Motor compensations, not sensory errors, regulate serial dependence in naturalistic speech production

Yuhan Lu[1,2,†], Xiaowei Tang[3,†], Zhenyan Xiao[2,4], Anqi Xu[5], Junxi Chen[3]**, Xing Tian[2,4,6]*

[1]Shanghai Key Laboratory of Brain Functional Genomics (Ministry of Education), School of Psychology and Cognitive Science, East China Normal University, Shanghai 200062, China

[2]NYU-ECNU Institute of Brain and Cognitive Science, New York University Shanghai, Shanghai 200062, China

[3]Department of Neurosurgery, Guangzhou Sanjiu Brain Hospital, Guanzhou 510510, China

[4]Shanghai Frontiers Science Center of Artificial Intelligence and Deep Learning; Division of Arts and Sciences, New York University Shanghai, Shanghai 200122, China

[5]School of Humanities and Social Sciences, Harbin Institute of Technology, Shenzhen 518055, China

[6]Lead Contact

[†]These authors have contributed equally to this work.

Corresponding:

*Xing Tian, E-mail: xing.tian@nyu.edu
**Junxi Chen, E-mail: chenjunxi2006@163.com

Abstract

Motor control is essential for organisms to efficiently interact with external world by maintaining accuracy and precision of actions and adapting to changes in the future. For example, speaking, one of the most complicated scenarios of motor control, utilize sensory errors and resultant motor compensation to precisely control articulation. However, how these distinct sensorimotor processes regulate subsequent production remains in debate. Here, we implemented a serial dependence paradigm by randomly perturbing pitch of auditory feedback in a sequence of trials to dissociate the effects of sensory errors and motor compensation on current production. We found that error-based motor compensation in preceding trials, rather than sensory errors, attracted the compensatory responses in the current trial. Critically, this serial dependence remained across difference lexical categories but only when the consecutive productions were in the same syntactic structure. The study provides ecologically valid evidence supporting that error-based motor experience is a crucial factor mediating adaptation and control.

Introduction

The accuracy and precision of motor output, as results of efficient interplay among motor, sensory and cognitive systems, are crucial for human survival and social interactions [1-3]. In the domain of speaking, for example, the control of vocal ford and articulators relies on real-time integration among the language system, speech motor system and sensory feedback from the external environment [4-7]. Specifically, the brain anticipates the sensory consequences of articulatory movement via the mechanisms of internal forward models [8-13], continuously compares the sensory predictions to actual somatosensory and auditory feedback, and adjusts articulation by correcting any sensory errors in either domain [14, 15]. Studies on auditory perturbation show that artificially shifting auditory feedback results in a sensory error in auditory cortical areas [16], which guides the motor area to compensate, forming a sensorimotor experience. Furthermore, the sensorimotor system continues to adapt even after the perturbation is removed, suggesting that preceding sensorimotor experience facilitates an ongoing recalibration to the external environment [17, 18].

It remains heavily debated how distinct types of sensorimotor experience regulate subsequent speech production. One hypothesis suggests that error-based motor compensation is incorporated into the feedforward motor program for future speech productions, driving learning and adaptation to reduce future errors [19, 20]. Supportive evidence demonstrates a correlation between the amount of compensation during a perturbation trial and the amount of adaptation observed in the subsequent trial [21]. Alternatively, another hypothesis posits that sensory errors are integrated into the internal

model, which predicts future sensory consequences of motor actions and guides subsequent movements [5, 22-24]. Saccadic and reaching studies have shown that adaptation can still occur even when sensory errors are presented with minimal motor compensation [25] or when errors are too rapid to allow compensation [26]. A recent computational model further supports this hypothesis, showing that sensory errors can be used to update the sensory-to-motor transformation to induce speech adaptation [27]. Here, we refer these two contrasting hypotheses to as the motor-compensation hypothesis and sensory-error hypothesis.

It is challenging, however, to dissociate sensory error and its motor compensation since they are strongly coupled. For example, speech motor adaptation studies use consistent perturbation to observe opposing compensations over trials [28]. The trends of both sensory error and motor compensation remain constant, causing their effects on subsequent trial overlapped. To overcome this experimental obstacle, we introduce a novel serial dependence paradigm that dissociates the influence of preceding motor responses and perception by independently manipulating the similarity between past and current stimuli [c.f., 29, 30]. Furthermore, previous studies focus on speech motor control in isolated vowels or words [31, 32] whereas little is known about how linguistic contexts, such as lexical categories and syntactic structures modulate the dependence effects in naturalistic speech production [33]. Specifying the contributions from distinct types of sensorimotor experiences to speech production control in ecologically valid scenarios is a fundamental gateway to holistically understand the interplay among cognitive, motor and sensory systems in regulating behaviors.

Here, the current study investigates how the history of sensory-error and motor-compensation regulates speech production in various lexical and syntactic contexts. We randomly perturbed auditory feedback while participants produced speech in each trial, and therefore dissociates the influence of previous motor compensations and sensory errors. Using three experiments involving sequences of vowels and words in isolation or in sentences, we consistently demonstrated that the subsequent production was attracted by the history of motor compensation rather than sensory errors, and the attractive effects were constrained within a syntactic structure.

