ASHA American Speech-Language-Hearing Association

JSLHR

Research Article

Modeling Responses to Auditory Feedback Perturbations in Adults, Children, and Children With Complex Speech Sound Disorders: Evidence for Impaired Auditory Self-Monitoring?

Hayo Terband[a] and **Frits van Brenk**[b,c]

[a] Department of Communication Sciences and Disorders, University of Iowa, Iowa City [b] Faculty of Humanities, Department of Languages, Literature and Communication & Institute for Language Sciences, Utrecht University, the Netherlands [c] Department of Communicative Disorders and Sciences, University at Buffalo, NY

ARTICLE INFO

Article History:
Received June 26, 2022
Revision received November 4, 2022
Accepted January 30, 2023

Editor-in-Chief: Cara E. Stepp
Editor: Soo-Eun Chang

https://doi.org/10.1044/2023_JSLHR-22-00379

ABSTRACT

Purpose: Previous studies have found that typically developing (TD) children were able to compensate and adapt to auditory feedback perturbations to a similar or larger degree compared to young adults, while children with speech sound disorder (SSD) were found to produce predominantly following responses. However, large individual differences lie underneath the group-level results. This study investigates possible mechanisms in responses to formant shifts by modeling parameters of feedback and feedforward control of speech production based on behavioral data.
Method: SimpleDIVA was used to model an existing dataset of compensation/adaptation behavior to auditory feedback perturbations collected from three groups of Dutch speakers: 50 young adults, twenty-three 4- to 8-year-old children with TD speech, and seven 4- to 8-year-old children with SSD. Between-groups and individual within-group differences in model outcome measures representing auditory and somatosensory feedback control gain and feedforward learning rate were assessed.
Results: Notable between-groups and within-group variation was found for all outcome measures. Data modeled for individual speakers yielded model fits with varying reliability. Auditory feedback control gain was negative in children with SSD and positive in both other groups. Somatosensory feedback control gain was negative for both groups of children and marginally negative for adults. Feedforward learning rate measures were highest in the children with TD speech followed by children with SSD, compared to adults.
Conclusions: The SimpleDIVA model was able to account for responses to the perturbation of auditory feedback other than corrective, as negative auditory feedback control gains were associated with following responses to vowel shifts. These preliminary findings are suggestive of impaired auditory self-monitoring in children with complex SSD. Possible mechanisms underlying the nature of following responses are discussed.

The acquisition and production of speech sounds is reliant on self-monitoring of the auditory signal. Studies using experimental paradigms in which auditory feedback is being perturbed in real time have highlighted that such unexpected spectral (usually formant) perturbations elicit an almost direct response. Speakers usually produce such responses in the opposite direction of the manipulation to attempt to resolve the apparent mismatch between the intended produced sound and the manipulated perceptual result (e.g., Houde & Jordan, 1998, 2002; Tourville et al., 2008; Villacorta et al., 2007). However, several studies reported that, when assessing individual speakers within a

---

Correspondence to Hayo Terband: hayo-terband@uiowa.edu. ***Disclosure:*** *The authors have declared that no competing financial or nonfinancial interests existed at the time of publication.*

given group, a sizeable proportion of individuals fail to show a consistent response; that is, some speakers fail to react to perturbed feedback or show perturbation changes that amplify and exacerbate the perturbation manipulation. These inconsistent responses have been reported for formant and pitch perturbations in adult speakers (e.g., Behroozmand et al., 2012; Burnett et al., 1998; Cai et al., 2010; Villacorta et al., 2007) and in typically developing (TD) children (van Brenk & Terband, 2020). This does not detract from the observation that TD children are able to notice and act on the feedback perturbation with similar performance compared to adults. In contrast, children with speech sound disorder (SSD) have shown predominantly following responses to formant perturbations (Terband et al., 2014). Children with SSD have difficulties accurately producing certain speech sounds beyond the typical age of acquisition for that particular speech sound. SSDs in children usually refer to an umbrella term encompassing phonological disorder (impaired comprehension of a language's sound system and the rules that govern the sound combinations), phonetic articulation disorder (atypical speech sound production characterized by substitutions, omissions, additions, or distortions), developmental stuttering (a neurodevelopmental disorder characterized by disruptions to fluency of speech), and childhood apraxia of speech (a neurological childhood SSD in which the precision and consistency of movements underlying speech are impaired in the absence of neuromuscular deficits; American Speech-Language-Hearing Association, 2007; Bauman-Waengler, 2020). As the process of successful correction of perturbations relies on a careful balancing of intertwined feedback and feedforward control mechanisms, it is feasible that children with SSD might have a disturbed balance between these control mechanisms. Such balance is essential for short-term learning and the acquisition of new speech sounds. Unraveling and identifying these faulty control mechanisms leading to disordered short-term learning is a further step in identifying some of the underlying deficits in pediatric SSD, as well as a possible guidance in diagnosis and treatment of this population. Studies investigating auditory feedback behavior in TD children and children with SSD are reviewed in the following section.

Auditory Feedback Perturbation in Children With and Without SSDs

Auditory feedback plays an important role in the acquisition of novel sensorimotor programs (Guenther et al., 1998; Menard et al., 2013), as evidenced by experiments quantifying measures of feedback control and motor learning in children. These experiments entail the shifting of a speaker's own vowel formants in real time, such that a mismatch is created between the articulatory action and the auditory result. The unexpected perturbation of auditory feedback during speech production elicits *compensation*, a response often in the opposite direction of the perturbation to maintain the intended auditory outcome. Adaptation experiments typically consist of a series of experimental trials starting with a baseline, followed by a ramp phase in which the perturbation is gradually introduced, a stay or hold phase where maximum perturbation is applied, and an end phase where perturbation is abruptly stopped. Such sustained application of formant shifts has been shown to also cause the speech motor system to *adapt* to the perturbation and modify the stored speech motor programs. The responses in the stay or hold phase are quantified to represent a combination of compensation (reflecting real-time feedback control and motor correction) and adaptation (reflecting motor learning in which the feedforward state of the vowel motoric plan is updated for future production control). The persistence of the changes in the produced formants is quantified by measuring the after-effect when the perturbation is abruptly removed in the end phase (Cheung et al., 2021; van Brenk & Terband, 2020; Villacorta et al., 2007).

Studies investigating sustained auditory feedback perturbations of formants in vowels have shown that although children before 4 years of age do not show compensation/adaptation (MacDonald et al., 2012), children aged 4–12 years are able to adapt to auditory feedback perturbation to a similar degree compared to young adults. In fact, a subgroup of children displayed stronger compensatory and amplifying effects in first and second formant (F1 and F2) productions compared to young adults, suggesting that learning mechanisms are stronger at a younger age while at the same time less directional due to less-ingrained existing representations (Cheung et al., 2021; van Brenk & Terband, 2020). With respect to feedback control and motor learning in children with SSD, the picture is less clear, at least partially due to the paucity in literature on auditory feedback perturbation in children with SSD. To our knowledge, thus far, only one study has addressed this line of research (Terband et al., 2014). Findings from this study indicated, as mentioned above, that TD children generally showed compensatory effects; that is, they successfully adjusted their vowel formants in the direction opposite to the vowel shifts introduced by perturbation. Children with complex SSD were also able to detect incongruencies of F1 and F2 in vowels during auditory feedback perturbation, but the majority of children with SSD adjusted their formants in the direction *following* the perturbation. They thus amplified the effects of the vowel shifts, particularly with respect to F1 and, to a lesser extent, F2. Similar to the children with typical development, the children with SSD also showed an aftereffect. These findings suggest that SSD does not involve an inability to update speech-motor representations but rather point in the direction of deficits in sensorimotor integration and impaired internal models. Such deficits may form the core of SSD, and their presence may underline the structural role of disturbed feedforward learning and feedback monitoring underlying impairments in children with SSD (Terband et al., 2014).

This study aimed to explore possible mechanisms in responses to formant shifts through a modeling experiment using the SimpleDIVA application, with which parameters of feedback and feedforward control quantifying short-term learning and acquisition effects can be modeled and extracted based on behavioral data from auditory feedback perturbation experiments (Kearney & Guenther, 2019; Kearney et al., 2020).

SimpleDIVA to Quantify Feedback and Feedforward Control

Derived from the Directions into Velocities of Articulators (DIVA) computational model of speech production (Tourville & Guenther, 2011), the application SimpleDIVA has been developed to quantify short-term speech sound learning and acquisition effects based on experimental data collected from auditory feedback perturbation tasks. The application provides a three-parameter mathematical model that quantifies the relative contribution of auditory feedback and somatosensory feedback components involved in online error correction and a feedforward component representing a learning or update rate parameter. Using behavioral data from auditory feedback perturbation experiments as input (e.g., realized vowel formant frequencies or fundamental frequencies), SimpleDIVA fits the optimal values for the feedback and feedforward components (Kearney et al., 2020). The SimpleDIVA installment used in this study (V1.3) allows for modeling parameters to become negative and thus is theoretically able to quantify and account for perturbation-following behavior.

Purpose

The quantification of parameters of feedback and feedforward control derived from auditory feedback perturbation data allows to derive predictions of impairments in speech motor learning in children with SSD. While previous studies have reported indirect evidence for impaired speech motor learning and feedback monitoring in 4- to 8-year-old children with SSD (Terband et al., 2014), to our knowledge, no study has directly quantified these measures in this particular population. The first aim of this study was to elucidate the underlying causes of following responses to formant shifts displayed by children with SSD during auditory feedback perturbation experiments. SimpleDIVA was used to model, quantify, and characterize feedback and feedforward control mechanisms in children with SSD, compared to TD peers and young adults. Furthermore, while previous studies focused on feedback and feedforward control parameters derived from mean group data, to date, these parameters have not yet been studied at the level of individual speakers, be it neurotypical adults, children with TD speech, or children with SSD. A second aim was to quantify control parameters for individual speakers, which allows us to explore further within-group and between-speakers characteristics, as well as to establish the strength of associations between model parameters and behavioral outcome measures for individual speakers, particularly with respect to the children with complex SSD. Here, associations between selected outcome measures of the assessment battery and model parameters were explored to gain a broader understanding of factors related to speech motor learning in both children with TD speech and children with SSD.

Method

Participants

The dataset consisted of speech materials produced by three groups of Dutch speakers: 50 adults (32 female, 18 male; age range: 18–29 years [*M* = 22.3, *SD* = 2.7]), 23 TD children (11 female, 12 male; age range: 4.0–8.7 years [*M* = 5.6, *SD* = 1.4]), and seven children with SSD (four female, three male; age range: 4.8–7.5 years [*M* = 5.7; *SD* = 1.0]). Detailed participant information is available in the works of Terband et al. (2014) and van Brenk and Terband (2020). The young adults were recruited through the Faculty of Humanities student participant pool of Utrecht University, and the TD children were recruited via local schools and acquaintances. None of the participants had current or previous speech or hearing problems. The seven children with SSD were referred by speech pathologists. None of the children with SSD suffered from hearing problems (pure-tone thresholds not exceeding 25 dB HL), language comprehension problems (a score less than 1 *SD* below population average), subnormal intelligence (IQ < 1 *SD* below population average), organic disorders in the orofacial area, gross motor disturbances, or dysarthria. The diagnosis of the children with SSD was established using standardized speech perception, production tests, and a short case history. Detailed background and diagnostic data of children with SSD are presented in Appendix A. The two groups of children were not significantly different in age, $t(26) = 0.163$, $p = .874$; gender, $X^2(1) = 0.190$, $p = .663$; receptive vocabulary (Word Comprehension Quotient

[WCQ] $t(26) = 1.522$, $p = .140$); auditory discrimination (words: $t(25) = -0.403$, $p = .690$; nonwords: $t(26) = -0.100$, $p = .921$), but the group of children with SSD scored significantly lower on intelligibility (Intelligibility In Context Scale [ICS, McLeod et al., 2013]: $t(26) = 3.702$, $p = .001$). The participants were recruited from the western region of the Netherlands and were native speakers of Standard Dutch. Written consent was sought from all adult participants and parents or caretakers of child participants prior to the study.

