Stuttering adults’ lack of pre-speech auditory modulation normalizes when speaking with delayed auditory feedback

Ayoub Daliri
Arizona State University, Tempe, AZ, United States

Ludo Max
University of Washington, Seattle, WA, United States
Haskins Laboratories, New Haven, CT, United States

Correspondence should be addressed to Ayoub Daliri, Ph.D., Department of Speech and Hearing Science, College of Health Solutions, Arizona State University, Coor Hall 3456, 975 S Myrtle Ave, Tempe, AZ 85287-0102, USA. E-mail: Ayoub.Daliri@asu.edu
Abstract

Auditory modulation during speech movement planning is limited in adults who stutter (AWS), but the functional relevance of the phenomenon itself remains unknown. We investigated for AWS and adults who do not stutter (AWNS) (a) a potential relationship between pre-speech auditory modulation and auditory feedback contributions to speech motor learning and (b) the effect on pre-speech auditory modulation of real-time vs. delayed auditory feedback. Experiment I used a sensorimotor adaptation paradigm to estimate auditory-motor speech learning. Using acoustic speech recordings, we quantified subjects’ formant frequency adjustments across trials when continually exposed to formant-shifted auditory feedback. In Experiment II, we used electroencephalography to determine the same subjects’ extent of pre-speech auditory modulation (reductions in auditory evoked potential N1 amplitude) when probe tones were delivered prior to speaking vs. not speaking. To manipulate subjects’ ability to monitor real-time feedback, we included speaking conditions with non-altered auditory feedback (NAF) and delayed auditory feedback (DAF). Experiment I showed that auditory-motor learning was limited for AWS vs. AWNS, and the extent of learning was negatively correlated with stuttering frequency. Experiment II yielded several key findings: (a) our prior finding of limited pre-speech auditory modulation in AWS was replicated; (b) DAF caused a decrease in auditory modulation for most AWNS but an increase for most AWS; and (c) for AWS, the amount of auditory modulation when speaking with DAF was positively correlated with stuttering frequency. Lastly, AWNS showed no correlation between pre-speech auditory modulation (Experiment II) and extent of auditory-motor learning (Experiment I) whereas AWS showed a negative correlation between these measures. Thus, findings suggest that AWS show deficits in both pre-speech auditory modulation and auditory-motor learning; however, limited pre-speech modulation is not directly related to limited auditory-motor adaptation; and in AWS, DAF paradoxically tends to normalize their otherwise limited pre-speech auditory modulation.
Keywords: Stuttering, Auditory modulation, Speech planning, Event-related potentials
**Introduction**

Structural neuroimaging studies of stuttering have revealed abnormalities in various fronto-parieto-temporal pathways (Chang, Erickson, Ambrose, Hasegawa-Johnson, & Ludlow, 2008; Chang, Horwitz, Ostuni, Reynolds, & Ludlow, 2011; Cykowski, Fox, Ingham, Ingham, & Robin, 2010; Kronfeld-Duenias, Amir, Ezrati-Vinacour, Civier, & Ben-Shachar, 2016; Sommer, Koch, Paulus, Weiller, & Buchel, 2002; Watkins, Smith, Davis, & Howell, 2008). Although the exact assignment of functional processes to the affected tracts is still open to debate (see for example Kronfeld-Duenias, Amir, Ezrati-Vinacour, Civier, & Ben-Shachar, 2017; Neef, Anwander, & Friederici, 2017), it appears that at least some of the pathways may impair the connectivity between speech (pre-)motor regions and auditory regions. This interpretation is consistent with behavioral studies demonstrating that stuttering individuals, as compared with nonstuttering individuals, show reduced compensatory motor responses to unexpected auditory feedback perturbations (Cai et al., 2012; Cai, Beal, Ghosh, Guenther, & Perkell, 2014; Daliri, Wieland, Cai, Guenther, & Chang, 2017; Loucks, Chon, & Han, 2012). Together, these lines of neuroimaging and behavioral research provide substantial support for the idea that stuttering is associated with deficits in the integration of auditory and motor information for speech production.

One aspect of sensorimotor integration involves predicting the sensory consequences of planned control signals (Kawato, 1999; Shadmehr, Smith, & Krakauer, 2010; Wolpert, Diedrichsen, & Flanagan, 2011; Wolpert, Miall, & Kawato, 1998). Predictions are used to optimize motor commands and to prepare task-relevant sensory systems for their subsequent roles in processing the sensory outcomes of those motor commands. In this context, we have recently investigated in adults who stutter (AWS) and adults who do not stutter (AWNS) a modulation of the auditory system that occurs during speech movement planning prior to the
initiation of muscle activity (Daliri & Max, 2015b). We recorded long-latency auditory evoked potentials (LLAEPs) in response to probe tones delivered during the speech movement planning phase as compared with no-speaking control conditions. Nonstuttering speakers showed a clear modulation of the auditory system (i.e., reduced amplitude of the N1 component in the LLAEP) during speech movement planning whereas stuttering speakers showed limited or no modulation of their auditory system (Daliri & Max, 2015b). These findings (later replicated, extended, and further interpreted in (Daliri & Max, 2015a) strongly suggest that stuttering is associated with deficits in using sensory prediction mechanisms to prime the auditory system for its role in processing upcoming auditory feedback. In other words, it is possible that neural activity in stuttering speakers’ pre-motor and/or motor systems involved in speech movement planning does not result in a correct, or complete, pre-setting of the auditory system for feedback monitoring once execution actually starts. As we have hypothesized elsewhere (Daliri & Max, 2015b; Max, 2004; but see also McClean, 1996; Zimmermann, 1980 for related hypotheses), an insufficient pre-movement priming of sensory systems (to ensure appropriate responses to self-generated sensory input) may trigger interfering motor responses that disrupt, rather than correctly update, ongoing movements.

However, the functional relevance of the pre-speech auditory modulation phenomenon itself is unknown at this time. It is particularly noteworthy that a reduction in the amplitude of a given LLAEP component (i.e., the scalp-recorded sum of electrical potentials from multiple intracranial sources and processes) does not necessarily indicate an actual suppression of the cognitive processes of interest (see, among others, Luck, 2014). Hence, it remains to be determined whether the neurophysiological processes underlying pre-speech auditory modulation serve to merely implement a *suppression* of auditory input (leading to a reduction in this sensory
channel’s contributions to error correction) vs. a more complex *optimization* of auditory processing (leading to precise contributions to error correction based on, for example, an adjustment in the excitation-suppression balance of auditory neurons with specific tuning characteristics—see Eliades & Wang, 2017). This is an absolutely critical question to start addressing if we are to understand if and how stuttering individuals’ lack of pre-speech auditory modulation relates to their demonstrated deficiencies in using the subsequent auditory feedback signal during speech production (Cai et al., 2012, 2014; Daliri et al., 2017; Loucks et al., 2012).