Results

Experiment 1: Serial dependence in repeated production of identical vowel

We first provided proof-of-concept of the serial dependent effects by investigating how the past sensory errors and the resultant motor compensation affected the subsequent production in the simplest setting when participants repeatedly produced a single vowel. Participants produced the vowel /a/ continuously for 2.5 sec in a trial and the pitch of the auditory feedback was unexpectedly downward or upward perturbed (*N* = 14; **Fig. 1A**). Through a long sequence of trials in which six levels of perturbation amount were randomly applied to each trial (**Fig. 1B**), we can separately quantify sensory error and motor compensation in each trial, as well as the relation between the preceding and current trials (i.e., serial dependence) in these two dimensions. We first measured motor responses to perturbation in the current trial (denoted as time *t*) independent from the influence of preceding trials. We averaged across trials with the same amount of perturbation, which eliminated the serial dependent effects because the perturbation amount of the immediately preceding trials was fully balanced. Motor compensations were opposed to the direction of auditory perturbation, started around 150 ms after the onset of the perturbation in the conditions of -300 cents, -180 cents, -60 cents, and 60 cents pitch shift (**Fig. 1C**, from left to right, respectively, *ts*(13) = 36.13, 56.36, 109.85, and -92.86, *ps* = 0.043, 0.021, 0.008, and 0.024, cluster-based permutation test), yet no effects were observed in the condition of 180-cent and 300-cent perturbations (*ps* > 0.203). To compare conditions, we selected 150~250 ms window that was commonly used to quantify the amount of motor compensation [21, 34]. Compensation amounts were not significantly differed across upward and downward

perturbations ($p > 0.067$, Friedman's chi-square test, FDR corrected).

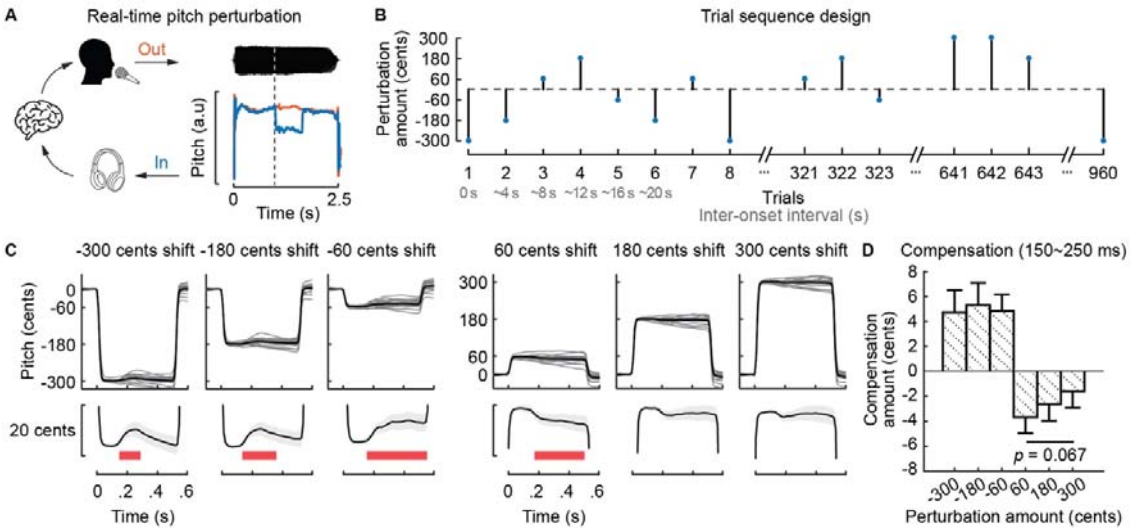

Figure 1. Motor compensation to pitch perturbation in the current trial. A, The trial procedure of pitch perturbation design. Real-time pitch perturbation during production of vowel /a/. The onset of perturbation is denoted by a dashed line. The pitch of auditory feedback sending to earphones (blue) is shifted from the original utterance recorded in the microphone (orange). **B**, The trial sequence of the experiment. Participants undergo 320 trials in each day of three consecutive days. The perturbation amounts and directions are randomly applied to each trial. **C**, Pitch traces with upward and downward perturbations and resultant compensation. The bold black lines denote the averaged pitch trace across trials for each participant (thin grey curves). Lower panels show magnified pitch trace from the above plots. The red horizontal bars denote the times of the significant effects of compensation ($p < 0.05$). **D**, Temporal averaged compensation in the time periods of interest (150-250 ms after the perturbation onset). In **C** and **D**, shaded areas and error bars denote the standard error of mean (SEM) across participants. *$p < 0.05$, **$p < 0.005$.

Next, we analyzed serial dependence effects from the immediately preceding trial, denoted as *t-1* (**Fig. 2A**). To test the motor-compensation and sensory-error hypotheses, we separately measured C_t as the function of compensation in the immediately preceding trial (i.e., C_{t-1}) (**Fig. 2B**, left) and of perturbation amount differences between the current and preceding trials (i.e., P_{t-1}-P_t) (**Fig. 2C**, left). The averaged data across trials and participants demonstrated that C_t was attracted towards both C_{t-1} and P_{t-1}-P_t (**Fig. 2BC,** right). When C_{t-1} or P_{t-1}-P_t was shifted upward, a more upward C_t was induced, and vice versa. A derivative-of-Gaussian-shaped (DoG) curve fitted both the push-pull psychometric curves (ps = 0.0001 for both, bootstrap, FDR corrected).

