AI translation · View original & related papers at
chinarxiv.org/items/chinaxiv-202001.00002

Mechanisms of Feedforward and Feedback Control Integration in the Speech Motor System

Authors: Cai Xiao, Zhang Qingfang, Zhang Qingfang

Date: 2020-01-02T00:00:00+00:00

Abstract

The final stage of speech production is articulatory movement, a stage that involves the integrated processing of feedforward and feedback control within the speech motor system. Feedforward control refers to the top-down retrieval and execution of motor commands for producing target speech sounds, whereas feedback control refers to the bottom-up adjustment of speech movements based on sensory feedback generated during vocalization, with sensory targets and sensory predictions serving as crucial links between the two. Drawing upon the DIVA (Directions Into Velocities of Articulators) neural computational model, the cognitive neural mechanisms underlying the integration of feedforward and feedback control are elaborated from the perspectives of both language acquisition and speech production. Building on previous research, this review specifically examines how auditory feedback assists individuals in the online control of speech movements and the updating of feedforward motor representations, as well as the cognitive implications of the P1-N1-P2 component complex in ERP research. Additionally, it summarizes various factors influencing speakers' feedforward and feedback control, including individual differences, training experience, and task context, and proposes key research questions that warrant focused attention in this field.

Full Text

The Integration Mechanisms of Feedforward and Feedback Control in the Speech Motor System

CAI Xiao; ZHANG Qingfang*
(Department of Psychology, Renmin University of China, Beijing 100872, China)

Abstract

The final stage of spoken production is articulation, which involves the integration of feedforward and feedback control in the speech motor system. Specifically, feedforward control refers to the top-down retrieval and execution of motor commands for producing target speech sounds, while feedback control refers to the bottom-up adjustment of speech movements based on sensory feedback generated during articulation. Sensory goals and sensory predictions serve as critical hubs linking these two systems. Based on the DIVA (directions into velocities of articulators) neural computational model, this paper elaborates on the cognitive and neural mechanisms underlying the integration of feedforward and feedback control across two stages: language acquisition and speech production. Building on previous research, we focus on how speakers utilize auditory feedback to control online speech movements and update feedforward motor representations, as well as the cognitive significance of the P1-N1-P2 component waves in ERP studies. Furthermore, we summarize various factors influencing speakers' feedforward and feedback control, including individual differences, training experience, and task contexts, and propose key research questions that should be prioritized in this field.

Key words: speech motor system; feedforward control; feedback control; auditory feedback

Speech represents one of the most important means of interpersonal communication and one of the most sophisticated motor skills mastered by humans. Speech production is generally considered to involve three processes: (1) conceptualization, which establishes communicative intentions and concepts to be expressed; (2) formulation, which converts these concepts into linguistic forms; and (3) articulation, which involves planning specific phonetic and articulatory gestures (Zhang & Yang, 2003). While conceptualization and formulation constitute the planning processes of speech production, articulation represents the execution process. Current psycholinguistic research has primarily focused on the cognitive mechanisms of the planning stage (Levelt, Roelofs, & Meyer, 1999), neglecting the complex mechanisms and theoretical elaboration involved in the execution process. Two meta-analyses of speech production have shown that many researchers treat terminal articulatory execution as simple, low-level motor output (Indefrey, 2011; Indefrey & Levelt, 2004). In reality, transforming articulatory plans into perceivable and comprehensible sound sequences through coordinated movements of the vocal apparatus is an extremely complex process involving high-level mental activities such as anticipating, comparing, and controlling articulatory movements. These mental activities require participation of extensive brain networks including motor cortex, sensory cortex, insula, and cerebellum (Golfinopoulos, Tourville, & Guenther, 2010). Given this complexity, researchers have proposed that speech production should also be studied from a motor control perspective to complement traditional planning-focused approaches, providing theoretical and practical guidance for a comprehensive understanding of speech production mechanisms (Hickok, 2012).

Early researchers focused on general motor control processes—how individuals produce precise, goal-directed movements—emphasizing the important role of sensory feedback in motor control (Wolpert, Diedrichsen, & Flanagan, 2011). In recent years, researchers have recognized that spoken production is essentially a refined form of motor control, where speakers judge the accuracy of vocal output through sensory feedback, particularly the more critical auditory feedback (Guenther, 2006; Guenther & Vladusich, 2012; Scheerer & Jones, 2012). As exploration of speech motor control has deepened, researchers have gradually recognized the cognitive significance and functional mechanisms of feedforward and feedback control. In summary, feedforward control refers to the top-down retrieval and execution of motor commands corresponding to target speech sounds from the speech motor system, while feedback control refers to the bottom-up processing of real-time sensory feedback information generated during speech production, with subsequent adjustment and correction of motor output based on speech errors detected in this feedback. These two systems do not operate independently but cooperate with each other to ensure smooth execution of speech production (Guenther, Ghosh, & Tourville, 2006; Parrell, Lammert, Ciccarelli, & Quatieri, 2019; Perkell, 2012; Tourville & Guenther, 2011). Although increasing numbers of researchers have examined the mechanisms and neural foundations of speech production from a motor control perspective (Cai, Ghosh, Guenther, & Perkell, 2010, 2011; Chang, Niziolek, Knight, Nagarajan, & Houde, 2013; Chen et al., 2015; Hickok, 2012; Scheerer & Jones, 2018), relevant research remains fragmented, lacking systematic discussion of the theoretical background, research hotspots, and potential issues in this field. Therefore, this paper discusses speech motor control from the perspectives of basic theories, current research status, and influencing factors regarding the integration of feedforward and feedback control.

2. Basic Theory of Feedforward and Feedback Control Integration

Speech motor control has attracted interest not only from psycholinguists and phoneticians but also from computer scientists and artificial intelligence experts who have devoted considerable effort to neural computational modeling. The most prominent among these is the DIVA (directions into velocities of articulators) neural computational model developed by Professor Guenther and his team. This model focuses not only on how individuals develop feedforward and feedback control capabilities during early language acquisition but also elaborates on how mature speakers integrate these two control systems during speech production to ensure normal speech output.

2.1 Integration of Feedforward and Feedback Control During Language Acquisition

The DIVA model proposes that newborn infants lack speech production capabilities, and that language acquisition requires two developmental stages—babbling and imitation—both involving integration of motor and sensory information.

Babbling Stage. Before individuals can produce speech sounds, they must first establish connections between motor commands for general sounds (i.e., non-speech-specific sounds) and their corresponding sensory feedback. During early babbling, infants begin experimenting with vocalizations, with auditory and somatosensory cortices processing the auditory and somatosensory feedback generated by specific articulatory gestures. After repeated attempts, motor and sensory information become paired. At this point, when auditory or somatosensory errors occur, individuals can transform sensory error signals into corrective motor commands based on these established motor-sensory associations (Guenther et al., 2006; Tourville & Guenther, 2011).

Imitation Stage. After establishing motor-sensory connections for general sounds, individuals enter the speech imitation stage and begin learning to produce specific speech sounds. At this point, speech information from other native speakers provided by the feedback system constitutes templates for imitation, and individuals gradually form auditory goals for speech—namely, the expected auditory feedback when correctly producing specific speech sounds (Tourville & Guenther, 2011; Kearney & Guenther, 2019). Because speech exhibits categorical perception—where a speech sound is perceived as that sound within a certain range of variation—researchers typically use the term "target region" to refer to the auditory goal (Guenther & Vladusich, 2012). Liu and Tian (2018) also noted that auditory goals are represented by auditory outcomes and can be stored in the auditory phonological system. Additionally, through somatosensory feedback from one' s own correct productions—including the positions of articulators (e.g., lips and jaw) and the degree of contact between different articulators (Parrell et al., 2019)—individuals form somatosensory goals, which represent the expected somatosensory feedback when correctly producing specific speech sounds (Tourville & Guenther, 2011). Thus, the feedback system is crucial during early language acquisition, helping individuals form sensory goal representations for native speech sounds in memory.

After acquiring auditory goals, individuals begin learning feedforward motor commands for speech production based on imitation mechanisms (Guenther et al., 2006; Perkell, 2012; Tourville & Guenther, 2011). Initially, because stable and reliable connections between speech sounds and corresponding motor commands have not yet formed, initial attempts at articulatory movements produce large errors, and speech production relies entirely on the feedback control system. However, with each vocalization, the auditory control system compares the actually produced speech with the auditory goal and corrects errors in articulatory movements based on this goal, ultimately enabling accurate acquisition of feedforward motor commands for speech sounds. After extensive practice, motor commands consistently produce the same sensory feedback without generating speech errors, strengthening the connection between motor and sensory information. At this point, speech production primarily relies on feedforward control. In summary, during early language acquisition, the core of feedforward and feedback control integration involves forming sensory goal representations for native speech sounds, acquiring feedforward motor commands for speech production, and establishing stable connections between motor commands and sensory feedback, gradually developing speech motor control skills similar to those of adult speakers.

