Do not cut off your tail: a mega-analysis of responses to auditory perturbation experiments

Article

Accepted Version

This is an Accepted Manuscript of an article published in the Journal of Speech, Language, & Hearing Research on October 18, 2023, available online: https://pubs.asha.org/doi/10.1044/2023_JSLHR-23-00315

Don’t Cut Off Your Tail: A Mega-Analysis of Responses to Auditory Perturbation Experiments

Hilary E. Miller[1], Elaine Kearney[1], Alfonso Nieto-Castañón[1], Riccardo Falsini[1], Defne Abur[1], Alexander Acosta[1], Sara-Ching Chao[2], Kimberly L. Dahl[1], Matthias Franken[3], Elizabeth S. Heller Murray[1], Fatemeh Mollaei[4], Caroline A. Niziolek[5], Benjamin Parrell[5], Tyler Perrachione[1], Dante J. Smith [1], Cara E. Stepp[1], Nicole Tomassi[6], and Frank H. Guenther[1]

[1]*Department of Speech, Language, and Hearing Sciences, Boston University, Boston, Massachusetts, USA*

[2]*College of Health Solutions, Arizona State University, Tempe, Arizona, USA*

[3]*Department of Psychology, McGill University, Montreal, Quebec, Canada*

[4]*School of Psychology & Clinical Language Sciences, University of Reading, Reading, England, United Kingdom*

[5]*Department of Communication Sciences and Disorders, University of Wisconsin-Madison, Madison, Wisconsin, USA*

[6]*Graduate Program for Neuroscience, Boston University, Boston, Massachusetts, USA*

Abstract

Purpose: The practice of removing “following” responses from speech perturbation analyses is increasingly common, despite no clear evidence as to whether these responses represent a unique response type. This study aimed to determine if the distribution of responses to auditory perturbation paradigms represents a bimodal distribution, consisting of two distinct response types, or a unimodal distribution.

Methods: This mega-analysis pooled data from 22 previous studies to examine the distribution and magnitude of responses to auditory perturbations across four tasks: adaptive pitch, adaptive formant, reflexive pitch, and reflexive formant. Data included at least 150 unique participants for each task, with studies comprising younger adult, older adult, and Parkinson’s disease populations. A Silverman’s unimodality test followed by a smoothed bootstrap resampling technique was performed for each task to evaluate the number of modes in each distribution. Wilcoxon signed-rank tests were also performed for each distribution to confirm significant compensation in response to the perturbation.

Results: Modality analyses were not significant ($p > .05$) for any group or task, indicating unimodal distributions. Our analyses also confirmed compensatory reflexive responses to pitch and formant perturbations across all groups, as well as adaptive responses to sustained formant perturbations. However, analyses of sustained pitch perturbations only revealed evidence of adaptation in studies with younger adults.

Conclusion: The demonstration of a clear unimodal distribution across all tasks suggests that following responses do not represent a distinct response pattern, but rather the tail of a unimodal distribution.

Introduction

Auditory feedback perturbation paradigms are a common method used to study the interplay between feedforward and feedback control of speech production. In this paradigm, some parameter of a participant's speech is modified in near real-time and played back to them via headphones such that the participant detects an error in their production. For example, a perturbation applied to the first formant (*F1*) of the vowel /ɛ/ can cause the speaker to hear the word "bid" or "bad" instead of the intended word "bed", depending on the direction and magnitude of the perturbation. To compensate for this error, participants adjust their speech so that what they hear in the headphones is closer to what they intended.

An important classification of perturbation paradigms concerns whether the paradigm elicits *reflexive* or *adaptive* responses. The typical reflexive paradigm involves unexpected perturbations that occur at random on a subset of trials, while normal feedback is presented on most trials. Speakers typically respond to these unexpected perturbations by adjusting their production to oppose the shift, partially correcting for the induced error within a trial (cf. Hantzsch et al., 2022, where the effect is also evident on the following trial). This has been well-documented in both reflexive pitch responses (where on average speakers raise their fundamental frequency to compensate for a downward pitch shift; e.g., Burnett et al., 1998; Kearney et al., 2022) as well as in reflexive formant responses (where on average speakers increase *F1* to compensate for a downward shift in a vowel's *F1;* e.g., Hantzsch et al., 2022).

In contrast to reflexive responses, which occur within an ongoing production, *sensorimotor adaptation* is characterized by motor responses that persist beyond the current production into future productions. The adaptive auditory perturbation paradigm involves consistent, predictable perturbations that are applied over many consecutive trials. A standard adaptive paradigm includes four phases: a *baseline* phase during which unaltered auditory feedback is presented to the participant, a *ramp* phase introducing a gradual shift in auditory feedback; a *hold* phase during which the feedback shift is applied consistently at the maximum level, and an *after-effect* phase during which feedback is unaltered. Similar to reflexive perturbations, on average, participants respond by adjusting their speech to oppose, or correct for, the shift in feedback (i.e., reduced *F1* in response to a consistent upward shift in *F1;* see Kearney et al., 2020; MacDonald et al., 2011). In addition to the within-trial correction, the adaptive paradigm induces a learning effect, where participants also pre-emptively adjust their productions on subsequent trials.

Although the predominant compensatory response to perturbations (both reflexive and adaptive) is an *opposing response*, there is clear literature supporting the presence of a small percentage of participants who adjust their speech in the same direction as the perturbation, or "follow" the perturbation (*following response*). This variation in response pattern has been documented under a variety of paradigm conditions, including in experiments that perturb fundamental frequency, vowel formants, and vocal intensity. In reflexive pitch perturbation studies, conditions shown to increase the proportion of responses that follow the direction of the perturbation include larger perturbation size (Burnett et al., 1998), more predictable perturbation direction (Behroozmand et al., 2012), and the direction of the pitch trajectory at perturbation onset (Franken et al., 2018). Although there remains little conclusive evidence to date as to the mechanisms that give rise to following responses, it has been proposed that they may represent a shift from an internal referent to an external one, or "target drift," wherein participants try to match the altered feedback instead of correcting for it (Hain et al., 2000; Larson & Robin, 2016; Terband et al., 2014), while an alternative hypothesis is that following responses simply reflect underlying fluctuations in voice or speech that are amplified when a perturbation is applied (consistent with Franken et al., 2018).

Despite our poor understanding of the mechanisms underlying this phenomenon, in some cases these following responses have been dubbed a separate phenomenon, leading to the practice of discarding these trials and/or participants from experimental analyses. In addition to following responses, non-responses—those that show no change in response to the perturbation—may also be excluded from the analysis. Treating the data in this way may be warranted if these different types of responses truly represent different populations or distinct underlying mechanisms. If, however, following responses form part of a normal distribution from a single population, the exclusion of part of the distribution will lead to over-estimation of experimental effects.

Removal of following or non-responses in reflexive studies began in the late 1990's (Burnett et al., 1998) and since then has become a widespread practice, with approximately 40% of published studies removing some data based on the type of response (59/147 reviewed studies). To the best of our knowledge, the practice of excluding responses in reflexive paradigms has been specific to studies employing fundamental frequency (f_o) perturbations, whereas this practice has not been implemented in studies using formant perturbations. Removal of following responses has similarly become a common practice for both adaptive pitch (e.g., Scheerer, Jacobson, et al., 2016) and adaptive formant studies (e.g., van den Bunt et al., 2017), despite limited evidence to support this practice. A literature review in preparation for this manuscript revealed that approximately 20% of published adaptive studies (20/97 reviewed studies, most of these adaptive formant studies) excluded some participants from their published analyses due to a following and/or non-response.

Despite the large number of studies published using auditory perturbation paradigms, prior work has been limited by the use of relatively small sample sizes as well as the noted high variability in participant responses that is often seen in perturbation studies (e.g., Behroozmand et al., 2012; Burnett et al., 1998). In the current study, we analyzed pooled data from previous perturbation studies across several research groups to investigate the nature of responses to auditory perturbations. This mega-analysis combines data from 22 unique studies and includes data from at least 150 participants for each of four perturbation tasks: reflexive f_o, reflexive formant, adaptive f_o, and adaptive formant. In addition, the data span three populations: neurotypical young adults (YA), neurotypical older adults (OA), and individuals with Parkinson's disease (PD). Our primary aim was to characterize the effect sizes and distribution of responses for each of the four perturbation tasks and across both neurotypical and disordered populations.

