Compensatory and adaptive responses to real-time formant shifts in adults and children

Frits van Brenk and Hayo Terband

Citation: The Journal of the Acoustical Society of America **147**, 2261 (2020); doi: 10.1121/10.0001018
View online: https://doi.org/10.1121/10.0001018
View Table of Contents: https://asa.scitation.org/toc/jas/147/4
Published by the Acoustical Society of America

JASA ARTICLE

Compensatory and adaptive responses to real-time formant shifts in adults and children

Frits van Brenk[1,a)] and Hayo Terband[2,b)]

[1]Department of Communicative Disorders and Sciences, University at Buffalo, 122 Cary Hall, 3435 Main Street, Buffalo, New York 14214, USA

[2]Utrecht Institute of Linguistics—OTS, Utrecht University, Trans 10, Room 0.25, 3512 JK Utrecht, The Netherlands

ABSTRACT:

Auditory feedback plays an important role in speech motor learning, yet, little is known about the strength of motor learning and feedback control in speech development. This study investigated compensatory and adaptive responses to auditory feedback perturbation in children (aged 4–9 years old) and young adults (aged 18–29 years old). Auditory feedback was perturbed by near-real-time shifting F1 and F2 of the vowel /ɪː/ during the production of consonant-vowel-consonant words. Children were able to compensate and adapt in a similar or larger degree compared to young adults. Higher token-to-token variability was found in children compared to adults but not disproportionately higher during the perturbation phases compared to the unperturbed baseline. The added challenge to auditory-motor integration did not influence production variability in children, and compensation and adaptation effects were found to be strong and sustainable. Significant group differences were absent in the proportions of speakers displaying a compensatory or adaptive response, an amplifying response, or no consistent response. Within these categories, children produced significantly stronger compensatory, adaptive, or amplifying responses, which could be explained by less-ingrained existing representations. The results are interpreted as both auditory-motor integration and learning capacities are stronger in young children compared to adults.

https://doi.org/10.1121/10.0001018

(Received 12 September 2019; revised 11 February 2020; accepted 17 March 2020; published online 16 April 2020)

[Editor: Anders Lofqvist] Pages: 2261–2270

I. INTRODUCTION

The accurate production of speech sounds is a sensorimotor accomplishment in which self-monitoring of the auditory signal plays an important role. Several studies have employed an auditory feedback perturbation paradigm to study the role of auditory feedback by measuring compensatory responses to manipulations of spectral properties in the speech signal. Such a paradigm involves the creation of an apparent mismatch between the speech sound the speaker intended to produce and what he/she hears back. The acoustic signal is recorded during speech production, and an acoustic cue is manipulated in real time and presented back to the speaker through headphones. These studies have demonstrated that the unexpected perturbation of auditory feedback during speech production elicits an *online compensatory response*, usually in the opposite direction of the perturbation to maintain the intended auditory outcome as well as possible (Cai *et al.*, 2010; Houde and Jordan, 2002; Villacorta *et al.*, 2007). This is a measure of real-time feedback control and motor correction. Furthermore, the *sustained* application of acoustic cue shifts has been shown to not only elicit an online compensatory response but also cause the speech motor system to adapt to the perturbation and modify the stored speech motor programs. This *adaptive response* during speech production reflects motor learning in which the feedforward state of the vowel motoric plans is updated for future production control and measured by the persistence of the responses when the perturbation is removed (Cai *et al.*, 2010; Villacorta *et al.*, 2007; Mitsuya *et al.*, 2015). Auditory feedback therefore serves both as a *guiding signal* in the online control and correction of speech movements (Perkell, 2012; Perkell *et al.*, 2007; Perkell *et al.*, 1997) as well as a *teaching signal* for the acquisition of speech motor programs (Guenther *et al.*, 1998; Guenther and Perkell, 2004; Perkell, 2012; Perkell *et al.*, 1997). As auditory feedback plays an important role in speech motor learning (Guenther *et al.*, 1998; MacDonald *et al.*, 2012; Ménard *et al.*, 2013; Ménard *et al.*, 2008), it is imperative to assess to what extent children compensate for and adapt to auditory feedback perturbations throughout their developmental trajectory. Thus far, little is known about the strength of feedback control and motor learning during speech development. A few studies have investigated responses to auditory feedback perturbations of first and second formants in vowels produced by young children. The findings of these studies suggest that crucial steps are made in the development of sensorimotor processing between the ages of 4 and 9 years old. MacDonald *et al.* (2012) investigated the ability to compensate for shifts of F1 and F2 in the unrounded

---

[a] ORCID: 0000-0003-4777-919X.

[b] Also at: Department of Languages, Literature and Communication, Utrecht University, Trans 10, Room 0.25, 3512 JK Utrecht, The Netherlands. Electronic mail: h.r.terband@uu.nl. ORCID: 0000-0001-7265-3711.

vowel /ɛ/ in 20 toddlers (2-year-olds) and 26 young children (4-year-olds) as compared to 26 adults. For the 2-year-olds, the results did not show a response to the perturbations, indicating that they do not yet self-regulate their vowel acoustics like adults or young children do. The authors speculate that error correction based on feedback of the child's own speech develops only after the internal representation of a sound category is robust enough. For the 4-year-olds, on the other hand, the results showed a compensatory response of similar strength as compared to the adults, albeit with a larger token-to-token variability (MacDonald *et al.*, 2012), indicating that the acquisition of speech sounds may be largely completed within this time frame. The ability of young talkers to change their motor output through immediate sensory feedback was further demonstrated by Shiller *et al.* (2010), who manipulated the spectral properties of /s/ toward /ʃ/ during the production of a series of monosyllabic words by 9–11-year-old children as compared to young adults and investigated compensatory changes in speech production as well as learning aftereffects in speech perception. The results showed that the children were able to compensate for the altered auditory feedback to a comparable degree as adults, albeit still with a larger token-to-token variability. Daliri *et al.* (2018) examined sensorimotor compensatory effects to auditory perturbation in fluent children (11 males, age range 7.1–11.4 years) and adults (8 males, age range 18.8–43.8 years) as control groups for children and adults who stutter. The results showed that both nonstuttering adults and children successfully compensated for F1 perturbation but not F2 perturbation. The strength of compensation was found to be comparable between the two groups, supporting the results found by MacDonald *et al.* (2012) in that children aged 4 years and older do employ ongoing feedback-based motor correction during formant perturbations.