Experimental Paradigm

For a detailed description of the experimental paradigm, we refer to Terband et al. (2014) and van Brenk and Terband (2020). The dataset was collected in a series of auditory feedback perturbation experiments. In the experimental paradigm, simultaneously, F1 was raised 25% and F2 was lowered 12.5% of the target vowel /ɪː/ in three consonant–vowel–consonant (CVC) words: /bɪːr/ "*bear*," /vɪːr/ "*feather*," and /pɪːr/ "*pear*." A paradigm was used comprising practice, start, ramp, stay/hold, and end phases. The practice phase consisted of nine trials, after which the total number of included experimental trials was 102 (27 start, 24 ramp, 27 stay/hold, 24 end) for all adults and four children above 7 years old (TD: $n = 3$; SSD: $n = 1$). Considering fatigue and attention loss, the paradigm of all other children (TD: $n = 20$; SSD: $n = 6$) comprised a shorter version of 66 experimental trials (15 start, 18 ramp, 18 stay/hold, 15 end). Participants were seated in front of a PC monitor showing images of the three target words. Given the children's age, vowels were required to be elicited by naming pictures of concepts instead of being read orthographically. An animated bird flying over one of the images cued the participant to produce the intended word. Images were displayed in a randomized block design to display either the bear, the pear, or the feather, ensuring masking of the identity of the upcoming target word to limit word preparation by the speaker.

Auditory feedback was manipulated using the software module Audapter, executed in MATLAB (Cai et al., 2010, 2012). Recordings were made by an externally powered lavalier microphone (Audio-Technica AT803b) connected to a Lenovo ThinkPad laptop. Audio was digitized at 16 kHz and 16-bit resolution. Over-ear headphones (Sennheiser HD 380 pro) were used to play back productions. As indicated above, F1 was raised 25% and F2 was lowered 12.5% of the near-close near-front lax vowel /ɪː/ in the three target words, yielding a more open and central vowel.

During the practice phase, participants familiarized themselves with the experimental paradigm and practiced the desired vowel duration (between 300 and 500 ms) and loudness (74–84 dB SPL at a 10-cm microphone distance) to optimize formant tracking and to ensure within-trial compensation could take place. The start phase served as a baseline without perturbation. In the ramp phase, 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/hold phase featured maximum perturbation, and perturbation was suspended at once in the end phase.

Behavioral Data Processing

For each production, the mean F1 and mean F2 were measured from steady-state portions of the produced vowels using custom scripts for Praat (Boersma & Weenink, 2013). Conforming with SimpleDIVA assumptions, formant values were extracted at least 150 ms into the vowel (Kearney et al., 2020). Following Daliri et al. (2018) and Kearney et al. (2020), the extracted formant values in Hertz used as input for modeling were averaged over the three trials within each block (i.e., three target words). The behavioral outcome measures were quantified using normalized formant values. For every speaker and for each of the three words, formant frequencies produced in the start phase were averaged. Then, 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. The resulting normalized mean formant ratio for each speaker was finally multiplied by 100 to arrive at percentages. The total behavioral response for the auditory perturbation was quantified by calculating the difference in normalized formant frequencies between the stay/hold and start phases. As described in the introduction, this measure is a product of online, within-trial compensation and trial-to-trial adaptation. The adaptation component was isolated from within-trial compensation by calculating the differences in normalized formant frequencies between the start phase and the end phase, where the perturbation is suddenly removed.

Modeling Adaptive Behavior

SimpleDIVA V1.3 (Guenther et al., 2019) was used to model the existing behavioral dataset of responses to auditory feedback perturbation (described in detail in Terband et al., 2014, and van Brenk & Terband, 2020). Adjustments for a perturbed auditory signal are dependent on the interaction between feedback control (detecting and correcting F1 and F2 errors within a trial) and feedforward control (updating of motor command for following trial). SimpleDIVA is a 3-free-parameter computational model based on the DIVA model (Guenther, 1994, 2016). SimpleDIVA estimates contributions of feedback and feedforward control mechanisms (auditory feedback control gain [α_A], somatosensory feedback control gain [α_S], and feedforward control/learning rate [λ_{FF}]) by modeling the produced formant values along with the perturbation trajectory.

F_i in a trial (n) is the sum of a feedforward command and sensory feedback-based correction:

$$F_{\mathrm{iproduced}}(n) = F_{i\mathrm{FF}}(n) + \Delta F_{i\mathrm{FB}}(n) \tag{1}$$

The feedback-based correction is based on both auditory and somatosensory errors (the difference between the target formant value [$F_{i\mathrm{T}}$] and the produced formant values based on auditory feedback [$F_{i\mathrm{AF}}$] and somatosensory feedback [$F_{i\mathrm{SF}}$], respectively) detected at the beginning of the production (before feedback control mechanisms contribute), scaled by the gains of the auditory and somatosensory feedback subsystems $\alpha_A$and α_S:

$$\Delta F_{i\mathrm{FB}}(n) = \alpha_A \times (F_{i\mathrm{T}} - F_{i\mathrm{AF}}(n)) + \alpha_S * (F_{i\mathrm{T}} - F_{i\mathrm{SF}}(n)) \tag{2}$$

The feedforward command for the next trial is updated by adding a fraction of the feedback-based corrective command from the current trial, characterized by the feedforward control/learning rate parameter λ_{FF}:

$$F_{i\mathrm{FF}}(n+1) = F_{i\mathrm{FF}}(n) + \lambda_{FF} \times \Delta F_{i\mathrm{FB}}(n) \tag{3}$$

The three parameters were estimated from the subject-averaged F1 and F2 frequency data as reported in the works of van Brenk and Terband (2020; adults and TD children) and Terband et al. (2014; children with SSD). We modeled behavior of speakers both on the group level (i.e., fitting the model to mean data across speakers) and individually (fitting the model to data of each speaker separately). SimpleDIVA modeling parameters α_A, α_S, and λ_{FF} were allowed to vary between −1 and 1. The default regularization terms of 0.001 combined with mean data fits were used to deal with between- and within-subject variability. Regularization functions as a damper preventing parameter values from drifting too far from "central" values (Kearney et al., 2020).

As mentioned earlier, 26 out of 30 children completed a shorter version of the experiment. Their experimental program included 66 trials instead of the 102 trials in the longer program for the adults and the four older children. To assess whether the different number of datapoints had an influence on parameter estimation by SimpleDIVA, the long programs of the adults were pruned to reduce the number of datapoints while maintaining a similar pattern of behavioral response. For each of the four phases (start – ramp – stay/hold – end), individual trials were proportionally removed at fixed and regular intervals. For every adult, model parameters for the long, regular program were correlated with model parameters for the pruned program. Additionally, we investigated a potential effect of program length by replacing the start, stay/hold, and end phases of the pruned program data with truncated phases to construct a short program similar to that of the younger children. The ramp phase was not truncated (but kept pruned) to keep the same amount of perturbation at the end of the ramp phase.

Statistical Analysis

The primary outcome measures of this study were the three estimated model parameters: auditory feedback control gain (α_A), somatosensory feedback control gain (α_S), and feedforward control/learning rate (λ_{FF}). First, we compared group level mean data fits. We then analyzed the parameters estimates from data fits of individual speakers.

Model fits were evaluated by means of the root-mean-square error (RMSE), a normalized value capturing goodness-of-fit of the model, and Pearson r, which describes the relationship between the data and model fit (Kearney et al., 2020). Model fits for individual data with an RMSE above .1 and/or an r below .3 were considered insufficient and were discarded for the remainder of the analyses.

The potential effects of a different number of datapoints and divergent program lengths on SimpleDIVA parameter fits were assessed by means of Cronbach's α. Modeling parameters obtained from individual participant data in the adult speaker group obtained from the original length paradigms were compared with parameters obtained from the shortened and truncated paradigms.

The three model parameters were then compared as outcome measures across groups. Group comparisons of parameter estimates for the individual data were carried out by means of a multivariate analysis of variance for the three model parameters (α_A, α_S, and λ_{FF}), followed up by univariate tests for each parameter separately. In addition, correlations between model parameter estimates were explored.

Subsequently, the strengths of relationships between the SimpleDIVA parameter estimates and the corresponding speakers' responses to F1 and F2 shifts in the perturbation experiment were investigated—both pooled over groups and for each group separately. For α_A and α_S, we calculated correlations between the model parameter estimates for each speaker and their mean amount of adaptation in the perturbation experiment. For λ_{FF}, however, we calculated correlations with the response magnitude (i.e., the absolute value) of mean adaptation in the perturbation experiment. This transformation to absolute values is necessary since λ_{FF}

determines the fraction of the feedback-based corrective command with which the motor command is updated for the next trial, regardless of whether the corrective command is positive (counteracting the perturbation) or negative (following or amplifying the perturbation).

Finally, for both pediatric groups, we calculated correlations of model outcome measures with scores on selected tasks from the standardized speech/language production and perception tests that are particularly demanding for the auditory and/or motor systems. These were the proportion of syllable-initial consonants correct (PCCI) during word and nonword repetition, word and nonword auditory discrimination, and sequential fast oral motor movements.

An α value of .05 was used for all analyses. Correlations were calculated by means of the relatively conservative Spearman ρ, given its robustness to outliers and the small sample sizes of the groups of children.

Results

Evaluation of Model Fits

A good model fit, expressed as *r*, was found for *mean* data of the groups of adults (*n* = 50, RMSE = 0.004, *r* = .90) and a moderate fit for TD children (*n* = 23, RMSE = 0.024, *r* = .52), but the model fit was weak for the SSD mean group data (*n* = 7, RMSE = 0.039, *r* = .41). Figures of group-data model fits are included in Appendix B.

Average model fits for *individual* data were weak to moderate for the adults (*n* = 50; RMSE: *M* = 0.025, *SD* = 0.006; *r*: *M* = .53, *SD* = .23), the TD children (*n* = 23; RMSE: *M* = 0.043, *SD* = 0.016; *r*: *M* = .48, *SD* = .27), and the children with SSD (*n* = 7; RMSE: *M* = 0.051, *SD* = 0.017; *r*: *M* = .58, *SD* = .14). The individually modeled data for seven adults and seven children in the TD group yielded model fits with an *r* below .3. These model fits were considered insufficient, and associated data were discarded for the remainder of the analyses. Examples of good, average, and poor model fits of individual data are included in Appendix C.

For the data of the remaining individuals, average model fits were moderate for the adults (*n* = 43; RMSE: *M* = 0.024, *SD* = 0.005; *r*: *M* = .59, *SD* = .18), the TD children (*n* = 16; RMSE: *M* = 0.048, *SD* = 0.016; *r*: *M* = .61, *SD* = .20), and the children with SSD (*n* = 7; RMSE: *M* = 0.051, *SD* = 0.017; *r*: *M* = .58, *SD* = .14). The two groups of children remained equivalent on age, *t*(21) = 0.139, *p* = .89; gender, $X^2(1) = 0.028$, *p* = .87; receptive vocabulary (WCQ: *t*(21) = 1.349, *p* = .19); and auditory discrimination (words: *t*(20) = −0.291, *p* = .77; nonwords: *t*(21) = −0.350, *p* = .73). The scores of the children with SSD remained lower on intelligibility (ICS: *t*(21) = 4.492, *p* < .001).

With respect to the potential effect of number of datapoints on outcome measures, the comparison between the model parameters for the original individual data of the adults and the pruned dataset with a reduced number of datapoints while maintaining a similar pattern of behavioral response yielded excellent reliability estimates for RMSE (Cronbach's α = .98), *r* (Cronbach's α = .96), α_A (Cronbach's α = .97) and α_S (Cronbach's α = .91) and good reliability estimates for λ_{FF} (Cronbach's α = .83). The comparison with the truncated dataset simulating the shorter program of the majority of the children yielded similar reliability estimates for RMSE (Cronbach's α = .95), *r* (Cronbach's α = .93), and α_A (Cronbach's α = .91). However, reliability was found to be moderate for α_S (Cronbach's α = .70) and λ_{FF} (Cronbach's α = .68).

Comparisons Between Speaker Groups

Group-level *mean* data fit SimpleDIVA parameter estimates are presented in Table 1. The results showed similar auditory feedback control gains (α_A) for adults and children with TD speech, while the estimate of this parameter was negative and almost twice as large for the children with SSD. The somatosensory feedback control gain estimates (α_S) were negative for both groups of children, while marginally negative for adults. The estimated feedforward control/learning rate (λ_{FF}) was positive for all groups. For the TD group, λ_{FF} was about 3 times higher compared to the group of adults and about 2 times higher compared to the SSD group.