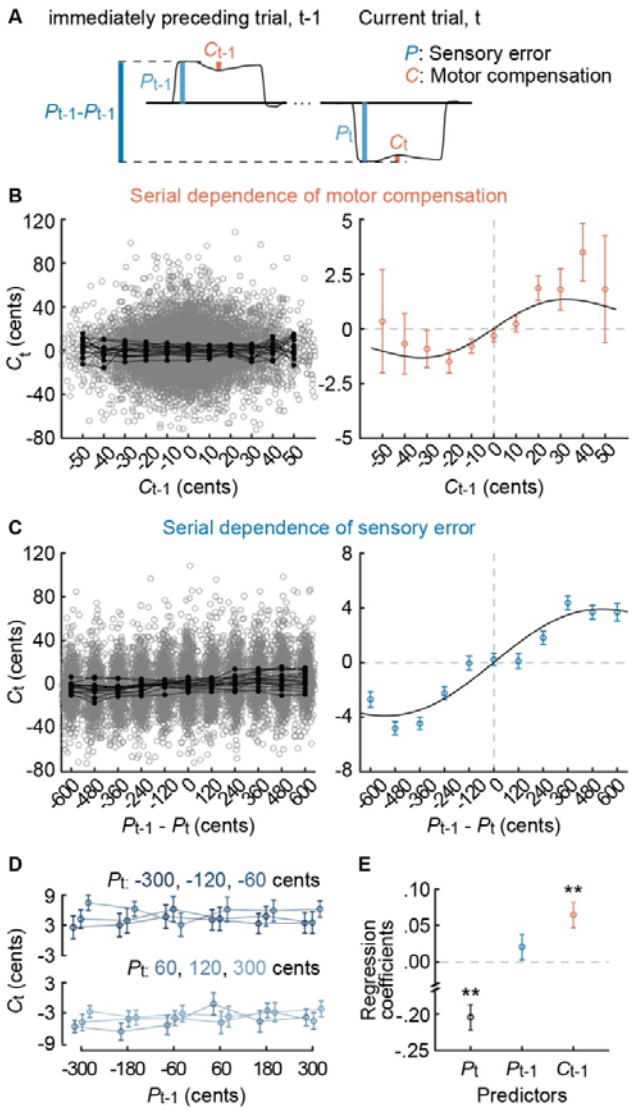

Figure 2. Attractive serial dependent effects of the immediately preceding trial on the compensatory responses of the current trial. A, Pitch trace of two adjacent trials. We measure the serial dependence of sensory error and compensation. In **B** and **C**, We plot compensation of the current trial, C_t, as a function of compensation in immediately preceding trial, C_{t-1}, and as a function of the perturbation differences between current and immediately preceding trial, P_{t-1}-P_t. In left plots, each dot represents one trial and each black curve represents one participant. Right plots show serial dependence effects after averaging across trials and participants. Color dots projecting to the same sign on the x and y axes

indicate that C_t is on the same direction of sensorimotor history of the preceding trial. **D**, C_t as a function of P_{t-1} does not show significant effects, suggesting the current compensation is not influenced by the perturbation amount of the immediately preceding trial. Error bars denote SEM across participants. **E**, The effects of current perturbation, preceding perturbation and preceding compensation on current compensation revealed by a generalized linear mixed-effect model. Regression coefficients for each predictor are extracted from model. Current perturbation and preceding compensation exert opposite influences on the current compensation. Error bars denote SEM across trials. $^{**}p < 0.005$.

Caution should be taken that the attractive effect of P_{t-1}-P_t could be driven by the reserve of the repulsive effect of P_t on C_t as shown in **Fig. 1CD.** Therefore, we further combined trials with the same P_t to assess the sequential effects of sensory errors. Main effect of P_{t-1} was not significant ($\chi^2(5) < 6.65$, $p > 0.247$; Friedman's chi-square test, FDR corrected), suggesting that the sensory error only contributed to the instant computation for compensation, but not directly to the future production. We further built a linear mixed-effect model to assess the contribution of P_t, P_{t-1}, and C_{t-1} to the C_t. The results revealed a significantly negative coefficient of P_t and a significantly positive coefficient of C_{t-1} ($t(11,680) = -21.8$ and 3.03, p = 0.0001 and 0.004, respectively, one-sample *t*-test, FDR corrected; **Fig. 2E**). No effect was found for P_{t-1} ($t(11,680) = 1.55$, $p = 0.160$). We confirmed that C_{t-1} but not P_{t-1} had an attractive serial dependent effect.

Experiment 2: Serial dependence across the production of different words

Naturalistic speech production involves variable vowel categories and thus in Experiment 2 we explored whether the attractive effect of C_{t-1} remained across different words. In this experiment, participants (*N* = 17) were asked to produce three different mono-syllabic words, i.e., 啊 (/a/), 咿 (/i/), and 呃 (/e/). We observed a similar attractive effect of C_{t-1} on C_t (p = 0.0001, two-sided bootstrap, FDR corrected; **Fig. 3A**). However, the amplitude of the fitted DoG curve was significantly lower when participants produced different words than when they produced the same vowels (p = 0.023, two-sided bootstrap, FDR corrected). An attractive effect of P_{t-1}-P_t was again significant (p = 0.0001; **Fig. 3B**), yet its effect was absent when P_t was controlled ($\chi^2(5) < 5.51$, $p > 0.163$; Friedman's chi-square test, FDR corrected; **Fig. 3C**) and in the results of a linear mixed effect model (from left to right: $t(14{,}994) = -14.36$, 1.59 and 2.11, $p = 0.0001$, 0.111, and 0.035, one-sample *t*-test, FDR corrected; **Fig. 3D**). These results suggested that the attractive effect of C_{t-1} remained but was weaker across different lexical categories.