Whereas the studies discussed so far investigated either real-time feedback control and correction through compensatory responses or motor learning through adaptive responses, a small number of studies have examined both feedback and feedforward aspects of production control simultaneously. For example, Ménard and colleagues (Ménard *et al.*, 2013; Ménard *et al.*, 2008) investigated the responses to perturbations in 4-year-old French speakers compared to adult speakers by using a lip-tube to alter F1 and F2 during the production of the rounded vowels /u/ and /y/. Their results indicated that the 4-year-olds were able to develop a compensatory strategy for online correction, but they were not able to adapt to the perturbation, i.e., to modify and update their stored representations as consistently as adults did. The authors interpret these findings in that whilst both speaker groups relied on an internal model used for compensation, children did not memorize and maintain the articulatory strategies built during perturbation as well as adults and were therefore less effective in using these when perturbation was removed (Ménard *et al.*, 2013; Ménard *et al.*, 2008). Alternatively, it might be that the children reacted differently to altered somatosensory feedback involved with the installation and removal of the lip tube, or the presence of the lip tube prevented the children from associating the compensatory articulation with the original speech sound mapping. Shiller and Rochon (2014) examined the effect of perceptual learning on adjustment to altered auditory effects in children aged 5–7 years old by using a single minimal pair contrast involving vowels /æ/ and /ɛ/. The results showed that, although not formally reported, children appeared to not only compensate for but also adapt to perturbation of F1 with children undergoing perceptual learning showing stronger responses to perturbation. In addition to experiments employing spectral perturbation, a number of studies have investigated the effects of pitch perturbations on sensorimotor learning in children. Scheerer *et al.* (2016) examined the effects of downward pitch perturbations in children aged 3–8 years old and young adults and reported further evidence of compensatory responses to auditory perturbation by children. Their study found that both groups demonstrated sensorimotor corrections, but compensatory responses to pitch shifts were higher in adults compared to children. In addition, evidence for adaptation effects were reported for both children and adults, but again no direct comparisons of adaptation strength were made across groups.

In summary, reliable adaptive responses to perturbed auditory feedback have been found in adults (e.g., Cai *et al.*, 2010; Houde and Jordan, 2002; Villacorta *et al.*, 2007) but have not been systematically investigated in children compared to adults (MacDonald *et al.*, 2012; Shiller *et al.*, 2010; Daliri *et al.*, 2018; Scheerer *et al.*, 2016). In a previous study comparing typically developing children with children with speech sound disorder, we have found indications of motor learning for both groups. However, the results also showed large differences in the strength of compensation and adaptation between individuals (Terband *et al.*, 2014). As our previous study did not include an adult control group, and considering that thus far no studies have investigated both compensatory and adaptive responses to perturbation of spectral properties of the speech signal in children, it is yet unknown how children's abilities of real-time feedback control and motor learning of spectral information compare to those of adult speakers.

Furthermore, previous studies have reported that a number of adult speakers within a given group under investigation do not show a consistent response; that is, some adult speakers do not react to perturbed feedback or show production changes that follow the direction of the perturbation. These inconsistent responses have been reported for formant perturbations (Villacorta *et al.*, 2007; Cai *et al.*, 2010) and pitch perturbations (Behroozmand *et al.*, 2012). However, it is yet unknown whether and to what degree children show inconsistent and divergent responses to perturbation and how their behavior compares to adult populations. In the current study, we therefore set out to further investigate to what extent children between 4 and 9 years of age compensate for real-time formant shifts (i.e., real-time feedback control) and to what extent they adapt to the perturbations and modify their stored speech motor programs (i.e., motor learning) in comparison with adult speakers.

In light of the above, the current experiment was designed to test the following three hypotheses. Because previous research has shown that children are able to compensate for oral motor lip tube perturbations (Ménard *et al.*, 2013; Ménard *et al.*, 2008), as well as able to compensate for auditory perturbations (Shiller and Rochon, 2014; Daliri *et al.*, 2018), children may be expected to possess the capacity to notice and act on auditory perturbations. However, as auditory-motor integration of children is still in a developmental phase, we predict that their real-time feedback control will not be as effective as for those with a matured speech motor system, i.e., young adults. As such, we predict children as a group to show, on average, smaller total compensatory responses to real-time formant shifts as compared to adults. Furthermore, children have been found to exhibit a larger speech motor learning plasticity that enables them to rapidly acquire and store newly learned auditory-articulatory mappings (Walsh *et al.*, 2006). Based on these results, we would expect children to show larger learning effects, that is, larger aftereffects of adaptation to altered feedback immediately after perturbation is suspended as compared to adults. Finally, although we expect children to be able to compensate for and adapt to perturbation, if children's motor skills are not yet developed enough in serially reproducing identical vocal tract configurations (cf. Ménard *et al.*, 2008) and produce less consistent motor responses to perturbed speech trials (cf. Shiller and Rochon, 2014), we predict that children will show higher token-to-token variability across trials compared to adults with disproportionately larger variability values during the compensation phase where the speech signal is perturbed.

II. METHOD AND MATERIALS

A. Participants

A total of 25 children participated in the study, but due to fatigue and attention loss two children were unable to complete the auditory perturbation task, and reliable data collection was unsuccessful, leaving 23 children for further analysis [11 female, 12 male; age range 4.0–8.7 years (mean, 5.6; standard deviation, SD, 1.4)]. In addition, 50 young adults [32 female, 18 male; age range 18–29 years (mean, 22.3; SD, 2.7)] participated in the study. None of the participants had current or previous speech or hearing problems. The children were recruited through local schools and acquaintances. The young adults were recruited through the Faculty of Arts participant pool of Utrecht University. All participants were recruited from the Randstad area of the Netherlands and native speakers of Standard Dutch. Written consent was obtained from all adult participants and parents or caretakers of child participants prior to starting the study.

B. Stimuli

The stimuli were three consonant-vowel-consonant (CVC) words: /bɪːr/ (bear), /vɪːr/ (feather), /pɪːr/ (pear), all containing a near-close near-front unrounded vowel. Auditory feedback was perturbed by shifting F1 and F2 of the near-close near-front lax vowel /ɪː/ during the production of three CVC words. Specifically, the first formant was raised 25%, and the second formant was raised 12.5%. A set of naming concepts containing the vowel /ɪː/ was constructed, which during perturbation was shifted toward the vowel /a/ and thus remained within the F1/F2 vowel space of Dutch speaking children. The CVC words with perturbed vowels are mapped to existing Dutch words as well: /baːr/ (bar), /vaːr/ (navigate), and /paːr/ (pair), making the chosen vowel contrasts ideally suited for this study.

C. Experimental procedure

The adults' data were collected at the Utrecht Institute of Linguistics (UIL)-OTS research laboratory, whereas the children's data were collected in a quiet room at the children's school or at a familiar local community centre. The participants were seated in front of a personal computer monitor showing pictures of the three target words. As the children were considerably younger in age than previous studies using the current experimental paradigm, the vowels were required to be elicited by children by naming pictures of concepts that were already in their vocabulary instead of being read from a screen. An animated bird flying over one of the pictures cued the participant to speak the intended word. Pictures were selected in a randomized block design representing the bear, the pear, or the feather. This ensured that the identity of the upcoming target word was masked until the last moment to limit word preparation and selection possibilities by the speaker.

