which more closely parallel manipulations done in pitch-based investigations.

One major difference between the compensation seen for pitch and formant targets is the degree of individual variability in response quality: nearly every subject in formant-perturbation experiments respond with compensation for the shift in formant frequency, but Burnett *et al.* (1997, 1998) have observed a qualitative behavioral split in responses to their pitch-shifting manipulations. When pitch is altered, the

majority of speakers counter the direction of the shift, but a subset of subjects do precisely the opposite, imitating the experimental feedback perturbation in their own speech. The proportion of those who counter versus imitate the pitch-shift varies across studies and with experimental parameters such as stimulus duration, the amplitude of the feedback, and the magnitude of the pitch-shift, but it seems that individual speakers are consistent in the quality of their response within a given paradigm. The reasons for the split and the motivation for each individual's response quality remain open questions, but speakers who imitate shifts in formant frequency shifts seem to be rarer (Villacorta *et al.*, 2007; Munhall *et al.*, 2009). Individual variation in the degree of compensation is still present, but the qualitative nature of speakers' responses is markedly more consistent in formant-shifting experiments than in manipulations of pitch.

As this review reflects, vocalic and fundamental frequency feedback alterations have been the main focus of perturbation experiments, with very little work investigating compensation for changes in consonant acoustics. The exception is a recent paper by Shiller *et al.* (2009) in which consonant acoustic feedback was manipulated and compensation similar to that described above was observed in response. In their experiment, speakers produced prolonged [s] fricatives whose spectral characteristics were systematically lowered, and speakers' productions were examined for compensatory responses. As in studies of vowel feedback perturbation, their subjects countered the shifting manipulation via a rise in average [s]-frication frequency. The specifics of individual compensation behavior were not described, but these results at least suggest that general responses to perturbation of consonant feedback are not radically different from compensation to vocalic targets. More research, particularly with natural speech, is needed on these understudied consonantal targets, however, to confirm this finding.

Another common thread in the majority of previous feedback compensation work is the use of acoustic perturbations that act globally in the temporal or phonetic domain, affecting either entire utterances or the entire acoustic signal at a particular point in time. Formant-shifting experiments generally operate over whole consonant-vowel-consonant words, affecting F1 and F2 virtually wherever they appear within the word. Stimuli with other, non-manipulated vowels might be present, but whenever an utterance contains a target vowel, the whole utterance is affected by the acoustic transformation. Pitch-shift manipulations, in contrast, typically occur within an utterance and affect only a temporal subset of its production, but the manipulation is phonetically global, altering the entire acoustic signal whenever it applies.

Certainly understanding speakers' responses to such global changes is important, but the effect of time-restricted, phonetically specific changes has not yet been investigated. New information about the role of acoustic speech feedback might be found in more restricted manipulations of speech acoustics that unambiguously apply only during the production of a particular segment and only to its relevant phonetic signatures. The real-time formant-shifting experiments by Purcell and Munhall (2006b) took steps in this direction by applying a time-restricted perturbation to formant-specific

acoustic alteration in long-duration vowel productions, but such a manipulation with more natural speech production is lacking in the literature.

III. CURRENT INVESTIGATION

This experiment aims to expand the feedback perturbation literature by exploring speaker responses to manipulation of consonant feedback and by applying that manipulation in a time- and acoustics-specific manner. Like Shiller *et al.* (2009), the distinction between the English anterior fricatives [s] and [ʃ] served as the target of the feedback manipulation. These fricatives are distinguished primarily by the frequency of their turbulent noise for English speakers (Hughes and Halle, 1956; Harris, 1958; Jongman *et al.*, 2000), with [s] having higher frequency frication (5000 to 6000 Hz) and [ʃ] having noise centered around lower frequencies (2500 to 3000 Hz) (Ladefoged, 2005). Fricative acoustics are extended in time and relatively constant, making them ideal consonant targets for transformation.

The fricative targets were also ideal for setting up a time-restricted and phonetically specific manipulation of their acoustics. The spectral characteristics of speakers' [ʃ] productions were altered by addition of gradually increasing bandwidths of high-frequency turbulent noise to their fricatives. Surrounding vowels were not affected, making the shift temporally local, and only the high-frequency energy of the fricatives were altered, making it phonetically "local" as well. No other aspect of their productions was changed, even the characteristics of high-frequency energy elsewhere in the utterance. The relationship between articulation and acoustic feedback, therefore, was altered only for the alveopalatal fricative [ʃ].

If speakers did not change their [ʃ] articulation in response to the acoustic manipulation, the added noise would quickly result in an ambiguous fricative between [ʃ] and [s]; if speakers lowered the frequency of their [ʃ] due to the added high-frequency noise, the bandwidths of noise—subsequently generated from low-frequency [ʃ]'s—would not reach [s]-like levels. The spectral characteristics of speakers' [ʃ] fricatives over the course of the experiment were measured, therefore, in order to assess compensation to the feedback perturbation. If a speaker responded to the acoustic manipulation, their [ʃ]-frequencies would change significantly over time.

