

Speakers with Cerebellar Ataxia Do Not Adapt Speech Segment Durations in Response to Durationally Altered Auditory Feedback

Robin Karlin^{1,2} and Benjamin Parrell^{2,3}

¹Department of Speech, Language, and Hearing Sciences, University of Missouri, Columbia, Missouri 65211, ²Waisman Center, University of Wisconsin – Madison, Madison, Wisconsin 53705, and ³Department of Communication Sciences and Disorders, University of Wisconsin – Madison, Madison, Wisconsin 53705

Sensorimotor adaptation (motor learning) is crucial to maintain movement accuracy. This is particularly true for temporal control, as duration and initiation timing are perceived only after the movement and thus cannot benefit from mid-movement adjustments. The cerebellum is critical for both sensorimotor adaptation and various temporal functions, but its role in temporal adaptation has not been investigated. In this study, we examine the effect of cerebellar damage on temporal adaptation in speech, a natural but complex task that requires millisecond precision. We use an altered auditory feedback paradigm to extend the perceived duration of the vowel /ε/ in the word “best.” Neurobiologically healthy controls (12 male, 22 female) adaptively shorten their vowels, but temporal adaptation is abolished in speakers with cerebellar damage (10 male, 24 female), who instead lengthen their vowels in response to the lengthening perturbation. This departs from findings in the spatial domain, where cerebellar damage merely reduces sensorimotor adaptation. A subset of speakers with cerebellar damage have lower perceptual acuity for vowel duration than controls; however, perceptual deficits did not drive the lack of adaptation. We suggest that temporal adaptation in speech is abolished due to cerebellar involvement in three aspects of temporal control: duration representation, temporal coordination, and temporal estimation. As speech motor control is highly complex, future research is needed to determine if the lack of adaptation observed in this study is due to combined demands on multiple cerebellar functions or critical damage to a single function that is common to all temporal control.

Key words: ataxic dysarthria; cerebellum; motor learning; speech; speech motor control; timing

Significance Statement

Speech is a complex motor activity that requires millisecond-level accuracy, thus providing a unique window to temporal control. When exposed to external perturbations that extend the perceived duration of vowels, speakers with cerebellar damage totally failed to adapt their speech to oppose the perturbation. Although a subset of speakers with cerebellar damage also had impaired perception of speech segment durations compared with controls, these perceptual deficits did not drive the lack of adaptation in the group. The total lack of adaptation departs from previous studies of sensorimotor adaptation in spatial control and emphasizes the critical role of the cerebellum in maintaining accurate movement timing.

Introduction

Speech requires millisecond-level accuracy. Differences in the duration of speech sounds mark lexical contrasts, signal word or phrase boundaries, and convey other paralinguistic information (Chen, 1970; Murray and Arnott, 1993; Byrd et al.,

2006; Shatzman and McQueen, 2006; Schertz and Clare, 2020; Maddieson, 2024). Ataxic dysarthria, the motor speech disorder associated with cerebellar damage (CD), is strongly associated with abnormal timing (Ackermann and Hertrich, 1997; Hartelius et al., 2000; Schalling and Hartelius, 2013;

Received Dec. 16, 2025; revised June 2, 2026; accepted June 16, 2026.

Author contributions: R.K. and B.P. designed research; R.K. performed research; R.K. analyzed data.

This work was supported by grants to B.P. (R01 DC017091) and R.K. (F32 DC019535) from the National Institute on Deafness and Other Communication Disorders, as well as a core grant to the Waisman Center from the National Institute of Child Health and Human Development (P50 HD105353) and an National Institutes of Health High-End Instrumentation grant (S10 OD030415). We also thank Kristie Spencer for providing dysarthria severity ratings.

The authors declare no competing financial interests.

Correspondence should be addressed to Robin Karlin at rkarlin@health.missouri.edu.

This paper contains supplemental material available at: <https://doi.org/10.1523/JNEUROSCI.2289-25.2026>

<https://doi.org/10.1523/JNEUROSCI.2289-25.2026>

Copyright © 2026 the authors

Spencer and France, 2016), suggesting that the cerebellum is critical for temporal accuracy in speech. In this paper, we leverage the unique temporal demands of speech to investigate the contributions of the cerebellum to accurate temporal control.

Movement accuracy is maintained in part by detecting and responding to mismatches between predicted and received sensory feedback. Errors can be corrected by compensating during the same movement or adapting control for future movements (Shadmehr and Krakauer, 2008). Studies in spatial control (e.g., examining the position of the hand in reaching tasks) show the cerebellum is critical for adaptation: people with CD show decreased adaptation in both nonspeech (Martin et al., 1996; Straube et al., 2001; Maschke et al., 2004; Smith and Shadmehr, 2005; Morton and Bastian, 2006; Tseng et al., 2007) and speech tasks (Parrell et al., 2017, 2021). Recent studies show that neurobiologically healthy speakers adapt speech sound durations (Mitsuya et al., 2014; Floegel et al., 2020; Oschkinat and Hoole, 2020, 2022; Karlin and Parrell, 2022), indicating that, like spatial control, temporal control uses sensory feedback loops. In fact, because the duration of a movement can only be determined after it ends, and the timing of movement initiation cannot be altered after the fact, adaptation is particularly critical for temporal control.

The process of adaptation requires both the detection of errors and the selection of motor commands that will reduce the error. In the temporal domain, both processes involve functions that are thought to rely on the cerebellum. First, the cerebellum is crucial for representation of temporal targets (Ivry, 1993; Zelaznik et al., 2005; Gerwig et al., 2007; Dimitrova et al., 2008; Timmann et al., 2008); degraded temporal targets may interfere with accurate selection of a set of motor commands and/or comparing received duration with a target duration. Second, CD is associated with increased perceptual thresholds for subsecond durations of both speech and nonspeech stimuli (Ivry and Keele, 1989; Ackermann and Hertrich, 1997; Ackermann et al., 1999; Grube et al., 2010; Teki et al., 2011, 2012); CD could thus impede detection of temporal errors. Finally, as the sequence of speech sounds is formed by a precisely overlapping set of articulatory movements, the duration of any speech segment (discrete unit of sound) may be affected by the initiation and/or offset timing of the surrounding movements. Evaluating the timing of movement initiation requires interpreting sensory input through a model of sensorimotor delay (Stetson et al., 2006; Sugano et al., 2010, 2012). The cerebellum has been implicated both in theoretical models of sensory delay (Miall et al., 1993) and experimental assays of adaptation to new delays (Cao et al., 2017); as such, deficits in delay calibration could impede the detection and/or correction of temporal errors.

The current study investigates how the cerebellum contributes to accurate duration control, as mediated by sensorimotor adaptation, perception of millisecond durations, and detection of sensory delay. We test this by temporally altering auditory feedback of speech and assessing changes in speech segment durations in speakers with CD and neurobiologically healthy controls. Given previous work on the heavy involvement of the cerebellum in both spatial adaptation and temporal control, we predict that speakers with CD will show impaired temporal adaptation of speech sound durations. To ensure that any changes in adaptation or effects of auditory delays are not due to perceptual deficits, we additionally assess potential CD deficits on the perception of vowel durations, as well as on the reaction to sensory delay.

Materials and Methods

Participants

We compare a group of participants with cerebellar damage (CD) and a group of age-matched, neurobiologically healthy controls. Participants with CD have a previous diagnosis of ataxia and did not report any history of speech, hearing, or neurological disorders other than those related to their ataxia diagnosis. The majority of participants were diagnosed with a subtype of spinocerebellar ataxia, a group of genetic disorders that affect the cerebellum and potentially other areas of the brain (Seidel et al., 2012). Control participants were age matched on a population basis to the CD group and did not report any history of speech, hearing, or neurological disorders. (See Table 1 for details.)

Prior to data analysis, four controls and two speakers with CD were excluded due to low performance on a composite cognitive assessment score suggestive of mild cognitive impairment (Carson et al., 2018), alone or in conjunction with low performance on other neuropsychological assessments (for more details, see below, Statistical analysis). An additional four speakers with CD and two controls were excluded due to an error in elicitation that led to the participant unnaturally extending

Table 1. Characteristics of included participants

Diagnosis	Age	Gender	SARA	Dysarthria severity	MoCA + CCAS
AOA2	46	F	24	2	26
CANVAS	70	F	20.5	3	23
Cerebellar tumor	52	F	9	0.5	26
SCA (unknown subtype)	25	F	11	1	25
SCA1	39	F	1	0	29
SCA2	69	F	8.5	2	28
SCA2	52	F	10	1	27
SCA2	69	M	10	2	23
SCA2	44	F	7.5	0.5	28
SCA2	65	M	11	1	27
SCA2	47	F	14	2	30
SCA2	60	M	12	1	25
SCA2	34	F	11.5	2	30
SCA2	68	F	17.5	2	24
SCA3	71	F	30	3	27
SCA3	64	F	6	2	27
SCA3	59	M	2.5	0	29
SCA3	42	F	9	0.5	27
SCA3	75	F	1	0.5	26
SCA3	35	F	1	0	30
SCA4	52	F	1	1	27
SCA4	35	M	10.5	0.5	23
SCA5	32	F	2	0	29
SCA5	56	F	12.5	1	28
SCA6	65	F	16	2	24
SCA6	74	M	9	1	23
SCA6	59	M	10	0.5	26
SCA6	65	F	1	0	27
SCA8	60	F	13.5	2	28
SCA8	39	M	2.5	0	26
SCA8	63	M	15.5	2	26
SCA8	73	F	19	2	26
SCA14	54	F	8.5	1	29
Idiopathic	53	M	27	2	24
Total CD	Median = 57.5, SD = 13.8	10 M, 24 F	10.7 (7.5)	—	26.6 (2.1)
Total control	Median = 59.5; SD = 14.5	12 M, 22 F	—	—	27.6 (1.8)

SARA is rated from 0 (no symptoms) to 30.

Dysarthria severity is rated from 0 (no dysarthria) to 4. MoCA + CCAS rated from 0 to 30 (no cognitive impairment).

AOA2, ataxia with oculomotor apraxia type 2; CANVAS, cerebellar ataxia with neuropathy and vestibular areflexia syndrome; SCA, spinocerebellar ataxia; SARA, scale for the assessment and rating of ataxia; MoCA, Montreal Cognitive Assessment; CCAS, cerebellar cognitive and affective syndrome.

the vowel for ~ 1 s. One more control was excluded due to failure to follow instructions. Four more speakers with CD were unable to complete the task due to very severe dysarthria. The final dataset for this study thus comprises 34 participants with CD (10 male, 24 female) and 34 neurobiologically healthy controls (12 male, 22 female).