Auditory feedback was manipulated using the Audapter software module (Cai *et al.*, 2012; Cai *et al.*, 2010), a custom MEX-based software module written in Microsoft Visual C++ (Microsoft, Redmond, WA) and executed under MATLAB (The Mathworks, Natick, MA). Recordings were made by an externally powered lavalier microphone (type Audio-Technica AT803b, Audio-Technica, Machida, Tokyo, Japan). The microphone signal was amplified, digitized to a sampling rate of 16 kHz and 16-bit resolution, and transferred to a Lenovo Thinkpad laptop (Lenovo, Hong Kong) via an external audio interface (MOTU MicroBook, MOTU, Cambridge, MA).

During perturbation conditions, the software tracked and shifted the formant frequencies of the vowel using autoregressive linear predictive coding and a dynamic-programming tracking algorithm and played back the target word over headphones (type Sennheiser HD 380 pro, Sennheiser, Wedemark, Germany) in near-real-time (estimated feedback delay 11–14 ms; Cai *et al.*, 2010; Cai *et al.*, 2012) at an average sound pressure level (SPL) calibrated at 78 dB. The auditory signal was perturbed by shifting F1 and F2 of the near-close near-front lax vowel /ɪː/ during the production of the three CVC words. The first formant was raised 25% and the second formant was lowered 12.5%, yielding a more open and more central vowel. The perturbation paradigm consisted of five phases (Fig. 1): a practice phase where participants were made familiar with the

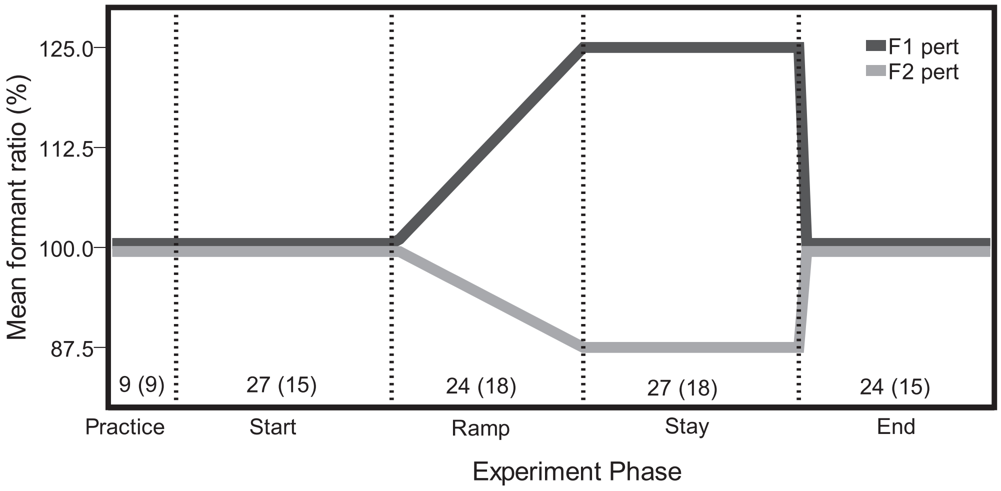


FIG. 1. Experimental paradigm comprising practice (familiarization), start (baseline; no perturbation), ramp (amount of perturbation is slowly increased), stay (hold; full perturbation), and end (release; no perturbation) with the number of trials for each phase (trial numbers of the short program are in parenthesis).

paradigm, a start phase which served as a baseline for unperturbed vowels, a ramp phase where the perturbation was linearly ramped to the maximum, a stay phase where maximum perturbation was applied, and an end phase where the perturbation was suspended.

During the practice phase, the participants were made familiar with the paradigm and practiced the desired vowel duration (approximately between 300 and 500 ms) and loudness (74–84 dB SPL at 10 cm microphone distance) to optimize formant tracking and shifting. The practice phase was of sufficient length for all speakers to be trained to remain within the duration and loudness boundaries, ensuring that for each speaker formant shifts were uniformly applied to the produced words. The start phase served as a baseline and did not contain formant shifts. Subsequently, in the ramp phase, the perturbation was linearly ramped to the maximum in which formants were altered stepwise by approximately 5–7 Hz for F1 and 13–17 Hz for F2 per trial. The stay phase then featured maximum perturbation. The perturbation was suspended at once in the end phase. The total number of trials was 111. Due to the potential occurrence of fatigue and attention loss, 19 children aged less than 7 years participated in a shorter version comprising 75 trials. For each trial, one of three words (bear, feather, or pear) was selected in a randomized block design such that within a block of three trials each of the words was randomly selected once. For example, the start phase of the regular program contained 27 trials, in which the target words "bear," "feather," and "pear" were each presented nine times.

D. Data processing and analysis

1. Compensation and adaptation

For each production, the mean first formant and mean second formant were measured from steady-state portions of the produced vowels using custom scripts for Praat (Boersma and Weenink, 2013). In order to remove speaker-specific and word-specific differences in formant values, the first and second formant frequencies were normalized across the different experimental phases as follows. First, for every speaker and each of the three words, formant frequencies produced in the start phase were averaged. Second, for each word, averaged formant values produced in the start phase were divided by the formant values of the words produced in the other experimental phases to obtain a normalized mean formant ratio for each speaker, which was subsequently multiplied by 100 to arrive at percentages.

The total compensatory response for the auditory perturbation was quantified by calculating the difference in normalized formant frequencies between the stay and start phases. As discussed in the Introduction, this total amount of compensation is a product of adaptation and online compensation and is used as a measure of sensorimotor control: the ability to notice and act on the mismatch between the motor command and the corresponding auditory result (Mitsuya *et al.*, 2015). The adaptive response to prolonged perturbation of vowels was isolated by removing the perturbation at once at the beginning of the end phase. The amount of adaptation was quantified by calculating the differences in normalized formant frequencies between the end phase and the start phase. This is a measure of motor learning: the ability to update motor command representations (Mitsuya *et al.*, 2015). Since this aftereffect of change in motor command is quickly followed by recovery (de-adaptation), the analysis of the end phase was limited to the first 11 trials after the perturbation was suspended.

First, the strength of compensation and adaptation was analyzed and compared between both groups and included all subjects. Statistical analyses were carried out using linear mixed models separately for F1 and F2 data with group (adults, children) and phase (start, stay, end) as fixed factors, subject as a random factor, and word (/bɪːr/, /vɪːr/, /pɪːr/) and repetition (trial numbers) as repeated factors. Significant main and interaction effects were further explored by means of univariate tests where appropriate or a pairwise comparison using Fisher's least significant difference test.