SimpleDIVA parameter estimates of *individual* speakers per group are presented in Figure 1. A multivariate analysis of variance did not reveal a significant between-groups difference.

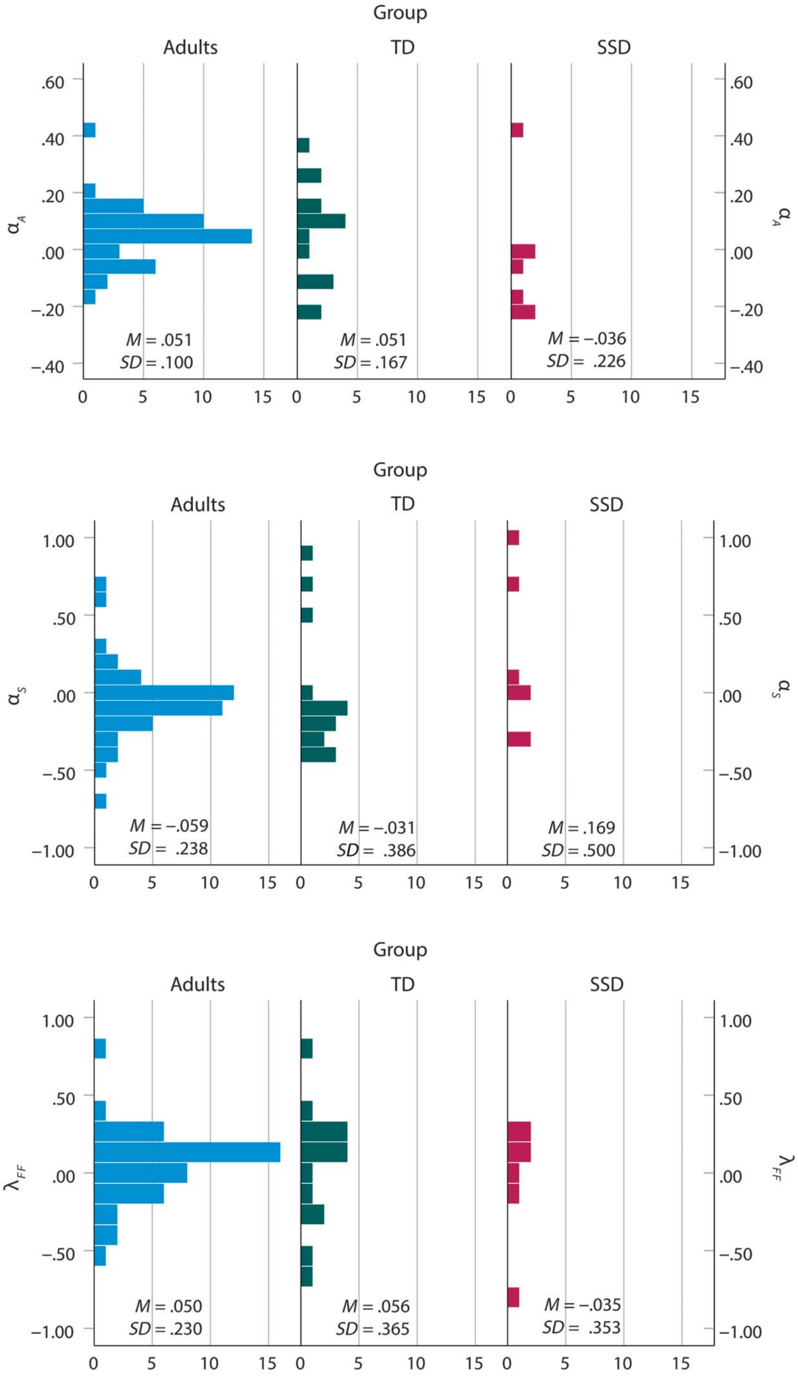
With respect to the correlation among model parameters, the results revealed a strong significant negative relationship between α_S and λ_{FF} for the children with SSD

Table 1. SimpleDIVA parameter estimates for mean group data from adults (means from *n* = 50), typically developing (TD) children (means from *n* = 23), and children with speech sound disorder (SSD; means from *n* = 7).

Group	α_A	α_S	λ_{FF}
Adults	0.042	−0.011	0.064
TD	0.049	−0.088	0.184
SSD	−0.082	−0.070	0.082

Note. α_A = auditory feedback control gain; α_S = somatosensory feedback control gain; λ_{FF} = feedforward control gain/learning rate.

Figure 1. Binned SimpleDIVA parameter estimates (top panel: auditory feedback control gain [α_A]; middle panel: somatosensory feedback control gain [α_S]; bottom panel: feedforward control gain/learning rate [λ_{FF}]) for individual speakers across groups (blue: adults [*n* = 43]; red: TD children [*n* = 16]; green: children with SSD [*n* = 7]). Bars indicate the number of speakers per bin. TD = typically developing; SSD = speech sound disorder.



(r_s = −.86, p = .014). These parameters were not correlated in the adults or the children with TD speech. No correlations were found involving α_A for any of the groups, and no correlations between model parameters were found across groups.

Correlations With Mean Changes in Produced Formants in the Perturbation Experiment

Pooled across groups, results showed a significant correlation between auditory feedback control gain (α_A)

and F1 change in the stay/hold phase compared to the start phase ($r_s = .78$, $p < .001$), F1 change in the end phase compared to the start phase ($r_s = .54$, $p < .001$), and F2 change in the stay/hold phase compared to the start phase ($r_s = .29$, $p = .017$) but not with F2 change in the end phase compared to the start phase ($r_s = .15$, $p = .58$). These correlations indicate that more positive α_A parameter values were correlated with stronger compensatory responses, and more negative α_A parameter values were correlated with stronger following responses. No significant correlations were found involving somatosensory feedback control gain (α_S) or feedforward control gain/learning rate (λ_{FF}).

The results for each group separately showed significant correlations for α_A and F1 change in the stay/hold phase compared to the start phase ($r_s = .71$, $p < .001$) and in the end phase compared to the start phase ($r_s = .38$, $p = .012$) for the adults, as well as for the magnitude of F1 change in the end phase compared to the start phase and λ_{FF} ($r_s = .31$, $p = .042$). For the TD children, F1 change compared to the start phase in both the stay/hold and end phases was found to be correlated with α_A ($r_s = .77$, $p < .001$, and $r_s = .61$, $p = .013$, respectively). The results for the children with SSD showed a similar pattern with a significant correlation between α_A and F1 change compared to the start phase in the stay/hold phase ($r_s = .82$, $p = .023$) and the end phase ($r_s = .79$, $p = .036$). No correlations were found involving α_S, λ_{FF}, or F2 data. Examples of model fits of individual children's data illustrating model parameters in relation to compensatory and amplifying behavior are included in Appendix D.

Correlations With Standardized Speech Production and Perception Test Scores

Oral motor movement skills were assessed in the children with SSD only. Correlations between the fitted model parameters and selected standardized test scores showed the following for this group. Significant strong correlations were found between fast oral motor movement skills and auditory feedback control gain (α_A; $r_s = .97$, $p = .001$), indicating that a lower or more negative parameter was associated with worse sequential fast oral motor movement skills. No correlation was found for somatosensory feedback control gain (α_S) or feedforward control gain/learning rate (λ_{FF}). In addition, no significant correlations were found involving other standardized test scores. For the TD group, none of the correlations reached significance.

Discussion

This study aimed to explore and quantify both between- and within-group differences in feedback and feedforward control parameters in young adults, 4- to 8-year-old children with TD speech, and children with SSD by modeling the speakers' responses to the perturbation of auditory feedback with the SimpleDIVA application.

Parameter Estimation and Model Fits

The overall findings of this study indicated that the parameter estimates derived from the current dataset were smaller in value (i.e., weaker feedback and feedforward gains) compared to similar studies (e.g., Coughler et al., 2021; Kearney et al., 2020). Regarding adult speakers, the single simulation study in the work of Kearney et al. (2020) comprising simultaneous F1 and F2 perturbation (based on the data from Daliri et al., 2018) found an α_S of zero, which is considered roughly similar to this study where α_S could take a negative value and was found marginally so for adults. However, α_A and λ_{FF} were considerably higher in the study by Kearney and colleagues (both 0.10 vs. 0.042 and 0.064, respectively, in this study, both based on mean group data from adults). These results are in line with the overall response effect that was found to be stronger in the study of Daliri et al. (2018; 21.3% for F1 and 3.8% for F2; as reported in Kearney et al., 2020) compared to the study by van Brenk and Terband (2020; 8.4% for F1 and 5.6% for F2). The exact reason for this difference in values does not appear to be straightforward. In the two studies, the perturbed vowels were very similar, Dutch /ɪ/ (near-close front unrounded vowel) versus English /ɛ/ (open-mid front unrounded vowel). Also, the formant shifts were the same (F1 raised 25% and F2 lowered 12.5%), and both shifts moved the sound into existing a neighboring, more open vowel. In both cases, the CVC sequence created with the perturbed vowels map to existing words. It might be that subtle differences in the relative importance of F1 and F2 in these Dutch and English vowel contrasts could lead speakers to respond differently to simultaneous F1 and F2 shifts, which is then reflected in differences in SimpleDIVA feedback and feedforward parameter estimates. Another explanation for these differences in feedback and feedforward gains across studies may lie in differentiating approaches to data processing. For example, when comparing the Daliri et al. (2018) and van Brenk and Terband (2020) studies, it might be observed that, whereas Daliri et al. excluded outliers (utterances that contained production errors and/or outlier formant frequencies, defined as frequencies outside the range of 2 *SD*s) comprising 7% of the trials, van Brenk and Terband excluded only data points of which the formant value could not be measured reliably (comprising 0.6% of all data points). This indicates that the current model fittings and analyses may be assumed to be based on a more capricious dataset, which is reflected in the SimpleDIVA model fit that is slightly lower in this study (*r* = .90 for the mean group data of the adults) compared to the model fit reported for the dataset of Daliri et al. (*r* = .95; Kearney et al., 2020). The present results illustrate that SimpleDIVA is able to model and provide plausible parameter estimates for datasets that are unfiltered and noisy (but therefore arguably ecologically more valid).

The current results also indicated that the number of fitted data points (i.e., the experiment length) does not have a significant effect on parameter estimates. The comparison between the model parameters for the original individual data of the adults and the pruned dataset yielded good to excellent reliability estimates. These results indicate that differences in parameter fits between speakers who completed the longer versus the shorter version of the experiment are not likely to be due to the difference in the number of datapoints. The comparison between the model parameters for the original individual data of the adults with the truncated dataset mimicking the shorter program of the children's experiment yielded moderate to excellent reliability estimates. These findings indicate that differences between groups are unlikely to stem from different lengths of exposure to the experimental phases. The parameters that fell behind in terms of reliability were α_S and λ_{FF} (whose reliability was found to be moderate albeit still considered acceptable, e.g., Field, 2013). In this respect, it should be noted that the results of the SimpleDIVA parameter estimates of data from individual speakers also show larger variability in α_S and λ_{FF} as compared to α_A (as illustrated in Figure 1). Perhaps the case is not one of lower reliability, but rather, the model might tolerate more variability in α_S and λ_{FF} compared to α_A, whether or not specific for the current experimental paradigm where auditory feedback is perturbed. In the DIVA model, the auditory feedback signal functions as the main teaching signal, driving the online correction for the auditory errors and providing the input for the trial-to-trial updates of the feedforward motor commands. α_S works as a damper on the α_A-based corrections as it attempts to keep the vocal tract in its typical somatosensory configuration (Kearney et al., 2020). λ_{FF} then determines the proportional update of the feedforward motor commands, regulating the fraction of the feedback-based corrective command that is added for the next trial. In other words, α_A regulates the main translation of the detected auditory error into a corrective motor command and is thus critical in this specific experimental paradigm. The damping function of α_S and the proportional update of the motor command λ_{FF} appear to be less critical and thus more tolerant to variability.

