

1 Effects of Sensorimotor Voice Training on Event-Related Potentials to Pitch-Shifted Auditory
2 Feedback
3
4

5 Sona Patel^{1,2*}, Karen Hebert³, Oleg Korzyukov^{1,4}, Charles R. Larson¹
6
7

8 ¹ Department of Communication Sciences and Disorders, Northwestern University, Evanston, IL,
9 United States of America
10

11 ² Department of Speech-Language Pathology, Seton Hall University, Nutley, NJ, United States
12 of America
13

14 ³ Department of Occupational Therapy, University of South Dakota, Vermillion, SD, United
15 States of America
16

17 ⁴ Department of Communication Sciences and Disorders, University of Wisconsin - Whitewater,
18 Whitewater, WI, United States of America
19
20

21 Short title: Sensorimotor Voice Training on ERPs to Pitch-Shifted Auditory Feedback
22

23 *Corresponding Author
24

25 Email: Sona.Patel@shu.edu
26

27 **Abstract**

28 The pitch perturbation technique is a validated technique that has been used for over 30
29 years to understand how people control their voice. This technique involves altering a person’s
30 voice pitch in real-time while they produce a vowel (commonly, a prolonged /a/ sound). Although
31 post-task changes in the voice have been observed in several studies (e.g., a change in mean
32 F0 across the duration of the experiment), the potential for using the pitch perturbation
33 technique as a training tool for voice pitch regulation and/or modification has not been explored.
34 The present study examined changes in event related potentials (ERPs) and voice pitch in three
35 groups of subjects due to altered voice auditory feedback following a brief, four-day training
36 period. Participants in the opposing group were trained to change their voice F0 in the opposite
37 direction of a pitch perturbation stimulus. Participants in the following group were trained to
38 change their voice F0 in the same direction as the pitch perturbation stimulus. Participants in the
39 non-varying group did not voluntarily change their pitch, but instead were asked to hold their
40 voice constant when they heard pitch perturbations. Results showed that all three types of
41 training affected the ERPs (N1 peak latency, P2 peak latency, and P2 peak amplitude) and the
42 response latency and magnitude of the voice pitch shift response in the baseline, “hold your
43 voice pitch steady” task (an indicator of voice pitch regulation). These results demonstrate that
44 participation in pitch-shifted auditory feedback tasks even for brief periods of time can modulate
45 the automatic tendency to compensate for alterations in voice pitch feedback and has
46 therapeutic potential.
47

48 *Keywords:* Vocal Training, Auditory Feedback, Pitch, ERPs, Voice Control
49

50 Introduction

51 Sensorimotor control is important for achieving accuracy of goal-directed movements
52 and involves active integration of sensory feedback and motor commands during an ongoing
53 movement [1]. Speech-motor control relies on integration of auditory feedback information in
54 order to adjust motor commands to correct for deviations from the intended production and
55 produce clear and fluent speech [2]. To examine the sensorimotor integration process for voice
56 and speech, altered auditory feedback can be used [3,4]. The pitch perturbation technique is an
57 established method for examining the sensorimotor system for voice control [3,5-9]. In this
58 technique the auditory feedback is manipulated by changing the voice pitch while a person is
59 speaking, which results in a perceived mismatch between the intended and perceived
60 vocalizations. Deviations in the auditory feedback from the intended voice pitch result in
61 predictable modifications to the voice. These modifications are typically compensatory and in
62 the opposite direction of the pitch shift (an “opposing” response), although voice changes in the
63 same direction of the shift (a “following” response) do occur [3,10-11]. The underlying
64 mechanisms or processes that result in an opposing or following response to auditory feedback
65 errors are unclear. An understanding of the nature of response types (opposing vs following) to
66 auditory errors could be useful for voice rehabilitation, as training protocols also take advantage
67 of auditory feedback in order to improve speech.

68 Although altered auditory feedback techniques have typically been used to examine
69 sensorimotor control of voice in healthy individuals, some evidence suggests that these
70 techniques might also be clinically useful. Delayed auditory feedback is a common therapeutic
71 technique for improving fluency in individuals who stutter and in individuals with Parkinson’s
72 disease [12-14]. Laukkanen [15] demonstrated that by shifting the pitch of auditory feedback
73 while participants read a text aloud, it was possible to change one’s habitual pitch. The authors
74 concluded that speaking repeatedly under the influence of auditory feedback changes a

75 person's voice, and as a result, this technique might be useful in voice training and therapy. A
76 number of studies have demonstrated that voice training can affect voice control [16]. For
77 example, vocal training for singing has been shown to affect pitch control [17-19]. These
78 changes are not only seen at the behavioral level during singing but also at the neurological
79 level [20,21]. Zarate and Zatorre [21] argue that activities involving activation of sensorimotor
80 and auditory areas are associated with changes in cortical regions as a result of musical
81 practice.