2.2 Integration of Feedforward and Feedback Control During Speech Production

For mature speakers, speech motor control during language production is more complex, involving feedforward and feedback systems and the hub connecting them. The DIVA model proposes that sensory goals stored in memory systems provide an interface for the cooperative interaction between feedforward and feedback control, while other models such as Task Dynamics (TD) (Saltzman & Munhall, 1989) and State Feedback Control (SFC) (Houde & Chang, 2015) explicitly state that sensory predictions estimated online through internal forward models constitute the critical hub linking feedforward and feedback control. Although the DIVA model does not incorporate internal forward models in its theoretical framework, it does not deny the existence of internal forward mechanisms in speech motor control (Guenther, 1995). Therefore, researchers have proposed that sensory feedback, sensory goals, and motor-based sensory predictions jointly facilitate individuals' ability to develop and maintain precise speech production (Guenther, 1995; Hickok, 2012; Hickok et al., 2011; Houde & Nagarajan, 2011). Based on the relationships among these three factors, we will elaborate on the functional mechanisms of feedforward and feedback control integration.

Open-loop feedforward control. Mature speakers' vocal movements begin with the feedforward control system. When individuals plan to produce specific phonemes or syllables from their speech repertoire, they top-down retrieve the corresponding feedforward motor commands for these speech sounds. These motor commands are sent to the articulators, guiding detailed positional movements of the upper and lower lips, jaw, tongue, and larynx to achieve articulation (see the gray-background portion of Figure 1 [Figure 1: see original paper]) (Guenther, 2016). Feedforward control transmits motor commands to the articulators but does not feed the sensory consequences of articulatory movements back to influence current motor control (Perkell, 2012). This process resembles open-loop control in systems theory—a control method where the controlling and controlled objects interact only unidirectionally without feedback information. Thus, feedforward control is essentially an open loop from motor commands to sound output. In the speech motor system, feedforward control offers clear advantages, as individuals can rapidly issue motor commands without processing complex sensory feedback, ensuring fluent speech production (Parrell et al., 2019). However, during early stages of native or second language acquisition, the accuracy of speech motor commands remains to be established, and errors in motor commands within an open-loop system inevitably lead to speech errors. Therefore, feedforward control is not suitable for early language acquisition stages (Guenther, 2006; Guenther & Vladusich, 2012). Additionally, when individuals are in unstable environments where interference is likely, feedforward control lacks the ability to monitor and correct errors in speech output, making it unsuitable for unstable contexts.

Hub linking feedforward and feedback control. The input to feedforward control systems is speech motor commands, while the input to feedback control systems is sensory feedback generated by articulatory movements. These two types of information cannot be directly compared, so speech motor control models must specify how feedforward control connects with feedback control. Researchers propose that feedforward control systems generate sensory representations for feedback control through two means: sensory goals (retrieved from memory) and motor-based sensory predictions (estimated online) (Liu & Tian, 2018; Tian & Poeppel, 2012).

First, the DIVA model proposes that each speech sound activates not only motor commands but also corresponding sensory goals in memory systems. Sensory goals encode the expected sensory feedback when speech sounds are correctly produced and can therefore be compared with actual sensory feedback. During speech production, individuals continuously compare sensory goals with sensory feedback to determine whether articulatory goals have been achieved (Guenther, 2016). Empirical and computational modeling evidence also indicates that discrepancies between sensory goals and actual sensory feedback can be used to update motor commands for target speech sounds (Guenther, 1994; Hickok, Houde, & Rong, 2011; Lametti, Krol, Shiller, & Ostry, 2014). It should be clarified that although sensory goals are not included in the feedforward control system shown in Figure 1, the top-down pathway from the speech repertoire to sensory goals actually belongs to feedforward control.

Furthermore, efficient feedforward control can anticipate sensory feedback through internal forward models (Franklin & Wolpert, 2011). This prospective nature also earns feedforward control the designation of predictive control (Parrell et al., 2019). At the cognitive level, sensory prediction refers to individuals' internal estimation of current vocal tract states and subsequent likely sensory feedback before actual sensory feedback is generated (Hickok, 2012; Tian & Poeppel, 2010, 2012). This prediction relies entirely on bidirectional connections between speech motor commands and sensory outputs established during language acquisition (Hickok, 2012; Tian & Poeppel, 2010, 2012, 2015; Tian, Zarate, & Poeppel, 2016). The neural basis of internal forward models is the efference copy—a copy of the motor command issued by the feedforward control system to the articulators that is replicated internally. Unlike the original motor command, the efference copy does not act on the articulators but is instead transmitted to auditory and somatosensory cortices to form motor-based auditory and somatosensory predictions (see red box in Figure 1) (Hickok, 2012; Hickok et al., 2011; Niziolek, Nagarajan, & Houde, 2013). Although sensory predictions are not included in the feedforward control system shown in Figure 1, the top-down pathway from feedforward commands to sensory predictions also belongs to feedforward control. Predictive control offers clear speed advantages, as individuals can directly anticipate sensory feedback from motor commands to instantly evaluate whether sensory goals have been achieved (Parrell et al., 2019). However, complex internal forward models are difficult to learn, and when the model itself is imprecise, the anticipated sensory feedback may not match actual sensory feedback. Without participation of the feedback control system under such circumstances, individuals cannot correct prediction errors.

Closed-loop feedback control. Another crucial mechanism in the speech motor system is feedback control, which fundamentally differs from feedforward control in its utilization of sensory feedback from articulatory movements for speech motor control (Parrell et al., 2019). This process resembles closed-loop control in systems theory—a control method where the output of the controlled object is fed back to the controlling object and influences it. Thus, feedback control is essentially a closed loop from motor commands to error signals and from error signals back to motor commands. Feedback control involves a series of processing stages (see the blue-background portion of Figure 1 [Figure 1: see original paper]): First, sensory cortices encode sensory feedback generated by articulatory movements. Next, individuals compare actual sensory feedback with sensory predictions estimated online through internal forward models. In most cases, these match, and speech production proceeds smoothly. Since no correction of current speech output is needed, sensory predictions suppress activation of sensory goals. However, in rare special cases where sensory feedback and sensory predictions do not match, sensory goals become fully activated to encode sensory errors because correction of speech output errors is required. Finally, error signals are transmitted to the feedback control module, which encodes corrective motor commands based on sensory-motor transformations. When sensory errors occur repeatedly, error-based corrective commands update the current feedforward motor representation for the speech sound (Guenther, 2016; Perkell, 2012). Thus, the advantage of feedback control is its suitability for unstable environments, ensuring normal and effective communication through continuous monitoring and correction of speech output (Hickok et al., 2011; Parrell et al., 2019). However, its speed disadvantage is also evident: the nervous system requires additional time to process sensory feedback signals and make appropriate motor adjustments based on feedback. Consequently, excessive reliance on sensory feedback for speech motor control can lead to slow movements and reduced stability (Civier, Tasko, & Guenther, 2010; Guenther, 2006; Perkell, 2012).

In summary, feedforward and feedback control systems each have their own advantages and disadvantages. Flexible speech motor control requires combining both systems, thereby preserving the speed of feedforward control while also addressing errors in speech production or unexpected external perturbations (Guenther, 2016; Parrell et al., 2019). Based on these processing mechanisms, the DIVA model includes numerous modules and corresponding parameters, constituting a large-scale process model. Professor Guenther has acknowledged that the current theoretical model includes many free parameters that should allow good fit to specific experimental data, yet empirical research has difficulty verifying the model' s assumptions. Therefore, Guenther' s team is working to develop a simplified, testable three-parameter computational model that includes only auditory feedback control gain (A), somatosensory feedback control gain (S), and feedforward control/learning rate (FF), aiming to assess the relative contributions of feedforward and feedback control systems during speech production (Kearney et al., submitted).

2.3 Neural Basis of Speech Motor Control

The brain network involved in speech production includes bilateral medial and lateral frontal cortex, parietal cortex, superior temporal cortex, thalamus, basal ganglia, and cerebellum (Bohland & Guenther, 2006). However, the specific roles of each brain region and their interaction patterns during speech production remain unclear. Based on numerous fMRI studies, the DIVA model identifies brain regions involved in various cognitive processes of feedforward and feedback control, providing neural-level insights for a more comprehensive understanding of speech motor control.

The DIVA model hypothesizes that the brain region responsible for processing the speech repertoire is located in the left-hemisphere ventral premotor cortex, primarily including the ventral precentral gyrus and adjacent posterior inferior frontal gyrus and anterior insular region. The premotor cortex participates in retrieving finely encoded articulatory motor programs for frequently used speech sounds, which constitute feedforward commands in speech motor control (Guenther, 2006, 2016; Guenther & Vladusich, 2012). Evidence for this hypothesis comes from numerous brain lesion studies of speech motor disorders (Kearney & Guenther, 2019). For example, researchers have found that patients with apraxia of speech exhibit normal speech production abilities before brain injury but develop deficits in motor planning and encoding after brain injury (typically in ventral premotor cortex regions) (Ballard, Tourville, & Robin, 2014; New et al., 2015).