First, we compared compensatory responses across the three participant groups for each of the four tasks to identify any group-specific behaviors. Parkinson's disease is perhaps one of the most frequently studied populations within the auditory perturbation literature; however, it remains an open question to what extent sensorimotor integration is impacted in Parkinson's disease, with prior studies showing mixed results as to whether individuals with PD differ in performance on various perturbation paradigms compared to older controls (e.g., Abur, Subaciute, Daliri, et al., 2021; Mollaei et al., 2013). Similarly, the impact of aging on sensorimotor control of voice and speech is not fully characterized; although work to date suggests potential differences in responses to altered auditory feedback across the lifespan (Ballard et al., 2018; Liu et al., 2010; Liu et al., 2011).

Second, we tested for significant compensation or adaptation in each group and task combination across prior studies to confirm that, on average, these methods elicit the expected compensatory response, and to determine the size of the effect. Third, we tested for the presence of following responses in each task. If following responses do in fact represent a sufficiently distinct population to justify exclusion from analyses, we expect the data will reflect a bimodal distribution, with clear peaks for both opposing and following response categories. Alternatively, if following responses do not constitute a unique response pattern, we expect the data should instead consist of a unimodal distribution, with the following responses representing one tail of the distribution that may cross zero depending on the mean and standard deviation of the distribution.

Method

Included studies

We pooled data across several previous studies examining reflexive and adaptive responses to auditory perturbations of f_o, *F1,* and *F2*. Studies included at least one of four tasks: reflexive f_o, reflexive formant, adaptive f_o, and adaptive formant. Data were classified as YA, OA, or PD based on participant characteristics (see Table 1). Data from neurotypical participants (NT) were divided by age into YA (for studies where the mean age was < 25 years old, with the majority of participants aged between 18-40) or OA (for studies where the mean age of neurotypical participants was > 60 years old, with the majority of participants aged between 45-85). All NT participants were native speakers of North American English or Dutch (reflexive pitch studies only) and had no history of speech, hearing, or neurological disorders. Participants in the PD group were all native speakers of North American English and presented with normal hearing thresholds. The mean age across all three PD studies was roughly matched for age with the OA group (i.e., mean age > 60), with the majority of participants ranging from 45-75.

For each task, we included only one dataset per participant. For example, many of the studies tested multiple perturbation magnitudes within a single task; we selected data for only one of the magnitudes that was most comparable to the other studies included in our analyses. Additionally, for any studies that included multiple perturbations per trial, we included only the first perturbation per trial in our analysis. None of the included studies removed any data due to response magnitude or direction (i.e., non-responses or following responses), either at the participant- or trial-level.

Table 1 details the studies included in the current analyses. The reflexive pitch analysis included data from ten studies, with a total of 351 participants (266 YA, 42 OA, 43 PD). Stimuli were primarily a single sustained vowel, although three of the studies used a sustained word instead. The total number of trials in a given study ranged from 80-240, with an average of 57% of trials perturbed. The magnitude of the perturbation was either 25 or 100 cents, and the

TABLE 1. Reflexive and adaptive studies included in analyses.

Task	Study ID	Reference	Group	N	Stimuli	Duration	# trials	# trials perturbed (%)*	Dimension	Perturbation magnitude	Pert. Method	Measure ment window
Reflexive *f*o	01	(Abur, Subaciute, Daliri, et al., 2021)	OA	28	/ɑ/	Sustained (2-3 s)	108	24 (22.2%)	*f*o	+100 cents	Eclipse	-
			PD	28								
	02	(Heller Murray & Stepp, 2020)	YA	20	/ɑ/	Sustained (>2 s)	120	120 (100%)	*f*o	±100 cents	Eclipse	-
	03	(Franken et al., 2018)	YA	39	/e/	Sustained (3s)	198	99 (50%)	*f*o	+25 cents	Audapter	-
	04	(Franken et al., 2019)	YA	44	/e/	Sustained (4s)	240	240 (100%)	*f*o	±100 cents	Eclipse	-
	05	(Franken et al., 2021)	YA	36	/e/	Sustained (4s)	50	50 (100%)	*f*o	±100 cents	Eclipse	-
	06	(Franken et al., 2022)	YA	59	/e/	Sustained (4s)	100	100 (100%)	*f*o	±100 cents	Eclipse	-
	07	(Smith et al., 2020)	YA	18	/i/	Sustained (>2 s)	80	20 (25%)	*f*o	-100 cents	Audapter	-
	08	(Mollaei et al., 2016)	OA	14	head	Sustained (2.5 s)	200	20 (10%)	*f*o	+100 cents	VoiceOne	-
			PD	15								

	09	(Tomassi et al., 2022)	YA	30	id	Sustained (1s)	144	36 (25%)	f_o	-100 cents	Eclipse	-
	22	(Acosta et al., 2023)	YA	21	bed, beck, bet, ben, beg	Sustained (2s)	180	60 (33%)	f_o	±100 cents	Audapter	-
Reflexive formants	01	(Abur, Subaciute, Daliri, et al., 2021)	OA	28	bid, tid, hid	Sustained (2-3 s)	108	24 (22.2%)	*F1*	+30%	Audapter	-
			PD	28								
	08	(Mollaei et al., 2016)	OA	12	head	Sustained (2.5 s)	200	20 (10%)	*F1*	+30%	VoiceOne	-
			PD	13								
	09	(Tomassi et al., 2022)	YA	30	id	Sustained (1s)	144	36 (25%)	*F1*	+30%	Audapter	-
	10	(Daliri et al., 2020)	YA	30	hep, head, heck	Naturalistic (450-700 ms)	315	35 (11%)	*F1*	+34.0% (SD = 12.6)	Audapter	-
	11	(Niziolek & Guenther, 2013)†	YA	8	bed, bet, dead, deb, debt, ped, tech, ted	Naturalistic (150-475 ms)	400	100 (25%)	*F1* & *F2*	±18.3% *F1* (SD = 5.7) ±7.1% *F2* (SD = 2.3)	Audapter	-
	12	(Niziolek et al., 2014)†	YA	14	head	Naturalistic (~300 ms)	800	400 (50%)	*F1* & *F2*	±18.0% *F1* (SD = 1.9) ±5.0% *F2* (SD = 2.1)	FUSP	-
	13	(Niziolek & Parrell, 2021)†	YA	39	bed, dead, head	Naturalistic (250-500 ms)	240	80 (33.3%)	*F1*	±22.8% (SD = 1.3)	Audapter	-

	14	(Parrell et al., 2017)[†]	OA	13	beck, bet, deck, debt, pet, tech	Naturalistic (400-1000 ms)	160	60-80 (37.5-50%)	*F1*	±23.7% (SD = 2.3)	FUSP	-
	15	(Parrell et al., 2021)[†]	OA	13	dead, fed, said, shed	Naturalistic (300-500 ms)	120	60 (50%)	*F1*	±25.3% (SD = 1.8)	Audapter	-
	22	(Acosta et al., 2023)	YA	20	bed, beck, bet, ben, beg	Sustained (2s)	180	60 (33%)	*F1*	±30%	Audapter	-
Adaptive f_o	01	(Abur, Subaciute, Daliri, et al., 2021)	OA	28	/ɑ/	Sustained (2-3 s)	108	30 (27.8%)	f_o	+100 cents	Eclipse	40-120 ms (early)
			PD	28								
	02	(Heller Murray & Stepp, 2020)	YA	20	/ɑ/	Sustained (3s)	120[§]	15 (25%)	f_o	±100 cents	Eclipse	40-120 ms (early)
	16	(Abur et al., 2018)	OA	18	/ɑ/	Sustained (3s)	320[§]	40 (25%)	f_o	±100 cents	Audapter	Entire vowel (mid)
			PD	17								
	21	(Dahl et al., 2023)	YA	24	/ɑ/	Sustained (3s)	64	17 (26.6%)	f_o	-200 cents	Eclipse	40-120 ms (early)
	22	(Acosta et al., 2023)	YA	19	bed, beck, bet, ben, beg	Sustained (2s)	540[§]	110 (40.7%)	f_o	±100 cents	Audapter	40-120 ms (early)
Adaptive	01	(Abur,	OA	28	bid, tid,	Sustained	108	30 (27.8%)	*F1*	+30%	Audapter	40-120

formants		Subaciute, Daliri, et al., 2021)	PD	28	hid	(2-3 s)						ms (early)
	15	(Parrell et al., 2021)	OA	13	head	Naturalistic (300-500 ms)	120	60 (50%)	*F1*	+25.4% (SD = 1.8)	Audapter	50-100 ms (early)
	17	(Daliri et al., 2018)	YA	14	bed, Ted, head	Naturalistic (300-700 ms)	90	36 (40%)	*F1* & *F2*	+25% *F1* -12.5% *F2*	Audapter	40-60% of vowel duration (mid)
	18	(Daliri & Dittman, 2019)	YA	30	bed, Ted, head	Naturalistic (400-600 ms)	150	60 (40%)	*F1* & *F2*	+24.5% *F1* (SD = 10.1) -8.3% *F2* (SD = 2.5)	Audapter	40-60% of vowel duration (mid)
	19	(Kearney et al., 2020)	YA	15	hep, head, heck	Naturalistic (400-600 ms)	180	45 (25%)	*F1*	+30%	Audapter	10-30% of vowel duration (early)
	20	(Scott et al., 2020)	YA	37	bed, dead, head	Naturalistic (400-600 ms)	180	60 (33%)	*F1*	+30%	Audapter	10-70% of vowel duration (mid)