82 The exact mechanism by which training paradigms that utilize altered auditory feedback
83 modify speech production patterns long-term remains unclear. One possibility is that these
84 paradigms involve sustained attention to the auditory feedback, resulting in improved attentional
85 control during sensorimotor integration and thus speech production. Tumber and colleagues
86 [22] demonstrated the role of attention on vocal compensations to pitch-shift modulations using
87 a dual-task paradigm. Compensatory voice responses were smaller under a dual-task condition
88 in which individuals had to monitor a visual stream of information for target letters while
89 vocalizing under pitch-shift conditions. This suggests that when less attention was available for
90 the pitch-shift task, individuals were less able to utilize the auditory feedback to change speech
91 production. Additionally, a study by Li and colleagues [23] found that working memory training
92 modified brain activation during a pitch-shift task. In this study participants trained on an
93 adaptive backwards digit span task for 12 days, and ERPs were measured before and after
94 training during a standard pitch-shift task. Their results showed modifications in the N1 and P2
95 amplitude recorded during pitch-shift perturbations following training. The working memory
96 training paradigm used in this study involved sustained attention to the auditory domain over
97 multiple training sessions.

98 Taken together, these studies suggest that training paradigms may improve voice
99 control and related brain processes by causing the individual to engage additional attentional
100 control mechanisms following a sustained focus on auditory attention processing. However, no

101 evidence exists on whether volitional changes to pitch-shifted feedback impact the automatic
102 error correction processes of voice control. Furthermore, no evidence exists to explain whether
103 simply speaking under pitch-shifted auditory feedback passively directs attention to the auditory
104 feedback, thereby impacting voice control and the related neural mechanisms. Thus, the
105 purpose of this study was to determine whether the neurological mechanisms for voice control
106 are modified due to brief training under pitch-shifted auditory feedback. To consistently change
107 one's voice pattern, it is necessary to go through a period of training that should lead to a stage
108 when the behavior becomes automatic. The automatic nature of this or any movement as a
109 function of training is likely to be reflected in the neural mechanisms underlying that movement
110 [21] and can be examined using event related potentials (ERPs) from electroencephalogram
111 (EEG) recordings. In this study the ERPs were recorded in response to shifts in the voice pitch
112 of one's auditory feedback while vocalizing to assess whether the altered auditory feedback
113 resulted in changes to the automatic compensatory response to alterations in voice pitch
114 feedback. The pattern of auditory-motor ERPs (i.e., the P50-N1-P2 ERP complex) obtained as
115 a result of speaking with altered auditory feedback have been shown to produce a consistent
116 pattern across studies [24-27] and have been reported to reflect the neural processing of voice
117 pitch feedback perturbations during vocalization [28].

118 In two variants of the vocal training task implemented in the present study, participants
119 volitionally changed their voice fundamental frequency (f_0) during the production of a steady
120 vowel sound. In the third variant, participants did not intentionally vary their voice pitch during
121 vowel production, but instead, were instructed to keep their voice pitch constant. These tasks
122 mirrored those implemented by Hain and colleagues [10] in which participants were also asked
123 to oppose the direction of the shift, follow the direction of the shift, or ignore the shift and
124 maintain a steady pitch. However, we implemented a between-groups design where each
125 group performed a single task: oppose the shift (the "opposing group"), follow the shift (the
126 "following group"), and ignore the shift (the "non-varying group"). The opposing and following

127 tasks were performed by different groups because differences in the vocal response during
128 each task were observed by Hain and colleagues [10] and also because a multiple baseline
129 approach would not be practical (3 tasks would require 6 EEG sessions, resulting in 15 days of
130 testing or 27 hours). In addition, both voice responses and auditory-motor ERPs were
131 measured during a baseline “ignore the pitch-shift” task in a pretest-posttest design, specifically,
132 before and after four training sessions. This allowed us to examine the effects of short training
133 intervals (a few sessions) as would be encountered in typical therapy sessions on voice pitch
134 regulation in typical individuals.

135 **Methods**

136 **Participants**

137 Thirty-eight participants were recruited from Northwestern University. All participants
138 were native speakers of American English and self-reported being right-hand dominant. They
139 all had normal hearing at octave intervals from 250 Hz to 8000 Hz at 20 dB HL [29] and passed
140 tests of central auditory processing (“CAP”; the Duration Pattern Sequence test and Pitch
141 Pattern Sequence test [30, 31]) with a score of at least 90% (18 of 20) correct. Participants
142 reported having no history of neurological, speech, or language disorders and minimal vocal
143 training (defined as less than three years of vocal training) and that they did not regularly sing in
144 a group (two times per week or less). Participant recruitment and testing procedures were
145 approved by the Northwestern University Institutional Review Board.