Data was collected at the University of Wisconsin–Madison or the University of California (San Francisco and Berkeley campuses). All procedures were carried out in accordance with the Institutional Review Board at the University of California, San Francisco.

Apparatus

Task 1. Audio recordings were made with a Sennheiser MKE 600, which was routed to a Scarlett 2i2 3rd generation audiobox and from there to a desktop computer. Audio feedback, mixed with white noise, was played back over Beyerdynamic DT 770 250 ohm closed-back circumaural headphones. Stimuli were displayed in white font on a black background on a computer monitor ~ 2 feet away from the participant. The task was implemented in MATLAB (MATLAB, 2023). The measured round-trip system delay between participant production and playback of auditory feedback was 16.0 ms.

In this task, the vowel / ϵ / in “best” was lengthened by an intended maximum target of 60 ms. A maximum target of 60 ms was chosen as it has previously been shown to provoke temporal adaptation in controls, without being so overt that it promotes explicit awareness of the perturbation (Karlín et al., 2021; Karlín and Parrell, 2022). Temporal perturbation of specific segments was implemented with Audapter (Cai et al., 2008; Tourville et al., 2013). Audapter produces targeted “time warping events” via an online status tracking (OST) capability, which uses root mean square intensity to detect changes in the acoustic signal, such as segment boundaries (Cai et al., 2008). OST settings were individualized for each participant during a pretest phase prior to the main production task using Audapter Viewer (Naber et al., 2024). When an event such as the onset of / ϵ / is detected, Audapter starts a *slowdown* period, which slows down a specified slice of the audio signal by a given factor. In this task, the slowdown period started 50 ms after the vowel onset was detected, to avoid extending the formant transition and giving a percept of “bwest”; the following 20 ms were played back at 0.25 original speed to produce 60 ms of lengthening. After the slowdown, Audapter enters a *stasis* period of a given duration, where the signal is played back unaltered, other than the delay induced by lengthening a prior slice of the signal. Audapter then enters a *catch-up* period, where it plays the signal back at a specified accelerated rate until it catches up to true time. In order to maintain durations for the remainder of the utterance, the stasis period was set based on the duration of / ϵ st/ plus three standard deviations, using a running average of the previous ten trials. Thus, the catch-up period occurred after the participant had finished the target word (see Fig. 1 for illustration of the perturbation), such that the final /st/ of “best” was delayed but replayed with an unaltered duration. Experimental code is available at <https://osf.io/rfq75/>.

The mean achieved perturbation of / ϵ / during the hold phase was 58.0 ms. A linear mixed effects model with participant as random intercept indicates that there was no significant difference between groups ($t_{(68,01)} = 0.54, p = 0.59$). The mean delay of the onset of /s/ was 58.1 ms; there was no significant difference between groups ($t_{(70,00)} = 0.26, p = 0.80$). As intended, the duration of /s/ was unaffected; the mean difference in the duration of this segment between the input and output signals was 0.1 ms. There was a significant difference between groups ($t_{(68,08)} = -2.24, p = 0.03$) but both groups received <1 ms of perturbation (CD mean = -0.7 ms, SD = 1.6 ms; control mean = 0.8 ms, SD = 3.5 ms).

Task 2. All participants also completed a perceptual task to assess duration thresholds on the vowel in “best”. Stimuli were played at a comfortable volume over Beyerdynamic DT 770 250 ohm closed-back circumaural headphones. The stimuli were synthesized from a token of “best” produced by a native speaker of American English. The vowel had a flat intonational contour to avoid effects of pitch contour on duration perception (Cumming, 2011). To avoid lengthening the formant transitions at the beginning and end of the vowel, which could affect both the naturalness and the duration percept of the vowel, the middle 73 ms of the

vowel was programmatically altered using the duration manipulation function in Praat (Boersma and Weenink, 2021) to produce tokens along a continuum of vowel durations from 100 to 200 ms in 1 ms steps (original token duration: 163 ms). Premade tokens were used to avoid potential pitfalls in manipulating the duration of tokens produced by the participants themselves, as speakers are often unable to consciously control things such as the flatness of their pitch or the duration of their formant transitions, both of which can affect duration percepts or produce unnatural tokens when manipulated.

Experimental design

Task 1: sensorimotor adaptation of speech sound durations. Participants produced the target word “best” on each trial. Based on previous studies that have shown that vowels show more robust adaptation responses (Oschkinat and Hoole, 2020, 2022), we tested adaptation by lengthening / ϵ /, with consequent delay of /st/. A representative trial is shown in Figure 1. The task consisted of 90 trials: a 20-trial baseline phase, where participants heard veridical feedback; a 20-trial ramp phase, where the perturbation was increased by 3 ms per trial; a 30-trial hold phase, where each trial received maximum 60 ms perturbation; and a 20-trial washout phase, where the perturbation was abruptly removed, such that participants received veridical feedback. Before the start of the task, the experimenter instructed participants to say “best” “like they are answering a question”.

Task 2: perceptual acuity of vowel duration. Participants completed a 3:1 adaptive staircase procedure, in which a correct response reduces the difference between tokens by a given step size, and an incorrect response increases the difference between tokens by that step size multiplied by three. A 3:1 ratio targets performance accuracy of 79%. Correct responses initially decreased the difference between tokens by 10 ms. The step size was reduced to 5 ms (5 ms down; 15 ms up) after the first two reversals, where reversal is defined as a trial that has the opposite correctness as the trial before it—that is, an inaccurate trial following an accurate trial or an accurate trial following an inaccurate trial.

The task for each trial was an AXB task, where tokens A and B are different from each other, and participants respond whether the X token is the same as A or B. Tokens were presented 1.25 s apart, with a jitter of ± 100 ms to prevent the use of rhythmic timing circuits (Teki et al., 2011, 2012). The difference between tokens started at 100 ms, the maximum possible. A 100 ms difference between tokens of 100 and 200 ms was chosen as a starting point as it is well above the Weber fraction for temporal perception in this range (Grondin et al., 2001); all participants were able to hear this difference. Tokens were presented symmetrically around 150 ms: thus, for a stimulus difference of 100 ms, the shorter token was 100 ms, and the longer token was 200 ms. The position (A vs B) of the shorter and longer stimuli was determined randomly for each trial, as was the correct answer. Participants continued the task until they completed 100 trials or 30 reversals, whichever came first (all participants reached 30 reversals before 100 trials). Experimental code is available at <https://osf.io/rfq75/>.

Additional assessments. Participants with CD completed assessments for CD symptom severity. For general motor symptoms, participants completed the Scale for the Assessment and Rating of Ataxia (SARA) (Schmitz-Hübsch et al., 2006). For speech symptoms, participants produced conversational speech, sentence repetition, passage reading, and a picture description task, following the Bogenhausen Dysarthria Scales for English (BoDyS-English; Ziegler et al., 2017; Staiger et al., 2023). Participants also completed a motor speech assessment, including diadochokinetic tasks (e.g., repetitions of “papapa...” or “pataka...”) and a sustained phonation task. Speech diagnosis and severity rating was performed by a licensed speech language pathologist with extensive experience working with ataxic dysarthria. Participants were rated on a scale from 0 to 4, where 0 represents no apparent symptoms of dysarthria and 4 represents severe dysarthria; no participants in our sample had severe dysarthria. All participants with dysarthria only showed symptoms of ataxic dysarthria (i.e., no other motor speech disorders).

SARA ratings were performed by the experimental team. See Table 1 for a summary of the participants in this study.

All participants also completed a cognitive assessment that consisted of the Cerebellar Cognitive Affective/Schmahmann Syndrome scale (CCAS scale; Hoche et al., 2018) and items from the Montreal Cognitive Assessment (MoCA; Nasreddine et al., 2005) scale that did not duplicate items from CCAS scale. Assessments were scored by the experimental team. The reported MoCA score is the score of the items from MoCA plus items from CCAS that duplicated MoCA items. Individual scores for participants with CD and group-level statistics for controls are reported in Table 1.

Participants also completed an automated Hughson–Westlake hearing screening at 250, 500, 1,000, 2,000, and 4,000 Hz. Due to the rarity and typical age range of the target disorder, we did not exclude participants based on hearing thresholds. Average hearing thresholds for each group are available in Table S1. No participant used an assistive hearing device (e.g., hearing aids).

Statistical analysis

Task 1. The audio from participants' productions was first automatically segmented using Montreal Forced Aligner (McAuliffe et al., 2017) and then hand corrected to obtain the duration of each segment in the target utterance. Raters were not blinded to group, as it is possible to discern the presence of ataxic dysarthria from looking at or listening to the speech signal, which is required for data processing. Raters were also not blinded to trial but were generally unaware of the hypotheses of the study. Raters followed the segmentation guidelines below in performing hand correction:

1. Onset of the vowel /ε/ is marked where regular voicing and formant structure start.
2. Onset of the /s/ is marked at the beginning of high frequency noise.
3. End of the /s/ (beginning of /t/) is marked at the beginning of a rapid decrease in amplitude, denoting the increased constriction for /t/.

The last 10 trials of the baseline phase serve as a baseline of comparison for changes across phases of the task. For the vowel /ε/, the last 10 trials of the hold phase serve as a measure of adaptation. Because the perturbation lengthened the vowel duration, any reduction in duration during the hold phase must be due to adaptive rather than reactive changes; since we perturbed the vowel by extending the end of the vowel past the real endpoint (rather than advancing the start of the vowel), the speaker has already moved onto producing /s/ when they receive the auditory error for the vowel and thus cannot reactively adjust the duration of the vowel within the trial (either lengthening or shortening). The first 10 trials of the washout phase serve as a measure of retention of adaptation. For /s/, the last 10 trials of the hold phase serve as a measure of reaction to delay, as there was no duration perturbation on this segment; the first 10 trials of the washout phase serve as a measure of retention of delay effects. Reported estimated means are the change in production compared with baseline.

Statistical analyses were conducted with linear mixed effects models using the lme4 package in R (Bates et al., 2014; R Core Team, 2024). Models included random intercepts for participant. By-participant random slopes for phase were not included due to failure to converge. Models were built incrementally, starting from the null model, and then adding fixed effects of phase [levels: baseline (intercept), hold, washout] and group [levels: control (intercept), CD], as well as the interaction between the two. Fixed effects were included in the final model if there was a significant improvement in model fit, as per chi-square and *p* values provided by the analysis of variance function to compare models (Kuznetsova et al., 2015). Effect sizes and post hoc comparisons were calculated using the emmeans package (Lenth, 2019); post hoc comparisons were corrected using the Tukey method. Statistical code is available at <https://osf.io/rfq75/>.