Then, to analyze participant behavior *within* the two groups, we tested each participant *individually* for how they performed at each stage during the experiment with respect to compensatory and adaptive responses. In order to do so, we established the significance of effects of phase (start, stay, end) as well as the direction of these effects. The possible effects included significantly changing the production in the opposite direction of the perturbation (the expected effect), no significant change, or significantly following the perturbation. These individual speaker analyses were carried out by means of linear mixed models separately for F1 and F2 data with phase as a fixed factor, word (/bɪːr/, /vɪːr/, /pɪːr/) as a random factor, and phase and repetition (trial numbers) as repeated factors. In four cases (F1 of CHILD 07 and CHILD 08, and F2 of CHILD 07 and CHILD 23) this model could not reach convergence and was replaced by a less complex model without repeated factors but maintained phase as a fixed factor. This enabled compensation and adaptation effects to continue to be evaluated. The level of significance of all analyses was set at $p < 0.05$.

2. Token-to-token variability

To investigate and compare the stability of compensation and adaptation across the two speaker groups, we calculated individual formant ratio values of each trial in the experimental phases start, stay, and end for every speaker. Token-to-token variability was defined as the SD of the formant ratios for each token, separated by the experimental phase. The SDs of all speakers within a group were used as an outcome measure to compare the two groups. The statistical analysis of token-to-token variability was carried out by means of a linear mixed model analysis assessing group, phase, and group by phase interaction effects separately for F1 and F2.

3. Individual differences in responses of compensation and adaptation

To assess whether the two groups differed in their proportion of responses to perturbation, each speaker was assessed in their behavior with respect to their response to perturbation. The *post hoc* results of the previously executed linear mixed model analyses were used for this purpose. Subsequently, the speaker results were recoded as 1 (going against perturbation), −1 (following perturbation), and 0 (no significant change). Chi-square tests were performed separately for F1 and F2 data and for compensation and adaptation effects.

III. RESULTS

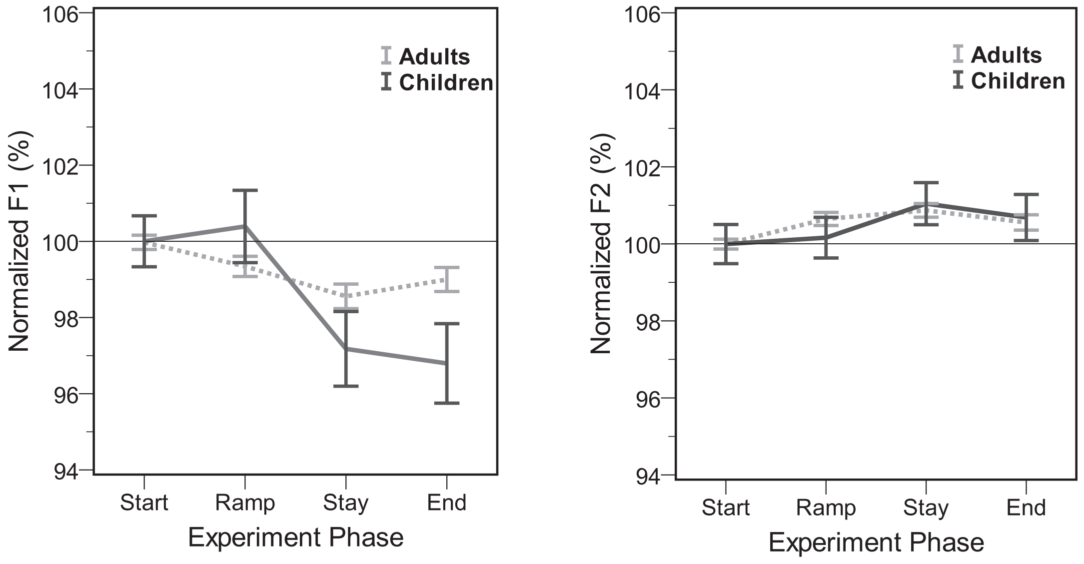
A. Normality and homoscedasticity

Shapiro's test of normality and Levene's test of homoscedasticity were applied to the main outcome measures prior to comparing the groups and phases by a series of statistical analyses. The results showed that both the requirements of normality and equality of variance were satisfied across all measures.

B. Compensation and adaptation across groups

Figure 2 presents normalized F1 and F2 values per group (23 children and 50 adults) for the start, ramp, stay, and end phases.

The results of the linear mixed model analyses showed the following. With respect to F1, a main effect of group was found [$F(1,4061) = 36.5$, $p < 0.001$], where. on average. children showed a larger response as compared to the adults. A main effect of phase was found [$F(2,3110) = 59.0$, $p < 0.001$]. Across groups, responses in the stay and end phases were significantly larger when compared to the start and ramp phases.



(a) (b)

FIG. 2. Mean and across-subject 95% confidence intervals of produced first and second formant frequencies, normalized to the mean values in the start phase, and compared across speaker groups (23 children and 50 adults). (a) F1; (b) F2. Adults, dashed grey line; children, solid black line.

A group × phase interaction effect $[F(2,3110) = 12.0$, $p < 0.001]$ showed that for some of the phases children showed a larger response compared to adults. To further investigate these results, a series of *post hoc* analyses were carried out. Separately for each group, it was found that both adults and children showed significant compensation and adaptation effects (all $p < 0.001$). When comparing the groups for each phase individually, it was found that children showed a stronger compensation response (stay phase; $p < 0.001$) and a stronger adaptation response (end phase; $p < 0.001$).

The statistical results for F2 showed a nonsignificant main effect of group $[F(1,4168) = 3.6$, $p = 0.059]$, when compared across all phases. A main effect of phase was present $[F(2,3133) = 23.7$, $p < 0.001]$ showing that responses in the stay and end phases were significantly larger when compared to the start and ramp phases. The interaction of group × phase was not significant $[F(2,3133) = 0.132$, $p = 0.877]$; both groups responded equally across the different phases. *Post hoc* analyses of each phase separately for each group indicated that adults showed significant compensation and adaptation (both $p < 0.001$), whilst children showed a significant compensation effect ($p < 0.001$) and a nonsignificant adaptation effect ($p = 0.053$). When comparing groups for each phase separately, it was found that there were no group differences in the stay phase ($p = 0.134$) or end phase ($p = 0.451$).

C. Token-to-token variability

Figure 3 shows the average token-to-token variability for each group, separated by experimental phase and formant. The statistical results of the token-to-token variability showed a significant effect of group for the first formant $[F(2,292) = 80.1$, $p < 0.001]$ and the second formant $[F(2,292) = 56.6$, $p < 0.001]$. For both formants, the group of children displayed higher token-to-token variability compared to the adults. A significant effect of phase was present with respect to F1 $[F(3,292) = 3.9$, $p = 0.009]$, where it was found that, pooled over both groups, variability in the stay phase was significantly higher compared to the start phase. No such effects were found for F2. Furthermore, significant group × phase interaction effects were absent for both formants.

