INTRODUCTION

Speech production is a highly complex task that requires coordinating rapid movements of numerous vocal tract muscles (Zemlin et al. 1998). Yet, the motor control demands of this important form of communication are met with relative ease by most speakers at a young age (Tsao and Weismer, 1997). Children entering grade school have typically learned an inventory of speech motor programs for most phonemes of their native language that they can combine to effectively produce long, complex sentences (McLeod and Bleile, 2003). Auditory feedback (AF) plays an important role in this learning process. The absence of AF at an early age profoundly disrupts normal speech development (Oller and Eilers, 1988) while a loss of hearing following normal speech development reduces intelligibility (Waldstein, 1990; Lane and Webster, 1991). By manipulating AF during speech, many studies have shown that as we speak, we continually monitor our vocal output, modulating it to meet the demands of our environment (e.g., Lane and Tranel, 1971; Summers et al., 1988) and correcting mismatches between what we intended to say and what we perceived ourselves saying (e.g., Larson et al., 2000; Heinks-Maldonado and Houde, 2005; Bauer et al., 2006; Purcell and Munhall, 2006a). These mismatches, or errors, provide a signal that is used to tune our speech motor programs (e.g., Houde et al, 2002; Jones and Munhall, 2005; Purcell and Munhall, 2006b; Villacorta et al., 2007).

Persistent developmental stuttering affects 1% of the adult population and 5% of children (Bloodstein and Ratner, 2008). It is characterized by disruptions of fluent speech in the form of sound and syllable repetitions, prolongations, and blocks and can severely interfere with verbal communication. The etiology of the disorder remains unknown and the long-term success of therapies remains limited. However, short-term reduction of stuttering symptoms has resulted from various AF manipulations, including noise masking (e.g., Sutton and Chase, 1961; Conture and Brayton, 1975), delayed auditory feedback (e.g., Webster et al., 1970; Stager et al., 1997) and whole-spectrum frequency shifts (e.g., Kalinowski et al. 1993, Ingham et al. 1997). Paradoxically, delayed AF disrupts the speech of typically fluent speakers, causing slowed speech or "stuttering like" disfluencies (e.g., Lee, 1951, Yates, 1963, Stuart et al., 2002). Such findings suggest that aberrant auditory-motor interactions may underlie persistent developmental stuttering.

Over the past several years, we have used unexpected perturbations of AF to investigate the neural mechanisms underlying AF-based speech motor learning and control. Using functional magnetic resonance imaging we found that compensation for unexpected perturbations of the 1st formant (F1) of speakers' AF resulted in right-lateralized increases in brain responses in lateral frontal cortex (Tourville, et al., 2008). Greater activity in lateral frontal cortex in during speech in persons who stutter (PWS) than in persons with fluent speech (PFS) has also been reported (e.g., Braun et al., 1997; De Nil et al., 2000). The similarity between brain responses associated with AF-based corrective movements and that of PWS speech further suggests that stuttering may be associated with abnormal auditory-motor interactions during speech. The finding prompted us to expand our efforts to characterize how AF influences speech production in normal development and to investigate its relationship with stuttering.

In this paper, we describe results from our recent studies of AF-based control of speech in both PFS and PWS. We also detail the expanded capabilities of the Audapter AF manipulation system, which enable highly configurable dynamic perturbations of the spatial and temporal content of a wide range of speech production parameters.

AUDITORY FEEDBACK BASED CONTROL OF SPEECH IN PFS AND PWS

Control of Static Intra-segmental Formants

Fluent speakers compensate for unexpected static perturbations of F1 during the vowel in mono-syllabic /CɛC/ by shifting the F1 of their vocal output in the direction opposite the perturbation (Tourville et al., 2008). For instance, speakers respond to an upward F1 perturbation that makes "head" sound like "had" by producing a word that sounds more like "hid" than she would normal produce. The compensatory F1 shift makes the word that the speaker hears sound more like the intended target. This response occurs within 165 ms of voicing onset, fast enough to permit online correction of the target /CɛC/ utterance (Hillenbrand et al., 2001). Using a static perturbation of F1 during a monopthong in isolated monosyllabic words, similar to that induced by Tourville et al., we observed that PWS, as a group, showed significantly weaker online formant correction compared to PFS matched in age, gender, handedness and level of education (Cai et al, 2012). The magnitudes of the Up and Down F1 perturbations were 20%. Twenty-one PWS and 18 PFS participated in the study. Both groups showed significant F1 changes that