The Role of Negative Feedback Parameters in Speech Development of Children

The overall group findings indicated some striking differences in parameter outcomes when comparing children with adults, which we suggest are reflective of underlying learning mechanisms in speech development. With respect to the TD children's mean data, the results showed similar α_As compared to the adults, while λ_{FF} was about 3 times higher compared to the adults. This higher learning rate in young children suggests a larger speech motor learning plasticity or less ingrained underlying motor programs, which enables them to rapidly acquire and adapt auditory–articulatory mappings (Walsh et al., 2006). The estimate of α_S was negative with half the magnitude of α_A, where in comparison α_S for the adults approached zero. In the DIVA model, a positive α_S works as a damper on (auditory) feedback-based corrections and a negative α_S would function to facilitate feedback-based corrections.[1] As such, these findings are in line with the hypothesis that the trade-off between auditory somatosensory feedback changes during development and might start with a predominant reliance on auditory feedback, moving away from auditory feedback toward somatosensory feedback during development and stabilizing into adolescence and adulthood (Daliri et al., 2018).

It is difficult to compare current results to previous findings, as the only study that we are aware of that modeled auditory feedback perturbation responses in children is the recent study by Coughler et al. (2021), which differs in three important aspects. First, their paradigm was different as only F1 was perturbed instead of F1 and F2 simultaneously, and F1 was shifted with a much larger magnitude compared to our dataset (+340 Hz in a positive shift condition and −230 Hz in a negative shift condition vs. in our study F1 was shifted +25%, equating to about +150 Hz on average). Second, the children in the study of Coughler et al. were 7–13 years old with an average age of 10 years, while the children in this study were 4- to

---

[1] A negative α_S works to facilitate auditory feedback-based corrections, as follows mathematically from SimpleDIVA Formula 2. The feedback-based correction is based on both auditory and somatosensory errors (the difference between the target formant value [F_{iT}] and the produced formant values based on auditory feedback [F_{iAF}] and somatosensory feedback [F_{iSF}], respectively) detected at the beginning of the production (before feedback control mechanisms contribute), scaled by the gains of the auditory and somatosensory feedback subsystems α_A and α_S. If auditory feedback is perturbed while somatosensory feedback is unaltered, any auditory feedback-based update of the motor command in reaction to the detected error will move the articulatory trajectory away from the somatosensory target. However, the thus instigated somatosensory error will trigger the somatosensory feedback subsystem. If α_S is positive, the somatosensory feedback subsystem will issue a motor command update to attempt to get the vocal tract back in the target somatosensory configuration. A positive α_S thus works as a damper on the α_A-based corrections. An α_S of zero indicates that the auditory feedback-driven response is not restrained by the somatosensory feedback system. If α_S is negative, it will update the motor command in the direction of the somatosensory error, moving toward the auditory feedback-based motor command update. A negative α_S thus works to facilitate feedback-based corrections.

8-year-olds averaging 5.5 years of age. Finally, Coughler et al. used an earlier version of SimpleDIVA in which the feedback and feedforward parameters were limited between 0 and 1, although it should be noted that based on their reported data and SimpleDIVA results, we consider it unlikely that α_S or any of the other parameters would be much different when allowed to become negative (values for α_S reported for the TD children by Coughler et al. were 0.28 and 0.22, depending on the condition). At this point, it is difficult to explain what the present finding of a negative α_S means in terms of cognitive processing. Functionally, it facilitates auditory feedback-based corrections. Our preliminary interpretation is that it might reflect a similar process as discussed with respect to negative auditory feedback control gain (α_A), further expanded on below.

Perhaps the most eye-catching result from the group-level fitted parameters is the negative auditory feedback control gain (α_A) for children with SSD, with a value almost twice the (positive) α_As for the TD children and adults. A negative α_A means that the auditory feedback controller has detected a difference between the target sound and the actual production, but this error led to an adjustment of articulation following the perturbation. Possible mechanisms underlying this following behavior are discussed in more detail below. Furthermore, the parameter estimates for the children with SSD featured an α_S nearing zero and a positive λ_{FF}. An α_S of zero indicates that the auditory feedback-driven response is not restrained by the somatosensory feedback system. In case of a negative α_A, a positive λ_{FF} signifies that the motor command is updated for the next trial with this following response and thus is part of, or contributes to, the following behavior.

Averaged Individual Speaker Modeling Results and Their Correlation With Behavioral Outcomes

Estimates of parameter values for individual speakers showed that the individual, uncurated data were not easy to model, with poor model fits in 7/50 adults and 7/23 children with TD speech. In order to explore individual speaker differences in modeling outcomes, model fits of individual speaker data and their distributions were compared between groups, while correlations with behavioral data were investigated within groups. To our knowledge, such an individualized approach has not been undertaken in SimpleDIVA modeling studies to date.[2] The results indicated an absence of significant differences for the averaged individual speaker results across groups for all three model parameters (see Figure 1). Furthermore, averaged individual trends for α_A were similar compared to the overall group findings, while α_S and λ_{FF} diverged across the two approaches, notably for the children with SSD. This indicates that modeling behavioral data at the group level might even out among speakers and thus obscure individual differences. Additionally, these findings might indicate that the speakers who were not included in the individual analysis (the speakers whose data could not be fitted sufficiently reliable) had a specific, differentiating pattern of model parameters (Lametti et al., 2012) and that leaving out these speakers thus influenced the average parameter fits. However, it is not trivial that the averages of the model parameters for individual speakers should correspond to the model parameters fitted for the mean group data. More research on the relation between modeling individual speakers and modeling mean group data is warranted.

With respect to correlations between model parameters for individual speakers and their responses to auditory feedback perturbation, an overall pattern emerged across all three groups in which higher (more positive) auditory feedback control gains were associated with stronger compensatory responses to the formant shifts and lower (more negative) α_A values were associated with stronger following responses. At the same time, adaptive behavior had no bearing on somatosensory feedback control gains or higher feedforward learning rates. This pattern of correlations was found consistently for F1, and not for F2. Firstly, the question arises why the results did not show a relation between model parameters and adaptation to vowel shifts in the F2 dimension. As described above, in the dataset modeled in this study, the response effects were considerably larger for F1 compared to F2 (Terband et al., 2014; van Brenk & Terband, 2020). Our interpretation is that F1 therefore had a larger influence on parameter estimations compared to F2. Secondly, adaptation effects were found to correlate only with α_A. Feedforward learning (as estimated by λ_{FF}) can only take effect when α_A (or α_S) is nonzero; without feedback control, no online corrective command will be generated necessary for updating the next trial. Overall, these findings underline the importance of online control for sensorimotor learning in speech production (e.g., Scheerer et al., 2016) and illustrate that sensorimotor learning is mediated by production variability and perceived sensory errors (e.g., Lametti et al., 2018).

For the children with SSD, the results showed that more negative and lower auditory feedback control gains (α_A) were associated with worse sequential fast oral motor movement skills, but no correlation was found for feedforward control gain/learning rate (λ_{FF}) or somatosensory feedback control gains (α_S), which were correlated to each

---

[2] Note that Kearney et al. (2022) modeled both group-mean data and individual subject responses to reflexive (randomly applied) perturbations of pitch in healthy adults to investigate the reliability of model fits and characterize individual differences.

other. In comparison, in our behavioral study, similarly strong correlations were found between fast sequential oral motor movement abilities and the amount of adaptation (Terband et al., 2014). No correlations were found with PCCI of word and nonword repetitions or with word and nonword auditory discrimination scores. Although the present results must be approached with reservation due to the small sample size, these findings correspond with previous studies reporting correlations between nonspeech oral motor skills and performance on speech tasks in children with a variety of developmental speech disorders in Dutch (e.g., Nijland et al., 2015; Terband et al., 2018; see also Diepeveen et al., 2019). Together, these results suggest that, in children with SSD, a sufficient level of general sequential oral motor skills might be a prerequisite for robust internal (forward and inverse) models (Chen et al., 2021), which in turn are prerequisites of goal-directed motor learning. The exact nature of these relations seems worthwhile for further investigation in future studies.

Possible Mechanisms Underlying Following Behavior

Perturbation amplifying or following responses to formant shifts are not unique to children with SSD, although studies explicitly reporting on amplifying responses are sparse. In our previous study, we found that about 12%–30% of the TD children and adults showed significant perturbation following behavior (van Brenk & Terband, 2020). This is also illustrated by the Simple-DIVA parameter estimates of individual speakers in this study in which negative α_As were found for speakers in the groups of TD children and adults. Our interpretation is that, in these cases, the auditory signal is processed as an external cue (cf. Hain et al., 2000, regarding following responses in fundamental frequency perturbation).

We believe that one or a combination of three causes might be underlying these particular findings—possible explanations that are not completely independent but mainly differ in causal origin. First, it might be instigated by the experimental setup in which the auditory signal is recorded, processed, and fed back online to the speaker via headphones. Although the analysis-perturbation-resynthesis-feedback procedure of the experimental setup used in this study has a small latency (about 27 ms for the total setup, of which 12–14 ms is due to software processing; Kim et al., 2020) and it is generally assumed that this latency is not noticeable, experiment debriefing revealed that around 65% of the adult participants indicated to have noticed manipulations of the stimuli, including 22% who reported to have consciously undertaken action (Terband & van Brenk, 2015). However, crosstab analysis showed no correlation between debriefing response and compensatory or following behavior during the experiment. While it could be the case in specific individuals that their awareness of the manipulation by the feedback latency caused their speech–sensorimotor system to process the auditory signal as an external cue, it cannot be the full story.

A second possible mechanism is based on the role of efference copies of the motor commands. The comparison between efferent and afferent signals plays a fundamental role in sensorimotor control and sensorimotor learning, in which the discrepancy between the intended (predicted) and actual (sensory) feedback could be used to identify the source of the production error (see, e.g., Wolpert et al., 2011, for a review). A neural marker for this process of comparing efference copies with auditory feedback is a phenomenon called speaking-induced suppression: the mechanism that auditory cortical responses to self-generated speech signals are suppressed when compared to their responses to external speech signals (e.g., Hirano et al., 1997; Houde et al., 2002; Numminen et al., 1999). Although the gradually introduced perturbation of formant shifts supposedly diminishes this effect (Sato & Shiller, 2018), previous studies have shown that speaking-induced suppression is weaker for vowel productions that are less prototypical with respect to their F1–F2 vowel space (Niziolek et al., 2013) and does not occur at all when altered feedback causes a mismatch between the speakers' feedback and their expectations based on the efference copy (Heinks-Maldonado et al., 2006; Houde et al., 2002). The specific type and degree of discrepancy, in either the time domain, the spectral domain, or both, could prevent the speech–sensorimotor system from identifying the auditory signal as self-produced, causing it to be either ignored or treated as an external referent.

Finally, a third possibility is that the mechanism of speaking-induced suppression itself is the root cause. Previous studies have found differences in the timing of speaking-induced suppression in both adults and children who stutter (Beal et al., 2010, 2011; Toyomura et al., 2020). More specifically, children who stutter showed delayed suppression of auditory evoked fields (but with a similar amplitude) compared to children with TD speech when producing vowels, but not when listening to vowels (Beal et al., 2011). In patients with Parkinson's disease, speaking-induced suppression has been found to be significantly reduced (Railo et al., 2019). The mechanism of speaking-induced suppression thus has been found to be deviant in several different motor speech disorders, in different ways. Considering the current findings regarding the children with SSD, where more negative auditory feedback control gains were associated with worse sequential fast oral motor movement skills in addition to a comparatively large negative α_A for the group-level data fit, we thus speculate that an atypicality in the speaking-induced suppression mechanism might also play a role in SSD—in particular when the impairment involves a motor component.

Conclusions

With the current SimpleDIVA installment allowing auditory and somatosensory feedback control gains and feedforward learning rates ranging from negative to positive, we were able to account for responses to the perturbation of auditory feedback other than corrective. The modeling of individual speaker data indicated that fitting such data is capricious, at least in the present, uncurated dataset (where no outliers were removed). Not all individual data yielded sufficiently reliable model fits. Focusing on the data where individual model fits were sufficient, we correlated the findings with the behavioral data and with outcome measures obtained from standardized assessment tasks, enabling fine-grained speaker profile analyses.