Thus, one purpose of the present study was to examine the relationship between individuals’ pre-speech auditory modulation and auditory feedback contributions to auditory-motor speech motor learning. In a first experiment, we used a sensorimotor adaptation paradigm to quantify, on an individual basis, the contributions of auditory feedback to speech motor learning. In this paradigm, the speakers’ auditory feedback was digitally altered in real time (i.e., the formant frequencies in the heard speech signal were shifted up during all trials in the perturbation phase of the experiment) while they produced monosyllabic words. We used the extent of auditory-motor adaptation (i.e., the speakers’ lowering of formant frequencies in their produced speech) as an estimate of learning. In a second experiment, we then recorded the same speakers’ LLAEPs in response to probe tones presented during speech movement planning in three conditions: (a) a speaking condition with typical auditory feedback (non-altered auditory feedback, NAF), (b) a speaking condition in which the usefulness of the auditory feedback signal for online monitoring was experimentally degraded (delayed auditory feedback, DAF), and (c) a no-speaking control condition (silent reading). DAF was included to provide additional insights into the possible relationship between pre-speech auditory modulation and contributions of the auditory feedback signal in speech production: we hypothesized that if this type of auditory
modulation serves to optimize auditory processing for feedback monitoring, then reducing the usefulness of auditory feedback would lead to a reduction in pre-speech auditory modulation.

**Experiment I**

**Participants**

After giving informed consent, 13 stuttering adults (11 males and 2 females; age 19–43 years, M = 28.71 years, SD = 7.96) and 13 nonstuttering adults (11 males and 2 females; age 20–45 years, M = 28.43 years, SD = 8.59) participated. All participants were native speakers of American English, right-handed (based on self-report), and naive to the purpose of the study. Nonstuttering participants were individually matched with the stuttering participants based on age (± 3 years) and sex. Eligibility criteria for all participants included (a) no history of psychological, neurological, or communication disorders (other than stuttering in the stuttering group), (b) not taking medications with a possible effect on sensorimotor functioning, and (c) pure-tone hearing thresholds of 20 dB HL or less at all octave frequencies 250–4000 Hz in both ears.

An American Speech-Language-Hearing Association (ASHA)-certified speech-language pathologist or a second-year speech-language-pathology graduate student with training and experience in the evaluation of stuttering confirmed the diagnosis of stuttering. All stuttering participants reported that their stuttering started during childhood (< 8 years of age). We used the Stuttering Severity Instrument, Fourth Edition (SSI–4; Riley, 2008) to determine each stuttering participant’s severity. Individual stuttering participant’s age, sex, SSI score, SSI stuttering severity classification, stuttering frequency (i.e., average frequency of stuttering across the SSI speaking and reading tasks) are listed in Table 1.
<table>
<thead>
<tr>
<th>Stuttering Participant</th>
<th>Age</th>
<th>Sex</th>
<th>SSI score</th>
<th>SSI severity</th>
<th>Stuttering frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1</td>
<td>38</td>
<td>male</td>
<td>16</td>
<td>very mild</td>
<td>1.77</td>
</tr>
<tr>
<td>S2</td>
<td>26</td>
<td>male</td>
<td>12</td>
<td>very mild</td>
<td>1.85</td>
</tr>
<tr>
<td>S3</td>
<td>22</td>
<td>male</td>
<td>20</td>
<td>mild</td>
<td>2.36</td>
</tr>
<tr>
<td>S4</td>
<td>41</td>
<td>male</td>
<td>15</td>
<td>very mild</td>
<td>3.51</td>
</tr>
<tr>
<td>S5</td>
<td>26</td>
<td>Male</td>
<td>27</td>
<td>moderate</td>
<td>5.29</td>
</tr>
<tr>
<td>S6</td>
<td>34</td>
<td>Male</td>
<td>17</td>
<td>very mild</td>
<td>5.50</td>
</tr>
<tr>
<td>S7(^a)</td>
<td>22</td>
<td>Female</td>
<td>29</td>
<td>moderate</td>
<td>5.67</td>
</tr>
<tr>
<td>S8</td>
<td>19</td>
<td>Female</td>
<td>29</td>
<td>moderate</td>
<td>6.59</td>
</tr>
<tr>
<td>S9</td>
<td>43</td>
<td>Male</td>
<td>28</td>
<td>moderate</td>
<td>7.75</td>
</tr>
<tr>
<td>S10</td>
<td>20</td>
<td>Male</td>
<td>21</td>
<td>mild</td>
<td>7.82</td>
</tr>
<tr>
<td>S11</td>
<td>28</td>
<td>Male</td>
<td>32</td>
<td>Severe</td>
<td>8.52</td>
</tr>
<tr>
<td>S12</td>
<td>32</td>
<td>Male</td>
<td>28</td>
<td>moderate</td>
<td>10.38</td>
</tr>
<tr>
<td>S13</td>
<td>22</td>
<td>Male</td>
<td>45</td>
<td>very severe</td>
<td>20.96</td>
</tr>
</tbody>
</table>

\(^a\) For participant S7, no recording of the reading task was available; S7’s SSI score, SSI severity, and stuttering frequency are based on the speaking task only.

Table 1. Individual participant information for the stuttering group (age, sex, SSI score, SSI severity, and stuttering frequency). Participants are rank-ordered based on stuttering frequency.

**Apparatus**

The experimental set-up is illustrated in Figure 1A. Participants were seated on a chair inside a sound-attenuated booth, approximately 1.5 m away from a computer monitor. The monitor was used to present target words and to provide visual feedback about participant’s
speech intensity. A microphone (SM58, Shure Incorporated, Niles, IL) was placed approximately 15 cm away from participants’ mouth to transduce speech output. After amplification (DPS II, ART ProAudio, Niagara Falls, NY), the microphone signal was routed through a speech processor (VoiceOne, TC-Helicon, Victoria, BC, Canada). The speech processor was used to alter all formant frequencies of the speech signal in real time (10 ms delay) during some trials in the experiment (see Procedure). The output of the speech processor was then amplified (S-phone, Samson Technologies Corp., Syosset, NY) and played-back to the participant through insert earphones (ER-3A, Etymotic Research Inc., Grove Village, IL). A custom-written MATLAB script (The MathWorks, Inc., Natick, MA) was used to visually present target words and color-coded feedback about speech intensity (see Procedure), to control the speech processor, and to record the microphone signal and the speech processor’s output at a sampling rate of 44100 Hz (16 bit).
Fig. 1. Instrumentation (A), timeline (B), and feedback manipulation (C, D) for Experiment I. The microphone signal was routed through signal processing equipment before being played back through insert earphones. After a baseline phase with unaltered auditory feedback, subjects heard feedback with all formant frequencies shifted upward by 250 cents during the adaptation phase. Unaltered feedback was restored for the after-effect phase. In C and D, solid blue lines indicate the original unaltered spectrum and first two formants for the subject’s speech whereas dashed blue lines indicate the same information for the real-time auditory feedback signal. The box in the center of the spectrogram (D) indicates the time window in
which the subject’s formant frequencies were measured (i.e., from 40% to 60% into the
duration of the vowel).