Due to the local nature of the acoustic manipulation and the potential for qualitatively different individual perturbation responses seen in the feedback perturbation literature,

the results of this experiment must be analyzed at a variety of levels. Formant-shifting studies have shown evidence of generalization in compensation behavior to novel words and non-manipulated vowels (Houde and Jordan, 2002; Villacorta *et al.*, 2007), but the restricted context of acoustic perturbation used in this study could result in more variable or restricted responses. Indeed, pilot work on this contrast has suggested that individual speakers respond in an idiosyncratic, restricted fashion more similar to the variability seen in the pitch-shifting literature (Burnett *et al.*, 1998), although such a pattern was not seen in temporally local manipulation of formant frequencies (Purcell and Munhall, 2006a). The generality of speakers' responses to the local manipulation here was therefore investigated explicitly, by examining any differences across the phonetic contexts of [ʃ]. The responses of each individual speaker were similarly explored.

Finally, this experiment introduces a new feedback perturbation methodology that might be more accessible to researchers with a wide variety of backgrounds. Although the traditional perturbation studies discussed in Sec. II have proven extremely valuable, they are limited to a relatively small community with a high level of technical proficiency. Perturbation studies as reported in the literature are conducted via real-time, online manipulation of speakers' acoustics; the manipulated speech signal must be fed back to the speaker with very little time lag or problematic interference will occur. Indeed, delayed speech feedback has been shown to severely compromise speakers' fluency (Howell *et al.*, 1987). Therefore, the time constraints on any manipulation of acoustic outputs in perturbation experiments are necessarily quite severe.

This problem has been solved in the past either by targeting relatively simple and consistent elements of speech acoustics, such as pitch, which can be quickly altered, or by implementing the signal processing in special hardware built to rapidly identify, isolate, and manipulate acoustic characteristics such as formant frequencies (e.g., Villacorta, 2005).

This study, by contrast, utilizes an "offline" signal manipulation that places fewer demands on the speed capabilities of the signal processing. Rather than analyzing and altering a speech sample in real-time, this method delays processing until the interval between trials, generating each trial's acoustic manipulation from the production that preceded it. As the schema in Fig. 1 below, the result is a system retains the sensitivity to changes in speaker behavior

FIG. 1. Procedural schema for the offline feedback perturbation method. (A) Final trial of baseline, without acoustic alteration. Speaker produces a vowel-[ʃ]-vowel nonword. (B) Production is recorded and the voiceless fricative is automatically excised from the file (vertical lines). (C) Power spectrum is computed from the fricative and its frequency bandwidth (schematic cutoffs shown by vertical lines) is noted. (D) Synthetic fricative containing turbulent noise starting at the topmost edge of the observed frequency bandwidth is created. (E) High-frequency fricative noise is played over headphones to the speaker during production of their next fricative. That natural production starts the cycle over at (A).

that is a key feature of real-time perturbation methods, but that sensitivity is delayed in time by one utterance.

In this investigation, therefore, the acoustics of speakers' [ʃ] productions—and only their [ʃ] productions—were altered, so that they were ambiguous between [s] and [ʃ] and the fricative frequencies produced in response to the manipulation were measured to assess speaker compensation. The feedback perturbation was implemented with a one-trial delay so that signal processing can take place in the interval between trials, and it affected only the relevant high-frequency acoustics of the target fricatives, rather than the entire utterance or acoustic signal. In order to assess the effects of the context-specific manipulation both across phonetic contexts and individuals, the results were analyzed at multiple levels of detail.

This experiment therefore fills a gap in the current acoustic feedback perturbation literature by addressing the effects of temporally and phonetically local feedback manipulation on both group and individual responses to altered consonant acoustics across phonetics contexts, and it does so using a widely accessible signal processing paradigm.

A. Subjects

Thirty-two native speakers of American English participated in this experiment in exchange for course credit. All participants were currently enrolled as undergraduates at Indiana University and completed a demographic and language background questionnaire. Twelve men and 20 women between the ages of 18 and 24 yr (mean, 19.5 yr) participated. No participant reported a known history of speech, language, or hearing disorders. Twenty-four speakers experienced the experimental perturbation of speech feedback acoustics while the remaining eight served as a control group, performing the same speech tasks without feedback perturbation.

B. Procedure

All aspects of the procedure described below were identical for speakers in the experimental and control groups, except that control speakers received no acoustic perturbations.

Speakers were seated in a sound-attenuating booth (Industrial Acoustics Company, Inc) in front of a TDN laptop computer (Technology Distribution Network, Inc., Tonawanda, NY) and were fitted with Radioshack Titanium headphones (Radioshack Corporation, Fort Worth, TX). The headphones were not designed to attenuate ambient sound and no attempt was made to lessen or eliminate bone conduction of speakers' productions. This access to noise in the environment was desirable in this set up because of the acoustic manipulation used (see Sec. III C), which involved a summation or merging of the speaker's own fricative production and the synthetic added noise.

Speakers in the experiment were instructed to read two-syllable nonwords in response to on-screen prompts. Each nonword contained the fricative [ʃ] immediately preceded and followed by one of three vowels: [i], [a], and [u]. The experimenter produced the three resulting stimuli—[iʃi], [aʃa], and [uʃu]—for each speaker to resolve any ambiguities in the standard orthography (“ee-shee,” “ah-shah,” and “oo-shoo,” respec-

tively). During the experiment, the three nonwords were elicited in randomized blocks of three so that each nonword occurred once in a group of three trials in random order. The maximum distance between identical elicitations, therefore, was five trials.