The DIVA model emphasizes multiple functions of the cerebellum. First, the cerebellum is essential for learning and updating fine feedforward motor commands, and its damage likely leads to motor disorders (Ito, 2000). Second, the cerebellum plays a role in forming sensory goals (O' Reilly, Mesulam, & Nobre, 2008). Third, the cerebellum may participate in feedback control, as cerebellar activity correlates with the magnitude and frequency of sensory errors, which in turn drive corrective motor commands (Grafton, Schmitt, Van Horn, & Diedrichsen, 2008). To address these controversies, Parrell et al. (2017) compared performance on feedforward and feedback control tasks between patients with cerebellar degeneration and normal controls, finding that patients with cerebellar degeneration showed reduced ability to maintain accurate feedforward commands but similar feedback control abilities compared to controls.

This supports the view that the cerebellum' s more critical role is updating and maintaining feedforward motor commands.

Neuroimaging studies have found that when experimenters artificially alter auditory feedback returned to participants' ears, significant activation is observed in temporal regions, indicating that these brain areas are involved in processing auditory errors (Fu et al., 2006; Tourville, Reilly, & Guenther, 2008; Toyomura et al., 2007). For example, Tourville et al. (2008) required participants to produce monosyllabic words (e.g., "beck" and "bet") while randomly altering the first formant of their acoustic signals (increased or decreased by 30%). Comparing brain activation between altered and normal auditory feedback conditions, the study found that auditory errors significantly activated the posterior superior temporal gyrus and transverse temporal gyrus, with peak activation located at the posterior end of the left transverse temporal gyrus. Numerous studies have also confirmed that the transverse temporal gyrus is responsible for monitoring auditory errors and encoding the degree of mismatch between auditory feedback and auditory predictions (Parkinson, Flagmeier, Manes, Larson, Rogers, & Robin, 2012; Zheng, Munhall, & Johnsrude, 2010). Hickok (2012) proposed in a theoretical model that the planum temporale integrates auditory and motor information, leading to this region being termed the auditory-motor interface (Hickok et al., 2011; Hickok, Okada, & Serences, 2009).

Feedback control primarily activates right-hemisphere brain regions (Kalpouzos & Nyberg, 2010). Tourville and Guenther (2011) first introduced a right-lateralized feedback control module in the DIVA model, responsible for generating corrective motor commands when speech errors occur. Early versions of the DIVA model predicted bilateral activity in motor cortex, primarily concentrated in primary motor cortex (Guenther et al., 2006). However, Tourville et al. (2008) found that artificially altering formants activated right premotor cortex during auditory feedback abnormalities, and Golfinopoulos et al. (2011) found that artificially altering lip or jaw positions similarly activated right premotor cortex during somatosensory feedback abnormalities.

In summary, the DIVA model elaborates on the integration mechanisms of feedforward and feedback control during both language acquisition and speech production from a developmental perspective. During language acquisition, individuals establish representations of sensory goals and motor commands for speech sounds and the connections between them. For mature speakers, feedforward control can either directly retrieve motor commands for target speech sounds or retrieve sensory goals from memory systems or generate sensory predictions corresponding to motor commands through internal forward models. Meanwhile, feedback control monitors speech production errors in real time and promptly corrects and updates feedforward motor representations. From a neural perspective, the speech motor system involves extensive bilateral brain regions. Future research should aim to reveal more refined brain networks and the connection patterns between different brain regions.

3. Current Research on Speech Motor Control

Early researchers typically investigated how auditory feedback participates in speech motor control at the behavioral level. In recent years, with the development of electrophysiological techniques (e.g., event-related potentials, a special type of brain-evoked potential recorded from the scalp) and brain imaging techniques (e.g., magnetoencephalography [MEG] and functional magnetic resonance imaging [fMRI]), many researchers have combined behavioral methods with ERP, MEG, and fMRI to explore the integration mechanisms of feedforward and feedback control, yielding fruitful results.

3.1 Functions of Auditory Feedback: Behavioral Studies

Current researchers have begun to examine the role of auditory feedback in language acquisition and speech production, proposing three primary functions: forming and maintaining speech motor skills, online control of speech production, and updating feedforward motor representations (Cai, 2012; Guenther et al., 2006; Civier et al., 2010).

The primary role of auditory feedback during language acquisition is to form speech motor skills. Research has shown that congenitally deaf children cannot acquire fluent speech production abilities (Cowie et al., 1982), and prelingually deaf children with mild to severe hearing loss typically cannot produce intelligible speech (Oller & Eilers, 1988). However, when hearing ability is restored through cochlear implants, speech production abilities are greatly facilitated (Tye-Murray & Spencer, 1995). This evidence demonstrates that normal hearing ability is a necessary condition for acquiring speech motor skills. Even after language acquisition, auditory feedback remains important for maintaining speech motor skills. Research has found that postlingual hearing loss leads to deterioration in many aspects of speech motor control, including speech rate, intensity, and fundamental frequency (Lane & Webster, 1991), while cochlear implantation promotes recovery of speech intelligibility (Gould et al., 2001).

Most of these studies have employed longitudinal designs focusing on slow changes in speech motor skills and thus cannot examine the role of auditory feedback in real-time speech motor control. Moreover, native speakers' motor commands for speech production are highly automated, with auditory predictions typically matching auditory feedback, making it difficult to investigate feedback-based motor control (Simmonds, Wise, & Leech, 2011). Early researchers used delayed auditory feedback paradigms, employing special equipment and software to delay the time for speakers' voices to reach their ears, which typically produces stuttering-like disfluencies. This paradigm interferes with the integration of motor information and auditory feedback, demonstrating that speech production is influenced by auditory feedback (Mitsuya, Munhall, & Purcell, 2017; Tian & Poeppel, 2015). However, some researchers have questioned whether the obvious alteration of natural auditory feedback patterns allows participants to become aware of the manipulation, advocating for more subtle experimental manipulations (Cai, 2012).

In recent years, modern acoustic signal processing technology has revolutionized research methods in speech motor control, giving rise to auditory feedback perturbation paradigms (Cai, Beal, Ghosh, Tiede, Guenther, & Perkell, 2012; Cai et al., 2010). Since auditory feedback manifests as explicit acoustic signals that are easily obtained and manipulated, researchers have begun to alter various acoustic parameters in auditory feedback in precisely controlled ways, with the most commonly used including intensity (Bauer, Mittal, Larson, & Hain, 2006; Patel, Reilly, Archibald, Cai, & Guenther, 2015), fundamental frequency (Chang et al., 2013; Franken et al., 2018a, 2018b), and formants (Cai et al., 2012; Daliri, Wieland, Cai, Guenther, & Chang, 2017). The primary advantage of perturbation paradigms is establishing causal relationships between auditory feedback perception and speech motor adjustments. Researchers have focused on several key questions:

First, how do participants adjust speech movements when auditory feedback perturbations occur? Numerous studies have found that when perturbation magnitude does not exceed threshold (i.e., the voice is still likely recognized as self-produced), speakers typically adjust speech movements in the direction opposite to the artificial perturbation, a phenomenon known as compensatory response. For example, when the fundamental frequency of feedback is artificially increased, speakers automatically lower their fundamental frequency (Chang et al., 2013). This opposite-direction compensatory response indicates that the goal of speech motor control is to maintain stable auditory feedback.

Second, what is the time course of motor adjustments based on auditory feedback errors, and is the integration process of feedforward and feedback control automatic or controlled? Research has found that speakers adjust speech movements very rapidly after auditory feedback perturbation, typically within 100-200 ms (Bauer et al., 2006; Franken et al., 2018a; Cai et al., 2012). Scheerer and Jones (2018) found that the feedback control system is sensitive even to small deviations in speech production, and post-experiment surveys indicated that participants were sometimes unaware of the perturbations and their compensatory responses (Cai et al., 2011; Parrell et al., 2017). Munhall et al. (2009) perturbed formants while informing participants about the specific manipulation and instructing them not to compensate. Results showed that even after being informed, participants still adjusted their movements in the direction opposite to the perturbation, indicating that compensatory responses are not easily influenced by conscious strategies. This evidence—rapid onset, unconscious correction, and resistance to strategic influence—all supports the view that the integration of feedforward and feedback control is an automatic process (Korzyukov et al., 2012a).

Third, what is the relationship between the magnitude of auditory feedback perturbation and the magnitude of speech motor adjustments, and what are the underlying cognitive mechanisms? Researchers have found that compensatory response magnitude typically constitutes only a small portion of the auditory feedback perturbation magnitude. For example, a 100-cent pitch perturbation in auditory feedback can only elicit compensatory responses smaller than 50 cents (Korzyukov et al., 2012a; Scheerer et al., 2013b), with some studies reporting only approximately 8-cent compensatory responses (Scheerer et al., 2013a). On one hand, this is because speech motor control relies on both feedforward and feedback control systems (Tourville & Guenther, 2011). On the other hand, since auditory feedback perturbation paradigms only create mismatches between expected and actual auditory feedback while expected and actual somatosensory feedback remain matched, the feedback control system must simultaneously process disrupted auditory feedback and consistent somatosensory feedback. This asymmetry between perturbation and compensation demonstrates that auditory feedback is not the sole information input in the speech motor system; feedforward control and somatosensory feedback also participate in speech motor control.