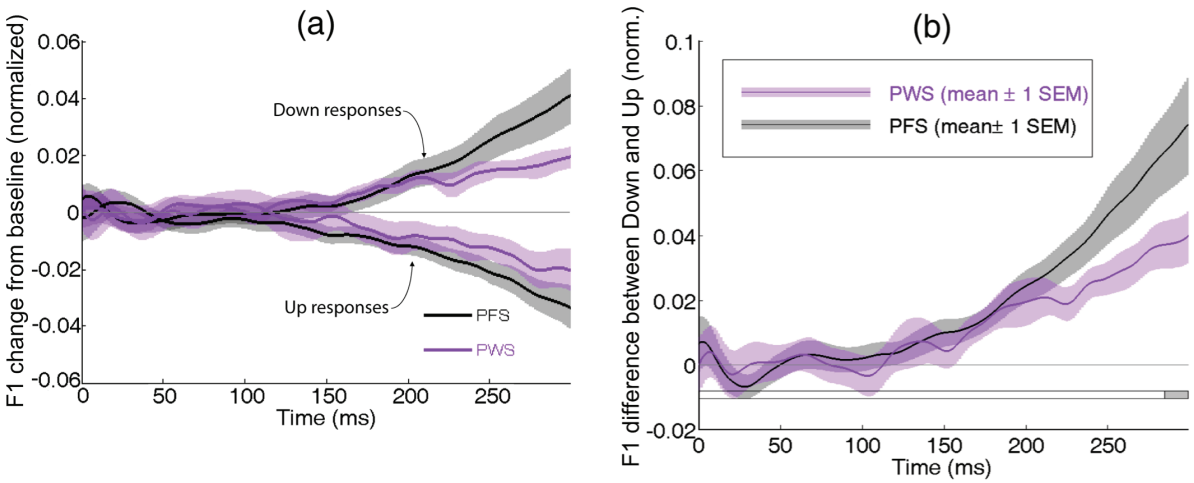

Figure 1. Comparison of the online compensatory responses to the perturbation of F1 during the monophthong [ɛ] by PWS and PFS. **(a)** Top two curves: mean differences between the F1 trajectories produced under the Down condition and the no-perturbation (noPert) baseline condition. Bottom two curves: mean differences between the F1 trajectories produced under the Up and noPert conditions. Time 0 corresponds to vowel onset. **(b)** Composite response curves: the contrasts between the Down and Up responses (as shown in Panel a) in the two groups. The gray area in the horizontal bar near the bottom of this panel shows time intervals in which the difference between the PWS and PFS reached significance at uncorrected $p<0.05$ (Adapted from Cai et al., 2012 with permission).

counteracted the perturbations (Figure 1a) with an onset latency of approximately 150 ms. However, the mean compensation magnitude of the PWS group was substantially lower (47%) than that of the PFS group, and this difference reached statistical significance at approximately 300 ms following perturbation onset ($p<0.05$, Figure 1b).

The F1 perturbation used by Cai et al. was similar to that of Tourville et al., but it was induced using a software solution, *Audapter*, that was more flexible and more robust than the hardware-based solution of the earlier study. The Audapter system was developed at MIT Speech Communication Group and Boston University Speech Lab for configurable, short-latency, on-line manipulation of AF of speech. The initial design of Audapter focused on perturbing formant frequencies during production of single words or pseudowords (Boucek, 2007; Cai et al., 2008). As diagrammed in Figure 2a, Audapter reads digitized microphone signals at a sampling rate of 12 kHz and a frame rate of 750 frames/s. It performs online linear prediction coding analysis (Markel and Gray, 1976), followed by a dynamic programming-based formant-tracking algorithm (Xia and Espy-Wilson, 2000) that estimates the first three formant frequencies. To alter the formant frequencies, infinite-impulse-response filters with zeros matching the estimated original formants and poles corresponding to the desired new formants are constructed and applied on the input waveform on the fly. The formant-shifted sounds are played back to the auditory system of the speaker through earphones or headphones with a latency of 10-20 ms. Formant perturbations during production of the phrase "head got bumped" are shown in Figure 2, Panels b and c. A simple upward F1 shift like that implemented by Cai et al. (2012) was applied to the word "head" (compare the dashed yellow and white lines) making it sound more like "had."