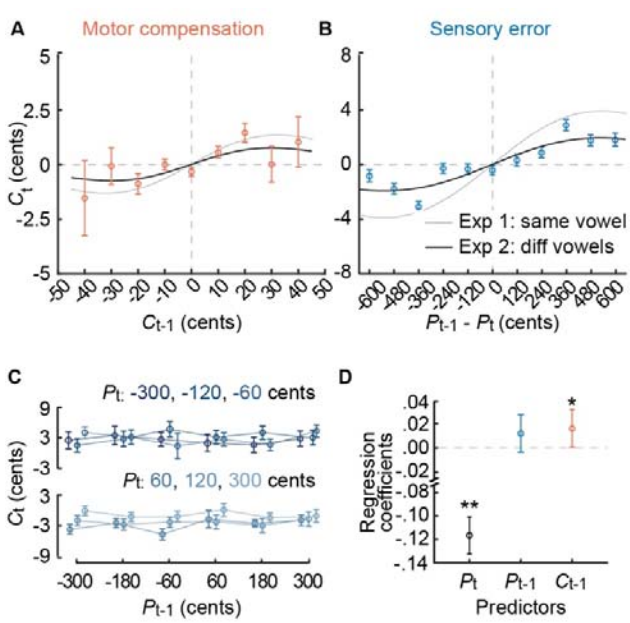

Figure 3. Attractive effects across different vowel categories. C_t as a function of C_{t-1} (**A**) and as a function of P_{t-1}-P_t (**B**) when pronouncing different vowels in a sequence. Black curves are fitting results in the Experiment 2. The gray curves are results from Experiment 1 (**Fig. 2BC**), superimposed here for comparison. The attractive serial dependent effects are weaker when producing different vowels, compared with that of same vowels. **C**, C_t as a function of P_{t-1}. No significant effects are found. **D**, Regression coefficients of P_t, P_{t-1} and C_{t-1} on C_t are extracted from a generalized linear mixed-effect model. Current perturbation and preceding compensation exert opposite influences on the current compensation, similar as that in Experiment 1 (**Fig. 2E**). Error bars denote SEM across trials. *p < 0.05, **p < 0.005.

Experiment 3: Serial dependence in the speech production within and across syntactic structures

In naturalistic speech production, not only does lexical categories change continuously, but also syntactic structures of sentences. Therefore, in Experiment 3, we investigated whether the serial dependence in speech production persisted within or across syntactic structures. Participants (N = 15) produced a five-word sentence while auditory pitch perturbation was randomly applied to the second, third, or fourth word (**Fig. 4A**). In some sentences, no pitch shift was applied. We quantified motor compensation in the preceding perturbed word and its adaptation in the current word, which could be within or across the syntactic structure. We validated that upward pitch perturbation applied on the word in a sentential context drove a significant compensatory response around 200 ms after perturbation onset, compared with the pitch trace without perturbation (60-cents shift, $t(14) = -103.52$, $p = 0.0005$; 180-cents

shift, $t(14) = -41.05$, $p = 0.005$; 300-cents shift, $t(14) = -40.30$, $p = 0.013$; **Fig. B**). Next, we extracted 200-300 ms mean pitch as compensation amount in the preceding perturbed word, i.e., C_{t-1}, and as adaptation amount in the current word, i.e., C_t. When the two words were across syntactic structures, serial dependence effect was not found ($r = 0.04$ and 0.03, $p = 0.167$ and 0.295, Spearman correlation; **Fig. CD**). Only when the two words were in the same syntactic structure, C_{t-1} attracted the pitch trace in the current word ($r = 0.09$, $p = 0.003$, Spearman correlation). These results suggested that serial dependence of motor compensation existed only when the adjacent words were within a syntactic structure.

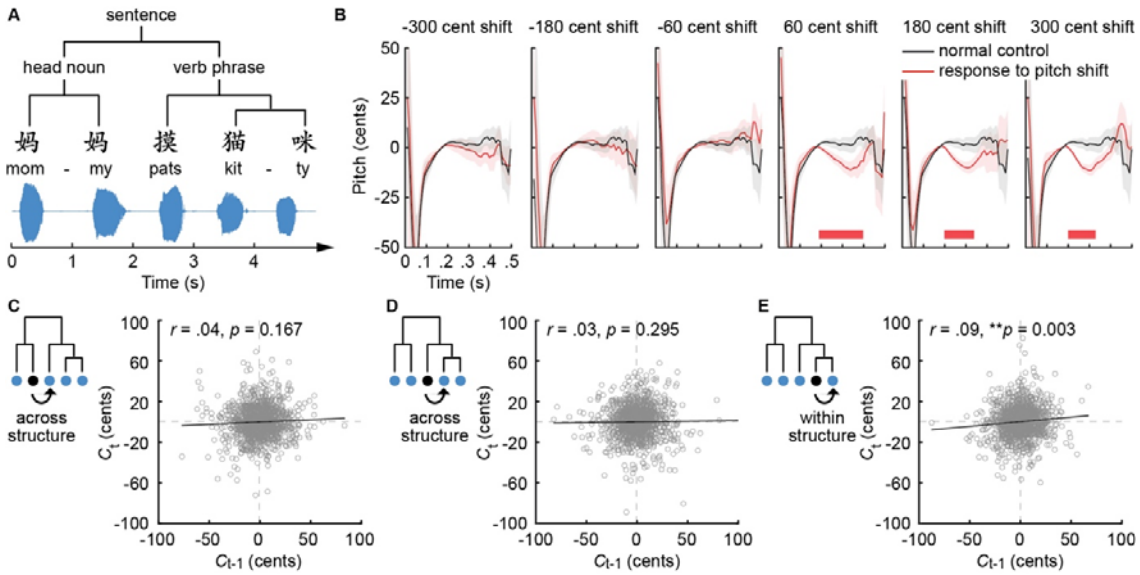

Figure 4. Attractive serial-dependence effects only within syntactic structures. A, In each trial, participants read a sentence randomly selected from three sentences that have the same syntactic structure. Each word is presented on the screen for 500 ms and disappears for 500 ms. **B**, Pitch traces of perturbed (red curves) and unperturbed word (black curves) averaged across the trials with the same conditions. Grey shaded areas denote SEM across participants. Red horizontal bars denote times of the significant effects

($p < 0.05$). **C-E**, Relations between the compensation amount of the preceding word (x-axis) and adaptation amount of the current word (y-axis). Mean pitch averaged across 200-300 ms are separately extracted from the perturbed word shown as black dot and from its subsequent word pointed by block arrow. Only when the two words are in one syntactic structure were the attractive serial dependent effects significant. **$p < 0.005$.