The third function of auditory feedback is updating feedforward motor representations. Delvaux and Soquet (2007) found that adult speech production remains influenced by surrounding linguistic environments, with people adjusting speech patterns (e.g., pitch, vowel characteristics) according to their environment. This indicates that after feedforward commands are established, they are not fixed but are continuously calibrated and updated through auditory feedback. Researchers have developed speech adaptation paradigms to investigate how long-term auditory feedback perturbations influence updates to feedforward motor representations (Cai et al., 2010; Daliri et al., 2017; Parrell et al., 2017). Currently, this paradigm is widely applied in research on fundamental frequency and formant control. A typical speech adaptation paradigm includes four consecutive phases (Daliri et al., 2017): (1) baseline phase—speaking under normal auditory feedback conditions; (2) ramp phase—speaking under auditory feedback perturbation conditions with gradually changing perturbation magnitude until reaching peak; (3) hold phase—continuously speaking under maximum-magnitude auditory feedback perturbation conditions to reshape speech motor representations across multiple trials; and (4) end phase—finally speaking under normal auditory feedback conditions. Research has found that during the hold phase, speakers show compensatory responses opposite to the perturbation direction compared to baseline, also called adaptive responses. When perturbations are withdrawn during the end phase, adaptive responses briefly persist, a phenomenon known as aftereffect (Daliri et al., 2017; Parrell et al., 2017). Since both baseline and end phases involve normal auditory feedback, the aftereffect in the end phase demonstrates that participants have updated feedforward motor representations for specific speech sounds. This shows that artificially induced auditory errors trigger the feedback control system to correct movements, and over longer timescales, corrective commands are incorporated into the feedforward control system to guide subsequent articulatory behavior.

Researchers have also examined whether speech adaptation generalizes to other unperturbed speech sounds. Cai et al. (2010) investigated how Mandarin speakers produce the triphthong /iau/ when facing formant perturbations and whether adaptive responses generalize to vowels with different temporal or spatial features, such as /uai/, /ia/, and /au/. Results showed that generalization patterns were widespread but weak, decreasing as similarity between speech sounds diminished. Reilly and Pettibone (2017) also found that repeatedly perturbing auditory feedback for specific vowels not only changed production of the perturbed vowel (adaptive response) but also changed production of nearby unperturbed vowels (generalization). This evidence indicates that feedforward motor commands for different vowels are not independently represented; otherwise, feedforward representation updates for one vowel based on auditory feedback errors would not affect production of other vowels. Therefore, researchers speculate that speech sounds with similar temporal and spatial features may share certain mechanisms responsible for computing speech motor trajectories, such that adjustment of motor-sensory mappings for one vowel leads to changes in motor encoding for other vowels. This poses a challenge to the DIVA model, which treats different speech sounds as independent entities with separately stored feedforward motor commands (Guenther et al., 2006; Tourville & Guenther, 2011) and therefore cannot explain generalization responses found across different vowels. Further research is needed to explain generalization between different speech sounds to refine the DIVA model.

3.2 Temporal Course of Feedforward and Feedback Control Integration: ERP and MEG Studies

Although numerous studies have revealed the mechanisms of speech motor control at the behavioral level, behavioral research can only rely on explicit motor adjustments and cannot examine the process of auditory feedback error monitoring. With the rise of cognitive neuroscience, many researchers have adopted auditory feedback perturbation paradigms combined with high-temporal-resolution techniques such as ERP and MEG to investigate the temporal course of feedforward and feedback control integration (Behroozmand et al., 2016; Chen et al., 2012a, 2012b; Franken et al., 2018b; Heinks-Maldonado, Mathalon, Gray, & Ford, 2005; Scheerer et al., 2013a, 2013b; Scheerer & Jones, 2014, 2018). Researchers have examined the relationship between auditory feedback perturbation magnitude and neural activity changes, identifying P1-N1-P2 component waves that reflect speech motor control (Chen et al., 2012b; Liu et al., 2011; Scheerer et al., 2013a).

First, researchers generally agree that the P1 component (corresponding to the M50 component in MEG experiments) reflects early monitoring of auditory stimulus changes. For example, Scheerer et al. (2013a) recorded ERPs while randomly perturbing auditory feedback and found that the P1 measure was sensitive only to whether perturbation occurred but not to perturbation magnitude. Korzyukov et al. (2012a) also found that P1 was evoked in an all-or-none fashion, reflecting monitoring of general auditory stimuli.

Compared to the P1 component, earlier researchers paid more attention to the language-specific N1 component wave (corresponding to the M1 or M100 component in MEG experiments) and the speech-induced suppression phenomenon (SIS). SIS is specifically manifested as: comparing participants listening to their own normal auditory feedback during vocalization versus listening to their own perturbed auditory feedback (Heinks-Maldonado et al., 2005, 2006), or comparing listening to their own normal auditory feedback during vocalization versus listening to others' voices (Heinks-Maldonado et al., 2005, 2006), or comparing listening to their own normal auditory feedback during vocalization versus passively listening to recordings of their own voice without vocalizing (Houde, Nagarajan, Sekihara, & Merzenich, 2002). In all cases, N1/M1 amplitudes are reduced. The SIS phenomenon is also supported by fMRI evidence, with studies finding reduced auditory cortex activity when participants listen to their own normal auditory feedback during vocalization compared to listening to their own perturbed auditory feedback (Parkinson et al., 2012; Zheng et al., 2010). Thus, SIS essentially reflects the mechanism of anticipating sensory feedback through internal forward models (Behroozmand et al., 2016; Heinks-Maldonado et al., 2005; Kort, Nagarajan, & Houde, 2014). When expected auditory feedback matches actual auditory feedback, the SIS phenomenon is evoked; when they do not match, auditory cortex requires additional resources to process speech errors. SIS is also considered an important mechanism for speakers to distinguish self-generated from externally generated speech.

Researchers believe that changes in N1 amplitude reflect individuals' monitoring of auditory feedback perturbations and discrimination between internally and externally generated speech (Behroozmand & Larson, 2011; Heinks-Maldonado et al., 2005; Liu et al., 2011; Scheerer et al., 2013a). In Scheerer et al.' s (2013a) study, any auditory feedback perturbation condition evoked larger N1 amplitudes than the no-perturbation condition, and amplitude changes were affected by perturbation magnitude, with larger perturbations (400 cents) evoking greater N1 amplitudes than relatively smaller perturbations (50-250 cents). This aligns with Liu et al.' s (2011) findings that 500-cent and 200-cent perturbations evoked larger N1 amplitudes than 100-cent perturbations (see also Behroozmand & Larson, 2011). Therefore, researchers speculate that the N1 component actually reflects speakers' comparison of auditory feedback with auditory predictions. When actual auditory feedback violates predictions but can still be recognized as self-produced speech, N1 amplitude increases in an all-or-none fashion. However, when large auditory feedback perturbations occur, speakers judge the auditory feedback as coming from others rather than themselves, evoking larger N1 amplitudes. N1 latency may reflect the efficiency of auditory feedback processing, with studies showing that N1 latency shortens as perturbation magnitude increases, indicating faster detection of large auditory feedback errors (Liu et al., 2011; Scheerer et al., 2013a).

Additionally, researchers have increasingly focused on neural manifestations of feedback-error-based motor control and have identified the P2 component wave (corresponding to the M2 or M200 component in MEG experiments). Unlike N1, P2 amplitude shows the most systematic variation according to perturbation magnitude. Scheerer et al. (2013a) found that within the 0-250 cent perturbation range, P2 amplitude increased with perturbation magnitude, but P2 amplitude did not continue to increase linearly—when perturbation magnitude exceeded 250 cents, amplitude began to decrease. Notably, compensatory responses at the behavioral level show patterns similar to P2 amplitude changes. When auditory feedback perturbations are small, compensatory responses increase with perturbation magnitude; however, when large auditory feedback perturbations occur, speakers judge the auditory feedback as coming from others rather than themselves and therefore do not produce compensatory responses or produce responses with smaller magnitude (Scheerer et al., 2013a; Tian & Poeppel, 2015). Moreover, both behavioral-EEG correlation and regression analyses have shown positive correlations between behavioral compensatory responses and P2 amplitude. Therefore, researchers believe that P2 may reflect computation of auditory errors and corresponding motor command corrections (Chen et al., 2015; Jones, Scheerer, & Tumber, 2013; Kort et al., 2014; Scheerer et al., 2013a; Scheerer & Jones, 2014).

A few researchers have begun using time-frequency analysis to explore how the brain integrates motor and auditory feedback information for speech motor control, finding that oscillatory activity in the theta and delta frequency bands is significant (Behroozmand et al., 2015; Cavanagh & Frank, 2014; Cruikshank, Singhal, Hueppelsheuser, & Caplan, 2012). For example, Cavanagh and Frank (2014) found that delta band (1-4 Hz) and theta band (5-8 Hz) activity reflects processing of novel, conflicting, and error information in perturbed auditory feedback, thereby marking cognitive control demands at the neural level. Behroozmand et al. (2015) further compared neural activity differences between musicians and non-musicians during pitch auditory feedback perturbation experiments. The study found that when participants produced compensatory responses to auditory feedback perturbations, this was accompanied by phase-locked theta band oscillatory activity in midfrontal regions, with musicians showing stronger theta band oscillations than non-musicians. One second after perturbation onset, non-phase-locked delta band oscillatory activity appeared in frontal regions, with musicians showing weaker delta band activity than non-musicians. Additionally, delta band oscillatory activity was related to participants' ability to readjust pitch to baseline levels after perturbation offset. Based on these findings, researchers speculate that theta band activity marks musicians' enhanced pitch processing ability at the neurophysiological level, reflecting the mechanism by which humans integrate auditory feedback information to control speech production (Behroozmand et al., 2015; Cruikshank et al., 2012). In contrast, delta band activity marks speech adaptation mechanisms—how individuals update feedforward motor commands based on auditory feedback errors to guide subsequent speech production (Behroozmand et al., 2015). Time-frequency studies supplement the neural mechanisms of feedforward and feedback control integration from different angles, and future research should continue to explore the cognitive significance of neural oscillatory activity.