146 All participants were randomly assigned to one of three training groups: opposing,
147 following, or non-varying (described in the next section). Of the participants recruited, one did
148 not complete the study, two were dropped as a result of not being able to perform the training
149 task, two were dropped due to technical errors in data collection, and four were dropped due to
150 artifacts in the EEG signals (mainly due to movements and sleepiness). As a result, a total of
151 29 participants remained: 10 participants in the opposing group (3 males, 7 females; mean 19.8

152 years), 9 participants in the following group (3 males, 6 females; mean 21.4 years), and 10
153 participants in the non-varying group (5 males, 5 females; mean 21.0 years).

154 **Procedures**

155 All testing took place in a double-walled, sound-treated booth. A visual display was
156 presented on the computer screen instructing the participant to vocalize an /a/ vowel. A
157 progress bar indicated the length of time to either “Get ready” or “Say aah”. Participant
158 vocalizations were recorded using an AKG boomset microphone (model C420; AKG, Vienna,
159 Austria). The voice was amplified with a 10 dB gain using a Mackie mixer (model 1202; Loud
160 Technologies, Woodinville, IL) and presented as real-time feedback using a Sennheiser headset
161 (Sennheiser Electronic Corporation, Old Lyme, CT) during training and Etymotic Research, Inc.,
162 (model ER-2) ear inserts (Etymotic Research, Inc., Elk Grove Village, IL) during the baseline
163 task pre- and post-training. During the vocalization the participant’s voice pitch was shifted
164 upward or downward by 100 cents (100 cents = 1 semitone) using an Eventide Eclipse
165 Harmonizer (Eventide, Little Ferry, NJ), creating perturbations in the real-time auditory
166 feedback. MIDI software (Max/MSP v. 5.0) was used to present the display and control
167 characteristics of the pitch-shift (direction randomization, timing, and magnitude). The
168 vocalizations, modified voice feedback signal, and control pulses (used as an indicator of the
169 direction of the pitch-shift) were digitized at 10 kHz, low-passed filtered at 5 kHz, and recorded
170 using LabChart Pro software (AD Instruments, Colorado Springs, CO).

171 To investigate the effects of volitional voice training on a person’s involuntary pitch-shift
172 response, a pretest-posttest design was used. Participants underwent a specific task (referred
173 to as the “baseline task”) before and after a training period (Days 1 and 5). Vocal training was
174 performed in four sessions, each on a different day within a two-week period (Days 2-5). The
175 total test time for this experiment was 5.5 hours, with no longer than 1.5 hours per session. All
176 participants were monetarily compensated for their participation.

177 **Training task**

178 During the training task a single 1000-ms long shift in pitch occurred during each
179 vocalization (either 100 cents down or up) with a random onset between 500 ms and 1000 ms
180 after voice onset. Participants were asked to dynamically change their pitch to either volitionally
181 oppose (the “opposing” group) or follow (the “following” group) the direction of the actual shift
182 depending on the group they were assigned to and maintain the new pitch level for the
183 remainder of their breath. Participants in a third group (the “non-varying” group) were simply
184 asked to ignore the changes in their auditory feedback and maintain a constant pitch and
185 loudness level (i.e., hold your voice steady). Thus, the non-varying group did not volitionally
186 change their voice in response to the stimuli. Participants performed a short practice session of
187 10 trials before testing. Each training session consisted of 4 blocks of 52 vocalizations.

188 **Baseline task**

189 For the baseline task (Days 1 and 5), all participants were first fitted with a 32-channel
190 Brain Products actiCAP active electrode cap that was connected to the actiCHamp amplifier
191 (Brain Products GmbH, Germany) for EEG recordings. In addition to recording voice samples,
192 event-related potentials were recorded using BrainVision Recorder software (Brain Products,
193 GmbH, Germany) at a sampling rate of 5 kHz and then low-pass filtered at 400 Hz.

194 Participants vocalized a steady “aah” sound while their pitch was shifted for five, 200-ms
195 segments within each vocalization. The first shift occurred randomly between 700 ms and 1000
196 ms after vocalization onset, and each successive pitch-shift occurred randomly between 700
197 and 900 ms after the onset of the previous shift. In this task participants in all three groups were
198 instructed to ignore changes to their voice and continue to say “aah” at a normal comfortable
199 level for the length of the progress bar. A total of 52 test vocalizations were recorded before
200 training and after training, which resulted in 260 trials for each measurement (52 vocalizations x

201 5 pitch shifts per vocalization) with an approximately equal number of upward and downward
202 pitch shifts.

203 **Data Analysis**

204 Since we were interested in the effects of training on the baseline task (the pitch-shift
205 reflex, an indicator of voice control), data analysis was performed for both ERP and voice data
206 during the baseline task on Day 1 (pre-training) and Day 5 (post-training). Data analysis was
207 not performed for the training task, as these results are reported elsewhere [8].