AF-Based Control of Dynamic Inter-segmental Formant Transitions

While it is informative to explore the auditory-motor interactions involved in achieving a static acoustic goal, isolated, intra-segmental AF perturbations fail to address the rapid transitions between sequentially ordered articulatory gestures with appropriate timing patterns. To explore this essential feature of speech production, Audapter was expanded to track the progression of multisyllabic and multiword utterances and to induce dynamic spatial and temporal AF perturbations on specific intervals within an utterance. Using these capabilities, we have found that the abnormality in the online auditory-motor interaction during speech articulation in PWS is not limited to the control of quasi-static articulation as described above. Twenty-nine PFS and 20 PWS participated in a separate perturbation experiment that involved manipulation of the timing of events in AF (Cai et al., 2011; Cai, 2012). Acceleration (Accel, e.g., Figure 3a) and Deceleration (Decel, e.g., Figure 3b) perturbations led to advancement and delay of the F2 minimum during the [u] sound in the word "owe", respectively, as the subjects produced the multiword utterance "I owe you a yo-yo." The PWS showed very low frequencies of stuttering due to

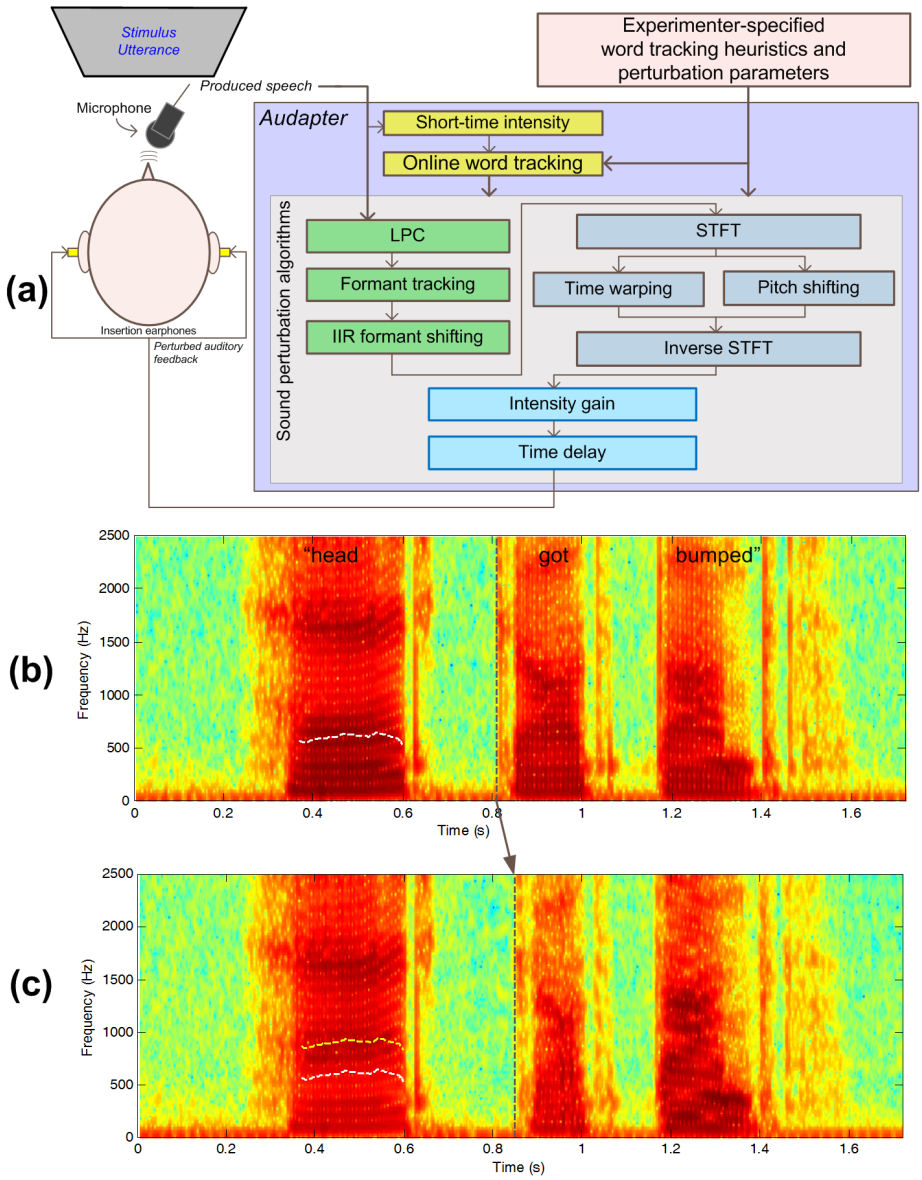

FIGURE 2. (a) A schematic diagram of the extended Audapter. **(b)** Spectrogram of the utterance "head got bumped," recorded from an adult male speaker inside a sound-attenuated booth. The dashed white curve shows the original F1 trajectory of the vowel in the first word, which can be compared with the perturbed F1 trajectory in Panel c. The vertical dashed line shows the timing of the onset stop consonant [g] in the second word, which can be compared with the perturbed timing of [g] in Panel c. **(c)** The perturbed version of the same utterance as in Panel b, generated in real time by Audapter. The vowel [ɛ] in the first word, "head", received a 250-Hz upward F1 shift (dashed yellow curve), which caused its AF to sound similar to [æ]. The original F1 trajectory (dashed white curve) identical to that shown in Panel b is overlaid for comparison. In