eNeuro

Research Article: New Research | Cognition and Behavior

The generic inhibitory function of corollary discharge in motor intention: evidence from the modulation effects of speech preparation on the late components of auditory neural responses

https://doi.org/10.1523/ENEURO.0309-22.2022

Cite as: eNeuro 2022; 10.1523/ENEURO.0309-22.2022

Received: 31 July 2022
Revised: 3 November 2022
Accepted: 14 November 2022

The generic inhibitory function of corollary discharge in motor intention: evidence from the modulation effects of speech preparation on the late components of auditory neural responses

Abbreviated Title: generic inhibition of corollary discharge

Xiaodan Zheng[1,2], Hao Zhu[2,3], Siqi Li[1,2] & Xing Tian[1,2,3, #]

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

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

[3]Division of Arts and Sciences, New York University Shanghai, China, 200122

Author Contributions: Zheng designed research, performed research, analyzed data, wrote the first draft of the paper, edited the paper and wrote the paper; Zhu and Li analyzed data and wrote the paper; Tian designed research, edited the paper and wrote the paper.

Correspondence to:
Xing Tian
Email: xing.tian@nyu.edu

Number of pages: 46; Number of figures: 6; Number of words for Abstract: 248
Number of words for Introduction: 646; Number of words for Significance Statement: 119; Number of words for Discussion: 1672
Acknowledgments: This study was supported by the National Natural Science Foundation of China 32071099, Natural Science Foundation of Shanghai 20ZR1472100, Program of Introducing Talents of Discipline to Universities, Base B16018, and NYU Shanghai Boost Fund.

Conflict of interest: The authors declare no competing financial interests
Funding sources: This study was supported by the National Natural Science Foundation of China 32071099, Natural Science Foundation of Shanghai 20ZR1472100, Program of Introducing Talents of Discipline to Universities, Base B16018, and NYU Shanghai Boost Fund.

Abstract

The importance of action-perception loops necessitates efficient computations linking motor and sensory systems. Corollary discharge (CD), a concept in motor-to-sensory transformation, has been proposed to predict the sensory consequences of actions for efficient motor and cognitive control. The predictive computation has been assumed to realize via inhibiting sensory reafference when actions are executed. Continuous control throughout the course of action demands inhibitory function ubiquitously on all potential reafference when sensory consequences are not available prior to execution. However, the temporal and functional characteristics of CD are unclear -- When does CD begin to operate? To what extent does CD inhibit sensory processes? How is the inhibitory function implemented in neural computation? Using a delayed articulation paradigm with three types of auditory probes (speech, non-speech, and non-human sounds) in an electroencephalography (EEG) experiment with 20 human participants (7 male), we found that preparing to speak without knowing what to say (general preparation) suppressed neural responses to each type of auditory probe, suggesting a generic inhibitory function of CD in motor intention. Moreover, power and phase coherence in low-frequency bands (1-8 Hz) were both suppressed, indicating that inhibition was mediated by dampening response amplitude and adding temporal variance to sensory processes. Furthermore, inhibition was stronger for sounds that humans can produce than non-human sounds, hinting that the generic inhibitory function of CD is regulated by the established motor-sensory associations. These

results suggest a functional and temporal granularity of corollary discharge that mediates multifaceted computations in motor and cognitive control.

keywords: sensorimotor integration, motor control, action-induced sensory suppression, internal forward model, agency

Significance Statement

The feeling and actual control of one's body are linked to the same phenomenon of sensorimotor interaction -- sensory processes of self-induced stimuli are attenuated by a copy of motor signals, coined as corollary discharge (*CD*). However, when, to what extent, and how *CD* inhibits sensory processes remain unclear. Using a delayed articulation paradigm in an EEG experiment, we found that *CD* inhibited all speech, non-speech and non-human sounds even when participants intended to speak, with stronger inhibition of the sounds that humans can produce. The inhibition was mediated by dampening response amplitude and adding temporal variance in low-frequency neural responses to sensory stimuli. These results suggest functional granularity of *CD* throughout the course of actions for motor control.