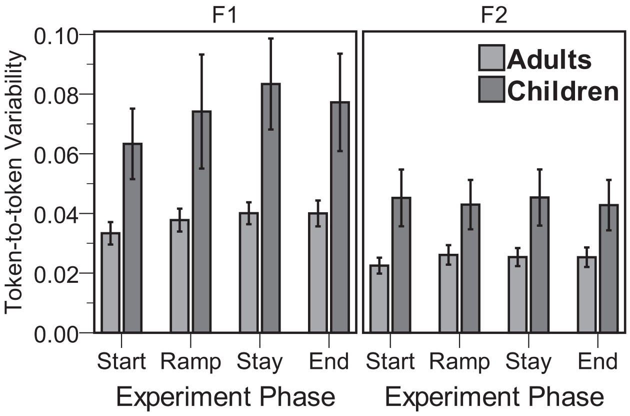


FIG. 3. Mean and across-subject token-to-token variability with 95% confidence intervals of both speaker groups in four experimental phases, separately for F1 and F2.

D. Individual speaker differences in responses of compensation and adaptation

The responses of individual speakers to auditory perturbation during the compensation and adaptation phases were categorized into whether individual speakers showed a significant change in the *same* direction of perturbation in the ramp and stay phases, the *opposite* direction of perturbation in the ramp and stay phases, or *no* significant change. The group comparisons are summarized in Fig. 4, shown by formant and experimental phase. The statistical results of the Chi-square tests revealed that there were no proportional differences in behavior across the two groups during compensation of F1: $\chi^2(2, N = 73) = 2.808$, $p = 0.246$ and F2: $\chi^2(2, N = 73) = 1.879$, $p = 0.391$, as well as adaptation of F1: $\chi^2(2, N = 73) = 4.179$, $p = 0.124$ and F2: $\chi^2(2, N = 73) = 2.528$, $p = 0.283$.

In a subsequent step, the strengths of the categorized response directions were compared across groups. The *post hoc* pairwise comparisons of a series of two-way analyses of variance (ANOVAs) with factors group and direction with normalized F1 and normalized F2 as dependent variables were used to explore differences between adults and children for each of the corrective, following, or neutral categories, separately for F1 and F2. For the F1 in the stay phase, no effect of group $[F(1,1784) = 2.23$, $p = 0.136]$ and a significant effect of direction $[F(2,1784) = 228.2$, $p < 0.001]$, as well as a significant group × direction interaction effect $[F(2,1784) - 21.5$, $p < 0.001]$, were found. The *post hoc* results indicated that both the counteracting ($p < 0.001$) and following ($p = 0.005$) responses in the stay phase were stronger in children compared to adults. No group difference was found in the neutral direction ($p = 0.710$). For the end phase, significant effects of group $[F(1,1562) = 9.34$, $p = 0.002]$, direction $[F(2,1562) = 220.4$, $p < 0.001]$, and group × direction $[F(2,1562) = 25.8$, $p < 0.001]$ were found. Counteracting responses were stronger in children compared to adults ($p < 0.001$), whereas following ($p = 0.338$) and neutral ($p = 0.664$) responses did not differ across groups.

The results for F2 in the stay phase showed no effect of group $[F(1,1783) = 3.00$, $p = 0.083]$, a significant effect of direction $[F(2,1783) = 144.9$, $p < 0.001]$, and a significant group × direction interaction $[F(2,1783) = 4.53$, $p = 0.011]$. *Post hoc* compensatory ($p = 0.064$) and following ($p = 0.201$) responses did not differ across groups, whilst a significant group difference was present for the neutral responses ($p = 0.001$), in which the children's responses were disposed toward a counteracting response. In the end phase, significant group $[F(1,1560) = 10.4$, $p = 0.001]$, direction $[F(2,1560) = 258.7$, $p < 0.001]$, and group × direction $[F(2,1560) = 14.5$, $p < 0.001]$ effects were found. Counteracting responses were stronger in children ($p < 0.001$) compared to adults, whereas group differences were absent for the following ($p = 0.100$) and neutral ($p = 0.146$) categories.

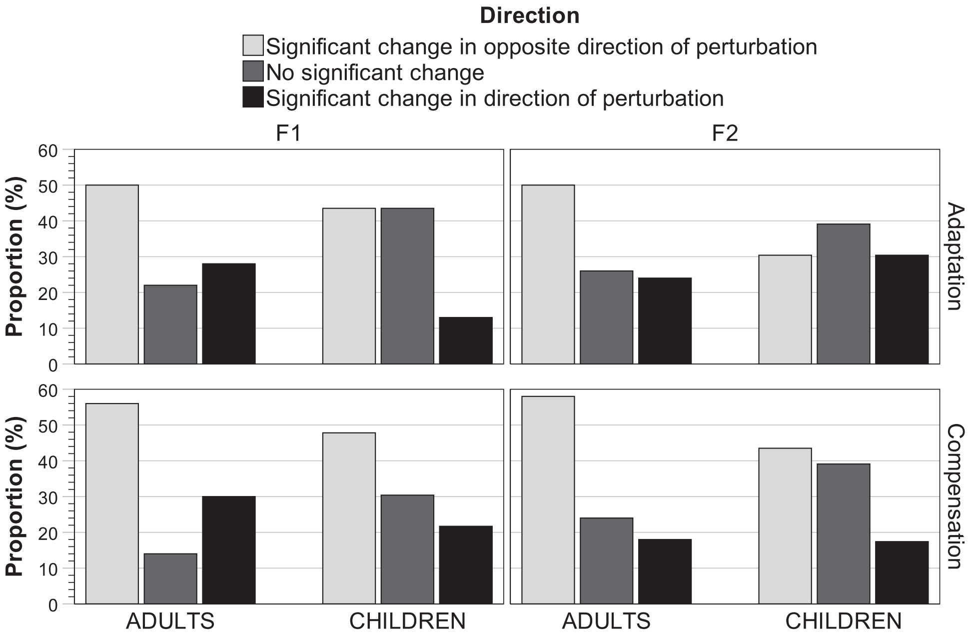


FIG. 4. Group comparisons of the proportion of individual speaker responses to compensation and adaptation, separately for F1 and F2.

IV. DISCUSSION

A. Summary of findings

In the present study, we investigated the nature of compensation for and adaptation to sustained auditory perturbation of F1 and F2 of the vowel /ɪː/ in CVC words in 4–9-year-old children compared to young adults (aged 19–29 years old). The presence and strength of compensation and adaptation effects were assessed by comparing the two groups in average formant ratios during the stay and end phases of the experiment. The results showed that the children were able to compensate and adapt in a similar or larger degree compared to the young adults. Whilst no overall group differences were found with respect to the distribution of response types (counteracting, following, or neutral behavior), the group of children showed stronger effects within the counteracting and following behavioral categories. Furthermore, we assessed the robustness of the responses by investigating token-to-token variability across both groups and categorizing individual responses to perturbation for each speaker. We found that, whereas the group of children demonstrated a higher overall token-to-token variability compared to the adults, both groups displayed comparable levels of relative variability across the four experimental phases. In the following, these aspects are expanded on.