Before each recording session, the amplification level of the microphone and headphone
amplifiers were calibrated such that speech input with a sound level of 68 dB SPL at the
microphone resulted in 73 dB SPL output in the earphones. For calibration, the intensity of the
auditory feedback in the earphones was measured using a 2 cc coupler (Type 4946, Brue &
Kjaer Inc., Norcross, GA) connected to a sound level meter (Type 2250 A Hand Held Analyzer
with Type 4947 ½” Pressure Field Microphone, Brue & Kjaer Inc., Norcross, GA).

Procedure

The experiment consisted of two sessions separated by a few minutes of rest. In each
session, participants completed 60 blocks of 3 trials. Each trial started with the presentation of a
target word (“bed,” “bud,” or “baud”; phonetically: /bɛd/, /bʌd/, /bɔd/) on the monitor for 2.5 s.
The three target words were selected to (1) cover distinct regions of the acoustic vowel space by
including a front vowel, a central vowel, and a back vowel; (2) facilitate learning by providing a
large number of repetitions for a limited set of practice items; and (3) verify reproducibility by
maintaining consistency with the set of vowels used in our previously published auditory-motor
adaptation studies (e.g., Max & Maffett, 2015; Max, Wallace, & Gracco, 2003).

Presentation order of the three words within each block was randomized, but the same list
was used for all participants. Participants read the words aloud, and were given visual feedback
(a colored horizontal bar at the bottom of the screen) to assist with maintaining a consistent
output intensity (red for > 71 dB SPL, blue for < 65 dB SPL, and green for 65–71 dB SPL). To
familiarize participants with the task and the visual feedback, 10 practice blocks (30 trials) were
presented prior to the recording sessions.
Each session included three phases with different types of trials (Fig. 1B): (a) 10 blocks (i.e., 30 trials) in a baseline phase in which participants produced the target words with unaltered auditory feedback, (b) 35 blocks (i.e., 105 trials) in an adaptation phase in which participants produced the target words while hearing auditory feedback in which all formant frequencies were shifted 250 cents up (1 octave = 1200 cents) (Fig. 1C–D), and (c) 15 blocks (i.e., 45 trials) in an after-effect phase in which participants produced the target words again with unaltered auditory feedback.

**Data extraction and statistical analysis**

Prior to data analysis, the acoustic signals were down-sampled to 22050 Hz. Based on visual inspection of spectrograms generated in a custom-written MATLAB program, the onset and offset of each vowel were marked manually. The trajectories of the first two formant frequencies (F1, F2) were extracted automatically using formant tracking modules available in Praat (Boersma, 2002) and The Snack Sound Toolkit (Sjölander, 2004) with the same parameter values for formant tracking. Both these formant trajectories were overlaid on the spectrogram and visually inspected to select the trajectory with fewer tracking errors. In the case of tracking errors, a custom written MATLAB script was used to search for peaks in the spectrogram within 100 Hz from the selected formant track. The final formant trajectories were smoothed using a robust local regression algorithm, visually inspected, and if necessary, manually corrected based on the spectrogram. The steady-state first formant (F1) and second formant (F2) for each trial were calculated by averaging the values obtained in the time window 40–60 % into the vowel (Fig. 1D).

To normalize baseline phase formant frequencies across speakers, all formant values were converted from Hz to cents for each speaker separately, using the following formula:
$F_{cents} = 1200 \log_2 \left( \frac{F_{Hz}}{B_{Hz}} \right)$, where $F_{Hz}$ corresponds to the F1 or F2 value (in Hz) for a given vowel, and $B_{Hz}$ corresponds to the average formant values (in Hz) across the speaker’s productions of the same vowel in the baseline phase (i.e., averaged across 10 trials). Thus, a speaker’s extent of auditory-motor adaptation (i.e., change in speech output relative to baseline) was expressed in the same unit (i.e., cent) as the auditory feedback alteration implemented by the speech processor.

Given that preliminary analyses did not reveal a difference between results from the first and second sessions, individual participant data for a given formant and a given vowel were averaged across the two sessions. To obtain the adaptation and after-effect data for statistical analysis, we then averaged for each participant and each vowel the formant values from the last five trials of the adaptation phase and the formant values from the last five trials from the after-effect phase, respectively (the last five trials of each phase were selected to quantify performance after it reached a relatively stable level).

We used the IBM SPSS Statistics 19 software package (IBM, Armonk, NY) for all statistical analyses. First, we addressed our primary aims of determining whether or not adaptation and after-effect occurred within each subject group by testing whether there were statistically significant changes from baseline for each of the formants in the adaptation phase and the after-effect phase. This was accomplished with a priori planned comparisons using uncorrected one sample $t$-tests (note that the average baseline formant frequency is zero due to the conversion from Hz to cents) applied to the data from all three vowels combined. For each formant separately, a repeated measures analysis of variance (ANOVA) was carried for the purpose of determining whether Group (stuttering and nonstuttering) as a between-subjects factor interacted with Vowel (/ɛ/, /ʌ/, and /ɔ/) and Phase (adaptation and after-effect) as repeated
measures. Given that all formant measures were normalized based on average values from the baseline trials (yielding an average value of zero for the baseline trials), the baseline was not included in these analyses. To account for potential violations of the sphericity assumption in conditions with more than two levels, degrees of freedom were adjusted using the Greenhouse-Geisser correction (Max & Onghena, 1999). Normality of the data (in each condition and each group) was confirmed using quantile-quantile probability plots and one-sample Kolmogorov-Smirnov tests. Statistically significant interactions in the ANOVA were followed up with post-hoc analyses, using \(t\)-tests with Bonferroni corrections for multiple comparisons.

For the stuttering group, we also used a Pearson correlation analysis to explore the possible relationship between participants’ stuttering frequency (average percent stuttered syllables across the speaking and reading tasks of the SSI–4, see Table 1) and extent of auditory-motor adaptation. For this purpose, we quantified in a single, normalized index how much participants adapted their speech output when auditory feedback was experimentally manipulated. We calculated this overall adaptation index for each individual participant by first averaging formant values at the end of the adaptation phase (already expressed in cents relative to baseline) across vowels and across F1 and F2, and then dividing this number by -250 (as complete compensation for the auditory feedback formant manipulation would require -250 cents downward adaptation). Thus, for this adaptation index, a greater positive value indicates more adaptation.