the second word, local dynamic time warping imposed in the vicinity of the word onset caused the onset stop [g] to be delayed by approximately 40 ms. Notice that the timing perturbation was restricted to the early part of this word and did not affect other parts of the utterance. The third word, "bumped", received a combined upward pitch (+1.5 semitones) and intensity (+9 dB) shift. Abbreviations: IIR = infinite impulse response, STFT = short-time Fourier transform.

the repetitive nature of the experimental design. The time shifts introduced by the perturbations into the AF were similar between the two groups ($p>0.4$).

The PFS group showed an asymmetric pattern of compensation under the Accel and Decel perturbations: under the Decel perturbation, timing of the local F2 minimum during [u] in the word "owe" and the local F2 maximum during [j] in the following word "you" were both delayed significantly ($p<0.025$) in the subjects' productions

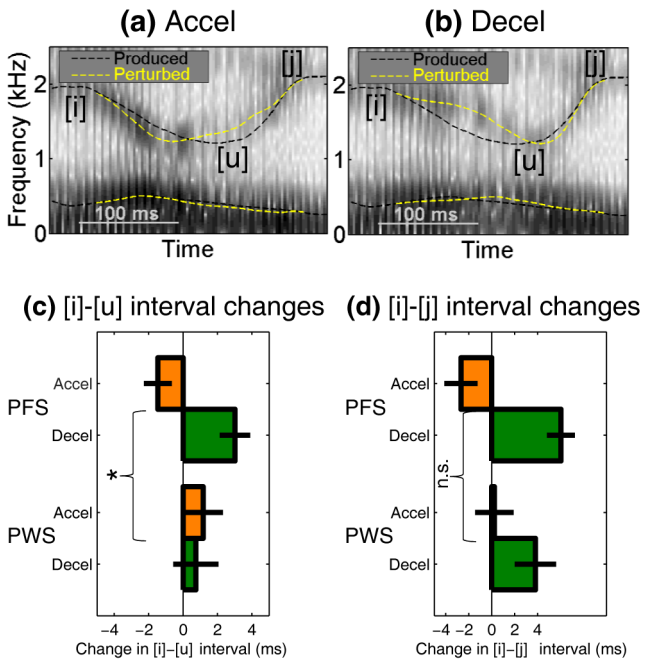

Figure 3. Online responses to the perturbation of temporal AF parameters in PWS and PFS. **(a)** An example of the Acceleration (Accel) perturbation, which advanced the local F2 minimum during [u] in time. **(b)** An example of the Deceleration (Decel) perturbation, which delays timing of the F2 minimum. In Panels a and b, the local maxima and minima during the three sounds [i], [u] and [j] are labeled. **(c)** Changes in the [i]-[u] time interval in the subjects production under the Accel and Decel perturbations from the noPert baseline. The asterisks indicates a significant interaction between Group and Perturbation according to an mixed ANOVA as well as a significant post hoc comparison of the Accel-Decel contrast ($p<0.05$; see text for details), which indicates a weaker-than-normal online adjustment of timing by the PWS at this time interval. **(d)** Changes in the [i]-[j] interval, in the same format as Panel c. The Group×Perturbation interaction did not reach significance, which indicates a gradual "catch-up" of the PWS group's response with the normal response (Adapted from Cai et al., 2011 and Cai et al. 2012, with permissions). Abbreviations: n.s.: non-significant.