YA = young adults; OA = older adults; PD = patients with Parkinson's disease.
*For adaptive studies, number of perturbed trials refers to number of trials with full perturbation magnitude (i.e., hold phase). For reflexive studies, number of perturbed trials includes only trials with the selected perturbation magnitude included in our analyses. Studies may also include additional trials of a different perturbation magnitude.
†Shared data only included subjects included in Hantzsch et al. (2022) analyses.
§Participants completed both an up-shifted and down-shifted adaptation run. Number of trials reported includes both conditions, with the number of trials in each run equal to half of the reported total.

perturbation was implemented using one of three real-time feedback perturbation systems: Audapter (Cai et al., 2008), VoiceOne (TC Helicon), or Eventide Eclipse hardware (Eventide Inc, Little Ferry, NJ, USA; for a review, see Heller Murray et al., 2019).

For the reflexive formant analysis, data were included from ten studies involving a total of 248 participants (141 YA, 66 OA, 41 PD). All stimuli consisted of single words, and the majority of studies required participants to produce stimuli in a naturalistic manner (i.e., <1 s duration) instead of a sustained production. The number of trials in a given study ranged from 108-800, with feedback perturbed on an average of 30% of trials. *F1* was the most commonly perturbed dimension in the included studies, but two studies perturbed both *F1* and *F2* (Niziolek et al., 2014; Niziolek & Guenther, 2013). The average magnitude of the *F1* perturbation in each study ranged from 18-34% while the average magnitude of the *F2* perturbation in each study ranged from 0-7%. The perturbation was implemented using one of three real-time feedback perturbation systems: Audapter (Cai et al., 2008), Feedback Utility for Speech Production (FUSP; Katseff et al., 2012), or VoiceOne (TC Helicon).

For the adaptive pitch analysis, data were included from five studies and a total of 154 participants (63 YA, 46 OA, 45 PD). Most studies required participants to produce a sustained vowel. Three of the five studies included both an up-shift and a down-shift run, while the others consisted of a single shifted run with only one shift direction. For these three studies, each participant's average response across the up-shift and down-shift runs was included in the mega-analysis. All studies also included a control run with no perturbation to account for potential drift in f_o over the duration of the paradigm. The number of trials in a given adaptation task ranged from 60-270, with an average of 29% of trials with the full perturbation magnitude, occurring during the hold phase. Studies implemented either a 100 or 200 cents perturbation using one of two real-time feedback perturbation systems: Audapter (Cai et al., 2008), and Eventide Eclipse hardware (Eventide Inc, Little Ferry, NJ, USA).

For the adaptive formant analysis, data were included from six studies and a total of 165 participants (96 YA, 41 OA, 28 PD). All studies used single words as stimuli and the majority were produced in a naturalistic manner (< 1 s duration). The number of trials in a given study ranged from 90-180, with the hold length consisting on average of 36% of total trials. Studies implemented either a pure *F1* perturbation (4/6 studies) or a combined *F1*/*F2* perturbation (2/6 studies). The average magnitude of the *F1* perturbation within each study ranged from about 25-30% and all studies implemented the perturbation using Audapter (Cai et al., 2008).

Data processing

For the reflexive studies, an average response for each trial was calculated during the 100-250 ms window post-perturbation. This window was selected in light of prior work showing that vocal responses consist of two responses: first, an initial, involuntary component beginning between 100-150 ms (Burnett et al., 1998) that is thought to reflect the feedback portion of the response as a result of auditory error detection and correction mechanisms, and then a second voluntary response that starts at around 300 ms (Hain et al., 2000). Therefore, our analysis window was selected to best capture the initial compensatory response, while ending prior to the beginning of the second, voluntary response. The same analysis window was used for both reflexive pitch and reflexive formant studies.

Next, a normalized compensation percentage was calculated for each participant on a trial-by-trial basis as the change in response from the average baseline and each trial's post-perturbation response average (across the 100-250 ms window post-perturbation), expressed as a percentage of the baseline. Due to limitations in data sharing and availability, for each dataset, we replicated the normalization steps performed originally by each study's authors, as described in the corresponding manuscript. In particular, we used each study's originally defined baseline period to normalize for differences across each participant's baseline productions. For studies where perturbation onset was delayed relative to speech onset (including most reflexive pitch studies), the pre-perturbation period was used as the baseline (typically 100 or 200 ms prior to perturbation onset, using the measurement window described in the original manuscript). If perturbation onset occurred at speech onset (most reflexive formant studies), unshifted trials were used as the baseline, again using the methods reported in the original publication; most studies either normalized to an average trajectory across all unperturbed trials or to the unperturbed trial immediately preceding the perturbed trial. For those formant studies that used multiple words as stimuli, by-word normalization was employed such that each trial was normalized to the average unperturbed mean for the corresponding target word to account for potential differences in the formant trajectory. For formant studies that perturbed both *F1* and *F2*, responses were projected into a single dimension by computing the Euclidean distance in *F1*-*F2* space relative to the perturbation direction, where compensation was the scalar projection of the response onto the shift vector, as described in the original publications.

For all pitch and formant studies, responses to upward shifts were multiplied by -1 to invert them so that a positive response always indicated a response opposing the perturbation direction, while a negative response indicated a following response in the direction of the perturbation. Then, an average compensation amount was estimated for each participant across combined up- and down-shift trials (if a perturbation was applied in both directions). Last, this average compensation amount was divided by the perturbation magnitude to calculate a normalized average percent compensation for each participant.

Only participants with a minimum of 15 usable shifted trials were included in our pooled analyses in order to ensure reliable estimates of compensation (Bauer & Larson, 2003), resulting in the removal of 4 participants: one from Study 05 (reflexive f_o) and three from Study 11 (reflexive *F1*). One additional participant from Study 12 (reflexive *F1*) was removed due to excessive removal of trials (>95%), primarily due to trial durations shorter than our analysis window.

For the adaptive studies, data were again expressed as a percentage of the baseline period (i.e., the trials prior to perturbation onset) for both pitch and formant studies in order to normalize data for comparison across participants. Due to limitations in data sharing and availability, calculations used a single average measure for each trial based on the measurement window reported in the original publication. Data were categorized as either "early" measurements (measuring <150 ms post-perturbation onset, prior to the onset of feedback-based corrections within that trial) or "mid" measurements (measuring roughly the mid-point of the vowel and capturing the involvement of both feedforward and feedback contributions within that trial), as noted in Table 1. Any adaptive formant studies that perturbed both *F1* and *F2* were projected into a single dimension by computing the Euclidean distance in *F1-F2* space relative to the perturbation direction. Consistent with the methods in the original publications (Daliri et al., 2018; Kearney et al., 2020; Scott et al., 2020), data for Studies 17, 19, and 20 were averaged across blocks of three trials (such that each block contained each stimuli one time) in order to control for differences in formant trajectories across different stimuli. All adaptive pitch studies were normalized to a control (no shift) condition to control for pitch drift across trials.

Then, we measured percent adaptation for each participant as the change in response from the average baseline trials to the average across the *first three trials* of the after-effect phase, as a percentage of the baseline average (to normalize for differences in each participant's baseline productions), and as a percentage of the perturbation magnitude (to normalize across studies). This measure captures any lasting adjustments to the motor output in response to repeated perceived errors (during the hold phase) that persist once feedback has returned to normal in the after-effect phase. To constitute adaptation, these changes should be evident after the perturbation has been removed (e.g., Houde & Jordan, 2002; Jones & Munhall, 2000; Purcell & Munhall, 2006). The first after-effect trials should capture the maximum adaptation amount prior to the wash-out of the perturbation effect over the course of the after-effect phase, due to the influence of the unperturbed feedback which will now serve to induce auditory errors driving productions back to their baseline values (i.e., de-adaptation).