**Results**

Figure 2A shows time series of both speaker groups’ average change in F1 (i.e., deviation from baseline in cents) throughout the experiment, and Figure 2B summarizes the descriptive F1 data obtained specifically from the adaptation and after-effect phases. In response to the 250-cent upward formant shift in the auditory feedback signal, the nonstuttering group’s F1 decreased in a
statistically significant manner in the adaptation phase \((t(12) = -6.584, p < .001)\), and remained lower than baseline in the after-effect phase \((t(12) = -4.444, p < .001)\). For the stuttering group, there was no statistically significant F1 decrease in either the adaptation phase \((t(12) = -1.636, p = .128)\) or the after-effect phase \((t(12) = -1.515, p = .156)\).

**Fig. 2.** Panels A (Formant 1) and C (Formant 2) show the stuttering (red) and nonstuttering (blue) groups’ average formant values (normalized to baseline and expressed in cents, then averaged across words with 3 different vowels) during successive phases of baseline with unaltered auditory feedback (trial blocks 1-10), adaptation to formant-shifted feedback (trial blocks 11-45), and after-effect with unaltered feedback (trial blocks 46-60). Shaded areas indicate the data at the end of adaptation and after-effect that were compared with baseline. Group mean data for trials in the shaded areas are shown in panels B and D for Formant 1 and Formant 2, respectively. Error bars correspond to standard errors. Whereas the nonstuttering group adapted and showed after-effect in both formants, the stuttering group did not significantly change their productions relative to baseline.
ANOVA results for the first formant showed that there were statistically significant main effects of Phase, $F(1, 24) = 12.630, p = .002,$ and Group, $F(1, 24) = 11.447, p = .002.$ However, these main effects were modified by a statistically significant Phase × Group interaction, $F(1, 24) = 7.749, p = .010.$ F1 for the nonstuttering group was lower at the end of the adaptation phase (i.e., where full adaptation has been reached) versus the end of the after-effect phase (i.e., where the initial after-effect have partially, but not completely, dissipated), $t(12) = -4.186, p = .001.$ F1 for the stuttering group did not differ between these two phases ($t(12) = -0.583, p = .571$) as no significant adaptation occurred ($t(12) = -1.636, p = .128$). No statistically significant main effect of Vowel ($F(1, 596, 38.315) = 0.225, p = .749$), two-way interactions (Vowel × Group , $F(1, 596, 38.315) = 1.369, p = .263$; and Phase × Vowel $F(1.763, 42.314) = 1.881, p = .169$), or three-way interactions (Phase × Vowel × Group) were found ($F(1.763, 42.314) = 1.267, p = .289$).

Time series of both groups’ overall change in F2 relative to the baseline are shown in Figure 2C, and the corresponding descriptive summary data are presented in Figure 2D. Similar to the situation for F1, planned comparisons for the nonstuttering individuals showed statistically significant F2 changes in both the adaptation phase ($t(12) = -7.304, p < .001$) and the after-effect phase ($t(12) = -6.092, p < .001$). For the stuttering individuals, on the other hand, F2 did not significantly change from baseline in the adaptation phase ($t(12) = -2.015, p = .060$) or the after-effect phase ($t(12) = -1.265, p = .230$).

Examining the F2 data with ANOVA, we again found a statistically significant main effect of Group, $F(1, 24) = 4.517, p = .044,$ but this time without a Group × Phase interaction ($F(1, 24) = 1.587, p = .220$): relative to baseline, F2 changed more for the nonstuttering group than the stuttering group. We also found a statistically significant main effect of Phase, $F(1,24) = 19.994, p < .001,$ as well as a significant Phase × Vowel interaction, $F(1.612, 38.683) = 4.219, p$
= .029. This interaction resulted from the fact that, in the adaptation phase, the F2 change was larger for vowel /ɔ/ than for vowel /ɛ/ (one-tailed t-test: \( t(25) = -2.648, p = .007 \)) and larger for vowel /ʌ/ than for vowel /ɛ/ (one-tailed t-test: \( t(25) = -2.272, p = .016 \)) whereas no significant difference among the vowels was found in the after-effect phase (/ɔ/ vs. /ɛ/: \( t(25) = 0.808, p = .426 \); /ɔ/ vs. /ʌ/: \( t(25) = 0.224, p = .824 \); /ɛ/ vs. /ʌ/: \( t(25) = 1.070, p = .295 \)). The main effect of Vowel approached significance \( (F(1.498, 35.958) = 3.391, p = .057) \). All two-way and three-way interactions not mentioned above were not statistically significant (Phase × Group: \( F(1, 24) = 1.587, p = .220 \); Vowel × Group: \( F(1.498, 35.958) = .479, p = .569 \); Phase × Vowel × Group: \( F(1.612, 38.683) = .851, p = .413 \)).

With regard to a possible relationship between auditory-motor adaptation during the adaptation experiment and stuttering frequency during a conversational speech sample recorded immediately prior to the experiment, we found a statistically significant negative correlation between the overall adaptation index and stuttering frequency \( (r = -.568, p = .043) \). Thus, participants with a higher stuttering frequency adapted less to the feedback alteration.

**Experiment II**

**Participants**

The same stuttering and individually matched nonstuttering individuals who participated in Experiment I were tested in a second experiment which used LLAEPs to quantify each individual’s extent of pre-speech auditory modulation. Stuttering participant S11 (together with the matched nonstuttering participant) was excluded from all analyses due to excessive noise in the electroencephalography (EEG) data that were recorded to derive LLAEPs. Four of the remaining 12 stuttering participants had previously participated in a similar but separate
recording session for our initial study on pre-speech auditory modulation in stuttering adults (Daliri & Max, 2015b).

**Apparatus**

Instrumentation was similar to that used in Experiment I with the exception that an EEG system was used to record brain responses to auditory stimulation. Participants were seated on a chair inside a sound-attenuated booth, approximately 1.5 m away from a computer monitor. A microphone (WL185, Shure Incorporated, Niles, IL) was placed 15 cm from the participant’s mouth to transduce speech output. The amplified microphone signal was routed to a speech processor (VoiceOne, TC-Helicon, Victoria, BC, Canada) which was used here only to delay the signal by 100 ms in some blocks of trials (see below). The output of the speech processor was then amplified and played-back to participants through insert earphones (ER-3A, Etymotic Research Inc., Grove Village, IL). The insert earphones were used to deliver binaural auditory stimuli (1 kHz, 40 ms duration, 10 ms rise/fall time, 75 dB SPL). Before each recording session, the microphone and headphone amplifiers were adjusted such that speech input with a sound level of 68 dB SPL at the microphone resulted in an output of 73 dB SPL in the earphones.