Stimulus elicitation used separate, regularly timed prompts for each nonword syllable. Subjects were free to respond to the syllable prompts in whatever manner felt most natural to them. No instructions were given with respect to the placement of stress or the duration of either syllable. Each trial began by showing speakers the nonword they would be asked to say at the top of the screen for 1 s. A countdown in the form of a stoplight prepared subjects for the first syllable elicitation prompt, which appeared 500 ms after the stoplight counter left the screen. The orthography of the first syllable of the nonword (e.g., “ah”) was presented in the middle of the screen for 250 ms, followed by 500 ms of blankness and 500 ms of the orthography of the second syllable (e.g., “shah”). The interval between trials incorporated processing time for the manipulation of the acoustics and was therefore variable, but no shorter than 1 s.

The regular timing of the prompts encouraged speakers to adopt a repeatable, rhythmic speaking style that constrained the natural variability in segment duration and timing relative to utterance onset. Five subjects from the experimental group were selected at random to verify the effectiveness of this technique. The naturally produced fricatives of these subjects showed a maximum trial-to-trial difference of 111 ms and an average difference of 25 ms. The production data from these five subjects were also used to estimate the accuracy of the durational match between natural and added synthetic fricative noise (see Sec. III C).

A custom program on the laptop, implemented in MATLAB (Mathwork, Ltd., Natick, MA), handled stimulus presentation, recording of speakers' utterances on a SigmaTel Audio microphone, (SigmaTel, Inc., Austin, TX) signal analysis, synthesis of frication noise to be added to speakers' next trial, and appropriate presentation of the created fricative over headphones. The entire session was simultaneously recorded via an Audio-Technica AT2041SP microphone (Audio-Technica US, Inc., Stow, OH) onto a Marantz PMD 660 solid state recorder (D&M Professional, Marantz America, Inc., Itasca, IL) at 48 kHz. Audio captured by the later microphone was then used in analysis of speakers' fricative acoustics.

The experimental session consisted of ten baseline trials, with no acoustic manipulation of speakers' productions, followed by 60 experimental trials. The baseline trials served the dual purpose of familiarizing speakers with the rhythmic elicitation of nonword syllables and enabling collection of several tokens of [ʃ] from each speaker in each vowel context before the feedback perturbation began. Due to the offline perturbation method being used in this study, the final feedback trial served as the input for the acoustic manipulation of the first experimental trial (see the schema in Fig. 1). All elicitations, baseline and experimental, were recorded and analyzed but only the experimental trials will be reported below.

C. Acoustic manipulation

Using a delayed analysis and synthesis method, speakers' high-frequency fricative acoustics were selectively

perturbed over the course of the experiment. Each vowel- $[\text{f}]$ -vowel nonword production was in response to regularly timed prompts for each syllable. As the subjects produced the second, fricative-initial syllable, synthesized high-frequency turbulent noise was played over their headphones. As a result, the subjects heard both their own fricative production and the high-frequency synthetic turbulent noise. The two noise sources together resulted in a higher average frequency of frication than the natural fricative alone, potentially raising its centroid frequency close to values characteristic of $[\text{s}]$ rather than $[\text{f}]$.

Added turbulent noise was synthesized in between trials, using the acoustic characteristics and duration of the previously naturally produced fricative as a guide. That is, on each trial a MATLAB script recorded the speaker's nonword production and excised the voiceless fricative from the vocalic context. A spectrum-based algorithm separated the fricative from the surrounding vowels by examining the spectral centers of gravity within consecutive 300-sample time windows—spectra with centroid frequencies above approximately 2000 Hz were identified as voiceless fricatives. Once the fricative was identified, its duration and power spectrum were computed and used to dictate the duration and acoustic characteristics, respectively, of the synthetic fricative noise.

The timing of the synthetic noise presentation with respect to the subject's natural fricative was crucial to the success of the manipulation, particularly since the added acoustics were not synthesized in real-time. The synthetic fricatives were presented over the headphones simultaneously with the visual presentation of the prompt for the second nonword syllable, a time that was predictable for subjects due to the consistent inter-prompt timing. Since the target fricative is the onset of that syllable, subjects produced it close to the appearance of the prompt. The duration of the synthetic turbulent noise was set to match the duration of the previously recorded natural $[\text{f}]$ fricative. Analysis of the duration match between natural and synthesized fricatives was completed for five randomly selected subjects (see Sec. III B). The durations were significantly different (paired t -test, $p < 0.01$), with the synthetic fricative being an average of 15.2 ms longer than its naturally produced counterpart. This slight, consistent overestimation was present for all five subjects. As a result, the synthetic fricative was more likely to overlap the entire naturally produced fricative on the next trial than if the duration match had been exact. In order to minimize the effect of duration mismatch, however, the synthetic fricatives were passed through an envelope filter that decreased the amplitude of the onset and offset, making them perceptually less abrupt. The combined effect of the presentation at the onset of the second syllable prompt and the consistent minor over-matching of fricative duration is a stimulus that merged effectively with speakers' naturally produced fricatives, without affecting the surrounding phonetic context.