Task 2. Thresholds are calculated as the average of the last 20 reversals (Kollmeier et al., 1988; Kaernbach, 1991; Leek et al., 1992; Garcia-Perez, 2011). One control participant had a perceptual threshold beyond the third quartile plus 1.5 times the interquartile range for their group

and was excluded from both perceptual analyses and correlation analyses between perception and other tasks. Group-level differences were assessed with an ANOVA.

Relationships between tasks. To examine the relationships between speech production, speech perception, and severity of ataxia and dysarthria symptoms, we also conducted linear regressions using the *lm()* function from the base stats package in R. To obtain a single measure per participant for each speech production task, we took the mean change from baseline from the last 10 trials of hold for both /ε/ and /s/ durations. We used hold rather than washout because they both reflect the maximum response in either learning (/ε/) or reaction (/s/), without potential de-adaptation effects—recall that shortening responses in /ε/ are necessarily adaptive, with no possible compensatory component.

For these correlation analyses, we shifted participants' thresholds by subtracting the minimum threshold in the study (21.75 ms). This ensures an interpretable intercept for these models, rather than a methodologically and biologically impossible 0 ms threshold.

Since it is possible that participants with perceptual thresholds higher than the magnitude of perturbation may not have perceived the feedback error, we also conducted secondary analyses of temporal adaptation where we divided speakers into two groups: those with high perceptual thresholds (defined as at or above 60 ms, the magnitude of perturbation for this study) and those with low perceptual thresholds (below 60 ms). Two controls fell into the high perceptual threshold group. One is the previously mentioned statistical outlier, who was removed from all perceptual analyses. The remaining control with a high perceptual threshold was removed from these secondary analyses only. Thus, these analyses were grouped as a three-level factor: speakers with CD and high thresholds; speakers with CD and low thresholds; and controls.

Results

Adaptive and reactive responses to temporal perturbation

Our primary goal is to determine the role of the cerebellum in temporal adaptation of motor behavior, using the lens of sensorimotor adaptation of temporal properties of speech. In this study, we follow previous work that has shown that neurobiologically healthy speakers adapt the durations of speech segments in response to auditory perturbations. Participants said the word “best” in every trial; a duration perturbation that lengthened the vowel /ε/ was gradually introduced and then sustained at a maximum of 60 ms for 30 trials (Fig. 1A; analysis windows highlighted). Linear mixed effects models were used to test effects of group (control vs CD) and phase (baseline, hold, washout) and the interaction of these two factors.

Overall, the results indicate that only control speakers showed adaptive shortening in response to the auditory perturbation (Fig. 2A,B). Both the main factors phase ($\chi^2_{(2)} = 68.62$, $p < 0.0001$) and group ($\chi^2_{(1)} = 11.94$, $p = 0.0005$) significantly improve model fit, as does the interaction between phase and group ($\chi^2_{(2)} = 133.97$, $p < 0.0001$). Speakers with CD had significantly longer vowel durations in the baseline phase compared with controls ($\chi^2_{(1)} = 218.55$, $p < 0.0001$; CD mean = 235.7 ms, SD = 51.0 ms; control mean = 198.8 ms, SD = 42.6 ms), which likely reflects the slowed speech that is common in ataxic dysarthria; however, the groups show considerable overlap, and the group differences are substantially smaller than the magnitude of the applied lengthening perturbation (Fig. 2B). Compared with baseline, control speakers had significantly shorter /ε/ during both the hold phase (-16.2 ± 2.7 ms, $p < 0.0001$, $t = 11.03$, $d = 0.85$) and during the washout phase (-16.0 ± 2.7 ms, $p < 0.0001$, $t = 10.94$, $d = 0.84$). There was no significant difference between the hold and washout phases ($p = 0.86$). In contrast, speakers with CD did not show any evidence of adaptation. Rather, the vowel duration was actually longer during the hold phase than in the

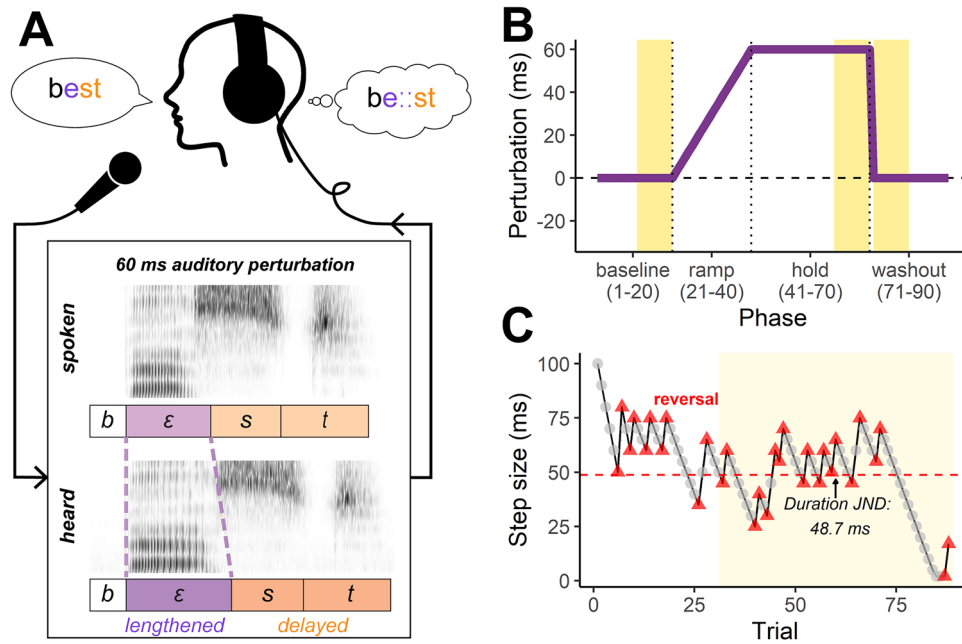


Figure 1. Experimental methods. **A**, Schematic of duration perturbation on the word “best” during the adaptation task, showing the lengthening of /ε/ (purple) and the delay of /st/ (orange). **B**, Schedule of the duration perturbation during the adaptation task. **C**, Example responses from the AXB perceptual staircase task for the duration of /ε/ in “best.” Reversals are indicated in red. The perceptual threshold is calculated as the mean of the last 20 reversals and is marked with a red dashed line.

baseline (8.12 ± 2.7 ms, $p < 0.0001$, $t = -5.53$, $d = -0.43$); productions during the washout phase did not significantly differ from baseline (-1.8 ± 2.7 ms, $p = 0.81$, $t = 1.25$). There was a significant difference between the hold and washout phases ($p < 0.0001$). Within each experimental phase, speakers with CD and controls were significantly different from each other: control speakers shortened their productions of /ε/, while speakers with CD did not in both the hold phase ($p < 0.0001$, $t = -6.40$, $d = -1.27$) and the washout phase ($p = 0.004$, $t = -3.74$, $d = -0.74$).

Because we generated the duration perturbation by delaying the end of the segment (rather than anticipating the beginning), it could be the case that speakers were detecting and learning something about the initiation timing of the /s/ following the /ε/, rather than only detecting and learning something about the duration of the /ε/ per se. Crucially, models of sensory delay would be required for making accurate predictions about the timing of sensory input (and thus to detect deviations from those predictions). As the cerebellum is thought to house these delay models, CD may impair the ability to detect feedback delays, leading to a failure to adapt in this group. To address this possibility, we measured changes in the production of /s/, whose onset was delayed by 60 ms in the auditory feedback participants received because of the lengthening perturbation applied to the vowel /ε/. As auditory feedback delays are known to cause online (reactive) slowing reactions (Karlín and Parrell, 2022; Malloy et al., 2022) and online reactions are generally not considered to be the domain of the cerebellum (see, e.g., typical compensation rates in people with CD; Nowak et al., 2004; Parrell et al., 2017, 2021; Houde et al., 2019; Li et al., 2019; Zimmet et al., 2020; Slis and Parrell, 2025), if the CD group shows appropriate compensatory slowing to the delayed onset of /s/, we can rule out the possibility that the failure to adapt vowel duration in the CD group arises from a failure to detect this delay.

Overall, the results indicate that control speakers and speakers with CD lengthened the production of /s/ during the hold phase by an equivalent amount, indicating that both groups were able

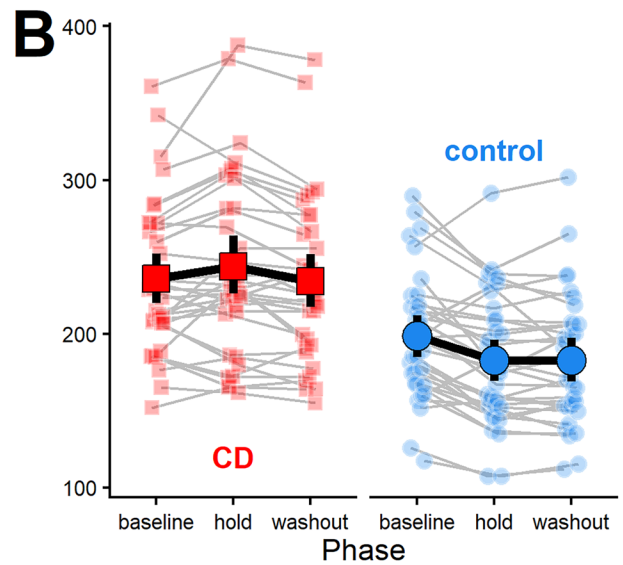
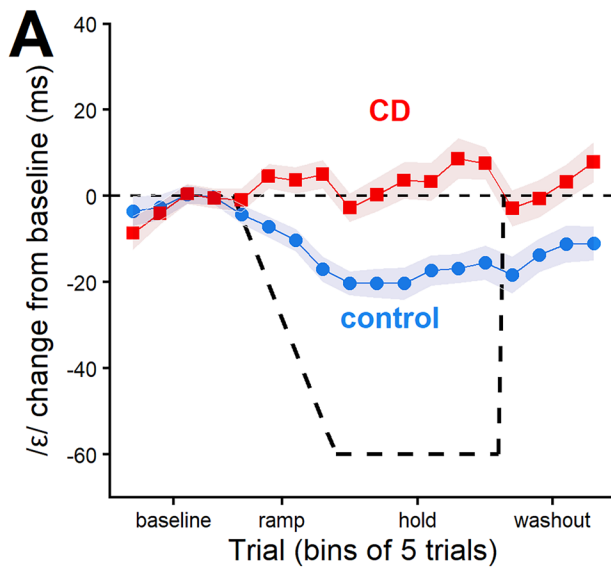
to detect the feedback delay. Adding phase as a fixed effect significantly improved model fit over the null model ($\chi^2_{(2)} = 357.11$, $p < 0.0001$), but adding group did not ($\chi^2_{(1)} = 3.6$, $p = 0.06$). Both control speakers and speakers with CD produced significantly longer /s/ during the hold phase (control: 16.5 ± 2.8 ms, $t = -8.85$, $d = -0.68$; CD: 27.1 ± 2.8 ms, $t = -14.62$, $d = -1.11$; both $p < 0.0001$ compared with baseline). Neither group retained this lengthening into the washout phase (Fig. 2C,D). During the washout phase, both groups returned approximately to baseline durations; controls slightly shortened /s/ compared with baseline (-6.7 ± 3.0 ms, $t = 3.66$, $p = 0.004$, $d = 0.28$), while there was no difference between baseline and washout productions for speakers with CD (3.7 ± 2.8 ms, $t = -20.4$, $p = 0.32$). Adding the interaction between phase and group significantly improved model fit ($\chi^2_{(2)} = 21.79$, $p < 0.0001$). There was a marginal difference between groups in both the hold phase ($p = 0.09$) and the washout phase ($p = 0.10$). Together, these two patterns lend support to the idea that the failure to adapt vowel duration in the CD group is not due to a failure to detect temporal delays.