We used an Active Two BioSemi system (BioSemi, Amsterdam, The Netherlands) to record continuous electroencephalographic (EEG), electromyographic (EMG), and electrooculographic (EOG) activity. Although only a subset of 17 data channels was used for the present purposes (see below), EEG recordings were made from 128 sites on the scalp according to an extension of the international 10-10 electrode system (Gilmore, 1994; Oostenveld & Praamstra, 2001). In addition, the signals from two electrodes placed over the left and right mastoids were recorded for off-line calculation of an averaged mastoid reference. A bipolar EOG signal was recorded using two electrodes placed below the lower eyelid and next to the outer
canthus of the left eye for off-line rejection of blinks and eye movements. Monopolar surface EMG signals were recorded from four electrodes placed on the skin overlying perioral muscles (right and left orbicularis oris superior and inferior) in order to detect upper and lower lip elevation/rounding. The acoustic signal from an additional microphone (SM58, Shure, Niles, IL) was recorded together with the EEG, EOG, and EMG signals (sampling rate 1024 Hz for each channel).

**Procedure**

Experiment II was conducted either on the same day as Experiment I or on a separate day (typically 3-15 days apart). The overall design as well as the within-trial timing of visual and auditory stimuli had been validated in our previous studies (Daliri & Max, 2015b; Max, Daniels, Curet, & Cronin, 2008). Here, the experiment included three conditions (Fig. 3A–C) with each condition consisting of three blocks of 90 trials. The order of all nine blocks was randomized for each participant. In one condition, participants read individual target words aloud after a go-signal while hearing non-delayed auditory feedback (NAF condition). In a second condition, participants read the words aloud after the go-signal while hearing delayed auditory feedback (DAF condition, 100 ms delay). The third condition served as a control condition in which participants read the words silently and without any orofacial movements.
Fig. 3. Schematic representation of experimental conditions involving speaking with non-delayed auditory feedback (NAF) (A), speaking with delayed auditory feedback (DAF) (B), and silent reading (C). A probe auditory stimulus (1 kHz; 40 ms duration; 75 dB SPL) was played on 40% of trials (tone trials) for each condition (D). No auditory stimulus was played in the remaining trials (no-tone trials). Evoked potentials were derived for the tone trials and no-tone trials separately (E), each calculated as the average evoked potential across 17 central electrodes (F). For each subject and in each condition, the averaged signal for no-tone trials (reflecting non-auditory processes) was subtracted from the averaged signal for tone trials (reflecting both auditory and non-auditory processes). Isolating the auditory activity of interest, the resulting signal provides the best estimate of the long latency auditory evoked potential in response to the probe tones.
Each trial started with the presentation of a word in white characters on a black background. After 600 ms, the color of the characters changed to green. This change in color served as the go-signal in the two conditions that involved speaking. For each trial, the word was randomly selected from a list of 90 words that were all 3–5 letters long, monosyllabic, of consonant-vowel-consonant structure, with an initial bilabial consonant, and with no consonant clusters in either initial or final position (e.g., bat, pet, moon). As shown in Figure 3D, 400 ms after presentation of the word in white characters, an auditory stimulus (the aforementioned 1 kHz pure tone) was presented binaurally through the insert earphones in 40% of the trials within each block (tone trials) whereas no auditory stimulus was presented in the remaining trials (no-tone trials). The time interval from the end of a trial to the beginning of the next trial was randomly selected from a set of five possible intervals (1500, 2000, 2500, 3000, 3500 ms).

**Signal processing and data extraction**

Signal processing and data extraction were carried out using the EEGLAB toolbox (Delorme & Makeig, 2004) and custom-written MATLAB scripts. First, all EEG signals were re-referenced to an offline average of the mastoid signals and digitally low-pass filtered at 50 Hz using a finite impulse response (FIR) filter (Kaiser windowed sinc FIR filter; deviation: 0.005; transition bandwidth: 1 Hz). To maintain zero-phase shift filters were applied bidirectionally (forward and backward; Tanner, Morgan-Short, & Luck, 2015). These signals were then epoched into segments from 100 ms before to 400 ms after the onset of the auditory stimulus in tone trials, and the equivalent time window in no-tone trials. Epochs were baseline-corrected by subtracting the average amplitude during the pre-stimulus period (100 ms) from the entire epoch. Epochs in which the amplitude exceeded ±100 µV were excluded. Epochs were further visually inspected to exclude trials with artifacts associated with excessive muscle activity, muscle
activity before the go-signal, blinking, or eye movements. The number of artifact-free epochs was highly similar for the nonstuttering group (Speaking-NAF $M = 96.25$, $SD = 9.73$; Speaking-DAF $M = 102.50$, $SD = 4.48$; Silent reading $M = 104.42$, $SD = 2.81$) as compared with the stuttering group (Speaking-NAF $M = 98.92$, $SD = 10.35$; Speaking-DAF $M = 101.26$, $SD = 6.07$; Silent reading $M = 102.92$, $SD = 4.02$).

Next, for each subject in each condition, the artifact-free epochs for all tone trials and those for all no-tone trials were averaged separately. The averaged signal for tone trials reflected brain activity related to auditory processing (evoked by the auditory stimuli) as well as non-auditory processing (e.g., activity related to motor, linguistic, cognitive, and visual processes associated with the task). The averaged signal for no-tone trials reflected only brain activity related to the same non-auditory processes. Thus, to isolate the auditory activity of interest, the averaged signal for no-tone trials was subtracted from the averaged signal for tone trials (Fig. 3E; for a discussion see Luck, 2014). The resultant signal, which was used as each channel’s final LLAEP after additional digital low-pass filtering at 15 Hz (Kaiser windowed sinc FIR filter; deviation: 0.005; transition bandwidth: 1 Hz), provides the best estimate of the auditory activity only (Baess et al., 2008; Baess, Horvath, Jacobsen, & Schroeger, 2011; Daliri & Max, 2015b; Luck, 2014; Martikainen, Kaneko, & Hari, 2005).

Data extraction focused on pre-speech modulation of the N1 and P2 components in the LLAEP. Our previous studies have shown that the extent of such modulation is largest in the frontocentral region (Daliri & Max, 2015b). Thus, we included here 17 electrodes located in this region of interest (ROI; Fig. 3F). The LLAEPs from these 17 channels were averaged for each subject in each condition, and the amplitude and latency of the N1 and P2 components were extracted. N1 was defined as the largest negative peak between 70 and 130 ms after onset of the
auditory stimulus. P2 was defined as the largest positive peak between 150 and 250 ms after onset of the auditory stimulus.