208 **EEG analysis**

209 The ERPs were obtained by averaging the recorded EEG signals using Brain Products'
210 Analyzer software, synchronized to the onset of the pitch-shift stimulus. Standard
211 preprocessing of the data was performed including filtering (1-50 Hz), segmentation (500 ms
212 segments were selected; 100 ms pre-shift and 400 ms post-shift), artifact rejection (on the
213 frontal channels and those epochs with amplitudes exceeding $\pm 50 \mu\text{V}$), normalization of the
214 mean value to 0, and averaging across all trials. In addition, the data were re-referenced to the
215 common reference instead of using the reference electrode, since preliminary results showed
216 high activation at the region of the reference electrode (FCz). This methodology allowed us to
217 make use of the electrode at the FCz location.

218 The N1 and P2 peak amplitudes and latencies were extracted for a subset of the
219 channels showing maximal negative and maximal positive responses, respectively. An
220 automatic search was performed to identify the global minima (N1 peak) and maxima (P2 peak),
221 during the time window of maximal activation. The following statistical tests were performed on
222 our four dependent variables (N1 peak amplitude, N1 peak latency, P2 peak amplitude, P2 peak
223 latency). First, four separate linear mixed models were run to identify any group (opposing,
224 following, non-varying) differences or within-subject effects of stimulus direction (up, down),
225 electrode site (Cz, Fz, FCz), and time (pre-training, post-training). We were primarily interested

226 in the main effect of time and the interaction of group x time to determine overall changes due to
227 training and whether any changes from pre-training to post-training varied by group, but we also
228 report the group x electrode, and group x stimulus direction interactions. Follow-up paired
229 samples t-tests were also performed within each group to confirm differences in time (from pre-
230 training to post-training) for each electrode and stimulus direction.

231 **Vocal analysis**

232 Voice samples were analyzed in Igor Pro (Wavemetrics, Inc., Lake Oswego, OR), which
233 called upon Praat [32] for f_0 detection. Praat was used to develop a wave representing the
234 voice f_0 contour, which was used in further analyses. These f_0 contours were first segmented
235 into individual trials of 1100 ms duration (400 ms prior to the pitch-shift onset and 700 ms
236 following the pitch-shift onset). Then outliers were removed from each trial using several
237 processes including normalization by setting the mean baseline voice pitch to 0 cents and
238 removal of extreme values in the vocalization wave prior to the pitch-shift (for *threshold* = 30
239 cents, where max cents > *threshold*, and min cents < *-threshold* were rejected) and in the entire
240 duration of each trial (for *threshold* = 1000 cents, where the whole wave was rejected if max
241 cents > *threshold* or min cents < *-threshold*). Only responses that opposed the direction of the
242 pitch shift were used. Finally, the trials were averaged within a participant for each condition
243 (+100-cent shifts, -100-cent shifts). The magnitude of the largest upward or downward
244 compensatory peak (“response magnitude”) and time that the peak reached maximum
245 amplitude (referred to as the “response latency”) was measured for each subject and submitted
246 to statistical testing using general linear mixed models.

247 **Results**

248 **ERP Results**

249 The ERPs showed that the maximal negative response occurred between 130-160 ms
250 pre-training and 115-135 ms post-training at the following frontal-central electrodes: Cz, FCz, Fz
251 (shown in Fig 1). The maximal positive response occurred between 210-250 ms pre-training
252 and post-training at the same three electrodes (shown in Fig 2). Each subject's N1 and P2 peak
253 information was extracted for these three channels using the above time windows. The grand
254 averaged ERPs for each group are shown for three electrode sites (Cz, FCz, and Fz) during the
255 pre-training and post-training phases in Fig 3.

256 **Fig 1. Mapping View of the N1 Response.** Mapping view of the grand averaged ERPs
257 from 130-160 ms pre-training (top row) and 115-135 ms post-training (bottom row) for the
258 following, opposing, and non-varying groups.

259 **Fig 2. Mapping View of the P2 Response.** Mapping view of the grand averaged ERPs
260 from 210-250 ms pre-training (top row) and 210-250 ms post-training (bottom row) for the
261 opposing, following, and non-varying groups.

262 **Fig 3. Grand-Averaged ERPs by Group at Pretest and Post-test.** Grand averages of the
263 ERPs at three electrode sites (Cz, FCz, and Fz) for all three groups: opposing (pre: red
264 dashed line; post: red solid line), following (pre: blue dashed line; post: blue solid line), and
265 non-varying (pre: black dotted line; post: black solid line).

266 Statistical analysis was performed in SPSS (v.17; SPSS Inc., Chicago, IL) to compare
267 the ERPs obtained before and after training and whether this difference was affected by four
268 fixed factors including the within-subject factors of time (pre-training vs post-training), stimulus
269 direction (up, down) and electrode (Fz, FCz, Cz) and the between-subjects factor of group
270 (opposing, following, non-varying). For this, separate linear mixed models were conducted for
271 N1 peak latency and N1 peak amplitude. Because Mauchly's test indicated that the assumption
272 of sphericity was violated ($p < 0.05$), Greenhouse-Geisser corrected estimates were used.
273 Comparisons show a significant effect of time on N1 peak latency, ($F(1,26) = 31.283, p < 0.05$),
274 but not N1 amplitude, ($F(1,26) = 0.096, p = 0.76$). None of the interaction terms for N1 peak

275 latency or amplitude (group x time, direction, or electrode) were significant. Thus, the N1 peak
276 occurred earlier post-training (129.32 ms) compared to pre-training (148.94 ms) across all
277 groups. The N1 peak latency and N1 peak amplitude are shown (at the Cz electrode) for all
278 three groups in Fig 4.