Finally, we examined the relationship between adaptive behavior (vowel length) and reactive changes (/s/ lengthening), as measured from the end of the hold phase. In a linear regression model that includes both group and change in /s/ duration as a predictor of change in /ε/ duration, there is not a significant relationship between change in /ε/ and change in /s/ ($p = 0.18$), nor is there a significant interaction between group and change in /s/ duration ($p = 0.49$). This aligns with previous findings in other domains that the magnitude of feedback control-based responses and predictive control-based responses is not correlated either in controls (Franken et al., 2019; Raharjo et al., 2021) or in CD (Parrell et al., 2017).

Temporal perception of /ε/

Previous literature has reported that perception of interval-based durations is impaired in CD, reflected in higher perceptual thresholds for some duration contrasts in speech, but not

Duration adaptation: /ε/



Reaction to delay: /s/

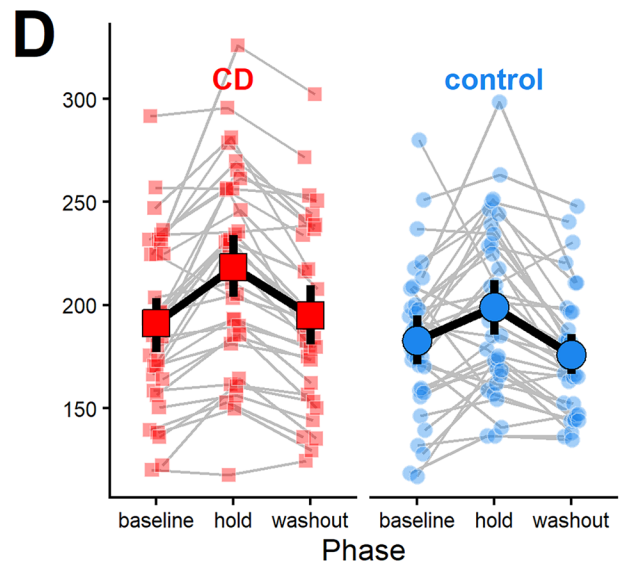
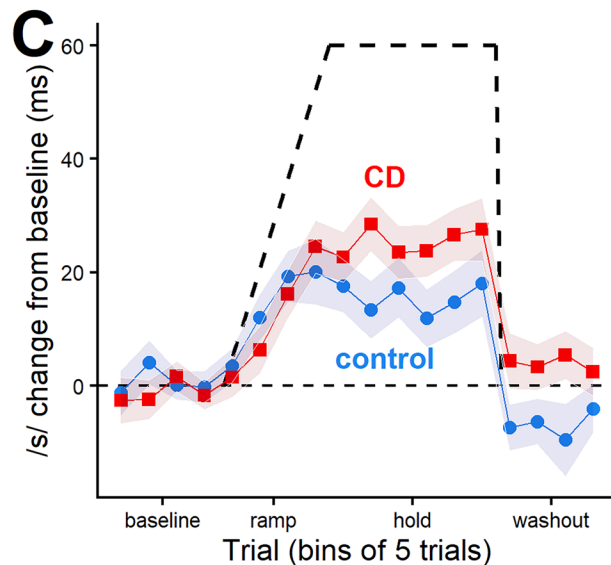


Figure 2. Task 1 results. **A**, Behavioral results showing the response to temporal perturbation of the vowel /ε/ in "best" in the CD group (red) and the control group (blue). Productions are grouped into bins of five trials. Error shading indicates 95% confidence interval. 100% response indicated with dashed black line. **B**, Individual duration of productions of the vowel /ε/ during the last 10 trials of the baseline and hold phases, as well as the first 10 trials of washout. Small dots show individual participants, and large dots show group means and standard errors. **C**, **D**, As for **A** and **B**, showing produced duration of the /s/ in "best." The dashed line in **C** represents the maximal expected response due to feedback-driven compensation, rather than any adaptive effect.

others (Ivry and Gopal, 1993; Ackermann et al., 1997). To date, perception of vowel duration has not yet been examined. As sensorimotor adaptation is dependent on the detection of sensory errors, it is possible that differences in sensorimotor adaptation we observed in Task 1 could have been driven by higher sensory thresholds for vowel duration perception in the CD group, leading to a failure to detect the duration error, rather than by an impairment in adapting to this error per se. To address this question, we first measured perceptual thresholds for the duration of /ε/ in "best" using an adaptive staircase test. Because all participants heard the same premade tokens, we perform this analysis on raw ms thresholds, rather than on a proportion of the duration of the stimuli. (For a proportional

analysis, which produces the same overall pattern of results, see Text S1 and Fig. S1).

Consistent with previous reports in other perceptual domains, the CD group had significantly higher perceptual thresholds for vowels ($M = 47.8$ ms, $SD = 14.9$ ms) compared with controls ($M = 37.7$ ms, $SD = 10.4$ ms, $F_{(1,65)} = 10.18$, $p = 0.002$, Cohen's $d = 0.78$; Fig. 3A). There is substantial overlap between groups; the difference between groups seems to be driven largely by a large group of speakers with CD (10/34) with thresholds above 60 ms (the magnitude of perturbation for this study), compared with just one control with a threshold above 60 ms (in addition to one participant that was excluded from these analyses due to being a statistical outlier).

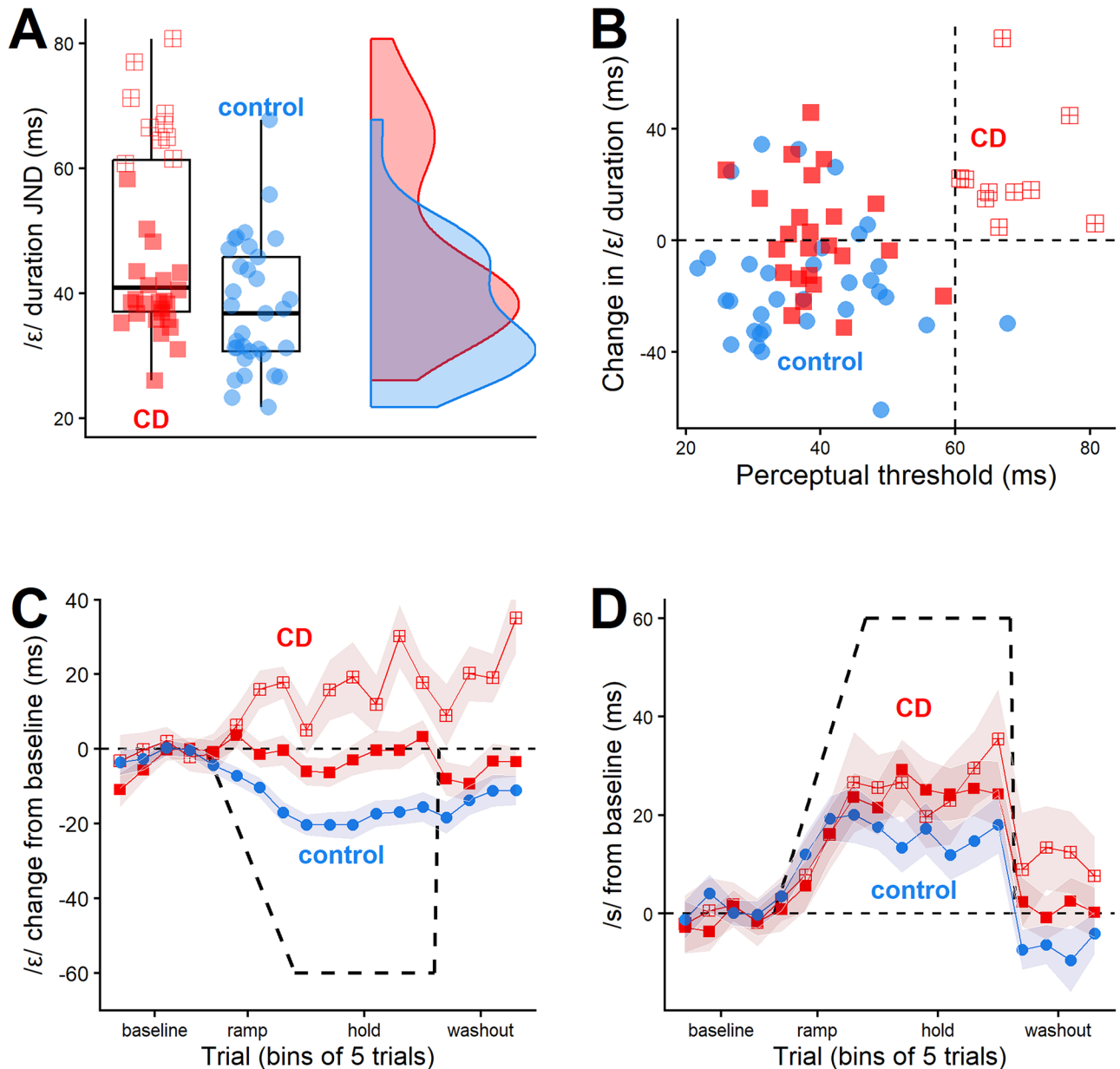


Figure 3. Task 2 results. **A**, Temporal perception thresholds in the CD group (red) and the control group (blue). Participants with CD that have perceptual thresholds above 60 ms are indicated with unfilled square markers. **B**, Relationship between perceptual thresholds for /ε/ duration and change from baseline in /ε/. **C**, **D**, Behavioral results for /ε/ and /s/, respectively, separating out the 10 participants with CD that have perceptual thresholds above 60 ms.