The complementary group level and individual level modeling approach suggested pertinent differences in speech–sensorimotor online control and learning behavior between TD children and adults, as well as between children with SSD and TD children. While, given the small group size of children with SSD, these preliminary results should be interpreted with caution, current findings allowed theorizing with respect to possible mechanisms underlying the nature of following responses displayed in children with SSD. It is speculated that the auditory signal might be processed as an external cue, thus preventing online compensation to perturbations and hindering the successful use of auditory feedback as a teaching signal for the acquisition and adaptation of speech motor programs, with a hypothesized important role of deviant speaking-induced suppression present in children with SSD. Future research should assess how children with SSD process efferent speech signals and evaluate a possible role of associated suppression mechanisms.

Data Availability Statement

The datasets generated and/or analyzed during this study are available from the corresponding author on reasonable request.

Acknowledgments

We would like to express our gratitude to the children and their parents or caretakers for their time and effort, Anniek van Doornik-van der Zee and Margoke Nijssen for their help with participant recruitment, and Shanqing Cai for his help with the Audapter software for the collection of the datasets used in this study. We gratefully thank Elaine Kearney and Riccardo Falsini for making the SimpleDIVA model available to others and, in particular, for their implementation of our requests for expanded functionality in the 1.3 version of the SimpleDIVA model.

References

American Speech-Language-Hearing Association. (2007). *Childhood apraxia of speech* [Technical report]. https://www.asha.org/policy/tr2007-00278/

Bauman-Waengler, J. (2020). *Articulation and phonology in speech sound disorders* (6th ed.). Pearson.

Beal, D. S., Cheyne, D. O., Gracco, V. L., Quraan, M. A., Taylor, M. J., & Luc, F. (2010). Auditory evoked fields to vocalization during passive listening and active generation in adults who stutter. *NeuroImage, 52*(4), 1645–1653. https://doi.org/10.1016/j.neuroimage.2010.04.277

Beal, D. S., Quraan, M. A., Cheyne, D. O., Taylor, M. J., Gracco, V. L., & Luc, F. (2011). Speech-induced suppression of evoked auditory fields in children who stutter. *NeuroImage, 54*(4), 2994–3003. https://doi.org/10.1016/j.neuroimage.2010.11.026

Behroozmand, R., Korzyukov, O., Sattler, L., & Larson, C. R. (2012). Opposing and following vocal responses to pitch-shifted auditory feedback: Evidence for different mechanisms of voice pitch control. *The Journal of the Acoustical Society of America, 132*(4), 2468–2477. https://doi.org/10.1121/1.4746984

Boersma, P., & Weenink, P. J. M. (2013). *Praat: Doing phonetics by computer* [Computer program]. https://www.praat.org/

Burnett, T. A., Freedland, M. B., Larson, C. R., & Hain, T. C. (1998). Voice F0 responses to manipulations in pitch feedback. *The Journal of the Acoustical Society of America, 103*(6), 3153–3161. https://doi.org/10.1121/1.423073

Cai, S., Beal, D. S., Ghosh, S. S., Tiede, M. K., Guenther, F. H., & 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), Article e41830. https://doi.org/10.1371/journal.pone.0041830

Cai, S., Ghosh, S. S., Guenther, F. H., & Perkell, J. (2010). Adaptive auditory feedback control of the production of formant trajectories in the Mandarin triphthong /iau/ and its pattern of generalization. *The Journal of the Acoustical Society of America, 128*(4), 2033–2048. https://doi.org/10.1121/1.3479539

Chen, T., Lammert, A. C., & Parrell, B. (2021). Modeling sensorimotor adaptation in speech through alterations to forward and inverse models. In H. Heřmanský & H. Černocký (Eds.), *Proceedings of Interspeech 2021* (pp. 3201–3205). International Speech Communication Association.

Cheung, S. T., Thompson, K., Chen, J. L., Yunusova, Y., & Beal, D. S. (2021). Response patterns to vowel formant perturbations in children. *The Journal of the Acoustical Society of America, 150*(4), 2647–2654. https://doi.org/10.1121/10.0006567

Coughler, C., Hamel, E. M., Cardy, J. O., Archibald, L. M., & Purcell, D. W. (2021). Compensation to altered auditory feedback in children with developmental language disorder and typical development. *Journal of Speech, Language, and*

Hearing Research, 64(6S), 2363–2376. https://doi.org/10.1044/2020_JSLHR-20-00374

Daliri, A., Wieland, E. A., Cai, S., Guenther, F. H., & Chang, S. E. (2018). Auditory-motor adaptation is reduced in adults who stutter but not in children who stutter. *Developmental Science, 21*(2), Article e12521. https://doi.org/10.1111/desc.12521

Diepeveen, S., van Haaften, L., Terband, H., De Swart, B., & Maassen, B. (2019). A standardized protocol for maximum repetition rate assessment in children. *Folia Phoniatrica et Logopaedica, 71*(5–6), 238–250. https://doi.org/10.1159/000500305

Dunn, L. H., & Dunn, L. M. (2005). *Picture Vocabulary Test-III-NL* [PPVT-III-NL] (L. Schlichting, trans.). Pearson.

Erlings-van Deurse, M., Freriks, A., Goudt-Bakker, K., Van der Meulen, S., & de Vries, L. (1993). *Dyspraxia programme: Oral motor movement assessment*. Swets & Zeitlinger.

Field, A. (2013). *Discovering statistics using IBM SPSS statistics* (4th ed.). Sage.

Guenther, F. H. (1994). A neural network model of speech acquisition and motor equivalent speech production. *Biological Cybernetics, 72*(1), 43–53. https://doi.org/10.1007/BF00206237

Guenther, F. H. (2016). *Neural control of speech*. MIT Press. https://doi.org/10.7551/mitpress/10471.001.0001

Guenther, F. H., Hampson, M., & Johnson, D. (1998). A theoretical investigation of reference frames for the planning of speech movements. *Psychological Review, 105*(4), 611–633. https://doi.org/10.1037/0033-295X.105.4.611-633

Guenther, F. H., Nieto-Castañón, A., Falsini, R., Elaine, K., & Weerathunge, H. R. (2019). *SimpleDIVA (version 1.3)*. http://sites.bu.edu/guentherlab/software/simplediva-app

Hain, T. C., Burnett, T. A., Kiran, S., Larson, C. R., Singh, S., & Kenney, M. K. (2000). Instructing subjects to make a voluntary response reveals the presence of two components to the audio-vocal reflex. *Experimental Brain Research, 130*(2), 133–141. https://doi.org/10.1007/s002219900237

Heinks-Maldonado, T. H., Nagarajan, S. S., & Houde, J. F. (2006). Magnetoencephalographic evidence for a precise forward model in speech production. *NeuroReport, 17*(13), 1375–1379. https://doi.org/10.1097/01.wnr.0000233102.43526.e9

Hirano, S., Kojima, H., Naito, Y., Honjo, I., Kamoto, Y., Okazawa, H., Ishizu, K., Yonekura, Y., Nagahama, Y., Fukuyama, H., & Konishi, J. (1997). Cortical processing mechanism for vocalization with auditory verbal feedback. *NeuroReport, 8*(9), 2379–2382. https://doi.org/10.1097/00001756-199707070-00055

Houde, J. F., & Jordan, M. I. (1998). Sensorimotor adaptation in speech production. *Science, 279*(5354), 1213–1216. https://doi.org/10.1126/science.279.5354.1213

Houde, J. F., & Jordan, M. I. (2002). Sensorimotor adaptation of speech I: Compensation and adaptation. *Journal of Speech, Language, and Hearing Research, 45*(2), 295–310. https://doi.org/10.1044/1092-4388(2002/023)

Houde, J. F., Nagarajan, S. S., Sekihara, K., & Merzenich, M. M. (2002). Modulation of the auditory cortex during speech: An MEG study. *Journal of Cognitive Neuroscience, 14*(8), 1125–1138. https://doi.org/10.1162/089892902760807140

Kay, J., Lesser, R., & Coltheart, M. (1995). *Psycholinguistic Assessment of Language Processing in Aphasia*. (R. Bastiaanse, M. Bosje, & E. G. Visch-Brink, trans.). Erlbaum. (Original work published 1992)

Kearney, E., & Guenther, F. H. (2019). Articulating: The neural mechanisms of speech production. *Language, Cognition and Neuroscience, 34*(9), 1214–1229. https://doi.org/10.1080/23273798.2019.1589541

Kearney, E., Nieto-Castañón, A., Falsini, R., Daliri, A., Murray, E. S. H., Smith, D. J., & Guenther, F. H. (2022). Quantitatively characterizing reflexive responses to pitch perturbations. *Frontiers in Human Neuroscience, 16,* 929687. https://doi.org/10.3389/fnhum.2022.929687

Kearney, E., Nieto-Castañón, A., Weerathunge, H. R., Falsini, R., Daliri, A., Abur, D., Ballard, K. J., Chang, S.-E., Chao, S.-C., Heller Murray, E. S., Scott T. L., & Guenther F. H. (2020). A simple 3-parameter model for examining adaptation in speech and voice production. *Frontiers in Psychology, 10,* 2995. https://doi.org/10.3389/fpsyg.2019.02995

Kim, K. S., Wang, H., & Max, L. (2020). It's about time: Minimizing hardware and software latencies in speech research with real-time auditory feedback. *Journal of Speech, Language, and Hearing Research, 63*(8), 2522–2534. https://doi.org/10.1044/2020_JSLHR-19-00419

Lametti, D. R., Nasir, S. M., & Ostry, D. J. (2012). Sensory preference in speech production revealed by simultaneous alteration of auditory and somatosensory feedback. *Journal of Neuroscience, 32*(27), 9351–9358. https://doi.org/10.1523/JNEUROSCI.0404-12.2012

Lametti, D. R., Smith, H. J., Watkins, K. E., & Shiller, D. M. (2018). Robust sensorimotor learning during variable sentence-level speech. *Current Biology, 28*(19), 3106–3113.e2. https://doi.org/10.1016/j.cub.2018.07.030

Maassen, B., van Haaften, L., Diepeveen, S., van den Engel-Hoek, L., Veenker, T., Terband, H., & De Swart, B. (2019). *Computer Articulatie-instrument (CAI)*. Boom Test Uitgevers.

MacDonald, E. N., Johnson, E. K., Forsythe, J., Plante, P., & Munhall, K. G. (2012). Children's development of self-regulation in speech production. *Current Biology, 22*(2), 113–117. https://doi.org/10.1016/j.cub.2011.11.052

McLeod, S., Harrison, L. J., & McCormack, J. (2013). *Schaal voor Verstaanbaarheid in de context* [Intelligibility in context scale: Dutch]. (J.C. van Doornik-van der Zee & H. Terband, trans.). Charles Sturt University. http://www.csu.edu.au/research/multilingual-speech/ics

Menard, L., Perrier, P., & Aubin, J. (2013). The role of auditory feedback in speech development: A study of compensation strategies for a lip-tube perturbation. *Proceedings of Meetings on Acoustics ICA2013, 19*(1), 060181.