In summary, changes in P1, N1, and P2 waveforms indicate that when auditory feedback perturbations occur, individuals can detect auditory information abnormalities early, recruit additional cognitive resources to process auditory errors, and adjust speech movements based on feedback information. Theta and delta frequency bands may participate in the integration of motor-sensory information.

4. Factors Influencing the Integration of Feedforward and Feedback Control

Normal operation of the speech motor system depends on cooperative interaction between feedforward and feedback control (Guenther et al., 2006; Perkell, 2012; Tourville & Guenther, 2011; Hickok, 2012). Therefore, the relative contributions of these two systems to speech motor control and their influencing factors constitute important research questions. Researchers have examined the relative weighting of feedforward and feedback control by observing the degree to which individuals are disrupted by auditory feedback perturbations. The experimental logic is that if participants rely less on feedforward control, they will depend more on sensory feedback and consequently be more susceptible to interference from auditory feedback perturbations. Numerous studies using compensatory responses or P1-N1-P2 component waves as measures have identified influencing factors in three main areas: (1) individual differences, including age, sex, vocal variability, and speech disorders; (2) training experience, including language and musical experience; and (3) task context, including predictability and attentional load.

4.1 Individual Differences

Feedforward and feedback control performance shows individual differences, and researchers have focused on the essential causes of these differences, particularly motor control deficits in populations with speech production disorders.

Age and sex. Integration of speech and motor representations begins with infant babbling and remains plastic throughout subsequent language acquisition to accommodate growth of articulatory organs, increases in muscle tissue, and changes in lung capacity (Guenther et al., 2006). Studies have examined differences in auditory feedback control between children and adults. For example, Liu et al. (2010a) found that English-speaking children aged 7-12 years showed longer compensatory response latencies than adults. Scheerer, Liu, and Jones (2013b) conducted a cross-sectional study examining participants aged 4-30 years and found that both compensatory responses and P1-N1-P2 amplitudes were affected by age. For children acquiring language, auditory feedback helps them establish feedforward representations, and corresponding weighting is increased. However, when development ceases, feedforward representations remain relatively stable, and information provided by auditory feedback becomes redundant. Therefore, increasing feedforward control weighting can enhance speech fluency and reduce external interference (Civier et al., 2010). Evidence suggests that females produce smaller compensatory responses than males in auditory feedback perturbation experiments, with shorter N1 and P2 latencies (Chen et al., 2010; Swink & Stuart, 2012). Li et al. (2018) found that young males evoked larger N1 and P2 amplitudes than young females. Researchers believe that physiological differences between sexes may account for these differences in feedforward and feedback control (Chen, Liu, Jones, Huang, & Liu, 2010; Kakimoto et al., 2016).

Vocal variability. In pitch perturbation tasks, experimenters typically require participants to perform sustained vowel production tasks. Vocal variability refers to the standard deviation of pitch variation during baseline conditions without auditory feedback perturbation. Some researchers have examined how individual differences in vocal variability influence auditory feedback control (Scheerer & Jones, 2012; Scheerer et al., 2013a). Scheerer and Jones (2012) found that participants with greater vocal variability produced larger compensatory responses to auditory feedback perturbations, revealing that instability in speech output leads individuals to increase dependence on auditory feedback input to maintain stable speech motor control.

Speech disorders. Based on existing neuroimaging research, the DIVA model identifies brain lesion regions associated with different speech disorders (Guenther, 2016). The severity of speech disorders depends on whether lesion regions affect the feedforward control system or the feedback control system. During language acquisition, the feedback control system is indispensable for forming feedforward commands. However, after language acquisition, the feedforward control system can issue motor commands and produce speech with minimal participation from the feedback control system. Therefore, for mature speakers, damage to brain regions responsible for the feedback control system has relatively limited impact on speech output. In contrast, damage to brain regions responsible for the feedforward control system causes obvious speech motor disorders (Kearney & Guenther, 2019). Common speech production disorders such as apraxia of speech, stuttering, and dysarthria have received extensive attention from researchers.

Apraxia of speech is a motor planning encoding disorder characterized by slow speech rate, distorted speech sounds, and abnormal prosody. Brain lesions in apraxia of speech patients are primarily located in left-hemisphere inferior frontal regions, particularly the ventral premotor cortex, which is responsible for retrieving finely encoded feedforward motor commands (Guenther, 2006, 2016). Therefore, apraxia of speech mainly results from deficits in the feedforward motor control system (Kearney & Guenther, 2019; Tourville & Guenther, 2011). Researchers have proposed that lesions in inferior frontal regions may also affect individuals' ability to retrieve sensory goals for speech sounds, and since the operational mechanism of feedback control involves comparing sensory goals with actual sensory feedback, damage to this brain region may consequently impair feedback control (Ballard et al., 2018; Kearney & Guenther, 2019). However, feedback control deficits have not been fully verified by empirical research. Maas et al. (2015) used a noise masking paradigm to investigate the nature of apraxia of speech, finding that patients' feedforward control was disrupted, causing feedback control to play a more prominent role (see also Iuzzini-Seigel, Hogan, Guarino, & Green, 2015), a conclusion also supported by computational simulation evidence (Terband, Rodd, & Maas, 2015).

Stuttering is a fluency disorder characterized by sound repetitions, sound prolongations, and silent pauses. The DIVA model suggests that stutterers show higher activation than normal controls in the main brain region comprising the feedback control module (right-hemisphere precentral and inferior frontal regions), which is responsible for computing corrective motor commands based on feedback errors. Therefore, stuttering mainly results from deficits or abnormalities in feedback control (Tourville & Guenther, 2011). At the behavioral level, Cai et al. (2012) confirmed through random perturbation of auditory feedback that stutterers have deficits in auditory-motor transformation functions (Daliri et al., 2017). An alternative view holds that stutterers' ability to directly retrieve motor commands from the feedforward control system is impaired, causing them to rely excessively on auditory feedback for motor control and affecting speech production fluency (Civier et al., 2010). At the electrophysiological level, Daliri and Max (2015a, 2015b) confirmed that stutterers have deficits in the general ability to anticipate auditory feedback, suggesting that their feedforward control system may be compromised.

4.2 Training Experience

Language experience. Research has found that experience with tonal languages reduces individuals' susceptibility to pitch perturbations. Because tonal languages use pitch changes to distinguish meaning, while non-tonal languages typically use pitch changes only to modify intonation, speakers of different languages develop language-specific pitch control abilities (Chen et al., 2012b; Liu et al., 2010b; Ning, Shih, & Loucks, 2014; Ning, Loucks, & Shih, 2015). English is a typical non-tonal language, while Mandarin Chinese is a typical tonal language. Ning et al. (2014) first compared the effects of tonal language experience on pitch control and extended the research question to second language learning (Mandarin L2 learners). Results showed that Mandarin native speakers produced the smallest compensatory responses, while L2 learners' performance patterns fell between those of Mandarin and English native speakers, revealing how language training experience shapes pitch feedforward and feedback control. Liu et al. (2010b) compared auditory feedback control differences between Cantonese and Mandarin native speakers. Although both languages are tonal, Cantonese has a more complex and refined tonal system (6 tones) than Mandarin (4 tones). Results showed that Cantonese native speakers produced smaller compensatory responses than Mandarin native speakers. These studies all confirm that refined tonal systems enable speakers to have stable pitch feedforward control abilities and thus be less susceptible to auditory feedback perturbations.

Musical experience. Researchers have found that musical training influences pitch feedforward and feedback control abilities. When experimenters randomly perturb pitch in auditory feedback during sustained vowel production tasks, experienced musicians produce smaller compensatory responses than non-musicians, indicating that musical training enhances pitch control stability and reduces interference from external feedback (Jones & Keough, 2008; Keough, Hawco, & Jones, 2013). Ning et al. (2015) compared four groups of participants during auditory feedback perturbation tasks: Mandarin native speakers, Mandarin L2 learners, trained singers, and speakers without tonal language experience. Results showed that Mandarin native speakers demonstrated the most stable pitch control in both vowel production and Mandarin tone tasks, while trained singers also showed more stable pitch control than speakers without tonal language experience in vowel production tasks, but no differences were found in language-specific tasks. These results indicate that experience controlling tone/pitch helps form more stable feedforward motor representations, but whether pitch feedforward control enhanced through musical training can generalize to language tasks remains to be verified.