In the CD group, there is some contribution of perceptual threshold to magnitude of adaptation in the vowel /ε/. A linear regression model with group, threshold (normalized to the minimum threshold in the study), and their interaction as predictive factors of /ε/ adaptation shows a significant interaction between group and threshold; speakers with CD (but not control speakers) show less shortening of /ε/ duration with higher thresholds ($\beta = 0.52$, $p = 0.04$). In the model with the interaction term, the effect of group is not significant ($p = 0.53$).

The effect of threshold on adaptation magnitude seems to be principally driven by the group of 10 speakers with CD with perceptual thresholds above 60 ms; all speakers with high perceptual thresholds numerically increased their vowel durations during hold phase. To compare these two groups of participants with CD, we ran a second linear mixed effects model looking at the effect of phase and group on adaptation, this time separating

the high and low threshold participants with CD into two groups to compare them to each other and to controls. For this analysis, we also removed the remaining control participant with a threshold above 60 ms (see Materials and Methods, Relationships between tasks for more details).

This model showed statistical separation between controls, speakers with CD and high thresholds, and speakers with CD and low thresholds (significant model improvement with interaction term; $\chi^2_{(4)} = 183.66$, $p < 0.0001$). Speakers with CD and lower perceptual thresholds did not significantly change their productions during the hold phase (1.50 ± 3.01 ms, $p = 0.99$ compared with baseline; $p = 0.003$ compared with CD speakers with high thresholds; $p = 0.003$ compared with controls). That is, even if speakers with CD had adequate perceptual acuity to perceive the perturbation, they still did not adapt. In contrast, speakers with thresholds above 60 ms followed the perturbation and

lengthened the vowel (24.0 ± 4.66 ms, $p < 0.0001$ compared with baseline; $p = 0.003$ compared with CD speakers with low thresholds, and $p < 0.0001$ compared with controls; Fig. 3B,C). This is not predicted by an inability to perceive that the vowel was too long.

To examine the effect of threshold on adaptation in just the group of speakers with CD and low thresholds, we ran another linear regression model with just the group of CD speakers with low thresholds and controls (high threshold control removed). For these groups, there is no contribution of perceptual threshold to the magnitude of adaptation. Only group is significantly predictive ($\beta = -16.12$, $p = 0.006$ in a model with group alone); adding threshold does not improve model prediction ($p = 0.43$), nor does the interaction between threshold and group ($p = 0.36$). In sum, perceptual thresholds do not predict magnitude of adaptation per se; rather, thresholds above 60 ms correlate with abnormal production behavior, possibly only in speakers with CD.

Although learning in this task requires shortening, which cannot be a reaction after receiving a sensory feedback error about duration, the measure of change during the first 10 trials of washout additionally avoids any possible contributions from reactions to any other aspect of the perturbation (such as misalignments of pitch, desire to speak clearly after an error, etc.). In the washout phase, controls produced the vowel /ε/ with a significantly shorter duration compared with baseline (-14.34 ± 2.61 ms; $p < 0.0001$) but with the same duration as hold ($p = 1.00$). In contrast, speakers with CD and high perceptual thresholds produced /ε/ significantly longer than during baseline (14.62 ± 4.66 ms, $p < 0.0001$), but also significantly shorter than during the hold phase ($p = 0.01$). Speakers with CD but lower perceptual thresholds also decreased /ε/ duration between hold and washout (washout: -8.72 ± 3.01 ms, $p < 0.0001$ compared with baseline and hold). While this is a statistically significant result, given the lack of any evidence of change in vowel duration in the hold phase in this group (Fig. 3C), it is unclear if this measure truly reflects adaptation. Nevertheless, we ran an additional linear model examining the relationship between change from baseline in washout productions and perceptual thresholds. When added as a second predictor alongside group, perceptual threshold does not explain additional variability in /ε/ productions during washout ($p = 0.16$); however, the interaction between group and threshold is significant. For speakers with CD, higher thresholds are associated with increases in vowel duration ($\beta = 0.65$, $p = 0.02$), which again is driven by the group of speakers with thresholds above 60 ms. In contrast, higher thresholds in controls are associated with decreases in vowel duration ($\beta = -1.01$, $p = 0.04$)—i.e., there is an inverse relationship between acuity and magnitude of learning in controls. As for the hold phase, this further supports the idea that there is no relationship between perceptual acuity and sensorimotor adaptation of speech segment duration.

As even speakers with CD with lower thresholds were still unable to adapt vowel duration, these results strongly suggest that the failure of the CD group to adapt their vowel durations was not due to an inability to detect the applied temporal error, but perhaps an inability to control duration more generally. Further examination of this group also reveals generally higher variability in /ε/ duration, particularly during the hold and washout phases (Fig. 3C). As a simple measure of individual variability, we calculated the standard deviation of vowel durations in the analysis windows for each participant. Variability is highest in the group with CD and high thresholds (hold mean: 18.56 ms,

washout mean: 18.52 ms), followed by the group with CD but lower thresholds (hold mean: 15.95 ms, washout mean: 14.42 ms), and finally lowest in control speakers (hold mean: 11.02 ms, washout mean: 11.93 ms).

In addition, although we used a cutoff of a 60 ms perceptual threshold for this study because it is the same as the magnitude of perturbation in Task 1, a *k*-means clustering analysis ($k = 2$) on the perceptual thresholds of the CD speakers yields one cluster with all 10 speakers with thresholds above 60 ms, plus one additional speaker with a threshold of 58.3 ms. It should also be noted that 60 ms was selected as the perturbation magnitude based on previous studies with healthy controls (Karlin and Parrell, 2022), not in anticipation of any characteristics of the present population. Thus, ~60 ms may be a meaningful perceptual threshold independent of the specifics of the auditory perturbation in Task 1, though further support would be needed to confirm this idea. Speakers with CD and high thresholds also generally have higher SARA scores compared with the rest of the CD group (Fig. S2; for additional information about the characteristics of this group of participants, see Text S2). The higher perceptual thresholds and temporal variability may both be general symptoms of more severe ataxia and possibly more extensive cerebellar damage, rather than causally linked to each other.

Correlations of temporal behavior to overall dysarthria severity

Since precise timing is required for accurate speech production, we would predict a relationship between dysarthria severity scores (reflecting at least partly atypical temporal patterns) and our measure of vowel adaptation, which measures the ability to maintain temporally accurate speech. Indeed, there is a significant correlation between dysarthria severity score and magnitude of /ε/ change ($R^2 = 0.15$, $p = 0.03$); speakers with less severe dysarthria are somewhat more likely to adapt than individuals with more severe dysarthria. Interestingly, however, the data strongly suggests that having any dysarthria is correlated with lack of adaptation, rather than dysarthria severity gradiently predicting adaptation ability. In our sample, six speakers were judged to have no symptoms of dysarthria; their mean change from baseline was -13.6 ms (SD = 8.9 ms), which is numerically similar to the model estimate of -16.0 ms for controls; all speakers in this group had shorter vowels at the end of the hold phase compared with baseline. The remaining speakers ($n = 28$) have a mean change from baseline of 12.7 ms (SD = 21.1 ms); even the group of speakers with negligible dysarthria symptoms ($n = 7$) has a mean change from baseline of 9.7 ms (SD = 9.5 ms; Fig. 4A).

Conversely, there was no significant relationship between dysarthria severity and perceptual acuity ($R^2 = 0.08$, $p = 0.11$; Fig. 4B). This reinforces the idea that any deficits in perceptual acuity of duration in CD are not a driving factor in durational impairments in speech production in this population.

No relationship between overall ataxia symptoms and temporal behavior

The overall lack of temporal adaptation in the CD group, even among participants that had perceptual thresholds below 60 ms (the magnitude of perturbation in this study), suggests that temporal deficits occur independently of other deficits in ataxia. We thus assessed the relationship between general ataxia severity and temporal adaptation. There was not a significant correlation between SARA scores (speech subscore removed) and changes in /ε/ duration ($R^2 = 0.06$, $p = 0.15$; Fig. 5A). Given that the CD

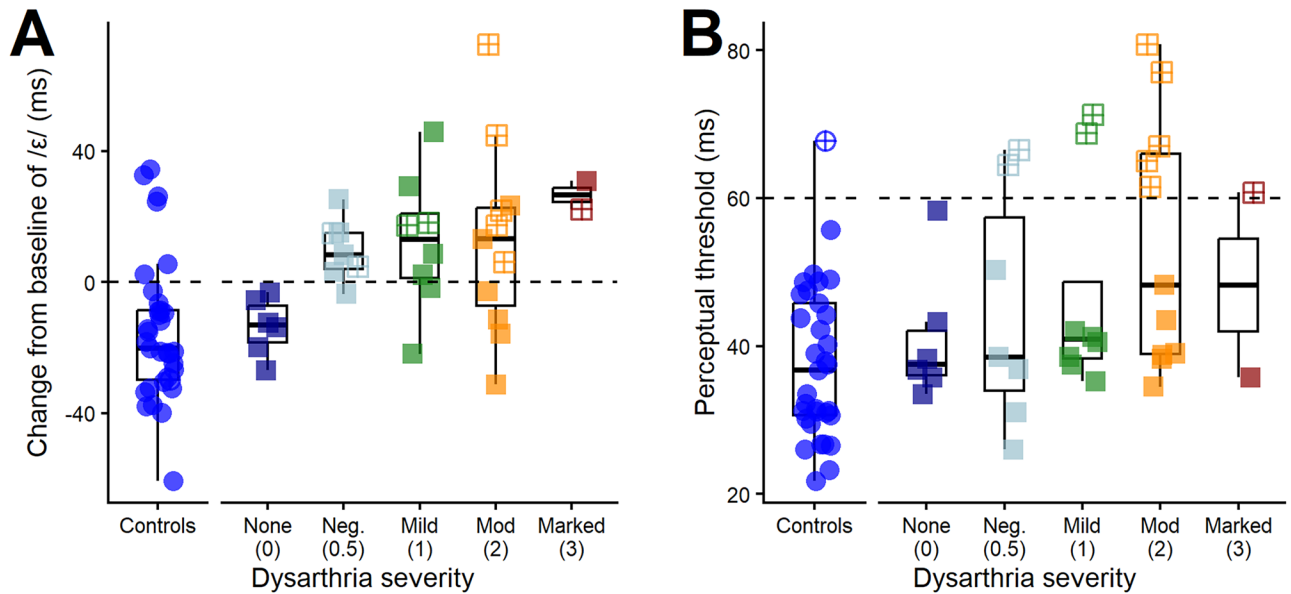


Figure 4. Relationships between temporal behavior and dysarthria severity. **A**, Magnitude of /ε/ adaptation, by dysarthria severity. **B**, Perceptual thresholds for /ε/ duration, by dysarthria severity. Both panels mark participants with thresholds above 60 ms with unfilled markers.