Statistical analysis

Separate analyses were run for each of the four tasks: reflexive f_o, reflexive formant, adaptive f_o, and adaptive formant data. First, the normalized compensatory responses for each task and age group were submitted to a Lilliefors test to confirm normality. Given the relatively high number of studies with non-normal distributions, non-parametric tests were used for subsequent analyses. Second, Kruskal-Wallis tests were conducted to test for differences in the mean compensation amount across the three groups for each of the tasks. Post-hoc Wilcoxon rank-sum tests were also conducted for all four tasks to separately test first for differences between the PD group and age-matched controls (OA), as well as to test for any effects of age by comparing the OA and YA groups.

Next, we conducted a Wilcoxon signed-rank test for each distribution to test the hypothesis that the median percent compensation for each group and task was different from 0, indicating significant compensation in response to the perturbation. Effect sizes were measured by calculating

Lastly, we tested the null hypothesis that the data were derived from two populations (i.e., separate following and opposing response types). This analysis was conducted separately for NT and PD groups and for each of the four tasks. We used a Silverman's unimodality test (Silverman, 1981) to evaluate the number of modes in each distribution from limited samples. This test uses kernel density estimates (Rosenblatt, 1956) with varying width to evaluate the critical width at which the probability density estimate from the sample distribution switches from unimodal to bimodal, followed by a smoothed bootstrap resampling technique to evaluate the significance of this critical value. A power analysis indicated the Silverman's unimodality test provides sufficient power (above 80%) at alpha = .05 to detect medium and large departures from unimodality with our NT sample sizes (i.e., mixture distributions with $1.5*D_0$ or larger effects, where D_0 is the minimum distance between the means of two gaussian distributions that results in a bimodal mixture distribution). See Supplemental Material S1 (Table S1) for further detail.

All statistical analyses were conducted in MATLAB (2020b, MathWorks) and were evaluated at alpha = .05. False discovery rate (FDR) corrections were applied to correct for multiple comparisons (Benjamini & Hochberg, 1995).

Results

Tests for group differences in compensation magnitude

Kruskal-Wallis tests revealed significant differences between the three groups only for the adaptive F1 task (reflexive f_o: $H(2) = 5.31$, $p = .070$; reflexive *F1*: $H(2) = 3.26$, $p = .196$; adaptive f_o: $H(2) = 0.35$, $p = .839$; adaptive *F1*: $H(2) = 6.84$, $p = .033$). Post-hoc rank sum tests for the adaptive F1 task revealed significant differences were present only between YA and PD groups ($p = .007$). Post-hoc rank sum tests conducted for all four tasks to separately test for the OA-PD comparison and OA-YA comparisons revealed no significant differences (see Table 2). Given no

TABLE 2. Results from rank sum tests for significant differences across groups for tests of older adult (OA)-Parkinson's disease (PD) and younger adult (YA)-OA contrasts. Medians for percent compensation for each group (with interquartile range shown in parentheses) are also listed, as well as both uncorrected *p*-values and with false discovery rate (FDR) corrections. For adaptive tasks, results from the first three trials of the after-effect phase are reported.

Task	**Median-YA (%)**	**Median-OA (%)**	**Median-PD (%)**	***p*-value (OA-PD)**	***p*-FDR (OA-PD)**	***p*-value (YA-OA)**	***p*-FDR (YA-OA)**
Reflexive f_o	6.31 (10.55)	10.29 (14.56)	12.05 (11.80)	.318	.743	.403	.743
Reflexive *F1*	1.00 (4.68)	2.41 (3.12)	1.54 (3.71)	.487	.743	.079	.391
Adaptive f_o	16.38 (77.12)	14.45 (92.01)	-1.14 (131.31)	.902	.902	.650	.743
Adaptive *F1*	21.07 (29.76)	19.55 (30.44)	6.02 (27.03)	.098	.391	.568	.743

significant age effects on any task, the OA and YA groups were combined to form a larger NT group for subsequent analyses.

Tests for significant compensation

Reflexive tasks

Wilcoxon signed rank tests revealed significant compensation (i.e., median response significantly different than 0) for all reflexive tasks in both NT and PD groups (results summarized in the first four rows of Table 3).

TABLE 3. Results from Wilcoxon signed rank tests for significant compensation for each task/group combination as well as summary statistics and effect sizes, as measured by rank-biserial correlation coefficient (*r*). Boldface values represent significant mean compensation. For adaptive tasks, results from analysis of adaptation during the first three trials of the after-effect phase are reported. FDR = false discovery rate, IQR = interquartile range, NT = neurotypical group; PD = Parkinson's disease group.

Task	**Group**	***p*-value**	***p*-FDR**	**Median**	**IQR**	***r***
Reflexive f_o	NT	**< .001**	**< .001**	6.66	11.13	0.817
Reflexive f_o	PD	**< .001**	**< .001**	12.05	11.80	0.795
Reflexive *F1*	NT	**< .001**	**< .001**	1.59	4.16	0.495
Reflexive *F1*	PD	**.0013**	**.0021**	1.54	3.71	0.575
Adaptive f_o	NT	.123	.140	15.42	83.10	0.170
Adaptive f_o	PD	.657	.657	-1.14	131.31	0.077
Adaptive *F1*	NT	**< .001**	**< .001**	20.65	29.67	0.787
Adaptive *F1*	PD	.0502	.067	6.02	27.03	0.424

Adaptive tasks

Analyses revealed significant adaptation for the *F1* task in the NT group, with findings approaching significance in the PD group (NT: $p < .001$, PD: $p = .0502$). However, neither group showed significant adaptation for the adaptive f_o task ($p > .123$). Results are summarized in the bottom four rows of Table 3.

For those studies that utilized an early measurement window (total of 91 NT participants for f_o analysis, 56 NT participants for *F1* analysis), we tested whether acoustic parameters measured during the early part of the vowel (< 150 ms) during both the after-effect phase (1st three trials) and the hold phase differed from the average baseline value. Prior research indicates a mean delay of 100-150 ms between perturbation onset and compensatory response onset (Burnett et al., 1998; Hain et al., 2000). Therefore, studies with a later measurement window (> 150 ms; coded as "mid" in Table 1) measure the feedback-based response to the perturbation in the current trial, in addition to any learned changes in the feedforward commands themselves. The early measurement studies provide a cleaner measurement of adaptation prior to the onset of any feedback-based corrective commands within that trial.

Significant compensation was seen for early measurement f_o and *F1* studies during the hold phase ($p < .001$; *F1:* median compensation of 24.93% of perturbation amount, IQR = 23.99; f_o*:* median of 23.09% of perturbation magnitude, IQR = 68.41). However, during the after-effect phase, only *F1* studies (early measurement) demonstrated adaptation (*F1:* $p < .001$, median of 19.43% of perturbation amount, IQR = 28.30; f_o*:* $p = .210$; median compensation of 14.45% of perturbation magnitude, IQR = 82.11).

We also verified that this was true for the separate YA and OA groups (early measurement studies only: total of 63 YA participants and 28 OA participants for adaptive pitch; 15 YA participants and 41 OA participants for adaptive formant). Adaptive formant analyses revealed significant adaptation for both groups in both hold and after-effect phases ($p < .001$). Adaptive pitch analyses again revealed significant compensation during the hold phase for both participant groups (YA: $p < .001$, median compensation of 28.20%, IQR = 68.16; OA: $p = .045$, median compensation of 13.18%, IQR = 80.45), but not during the first three after-effect trials (YA: $p = .112$, median = 16.38%, IQR = 77.12; OA: $p = .964$, median = 7.77%, IQR = 111.11).

However, further inspection revealed different behavior across individual pitch adaptation studies (see Supplementary Material S2 and S3 for descriptive statistics and plots for each study).

Therefore, we repeated the same analyses separately for each adaptive pitch study, which revealed significant adaptation, as measured across the first three after-effect trials, for two of the three YA studies: Study 02 (p = .0057, median = 23.91% of perturbation amount, IQR = 32.95) and Study 21 (p = .0025, median = 38.61% of perturbation amount, IQR = 52.75). For the third study, Study 22, analyses revealed the median response was significantly different from zero but in the following direction (p = .0070, median = -34.52% of perturbation amount, IQR = 53.56). These analyses suggest that our findings in the pooled analysis reflect variability across adaptive pitch paradigms, such that in some studies YA speakers show significant adaptation on average while other studies do not.