The specific acoustic characteristics of the added turbulent noise were also important for the manipulation. Synthetic turbulent noise was used to raise the spectral center of gravity across the natural and synthetic fricatives, so the added turbulent noise needed to be of higher frequency than

FIG. 2. Summary of the acoustic manipulation. (A) Natural nonword production, (B) synthetic turbulent noise added to the production in (A), and (C) simulated summation of the natural and synthetic signals in (A) and (B). Right column shows the synchronized waveforms for each item, left column contains power spectra (100 Hz bandwidth cepstral smoothing) from the fricative in each waveform.

the natural fricative, but similar enough that the two noise sources would perceptually merge. The power spectrum generated from the previous naturally produced fricative was therefore used to determine lower boundary of the frequencies present in the synthetic noise. The point at which the natural spectrum dropped off in strength served as the synthesis starting point. If speakers raised or lowered the frequencies of their $[\text{f}]$ fricatives, the reference point for the frequency of the synthesized fricative would likewise raise or lower.

The total bandwidth of synthetic noise increased over the course of the experiment by steps of 250 Hz every four trials, so that the upper boundary of the frequency of synthetic noise was initially 250 Hz above the lower boundary (determined by the $[\text{f}]$ power spectrum) and ended 3750 Hz above. Synthesis of turbulent noise was achieved by addition of low-amplitude sinusoids of randomly selected frequencies within the appropriate bandwidth; resulting amplitude was held constant at approximately 50 dB across trials, for all subjects.

Figure 2 summarizes the acoustic manipulation used in this experiment. Panels on the right show waveforms from a speaker's natural vowel-fricative-vowel nonword production, the synchronized synthetic noise played over the headphones, and a simulation of the resulting perceptual summation of the natural and synthetic signals. Panels on the left show the spectral characteristics of the naturally produced alveopalatal fricative, the synthetic fricative (of the maximum bandwidth, 3750 Hz) generated from the previous trial, and the simulated altered fricative resulting from the combination of the two turbulent noise sources. Figure 3

FIG. 3. Spectrum from a naturally produced alveopalatal fricative (dashed line) overlaid with the spectrum from a simulation of the same fricative after it has been merged with the added synthetic noise (solid line). The merged spectrum illustrates the effect of the acoustic manipulation, raising the fricative centroid from 4526.1 to 5243.8 Hz.

highlights the difference between the natural (dashed line) and altered (solid line) fricative spectra and gives the centroid frequencies for each fricative (4526.1 and 5243.8 Hz, respectively).

D. Data analysis

Speakers' [ʃ] fricatives were excised manually by the author from the digital recordings of each trial and analyzed using PRAAT 5.0.40 (Boersma and Weenink, 2008). For each fricative, a centroid frequency or spectral center of gravity was then computed from a frequency spectrum. These centroids represent the average frequency present in the spectrum have been shown to largely determine the perceptual and productive distinction between the anterior fricatives [ʃ] and [s] for native speakers of English (Harris, 1958; Jongman *et al.*, 2000).

If speakers adjusted their productions of [ʃ] in response to the experimental perturbation of their fricative acoustics, significant changes should be observed in their centroid frequencies over time. Subjects' responses to the perturbation were therefore assessed via unpaired *t*-tests ($\alpha = 0.05$) comparing the distribution of centroids produced in the first ten experimental trials to those from the last ten experimental trials. Large changes in centroid frequency therefore resulted in statistically significant *t*-scores. These comparisons of means were conducted across- and within-subjects on the raw centroid frequencies collected. In addition to these analyses, correlations were also calculated between centroid frequency throughout the experiment and trial number (as an index of time). Whenever correlations are reported across subjects, normalized centroid values that converted the frequencies for each subject to *z*-scores were used in the analysis. Speakers vary widely in their typical fricative frequencies, but normalization allowed their data to be pooled effectively for the group correlations.

IV. RESULTS

The results below report the data from 20 speakers in the experimental group; four subjects were eliminated from

TABLE I. Individual changes in fricative centroid from the first ten manipulation trials to the last ten. *P*-values from unpaired two-tailed *t*-tests.

Subject	Change in centroid (end-beginning, Hz)	Response type	<i>p</i> -value
3	−849.8 ^a	counter	0.016
11	−252.8 ^b	counter	0.000
15	−392.2 ^a	counter	0.030
16	−397.5 ^c	counter	0.063
20	−265.6 ^b	counter	0.003
5	+428.0 ^a	imitate	0.029
6	+222.7 ^b	imitate	0.001
9	+697.1 ^b	imitate	0.000
12	+342.8 ^a	imitate	0.014
17	+218.7 ^c	imitate	0.067
18	+186.1 ^a	imitate	0.032
19	+1066.8 ^a	imitate	0.022
23	+536.3	imitate	0.055
1	−95.3	(no response)	0.248
2	+46.6	(no response)	0.417
4	−98.3	(no response)	0.536
7	−158.4	(no response)	0.190
21	−225.3	(no response)	0.379
22	+67.5	(no response)	0.690
24	−272.7	(no response)	0.437

^a*p* < 0.05.