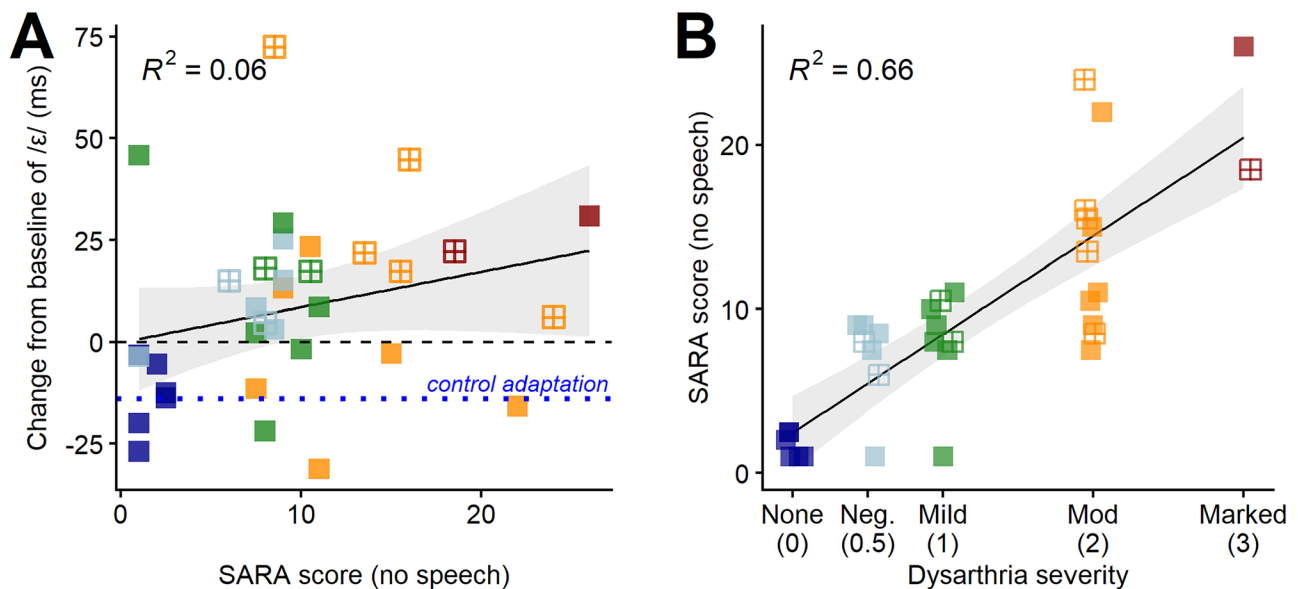


Figure 5. Relationships between speech-specific symptoms and ataxia symptom severity. Ataxia symptom severity is the SARA score with the speech component removed. Markers are colored according to dysarthria severity (dark blue, none; light blue, negligible; green, mild; gold, moderate; red, marked). Unfilled markers indicate perceptual thresholds above 60 ms. **A**, Severity of nonspeech ataxia symptoms does not predict change in /ε/ duration (adaptation of vowel duration). The mean adaptation magnitude of controls is marked with a blue dotted line. **B**, Dysarthria severity scores are highly correlated with the severity of nonspeech ataxia symptoms.

group overall showed no adaptation, this indicates that abolition of temporal adaptation can accompany even just mild to moderate ataxia symptoms in the rest of the body.

One possible explanation for this lack of relationship is that we extracted temporal measures from the speech domain, which principally relies on movements of the tongue, lips, jaw, and larynx, while SARA scores are principally influenced by movements of the arms, hands, legs, and torso. Given the somatotopic organization of the cerebellum (Nitschke et al., 1996; Grodd et al., 2001; Glickstein et al., 2011), it is possible that the lack of relationship here is simply due to different degrees of degeneration in the areas of the cerebellum that are responsible for speech

articulators versus areas that are responsible for other motor effectors. However, there is a significant correlation between SARA scores (with the speech component removed) and overall dysarthria severity ($R^2 = 0.66$, $p < 0.0001$; Fig. 5B), suggesting that somatotopy is not the driving factor behind the separation of (speech) timing behavior and general ataxia symptoms. Instead, this is likely a reflection of a shared impairment in spatial aspects of movement control in both ratings—e.g., an overall tendency toward dysmetria would result in higher scores in the manual precision tasks in SARA, as well as a percept of more severe articulatory imprecision in speech. This is a particularly likely interpretation for the population in this study, which is

principally composed of patients with degenerative disorders, rather than focal lesions. Together with the lack of strong relationship between dysarthria severity and temporal behaviors, this suggests that the timing functions investigated in this study comprise a distinct cerebellar domain from spatial control.

Discussion

Overall, these results indicate that speakers with CD cannot adapt vowel duration in response to perceived errors; as a group, they even showed a statistically significant following response. This contrasts with age-matched, neurobiologically healthy controls, who do adaptively shorten their vowels in response to lengthening perturbations. These results strongly suggest that adapting movement timing relies critically on the cerebellum, similar to the spatial domain. Strikingly, the absence of adaptation in the CD group departs from previous studies on spatial control in both speech and nonspeech: although multiple studies in those domains have demonstrated decreased sensorimotor adaptation in people with CD (Martin et al., 1996; Straube et al., 2001; Maschke et al., 2004; Smith and Shadmehr, 2005; Morton and Bastian, 2006; Tseng et al., 2007; Golla et al., 2008; Schlerf et al., 2013; Parrell et al., 2017, 2021) or transiently disrupted cerebellar function (Jenkinson and Miall, 2010; Weightman et al., 2021), complete abolishment of adaptation has not been previously documented.

One potentially appealing explanation for this difference is that the cerebellum is involved in both the perceptual and learning components of temporal adaptation, but only the learning component of spatial adaptation. In this study, speakers with CD as a group had impaired temporal perception of vowels compared with controls, and there was a correlation between perceptual thresholds and adaptation magnitude in the CD group. However, both the group difference and the within-group correlation were driven by 10 participants with thresholds above 60 ms, which was the magnitude of perturbation. All 10 showed following responses, whereas participants with CD but adequate acuity still failed to adapt. This pattern would not be predicted if perceptual deficits drove deficits in temporal adaptation: first, perceptual deficits should generate no response, rather than a following response. Second, adequate acuity should generate some adaptation, rather than no adaptation.

In this study, an alternative to perceiving duration errors could be perceiving errors in initiation time for /s/, because the elongation perturbation of the vowel /ε/ also produces delayed auditory feedback for the following /s/. This, in turn, could reflect an error in timing the initiation of /s/. If models of sensory delay are contained in and maintained by the cerebellum (Miall et al., 1993; Cao et al., 2017), speakers with CD could have inaccurate delay models, leading them to not notice the delay. Although we did not explicitly test perceptual thresholds for delay, both speakers with CD and controls showed robust lengthening in response to the delayed /s/ and immediately returned to baseline when delay was removed. The statistically identical behavior between groups suggests that controls and speakers with CD could detect the delay to the same degree. This, combined with the duration perception results, indicates that the abolition of temporal adaptation in speakers with CD is not driven by perceptual impairments. However, even if speakers with CD can detect delays, they may not be able to update movement timing accordingly.

Several other explanations might account for the difference in adaptation impairment in spatial versus temporal control. One possibility is that adapting to an elongation perturbation required

shortening vowel production, which may be intrinsically more difficult for speakers with CD given their generally slower movements. In spatial adaptation studies, there is no physical limit for corrective movements. In this study, if participants were already speaking as quickly as possible, it would be difficult to further shorten their speech sounds. However, the words in this task were produced in isolation, which invokes slower speech (with longer vowel durations) compared with speech in full phrases. Baseline durations also overlapped between groups, suggesting that speakers with CD did not have particularly slow speech movements. Given this, it is unlikely that they were speaking at their maximum rate and thus unlikely that adaptation was prevented solely due to limits on speed. A speech rate constraint also would not explain the statistically significant following response in the CD group.

Another possibility is that the presence of dysarthria induces changes in speech motor control strategies that result in longer speech durations in the presence of errors, masking any adaptation in this group. For example, speakers with dysarthria may be more attuned to errors in their speech and slow down when they perceive an error, either globally as an attempt to speak more clearly and/or to give more time for feedback-based spatial corrections. Although slowing down may enable improved spatial accuracy as movements have more time to correct for errors (see Text S3 and Fig. S3 for spatial changes), this strategy is maladaptive for the lengthening perturbation applied here; slowing down would override adaptive shortening. This account could explain the unexpected following (lengthening) behavior seen in the CD group and would fit with the preserved adaptation in individuals with CD but without dysarthria. However, the lengthening was mostly seen in individuals with perceptual duration thresholds above the applied perturbation, suggesting that individuals with more typical thresholds did not slow down but merely failed to adapt. Moreover, the increased vowel duration observed in the CD group (~3% of baseline) is substantially smaller than increases seen in clear speech (~33%; Mefferd and Dietrich, 2020). Taken together, this suggests that attempts to speak more clearly are not driving the failure to adapt in the CD group.

A major barrier to explaining the lack of temporal adaptation in speakers with CD is poor understanding of the functions involved in producing speech timing. Although models of speech production often consider timing (Browman and Goldstein, 1986; Saltzman, 1986; Bohland et al., 2010; Tourville and Guenther, 2011), none elaborate on achieving durational targets. In contrast, models of timing production are more developed for manual control (Wing and Kristofferson, 1973; Ivry and Keele, 1989; Zelaznik et al., 2005). These models provide three candidate functions to explain the lack of temporal adaptation we observed in speakers with CD.

The first two candidates affect the representation of temporal targets, which in speech may be concrete millisecond targets and/or derived from abstract targets like coordination between movements. In either case, the cerebellum is likely involved. In models of rhythmic tapping, interval representation serves as an internal “clock” to trigger movement and is impaired in people with CD (Wing and Kristofferson, 1973; Ivry, 1993; Ivry et al., 2002). A similar “clock” could be recruited to signal the end of a held movement, such as speech movements with a plateau (Smith, 1992; Burroni et al., 2025). Specifically, duration representation may contribute to a temporal error calculation when implementing a motor plan, a direct analog of reducing the error between the desired and current states in state feedback models of spatial

control (Todorov and Jordan, 2002; Scott, 2004; Parrell et al., 2019). That is, even after a spatial target is achieved, movement release is triggered only after the temporal interval has passed and temporal error is minimized. Poor duration representation may impair accurately triggering the release.