**Statistical analyses**

For each dependent variable, we conducted a repeated measures ANOVA with Group (stuttering, nonstuttering) as a between-subjects variable and Condition (NAF, DAF, silent reading) as a within-subjects variable. Degrees of freedom for the within-subjects comparisons were adjusted with Greenhouse-Geisser corrections to account for potential violations of the sphericity assumption (Max & Onghena, 1999). Given the specific hypotheses under investigation, we also conducted *a priori* planned comparisons with uncorrected one-sample *t*-tests to examine whether or not there was significant N1 or P2 modulation (i.e., a deviation from zero for the difference in peak amplitude between speaking with either NAF or DAF versus silent reading) for each group separately. For the stuttering group, Pearson correlation coefficients were calculated to explore a possible relationship between participants’ stuttering frequency during a clinical assessment prior to the experiment and any of the dependent variables or the derived magnitude of N1 and P2 modulation. In addition, we used Pearson correlation coefficients also to examine a possible relationship between auditory-motor adaptation (using the adaptation index calculated in Experiment I) and pre-speech auditory modulation (using the extent of auditory modulation in Experiment II). All statistical analyses were performed with the IBM SPSS Statistics 19 software package.

**Results**

*Pre-speech auditory modulation*

Figure 4 shows the grand-average LLAEPs from all conditions for the nonstuttering group and the stuttering group (panels A and B, respectively). There were no statistically significant effects for N1 latency (Group: \(F(1.22) = 0.140, p = .712\); Condition: \(F(1.417, \ldots)\).
auditory modulation

$31.171) = 0.725, p = .447$; Group × Condition: $F(1.417, 31.171) = 0.271, p = .687$, but N1 amplitude was associated with a statistically significant main effect of Condition, $F(1.571, 34.568) = 28.594, p < .001$, and a statistically significant Group by Condition interaction, $F(1.571, 34.568) = 15.142, p < .001$. The main effect of Group was not statistically significant ($F(1,22) = .001, p = .992$). Thus, N1 amplitude differed across the NAF, DAF, and silent reading conditions, and the size of such differences varied between the stuttering and nonstuttering groups. These effects are graphically illustrated in panels 4C and 4D. Panel 4C shows the average N1 amplitude in each of the conditions for the two groups (negative voltages are down). Panel 4D shows the calculated amount of N1 modulation (i.e., relative to the silent reading condition) in the NAF and DAF speaking conditions for the nonstuttering and stuttering groups as well as individual subject changes in the extent of N1 modulation from the NAF to the DAF condition. Note that pre-speech modulation is shown as a positive value when the negative N1 amplitude is reduced prior to speaking relative to the control condition.

Most importantly, the planned pair-wise comparisons indicated that for the nonstuttering group the amount of N1 amplitude modulation relative to silent reading was statistically significant prior to both speaking with NAF, $t(11) = 13.405, p < .001$, and prior to speaking with DAF, $t(11) = 8.011, p < .001$. For the stuttering group, on the other hand, the same comparisons showed that statistically significant N1 modulation occurred only when speaking with DAF, $t(11) = 2.555, p = .027$, but not when speaking with NAF ($t(11) = -0.262, p = .798$). Given this finding that speaking with DAF caused stuttering individuals to start showing auditory modulation (i.e., increasing their otherwise very limited amount of modulation to similar levels as observed in the nonstuttering individuals), Figure 4E presents an additional illustration of the individual subject data in terms of direction and extent of change in N1 modulation when
AUDITORY MODULATION

speaking with DAF vs. NAF. The figure clearly shows a trend for the presence of delayed auditory feedback to decrease the amount of N1 modulation in nonstuttering speakers (8 of 12 subjects) but to increase it in stuttering speakers (9 of 12 subjects).

Fig. 4. Long latency auditory evoked potential for the nonstuttering group (A) and the stuttering group (B) prior to speaking with non-altered auditory feedback (NAF, green), prior to speaking with delayed auditory feedback (DAF, orange), and during silent reading (black). Group mean N1 amplitudes are shown in C (error bars correspond to standard errors). The corresponding amount of pre-speech N1 modulation (amplitude decrease prior to speaking vs.
silent reading) in both the NAF and DAF speaking condition is shown as box plots in C. The additional blue and red lines between the boxes indicate nonstuttering and stuttering individual subject changes in N1 modulation from the NAF to the DAF condition. In panel E, these individual subject changes in pre-speech N1 modulation when speaking with DAF are represented as decreases vs. increases as compared with the NAF condition. Stuttering subjects showed a lack of pre-speech N1 modulation in the NAF condition, but most increased the amount of modulation in the DAF condition. Most nonstuttering subjects decreased their N1 modulation when speaking with DAF.

Examining P2 amplitude, we again found a significant main effect of Condition, $F(1.990, 43.775) = 3.549, p = .037$, where P2 amplitude in the silent reading condition was marginally larger than P2 amplitudes in the NAF ($t(23) = 2.079, p = .049$, before Bonferroni correction) or DAF speaking conditions ($t(23) = 2.292, p = .032$, before Bonferroni correction). For this later auditory component neither the main effect of Group ($F(1,22) = 1.142, p = .297$) nor the Group × Condition interaction ($F(1.990, 43.775) = 2.563, p = .089$) were statistically significant. Given the absence of any interactions involving the Group factor, no pair-wise comparisons of conditions were conducted. P2 latency showed no statistically significant effects (Group: $F(1,22) = 0.163, p = .690$; Condition: $F(1.340, 29.486) = 1.867, p = .181$; Group × Condition: $F(1.340, 29.486) = 1.181, p = .764$).

For the stuttering group, correlation analyses were completed with stuttering frequency data from a preceding clinical assessment and the amount of pre-speech N1 amplitude modulation during the NAF and DAF speaking conditions (with higher positive values for more pre-speech modulation, as in Fig. 4D). There was no statistically significant correlation between stuttering frequency and amount of pre-speech auditory modulation when speaking with NAF ($r = .230, p = .472$), but there was a statistically significant correlation between these variables when speaking with DAF, $r = .655, p = .021$. Thus, when normal auditory feedback is replaced with a delayed feedback signal, individuals with more severe stuttering showed the most N1
modulation. Stuttering frequency was not significantly correlated with P2 modulation in either the NAF ($r = .184, p = .567$) or DAF condition ($r = -.089, p = .783$).