compared to the noPert baseline, whereas timing changes were much smaller and did not reach significance under the Accel perturbations. In comparison, the PWS group showed significantly weaker changes in the [i]-[u] interval than the PFS group. When the [i]-[u] interval change was analyzed as the dependent variable, a two-way mixed analysis of variance (ANOVA) indicated a significant interaction of the between-subject factor of Group ({PWS, PFS}) and the within-subject factor of Perturbation ({Accel, Decel}) ($F_{1,47}=7.19$; $p=0.010$). A *post hoc* between-group comparison of the Decel-Accel contrast of the [i]-[u] interval also reached significance ($p=0.010$, t-test, Figure 4c). However, when the timing of the onset of the following word [j] (i.e., the [i]-[j] interval) was used as the dependent variable, the Group×Perturbation interaction did not reach significance ($F_{1,47}=2.69$; $p=0.11$). In fact, the PWS group showed [i]-[j] interval changes that were more similar to the normal pattern compared to their [i]-[u] interval changes (Figure 4d), which demonstrated a trend of the PWS to "catch up" in timing correction magnitude at longer latencies following the perturbation. Based on these findings, we suggest that online control of the timing parameters of articulation is not completely dysfunctional in PWS, but is instead limited in operational speed, which may be related to well-known phenomena in stuttering such as the fluency-enhancing effect of speaking at slower rates.

ADDITIONAL CAPABILITIES OF THE AUDAPTER SYSTEM

The above-mentioned fine-scale manipulation of timing is based on time-varying perturbations of formant frequencies and hence can only be used on utterances with continuous voicing, such as the sentence used in Cai et al. (2011). To enable fine-scale temporal manipulations on more generic types of utterances, we incorporated a phase vocoder (Bernsee, 2005) into Audapter. The phase vocoder first applies a short-time Fourier transform (STFT) on the input speech signal and stores the frame-by-frame Fourier spectra in memory. Through linear interpolation across the spectral frames and inverse STFT re-synthesis, Audapter can achieve arbitrary dilation (deceleration) and

compression (acceleration) along the time axis to achieve any user-specified time-warping that does not violate causality. The second word in Panels b and c of Figure 2 shows an example of local time warping, which delayed the initial consonant [g] by approximately 40 ms but did not affect the timing of other words within the utterance. This new AF manipulation type can be used to study the role of AF in the online control of speech movements during utterances that contain any sound type (stops, fricatives, etc.) and to test the generalizability of the finding by Cai et al. (2011).

Another function of the newly incorporated phase vocoder in Audapter is pitch shifting. This is achieved by stretching and interpolating the STFT spectra along the frequency axis. Similar to the approach used in Patel et al. (2011), Audapter can apply the pitch perturbation to a specific part of a multiword utterance (e.g., the third word Figure 2, Panels b and c). This new technique will be useful for further investigations on the control of voicing and prosody based on AF during multiword connected speech.

In addition to the manipulations of formants, pitch and fine-scale timing, the Audapter can also perturb several additional speech parameters, including the overall timing (delay), as well as the overall or local intensity. Further, these various types of perturbations can be combined on a single utterance. Panels b and c of Figure 2 demonstrate combinatorial application of the multiple AF manipulations: a 250-Hz F1 perturbation is imposed on the first word of the utterance, the onset stop consonant of the second word is be delayed by 40 ms, followed by an upward pitch shift and increased intensity on the third. Such flexibility is made possible by Audapter's configurable online heuristics based on short-time intensity and spectrum analysis to track the progress of articulation during an utterance, and control of the onset and termination of the various types of perturbations.

CONCLUSIONS

We have demonstrated that AF-based control of speech by PWS differs from that of fluent speakers. Compensatory responses to static intra-segmental F1 perturbations were smaller in PWS and responses to inter-segmental temporal perturbations were slower. These findings indicate that auditory-motor integration is anomalous in PWS. Future work is required to understand the relationship between auditory-motor integration deficits and the core symptoms of the PDS. Determining the neural substrates of these behavioral differences could clarify whether the deficit seen in PWS is related to encoding a sensory error, translating the sensory error into a corrective motor command, or in implementing the motor command.