Second, timing in complex movements like speech may be represented as coordination between multiple effectors. Despite the apparently linear sequence of speech segments, speech movements are coordinated with precise temporal overlap. In this study, the key coordination is when the /s/ movements occur relative to the vowel /ε/: starting the /s/ early would produce a short vowel, while starting the /s/ late would produce a long vowel. Diedrichsen et al. (2007) demonstrated that people learned duration targets by starting the second movement after a fixed percentage of the first movement—though instructions and feedback during learning referenced milliseconds—and that the cerebellum was critical to determine when the target proportion of movement was completed. Thus, CD could impair duration production via impaired coordination.

A third factor is that movement itself takes time, which varies based on factors such as distance to target and demands on spatial precision (Hancock and Newell, 1985; Zelaznik et al., 1988; Lammert et al., 2018). Thus, in tasks that coordinate target achievement across effectors, calculating temporal error is not as simple as waiting for the target interval/proportion to elapse but rather requires temporal estimation. Recent studies suggest that speakers estimate the time needed to complete movements given the current estimated state of the effectors (Karlin, 2018, 2022; Shaw and Chen, 2019). People with CD have impaired performance in similar tasks like projectile interception (Bares et al., 2007, 2011), indicating that the cerebellum is crucial for temporal estimation. Decreased ability to estimate the correct initiation of movements could further impair duration production and adaptation.

In summary, we showed that CD abolishes temporal adaptation in speech, a uniquely natural task that coordinates multiple effectors with millisecond-level precision. Although we found that CD impairs perception of speech durations, these deficits did not drive the lack of adaptation. It is unclear if abolished temporal learning is unique to speech, perhaps due to a unique combination of demands, or if CD critically impairs all temporal adaptation. To fully resolve this question, future work is needed examining the role of cerebellar timing functions in the adaptation of simple and complex timing tasks.

References

- Ackermann H, Hertrich I (1997) Voice onset time in ataxic dysarthria. *Brain Lang* 56:321–333.
- Ackermann H, Gräber S, Hertrich I, Daum I (1997) Categorical speech perception in cerebellar disorders. *Brain Lang* 60:323–331.
- Ackermann H, Gräber S, Hertrich I, Daum I (1999) Cerebellar contributions to the perception of temporal cues within the speech and nonspeech domain. *Brain Lang* 67:228–241.
- Bares M, Lungu O, Liu T, Waechter T, Gomez CM, Ashe J (2007) Impaired predictive motor timing in patients with cerebellar disorders. *Exp Brain Res* 180:355–365.
- Bares M, Lungu OV, Liu T, Waechter T, Gomez CM, Ashe J (2011) The neural substrate of predictive motor timing in spinocerebellar ataxia. *Cerebellum* 10:233–244.
- Bates D, et al. (2014) Lme4: linear mixed-effects models using Eigen and S4. R Package Version 1:1–23.
- Boersma P, Weenink D (2021) *Praat: doing phonetics by computer* (Version 6.1.38) [Computer software]. <http://www.fon.hum.uva.nl/praat/>
- Bohland JW, Bullock D, Guenther FH (2010) Neural representations and mechanisms for the performance of simple speech sequences. *J Cogn Neurosci* 22:1504–1529.
- Browman CP, Goldstein LM (1986) Towards an articulatory phonology. *Phonology* 3:219–252.
- Burroni F, Kawahara S, Shaw JA (2025) Articulatory correlates of consonantal length contrasts: the case of Japanese mimetic geminates. *JASA Exp Lett* 5: 015201.
- Byrd D, Krivokapić J, Lee S (2006) How far, how long: on the temporal scope of prosodic boundary effects. *J Acoust Soc Am* 120:1589–1599.
- Cai S, Boucek M, Ghosh SS, Guenther FH, Perkell JS (2008) A system for online dynamic perturbation of formant trajectories and results from perturbations of the Mandarin triphthong /iau/. *Proceedings of the 8th ISSP*, 65–68.
- Cao L, Veniero D, Thut G, Gross J (2017) Role of the cerebellum in adaptation to delayed action effects. *Curr Biol* 27:2442–2451.
- Carson N, Leach L, Murphy KJ (2018) A re-examination of Montreal Cognitive Assessment (MoCA) cutoff scores. *Int J Geriatr Psychiatry* 33:379–388.
- Chen M (1970) Vowel length variation as a function of the voicing of the consonant environment. *Phonetica* 22:129–159.
- Cumming R (2011) The effect of dynamic fundamental frequency on the perception of duration. *J Phon* 39:375–387.
- Diedrichsen J, Criscimagna-Hemminger SE, Shadmehr R (2007) Dissociating timing and coordination as functions of the cerebellum. *J Neurosci* 27: 6291–6301.
- Dimitrova A, Gerwig M, Brol B, Gizewski ER, Forsting M, Beck A, Aurich V, Kolb FP, Timmann D (2008) Correlation of cerebellar volume with eyeblink conditioning in healthy subjects and in patients with cerebellar cortical degeneration. *Brain Res* 1198:73–84.
- Floegel M, Fuchs S, Kell CA (2020) Differential contributions of the two cerebral hemispheres to temporal and spectral speech feedback control. *Nat Commun* 11:1–12.
- Fowler CA, Housum J (1987) Talkers' signaling of "new" and "old" words in speech and listeners' perception and use of the distinction. *J Mem Lang* 26:489–504.
- Franken MK, Acheson DJ, McQueen JM, Hagoort P, Eisner F (2019) Consistency influences altered auditory feedback processing. *Q J Exp Psychol* 72:2371–2379.
- Garcia-Perez MA (2011) A cautionary note on the use of the adaptive up-down method. *J Acoust Soc Am* 130:2098–2107.
- Gerwig M, Kolb F, Timmann D (2007) The involvement of the human cerebellum in eyeblink conditioning. *Cerebellum* 6:38–57.
- Glickstein M, Sultan F, Voogd J (2011) Functional localization in the cerebellum. *Cortex* 47:59–80.
- Golla H, Tziridis K, Haarmeier T, Catz N, Barash S, Thier P (2008) Reduced saccadic resilience and impaired saccadic adaptation due to cerebellar disease. *Eur J Neurosci* 27:132–144.
- Grodd W, Hülsmann E, Lotze M, Wildgruber D, Erb M (2001) Sensorimotor mapping of the human cerebellum: fMRI evidence of somatotopic organization. *Hum Brain Mapp* 13:55–73.
- Grondin S, Ouellet B, Roussel M-È (2001) About optimal timing and stability of Weber fraction for duration discrimination. *Acoustic Sci Technol* 22: 370–372.
- Grube M, Cooper FE, Chinnery PF, Griffiths TD (2010) Dissociation of duration-based and beat-based auditory timing in cerebellar degeneration. *Proc Natl Acad Sci U S A* 107:11597–11601.
- Hancock PA, Newell KM (1985) The movement speed-accuracy relationship in space-time. In: *Motor behavior: programming, control, and acquisition* (Heuer H, Kleinbeck U, Schmidt KH, eds), pp 153–188. Berlin: Springer.
- Hartelius L, Runmarker B, Andersen O, Nord L (2000) Temporal speech characteristics of individuals with multiple sclerosis and ataxic dysarthria: 'scanning speech' revisited. *Folia Phoniatr Logop* 52:228–238.
- Hillenbrand JM, Clark MJ, Houde RA (2000) Some effects of duration on vowel recognition. *J Acoust Soc Am* 108:3013–3022.
- Hoche F, Guell X, Vangel MG, Sherman JC, Schmahmann JD (2018) The cerebellar cognitive affective/Schmahmann syndrome scale. *Brain* 141:248–270.
- Houde JF, Gill JS, Agnew Z, Kothare H, Hickok G, Parrell B, Ivry RB, Nagarajan SS (2019) Abnormally increased vocal responses to pitch feedback perturbations in patients with cerebellar degeneration. *J Acoust Soc Am* 145:EL372–EL378.
- House AS, Fairbanks G (1953) The influence of consonant environment upon the secondary acoustical characteristics of vowels. *J Acoust Soc Am* 25: 105–113.