4.3 Task Context

Predictability. The more stable the frequency of specific events, the stronger the ability to predict their occurrence. In speech motor control, feedforward control can anticipate auditory feedback through internal forward models. Therefore, researchers hypothesized that if auditory feedback is perturbed in a predictable manner, participants might increase feedforward control weighting and consequently reduce feedback interference. This hypothesis has received substantial support (Chen et al., 2012a; Korzyukov et al., 2012b; Scheerer & Jones, 2014). Korzyukov et al. (2012b) found that predictability of perturbation direction evoked smaller N1 amplitudes. Scheerer and Jones (2014) found that predictability of perturbation magnitude resulted in smaller compensatory response magnitudes, reduced N1 amplitudes, and faster compensatory response and N1 latencies. These findings confirm that predictability increases feedforward control system weighting.

Attentional load. In daily life, speakers may need to process information from other modalities while receiving auditory feedback, but attentional resources are limited. Therefore, some researchers have proposed that attentional load may influence auditory feedback processing. Tumber, Scheerer, and Jones (2014) randomly perturbed auditory feedback under single-task and dual-task conditions. The single-task condition required participants to passively watch visual cues while vocalizing, while the dual-task condition required participants to identify letter strings while vocalizing to increase attentional load. Results showed that single-task conditions produced larger compensatory responses than dual-task conditions, indicating that when participants' attention was divided, less attention was allocated to processing auditory errors. Liu et al. (2015) further validated this view, finding that when participants selectively attended to auditory feedback, they produced larger compensatory responses. However, Alsius, Mitsuya, and Munhall (2013) found no differences in compensatory responses between focused and divided attention tasks. Given these contradictory results, future research needs to explore the mechanisms of attentional modulation of auditory feedback control.

5. Summary and Outlook

Understanding how feedforward and feedback control systems cooperate in the brain to ensure normal speech production has important theoretical and practical significance. Over the past decade, researchers have devoted considerable effort to constructing detailed theoretical frameworks and neural networks for speech motor control, particularly focusing on how individuals use auditory feedback to adjust speech movements online and update feedforward motor representations. Speakers' ability to directly retrieve motor commands through feedforward control and to correct motor commands using sensory feedback through feedback control is influenced by multiple factors. Future research can further explore the following directions:

First, most studies of online speech motor control have focused on relatively steady-state monophthongs such as /a/ or /u/ (Chen et al., 2012b; Scheerer et al., 2013a). Typical pitch perturbation experiments often require participants to prolong pronunciation, exaggerating the static features of vowels. However, articulatory movements in everyday communication cause vocal tract deformation, such as transitions between consonants and vowels, making dynamic change over time an important feature of speech. In contrast, prolonged static articulation rarely occurs in natural continuous speech, resulting in low ecological validity. Future research should not only focus on simple vowels but also examine natural language to achieve more comprehensive understanding of speech motor control characteristics.

Second, from a theoretical modeling perspective, although DIVA is one of the most comprehensive and reasonable models in the speech motor control field, it still has several limitations. First, it primarily focuses on articulatory movements of independent small units such as syllables or frequently used phrases, neglecting how the speech motor system controls articulatory transitions in multisyllabic units. Second, the DIVA model cannot explain generalization effects found in auditory feedback perturbation experiments—why perturbing auditory feedback for specific speech sounds leads to articulatory adjustments in adjacent speech sounds. Third, psycholinguistic models propose that speech production also includes stages such as concept selection, lexical selection, and word form encoding (Levelt et al., 1999), and Hickok (2012) noted that the DIVA model focuses more on dynamics, motor trajectories, and feedback control while neglecting connections with traditional psycholinguistic perspectives. Finally, the model does not address prosodic control in speech production, even though pitch, intensity, duration, and rhythm convey important linguistic and emotional information (Tourville & Guenther, 2011). Currently, Professor Guenther' s team is developing the GODIVA (Gradient-order) model. Future researchers also need to explore speech motor control mechanisms at larger linguistic units and suprasegmental levels.

Third, regarding influencing factors, individual differences, training experience, and task contexts have all been shown to be closely related to the integration of feedforward and feedback control. However, some findings in this field remain inconsistent. Moreover, how these factors influence or jointly influence speech motor control is far from fully understood. Future research should focus on examining factors influencing the integration of feedforward and feedback control and their theoretical underpinnings.

Fourth, from a language typology perspective, research on second language motor control in bilinguals is just beginning, and exploration of differences from the native language urgently needs strengthening. Simmonds et al. (2011) proposed that although native and second languages share perceptual-motor brain networks, theoretically, motor commands in the native language are highly automated and efficiently integrate information from feedforward motor, auditory feedback, and somatosensory feedback systems. In contrast, motor commands in the second language are less familiar, and auditory feedback frequently mismatches internal representations, potentially requiring greater involvement of sensory feedback control. Furthermore, the accent problem commonly found in late bilinguals essentially reflects that motor-sensory information integration in the second language is difficult to achieve native-like levels. Therefore, we should conduct empirical research to reveal general principles of second language speech motor control, providing theoretical guidance for promoting second language pronunciation teaching.

Fifth, exploring the neural mechanisms of speech motor control. Early explorations of speech motor systems and cognitive mechanisms primarily came from acoustic research. In recent years, researchers have attempted to use brain imaging techniques to investigate the temporal course and neural basis of auditory processing and motor correction. However, the greatest challenge in overt naming tasks is movement-related artifacts caused by vocalization, which researchers believe may mask neural activity changes induced by experimental manipulations. In light of this, some researchers have required participants to minimize articulatory movements during speech production (Chen et al., 2012b), adopted strict data rejection criteria (Scheerer et al., 2013a, 2013b), or used silent naming tasks (Tian & Poeppel, 2015). Future research should focus on how to combine behavioral, ERP, MEG, and fMRI techniques with multiple experimental paradigms to effectively investigate the cognitive and neural mechanisms underlying the integration of feedforward and feedback control.

References

Zhang, Q., & Yang, Y. (2003). Theories of lexical access in speech production. *Advances in Psychological Science*, *11*(1), 6–11.

Alsius, A., Mitsuya, T., & Munhall, K. (2013). Does compensation in auditory feedback require attention? *Journal of the Acoustical Society of America*, *133*(5), 3342.

Ballard, K. J., Halaki, M., Sowman, P. F., Kha, A., Daliri, A., Robin, D., ⋯Guenther, F. (2018). An investigation of compensation and adaptation to auditory perturbations in individuals with acquired apraxia of speech. *Frontiers in Human Neuroscience*, *12*, 510.

Ballard, K. J., Tourville, J., & Robin, D. A. (2014). Behavioural, computational, and neuroimaging studies of acquired apraxia of speech. *Frontiers in Human Neuroscience*, *8*, 892.

Bauer, J. J., Mittal, J., Larson, C. R., & Hain, T. C. (2006). Vocal responses to unanticipated perturbations in voice loudness feedback: An automatic mechanism for stabilizing voice amplitude. *Journal of the Acoustical Society of America*, *119*(4), 2363–2371.

Behroozmand, R., Ibrahim, N., Korzyukov, O., Robin, D. A., & Larson, C. R. (2015). Functional role of delta and theta band oscillations for auditory feedback processing during vocal pitch motor control. *Frontiers in Neuroscience*, *9*, 109.

Behroozmand, R., & Larson, C. (2011). Error-dependent modulation of speech-induced auditory suppression for pitch-shifted voice feedback. *BMC Neuroscience*, *12*(1), 54–63.

Behroozmand, R., Sangtian, S., Korzyukov, O., & Larson, C. R. (2016). A temporal predictive code for voice motor control: Evidence from ERP and behavioral responses to pitch-shifted auditory feedback. *Brain Research*, *1636*, 1–12.

Bohland, J. W., & Guenther, F. H. (2006). An fMRI investigation of syllable sequence production. *NeuroImage*, *32*(2), 821–841.

Cai, S. (2012). *Online control of articulation based on auditory feedback in normal speech and stuttering: Behavioral and modeling studies* (Unpublished doctoral dissertation). Massachusetts Institute of Technology, Cambridge.

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

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

Cai, S., Ghosh, S. S., Guenther, F. H., & Perkell, J. S. (2011). Focal manipulations of formant trajectories reveal a role of auditory feedback in the online control of both within-syllable and between-syllable speech timing. *Journal of Neuroscience*, *31*(45), 16483–16490.

Cavanagh, J. F., & Frank, M. J. (2014). Frontal theta as a mechanism for cognitive control. *Trends in Cognitive Sciences*, *18*(8), 414–421.

Chang, E. F., Niziolek, C. A., Knight, R. T., Nagarajan, S. S., & Houde, J. F. (2013). Human cortical sensorimotor network underlying feedback control of vocal pitch. *Proceedings of the National Academy of Sciences*, *110*(7), 2657-2662.

Chen, Z., Chen, X., Liu, P., Huang, D., & Liu, H. (2012a). Effect of temporal predictability on the neural processing of self-triggered auditory stimulation during vocalization. *BMC Neuroscience*, *13*(1), 1–10.

Chen, Z., Liu, P., Jones, J. A., Huang, D., & Liu, H. (2010). Sex-related differences in vocal responses to pitch feedback perturbations during sustained vocalization. *Journal of the Acoustical Society of America*, *128*(6), EL355-EL360.