B. Compensation and adaptation across groups

Within the group of children, both the younger children with a shorter experimental program as well as the older children participating in the longer paradigm showed significant responses to auditory perturbation. The presence of significant compensation and adaptation effects in the group of children indicates that the length of the ramp and stay phases were adequate; that is, the paradigm contained enough trials to induce short-term training and learning effects, even during the shorter program designed for the younger among the children. This is further supported by considering the stronger first formant compensatory and adaptive responses in the group of children compared to the averaged responses in the adult group.

1. First formant

With respect to the first formant, it was found that the effects of compensation and adaptation were stronger for children as compared to the young adults. The stronger compensatory responses in the group of children suggest that the children may exhibit stronger training effects as compared to adults. Although we did not predict higher compensatory effects for children, the fact that they were able to compensate for real-time formant shifts to at least a similar degree supports our hypothesis that 4-year-olds are already capable of real-time feedback control in a similar way as adults. Daliri *et al.* (2018) reported an absence of age-related group differences in the strength of response during the stay phase; that is, in their study children did not show stronger compensatory effects compared to adults. This may be mediated by the considerably smaller group sizes they employed (20 children and 14 adults versus 23 children and 50 adults in the present study) as well as the older age of the children participating in their study (age range 7–11 years). Evidence for stronger learning effects can be derived from the adaptation results, in which the children display stronger adaptation responses, supporting our hypothesis of stronger learning effects, greater plasticity, and increased capacity of modifying stored speech motor programs. These adaption results suggest that sensorimotor programs are less ingrained in children and therefore more readily updated based on new information. These results may also suggest a stronger reliance on auditory feedback for children as compared to adults, who in turn might rely relatively more on somatosensory feedback (Lametti *et al.*, 2012). At any rate, the current findings underscore the important role of auditory feedback in speech sound acquisition and the control of speech (Houde and Nagarajan, 2011).

2. Second formant

No specific group-differentiating compensation and adaptation characteristics were found for the second formant. Inspection of the perturbation strengths for both formants indicated that the overall response effect was stronger for F1 (8.4%) compared to F2 (5.6%) when averaged over participants (both adults and children), something that has been reported in other studies (Cai *et al.*, 2010; Daliri *et al.*, 2018). Overall reported discrimination thresholds, expressed as just noticeable differences (JNDs) of adults, vary between 1% and 2.9% for F1, and 1% and 1.3% for F2 (Kewley-Port *et al.*, 1996; Lyzenga and Horst, 1998). Little is known about children's discrimination thresholds for F1 and F2. Bradlow *et al.* (1999) and Koch *et al.* (1999) reported a discrimination threshold for F3 of about 3.2%–3.7% for children, whereas Koch *et al.* (1999) reported a threshold of about 3.4% for young adults, suggesting that children operate largely within the same discrimination range as adults. As the extent of perturbation (25% for F1, 12.5% for F2) was much larger than the reported JNDs of both F1 and F2 frequencies, it can be ruled out that the perturbation effect of F2 was too weak. A possible explanation for the differences in compensatory and adaptive strength between F1 and F2 might be found in the manipulation direction of the vowel used in this study. The resulting sound after perturbation of the near-close near-front lax vowel /ɪː/ came closest to the Dutch open-mid vowel /ɛ/, which, on average, differs from /ɪː/ largely in F2 and less so in F1 (Adank *et al.*, 2004). Therefore, listeners might have a stronger auditory acuity for F1 changes, resulting in stronger responses to F1 perturbations (Niziolek and Guenther, 2013). As seen above and similar to findings in previous studies (Houde and Jordan, 2002; Villacorta *et al.*, 2007; Shiller *et al.*, 2010), compensation and adaptation in motor output are incomplete. An important reason for incomplete responses specific for the current study is the fact that neutral and following responses were not discarded but contributed to the group averages. These responses reduced the overall compensatory strengths. Additional reasons have been discussed in Shiller *et al.* (2009), for example. The incomplete responses might be due to articulatory constraints, speaker-specific variation in auditory acuity, and/or in reliance on feedback versus feedforward control mechanisms, as well as due to the possible impact of unaltered somatosensory targets that prevent compensation taking place, i.e., whilst auditory feedback indicates that the segment being produced is incorrect, somatosensory feedback indicates that the execution of articulatory plans is on target, creating an incongruence in the feedback systems that is only partially resolved during compensation (Shiller *et al.*, 2009; Mitsuya *et al.*, 2015). In addition, it has been found that close vowels, including the close-mid /ɪː/ vowel used in the current study, elicit smaller compensatory responses compared to open vowels. For close vowel types, the strong lingual sensation might supply the feedback controller with rich somatosensory information specifying well-defined articulatory postures and consequently down-weighing the auditory feedback for controlling the production of these vowels (Mitsuya *et al.*, 2015).

C. Token-to-token variability across groups

Many studies on auditory perturbation of formants report higher within-speaker token-to-token variability in children compared to adults (Ménard *et al.*, 2008; MacDonald *et al.*, 2012; Shiller *et al.*, 2010, among others). In accordance with these studies, the current study found higher token-to-token variability in both F1 and F2 in the group of children. It has been speculated that the higher variability in young talkers might be because children are less skilled in consistently producing identical vocal tract configurations (Ménard *et al.*, 2008). Previous studies have also suggested that children may be disproportionately affected by an altered acoustic signal compared to adult speakers, showing even higher variability in token-to-token production (Hazan and Barrett, 2000; Shiller and Rochon, 2014). The findings of the current study do not support this notion; for the group of children, a significant effect of phase was absent, along with the interaction effect of group × phase, showing that token-to-token variability was not disproportionately larger in the group of children during the stay or end phase compared to the start phase. The added challenge of perturbed auditory feedback to auditory-motor integration did not negatively affect the consistency of vowel productions in young talkers. The results imply that outcome measures of token-to-token variability do not reflect the demands of auditory-motor integration in typically developing children but rather express a general immaturity in speech motor execution leading to the presence of background noise across all experimental phases. In other words, the higher variability is not borne out of task difficulties, perceptual limitations, or underdeveloped sensorimotor integration (although some of these issues might still exist), but the higher variability exists because of variability in production (MacDonald *et al.*, 2012; Scheerer *et al.*, 2016). In addition, the higher production variability in children might potentially affect their perceived reliability of the sensory input and impact sensory learning, which in turn may lead to reduced perceptual compensation effects (Shiller *et al.*, 2010). Because the children in the current study did not show smaller compensatory effects, it may be assumed that the increased production variability did not influence the children' real-time feedback control in a meaningful way. Taken together, as evidenced from their strong adaptive responses, we therefore pose that 4–9-year-olds have adultlike internal models that can be utilized for speech motor learning.