^b*p* < 0.01.

^cMarginal: *p* < 0.07.

analysis due to equipment malfunction (*n* = 3) and mispronunciation of nonword targets ([uʃu] as [ouʃou]; *n* = 1). Across all subjects' centroid frequencies from all three vowel contexts, the *t*-test comparing the first and last ten manipulation trials was not significant (*p*'s > 0.05). Within-subjects *t*-tests of specific individuals' centroid frequencies, however, did reveal significant changes in [ʃ] frequency over time for ten of the 20 speakers and marginally significant changes (*p*'s < 0.068) for three additional speakers (Table I). Of the significant and marginally significant results, the largest change in fricative centroid (averaged across the block of ten trials) was +1066.8 Hz (*p* = 0.022) and the smallest was +186.1 Hz (*p* = 0.032).

Table I also shows that, for the speakers whose fricatives changed significantly in response to the perturbation, the behavioral split between those who imitated the feedback perturbation and those who countered it was relatively even. That is, some of the talkers (*n* = 5) lowered their [ʃ] frequencies in a countering response to rising frequencies in fricative feedback, and a slightly larger group (*n* = 8) imitated the rising frequency of their fricative feedback by raising the frequency of their [ʃ]'s over time. While the two groups are not even, the opposite nature of their responses could explain the lack of main effect across the entire experimental group.

In order to help determine whether the significant changes in [ʃ] centroid frequencies were the result of noise in the data or a truly consistent behavioral response on the part of the speakers, each speaker's data were analyzed by stimulus item. This level of analysis was planned in order to investigate potential context-specific effects of the local acoustic manipulation, and independently motivated by pilot

TABLE II. Individual changes in fricative centroid by context. Significant or marginally significant changes are in bold.

Subject	[i]		[a]		[u]	
	Context change (Hz)	Response type	Context change (Hz)	Response type	Context change (Hz)	Response type
1	-263.9		+94.5		-31.6	
2	+171.9		-99.9		+101.7	
3	-849.8^b	counter	-120.0		-96.4	
4	-80.4		-101.0		+309.6	
5	+409.9		+370.0		+154.4	
6	+66.0		+283.9^c	imitate	+260.2^a	imitate
7	-71.5		-323.2^a	counter	+59.9	
9	+429.9		+579.5^a	imitate	+547.2^a	imitate
11	-138.5		-305.8^a	counter	-373.2^c	counter
12	+404.7		+284.4		+85.2	
15	-217.0		-1.4		-1063.0^a	counter
16	-507.2		-832.3^b	counter	+194.6	
17	+33.0		+78.0		+720.2	
18	+645.1^a	imitate	+33.5		+121.7	
19	-289.8		+373.3		+754.0	
20	-166.3^c	counter	-292.5		-291.6^c	counter
21	+67.6		-421.3		-25.9	
22	+68.6		-226.7		+107.5	
23	+399.0		+252.7		-36.8	
24	+506.1^a	imitate	+4.8		-400.1	

^a $p < 0.05$.

^b $p < 0.01$.

^cMarginal: $p < 0.07$.

data suggesting a significant effect of the vocalic context on speakers' behavior.

Results of t -tests on the centroids of each speaker (even those whose responses were not significant in Table I) within each vocalic context, [i_i], [a_a], or [u_u], are given in Table II. Most speakers whose across-context centroids significantly changed between the first and last ten experimental trials (Table I) also showed significant changes within at least one phonetic context, and two additional speakers' fricative acoustics were also found to change significantly at this level of analysis. A total of 15 subjects, therefore, showed significant alteration of fricative acoustics at this context-specific level. Most importantly, the nature of an individual's response—imitating or countering the perturbation—was consistent across contexts for each of these 15 speakers. This consistency suggests that the observed changes in [ʃ] production are indeed a systematic response to the acoustic perturbation, rather than a coincidence of normal variability in fricative acoustics.

Based on the consistency of individual speakers' response quality, the experimental participants were split into three groups: those who imitated the fricative frequency perturbation ($n = 9$) at any level of analysis; those who countered it ($n = 6$); and those whose fricatives did not change significantly over the course of the experiment, at any level of analysis in any context ($n = 5$). Between-subjects unpaired t -tests were conducted on the raw centroids of each qualitative response group. As the data in Table III show, imitating and countering speakers now showed significant changes to their across-context fricative acoustics (imitation subgroup change was +396.0 Hz, $p < 0.01$; countering subgroup change was -332.5 Hz, $p < .001$), while the unresponsive group's change was not significant ($p = 0.698$). The trajectories of change in each group over the course of the entire manipulation are shown in Fig. 4, plotting normalized centroid z -scores against trial number as an index of time.

When speakers whose responses differ qualitatively are grouped appropriately, therefore, their pooled responses appear to show no context-specificity; changes in fricative centroid occur in all three vocalic contexts and the individual variability is washed out. Comparison of means via t -tests run on the context-specific centroid frequencies for each subgroup confirmed the observed uniformity in the group data, revealing significant changes in fricative acoustics for all vocalic contexts in both groups (Table III).