Tests for bimodal distributions

Histograms indicating distributions of compensation magnitude for the reflexive and adaptive tasks are provided in Figures 1 and 2, respectively. Silverman tests for multimodality (summarized in Table 4) revealed non-significant findings in NT and PD groups for all tasks, indicating a unimodal distribution for all analyzed tasks (p > .063, p-FDR > .372). See Supplemental Material S4 (Table S6) for supplemental analyses confirming similar findings for both OA and YA groups when analyzed separately.

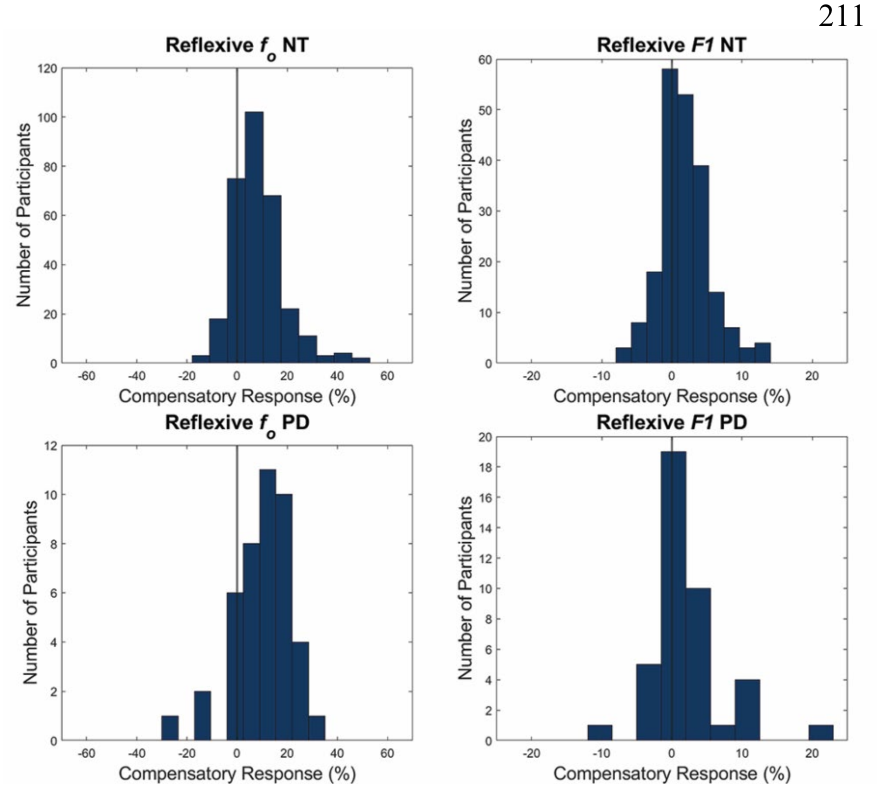

Fig. 1. Distribution of responses for reflexive tasks. Bin width is set individually for each plot so that results are distributed into ten bins. Bin width is as follows: reflexive pitch (f_o) combined older adult and younger adult neurotypical group (NT) = 7.1, reflexive f_o Parkinson's disease group (PD) = 6.5, reflexive formant (*F1*) NT group = 2.2, reflexive *F1* PD group = 3.5.

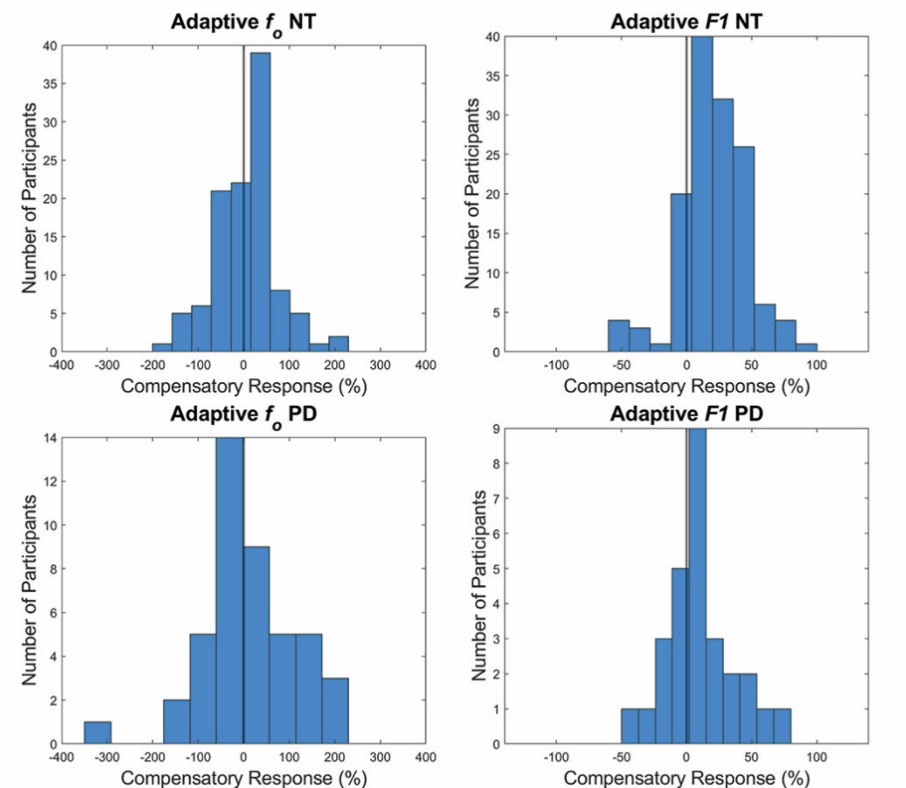

Fig. 2. Distribution of results for adaptive tasks. Bin width is set individually for each plot so that results are distributed into ten bins for each histogram. Bin width is as follows: adaptive pitch (f_o) neurotypical (NT) = 43, adaptive f_o Parkinson's disease group (PD) = 58, adaptive formant (*F1*) NT = 16, adaptive F1 PD = 13.

TABLE 4. Results of modality tests for each task and group: *p*-values below .05 would indicate evidence of a distribution with two or more modes. Also shown for each analysis is the percentage of responses that were less than 0 (i.e., response followed the direction of the perturbation). For adaptive tasks, reported results are for analysis of the average adaptive response across the first three trials of the after-effect phase. f_o = pitch perturbation; *F1* = formant perturbation; NT = neurotypical group; PD = Parkinson's disease group.

Task	**Group**	***p*-value**	***p*-FDR**	**Percentage of Responses <0**
Reflexive f_o	NT	.339	.527	12.90
Reflexive f_o	PD	.395	.527	16.28
Reflexive *F1*	NT	.482	.551	28.44
Reflexive *F1*	PD	.063	.372	21.95
Adaptive f_o	NT	.151	.372	37.27
Adaptive f_o	PD	.109	.372	50.00
Adaptive *F1*	NT	.186	.372	15.33
Adaptive *F1*	PD	.877	.877	32.14

Discussion

No evidence for a distinct population of "followers"

Our analyses showed no evidence of a bimodal distribution for any of the analyzed tasks. This finding is highly relevant to the field of speech motor control since removal of following responses has been a common practice in auditory perturbation studies, particularly in reflexive pitch studies, despite a lack of evidence establishing these data as a clearly distinct phenomenon. Our literature review found that the amount of data excluded varies widely across reflexive pitch studies; some studies excluded as little as 1% of the data due to nonresponses (e.g., Hain et al., 2000) whereas others excluded up to 44% of the data due to following responses (e.g., Liu et al., 2020). Many studies, however, fail to report the percentage of data that are removed from the analysis, particularly in the case of studies that also removed non-responses, making it challenging to assess the proportion of responses that one can typically expect to oppose, follow, or not respond to a given paradigm. Of further concern, prior studies have varied in the level of detail provided to replicate their exclusion procedure as well as in the exact method used for classifying opposing and following responses.

Another variation in data exclusion practices in reflexive studies is whether the exclusion is conducted at the individual trial level (e.g., Tang et al., 2018) or applied to the average trajectory at the participant-level (e.g., Li et al., 2016). Our analyses only examined followers on a participant-level; a unimodal distribution at the participant level is not necessarily the result of a unimodal distribution at a within-participant trial level. Factors that may influence response direction and explain the consistent presence of some percentage of following responses at the trial level have been explored by Franken and colleagues (2018). Their study demonstrated that opposing and following responses in f_o are influenced by ongoing fluctuations in the state of the vocal tract just prior to perturbation onset that can 1) mask opposing responses in "following" trials and 2) exaggerate opposing responses in "opposing" trials. This variability in trial-by-trial data is an additional source of variability beyond the participant-level distributions shown in the current study and warrants further research.