D. Individual speaker differences in responses of compensation and adaptation

Differences between the two speaker groups were further explored by investigating speaker-individual responses to perturbation within each group. Although the results showed that across groups the proportion of speakers were not significantly different in terms of responses to auditory feedback perturbation, a considerable between-speaker variation was noted. A significant proportion of both groups did not show a response or showed a response that enhanced the perturbation. Proportions of individual responses opposing perturbation varied between 30% (F2 adaptation in children) and 58% (F2 compensation in adults). In comparison, Cai *et al.* (2010) reported that 58% of their control speakers (only adults) performed as expected. The process of compensation and adaptation has been found to be largely automatic, and speakers have been found to show compensation even when explicitly instructed not to compensate for the perturbation (Munhall *et al.*, 2009). It is therefore highly unlikely that an amplifying response to the feedback perturbation was related to speaker-specific intentional differences in task interpretation or differences in overt task execution strategies. Test-retest experiments at sufficiently spaced-out time intervals documenting participants' consistency in response direction and strength to perturbation might shed further light on this issue.

Hain *et al.* (2000) investigated opposing and following responses to randomly applied voice F0 shifts and speculated that findings of following responses occur in conditions when the audio-vocal system treats the altered feedback as an external referent and attempts to match it by changing the vocal output to follow the feedback (Hain *et al.*, 2000, cf. Behroozmand *et al.*, 2012). A similar mechanism might be applicable with respect to the online monitoring and control of formants; however, since the present study did not feature randomly applied full perturbations but did feature a gradual introduction of the perturbation in the ramp phase, it is highly unlikely that the feedback was interpreted as an external referent. A possible explanation for the perturbation-enhancing behavior might be that the implemented formant shifts have caused a "target drift"; That is, the sensory motor system interprets the formant shifts as adjustments of the intended auditory outcome (Terband *et al.*, 2014). The auditory targets are then updated accordingly, and the shifted formant settings become the auditory target for the next trial. Findings on vocal control by Behroozmand *et al.* (2012) suggest that following and opposing responses are mediated by different neural mechanisms, as per their study, opposing responses were affected by stimulus magnitude, whereas following responses were not. The absence of such an effect for following responses suggests that the neural mechanisms controlling these responses are different from those controlling the opposing responses (Behroozmand *et al.*, 2012). The auditory-motor system might have used the efference copies of motor commands to estimate the degree of discrepancy between intended (predicted) and actual (sensory) feedback in order to identify the source of the production error (see Wolpert *et al.*, 2011, for a review). In the present study, speakers might have employed a strategy that was not aimed at neutralizing but rather at matching the perceived formant error without being controlled by efference mechanisms. The current results and their tentative interpretation suggest that future studies should include varying magnitudes of perturbation to explore this issue further.

The analysis of group differences at the level of individual response categories revealed that children generally displayed stronger corrective and amplifying responses to perturbations. An explanation for the differences in responses to perturbation across age groups might be found in the different strategies that were employed with respect to focusing on somatosensory feedback versus focusing on auditory feedback (Katseff *et al.*, 2012). During development, this feedback trade-off might start with a predominant reliance on auditory feedback, shift toward reliance on somatosensory feedback, and stabilize into adolescence and adulthood (Daliri *et al.*, 2018). If children bear a greater reliance on auditory feedback, their reaction to auditory manipulations (whether following or counteracting) may be expected to be stronger compared to adults. When considering the effect of formant perturbations on the feedback subsystems, it is assumed that formant perturbations create errors in the auditory feedback control subsystem but not in the somatosensory feedback subsystem. Compensation to auditory perturbation will lead to changes in somatosensory feedback, resulting in perceived somatosensory errors and subsequent corrective commands aimed at counteracting the compensatory response to feedback (Guenther, 2016). If children rely less on somatosensory feedback, corrective responses originating from the somatosensory feedback subsystem will be weaker, enabling stronger following or counteracting reactions to auditory perturbation to take place. Future studies should investigate potential trade-offs between different types of feedback across populations of different ages, cf. Mitsuya *et al.* (2015) and Max and Maffett (2015).

V. CONCLUSIONS

The results of this auditory feedback perturbation study showed that Dutch children aged 4–9 years old were able to compensate and adapt in a similar or larger degree compared to adults, possibly indicating a weaker engrainment of existing speech sound representations and stronger learning effects in children. Although both token-to-token and within-token variability in F1 and F2 was higher in children compared to the adults, variability was not disproportionately higher in the perturbed trials. Hence, the added challenge of perturbed auditory feedback to auditory-motor integration did not influence variability, implying that outcome measures of token-to-token variability do not reflect the demands of auditory-motor integration. At the speaker level, the consistency of individual responses to auditory perturbation was inventoried. Whereas large individual differences were found across speakers, the proportion of speakers displaying a consistent compensatory and adaptive response was similar in both groups under investigation. At different phases in the experimental paradigm, children showed stronger responses to perturbation, irrespective of direction. Taken together, the results of this study indicate that mechanisms of auditory feedback control and motor learning are stronger in young children compared to adults, underlining the still higher plasticity of the auditory-motor system during childhood.

ACKNOWLEDGMENTS

This study was funded by the Netherlands Organization for Scientific Research (NWO-VENI Grant No. 275-89-016 awarded to H.T.). The authors gratefully thank the participants, especially the children and their parents or caretakers for their time and effort, Anniek van Doornik-van der Zee and Margoke Nijjssen for their help with participant recruitment, and Shanqing Cai for his help with the Audapter software.

Adank, P., Van Hout, R., and Smits, R. (**2004**). “An acoustic description of the vowels of Northern and Southern Standard Dutch,” J. Acoust. Soc. Am. **116**(3), 1729–1738.

Behroozmand, R., Korzyukov, O., Sattler, L., and Larson, C. R. (**2012**). “Opposing and following vocal responses to pitch-shifted auditory feedback: Evidence for different mechanisms of voice pitch control,” J. Acoust. Soc. Am. **132**(4), 2468–2477.

Boersma, P., and Weenink, D. (**2013**). “Praat: Doing phonetics by computer (version 5.3.42) [computer program],” http://www.praat.org/ (Last viewed March 10, 2013).

Bradlow, A. R., Kraus, N., Nicol, T. G., McGee, T. J., Cunningham, J., Zecker, S. G., and Carrell, T. D. (**1999**). “Effects of lengthened formant transition duration on discrimination and neural representation of synthetic CV syllables by normal and learning-disabled children,” J. Acoust. Soc. Am. **106**(4), 2086–2096.

Cai, S., Beal, D. S., Ghosh, S. S., Tiede, M. K., Guenther, F. H., and Perkell, J. S. (**2012**). “Weak responses to auditory feedback perturbation during articulation in persons who stutter: Evidence for abnormal auditory-motor transformation,” PLoS One **7**(7), e41830.

Cai, S., Ghosh, S. S., Guenther, F. H., and Perkell, J. (**2010**). “Adaptive auditory feedback control of the production of formant trajectories in the Mandarin triphthong /iau/ and its pattern of generalization,” J. Acoust. Soc. Am. **128**(4), 2033–2048.