Discussion

We investigated how prior sensorimotor experiences affected subsequent speech production. We showed that motor compensation in one trial had an attractive effect on the following trial, whereas sensory errors did not have any serial dependence effects. Crucially, the attractive effect of motor compensation was modulated by both lexical and syntactic contexts. It persisted across different vowels but only when they are embedded within the same syntactic structure. These results suggest that distinct sensorimotor processes are constrained by linguistic system in regulating speech production.

Here, we introduce a serial dependence paradigm in the domain of speech production to distinguish the contribution of sensory-error and motor-compensation experiences in regulating speech production. This paradigm has been implemented recently in many cognitive domains, such as perception [29, 30, 35] and motor control [36, 37], that successfully dissociate the influence of past perception and motor response on the current task. In the past decades, studies on speech production have used unexpected perturbation paradigm and motor adaptation paradigm; Using this paradigm, we demonstrate that motor-compensation history in the preceding one trial attracts the current pitch production. Our findings of dissociations between sensory-errors and motor compensation offer new insights into speech motor control and learning. Previous studies have demonstrated that motor responses to auditory perturbations are not always compensatory in the opposite direction of the perturbation; sometimes, the reaction follows the perturbation [38]. This type of following response tends to occur when the perturbation is large [39] or more predictable in both

direction and onset [40, 41]. The current study shows that the production responses are determined by the motor responses in the preceding trials. That is, if the prior motor responses resulted in a raised pitch, the subsequent responses continued in the same direction, even in the presence of an upward perturbation, leading to a following response. Experiences of motor compensation may be effective in the late process after error correction. Whereas, experience of sensory-errors may influence speech motor actions more indirectly by modulating perceptual sensitivity to auditory feedback processing [17, 42]. Therefore, sensory-error and motor-compensation play distinct roles in speech production control - sensory errors indirectly influence production through perceptual mechanisms, whereas motor compensation has a direct impact on subsequent motor responses. Together with the long-lasting effects in the motor domain, presumably via longer time constants compared with those in the perceptual system, emphasizes the importance of motor configuration and experiences in the adaptation and learning process.

The extension of serial dependent effects into naturalistic speech production reveals the cognitive constraints in regulating action. Few studies have explored the effects of perturbation in continuous speech [43, 44]. Analyzing pitch contours in continuous speech is challenging due to potential interference from consonantal perturbations [45] or sentence intonation [46]. Our study carefully selects different combinations of consonants and vowels and shows that the attractive effects of motor compensation persist only within a syntactic structure. Although sensorimotor experiences from the previous syllable can affect the production of next one, the serial dependent effects are diminished when moving to a new

syntactic structure. The linguistic constraints on speech production control may root in the encoding process in which the unit of sensory-to-motor transformation is chunked according to the syntactic structures [47, 48]. Hence, the controlling process via the internal forward models [11-13] and serial dependence effects are in the size of the encoding unit within the same syntactic structure.

In summary, implementing a novel serial-dependence paradigm in the domain of speaking, we found distinct functions of sensory-errors and motor-compensation in speech production control. The serial dependent effects were constrained by lexical and syntactic contexts of speech production. These results suggest that error-based motor experience is a crucial factor mediating adaptation and control. Systematic interplay among cognitive, motor and sensory systems regulates behaviors.

Methods

Participants

In Experiments 1-3, we recruited 14 participants (9 females, 18-27 years old, mean 21 years old), 17 participants (11 females, 19-32 years old, mean 23 years old), and 15 participants (9 females, 19-25 years old, mean 22 years old), respectively. They were college students with self-reported normal hearing and speech. All participants were right-handed and were native Chinese speakers. Each participant would not attend more than one experiment. Written informed consent was obtained from each participant before the experiment. The experimental procedures were approved by the Research Ethics Committee at the East China Normal University and New York University Shanghai.

Procedure

Experiment 1

Participants were asked to produce the vowel /a/ in a total of 960 trials over three days. Each trial began with a white fixation cross, displayed at the center of the screen for a randomized duration of 0.8–1.2 s (uniform distribution), signaling participants to be prepared for vocalization. When the cross turned green, participants were instructed to vocalize no more than 500 ms and sustain the /a/ sound with a steady and natural tone for 2.5 s until the cross disappeared. A pitch perturbation for a duration of 500 ms was randomly introduced 0.8–1.2 s after vocal onset. The direction (i.e., upward or downward) and magnitude (i.e., 60, 180, or 300 cents) of the pitch shift were randomly assigned across the 960 trials for each participant. At the end of each trial, participants received visual feedback: "Good," "Too

slow," or "Too short." Participants completed 95.86% ± 0.02% (s.d.) of trials that met criteria. Each trial was followed by a silent interval with a random duration of 1–2 s (uniform distribution), before the next trial began.

Participants completed 320 trials in 10 blocks each day, with self-paced rest periods of at least 1 min between blocks. At the beginning of each session, participants underwent a training block of 10 trials. The mean pitch from these trials was used to define the pitch range in the Audapter toolbox (see *Auditory feedback perturbation* below).

Experiment 2

The procedure for Experiment 2 was identical to Experiment 1, except participants were required to produce three different words—啊 (/i/), 咿 (/a/), and 呃 (/e/)—randomly assigned in each trial. A white letter corresponding to the word was presented at the center of the screen, and upon turning green, participants were instructed to vocalize a word. Participants completed 94.82% ± 0.11% of trials that met the criteria.