- Ivry RB (1993) Cerebellar involvement in the explicit representation of temporal information. *Ann N Y Acad Sci* 682:214–230.
- Ivry RB, Gopal H (1993) Speech production and perception in patients with cerebellar lesions. In: *Attention and performance XIV (silver jubilee volume) synergies in experimental psychology, artificial intelligence, and cognitive neuroscience* (Meyer DE, Kornblum S, eds), pp 771–802. Cambridge, MA: MIT Press.
- Ivry RB, Keele SW (1989) Timing functions of the cerebellum. *J Cogn Neurosci* 1:136–152.
- Ivry RB, Spencer RM, Zelaznik HN, Diedrichsen J (2002) The cerebellum and event timing. *Ann N Y Acad Sci* 978:302–317.
- Jenkinson N, Miall RC (2010) Disruption of saccadic adaptation with repetitive transcranial magnetic stimulation of the posterior cerebellum in humans. *Cerebellum* 9:548–555.
- Kaernbach C (1991) Simple adaptive testing with the weighted up-down method. *Percept Psychophys* 49:227–229.
- Karlin R (2018) Towards an articulatory model of tone: a cross-linguistic investigation. [PhD Thesis]. Cornell University.
- Karlin R (2022) Expanding the gestural model of lexical tone: evidence from two dialects of Serbian. *Lab Phonol* 13:1–43.
- Karlin R, Parrell B (2022) Speakers monitor auditory feedback for temporal alignment and linguistically relevant duration. *J Acoust Soc Am* 152:3142–3154.
- Karlin R, Naber C, Parrell B (2021) Auditory feedback is used for adaptation and compensation in speech timing. *J Speech Lang Hear Res* 64:3361–3381.
- Kollmeier B, Gilkey RH, Sieben UK (1988) Adaptive staircase techniques in psychoacoustics: a comparison of human data and a mathematical model. *J Acoust Soc Am* 83:1852–1862.
- Kuznetsova A, Brockhoff PB, Christensen RHB (2015) Package 'lmerTest'. R Package Version 2.0.
- Lammert AC, Shadle CH, Narayanan SS, Quatieri TF (2018) Speed-accuracy tradeoffs in human speech production. *PLoS One* 13:e0202180.
- Leek MR, Hanna TE, Marshall L (1992) Estimation of psychometric functions from adaptive tracking procedures. *Percept Psychophys* 51:247–256.
- Lenth R (2019) *emmeans: estimated marginal means, aka least-squares means*. <https://CRAN.R-project.org/package=emmeans>
- Li W, Zhuang J, Guo Z, Jones JA, Xu Z, Liu H (2019) Cerebellar contribution to auditory feedback control of speech production: evidence from patients with spinocerebellar ataxia. *Hum Brain Mapp* 40:4748–4758.
- Maddieson I (2024) *Quantity and syllable structure: cross-linguistic patterns*. Vowel and consonant quantity in Germanic, Indo-European and beyond, University of Zurich.
- Malloy JR, Nistal D, Heyne M, Tardif MC, Bohland JW (2022) Delayed auditory feedback elicits specific patterns of serial order errors in a paced syllable sequence production task. *J Speech Lang Hear Res* 65:1800–1821.
- Martin TA, Keating JG, Goodkin HP, Bastian AJ, Thach WT (1996) Throwing while looking through prisms: I. Focal olivocerebellar lesions impair adaptation. *Brain* 119:1183–1198.
- Maschke M, Gomez CM, Ebner TJ, Konczak J (2004) Hereditary cerebellar ataxia progressively impairs force adaptation during goal-directed arm movements. *J Neurophysiol* 91:230–238.
- MATLAB (2023) [Computer software]. Natick, MA: The Mathworks, Inc.
- McAuliffe M, Socolof M, Mihuc S, Wagner M, Sonderegger M (2017) Montreal forced aligner. Available at: <http://montrealcorpus-tools.github.io/Montreal-Forced-Aligner/>.
- Mefferd AS, Dietrich MS (2020) Tongue- and jaw-specific articulatory changes and their acoustic consequences in talkers with dysarthria due to amyotrophic lateral sclerosis: effects of loud, clear, and slow speech. *J Speech Lang Hear Res* 63:2625–2636.
- Miall RC, Weir DJ, Wolpert DM, Stein J (1993) Is the cerebellum a smith predictor? *J Mot Behav* 25:203–216.
- Mitsuya T, MacDonald EN, Munhall KG (2014) Temporal control and compensation for perturbed voicing feedback. *J Acoust Soc Am* 135:2986–2994.
- Morton SM, Bastian AJ (2006) Cerebellar contributions to locomotor adaptations during splitbelt treadmill walking. *J Neurosci* 26:9107–9116.
- Murray IR, Arnott JL (1993) Toward the simulation of emotion in synthetic speech: a review of the literature on human vocal emotion. *J Acoust Soc Am* 93:1097–1108.
- Naber C, Karlin R, Parrell B, Niziolek CA (2024) Improving boundary detection and expanding perturbation capabilities in Audapter [Poster]. Madonna Rehabilitation Hospitals Motor Speech Conference, San Diego, CA.
- Nasreddine ZS, Phillips NA, Bédirian V, Charbonneau S, Whitehead V, Collin I, Cummings JL, Chertkow H (2005) The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. *J Am Geriatr Soc* 53:695–699.
- Nitschke MF, et al. (1996) Somatotopic motor representation in the human anterior cerebellum. *Brain* 119:1023–1029.
- Niziolek CA (2021) *wave_viewer: first release* (Version 1.1) [Computer software].
- Nowak DA, Hermsdörfer J, Rost K, Timmann D, Topka H (2004) Predictive and reactive finger force control during catching in cerebellar degeneration. *Cerebellum* 3:227–235.
- Oschkinat M, Hoole P (2020) Compensation to real-time temporal auditory feedback perturbation depends on syllable position. *J Acoust Soc Am* 148:1478–1495.
- Oschkinat M, Hoole P (2022) Reactive feedback control and adaptation to perturbed speech timing in stressed and unstressed syllables. *J Phon* 91:101133.
- Parrell B, Agnew Z, Nagarajan S, Houde J, Ivry RB (2017) Impaired feedforward control and enhanced feedback control of speech in patients with cerebellar degeneration. *J Neurosci* 37:9249–9258.
- Parrell B, Ramanarayanan V, Nagarajan S, Houde J (2019) The FACTS model of speech motor control: fusing state estimation and task-based control. *PLoS Comput Biol* 15:e1007321.
- Parrell B, Kim HE, Breska A, Saxena A, Ivry RB (2021) Differential effects of cerebellar degeneration on feedforward versus feedback control across speech and reaching movements. *J Neurosci* 41:8779–8789.
- Raharjo I, Kothare H, Nagarajan SS, Houde JF (2021) Speech compensation responses and sensorimotor adaptation to formant feedback perturbations. *J Acoust Soc Am* 149:1147–1161.
- R Core Team (2024) *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>
- Saltzman E (1986) Task dynamic coordination of the speech articulators: a preliminary model. *Experimental brain research series* 15:129–144.
- Schalling E, Hartelius L (2013) Speech in spinocerebellar ataxia. *Brain Lang* 127:317–322.
- Schertz J, Clare EJ (2020) Phonetic cue weighting in perception and production. *Wiley Interdiscip Rev Cogn Sci* 11:e1521.
- Schlerf JE, Xu J, Kleffuss NM, Griffiths TL, Ivry RB (2013) Individuals with cerebellar degeneration show similar adaptation deficits with large and small visuomotor errors. *J Neurophysiol* 109:1164–1173.
- Schmitz-Hübsch T, et al. (2006) Scale for the assessment and rating of ataxia: development of a new clinical scale. *Neurology* 66:1717–1720.
- Scott SH (2004) Optimal feedback control and the neural basis of volitional motor control. *Nat Rev Neurosci* 5:532–545.
- Seidel K, Siwanto S, Brunt ER, Den Dunnen W, Korf H-W, Rüb U (2012) Brain pathology of spinocerebellar ataxias. *Acta Neuropathol* 124:1–21.
- Shadmehr R, Krakauer JW (2008) A computational neuroanatomy for motor control. *Exp Brain Res* 185:359–381.
- Shatzman KB, McQueen JM (2006) Segment duration as a cue to word boundaries in spoken-word recognition. *Percept Psychophys* 68:1–16.
- Shaw JA, Chen W (2019) Spatially conditioned speech timing: evidence and implications. *Front Psychol* 10:2726.
- Shields LW, Balota DA (1991) Repetition and associative context effects in speech production. *Lang Speech* 34:47–55.
- Slis A, Parrell B (2025) Sensorimotor adaptation of vocal pitch is severely impaired in cerebellar ataxia. *bioRxiv*:2025–10.
- Smith CL (1992) *The timing of vowel and consonant gestures*. New Haven, CT: Yale University.
- Smith MA, Shadmehr R (2005) Intact ability to learn internal models of arm dynamics in Huntington's disease but not cerebellar degeneration. *J Neurophysiol* 93:2809–2821.
- Solé M-J, Ohala JJ (2010) What is and what is not under the control of the speaker: intrinsic vowel duration. *Papers Laborat Phonol* 10:607–655.
- Spencer KA, France AA (2016) Perceptual ratings of subgroups of ataxic dysarthria. *Int J Lang Commun Dis* 51:430–441.
- Staiger A, Schroeter ML, Ziegler W, Pino D, Regenbrecht F, Schölderle T, Rieger T, Riedl L, Müller-Sarnowski F, Diehl-Schmid J (2023) Speech motor profiles in primary progressive aphasia. *Am J Speech Lang Pathol* 32:1296–1321.

- Stetson C, Cui X, Montague PR, Eagleman DM (2006) Motor-sensory recalibration leads to an illusory reversal of action and sensation. *Neuron* 51:651–659.
- Straube A, Deubel H, Ditterich J, Eggert T (2001) Cerebellar lesions impair rapid saccade amplitude adaptation. *Neurology* 57:2105–2108.
- Sugano Y, Keetels M, Vroomen J (2010) Adaptation to motor-visual and motor-auditory temporal lags transfer across modalities. *Exp Brain Res* 201:393–399.
- Sugano Y, Keetels M, Vroomen J (2012) The build-up and transfer of sensorimotor temporal recalibration measured via a synchronization task. *Front Psychol* 3:246.
- Tang D-L (2020) Auditory-motor interactions during speech perceptions and production. [PhD Thesis]. University of Oxford.
- Teki S, Grube M, Kumar S, Griffiths TD (2011) Distinct neural substrates of duration-based and beat-based auditory timing. *J Neurosci* 31:3805–3812.
- Teki S, Grube M, Griffiths TD (2012) A unified model of time perception accounts for duration-based and beat-based timing mechanisms. *Front Integr Neurosci* 5:90.
- Timmann D, Brandauer B, Hermsdörfer J, Ilg W, Konczak J, Gerwig M, Gizewski E, Schoch B (2008) Lesion-symptom mapping of the human cerebellum. *Cerebellum* 7:602–606.
- Todorov E, Jordan MI (2002) Optimal feedback control as a theory of motor coordination. *Nat Neurosci* 5:1226–1235.
- Tourville JA, Cai S, Guenther F (2013) Exploring auditory-motor interactions in normal and disordered speech. *Proc Meet Acoustic ICA2013* 19:060180.
- Tourville JA, Guenther FH (2011) The DIVA model: a neural theory of speech acquisition and production. *Lang Cogn Process* 26:952–981.
- Tseng Y, Diedrichsen J, Krakauer JW, Shadmehr R, Bastian AJ (2007) Sensory prediction errors drive cerebellum-dependent adaptation of reaching. *J Neurophysiol* 98:54–62.
- Weightman M, Brittain J-S, Miall RC, Jenkinson N (2021) Direct and indirect effects of cathodal cerebellar TDCS on visuomotor adaptation of hand and arm movements. *Sci Rep* 11:4464.
- Wing AM, Kristofferson AB (1973) Response delays and the timing of discrete motor responses. *Percept Psychophys* 14:5–12.
- Zelaznik HN, Mone S, McCabe GP, Thaman C (1988) Role of temporal and spatial precision in determining the nature of the speed-accuracy trade-off in aimed-hand movements. *J Exp Psychol Hum Percept Perform* 14:221.
- Zelaznik HN, Spencer RM, Ivry RB, Baria A, Bloom M, Dolansky L, Justice S, Patterson K, Whetter E (2005) Timing variability in circle drawing and tapping: probing the relationship between event and emergent timing. *J Mot Behav* 37:395–403.
- Ziegler W, Staiger A, Schölderle T, Vogel M (2017) Gauging the auditory dimensions of dysarthric impairment: reliability and construct validity of the Bogenhausen Dysarthria Scales (BoDyS). *J Speech Lang Hear Res* 60:1516–1534.
- Zimmer AM, Cao D, Bastian AJ, Cowan NJ (2020) Cerebellar patients have intact feedback control that can be leveraged to improve reaching. *Elife* 9:e53246.