Daliri, A., Wieland, E. A., Cai, S., Guenther, F. H., and Chang, S. E. (**2018**). “Auditory-motor adaptation is reduced in adults who stutter but not in children who stutter,” Dev. Sci. **21**(2), e12521.

Guenther, F. H. (**2016**). *Neural Control of Speech* (MIT Press, Cambridge, MA), 410 pp.

Guenther, F. H., Hampson, M., and Johnson, D. (**1998**). “A theoretical investigation of reference frames for the planning of speech movements,” Psychol. Re. **105**(4), 611–633.

Guenther, F. H., and Perkell, J. S. (**2004**). “A neural model of speech production and its application to studies of the role of auditory feedback in speech,” in *Speech Motor Control in Normal and Disordered Speech*, edited by B. Maassen, R. Kent, H. F. M. Peters, P. H. H. M. van Lieshout, and W. Hulstijn (Oxford University Press, Oxford, UK), pp. 29–50.

Hain, T. C., Larson, C. R., Burnett, T. A., Kiran, S., and Singh, S. (**2000**). “Instructing participants to make a voluntary response reveals the presence of two vocal responses to pitch-shift stimuli,” Exp. Brain Res. **130**, 133–141.

Hazan, V., and Barrett, S. (**2000**). “The development of phonemic categorization in children aged 6–12,” J. Phonetics **28**(4), 377–396.

Houde, J. F., and Jordan, M. I. (**2002**). “Sensorimotor adaptation of speech I: Compensation and adaptation,” J. Speech Lang. Hear. Res. **45**(2), 295–310.

Houde, J. F., and Nagarajan, S. S. (**2011**). “Speech production as state feedback control,” Front. Hum. Neurosci. **5**, 82.

Katseff, S., Houde, J., and Johnson, K. (**2012**). “Partial compensation for altered auditory feedback: A tradeoff with somatosensory feedback?,” Lang. Speech **55**(2), 295–308.

Kewley-Port, D., Li, X., Zheng, Y., and Neel, A. T. (**1996**). “Fundamental frequency effects on thresholds for vowel formant discrimination,” J. Acoust. Soc. Am. **100**(4), 2462–2470.

Koch, D. B., McGee, T. J., Bradlow, A. R., and Kraus, N. (**1999**). “Acoustic-phonetic approach toward understanding neural processes and speech perception,” J. Am. Acad. Audiol. **10**(6), 304–318.

Lametti, D. R., Nasir, S. M., and Ostry, D. J. (**2012**). “Sensory preference in speech production revealed by simultaneous alteration of auditory and somatosensory feedback,” J. Neurosci. **32**(27), 9351–9358.

Lyzenga, J., and Horst, J. W. (**1998**). “Frequency discrimination of stylized synthetic vowels with two formants,” J. Acoust. Soc. Am. **104**(5), 2956–2966.

MacDonald, E. N., Johnson, E. K., Forsythe, J., Plante, P., and Munhall, K. G. (**2012**). “Children’s development of self-regulation in speech production,” Curr. Biol. **22**(2), 113–117.

Max, L., and Maffett, D. G. (**2015**). “Feedback delays eliminate auditory-motor learning in speech production,” Neurosci. Lett. **591**, 25–29.

Ménard, L., Perrier, P., and Aubin, J. (**2013**). “The role of auditory feedback in speech development: A study of compensation strategies for a lip-tube perturbation,” Proc. Mtgs. Acoust. **19**, 060181.

Ménard, L., Perrier, P., Aubin, J., Savariaux, C., and Thibeault, M. (**2008**). “Compensation strategies for a lip-tube perturbation of French [u]: An acoustic and perceptual study of 4-year-old children,” J. Acoust. Soc. Am. **124**(2), 1192–1206.

Mitsuya, T., MacDonald, E. N., Munhall, K. G., and Purcell, D. W. (**2015**). “Formant compensation for auditory feedback with English vowels,” J. Acoust. Soc. Am. **138**(1), 413–424.

Munhall, K. G., MacDonald, E. N., Byrne, S. K., and Johnsrude, I. (**2009**). “Talkers alter vowel production in response to real-time formant perturbation even when instructed not to compensate,” J. Acoust. Soc. Am. **125**(1), 384–390.

Niziolek, C. A., and Guenther, F. H. (**2013**). “Vowel category boundaries enhance cortical and behavioral responses to speech feedback alterations,” J. Neurosci. **33**(29), 12090–12098.

Perkell, J. (**2012**). “Movement goals and feedback and feedforward control mechanisms in speech production,” J. Neurolinguist. **25**(5), 382–407.

Perkell, J., Denny, M., Lane, H., Guenther, F., Matthies, M. L., Tiede, M., Vick, J., Zandipour, M., and Burton, E. (**2007**). “Effects of masking noise on vowel and sibilant contrasts in normal-hearing speakers and postlingually deafened cochlear implant users,” J. Acoust. Soc. Am. **121**(1), 505–518.

Perkell, J., Matthies, M., Lane, H., Guenther, F., Wilhelms-Tricarico, R., Wozniak, J., and Guiod, P. (**1997**). “Speech motor control: Acoustic goals, saturation effects, auditory feedback and internal models,” Speech Commun. **22**(2-3), 227–250.

Scheerer, N. E., Jacobson, D. S., and Jones, J. A. (**2016**). “Sensorimotor learning in children and adults: Exposure to frequency-altered auditory feedback during speech production,” Neuroscience **314**, 106–115.

Shiller, D., Gracco, V., and Rvachew, S. (**2010**). “Auditory-motor learning during speech production in 9–11-year-old children,” PLoS One **5**(9), e12975.

Shiller, D., Sato, M., Gracco, V., and Baum, S. (**2009**). “Perceptual recalibration of speech sounds following speech motor learning,” J. Acoust. Soc. Am. **125**, 1103–1113.

Shiller, D. M., and Rochon, M. L. (**2014**). “Auditory-perceptual learning improves speech motor adaptation in children,” J. Exp. Psychol.: Hum. Percept. Perform. **40**(4), 1308–1315.

Terband, H., van Brenk, F., and van Doornik-van der Zee, J. C. (**2014**). “Auditory feedback perturbation in children with developmental speech disorders,” J. Commun. Disord. **51**, 64–77.

Villacorta, V. M., Perkell, J. S., and Guenther, F. H. (**2007**). “Sensorimotor adaptation to feedback perturbations of vowel acoustics and its relation to perception,” J. Acoust. Soc. Am. **122**(4), 2306–2319.

Walsh, B., Smith, A., and Weber-Fox, C. (**2006**). “Short-term plasticity in children’s speech motor systems,” Dev. Psychobiol. **48**(8), 660–674.

Wolpert, D. M., Diedrichsen, J., and Flanagan, J. R. (**2011**). “Principles of sensorimotor learning,” Nat. Rev. Neurosci. **12**(12), 739–751.