Experiment 3

The procedure for Experiment 3 followed that of Experiment 1, with two differences. First, participants completed 300 trials in a single day, and they were required to produce sentences rather than isolated vowels. In each trial, one of three five-word sentences (e.g., ▪▪▪猫咪 [ma1 ma1 lu1 mao1 mi1], ㆎㆎ邀巫医 [ya1 ya1 yao1 wu1 yi1], 依依摸▪▪ [yi1 yi1 mo1 wu1 ya1]) was randomly selected (uniform distribution), with each mono-syllabic

word appearing sequentially at the center of the screen for 500 ms, followed by a 500 ms blank screen. The whole duration for a sentence production was 5 s. During vocalization, a whole-utterance pitch perturbation (upward or downward perturbation at 60, 180, 300 cents, randomly assigned) was applied to the second-, third-, or fourth-word during vocalization (uniformed distribution). In 20% of the sentences no perturbation was applied. Participants were instructed to vocalize each word within 500 ms after it appeared and to avoid coarticulation between words. The speech materials and task were carefully chosen to avoid effect of consonantal obstruction [49] and effect of articulators' dynamics on pitch contour, respectively. Participants completed 94.44% ± 0.07% of trials that met the criteria.

Auditory feedback perturbation

Participants' speech was recorded by a Shure beta 58A microphone that was placed 5 centimeters away from their mouth. The acoustic signals from the microphone were pre-amplified and then provided to an audio interface (MOTU Microbook IIc), collected at 48k Hz sampling rate. The recorded speech was real-time manipulated (i.e., perturbation) in the audio interface, and the processed signal was split and sent both to the computer and to the headphones. The speech processed in the audio interface was performed using Audapter toolbox [50]. The pitch of speech in each trial was shifted using the time-domain method. A 10-ms linear ramp was added at boundaries to attenuated possible abrupt jumps in the speech. A priori pitch range was set as +/-50 Hz relative to the mean pitch of a vocal /a/ for each participant. The rest of the parameters were set as default including algorithm, frame length, and cepstral liftering.

Acoustic analysis

In the experiments 1 and 2, we first removed silence from the recordings of produced vowels in each trial using *vad* function. Visual inspection was performed to make sure noise did not affect the voice activity detection. Pitch tracks (at 100-Hz sampling rate) were measured using Praat-based script ProsodyPro [51]. Default parameters were used. Perturbation onset in the pitch trace was obtained by a customized automatic algorithm combined with manually inspection. To convert hertz to cent scale, we defined baseline pitch by taking the mean of pitch trace in the -400 to -200 ms before perturbation onset. Pitch trace was transformed according to the formula: $cents = 1200 * \log_2(\frac{pitch}{pitch_{baseline}})$. A trial rejection was performed if the pitch trace was varied beyond its perturbation amount minus 100 cents (for downward-perturbation trials) or plus 100 cents (for upward-perturbation trials). To extract perturbation and compensation amounts for DoG-curve fitting and linear mixed effect model, the pitch trace in the 50-80 ms and the 150-250 ms windows were averaged, respectively, for each trial [21]. Compensation amounts for all trials and all participants were corrected by subtracting the grand average. In the experiment 3, we applied the same pipeline except that only the perturbed syllable and its subsequent syllable were extracted to measure their pitch trace. Since pitch boundary was not stable, which was visually inspected between 0-100 ms, baseline pitch was defined as mean pitch in the 150-200 ms time window. Compensation amount for perturbed word and adaptation amount for subsequent word were both defined as mean pitch in the 200-250 ms time window.

Statistical analysis and significance tests

Compensatory responses

For participant-level analysis, trials with the same perturbation condition were averaged, regardless of vowel categories in the experiment 2 and 3. To assess whether there was a compensatory response to perturbation, a two-sided cluster-based permutation test [52] was used to test averaged pitch traces against perturbation amount or between perturbed and normal conditions (**Fig. 1C** and **4B**). A one-sided one-sample t test was used to test motor compensation extracted from time of interests against 0 (**Fig. 1D)**. Furthermore, to the main effect of P_{t-1} on C_t, a Friedman's chi-square test was used for multiple condition comparison (**Fig. 2D** and **3C**)**.** Spearman correlation was used to find relation between C_{t-1} and C_t within and across syntactic structures (**Fig. 4C-E**). To correct for multiple comparisons, FDR correction was applied.

Serial dependence effect

The serial dependence paradigm quantified the similarity between auditory perturbations in two consecutive trials, i.e., P_{t-1}-P_t, and motor responses in the preceding one trial, i.e., C_{t-1}, to access their serial effect on the current motor compensation, i.e., C_t. To explore serial-dependence effect of sensory error on C_t, C_t was expressed as a function of the difference between P_{t-1} and P_t. For positive values of this difference, preceding perturbation P_{t-1} was higher than the perturbation of the current trial P_t; similarly, positive compensation amount denoted a trial in which participants compensated upward in response to perturbation. Consequently, data points in the scatter plot (**Fig. 2BC**) that had x- and y- value in the same

sign indicated that C_t was in the direction of P_{t-1}. In addition, to explore serial-dependence effect of motor compensation on C_t, C_t was expressed as a function of the compensation amount in the preceding trial C_{t-1}. Following the same principle, the positive value of previous and current compensation, which had x- and y-value in the same sign, denoted a trial in which C_t was in the direction of C_{t-1}.

We assess the systematic serial dependence by polling the compensation amount of all trials and fitting the DoG-shape curve, a well-found psychophysical function in previous serial-dependence studies [29, 35]. The DoG was given as $y = x * a * w * c * e^{-(wx)^2}$, where *a* was the curve amplitude, *w* was the curve width, and *c* was the constant $\sqrt{2}/e^{-0.5}$. To assess sensory-error serial dependence effects, *x* was P_{t-1}-P_t, and width parameter *w* of the curve was constrained to a range of plausible value (*w* = 0.001-0.005, corresponding to the curve peak between 130 and 530 cents perturbation difference). To assess motor-compensation serial dependence effects, *x* was C_{t-1}, and width parameter *w* of the curve was constrained to a range of plausible value (*w* = 0.019-0.071, corresponding to the curve peak between 10 and 36 perturbation pitch difference). The model fitting used the function *fmincon* with default parameters. We employed a permutation test to evaluate the significance of the DoG models following the methodology described in Fritsche, et al. [53].