279 **Fig 4. N1 Latency and Amplitude.** The a) mean N1 peak latency and b) mean N1 peak
280 amplitude from the ERPs at Cz for the opposing (Opp), following (Fol), and non-varying
281 (Non) groups pre-training (solid bars) compared to post-training (slanted lines). Bars
282 represent standard error of the mean.

283 A second set of tests was performed to examine the effects of the same four factors
284 (time, direction, electrode, group) on P2 peak latency and P2 amplitude. Greenhouse-Geisser
285 corrected comparisons show a significant effect of time (pre-training vs post-training) on both P2
286 peak latency, ($F(1,30) = 26.145$, $p < 0.05$), and amplitude, ($F(1,30) = 25.250$, $p < 0.05$).
287 Specifically, the P2 peak occurred earlier post-training (230.87 ms) and with greater amplitude
288 (1.890 V) compared to pre-training (250.05 ms, 1.448 V). Of the interaction terms, only the
289 electrode by group interaction was significant for both P2 peak latency, ($F(3,60) = 3.269$, $p <$
290 0.05) and amplitude, ($F(2.838,60) = 5.008$, $p < 0.05$). The mean P2 peak latency and mean P2
291 peak amplitude (at the Cz electrode) are shown for all three groups in Fig 5. The finding of a
292 significant effect of time and not the time by group interaction demonstrates a consistent change
293 in the N1 peak latency and P2 peak latency and amplitude from pre-training to post-training
294 across all groups.

295 **Fig 5. P2 Latency and Amplitude.** The a) mean P2 peak latency and b) mean P2 peak
296 amplitude from the ERPs at Cz for the opposing (Opp), following (Fol), and non-varying
297 (Non) groups pre-training (solid bars) compared to post-training (slanted lines). Bars
298 represent standard error of the mean.

299 **Voice Responses**

300 The grand averages of the voice f_0 response contours pre-training and post-training are
301 shown for each group in Fig 6 (Panel A: opposing, Panel B: following, and Panel C: non-
302 varying). The voice responses to upward pitch-shifts and downward pitch-shifts are displayed in
303 separate graphs within each panel. The dashed vertical line is the onset time of the pitch-shift
304 stimulus. All groups demonstrate changes in the magnitude of the voice pitch responses from
305 pre-training to post-training.

306 **Fig 6. Grand-Averaged Vocal Responses by Group.** Grand-averaged vocal
307 responses to upward (top) and downward (bottom) pitch shifts for the a) opposing, b)
308 following, and c) non-varying groups (blue line represents pre-training responses, red
309 line represents post-training responses).

310 Differences in the voice responses were examined between groups for two measures: 1)
311 voice response latency and 2) magnitude of the largest upward or downward compensatory
312 peak. Linear mixed models were used to test differences in voice response latency with the
313 between-subjects factor of training group (opposing, following, non-varying) and within-subjects
314 factors of time (pre-training, post-training) and direction (up, down). Levene's test revealed
315 differences in the between-group variances at pre- and post-training ($F(2,69) = 3.927, p < 0.05$)
316 and $F(2,69) = 6.016, p < 0.05$, respectively). Therefore, all values were log-transformed prior to
317 further analysis. Main effects of direction ($F(1,11) = 5.541, p < 0.05$) and time ($F(1,11) = 5.262,$
318 $p < 0.05$) were observed. This finding indicates that the latency of the response to the pitch-
319 shift stimulus for all groups changed from the pre- to post-training period similarly.

320 Next, differences in the magnitude of responses were examined pre- and post-training
321 for each group. Levene's test was not significant at pre- and post-training ($F(2,69) = 1.445, p >$
322 0.05) and ($F(2,69) = 1.494, p > 0.05$), respectively), so log transformation was not necessary.
323 Since direction was not a significant factor ($F(1,71) = 0.890, p > 0.05$), the up and down
324 responses were aggregated, resulting in fixed factors of group (opposing, following, non-

325 varying) and time (pre-training, post-training). A linear mixed model was used to examine
326 differences in the absolute values of the peak magnitude of the responses. Results showed a
327 main effect of time ($F(1,23) = 7.651, p < 0.05$) but not group ($F(2,46) = 2.195, p > 0.05$), which
328 suggests that response magnitudes for all groups changed from the pre- to post-training period
329 similarly.