We have also described the Audapter AF manipulation system, which is now capable of perturbing several key acoustic parameters of speech, including formant frequencies, intensity, and pitch, and the overall delay and fine-scale timing of AF. It can be used not only on isolated speech sounds or monosyllabic words, but also on connected multiword utterances. To the best of our knowledge, with these extensions, Audapter stands as the most comprehensive and flexible tool for manipulating speech AF. We believe Audapter is a powerful and valuable tool for speech scientists who work on deepening our understanding of auditory-motor interactions in speech production beyond simple parameters (e.g., prosody, accent and fluency) that will facilitate research on the roles of AF in the speech motor system in both its normal (e.g., Cai et al., 2010; Cai et al., 2011) and disordered states (e.g., Cai et al., 2012).

ACKNOWLEDGMENTS

This work was supported by NIDCD grants R01-DC007683 and R01-DC002852 (PI: Frank Guenther)

REFERENCES

Bauer, J. J., Mittal, J., Larson, C. R., and Hain, T. C. **(2006)**. "Vocal responses to unanticipated perturbations in voice loudness feedback: an automatic mechanism for stabilizing voice amplitude," J. Acoust. Soc. Am. **119**, 2363-2371.

Bloodstein, O., and Ratner. N.B. **(2008)**. *A handbook on stuttering* (Thomson/Delmar Learning, Clifton Park, NY).

Bernsee, S.M. **(2005)**. "Time Stretching And Pitch Shifting of Audio Signals – An Overview," http://www.dspdimension.com/admin/time-pitch-overview/ (Accessed Jan. 12, 2013).

Boucek, M. **(2007)**. "The nature of planned acoustic trajectories", *Unpublished master's thesis.* Karlsruhe Institute of Technology, Karlsruhe, Baden-Württemberg, Germany.

Braun, A.R., Varga, M., Stager, S., Schulz, G., Selbie, S., Maisog, J.M., Carson, R.E., Ludlow, C.L. **(1997)**. **"**Altered patterns of cerebral activity during speech and language production in developmental stuttering. An H2(15)O positron emission tomography study," Brain **120,** 761-784.

Cai, S., Boucek, M., Ghosh, S.S., Guenther, F.H., and Perkell, J.S. **(2008)**. "A system for online dynamic perturbation of formant frequencies and results from perturbation of the Mandarin triphthong /iau/," *8th Intl Seminar on Speech Production* (Strasbourg, France), pp. 65-68.

Cai, S., Ghosh, S.S., Guenther, F.H., and Perkell, J.S. **(2010)**. "Adaptive auditory feedback control of the production of the formant trajectories in the Mandarin triphthong /iau/ and its patterns of generalization," J. Acoust. Soc. Am. **128**, 2033-2048.

Cai, S., Ghosh, S.S., Guenther, F.H., and Perkell, J.S. **(2011)**. "Focal manipulations of formant trajectories reveal a role of auditory feedback in the online control of both within-syllable and between-syllable speech timing," J. Neurosci. **31**, 16483-16490.

Cai, S. **(2012)**. *Online Control of Articulation Based on Auditory Feedback in Normal Speech and Stuttering: Behavioral and Modeling Studies*. Ph.D. Dissertation, Harvard-MIT Division of Health Science and Technology, Massachusetts Institute of Technology, Cambridge, MA, USA.

Cai, S., Beal, D.S., Ghosh, S.S., Tiede, M.K., Guenther, F.H., and Perkell, J.S. **(2012)**. "Weak responses to auditory feedback perturbation during articulation in persons who stutter: Evidence for abnormal auditory-motor transformation," PLoS ONE. **7**(7), 41830.

Conture, E.G., and Brayton, E.R. **(1975)**. "Influence of noise on stutterers different disfluency types," J. Speech Hear. Res. **18**, 381-384.

De Nil, L.F., Kroll, R.M., Kapur, S., Houle, S. **(2000)**. "A positron emission tomography study of silent and oral single word reading in stuttering and nonstuttering adults," J. Speech Lang. Hear. R. **43**, 1038-1053.

Heinks-Maldonado, T. H., and Houde, J. F. **(2005)**. "Compensatory responses to brief perturbations of speech amplitude," Acoust. Res. Lett. Onl. **6**, 131-137.

Hillenbrand, J.M., Clark, M.J. and Nearey, T.M. **(2001)**. "Effects of consonant environment on vowel formant patterns," J. Acoust. Soc. Am. **109**, 748-763.

Houde, J. F., and Jordan, M. I. **(2002)**. "Sensorimotor adaptation of speech I: Compensation and adaptation," J. Speech Lang. Hear. R. **45**, 295-310.