Generalized linear mixed-effect model

To quantitatively test relations of C_t to P_t, P_{t-1} and C_{t-1}, we built a generalized linear mixed-effect models (**Fig. 2E** and **3D**). Model assumed that P_t, P_{t-1} and C_{t-1} together influenced C_t.

The function 'fitglme' was performed and accounted for group-level fixed effects and random effects because of vocal pitch variation across participants. The model was specified as $C_{t\ ij} = f(\beta_{0j} + \beta_{1j}P_{t\ ij} + \beta_{2j}P_{t-1\ ij} + \beta_{3j}C_{t-1\ ij} + 1|subjects)$, where i indicated participant index and j indicated j^{th} trial. Regression coefficient of each predictor, i.e., β, was accessed by the *t*-statistics against the null hypothesis test that the coefficient was equal to 0. FDR correction was applied.

Data and code availability: All data and code that support the findings of this study will be available after acceptance.

Acknowledgments: We thank Bao Hong for thoughtful comments on previous versions of the manuscript, Tingting Wang and Huixuan Liu for assisting in collecting data. This study is supported by National Natural Science Foundation of China 32071099 and 32271101, Program of Introducing Talents of Discipline to Universities, Base B16018, and NYU Shanghai Boost Fund to X.T.

Author contributions: Y.L., X.T., and J.C. designed the experiment. Y.L., W.T., and Z.X. collected the data, Y.L., Z.X., A.X., and W.T., performed the analyses, Y.L., and X.T. drafted, reviewed, and corrected the manuscript. X.T. and J.C. supervised the project.

Competing interests: The authors declare no competing financial interests.

Requests for further information and resources should be directed to and will be fulfilled by the lead contact, Xing Tian (xing.tian@nyu.edu).

References

1. Rosenbaum, D.A. (2009). Human motor control, (Academic press).
2. Ivanenko, Y.P., Poppele, R.E., and Lacquaniti, F. (2006). Motor control programs and walking. The Neuroscientist *12*, 339-348.
3. Olivier, E., Davare, M., Andres, M., and Fadiga, L. (2007). Precision grasping in humans: from motor control to cognition. Current opinion in neurobiology *17*, 644-648.
4. Hickok, G. (2012). Computational neuroanatomy of speech production. Nat Rev Neurosci *13*, 135-145.
5. Houde, J.F., and Nagarajan, S.S. (2011). Speech production as state feedback control. Front Hum Neurosci *5*, 82.
6. Guenther, F.H. (2016). Neural control of speech, (Mit Press).
7. Tian, X., and Poeppel, D. (2010). Mental imagery of speech and movement implicates the dynamics of internal forward models. Frontiers in psychology *1*, 166.
8. Kawato, M. (1999). Internal models for motor control and trajectory planning. Current opinion in neurobiology *9*, 718-727.
9. Wolpert, D.M., and Ghahramani, Z. (2000). Computational principles of movement neuroscience. Nature neuroscience *3*, 1212-1217.
10. Schubotz, R.I. (2007). Prediction of external events with our motor system: towards a new framework. Trends in cognitive sciences *11*, 211-218.
11. Tian, X., Ding, N., Teng, X., Bai, F., and Poeppel, D. (2018). Imagined speech influences perceived loudness of sound. Nature Human Behaviour *2*, 225-234.
12. Li, S., Zhu, H., and Tian, X. (2020). Corollary Discharge Versus Efference Copy: Distinct Neural Signals in Speech Preparation Differentially Modulate Auditory Responses. Cereb Cortex *30*, 5806-5820.
13. Li, Y., Luo, H., and Tian, X. (2020). Mental operations in rhythm: Motor-to-sensory transformation mediates imagined singing. PLoS biology *18*, e3000504.
14. Lametti, D.R., Nasir, S.M., and Ostry, D.J. (2012). Sensory preference in speech production revealed by simultaneous alteration of auditory and somatosensory feedback. J Neurosci *32*, 9351-9358.
15. Larson, C.R., Altman, K.W., Liu, H., and Hain, T.C. (2008). Interactions between auditory and somatosensory feedback for voice F0 control. Exp Brain Res *187*, 613-621.
16. Chang, E.F., Niziolek, C.A., Knight, R.T., Nagarajan, S.S., and Houde, J.F. (2013). Human cortical sensorimotor network underlying feedback control of vocal pitch. Proc Natl Acad Sci U S A *110*, 2653-2658.
17. Tang, D.L., Parrell, B., and Niziolek, C.A. (2022). Movement variability can be modulated in speech production. J Neurophysiol *128*, 1469-1482.
18. Houde, J.F., and Jordan, M.I. (1998). Sensorimotor adaptation in speech production. Science *279*, 1213-1216.
19. Tourville, J.A., and Guenther, F.H. (2011). The DIVA model: A neural theory of speech acquisition and production. Lang Cogn Process *26*, 952-981.
20. Kawato, M., Furukawa, K., and Suzuki, R. (1987). A hierarchical neural-network model for control and learning of voluntary movement. Biological cybernetics *57*, 169-185.
21. Hantzsch, L., Parrell, B., and Niziolek, C.A. (2022). A single exposure to altered auditory feedback causes observable sensorimotor adaptation in speech. Elife *11*.
22. Shadmehr, R., Smith, M.A., and Krakauer, J.W. (2010). Error correction, sensory prediction, and adaptation in motor control. Annual review of neuroscience *33*, 89-108.