Relationship between auditory-motor adaptation (Experiment I) and pre-speech auditory modulation (Experiment II)

To determine the potential relationship between speech adjustments based on auditory feedback and pre-speech auditory modulation, we examined the correlation between the overall index of auditory-motor adaptation to formant-shifted feedback from Experiment I (as mentioned above, this index represented subjects’ extent of formant adaptation at the end of the perturbation phase averaged across F1 and F2 and across three target words) and the extent of auditory N1 modulation prior to speaking with NAF in Experiment II (given that the auditory-motor adaptation task was also completed with non-delayed feedback). For the nonstuttering group, the adaptation index was not statistically significantly correlated with the amount of N1 modulation ($r = .054, p = .869$); for the stuttering group, on the other hand, there was a statistically significant negative correlation between these two variables, $r = -0.655, p = 0.021$ (Fig. 5). Stuttering subjects who showed more pre-speech auditory modulation showed less auditory-motor adaptation (even though it should be noted that, in comparison, nonstuttering subjects showed more pre-speech auditory modulation but also more, rather than less, auditory-motor adaptation than the stuttering subjects). Neither group showed a significant correlation between the adaptation index and P2 modulation (nonstuttering group: $r = .296, p = .349$; stuttering group: $r = -.216, p = .501$).
Fig. 5. Relationship between the extent of pre-speech auditory N1 modulation when speaking with NAF and an index representing the extent of auditory-motor adaptation in a separate task on a different day. Stuttering subjects showed less N1 modulation than any of the nonstuttering subjects and a negative correlation between N1 modulation and adaptation. For the nonstuttering subjects, the two variables were uncorrelated.

General discussion

In previous work, we demonstrated that typical adults modulate aspects of auditory processing during the planning phase preceding speech movement onset but that this modulation is lacking in adults who stutter (Daliri & Max, 2015a, 2015b; Max et al., 2008). In the same previous publications, we also suggested a number of potential explanations for this lack of pre-speech auditory modulation in individuals who stutter (Daliri & Max, 2015a, 2015b). Nevertheless, even for nonstuttering speakers the functional relevance of this pre-speech auditory modulation remains unknown. We therefore investigated here a potential relationship between auditory modulation during speech planning and auditory feedback contributions to speech auditory-motor learning. As an additional exploration of the phenomenon’s functional role, we
studied the importance of being able to monitor auditory feedback in real time. For this purpose, we tested pre-speech auditory modulation both in speaking conditions with NAF and DAF.

Several important findings emerged from the separate experiments and their combined consideration. First, in agreement with previous studies showing reduced immediate compensation in response to unexpectedly altered auditory feedback in AWS vs. AWNS (Cai et al., 2012, 2014; Loucks et al., 2012), results from Experiment I show that auditory-motor learning with predictably altered feedback is also limited for AWS vs. AWNS. In fact, whereas the group of AWNS adapted to an extent that, across all three target words combined, compensated for ~28% (F2) to ~36% (F1) of the auditory perturbation (which shifted F1 and F2 by 250 cents), AWS showed only small and statistically nonsignificant changes in their acoustic speech output relative to the formant-shifted auditory feedback (~6–12%). Additionally, for the stuttering participants, the extent of this auditory-motor learning was negatively correlated with their stuttering frequency during a conversational speech sample. Hence, our results strongly suggest that during speech movement planning the central nervous system of AWS does not fully account for the auditory consequences experienced during preceding utterances. That is, when productions resulted in auditory feedback that was inconsistent with the speaker’s usual motor-to-auditory transformation (due to our experimental manipulation of the feedback signal), stuttering speakers differed from control speakers by showing much more limited adjustments of their acoustic output in subsequent productions, and this limitation was more severe in those individuals with more severe stuttering. To gain more insight into this sensorimotor limitation, future studies should aim to distinguish among several potential explanations. Some possibilities are that stuttering speakers’ reduced extent of auditory-motor adaptation is due to (a) inaccurate feedback processing within the auditory system itself, (b) a reduced reliance on auditory
feedback to determine production accuracy, (c) difficulties with the updating of internal representations (forward models) of the mapping between vocal tract configurations and their acoustic consequences, (d) differences in the specification of auditory targets during speech planning, or (e) a combination of two or more of these possibilities.

As a second important finding, Experiment II replicated the results from our previous studies (Daliri & Max, 2015a, 2015b) by showing again that modulation of the auditory response to probe tones presented during the planning phase prior to speech onset is also limited in AWS vs. AWNS. A major component of the rationale for the present work was to start determining the functional relevance of the pre-speech auditory modulation phenomenon. We had hypothesized that if the reduction in auditory N1 amplitude prior to speech onset represents a partial suppression, or attenuation, of the auditory response, then subjects who show more pre-speech modulation might rely less on auditory feedback to evaluate the accuracy of their productions and, thus, they can be expected to show less auditory-motor adaptation. If, on the other hand, the reduction in N1 amplitude does not reflect simply a global scaling down of the auditory response but rather a more purposeful and predictive “fine-tuning” of auditory processes involved in error monitoring then (a) subjects who show more pre-speech modulation might be expected to rely more on auditory feedback to evaluate produced utterances and, thus, show more auditory-motor adaptation, and (b) pre-speech auditory modulation can be expected to diminish in a DAF condition with a 100 ms feedback delay that disrupts normal feedback monitoring and auditory-motor adaptation (Max & Maffett, 2015).

The pattern of findings with regard to these issues—the relation of pre-speech auditory modulation to speech auditory-motor learning and the effect of DAF on such auditory modulation—was more complex than predicted by our hypotheses. AWNS showed no
statistically significant across-subjects correlation between the extent of pre-speech auditory modulation in Experiment II (NAF condition) and the extent of auditory-motor learning in Experiment I. The AWS group did show a correlation between these two measures, but the relationship was characterized by a negative correlation coefficient: AWS who modulated their auditory system more during speech planning, compensated less for the formant shift manipulation in the auditory-motor adaptation task. At first sight, the latter finding for AWS might seem consistent with the notion that pre-speech auditory modulation reflects a partial suppression of the neural response to auditory input in advance of speech production (and thus prior to receiving an auditory feedback signal). However, other aspects of the data strongly argue against this simple interpretation: (a) as mentioned, for nonstuttering speakers, the extent of pre-speech auditory modulation was not related to the extent of auditory-motor adaptation, and (b) the nonstuttering speakers as a group showed more pre-speech auditory modulation than the stuttering speakers but also more (rather than less) auditory-motor adaptation in the formant shift task. In other words, based on the group-level data alone, one could argue that the overall results are actually consistent with the hypothesis that auditory modulation during speech planning enhances feedback monitoring for auditory-motor learning.