In adaptive studies, removal of following responses has primarily been conducted on a by-participant basis. Consistent with the reflexive literature, though, there is no clear consensus across adaptive studies as far as the method by which a participant is identified as a follower. Methods reported in prior studies include the use of t-tests (with variation in the exact statistical parameters), simple subtraction, or the use of a pre-specified threshold or percent change. Again, these methods are not always reported with sufficient detail to be replicable.

Regardless of the exact methods employed, our analyses do not support the claim that followers represent a distinct population or that removal or separate analysis of following responses is warranted. The removal of any responses from a unimodal distribution based on an arbitrary cut-point threatens the validity of results. For example, a group with a lower mean may also have a larger proportion of responses with a negative value, but removal of these responses would artificially inflate the group mean and potentially mask group differences. Similarly, a more variable group may also have a larger portion of the responses with a negative value; removal of these responses as following responses will therefore have a greater impact on the group mean for the variable group and could inflate group differences. These concerns may be particularly problematic for disordered populations where the speech motor system is inherently more variable.

Given the limited sample size in our PD group, the modality analyses may not be sufficiently sensitive to detect bimodality at this sample size. Therefore, while we did not find any evidence of a distinct following response in our disordered population (PD) in this study, further analysis of this question in clinical populations is warranted. It is possible, for example, that some speech disorders are characterized by highly abnormal responses to perturbations, such as following responses. Prior studies of both pitch and formant adaptation have found an increased number of following responses in clinical populations compared to controls. Children with disordered speech demonstrate an increased number of responses in the direction of the formant perturbation (Terband et al., 2014), while in hyperfunctional voice disorders, a greater percentage of patients followed the direction of the pitch perturbation compared to a control group, with a notable degree of heterogeneity within the patient group (Abur, Subaciute, Kapsner-Smith, et al., 2021). However, in both studies, the clinical population had higher response variability than controls, and this higher variability alone would be expected to result in more following responses (i.e., more responses in the left-hand tail of the response distribution) even in the absence of a group difference in mean compensation. Nonetheless, whether some speech disorders are characterized by abnormal responses to auditory perturbations, including following responses, remains a topic for elucidation by future studies.

Comparison of compensation magnitudes across tasks

Our analyses demonstrated significant compensation in the direction opposite the perturbation in both reflexive pitch and reflexive formant perturbations for both NT and PD groups. However, effect sizes were larger for reflexive pitch responses in both groups (effect sizes of ~0.8), while effect sizes for reflexive formant responses were more moderate (effect sizes of ~0.5). This difference in effect size is consistent with prior work showing that articulatory accuracy (and therefore formant values) in adult speakers is less strongly influenced by feedback than pitch (Perkell et al., 2007). Notably, response magnitudes for the reflexive tasks may be lower than reported in prior literature due to our selection of an earlier analysis window (100-250 ms) than is often reported in the literature (e.g., Daliri et al., 2020; Mollaei et al., 2016). For example, data from Daliri et al. (2020) shows that reflexive formant compensation magnitudes continue to increase after 250 ms across a number of perturbation magnitudes and directions. This earlier time window was selected to allow us to include the maximum amount of data, including from studies that asked participants to produce naturalistic words of relatively short duration. However, this time window likely begins prior to the initiation of a compensatory response in many participants, overall reducing the average response since this is averaged across some period of non-response as well.

Consistent with prior literature, our analyses of adaptive *F1* perturbation studies revealed strong evidence of adaptation. However, our analysis of f_o adaptation studies found mixed evidence of adaptation and only in the YA group. This finding is consistent with evidence that vocal motor control declines with age (e.g., Liu et al., 2011). Interestingly, the OA group showed significant compensation during the hold phase, but this did not persist into the early after-effect trials. The most salient implication of this finding is that compensatory pitch adjustments in older populations appear to be auditory-feedback-based, as any such adjustments disappear rapidly when feedback returns to normal. This finding aligns well with the concept that formants are *segmental* parameters (i.e., parameters that can be used to distinguish phonemes) that remain stable over time while pitch is a *postural* parameter, along with loudness, that can change rapidly when listening conditions change (Perkell et al., 2007).

This discrepancy in estimations of adaptation across our two measurement windows (early measurements in the hold phase compared to the first three after-effect trials) has important implications for future perturbation work. Critically, measurement of adaptation during the after-effect trials ensures that the learned transformation truly persists beyond the application of altered auditory feedback since, by definition, adaptation refers to changes that persist into future movements. However, some degree of unlearning is expected to occur during the after-effect phase, in response to the return to unaltered feedback. Our selection of a relatively short sample window (three trials) attempted to limit this unlearning in order to measure a near-maximal amount of adaptation. Though it should maximize adaptation, the small sample size introduces further limitations as it likely contributes to the large variability observed in some analyses. One alternative to these two measurement windows is the estimation of adaptation instead using auditory-noise-masked trials (a technique not included in the current mega-analysis due to a paucity of available data). Further work should determine how the use of auditory-masking trials compares to the two measurements used in this analysis in order to identify recommendations for future measurement of adaptation in perturbation paradigms.

The YA group varied in performance across adaptive pitch studies, with evidence of significant adaptation present in two of the three studies (Study 02 and Study 21), but not in Study 22. The variability in results across these studies raises interesting questions as to what factors may be necessary for pitch adaptation to occur. Study 22 differed from the other two studies in several parameters, including the number of trials in the paradigm (270 compared to 60-64), the absence of a ramp phase, the use of words as stimuli instead of a sustained vowel, and the implementation of a new time-domain pitch shift algorithm within the Audapter software (which changes only f_o and not formants, unlike most pitch perturbation studies). One comparable study which also used a long hold phase (180 trials) without a ramp phase also reported a large proportion of non-responders (14/30 participants; Scheerer, Tumber, et al., 2016), suggesting these factors may influence participant response in adaptive F0 studies. However, other prior studies have shown significant adaptation using paradigms without a ramp phase (e.g., Hawco & Jones, 2010) or with comparably long hold phases (e.g., Behroozmand & Sangtian, 2018).

Although our data preclude a full analysis of the impact of stimulus type (sustained vowel versus word) on the likelihood of a following response, our data set included sufficient data from both stimulus types to examine this question in reflexive pitch studies. Interestingly, the use of words as stimuli was more likely to elicit a following response than when sustained vowels were used as stimuli ($X^2(1, 308) = 9.86$, $p = .0017$; 25% of 65 NT speakers in studies with word stimuli exhibited following responses, compared to 10% of 243 NT speakers in sustained vowel studies).

Our data set precluded analysis of this question in formant perturbations since all included studies perturbed words. Nonetheless, our reflexive pitch results suggest that stimulus choice may have an impact on response that could explain the discrepancy in results in Study 22.

Given the relatively small sample size in Study 22 (n=19), it is also possible that this study may have simply sampled a larger number of participants from the left tail of the distribution, resulting in an average response "following" the perturbation. In fact, a closer look at this study shows there was not a consistent following response across all participants, and 3 of the 19 participants do in fact show a robust opposing response (range = 37-53%, comparable to median adaptation observed in Study 02 and Study 21; see Supplementary Material S2 and S3 for descriptive data and plots for each study). In sum, further research is needed to fully determine the conditions in which f_o adaptation may occur.

Perturbation responses in Parkinson's disease

Lastly, our finding of no significant differences between PD and OA groups may seem somewhat surprising, given that some prior studies (including studies involving the current authors) have reported group differences between PD and OA populations (e.g., Abur et al., 2018; Mollaei et al., 2016). One primary factor that may explain this discrepancy in results is medication status. Two prior speech perturbation studies that collected data while patients were receiving levodopa (L-dopa) therapy both found no differences in the magnitude of compensatory responses compared to age-matched controls (Abur, Subaciute, Daliri, et al., 2021; Kiran & Larson, 2001), consistent with our findings, while others have reported differences in auditory perturbation responses in PD while off medication (e.g., Mollaei et al., 2013, 2016). Notably, the participants in the latest of these studies, by Abur and colleagues (2021), made up the majority of the PD group for the current mega-analysis, potentially driving our null finding regarding group differences. As discussed by Abur and colleagues (2021), studying speech motor control while individuals are on L-dopa medication patients has high ecological validity, since almost all individuals with PD are prescribed L-dopa, but speech symptoms typically persist despite the medication.