23. Wolpert, D.M., and Kawato, M. (1998). Multiple paired forward and inverse models for motor control. Neural networks *11*, 1317-1329.
24. Daliri, A., and Dittman, J. (2019). Successful auditory motor adaptation requires task-relevant auditory errors. J Neurophysiol *122*, 552-562.
25. Noto, C.T., and Robinson, F.R. (2001). Visual error is the stimulus for saccade gain adaptation. Cognitive brain research *12*, 301-305.
26. Tseng, Y.-w., Diedrichsen, J., Krakauer, J.W., Shadmehr, R., and Bastian, A.J. (2007). Sensory prediction errors drive cerebellum-dependent adaptation of reaching. Journal of neurophysiology *98*, 54-62.
27. Kim, K.S., Gaines, J.L., Parrell, B., Ramanarayanan, V., Nagarajan, S.S., and Houde, J.F. (2023). Mechanisms of sensorimotor adaptation in a hierarchical state feedback control model of speech. PLoS Computational Biology *19*, e1011244.
28. Raharjo, I., Kothare, H., Nagarajan, S.S., and Houde, J.F. (2021). Speech compensation responses and sensorimotor adaptation to formant feedback perturbations. J Acoust Soc Am *149*, 1147.
29. Fischer, J., and Whitney, D. (2014). Serial dependence in visual perception. Nat Neurosci *17*, 738-743.
30. Cicchini, G.M., Mikellidou, K., and Burr, D.C. (2023). Serial Dependence in Perception. Annu Rev Psychol.
31. Niziolek, C.A., and Guenther, F.H. (2013). Vowel category boundaries enhance cortical and behavioral responses to speech feedback alterations. J Neurosci *33*, 12090-12098.
32. Smith, D.J., Stepp, C., Guenther, F.H., and Kearney, E. (2020). Contributions of Auditory and Somatosensory Feedback to Vocal Motor Control. J Speech Lang Hear Res *63*, 2039-2053.
33. Zeng, Y., Niziolek, C.A., and Parrell, B. (2024). Simultaneous acquisition of multiple auditory-motor transformations reveals supra-syllabic motor planning in speech production. psyarxiv.
34. Burnett, T.A., Freedland, M.B., Larson, C.R., and Hain, T.C. (1998). Voice F0 responses to manipulations in pitch feedback. J Acoust Soc Am *103*, 3153-3161.
35. Liberman, A., Fischer, J., and Whitney, D. (2014). Serial dependence in the perception of faces. Curr Biol *24*, 2569-2574.
36. Goettker, A., and Stewart, E.E. (2022). Serial dependence for oculomotor control depends on early sensory signals. Current Biology *32*, 2956-2961. e2953.
37. Zhang, H., and Alais, D. (2020). Individual difference in serial dependence results from opposite influences of perceptual choices and motor responses. Journal of Vision *20*, 2-2.
38. Miller, H.E., Kearney, E., Nieto-Castanon, A., Falsini, R., Abur, D., Acosta, A., Chao, S.C., Dahl, K.L., Franken, M., Heller Murray, E.S., et al. (2023). Do Not Cut Off Your Tail: A Mega-Analysis of Responses to Auditory Perturbation Experiments. J Speech Lang Hear Res *66*, 4315-4331.
39. Burnett, T.A., Senner, J.E., and Larson, C.R. (1997). Voice F0 responses to pitch-shifted auditory feedback: A preliminary study. Journal of Voice *11*, 202-211.
40. Behroozmand, R., Korzyukov, O., Sattler, L., and Larson, C.R. (2012). Opposing and following vocal responses to pitch-shifted auditory feedback: evidence for different mechanisms of voice pitch control. J Acoust Soc Am *132*, 2468-2477.
41. Franken, M.K., Acheson, D.J., McQueen, J.M., Hagoort, P., and Eisner, F. (2018). Opposing and following responses in sensorimotor speech control: Why responses go both ways. Psychonomic Bulletin & Review *25*, 1458-1467.
42. Niziolek, C.A., and Parrell, B. (2021). Responses to Auditory Feedback Manipulations in Speech May Be Affected by Previous Exposure to Auditory Errors. J Speech Lang Hear Res *64*, 2169-2181.
43. Patel, R., Niziolek, C., Reilly, K., and Guenther, F.H. (2011). Prosodic adaptations to pitch perturbation

in running speech.

44. Patel, R., Reilly, K.J., Archibald, E., Cai, S., and Guenther, F.H. (2015). Responses to intensity-shifted auditory feedback during running speech. Journal of Speech, Language, and Hearing Research *58*, 1687-1694.

45. Ohala, J.J. (1974). A mathematical model of speech aerodynamics. Annual Report of the Institute of Phonetics University of Copenhagen *8*, 11-22.

46. Daneš, F. (1960). Sentence intonation from a functional point of view. Word *16*, 34-54.

47. Lu, Y., Jin, P., Ding, N., and Tian, X. (2022). Delta-band neural tracking primarily reflects rule-based chunking instead of semantic relatedness between words. Cerebral Cortex.

48. Lu, Y., Jin, P., Pan, X., and Ding, N. (2022). Delta-band neural activity primarily tracks sentences instead of semantic properties of words. NeuroImage, 118979.

49. Hanson, H.M. (2009). Effects of obstruent consonants on fundamental frequency at vowel onset in Englisha). The Journal of the Acoustical Society of America *125*, 425-441.

50. Cai, S., Boucek, M., Ghosh, S.S., Guenther, F.H., and 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 ISSP, 65-68.

51. Xu, Y. (2013). ProsodyPro—A tool for large-scale systematic prosody analysis. (Laboratoire Parole et Langage, France).

52. Maris, E., and Oostenveld, R. (2007). Nonparametric statistical testing of EEG-and MEG-data. Journal of neuroscience methods *164*, 177-190.

53. Fritsche, M., Spaak, E., and de Lange, F.P. (2020). A Bayesian and efficient observer model explains concurrent attractive and repulsive history biases in visual perception. Elife *9*.