Chen, Z., Liu, P., Wang, E. Q., Larson, C. R., Huang, D., & Liu, H. (2012b). ERP correlates of language-specific processing of auditory pitch feedback during self-vocalization. *Brain and Language*, *121*(1), 25–34.

Chen, Z., Wong, F. C. K., Jones, J. A., Li, W., Liu, P., Chen, X., et al. (2015). Transfer effect of speech-sound learning on auditory-motor processing of perceived vocal pitch errors. *Scientific Reports*, *5*, 13134.

Civier, O., Tasko, S. M., & Guenther, F. H. (2010). Overreliance on auditory feedback may lead to sound/syllable repetitions: Simulations of stuttering and fluency-inducing conditions with a neural model of speech production. *Journal of Fluency Disorders*, *35*(3), 246–279.

Cowie, R., Douglas-Cowie, E., & Kerr, A. G. (1982). A study of speech deterioration in post-lingually deafened adults. *The Journal of Laryngology & Otology*, *96*(02), 101–112.

Cruikshank, L. C., Singhal, A., Hueppelsheuser, M., & Caplan, J. B. (2012). Theta oscillations reflect a putative neural mechanism for human sensorimotor integration. *Journal of Neurophysiology*, *107*(1), 65–77.

Daliri, A., & Max, L. (2015a). Electrophysiological evidence for a general auditory prediction deficit in adults who stutter. *Brain and Language*, *150*, 37-44.

Daliri, A., & Max, L. (2015b). Modulation of auditory processing during speech movement planning is limited in adults who stutter. *Brain and Language*, *143*, 59–68.

Daliri, A., Wieland, E. A., Cai, S., Guenther, F. H., & Chang, S.-E. (2017). Auditory-motor adaptation is reduced in adults who stutter but not in children who stutter. *Developmental Science*, *21*(2), e12521.

Delvaux, V., & Soquet, A. (2007). The influence of ambient speech on adult speech productions through unintentional imitation. *Phonetica*, *64*(2-3), 145-

173.

Franken, M. K., Acheson, D. J., Mcqueen, J. M., Peter, H., & Frank, E. (2018a). Opposing and following responses in sensorimotor speech control: Why responses go both ways. *Psychonomic Bulletin & Review*, *25*(4), 1458-1467.

Franken, M. K., Frank, E., Acheson, D. J., Mcqueen, J. M., Peter, H., & Jan-Mathijs, S. (2018b). Self-monitoring in the cerebral cortex: Neural responses to small pitch shifts in auditory feedback during speech production. *NeuroImage*, *179*, 326-336.

Franklin, D. W., & Wolpert, D. M. (2011). Computational mechanisms of sensorimotor control. *Neuron*, *72*(3), 425-442.

Fu, C. H., Vythelingum, G. N., Brammer, M. J., Williams, S. C., Amaro, E., Jr., Andrew, C. M., et al. (2006). An fMRI study of verbal self-monitoring: Neural correlates of auditory verbal feedback. *Cerebral Cortex*, *16*(7), 969-977.

Golfinopoulos, E., Tourville, J. A., & Guenther, F. H. (2010). The integration of large-scale neural network modeling and functional brain imaging in speech motor control. *NeuroImage*, *52*(3), 862-874.

Golfinopoulos, E., Tourville, J. A., Bohland, J. W., Ghosh, S. S., Nieto-Castanon, A., & Guenther, F. H. (2011). fMRI investigation of unexpected somatosensory feedback perturbation during speech. *NeuroImage*, *55*(3), 1324-1338.

Gould, J., Lane, H., Vick, J., Perkell, J. S., Matthies, M. L., & Zandipour, M. (2001). Changes in speech intelligibility of postlingually deaf adults after cochlear implantation. *Ear and Hearing*, *22*(6), 453-460.

Grafton, S. T., Schmitt, P., Van Horn, J., & Diedrichsen, J. (2008). Neural substrates of visuomotor learning based on improved feedback and prediction. *NeuroImage*, *39*(3), 1383-1395.

Guenther, F. H. (1994). A neural network model of speech acquisition and motor equivalent speech production. *Biological Cybernetics*, *72*(1), 43-53.

Guenther, F. H. (1995). Speech sound acquisition, coarticulation, and rate effects in a neural network model of speech production. *Psychological Review*, *102*(3), 594-621.

Guenther, F. (2006). Cortical interactions underlying the production of speech sounds. *Journal of Communication Disorders*, *39*(5), 350-365.

Guenther, F. H. (2016). *Neural control of speech.* Cambridge, MA: MIT Press.

Guenther, F. H., Ghosh, S. S., & Tourville, J. A. (2006). Neural modeling and imaging of the cortical interactions underlying syllable production. *Brain and Language*, *96*(3), 280-301.

Guenther, F., & Vladusich, T. (2012). A neural theory of speech acquisition and production. *Journal of Neurolinguistics*, *25*(5), 408-422.

Heinks-Maldonado, T. H., Mathalon, D. H., Gray, M., & Ford, J. M. (2005). Fine-tuning of auditory cortex during speech production. *Psychophysiology*, *42*(2), 180-190.

Heinks-Maldonado, T. H., Nagarajan, S. S., & Houde, J. F. (2006). Magnetoencephalographic evidence for a precise forward model in speech production. *Neuroreport*, *17*(13), 1375-1379.

Hickok, G. (2012). Computational neuroanatomy of speech production. *Nature Reviews Neuroscience*, *13*(2), 135-145.

Hickok, G., Houde, J., & Rong, F. (2011). Sensorimotor integration in speech processing: Computational basis and neural organization. *Neuron*, *69*(3), 407-422.

Hickok, G., Okada, K., & Serences, J. T. (2009). Area Spt in the human planum temporale supports sensory-motor integration for speech processing. *Journal of Neurophysiology*, *101*(5), 2725-2732.

Houde, J. F., & Chang, E. F. (2015). The cortical computations underlying feedback control in vocal production. *Current Opinion in Neurobiology*, *33*, 174-181.

Houde, J. F., & Nagarajan, S. S. (2011). Speech production as state feedback control. *Frontiers in Human Neuroscience*, *5*, 82.

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

Indefrey, P. (2011). The spatio and temporal signatures of word production components: A critical update. *Frontiers in Psychology*, *2*, 255.

Indefrey, P., & Levelt, W. J. M. (2004). The spatial and temporal signatures of word production components. *Cognition*, *92*(1-2), 101-144.

Ito, M. (2000). Mechanisms of motor learning in the cerebellum. *Brain Research*, *886*(1-2), 237-245.

Iuzzini-Seigel, J., Hogan, T. P., Guarino, A. J., & Green, J. R. (2015). Reliance on auditory feedback in children with childhood apraxia of speech. *Journal of Communication Disorders*, *54*, 32-42.

Jones, J. A., & Keough, D. (2008). Auditory-motor mapping for pitch control in singers and nonsingers. *Experimental Brain Research*, *190*(3), 279-287.

Jones, J. A., Scheerer, N., & Tumber, A. (2013). The relationship between vocal pitch feedback error and event-related brain potentials. *Proceedings of Meetings on Acoustics*, *19*, 060151.

Kakimoto, A., Ito, S., Okada, H., Nishizawa, S., Minoshima, S., & Ouchi, Y. (2016). Age-related sex-specific changes in brain metabolism and morphology. *Journal of Nuclear Medicine*, *57*(2), 221-225.

Kalpouzos, G., & Nyberg, L. (2010). Asymmetry of Memory in the Brain. In K. Hugdahl & R. Westerhausen (Eds.), *The two halves of the brain: Information processing in the cerebral hemispheres* (pp. 499-530). MIT Press.

Kearney, E., & Guenther, F. H. (2019). Articulating: The neural mechanisms of speech production. *Language, Cognition and Neuroscience, 1-16.*

Kearney, E., Nieto-Castoñón, A., Weerathunge, H.R., Falsini, R., Daliri, A., Abur, D., ···Guenther, F.H. (submitted). A simple 3-parameter model for examining adaptation in speech and voice production.

Keough, D., Hawco, C., & Jones, J. A. (2013). Auditory-motor adaptation to frequency-altered auditory feedback occurs when participants ignore feedback. *BMC Neuroscience*, *14*(1), 25-35.

Kort, N. S., Nagarajan, S. S., & Houde, J. F. (2014). A bilateral cortical network responds to pitch perturbations in speech feedback. *NeuroImage*, *86*, 525-535.

Korzyukov, O., Karvelis, L., Behroozmand, R., & Larson, C. R. (2012a). ERP correlates of auditory processing during automatic correction of unexpected perturbations in voice auditory feedback. *International Journal of Psychophysiology*, *83*(1), 71-78.

Korzyukov, O., Sattler, L., Behroozmand, R., & Larson, C. R. (2012b). Neuronal mechanisms of voice control are affected by implicit expectancy of externally triggered perturbations in auditory feedback. *PLoS One*, *7*(7), e41216.

Lametti, D. R., Krol, S. A., Shiller, D. M., & Ostry, D. J. (2014). Brief periods of auditory perceptual training can determine the sensory targets of speech motor learning. *Psychological Science*, *25*(7), 1325-1336.