Nijland, L., Terband, H., & Maassen, B. (2015). Cognitive functions in childhood apraxia of speech. *Journal of Speech, Language, and Hearing Research, 58*(3), 550–565. https://doi.org/10.1044/2015_JSLHR-S-14-0084

Niziolek, C. A., Nagarajan, S. S., & Houde, J. F. (2013). What does motor efference copy represent? Evidence from speech production. *Journal of Neuroscience, 33*(41), 16110–16116. https://doi.org/10.1523/JNEUROSCI.2137-13.2013

Numminen, J., Salmelin, R., & Hari, R. (1999). Subject's own speech reduces reactivity of the human auditory cortex. *Neuroscience Letters, 265*(2), 119–122. https://doi.org/10.1016/S0304-3940(99)00218-9

Railo, H., Nokelainen, N., Savolainen, S., & Kaasinen, V. (2019). *Reduced speaking-induced suppression of auditory evoked-potentials in early Parkinson's disease suggests deficits in monitoring self-produced speech. bioRxiv.* https://www.biorxiv.org/content/biorxiv/early/2019/10/30/823674.full.pdf

Sato, M., & Shiller, D. M. (2018). Auditory prediction during speaking and listening. *Brain and Language, 187,* 92–103. https://doi.org/10.1016/j.bandl.2018.01.008

Scheerer, N. E., Jacobson, D. S., & Jones, J. A. (2016). Sensorimotor learning in children and adults: Exposure to frequency-altered auditory feedback during speech production. *Neuroscience, 314,* 106–115. https://doi.org/10.1016/j.neuroscience.2015.11.037

Terband, H., Spruit, M., & Maassen, B. (2018). Speech impairment in boys with fetal alcohol spectrum disorders. *American Journal of Speech-Language Pathology, 27*(4), 1405–1425. https://doi.org/10.1044/2018_AJSLP-17-0013

Terband, H., & van Brenk, F. (2015). Compensatory and adaptive responses to real-time formant shifts in adults and children. *Proceedings of the 18th Congress of Phonetic Sciences.*

Terband, H., van Brenk, F., & van Doornik-van der Zee, A. (2014). Auditory feedback perturbation in children with developmental speech sound disorders. *Journal of Communication Disorders, 51,* 64–77. https://doi.org/10.1016/j.jcomdis.2014.06.009

Tourville, J. A., & Guenther, F. H. (2011). The DIVA model: A neural theory of speech acquisition and production. *Language and Cognitive Processes, 26*(7), 952–981. https://doi.org/10.1080/01690960903498424

Tourville, J. A., Reilly, K. J., & Guenther, F. H. (2008). Neural mechanisms underlying auditory feedback control of speech. *NeuroImage, 39*(3), 1429–1443. https://doi.org/10.1016/j.neuroimage.2007.09.054

Toyomura, A., Miyashiro, D., Kuriki, S., & Sowman, P. F. (2020). Speech-induced suppression for delayed auditory feedback in adults who do and do not stutter. *Frontiers in Human Congress of Phonetic Sciences Neuroscience, 14,* 150. https://doi.org/10.3389/fnhum.2020.00150

van Brenk, F., & Terband, H. (2020). Compensatory and adaptive responses to real-time formant shifts in adults and children. *The Journal of the Acoustical Society of America, 147*(4), 2261–2270. https://doi.org/10.1121/10.0001018

Villacorta, V. M., Perkell, J. S., & Guenther, F. H. (2007). Sensorimotor adaptation to feedback perturbations of vowel acoustics and its relation to perception. *The Journal of the Acoustical Society of America, 122*(4), 2306–2319. https://doi.org/10.1121/1.2773966

Walsh, B., Smith, A., & Weber-Fox, C. (2006). Short-term plasticity in children's speech motor systems. *Developmental Psychobiology, 48*(8), 660–674. https://doi.org/10.1002/dev.20185

Wolpert, D. M., Diedrichsen, J., & Flanagan, J. R. (2011). Principles of sensorimotor learning. *Nature Reviews Neuroscience, 12*(12), 739–751. https://doi.org/10.1038/nrn3112

Appendix A (p. 1 of 2)

Demographic Information and Background Diagnostic Data of the Children With SSD That Participated in the Study

Table A1. Demographic information and results on the standardized speech production, speech perception, and oral motor tasks (Terband et al., 2014).

ID	Diagnosis	Age (y)	Sex	WCQ (PPVT)	Intelligibility (ICS)	Auditory discrimination (PALPA)		Diadochokinesis (DDK; pataka CAI)		Oral-motor mov. Assessment
						Words	Nonwords	Score	Judgment	Iso – seq – seq
						(% correct)	(% correct)			fast (% correct)
SSD1	PD	5.8	m	127	4.00	94	100	1	2	85 – 83 – 60
SSD2	PD/DS	4.8	m	98	4.43	97	81	1	3	85 – 78 – 40
SSD3	PD + PAD	7.5	f	106	4.00	94	86	1	3	92 – 94 – 50
SSD5	PD	6.6	f	84	3.42	64	44	1	3	77 – 67 – 40
SSD8	CAS/PD	6.1	m	106	3.14	92	69	0	0	96 – 95 – 50
SSD10	PD + PAD	5.0	f	104	4.30	81	83	1	3	85 – 100 – 80
SSD14	CAS/PD	5.0	f	115	3.86	94	86	0	0	77 – 78 – 60

ID	Diagnosis	Picture naming (60 words CAI)				Word repetition (WR; 10 words CAI)				Nonword repetition (10 nonwords similar to WR CAI)			
		PCCI	PCCCI	PSSC	atyp./typ. Sub.proc.	PCCI	PCCCI	PSSC	atyp./typ. Sub.proc.	PCCI	PCCCI	PSSC	atyp./typ. Sub.proc.
SSD1	PD	.94	.70	.96	3/0	.98	.50	1.00	1/0	.88	.27	.94	4/1
SSD2	PD/DS	1.00	1.00	.99	0/0	1.00	.91	.97	0/0	.79	.36	.88	11/5
SSD3	PD + PAD	1.00	.96	.97	0/0	1.00	.95	1.00	0/0	.88	.95	.95	13/5
SSD5	PD	.81	.39	.82	4/8	.69	.77	.82	5/7	.65	.50	.73	18/24
SSD8	CAS/PD	.69	.26	.81	22/14	.59	.41	.59	15/9	.71	.09	.72	18/24
SSD10	PD + PAD	.91	.83	.94	7/3	.89	.77	1.00	6/0	.91	.95	.97	9/1
SSD14	CAS/PD	.57	.13	.80	28/9	.48	.23	.62	23/11	.71	.27	.80	26/10

Note. A forward slash (/) indicates a mix of symptoms unable to be classified as either of the listed diagnoses at the time/stage of development. A plus sign (+) indicates a symptom profile that could be classified as both diagnoses. Diagnosis: PD = phonological disorder; DS = developmental stuttering; PAD = phonetic articulation disorder; CAS = childhood apraxia of speech. y = year; WCQ = Word Comprehension Quotient; PPVT = Peabody Picture Vocabulary Test; ICS = Intelligibility in Context Scale; PALPA = Psycholinguistic Assessments of Language Processing in Aphasia; CAI = Computer Articulation Instrument; m = male; f = female; PCCI = proportion of syllable-initial consonants correct; PSSC = proportion of syllable structures correct; atyp./typ. sub.proc. = atypical/typical substitution processes.

Appendix A (p. 2 of 2)

Demographic Information and Background Diagnostic Data of the Children With SSD That Participated in the Study

Table A2. Description of the standardized speech production, speech perception, and oral motor tasks (Terband et al., 2014).

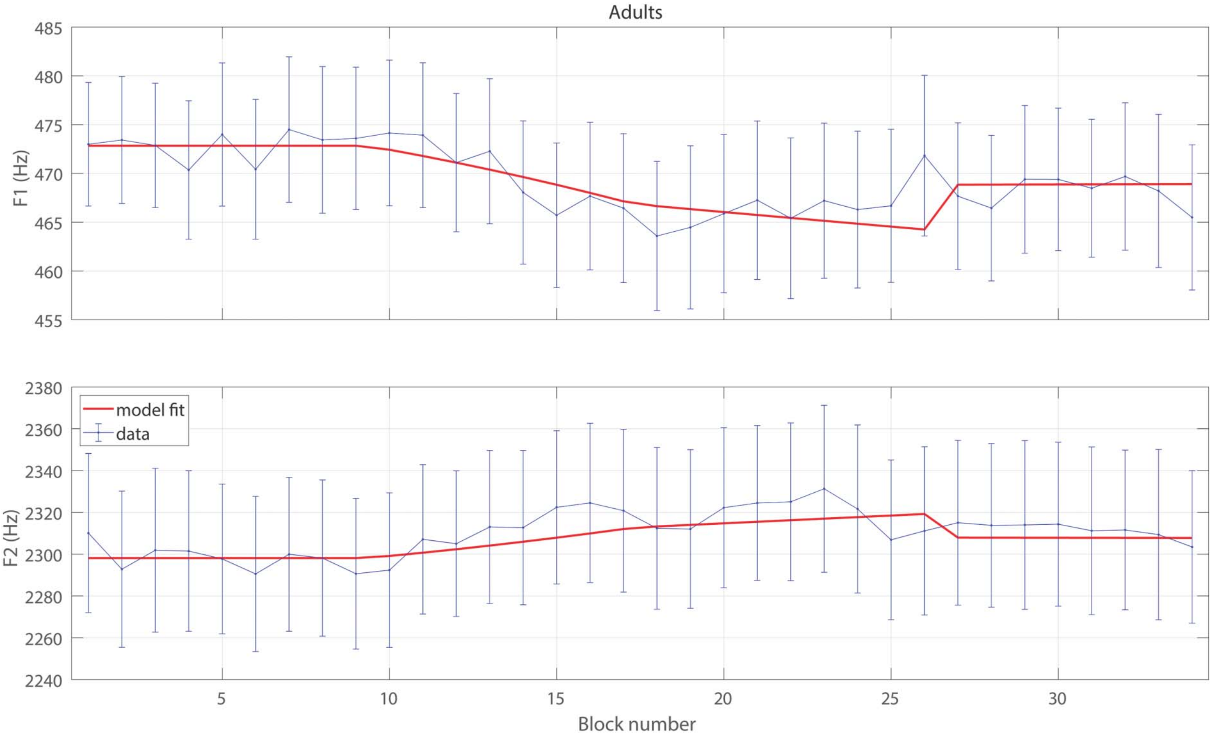
Task/assessment		**Description**
WCQ (PVVT-III-NL)		Dutch Version of The Peabody Picture Vocabulary Test of receptive vocabulary (Dunn & Dunn, 2005), with high correlations to verbal intelligence. This test is given verbally. The children were asked to point to one out of four pictures corresponding to the stimulus word. Score is expressed by the Word Comprehension Quotient (WCQ; *M* = 100, *SD* = 15).
Intelligibility (ICS)		The Intelligibility in Context Scale (McLeod et al., 2013) is a quick parent report measure of children's intelligibility. The 7-item questionnaire rates the degree to which children's speech is understood by different communication partners (parents, immediate family, extended family, friends, acquaintances, teachers, and strangers) on a 5-point scale. A higher score denotes better intelligibility.
Auditory discrimination (PALPA)		Auditory discrimination task from the Dutch translation of the PALPA (Kay et al., 1992/1995) adapted for children. Score is percentage correct.
	Words	36 pairs of CVC-words that were either the same (18 pairs), differed on one consonant (initial or final; 12 pairs), or were metatheses of each other (6 pairs, e.g., "lor" vs "rol").
	Nonwords	36 pairs of CVC-nonwords that were either the same (18 pairs),differed on one consonant (initial or final; 12 pairs), or were metatheses of each other (6 pairs, e.g., "tus" vs "sut").
Diadochokinesis (CAI)		Maximum performance task using utterances of [pataka]. The children were first asked to produce "pataka" once, and when they succeeded, they were asked to produce "pataka" in a sequence of several repetitions of "pataka." After that, the children were asked to speed up while producing a sequence of "pataka." This task is administered with the Computer Articulation Instrument (CAI; Maassen et al., 2019).
	PTK-score	1 = [pataka] could be produced; 0 = [pataka] could not be produced.
	PTK-judgment	4 = perfect; 3 = [pataka] in sequence in normal rate, but no acceleration; 2 = [pataka] in sequence incorrect ([t] or [k] could not be pronounced), but speeding up on two different consonants ([pata], [taka]) was possible; 1 = no fluent [pataka], not in sequence; 0 = no [pataka] production either in isolation or in a sequence of two.
Oral motor Movement Assessment		Oral motor assessment from the Dutch Dyspraxia Program (Erlings-van Deurse et al., 1993). For each element: unable = 0; deviant = 1; able = 2. Scores expressed in percentage per subtask.
	Isolation	Positioning of the lips in rest; close lips around straw; lip protrusion; lip spreading; bite lower lip; clench straw with lips; stick out the tongue; move tongue left; move tongue right; move tongue up; move tongue down; push straw with tongue horizontally; push straw with tongue vertically.
	Sequential	Lips protrude and spread; bite and release lower lip; move tongue left and right; lick lips; touch teeth front and back with tongue; move tongue up and down; open and close the jaw; check whether velum closes when blowing; check whether velum closes when sucking.
	Seq. fast	In high tempo: Lips protrude and spread; move tongue left and right; lick lips; move tongue up and down; open and close the jaw.
Picture naming (60 words CAI)		This task consists of 60 images depicting 50 words with different consonants, consonant clusters, and vowels at various positions (initial, medial, final) and 10 words with complex consonant patterns. This task is administered with the CAI (Maassen et al., 2019).
Word repetition (10 words CAI)		Repetition task using the same 10 words with complex consonant patterns as in *picture naming*. The words were presented through headphones, and the children were asked to repeat them. This task is administered with the CAI (Maassen et al., 2019).
Nonword repetition (33 nonwords CAI)		Same task as *word repetition,* using 33 multisyllabic nonword stimuli consisting of syllables that do not exist as words in Dutch. The first 23 nonwords have syllable structures similar to the multisyllabic stimuli of the picture naming task, while the last 10 feature complex consonant patterns resembling the stimuli in the word repetition task. This task is administered with the CAI (Maassen et al., 2019), similar to *word repetition*.