Moreover, the notion that the central nervous system would start applying a general suppression of auditory pathways during the planning phase prior to speech onset is also difficult to reconcile with studies demonstrating that auditory feedback is, in fact, the primary modality driving speech motor learning (e.g., Feng et al., 2011). Thus, taken together, these present and prior findings make it more likely that pre-speech modulation does not reflect a global attenuation of auditory processing, but the phenomenon also appears not directly correlated—at least not within a group of typical speakers—to the auditory monitoring that contributes to
auditory-motor adaptation across multiple productions. It may be that the inter-individual variability in pre-speech auditory modulation (or limitations in the ability to quantify the phenomenon with EEG-based evoked potentials) prevents the detection of a participant-level correlation with auditory-motor learning while allowing the detection of a more global relationship at the group level (stuttering vs. nonstuttering speakers). As one alternative hypothesis, it is possible that pre-speech auditory modulation may serve a role in aspects of auditory monitoring that were not tested here (i.e., different from those sub-serving auditory-motor adaptation). For example, auditory monitoring also plays an important role in the online error evaluation processes that lead to immediate, within-trial articulatory compensation (for evidence of such online compensation see Kim and Max, 2014, for unperturbed productions and, for example, Niziolek and Guenther, 2013, for productions with altered auditory feedback). Another alternative hypothesis is that limited pre-speech auditory modulation may negatively affect speech motor learning only during childhood speech development. Auditory feedback is critical for speech acquisition during childhood (Guenther & Vladusich, 2012; Kuhl, 2004), and thus any limitation in neural processes sub-serving feedback monitoring may negatively impact fluent speech production. Lastly, it should be recalled that both our tasks involved fluent productions of monosyllabic words—i.e., subjects produced only simple utterances that are typically not associated with stuttering and that may be characterized by a rather low reliance on auditory feedback in general. Thus, additional insights might also be gained from examining auditory modulation prior to the production of longer utterances and, if possible, with both fluent and stuttered utterances. Clearly, numerous critical questions regarding the functional relevance and theoretical importance of pre-movement sensory modulation remain to be determined in future studies.
Nevertheless, as a third major finding of the present study, data from the different feedback conditions in Experiment II showed, for the first time, that producing speech with a 100 ms delay in the auditory feedback signal (DAF condition) caused a decrease in pre-speech auditory modulation for most AWNS (8 of 12 participants) but an increase in the same phenomenon for most AWS (9 of 12 participants). In fact, the stuttering group showed a statistically significant modulation of the auditory N1 response only before speaking with DAF. Thus, in AWS, the delayed feedback signal paradoxically tends to normalize their otherwise limited pre-speech auditory modulation. This is an intriguing finding, with potentially important implications, in light of the very extensive literature demonstrating that many AWS experience a decrease in the frequency of their stuttering symptoms when producing connected speech with such delayed auditory feedback (for a review, see Bloodstein and Bernstein-Ratner, 2008).

Whether the DAF-based increase in auditory modulation demonstrated here and the DAF-based fluency enhancements documented in previous studies are directly or indirectly related is yet another new question that warrants further exploration. It is certainly worth noting in this context that, in our group of AWS, the amount of auditory modulation when speaking with DAF was positively correlated with stuttering frequency during conversational speech, and that other researchers have argued that the fluency-enhancing benefits of DAF are also more substantial in individuals with more severe stuttering (see Foundas, Mock, Corey, Golob, & Conture, 2013; Lincoln, Packman, & Onslow, 2006).

One potential study limitation that deserves mentioning relates to the low ratio of female to male participants in each group. Although this ratio was kept low by design (the female to male ratio is also small in the overall population of individuals who stutter), it could be considered problematic in light of previous reports that DAF might have different effects on
male vs. female typical speakers (e.g., Bachrach, 1964; Corey & Cuddapah, 2008; Sutton, Roehrig, & Kramer, 1963). It should be noted, however, that those studies only examined DAF-induced increases in typical participant’s speech disfluencies and/or errors, and those findings may not generalize to electrophysiological brain recordings such as in the present study. In addition, there also have been several previous studies that failed to find such a DAF-based sex difference in speech disfluencies or errors (Chon, Kraft, Zhang, Loucks, & Ambrose, 2013; Sasisekaran, 2012; Stuart & Kalinowski, 2015; Van Borsel, Sunaert, & Engelen, 2005). Thus, the exact influence of DAF on male vs. female speakers remains unknown, and the experimental design of the present study also does not allow us to draw any conclusions in this regard. We can only state that in terms of both auditory-motor learning and pre-speech auditory modulation, the data from our stuttering and nonstuttering female participants fell within the same range as those of the male participants in the same group.

In sum, both pre-speech auditory modulation and auditory-motor learning are limited in adults who stutter, and the extent of auditory-motor learning is negatively correlated with stuttering frequency. Consistent with the hypothesis that pre-speech auditory modulation serves a role in feedback monitoring, speaking with delayed auditory feedback causes a decrease in modulation for most nonstuttering adults. However, speaking with delayed auditory feedback causes an increase in modulation for most stuttering adults, and there is a positive correlation between the amount of pre-speech auditory modulation and stuttering frequency. These findings contribute several new insights into the sensorimotor mechanisms underlying stuttering, and, importantly, they suggest several new lines of inquiry with theoretical as well as clinical implications.
Acknowledgements

This research was supported by grants R01 DC007603, R01 DC014510, and P30 DC004661 from the National Institute on Deafness and Other Communication Disorders and grant MOP-137001 from the Canadian Institutes of Health Research. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute on Deafness and Other Communication Disorders, the National Institutes of Health, or the Canadian Institutes of Health Research.
References


processing during the planning of speech movements. In *Proceedings of the 8th International Seminar on Speech Production* (pp. 41–44). Strasbourg, France.


Sasisekaran, J. (2012). Effects of delayed auditory feedback on speech kinematics in fluent


https://doi.org/10.1146/annurev-neuro-060909-153135


https://doi.org/10.1093/brain/awm241


Figure 2
Figure 3

A. Silent Reading & Movement Planning
   \text{word} \rightarrow \text{word} \rightarrow \text{Speaking With NAF}

B. Silent Reading & Movement Planning
   \text{word} \rightarrow \text{word} \rightarrow \text{Speaking With DAF}

C. Silent Reading
   \text{word} \rightarrow \text{word}

D. Silent Reading
   \text{word} \rightarrow \text{word}
   \text{Auditory Stimulus (40\% of Trials)}

E. \begin{align*}
   &10 \mu V \\
   &-100 \text{ ms} \rightarrow 400 \text{ ms} \rightarrow -10 \mu V \\
   \text{No-Tone Trials} & \quad \text{Tone Trials} \\
   \text{Tone Trials Minus No-Tone Trials}
\end{align*}

F. \text{Brain activity map}
Figure 4

A 12 µV

-100 ms - 400 ms

B 12 µV

-100 ms - 400 ms

C

N1 Amplitude (µV)

-12.0 - 0.0

Nonstuttering Stuttering

D

N1 Modulation (µV)

-2.0 - 4.0

Nonstuttering Stuttering

E

Change in N1 Modulation (µV)

-2.0 - 4.0

DAF Relative to NAF

Increase in N1 Modulation (DAF Relative to NAF)

Decrease in N1 Modulation (DAF Relative to NAF)
Figure 5

Nonstuttering\(r = 0.05, p = .87\)

Stuttering\(r = -0.65, p = .02\)