Underlying the group context-generalty, however, is a strong effect of vocalic context, as seen in Table II and summarized in Table IV. Fifteen of the 20 speakers experiencing the acoustic manipulation responded by altering the characteristics of their fricative production, but the majority of speakers systematically altered their [ʃ] productions in only a subset of the vocalic contexts, either two ($n = 4$) or in a single vowel environment ($n = 6$). Changes in particular contexts did not reach significance for the remaining five speakers with significant across-context change, presumably because their responses were more general and distributed across stimuli. The centroid trajectories for one striking subject with context-restricted responses are shown in Fig. 5, where the systematicity seen in the [a_a] stimulus context is clearly absent from centroids produced in [i_i] and [u_u] nonwords.

Despite the idiosyncratic context-specific behavior in the three vowel contexts, speakers' responses across contexts were not random: the r -values seen in time/centroid correlations for each context are themselves significantly correlated across subjects, revealing a pattern of directionality and

TABLE III. Between-subjects changes in fricative centroid frequency for speaker subgroups with different qualitative responses to the feedback perturbation.

Qualitative subgroup	All contexts change (Hz)	All contexts p -value	[i] Context change (Hz)	[a] Context change (Hz)	[u] Context change (Hz)
Imitating	+396.0	0.001	+219.3 ^a	+226.3 ^a	+225.9 ^b
Countering	-332.5	0.006	-325.0 ^a	-312.5 ^a	-261.6 ^b
No response	-51.2	0.698	-7.2	-150.9	+92.3

^a $p < 0.01$.

^b $p < 0.05$.

FIG. 4. Normalized centroid frequencies for [j] produced over the course of the experimental perturbation (trial numbers 10–70) by speakers in each of three response quality groups: those who imitated the acoustic manipulation, those who countered it, and those whose response did not change. Changes in centroid were significant (t -tests comparing first and last ten experimental trials, p 's < 0.05) for the countering and imitating subgroups and not significant for the no response group ($p > 0.05$). Linear regressions representing the direction of change in fricative frequency are also shown on each graph.

variance. Correlations were significant between responses in [i_i] and [a_a] contexts ($r = 0.629$, $p < 0.01$) and between responses in [a_a] and [u_u] contexts ($r = 0.532$, $p < 0.01$).

Fricatives produced by the eight control subjects, who completed the identical experimental procedure without the

perturbation of their acoustic outputs, were also analyzed. Unpaired t -tests of overall centroid means in the same two blocks of ten trials were not significant for the group or individuals (all $p > 0.05$).

V. DISCUSSION

This study succeeded in using an offline feedback perturbation to induce systematic changes in subjects' speech production. Although a qualitative behavioral split between those who countered the artificial rise in fricative frequency and those who imitated it obscured the systematicity across the whole experimental group, the responses within each behavioral subgroup were significant and consistent across the three phonetic contexts elicited in the experiment. The lack of any significant change over time in the fricative acoustics of control speakers, who experienced no feedback perturbation, rules out the possibility that the observed responses are the coincidental result of normal variation in fricative acoustics or the result of articulatory fatigue from repeated elicitations. Based on these data, therefore, we can conclude that the less technically demanding delay in signal processing and presentation used in this study still resulted in systematic speaker alterations in fricative production in the face of acoustic feedback perturbation.

This experiment differs from the majority of previous work in the paradigm by altering speaker acoustics in a

TABLE IV. Response profiles by stimulus phonetic context for speakers in the imitating and countering subgroups.

Qualitative subgroup	Subject	Number of response contexts	Vowel contexts with significant response
Imitating	5	0	Overall
	6	2	[a_a], [u_u]
	9	2	[a_a], [u_u]
	12	0	Overall
	17	0	overall
	18	1	[i_i]
	19	0	overall
	23	0	overall
	24	1	[i_i]
Countering	3	1	[i_i]
	7	1	[a_a]
	11	2	[a_a], [u_u]
	15	1	[u_u]
	16	1	[a_a]
	20	2	[i_i], [u_u]

FIG. 5. [ʃ] centroid frequencies over the course of the acoustic perturbation for a single subject whose productions countered the manipulation. Although this speaker's response was significant in the context of [a] (change of -323 Hz, $p = 0.047$), no significant change in fricative acoustics was observed in the context of either [i] or [u] (p 's > 0.05).

temporally and phonetically local manner. The high-frequency spectral characteristics of the anterior fricative [ʃ] alone were manipulated, leaving other high-frequency acoustic information and the rest of the utterance unchanged. This distinction appears to have affected the nature of the perturbation response: whereas speaker responses to global changes in previous feedback perturbation experiments have been characterized by generalization of compensation effects to novel lexical items and vowels, the responses observed in our study were more idiosyncratic and restricted. In fact, 10 of the 15 speakers who responded to the acoustic manipulation systematically altered their fricative acoustics in only one or two vocalic contexts, while none showed significant changes in [ʃ] spanning all three. Whether this specificity is a direct result of the temporal and phonetic specificity of the feedback manipulation or is due to the consonantal linguistic target remains an open question, as the only other study targeting consonants did not report the behavior of individual participants (Shiller *et al.*, 2009). Regardless, this ability for speakers to differentially alter their production of [ʃ] across the three contexts is surprising, especially given that the contexts were produced in random, interleaved order.