Note. PPVT-III-NL = Peabody Picture Vocabulary Test; ICS = Intelligibility in Context Scale; PALPA = Psycholinguistic Assessments of Language Processing in Aphasia; CAI = Computer Articulation Instrument.

Appendix B (p. 1 of 3)

Model Fits of Group Data

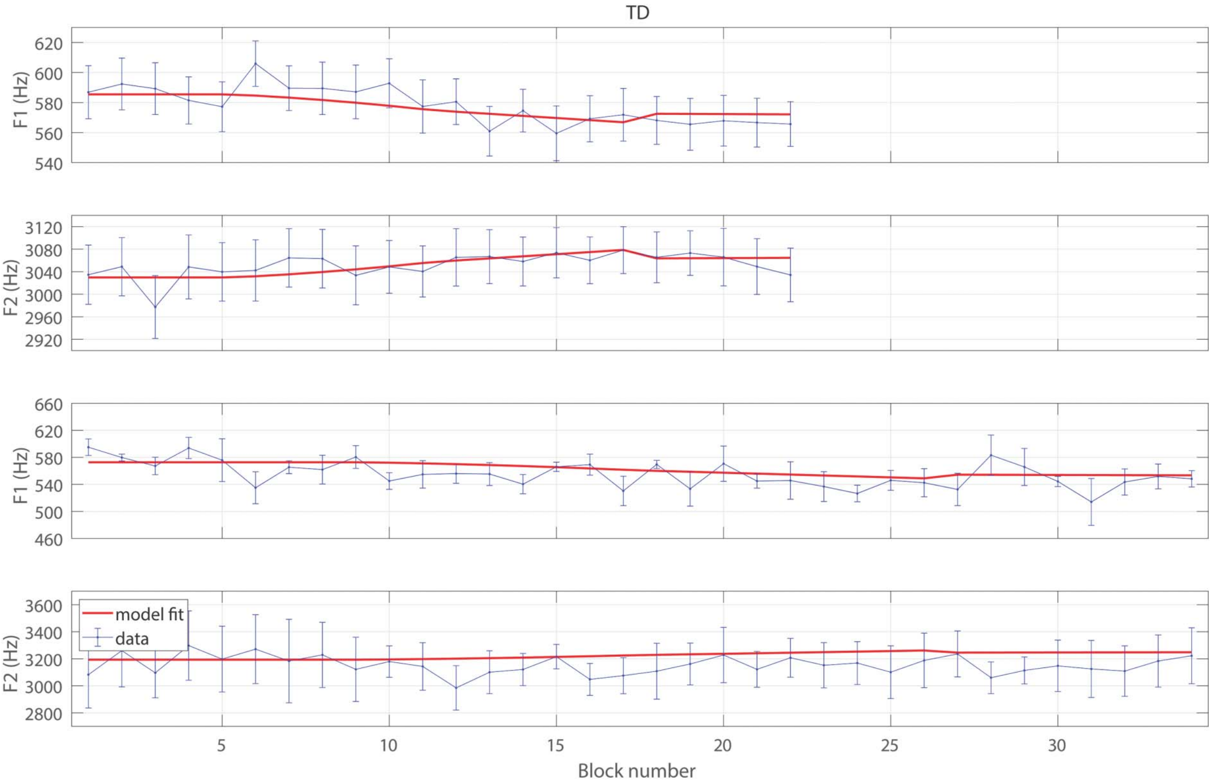
Figure B1. Group-level model fits of the dataset of adults with perturbations simultaneously applied to both F1 and F2. Mean and standard error of experimental data in blue; model fit in red.



Appendix B (p. 2 of 3)

Model Fits of Group Data

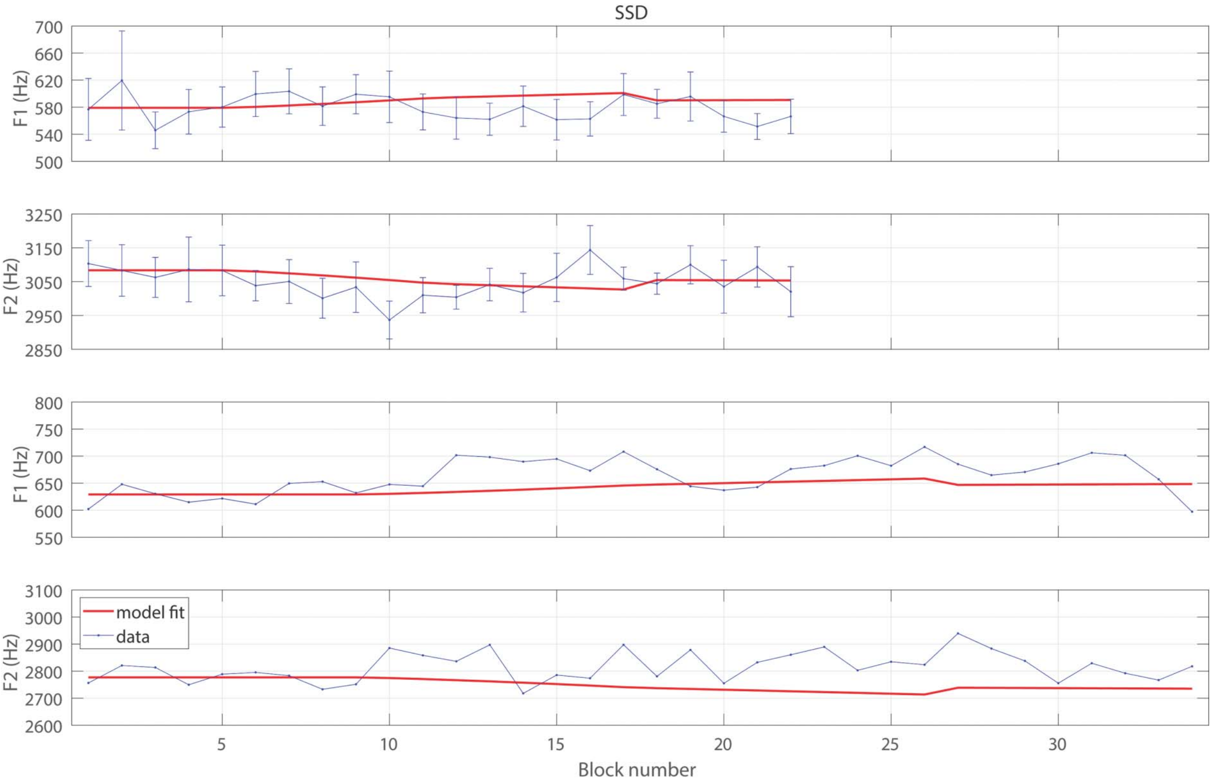
Figure B2. Group-level model fits of the dataset of children with TD speech with perturbations simultaneously applied to both F1 and F2. Mean and standard error of experimental data in blue; model fit in red.



Appendix B (p. 3 of 3)

Model Fits of Group Data

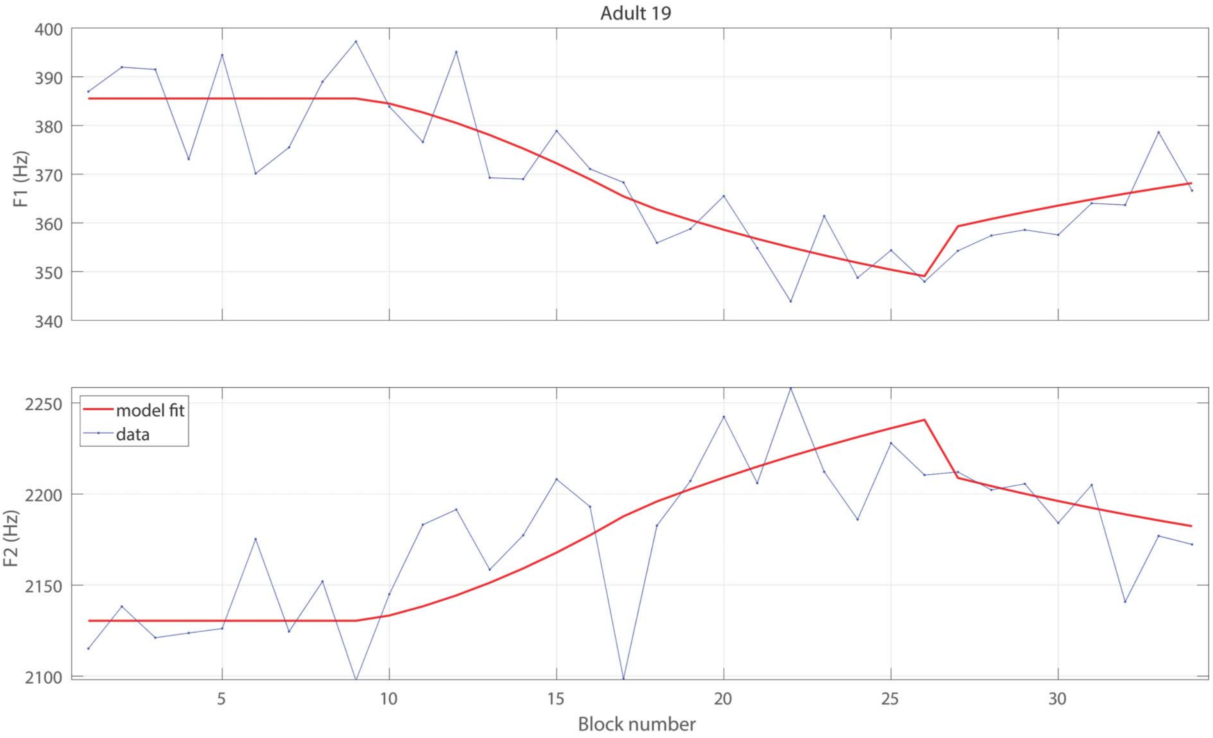
Figure B3. Group-level model fits of the dataset of children with SSD with perturbations simultaneously applied to both F1 and F2. Mean and standard error of experimental data in blue; model fit in red.