Other experimental parameters, including speech severity of participants, sample size, and perturbation magnitude may also be important factors to consider. For example, Mollaei and colleagues (2016) found group differences in percent compensation in an off medication PD group only for small (15%) shifts; however, the current mega-analysis includes only the large (30%) shift data from this study to be consistent with the perturbation magnitude used for other included formant studies.

The current work was also limited by the relatively small sample size for the PD group, which precluded separate analysis of on and off medication status. However, our findings indicate the need for caution when drawing conclusions regarding possible anomalies in the performance of PD patients in reflexive and adaptive auditory perturbation paradigms. Further study of variables such as medication status is needed before firm conclusions can be drawn as to potential differences in speech motor control in PD.

Conclusion

This mega-analysis is the largest analysis of auditory perturbation responses to date. Our analyses confirmed compensatory reflexive responses to f_o and *F1* perturbations as well as adaptive responses to sustained *F1* perturbations. We also found evidence of adaptation to sustained f_o perturbations only in YA studies, suggesting potential age-related differences in vocal motor control in OA and PD groups. Another key finding from this mega-analysis is the failure to identify a bimodal distribution in any of the four tasks, suggesting that "followers" who change their productions in the same direction as the perturbation represent the left-hand tail of a unimodal response distribution with a positive mean rather than a unique class of responders. This finding calls into question the common practice of removing "followers" from published analyses of reflexive and adaptive perturbation studies. Finally, we found no significant differences in response magnitude between individuals with Parkinson's disease and older neurotypical adults across the four analyzed tasks.

Data Availability Statement

The datasets analyzed during the current study are available from the authors on reasonable request.

Acknowledgements

We are grateful to Soo-Eun Chang, Ayoub Daliri, and Terri Scott for kindly sharing their data for inclusion in this study. We also thank Hannah Indiviglio for her assistance with data extraction. This research was supported by National Institute on Deafness and Other

Communication Disorders Grants R01 DC002852 (F.H. Guenther, PI), R01 DC007683 (F.H. Guenther, PI), R01 DC015570 (C.E. Stepp, PI), R01 DC016270 (F.H. Guenther and C.E. Stepp, MPIs), R01 DC011277 (S-E Chang, PI), R01 DC017091 (B. Parrell, PI), R00 DC014520 (C.A. Niziolek, PI), R21 DC017563 (A. Daliri, PI), T32 DC013017 (C. Moore and C.E. Stepp, MPIs), T90 DA032484, (B. Shinn-Cunningham, PI), P50 DC015446 (R. Hillman, PI), F31 DC016197 (E.S. Heller Murray, PI), F31 DC019032 (D. Abur, PI), and F31 DC020352 (H. Miller, PI). Additionally, this work was supported by the Natural Sciences and Engineering Council of Canada (NSERC; awarded to S. Baum and V. Gracco), ASHFoundation New Century Doctoral Scholarships (D. Abur, K. Dahl, and H. Miller), a Stetson Scholarship from the Acoustical Society of America (H. Miller), Graduate Fellow Awards from the Rafik B. Hariri Institute for Computing and Computational Science and Engineering (D. Abur and H. Miller), and a Dudley Allen Sargent Research Fund Award (D. Abur).

References

Abur, D., Lester-Smith, R. A., Daliri, A., Lupiani, A. A., Guenther, F. H., & Stepp, C. E. (2018). Sensorimotor adaptation of voice fundamental frequency in Parkinson's disease. *PLoS ONE*, *13*(1), e0191839. https://doi.org/10.1371/journal.pone.0191839

Abur, D., Subaciute, A., Daliri, A., Lester-Smith, R. A., Lupiani, A. A., Cilento, D., Enos, N. M., Weerathunge, H. R., Tardif, M. C., & Stepp, C. E. (2021). Feedback and feedforward auditory-motor processes for voice and articulation in Parkinson's disease. *Journal of Speech, Language, and Hearing Research*. https://doi.org/10.1044/2021_JSLHR-21-00153

Abur, D., Subaciute, A., Kapsner-Smith, M., Segina, R. K., Tracy, L. F., Noordzij, J. P., & Stepp, C. E. (2021). Impaired auditory discrimination and auditory-motor integration in hyperfunctional voice disorders. *Scientific Reports*, *11*(1), 13123. https://doi.org/10.1038/s41598-021-92250-8

Acosta, A., Kearney, E., & Guenther, F. H. (2023, June). *Pitch and formant targets vary across days as seen in reflexive responses to auditory perturbations* [Poster]. Boston Speech Motor Control Symposium, Boston, MA.

Ballard, K. J., Halaki, M., Sowman, P. F., Kha, A., Daliri, A., Robin, D. A., Tourville, J. A., & Guenther, F. H. (2018). An investigation of compensation and adaptation to auditory perturbations in individuals with acquired apraxia of speech. *Frontiers in Human Neuroscience*, *12*, 510. https://doi.org/10.3389/fnhum.2018.00510

Bauer, J. J., & Larson, C. R. (2003). Audio-vocal responses to repetitive pitch-shift stimulation during a sustained vocalization: Improvements in methodology for the pitch-shifting technique. *Journal of the Acoustical Society of America*, *114*(2), 1048–1054. https://doi.org/10.1121/1.1592161

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. *Journal of the Acoustical Society of America*, *132*(4), 2468–2477. https://doi.org/10.1121/1.4746984

Behroozmand, R., & Sangtian, S. (2018). Neural bases of sensorimotor adaptation in the vocal motor system. *Experimental Brain Research*, *236*, 1881–1895. https://doi.org/10.1007/s00221-018-5272-9

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, *57*(1), 289–300.

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

Cai, S., Boucek, M., Ghosh, S. S., Guenther, F. H., & Perkell, J. S. (2008). A system for online dynamic perturbation of formant trajectories and results from perturbations of the Mandarin triphthong /iau/. *Proceedings of the 8th International Seminar on Speech Production*, 65–68.

Cureton, E. E. (1956). Rank-biserial correlation. *Psychometrika*, *21*(3), 287–290. https://doi.org/10.1007/BF02289138

Dahl, K. L., Díaz Cádiz, M., Zuk, J., Guenther, F. H., & Stepp, C. E. (2023). Controlling pitch for prosody: Sensorimotor adaptation in linguistically meaningful contexts. Manuscript submitted for publication.

Daliri, A., Chao, S.-C., & Fitzgerald, L. C. (2020). Compensatory responses to formant perturbations proportionally decrease as perturbations increase. *Journal of Speech, Language and Hearing Research*. https://doi.org/10.1044/2020_JSLHR-19-00422

Daliri, A., & Dittman, J. (2019). Successful auditory motor adaptation requires task-relevant auditory errors. *J Neurophysiol*, *122*, 552–562. https://doi.org/10.1152/jn.00662.2018

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), 1–11. https://doi.org/10.1111/desc.12521

Franken, M. K., Acheson, D. J., McQueen, J. M., Hagoort, P., & Eisner, F. (2018). Opposing and following responses in sensorimotor speech control: Why responses go both ways. *Psychonomic Bulletin & Review*, *25*, 1458–1467. https://doi.org/10.3758/s13423-018-1494-x

Franken, M. K., Hartsuiker, R. J., Johansson, P., Hall, L., & Lind, A. (2021). Speaking with an alien voice: Flexible sense of agency during vocal production. *Journal of Experimental Psychology: Human Perception and Performance*, *47*(4), 479–494. https://doi.org/10.1037/xhp0000799

Franken, M. K., Hartsuiker, R. J., Johansson, P., Hall, L., & Lind, A. (2022). Don't blame yourself: Conscious source monitoring modulates feedback control during speech production. *Quarterly Journal of Experimental Psychology*, 17470218221075632. https://doi.org/10.1177/17470218221075632

Franken, M. K., Hartsuiker, R. J., Johansson, P., Hall, L., Wartenberg, T., & Lind, A. (2019). Does passive sound attenuation affect responses to pitch-shifted auditory feedback? *Journal of the Acoustical Society of America*, *146*(6), 4108–4121. https://doi.org/10.1121/1.5134449

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

Hantzsch, L., Parrell, B., & Niziolek, C. A. (2022). A single exposure to altered auditory feedback causes observable sensorimotor adaptation in speech. *eLife*, *11*, e73694. https://doi.org/10.7554/eLife.73694