Moreover, as the data in Table IV show, the specific context in which speakers responded to feedback perturba-

tion does not appear to be consistent across individuals. That is, a similar number of speakers responded in each of the three contexts ($n = 4$ for [iʃi], $n = 5$ for [aʃa], and $n = 5$ for [uʃu]) and no relationship between speakers' response quality—imitating or countering—and response context is apparent. Several hypotheses for the motivation of individuals' response contexts were tested and rejected by the author, including a ceiling or floor effect of fricative frequency (which might potentially determine the contexts in which fricatives were free to be manipulated), a relationship to the speakers' perceptual sensitivity to or use of vocalic context in categorization of the [ʃ]/[s] continuum (explored in a perceptual experiment conducted with these same participants), or the amplitude of the speakers' fricatives [as might be expected from Burnett *et al.* (1998)], either as absolute amplitude or amplitude relative to the speaker's average vowel sound pressure level (SPL). In none of these cases was there any relationship between the effect in question and the degree or type of changes in fricative centroid observed here in response to altered acoustic feedback. Further research is necessary, therefore, to determine the causes underlying speakers' particular response contexts.

Despite the specificity, speakers' responses in each context are not entirely unrelated. The direction and degree of centroid frequency change over time (represented by the r -value of the correlation between centroid and trial number) is significantly correlated in [i] and [a] contexts and in [a] and [u] contexts, meaning that some proportion of the response variance is shared between [i] and [a] and between [a] and [u]. The lack of shared variance between [i] and [u] contexts indicates that whatever they each share with responses in the [a] context does not overlap. These specific relationships are linguistically logical: in terms of distinctive phonological features, [i] and [a] are both [-round], while [a] and [u] are both [+back]. Vowel rounding and backness are well-known factors affecting fricative perception and production (Mann and Repp, 1980; Whalen, 1981; Nittrouer and Studdert-Kennedy, 1987), while vowel height, the feature shared between [i] and [u] but distinct from [a], has not been similarly implicated. Functional relationships between phonetic contexts do play a role in speakers' responses to perturbation of fricative acoustics, therefore, but they are not the sole determiners of the complex behavior.

Finally, the results revealed a relatively even split between speakers who imitate or counter feedback perturbation. A similar level of division has also been observed in pitch-shifting studies (Burnett *et al.*, 1998), but many fewer speakers imitating acoustic manipulations have been seen previously in experiments targeting segmental acoustics (Villacorta *et al.*, 2007; Munhall *et al.*, 2009). In the only other perturbation experiment to date that targets consonant feedback, Shiller *et al.* (2009) do not report the nature of individual responses to their manipulation of [s] acoustics, but did obtain an effect when averaging across speakers, suggesting at the least a group bias toward countering the manipulation. Whether the greater proportion of speakers imitating the acoustic manipulation seen in this study results from its novel methodology, its consonant target or its local acoustic perturbation, therefore, must be investigated more fully in future research.

VI. CONCLUSION

This study extends the reach of acoustic feedback perturbation studies, first by investigating the effects of a temporally and phonetically local acoustic feedback manipulation and also by using a relatively undemanding signal processing method to alter speakers' fricative acoustics. The offline method successfully produced systematic alterations of participants' speech, and the local manipulation resulted in changes to fricative production that were more variable, idiosyncratic and restricted than the compensation seen in global acoustic alterations. Speakers were equally likely to imitate or counter the fricative acoustic manipulation and the majority did so differentially across phonetic contexts. While speaker behavior across contexts was not entirely unrelated, many questions remain regarding the mechanism of such context-specific manipulation of fricative production, the causes underlying the particular response profiles of individual talkers and the difference in speaker response for segment-specific versus global acoustic feedback manipulation targets.

ACKNOWLEDGMENTS

The author would like to thank Ken de Jong for his suggestions and assistance during data analysis and Anders Löfqvist and two anonymous reviewers for their helpful comments. Research supported by the National Science Foundation Graduate Research Fellowship Program and NIH NIDCD Research Grant No. R01-DC0011.