Appendix C (p. 1 of 4)

Examples of Model Fits of Individual Subject Data

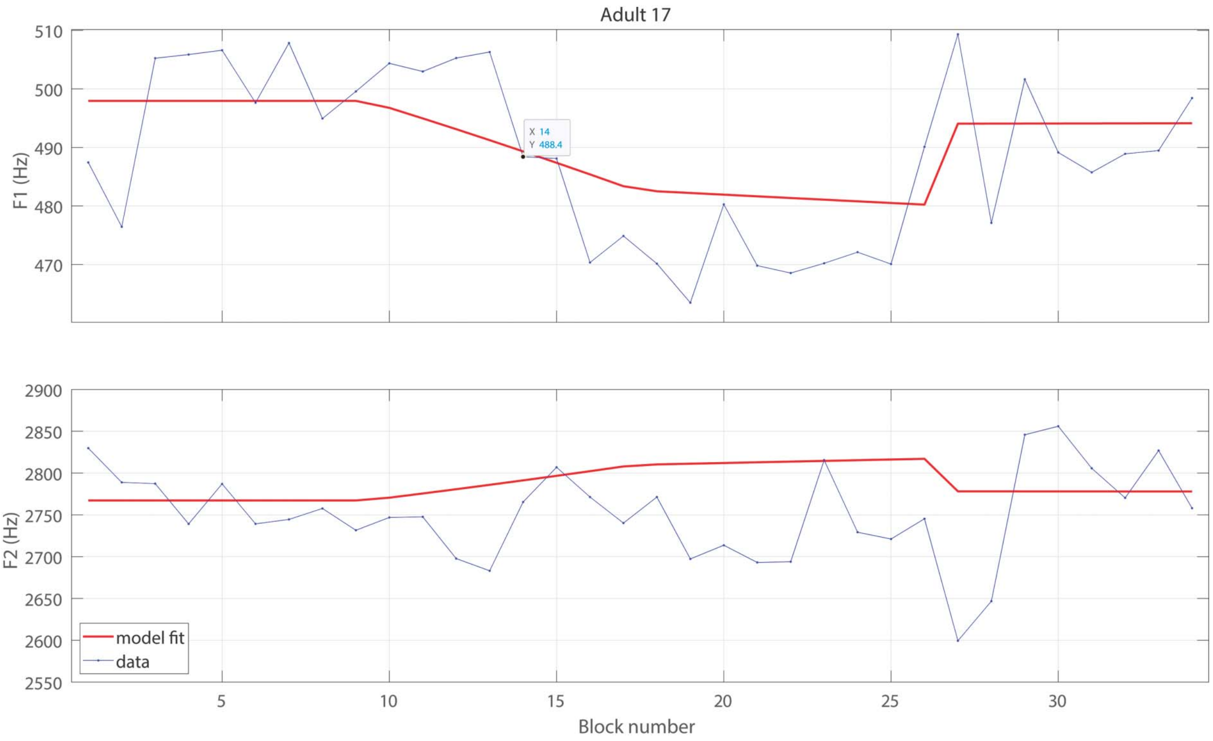
Figure C1. Model fit for the data of Adult 19 with perturbations simultaneously applied to both F1 and F2, exemplifying a good fit (*r* = .93). Mean and standard error of experimental data in blue; model fit in red.



Appendix C (p. 2 of 4)

Examples of Model Fits of Individual Subject Data

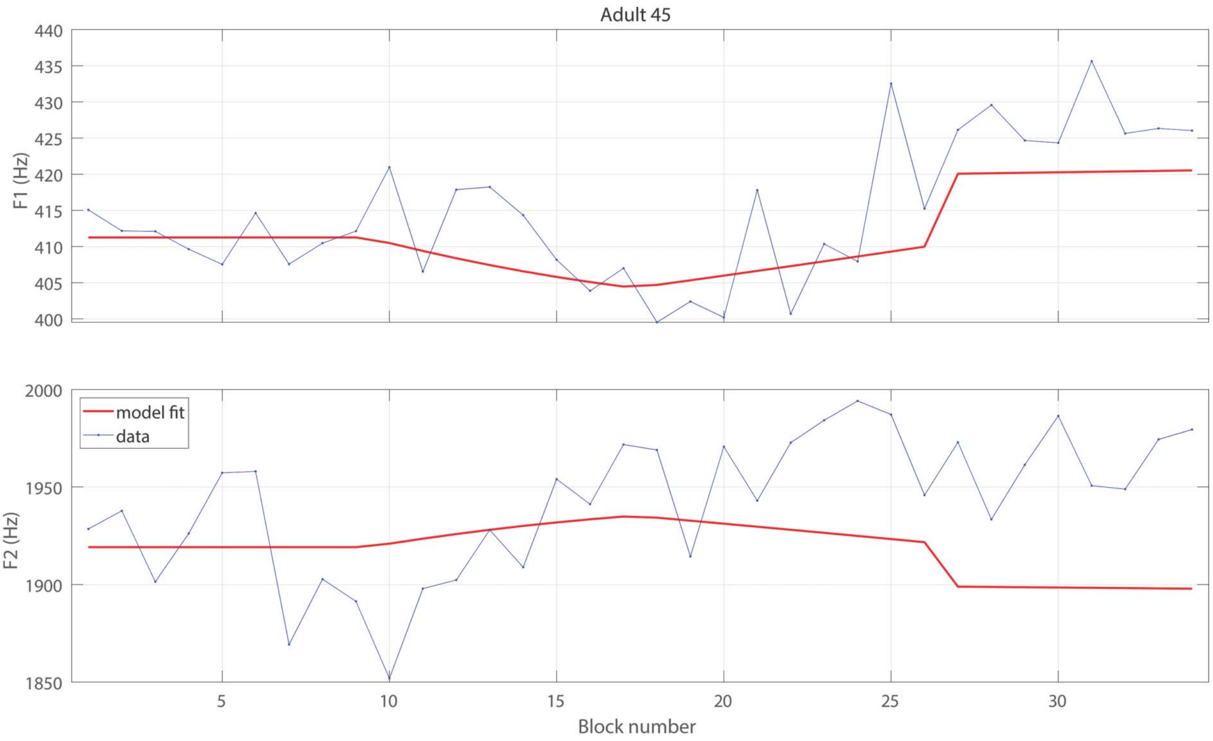
Figure C2. Model fit for the data of Adult 17 with perturbations simultaneously applied to both F1 and F2, exemplifying a medium fit (*r* = .52). Mean and standard error of experimental data in blue; model fit in red.



Appendix C (p. 3 of 4)

Examples of Model Fits of Individual Subject Data

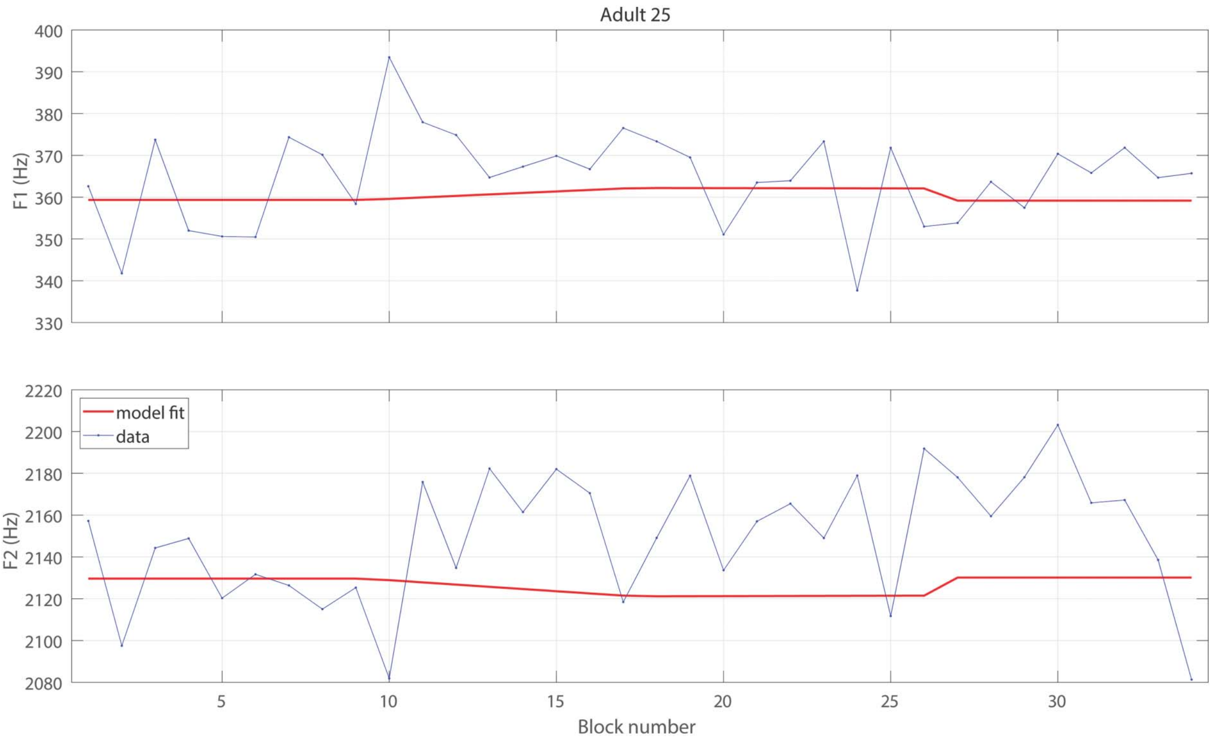
Figure C3. Model fit for the data of Adult 45 with perturbations simultaneously applied to both F1 and F2, exemplifying a medium fit (r = .49). Mean and standard error of experimental data in blue; model fit in red.



Appendix C (p. 4 of 4)

Examples of Model Fits of Individual Subject Data

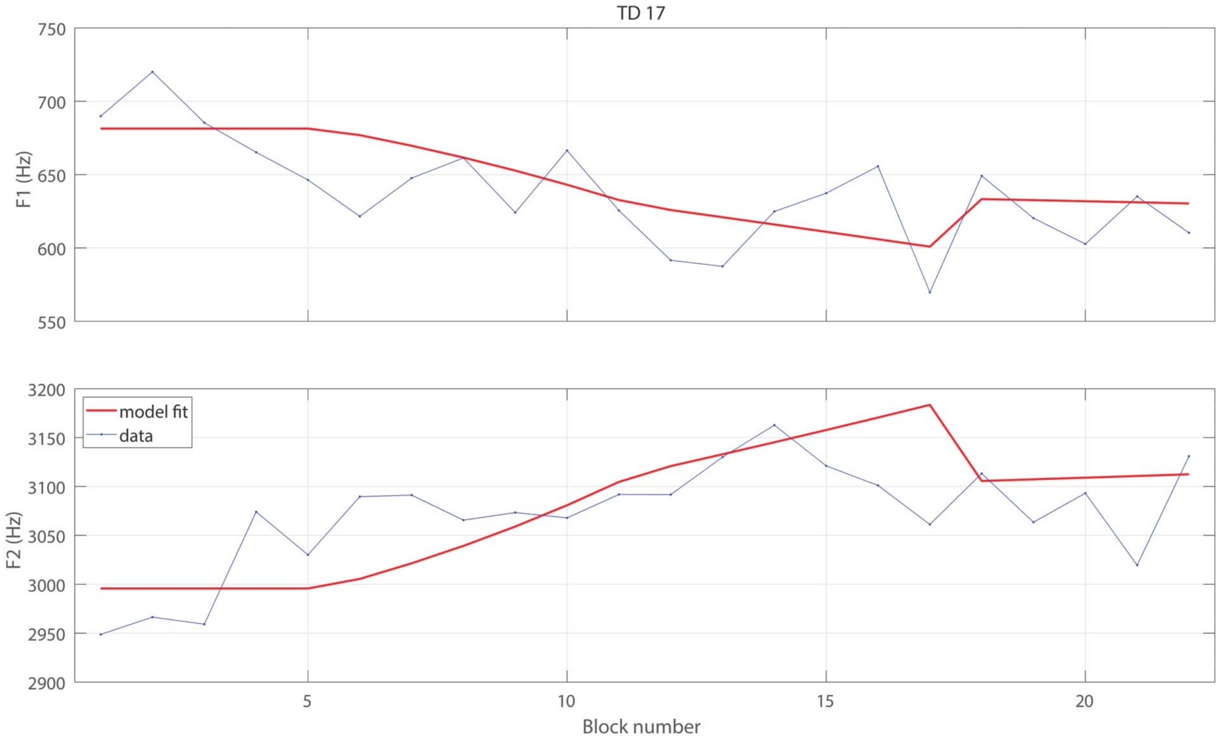
Figure C4. Model fit for the data of Adult 25 with perturbations simultaneously applied to both F1 and F2, exemplifying a bad fit (*r* = .06). Mean and standard error of experimental data in blue; model fit in red.



Appendix D (p. 1 of 2)

Examples of Model Fits of Individual Subject Data Illustrating Compensatory and Amplifying Behavior

Figure D1. Model fit for the data of a child with TD speech (TD17) with perturbations simultaneously applied to both F1 and F2, exemplifying compensatory behavior (r = .86; α_A = .24; α_S = −.37; λ_{FF} = .11). Mean and standard error of experimental data in blue; model fit in red.



Appendix D (p. 2 of 2)

Examples of Model Fits of Individual Subject Data Illustrating Compensatory and Amplifying Behavior

Figure D2. Model fit for the data of a child with SSD (SSD2) with perturbations simultaneously applied to both F1 and F2, exemplifying a following/amplifying response (r = .81; α_A = −.23; α_S = −.31; λ_{FF} = .08). Mean and standard error of experimental data in blue; model fit in red.