330 Discussion

331 The purpose of the present study was to determine whether short auditory feedback
332 training intervals could modify voice pitch regulation and affect the corresponding ERPs.
333 Subjects were trained over a short, four-day period to respond to unpredictable perturbations in
334 the pitch of their auditory feedback as they produced a prolonged vowel sound (/a/). Three
335 types of training were implemented including (a) instructing subjects to change their voice f_0 in
336 the opposite direction of pitch-shifted auditory feedback stimuli (opposing the shift), (b)
337 instructing subjects to change their voice f_0 in the same direction of the pitch-shifted auditory
338 feedback stimuli (following the shift), or (c) by simply instructing subjects to ignore all pitch shifts
339 and maintain a steady voice pitch (non-varying). A pretest-posttest design was used to assess
340 changes in voice control before and after the training period. The outcome measures included
341 magnitude and latency of the compensatory voice response and the magnitude and timing of
342 the corresponding ERPs to the baseline task, i.e., an involuntary pitch-shift task, where
343 participants were asked to hold a constant pitch and loudness (similar to the non-varying task).
344 The resulting involuntary pitch-shift response is an indicator of voice control.

345 Our results show differences in the voice responses and the corresponding ERPs during
346 the baseline task as a result of training. Differences in the ERPs were seen in N1 peak latency,
347 P2 peak latency, and P2 peak amplitude. Specifically, both the N1 and P2 peaks occurred
348 earlier post-training compared to pre-training, and the P2 peak magnitude was enhanced post-
349 training compared to pre-training. These results are consistent with the findings of Li and

350 colleagues [23] who report a decrease in N1 and an increase in P2 magnitude post-training.
351 The finding of systematic changes in the neural response suggests that the trained motor
352 behavior (due to each of the three training tasks) may have become more automatic [21] and
353 the processing of auditory information has become more efficient [23]. These results were
354 complemented by the voice changes, which revealed significant changes in response latency
355 and magnitude in that the peak responses occurred earlier and with a reduced amplitude post-
356 training compared to pre-training. We suspect the reduction in amplitude of the corrective
357 response to pitch-shifts indicates a greater control of the voice, as others have shown larger
358 amplitudes of vocal responses in pathological conditions such as Parkinson's disease, where
359 vocal control is abnormal [33].

360 The present results confirmed our predictions that ERPs are modified following the two
361 dynamic-response training tasks. Surprisingly, a similar pattern of ERPs was observed for the
362 hold-your-voice-steady ("non-varying") task. Some studies have shown that voluntary
363 responses to the perturbation paradigm engage different mechanisms than involuntary
364 responses in that voluntary responses to pitch shifts can have both involuntary and voluntary
365 components [8,10,16]. The involuntary component of responses often results in latency times
366 that are shorter (~100-150 ms) than the voluntary component (~250-600 ms), and are thought to
367 reflect automatic neural processing used by the audio-vocal system to correct for any errors.
368 On the other hand, the voluntary component is thought to represent higher cognitive
369 mechanisms used at a more conscious level to control voice f_0 , such as in speaking and singing
370 tasks [8,10,16]. In the present study, we investigated involuntary and voluntary responses as
371 individual conditions, rather than as components of the voluntary response. Results show that
372 practicing to ignore auditory changes in pitch and hold your voice pitch steady also produced
373 differences in ERP/voice responses, potentially because the non-varying task invoked similar
374 cognitive processes used for voice error detection and correction as the other dynamic or
375 volitional tasks. In other words, both dynamic response and hold-your-voice-steady tasks

376 resulted in training an underlying process that positively affected voice pitch regulation. Further,
377 this task activated strong enough processes to produce a change in the neurological
378 mechanism thought to be involved in controlling the voice [5,23].

379 The results of this study can be used to support the development of brief training
380 interventions for voice modulation. Popular voice therapy programs such as the LSVT® have
381 been shown to be effective in helping individuals with PD to improve their vocal communication
382 by raising voice loudness [34,35]. However, the standard initial treatment program for LSVT®
383 requires a minimum of 16 sessions over a four-week period. Similar to the LSVT®, the training
384 tasks described in the present study also required subjects to monitor their vocal output and
385 modify their voice f_0 based on their auditory feedback. This study found changes in voice
386 control and underlying brain mechanisms supporting speech production in only four brief
387 training sessions. Our findings of behavioral and brain changes due to training suggest that
388 brief voice control paradigms modulate the neurological processes for voice production and may
389 be valuable in applications for individuals with neurological voice disorders, such as patients
390 with PD.