Hawco, C. S., & Jones, J. A. (2010). Multiple instances of vocal sensorimotor adaptation to frequency-altered feedback within a single experimental session. *Journal of the Acoustical Society of America Express Letters*, *127*(1), 13–18. https://doi.org/10.1121/1.3272633

Heller Murray, E. S., Lupiani, A. A., Kolin, K. R., Segina, R. K., & Stepp, C. E. (2019). Pitch shifting with the commercially available Eventide Eclipse: Intended and unintended changes to the speech signal. *Journal of Speech, Language, and Hearing Research*, *62*(7), 2270–2279. https://doi.org/10.1044/2019_JSLHR-S-18-0408

Heller Murray, E. S., & Stepp, C. E. (2020). Relationships between vocal pitch perception and production: A developmental perspective. *Scientific Reports*, *10*, 3912. https://doi.org/10.1038/s41598-020-60756-2

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

Jones, J. A., & Munhall, K. G. (2000). Perceptual calibration of F0 production: Evidence from feedback perturbation. *The Journal of the Acoustical Society of America*, *108*, 1246. https://doi.org/10.1121/1.1288414

Katseff, S., Houde, J. F., & Johnson, K. (2012). Partial compensation for altered auditory feedback: A tradeoff with somatosensory feedback? *Language and Speech*, *55*(2), 295–308. https://doi.org/10.1177/0023830911417802

Kearney, E., Nieto-Castañón, A., Falsini, R., Daliri, A., Heller Murray, E. S., 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-Castañon, 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*(January), 1–16. https://doi.org/10.3389/fpsyg.2019.02995

Kiran, S., & Larson, C. R. (2001). Effect of duration of pitch-shifted feedback on vocal responses in patients with Parkinson's Disease. *Journal of Speech, Language, and Hearing Research*, *44*, 975–987. https://doi.org/10.1044/1092-4388(2001/076

Larson, C. R., & Robin, D. A. (2016). Sensory processing: Advances in understanding structure and function of pitch-shifted auditory feedback in voice control. *AIMS Neuroscience*, *3*(1), 22–39. https://doi.org/10.3934/Neuroscience.2016.1.22

Li, W., Chen, Z., Yan, N., Jones, J. A., Guo, Z., Huang, X., Chen, S., Liu, P., & Liu, H. (2016). Temporal lobe epilepsy alters auditory-motor integration for voice control. *Scientific Reports*. https://doi.org/10.1038/srep28909

Liu, D., Dai, G., Liu, C., Guo, Z., Xu, Z., Jones, J. A., Liu, P., & Liu, H. (2020). Top–down inhibitory mechanisms underlying auditory–motor integration for voice control: Evidence by TMS. *Cerebral Cortex*, *30*(8), 4515–4527. https://doi.org/10.1093/cercor/bhaa054

Liu, H., Russo, N. M., & Larson, C. R. (2010). Age-related differences in vocal responses to pitch feedback perturbations: A preliminary study. *The Journal of the Acoustical Society of America*, *127*(2), 1042–1046. https://doi.org/10.1121/1.3273880

Liu, P., Chen, Z., Jones, J. A., Huang, D., & Liu, H. (2011). Auditory feedback control of vocal pitch during sustained vocalization: A cross-sectional study of adult aging. *PLoS ONE*, *6*(7), e22791. https://doi.org/10.1371/journal.pone.0022791

MacDonald, E. N., Purcell, D. W., & Munhall, K. G. (2011). Probing the independence of formant control using altered auditory feedback. *Journal of the Acoustical Society of America*, *129*(2), 955–965. https://doi.org/10.1121/1.3531932

Mollaei, F., Shiller, D. M., Baum, S. R., & Gracco, V. L. (2016). Sensorimotor control of vocal pitch and formant frequencies in Parkinson's disease. *Brain Research*, *1646*, 269–277. https://doi.org/10.1016/j.brainres.2016.06.013

Mollaei, F., Shiller, D. M., & Gracco, V. L. (2013). Sensorimotor adaptation of speech in Parkinson's disease. *Movement Disorders*, *28*(12), 1668–1674. https://doi.org/10.1002/mds.25588

Niziolek, C. A., & Guenther, F. H. (2013). Vowel category boundaries enhance cortical and behavioral responses to speech feedback alterations. *Journal of Neuroscience*, *33*(29), 12090–12098. https://doi.org/10.1523/JNEUROSCI.1008-13.2013

Niziolek, C. A., Nagarajan, S. S., & Houde, J. F. (2014). Sensorimotor adaptation in speech and its effects on auditory monitoring. *Neuroscience Meeting Planner*, Program No. 631.14.

Niziolek, C. A., & Parrell, B. (2021). Responses to auditory feedback manipulations in speech may be affected by previous exposure to auditory errors. *Journal of Speech, Language, and Hearing Research*, *64*(6S), 2169–2181. https://doi.org/10.1044/2020_JSLHR-20-00263

Parrell, B., Agnew, Z., Nagarajan, S. S., Houde, J. F., & Ivry, R. B. (2017). Impaired feedforward control and enhanced feedback control of speech in patients with cerebellar degeneration. *Journal of Neuroscience*, *37*(38), 9249–9258. https://doi.org/10.1523/JNEUROSCI.3363-16.2017

Parrell, B., Kim, H. E., Breska, A., Saxena, A., & Ivry, R. B. (2021). Differential effects of cerebellar degeneration on feedforward versus feedback control across speech and reaching movements. *Journal of Neuroscience*, *41*(42), 8779–8789. https://doi.org/10.1101/2021.04.05.438454

Perkell, J. S., Lane, H., Denny, M., Matthies, M. L., Tiede, M., Zandipour, M., Vick, J., & Burton, E. (2007). Time course of speech changes in response to unanticipated short-term changes in hearing state. *The Journal of the Acoustical Society of America*, *121*(4), 2296–2311. https://doi.org/10.1121/1.2642349

Purcell, D. W., & Munhall, K. G. (2006). Adaptive control of vowel formant frequency: Evidence from real-time formant manipulation. *Journal of the Acoustical Society of America*, *120*(2), 966–977. https://doi.org/10.1121/1.2217714

Rosenblatt, M. (1956). Remarks on some nonparametric estimates of a density function. *Annals of Mathematical Statistics*, *27*, 832–837.

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

Scheerer, N. E., Tumber, A. K., & Jones, J. A. (2016). Attentional demands modulate sensorimotor learning induced by persistent exposure to changes in auditory feedback. *Journal of Neurophysiology*, *115*, 826–832. https://doi.org/10.1152/jn.00799.2015.-Hearing

Scott, T. L., Haenchen, L., Daliri, A., Chartove, J., Guenther, F. H., & Perrachione, T. K. (2020). Noninvasive neurostimulation of left ventral motor cortex enhances sensorimotor adaptation in speech production. *Brain and Language*, *209*. https://doi.org/10.1016/j.bandl.2020.104840

Silverman, B. W. (1981). Using kernel density estimates to investigate multimodality. *Journal of the Royal Statistical Society. Series B (Methodological)*, *43*(1), 97–99.

Smith, D. J., Stepp, C., Guenther, F. H., & Kearney, E. (2020). Contributions of auditory and somatosensory feedback to vocal motor control. *Journal of Speech, Language, and Hearing Research*, *63*(7), 2039–2053. https://doi.org/10.1044/2020_JSLHR-19-00296

Tang, X., Chen, N., Zhang, S., Jones, J. A., Zhang, B., Li, J., Liu, P., & Liu, H. (2018). Predicting auditory feedback control of speech production from subregional shape of subcortical structures. *Human Brain Mapping*, *39*(1), 459–471. https://doi.org/10.1002/hbm.23855

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

Tomassi, N. E., Weerathunge, H. R., Cushman, M. R., Bohland, J. W., & Stepp, C. E. (2022). Assessing ecologically valid methods of auditory feedback measurement in individuals with typical speech. *Journal of Speech, Language, and Hearing Research*, *65*(1), 121–135. https://doi.org/10.1044/2021_JSLHR-21-00377 van den Bunt, M. R., Groen, M. A., Ito, T., Francisco, A. A., Gracco, V. L., Pugh, K. R., & Verhoeven, L. (2017). Increased response to altered auditory feedback in dyslexia: A weaker sensorimotor magnet implied in the phonological deficit. *Journal of Speech, Language, and Hearing Research*, *60*(3), 654–667. https://doi.org/10.1044/2016_JSLHR-L-16-0201