- Baum, S. R., and McFarland, D. H. (1997). "The development of speech adaptation to an artificial palate," *J. Acoust. Soc. Am.* **102**, 2353–2359.
- Boersma, P., and Weenink, D. (2008). "Praat: Doing phonetics by computer," Version 5.0.40 [Computer program], Institute of Phonetic Sciences, Amsterdam.
- Burnett, T. A., Senner, J., and Larson, C. R. (1997). "Voice responses to pitch-shifted auditory feedback: A preliminary study," *J. Voice* **11**, 202–211.
- Burnett, T. A., Freedland, M. B., Larson, C. R., and Hain, T. C. (1998). "Voice F_0 responses to manipulations in pitch feedback," *J. Acoust. Soc. Am.* **103**, 3153–3161.
- Chen, S. H., Liu, H., Xu, Y., and Larson, C. R. (2007). "Voice F_0 responses to pitch-shifted voice feedback during English speech," *J. Acoust. Soc. Am.* **121**, 1157–1163.
- Donath, T. M., Natke, U., and Kalveram, K. T. (2002). "Effects of frequency-shifted auditory feedback on voice F_0 contours in syllables," *J. Acoust. Soc. Am.* **111**, 357–366.
- Elman, J. L. (1981). "Effects of frequency-shifted feedback on the pitch of vocal productions," *J. Acoust. Soc. Am.* **70**, 45–50.
- Fowler, C. A., and Turvey, M. T. (1980). "Immediate compensation in bite-block speech," *Phonetica* **37**, 306–326.
- Harris, K. S. (1958). "Cues for the discrimination of American English fricatives in spoken syllables," *Lang. Speech* **1**, 1–7.
- Houde, J. F., and Jordan, M. I. (1998). "Sensorimotor adaptation in speech production," *Science* **279**, 1213–1216.
- Houde, J. F., and Jordan, M. I. (2002). "Sensorimotor adaptation of speech I: compensation and adaptation," *J. Speech Lang. Hear. Res.* **45**, 295–310.
- Howell, P., El-Yaniv, N., and Powell, D. J. (1987). "Factors affecting fluency in stutterers when speaking under altered auditory feedback," in *Speech Motor Dynamics in Stuttering*, edited by H. Peters and W. Hulstijn (Springer Press, New York), pp. 361–369.
- Hughes, G. W., and Halle, M. (1956). "Spectral properties of fricative consonants," *J. Acoust. Soc. Am.* **28**, 303–310.
- Jones, J. A., and Munhall, K. G. (2000). "Perceptual calibration of F_0 production: Evidence from feedback perturbation," *J. Acoust. Soc. Am.* **108**, 1246–1251.
- Jones, J. A., and Munhall, K. G. (2002). "The role of auditory feedback during phonation: Studies of Mandarin tone production," *J. Phonetics* **30**, 303–320.
- Jongman, A., Wayland, R., and Wong, S. (2000). "Acoustic characteristics of English fricatives," *J. Acoust. Soc. Am.* **108**, 1252–1263.
- Kalinowski, H., Armson, J., Roland-Mieszkowski, M., and Stuart, A. (1993). "Effects of alterations in auditory feedback and speech rate on stuttering frequency," *Lang. Speech* **36**, 1–16.
- Kelso, J. S., Tuller, B., Vatikiotis-Bateson, E., and Fowler, C. A. (1984). "Functionally specific articulatory cooperation following jaw perturbations during speech: Evidence for coordinative structures," *J. Exp. Psychol. Hum Percept. Perform.* **10**, 812–832.
- Ladefoged, P. (2005). *A Course in Phonetics*, 5th ed. (Thomson/Wadsworth, Boston, MA), pp. 191–197.
- Mann, V. A., and Repp, B. H. (1980). "Influence of vocalic context on perception of the [sh]-[s] distinction," *Percept. Psychophys.* **28**, 213–228.
- Munhall, K. G., MacDonald, E. N., Byrne, S. K., and Johnsrude, I. (2009). "Talkers alter vowel production in response to real-time formant perturbation even when instructed not to compensate," *J. Acoust. Soc. Am.* **125**, 384–390.
- Nittrouer, S., and Studdert-Kennedy, M. (1987). "The role of coarticulatory effects in the perception of fricatives by children and adults," *J. Speech Hear. Res.* **30**, 319–329.
- Purcell, D. W., and Munhall, K. G. (2006a). "Compensation following real-time manipulation of formants in isolated vowels," *J. Acoust. Soc. Am.* **119**, 2288–2297.
- Purcell, D. W., and Munhall, K. G. (2006b). "Adaptive control of vowel formant frequency: Evidence from real-time formant manipulation," *J. Acoust. Soc. Am.* **120**, 966–977.
- Shiller, D. M., Sato, M., Gracco, V. L., and Baum, S. R. (2009). "Perceptual recalibration of speech sounds following speech motor learning," *J. Acoust. Soc. Am.* **126**, 1103–1113.
- Sivasankar, M., Bauer, J. J., Babu, T., and Larson, C. R. (2005). "Voice responses to changes in pitch of voice or tone auditory feedback," *J. Acoust. Soc. Am.* **117**, 850–857.
- Villacorta, V. M. (2005). "Speech sensorimotor adaptation to acoustic perturbations in the first formant of vowels and its relation to perception," Unpublished doctoral thesis. MIT, Cambridge, MA.
- Villacorta, V. M., Perkell, J., and Guenther, F. H. (2007). "Sensorimotor adaptation to feedback perturbations of vowel acoustics and its relation to perception," *J. Acoust. Soc. Am.* **122**, 2306–2319.
- Whalen, D. H. (1981). "Effects of vocalic formant transitions from vowel quality on the English [s]-[ʃ] boundary," *J. Acoust. Soc. Am.* **69**, 275–282.
- Xu, Y., Larson, C. R., Bauer, J. J., and Hain, T. C. (2004). "Compensation for pitch-shifted auditory feedback during the production of Mandarin tone sequences," *J. Acoust. Soc. Am.* **116**, 1168–1178.