391 **Conclusions**

392 In the present study we hypothesized that training individuals to produce a vocal-motor
393 behavior in response to changes in auditory-sensory feedback would affect the ERPs and voice
394 pitch regulation (also assessed as the pitch shift response). Three types of training were
395 implemented where subjects changed their pitch in the opposite direction to a shift in their
396 auditory feedback while vocalizing a prolonged /a/ vowel or in the same direction to the same
397 shift, or subjects maintained a steady voice pitch with no volitional intent to change F_0 .
398 Effectiveness of training was evaluated by comparing the voice and ERP responses during the
399 baseline task, before and after 4 days of training. Results revealed differences in both the
400 ERPs and voice responses after training for all training tasks. Differences in the ERPs were

401 seen in N1 peak latency, P2 peak latency, and P2 peak amplitude, and voice changes were
402 seen in response latency and magnitude. in both ERPs and voice responses, the peak
403 responses occurred earlier and with a reduced amplitude post-training compared to pre-training.
404 These results suggest that active participation in a vocal task involving the use of altered
405 auditory feedback even for brief periods of time can result in changes in neural activation
406 patterns.

407 **Acknowledgements**

408 The authors would like to thank Chun Liang Chan for his help with computer
409 programming. The authors have no conflicts of interest to disclose. Information regarding
410 reprints can be directed to Sona Patel, Department of Speech-Language Pathology, 123 Metro
411 Blvd, Nutley, NJ 07110.

412 **References**

- 413 1. Riemann BL, Lephart SM. The sensorimotor system, part I: the physiologic basis of
414 functional joint stability. *J Athl Train.* 2002;37(1):71–79.
- 415 2. Gracco VL. Sensorimotor mechanisms in speech motor control. In: *Speech Motor*
416 *Control and Stuttering.* Amsterdam: North Holland: Elsevier; 1991. pp. 53-78.
- 417 3. Burnett TA, Freedland MB, Larson CR, Hain TC. Voice F0 Responses to Manipulations
418 in Pitch Feedback. *J Acoust Soc Am.* 1998;103(6): 3153-3161. doi: 10.1121/1.423073.
- 419 4. Donath TM, Natke U, Kalveram KT. Effects of frequency-shifted auditory feedback on
420 voice F₀ contours in syllables. *J Acoust Soc Am.* 2002;111(1): 357-366. doi:
421 10.1121/1.1424870.
- 422 5. Behroozmand R, Korzyukov O, Sattler L, Larson CR. Opposing and following vocal
423 responses to pitch-shifted auditory feedback: evidence for different mechanisms of voice
424 pitch control. *J Acoust Soc Am.* 2012;132(4): 2468-2477. doi: 10.1121/1.4746984.

- 425 6. Kawahara H. Interactions between speech production and perception under auditory
426 feedback perturbations on fundamental frequencies. *Journal of the Acoustical Society of*
427 *Japan*, 1994;15(3):201-202. doi: 10.1250/ast.15.201.
- 428 7. Larson CR, Burnett TA, Bauer JJ, Kiran S, Hain TC. Comparisons of voice F_0 responses
429 to pitch-shift onset and offset conditions. *J Acoust Soc Am.*2001;110(6):2845-2848. doi:
430 10.1121/1.1417527.
- 431 8. Patel S, Gao L, Wang S, Gou C, Manes J, Robin DA, et al. Comparison of volitional
432 opposing and following responses across speakers with different vocal histories. *J*
433 *Acoust Soc Am.* 2019;146(6): 4244-4254. doi: 10.1121/1.5134769.
- 434 9. Scheerer NE, Jones JA. The relationship between vocal accuracy and variability to the
435 level of compensation to altered auditory feedback. *Neurosci Lett.* 2012;529(2):128-132.
436 doi: 10.1016/j.neulet.2012.09.012.
- 437 10. Hain TC, Burnett TA, Kiran S, Larson CR, Singh S, Kenney MK. Instructing subjects to
438 make a voluntary response reveals the presence of two components to the audio-vocal
439 reflex. *Exp Brain Res.* 2000;130(2):133-141. doi: 10.1007/s002219900237.
- 440 11. Bauer JJ, Larson, CR. Audio-vocal responses to repetitive pitch-shift stimulation during a
441 sustained vocalization: Improvements in methodology for the pitch-shifting technique. *J*
442 *Acoust Soc Am.* 2003;114(2):1048-1054. doi: 10.1121/1.1592161.
- 443 12. Ward D. *Stuttering and cluttering: frameworks for understanding and treatment:*
444 *Psychology Press; 2008.*
- 445 13. Brendel B, Lowit A, Howell P. The effects of delayed and frequency shifted feedback on
446 speakers with Parkinson disease. *J Med Speech Lang Pathol*, 2004;12(4): 131-138.
- 447 14. Soderberg GA. Delayed auditory feedback and the speech of stutterers: A review of
448 studies. *J Speech Hear Disord.* 1969;34(1):20-29. doi: 10.1044/jshd.3401.20.
- 449 15. Laukkanen AM. Artificial Pitch Changing in Auditory Feedback as a Possible Method in
450 Voice Training and Therapy. *Folia Phoniatica Et Logopaedica*, 1994;46(2): 86-96.