Ingham, R.J., Moglia, R.A., Frank, P., Ingham, J.C., and Cordes, A.K. **(1997)**. "Experimental investigation of the effects of frequency-altered auditory feedback on the speech of adults who stutter," J. Speech Lang. Hear. R. **40**, 361-372.

Jones, J. A., and Munhall, K. G. **(2005)**. "Remapping auditory-motor representations in voice production," Curr. Biol. **15**, 1768-1772.

Kalinowski, J., Armson, J., Roland-Mieszkowski, M., Stuart, A., and Gracco, V.L. **(1993).** "Effects of alterations in auditory feedback and speech rate on stuttering frequency," Lang. Speech **36**, 1-16.

Lane, H., and Tranel, B. **(1971)**. "The Lombard sign and the role of hearing in speech," J. Speech Lang. Hear. R. **14**, 677-709.

Lane, H., and Webster, J.W. **(1991)**. "Speech deterioration in postlingually deafened adults," J. Acoust Soc. Am. **89**, 859-866.

Larson, C. R., Burnett, T. A., Kiran, S., and Hain, T. C. **(2000)**. Effects of pitch-shift velocity on voice F0 responses. J. Acoust. Soc. Am. **107**, 559-564.

Lee, B.S. **(1951).** "Artificial stutter," J. Speech Disord. **16**, 53-55.

Markel, J.D., and Gray, A.H. **(1976)**. *Linear Prediction of Speech* (Springer-Verlag).

McLeod, S., and Bleile, K. **(2003)** "Neurological and developmental foundations of speech acquisition," *Program of the American Speech-Language-Hearing Association Convention* (Chicago, Illinois).

Oller, D.K., and Eilers, R.E. **(1988)**. "The role of audition in infant babbling," Child Dev. **59**, 441-449.

Patel, R., Niziolek, C., Reilly, K., and Guenther, F.H. **(2011)**. "Prosodic adaptations to pitch perturbation in running speech," J. Speech Lang. Hear. R. **54**, 1051.

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.

Stager, S.V., Denman, D.W., and Ludlow, C.L. **(1997)**. "Modifications in aerodynamic variables by persons who stutter under fluency-evoking conditions," J. Speech Lang. Hear. R. **40**, 832-847.

Stuart, A., Kalinowski, J., Rastatter, M. P., and Lynch, K. **(2002)**. "Effect of delayed auditory feedback on normal speakers at two speech rates," J. Acoust. Soc. Am. **111**, 2237-2241.

Summers, W. V., Pisoni, D. B., Bernacki, R. H., Pedlow, R. I., and Stokes, M. A. **(1988)**. "Effects of noise on speech production: acoustic and perceptual analyses." J. Acoust. Soc. Am. **84**, 917-928.

Sutton, C., and Chase, R. **(1961)**. "White noise and stuttering," J. Speech Hear. Disord. **4**, 72.

Tourville, J.A., Reilly, K.J., and Guenther, F.H. **(2008)**. "Neural mechanisms underlying auditory feedback control of speech," Neuroimage **39**, 1429-1443.

Tsao, Y.C., and Weismer, G. **(1997)**. "Interspeaker variation in habitual speaking rate: evidence for a neuromuscular component," J. Speech Lang. Hear. R. **40**, 858-866.

Villacorta, V. M., Perkell, J. S., and Guenther, F. H. **(2007)**. Sensorimotor adaptation to feedback perturbations on vowel acoustics and its relation to perception. J. Acoust. Soc. Am. **122**, 2306-2319.

Waldstein, R.S. **(1990)**. "Effects of postlingual deafness on speech production - Implications for the role of auditory feedback." J. Acoust. Soc. Am. **88**, 2099-2114.

Webster, R.L., Schumacher, S.J., and Lubker, B.B. **(1970)**. "Changes in stuttering frequency as a function of various intervals of delayed auditory feedback," J. Abnorm. Psychol. **75**, 45-49.

Xia, K., and Espy-Wilson, C.Y. **(2000)**. "A new strategy of formant tracking based on dynamic programming," *Proceedings of the 6th International Conference on Spoken Language Processing* (ICSLP2000, Beijing, China). **3,** 55-58.
Yates, A. J. **(1963)**. "Delayed auditory feedback," Psychol. Bull, **60**, 213-232.
Zemlin, W.R. **(1998)**. *Speech and hearing science anatomy and physiology* (Allyn and Bacon, Needham Heights, Massachusetts).