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

Lane, H., & Webster, J. W. (1991). Speech deterioration in postlingually deafened adults. *Journal of the Acoustical Society of America*, *89*(2), 859-866.

Levelt, W. J. M., Roelofs, A., & Meyer, A.S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, *22*(1), 1-75.

Li, J., Hu, H., Chen, N., Jones, J. A., Wu, D., Liu, P., & Liu, H. (2018). Aging and sex influence cortical auditory-motor integration for speech control. *Frontiers in Neuroscience*, *12*, 749.

Liu, H., Meshman, M., Behroozmand, R., & Larson, C. R. (2011). Differential effects of perturbation direction and magnitude on the neural processing of voice pitch feedback. *Clinical Neurophysiology*, *122*(5), 951-957.

Liu, H., Russo, N., & Larson, C. R. (2010a). Age-related differences in vocal responses to pitch feedback perturbations: A preliminary study. *Journal of the Acoustical Society of America*, *127*(2), 1042-1046.

Liu, H., Wang, E. Q., Chen, Z., Liu, P., Larson, C. R., & Huang, D. (2010b). Effect of tonal native language on voice fundamental frequency responses to pitch feedback perturbations during vocalization. *Journal of the Acoustical Society of America*, *128*(6), 3739–3746.

Liu, X., & Tian, X. (2018). The functional relations among motor-based prediction, sensory goals and feedback in learning non-native speech sounds: Evidence from adult Mandarin Chinese speakers with an auditory feedback masking paradigm. *Scientific Reports*, *8*(1), 11910.

Liu, Y., Hu, H., Jones, J., Guo, Z., Li, W., Chen, X., et al. (2015). Selective and divided attention modulates auditory-vocal integration in the processing of pitch feedback errors. *European Journal of Neuroscience*, *42*(3), 1895–1905.

Maas, E., Mailend, M. L., & Guenther, F. H. (2015). Feedforward and feedback control in apraxia of speech: Effects of noise masking on vowel production. *Journal of Speech, Language, and Hearing Research*, *58*(2), 185–196.

Mitsuya, T., Munhall, K. G., & Purcell, D. W. (2017). Modulation of auditory-motor learning in response to formant perturbation as a function of delayed auditory feedback. *Journal of the Acoustical Society of America*, *141*(4), 3023.

Munhall, K. G., Macdonald, E. N., Byrne, S. K., & Johnsrude, I. (2009). Talkers alter vowel production in response to real-time formant perturbation even when instructed not to compensate. *Journal of the Acoustical Society of America*, *125*(1), 384–390.

New, A. B., Robin, D. A., Parkinson, A. L., Duffy, J. R., McNeil, M. R., Piguet, O., et al. (2015). Altered resting-state network connectivity in stroke patients with and without apraxia of speech. *NeuroImage: Clinical*, *8*, 387–395.

Ning, L.-H., Loucks, T. M., & Shih, C. (2015). The effects of language learning and vocal training on sensorimotor control of lexical tone. *Journal of Phonetics*, *51*, 50–69.

Ning, L.-H., Shih, C., & Loucks, T. M. (2014). Mandarin tone learning in L2 adults: A test of perceptual and sensorimotor contributions. *Speech Communication*, *63-64*, 55–69.

Niziolek, C. A., Nagarajan, S. S., & Houde, J. F. (2013). What does motor efference copy represent? Evidence from speech production. *Journal of Neuroscience*, *33*(41), 16110–16116.

Oller, D. K., & Eilers, R. E. (1988). The role of audition in infant babbling. *Child Development*, *59*(2), 441–449.

O' Reilly, J. X., Mesulam, M. M., & Nobre, A. C. (2008). The cerebellum predicts the timing of perceptual events. *Journal of Neuroscience*, *28*(9), 2252–2260.

Parkinson, A. L., Flagmeier, S. G., Manes, J. L., Larson, C. R., Rogers, B., & Robin, D. A. (2012). Understanding the neural mechanisms involved in sensory control of voice production. *NeuroImage*, *61*(1), 314-322.

Parrell, B., Agnew, Z., Nagarajan, S., Houde, J., & Ivry, R. B. (2017). Impaired feedforward control and enhanced feedback control of speech in patients with cerebellar degeneration. *The Journal of Neuroscience*, *37*(38), 9249-9258.

Parrell, B., Lammert, A. C., Ciccarelli, G., & Quatieri, T. F. (2019). Current models of speech motor control: A control-theoretic overview of architectures and properties. *Journal of the Acoustical Society of America*, *145*(3), 1456-1481.

Patel, R., Reilly, K. J., Archibald, E., Cai, S., & Guenther, F. H. (2015). Responses to intensity-shifted auditory feedback during running speech. *Journal of Speech Language and Hearing Research*, *58*(6), 1687-1694.

Perkell, J. S. (2012). Movement goals and feedback and feedforward control mechanisms in speech production. *Journal of Neurolinguistics*, *25*(5), 382-407.

Reilly, K. J., & Pettibone, C. (2017). Vowel generalization and its relation to adaptation during perturbations of auditory feedback. *Journal of Neurophysiology*, *118*(5), 2925-2934.

Saltzman, E., & Munhall, K. (1989). A dynamical approach to gestural patterning in speech production. *Ecological Psychology*, *1*(4), 333-382.

Scheerer, N. E., Behich, J., Liu, H., & Jones, J. A. (2013a). ERP correlates of the magnitude of pitch errors detected in the human voice. *Neuroscience*, *240*, 176-185.

Scheerer, N. E., & Jones, J. A. (2012). The relationship between vocal accuracy and variability to the level of compensation to altered auditory feedback. *Neuroscience Letters*, *529*(2), 128-132.

Scheerer, N. E., Liu, H., & Jones, J. A. (2013b). The developmental trajectory of vocal and event-related potential responses to frequency-altered auditory feedback. *European Journal of Neuroscience*, *38*(8), 3189-3200.

Scheerer, N. E., & Jones, J. A. (2014). The predictability of frequency-altered auditory feedback changes the weighting of feedback and feedforward input for speech motor control. *European Journal of Neuroscience*, *40*(12), 3793-3806.

Scheerer, N. E., & Jones, J. A. (2018). Detecting our own vocal errors: An event-related study of the thresholds for perceiving and compensating for vocal pitch errors. *Neuropsychologia*, *114*, 158-167.

Simmonds, A. J., Wise, R. J., & Leech, R. (2011). Two tongues, one brain: Imaging bilingual speech production. *Frontiers in Psychology*, *2*, 166.

Swink, S., & Stuart, A. (2012). The effect of gender on the N1-P2 auditory complex while listening and speaking with altered auditory feedback. *Brain and Language*, *122*(1), 25-33.

Terband, H., Rodd, J., & Maas, E. (2015). Simulations of feedforward and feedback control in apraxia of speech (AOS): Effects of noise masking on vowel production in the DIVA model. *The 18th International Congress of Phonetic Sciences*.

Tian, X., & Poeppel, D. (2010). Mental imagery of speech and movement implicates the dynamics of internal forward models. *Frontiers in Psychology*, *1*, 166.

Tian, X., & Poeppel, D. (2012). Mental imagery of speech: Linking motor and perceptual systems through internal simulation and estimation. *Frontiers in Human Neuroscience*, *6*, 314.

Tian, X., & Poeppel, D. (2015). Dynamics of Self-monitoring and Error Detection in Speech Production: Evidence from Mental Imagery and MEG. *Journal of Cognitive Neuroscience*, *27*(2), 352-364.

Tian, X., Zarate, J. M., & Poeppel, D. (2016). Mental imagery of speech implicates two mechanisms of perceptual reactivation. *CORTEX*, *77*(7), 1-12.

Tourville, J. A., & Guenther, F. H. (2011). The DIVA model: A neural theory of speech acquisition and production. *Language and Cognitive Processes*, *26*(7), 952-981.

Tourville, J. A., Reilly, K. J., & Guenther, F. H. (2008). Neural mechanisms underlying auditory feedback control of speech. *NeuroImage*, *39*(3), 1429-1443.

Toyomura, A., Koyama, S., Miyamaoto, T., Terao, A., Omori, T., Murohashi, H., et al. (2007). Neural correlates of auditory feedback control in human. *Neuroscience*, *146*(2), 499-503.

Tumber, A. K., Scheerer, N. E., & Jones, J. A. (2014). Attentional demands influence vocal compensations to pitch errors heard in auditory feedback. *PLoS ONE*, *9*(10), e109968.

Tye-Murray, N., & Spencer, L. (1995). Acquisition of speech by children who have prolonged cochlear implant experience. *Journal of Speech and Hearing Research*, *38*(2), 327-337.

Wolpert, D. M., Diedrichsen, J., & Flanagan, J. R. (2011). Principles of sensorimotor learning. *Nature Reviews Neuroscience*, *12*(12), 739-751.

Zheng, Z. Z., Munhall, K. G., & Johnsrude, I. S. (2010). Functional overlap between regions involved in speech perception and in monitoring one's own voice during speech production. *Journal of Cognitive Neuroscience*, *22*(8), 1770-1781.

Note: Figure translations are in progress. See original paper for figures.

Source: ChinaXiv —Machine translation. Verify with original.