- 451 16. Larson CR. Cross-modality influences in speech motor control: The use of pitch shifting
452 for the study of F0 control. *J Commun Disord.*1998;31(6):489-503. doi: 10.1016/s0021-
453 9924(98)00021-5.
- 454 17. Jones JA, Keough D. Auditory-motor mapping for pitch control in singers and
455 nonsingers. *Exp Brain Res.* 2008;190(3): 279-287. doi: 10.1007/s00221-008-1473-y.
- 456 18. Natke U, Donath TM, Kalveram KT. Control of voice fundamental frequency in speaking
457 versus singing. *J Acoust Soc Am.* 2003;113:1587-1593. doi: 10.1121/1.1543928.
- 458 19. Sundberg J. *The Science of the Singing Voice.* Dekalb, IL: Northern Illinois University
459 Press; 1987.
- 460 20. Zarate JM. The neural control of singing. *Front Hum Neurosci.* 2013;7:237. doi:
461 10.3389/fnhum.2013.00237.
- 462 21. Zarate JM, Zatorre RJ. Experience-dependent neural substrates involved in vocal pitch
463 regulation during singing. *NeuroImage.* 2008;40(4):1871-1887. doi:
464 10.1016/j.neuroimage.2008.01.026.
- 465 22. Tumber AK, Scheerer NE, Jones JA. Attentional demands influence vocal
466 compensations to pitch errors heard in auditory feedback. *PLoS ONE.*
467 2014;9(10):e109968. doi: 10.1371/journal.pone.0109968.
- 468 23. Li W, Guo Z, Jones JA, Huang X, Chen X, Liu P, et al. Training of working memory
469 impacts neural processing of vocal pitch regulation. *Sci Rep.*2015;5:16562.doi:
470 10.1038/srep16562.
- 471 24. Behroozmand R, Karvelis L, Liu H, Larson CR. Vocalization-induced enhancement of
472 the auditory cortex responsiveness during voice F0 feedback perturbation. *Clin*
473 *Neurophysiol.* 2009;120(7): 1303-1312. doi: 10.1016/j.clinph.2009.04.022.
- 474 25. Behroozmand R, Larson CR. Error-dependent modulation of speech-induced auditory
475 suppression for pitch-shifted voice feedback. *BMC neuroscience,* 2011;12:54. doi:
476 10.1186/1471-2202-12-54.

- 477 26. Korzyukov O, Karvelis L, Behroozmand R, Larson CR. ERP correlates of auditory
478 processing during automatic correction of unexpected perturbations in voice auditory
479 feedback. *Int J Psychophysiol.* 2012;83(1): 71-78. doi: 10.1016/j.ijpsycho.2011.10.006.
- 480 27. Korzyukov O, Tapaskar N, Pflieger ME, Behroozmand R, Lodhavia A, Patel S, et al.
481 Event related potentials study of aberrations in voice control mechanisms in adults with
482 attention deficit hyperactivity disorder. *Clin Neurophysiol.*2015;126(6):1159-1170. doi:
483 10.1016/j.clinph.2014.09.016.
- 484 28. Liu H, Meshman M, Behroozmand R, Larson CR. Differential effects of perturbation
485 direction and magnitude on the neural processing of voice pitch feedback. *Clin*
486 *Neurophysiol.* 2011;122(5):951-957. doi: 10.1016/j.clinph.2010.08.010.
- 487 29. ANSI. Methods for manual pure-tone threshold audiometry. American National
488 Standards Institute (Vol. ANSI S3.21-2004). New York, NY; 2004
- 489 30. Musiek FE, Baran JA, Pinheiro ML. Duration pattern recognition in normal subjects and
490 patients with cerebral and cochlear lesions. *Audiology.*1990;29(6):304-313. doi:
491 10.3109/00206099009072861.
- 492 31. Musiek FE, Pinheiro ML. Frequency patterns in cochlear, brainstem, and cerebral
493 lesions. *Audiology.*1987;26(2):79-88.
- 494 32. Boersma P, Weenink D. Praat, a system for doing phonetics by computer. *Glott*
495 *International,* 2001;5(9/10): 341-345.
- 496 33. Liu H, Wang EQ, Metman LV, Larson CR. Vocal responses to perturbations in voice
497 auditory feedback in individuals with Parkinson's disease. *PLoS One.* 2012;7(3):e33629.
498 doi: 10.1371/journal.pone.0033629.
- 499 34. Ramig LO, Fox C, Sapir S. Parkinson's disease: speech and voice disorders and their
500 treatment with the Lee Silverman Voice Treatment. *Semin Speech Lang.*
501 2004;25(2):169-180. doi: 10.1055/s-2004-825653.

502 35. Sapir S, Ramig LO, Fox CM. Intensive voice treatment in Parkinson's disease: Lee
503 Silverman Voice Treatment. *Expert Rev Neurother*. 2011;11(6):815-830. doi:
504 10.1586/ern.11.43.

505