Introduction

The efficient interplay of action and perception is an adaptive trait in any organism for

survival. The importance manifests through evolution and engraves dedicated neural computational pathways linking motor and sensory systems (Crapse and Sommer, 2008). One of such functional computations has been theorized as the internal forward model (von Helmholtz, 1910; Wolpert and Ghahramani, 2000) -- a copy of motor signals, coined as 'corollary discharge' (CD) (Sperry, 1950) or 'efference copy' (EC) (von Holst and Mittelstaedt, 1950), transmits to sensory systems to predict the sensory consequences of actions (Kawato, 1999; Schubotz, 2007). Such predictive functions of the internal forward model have been implied as canonical computations mediating visual perception (Ross et al., 2001; Sommer and Wurtz, 2006), motor control (Miall and Wolpert, 1996), speech production (Guenther, 1995; Houde and Nagarajan, 2011; Hickok, 2012), and higher-order cognitive functions such as mental imagery and agency (Desmurget et al., 2009; Tian and Poeppel, 2010; Kilteni et al., 2018).

The operation of the internal forward model has been assumed to rely on the inhibitory modulation of the CD and EC on sensory processing (Blakemore and Decety, 2001; Houde et al., 2002; Tian et al., 2018) (but also see exceptions of enhancement modulation in recent empirical and theoretical studies (Li et al., 2020; Press et al., 2022)). Recently, an updated theoretical framework has been proposed by considering distinct modulatory functions of CD and EC throughout the time course of actions (Li et al., 2020). Specifically, EC is available after motor encoding and includes detailed action codes that selectively enhance the processing sensitivity of the sensory reafference. Whereas, CD exerts an inhibitory function and is available throughout the

course of actions (Fig. 1). CD does not depend on specific information and is available as early as in motor intention to inhibit sensory consequences caused by all possible actions that an agent can perform – the generic inhibitory function of CD.

The early onset of CD has been supported by empirical results, but the generic inhibition is equivocal. In motor intention when participants prepared to speak but did not know what to say (general preparation), CD was generated in this earliest stage of actions and suppressed the neural responses to auditory syllables but not pure tones (Li et al., 2020). The mixed results could be because the CD induced in that study was not 'general' enough -- participants were only asked to pronounce syllables in subsequent articulation tasks, and hence the CD in general preparation could contain categorical information. Moreover, the generic inhibition of CD may be constrained by the distance between sensory feedback and the possible sensory consequences caused by the repertoire of actions that an agent can perform. Pure tones only partially overlap with features of the tones that humans articulate and hence could be less inhibited than sounds that humans normally produce (shorter blue bar for non-human sound in Fig. 1). A recent study found that the strength of suppression to auditory responses decreased as the frequency of tones deviated from the standard frequency of action consequence (Schneider et al., 2018). This evidence offers hints supporting our conjecture of the gradient suppression effects.

How CD exerts the inhibitory function is also unclear. Auditory processes can

operate in temporal or rate codes (Lu et al., 2001). The modulation effects can be manifested by altering the magnitude or temporal aspects of responses (Grill-Spector et al., 2006). Specifically, the effects can be a result of direct gain modulation on response magnitude. Numerous studies have demonstrated that manual actions and speaking dampen the amplitude of neural responses to sounds (Houde et al., 2002; Baess et al., 2011). Whereas in the temporal dimension, it has been suggested that the phase of neural oscillations can be reset and aligned with upcoming external stimuli to boost the sensitivity of neural encoding (Schroeder and Lakatos, 2009; Giraud and Poeppel, 2012; Tomassini et al., 2017; Teng et al., 2020). If CD influences the alignment between neural phase and auditory stimuli, similar suppression effects can be achieved. Therefore, the generic inhibition of CD can potentially dampen response power or increase temporal variance in responses to sensory feedback.

To examine the hypothesis of generic inhibitory function and neural mechanisms of CD (Li et al., 2020), we adopted the delayed articulation paradigm and excluded categorical information from CD in general preparation by asking participants to produce three types of sounds in subsequent articulation task — a speech sound of syllable /ba/, a non-speech sound of cough, and a humming tone that simulated a non-human sound of pure tone. According to the hypothesis of generic inhibition, general preparation would suppress the neural responses to all types of auditory probes, but less to pure tone. Moreover, the time-frequency analysis would reveal whether the inhibitory function was realized by dampening response amplitude or increasing

temporal variance in sensory processes.

Materials and Methods

Participants

Twenty right-handed volunteers (7 males; aged 19 - 25 years; *Mean* = 22.2) participated in the experiment. The sample size was determined as the same number of participants in the target comparison study that used similar paradigms (Li et al., 2020). All participants had normal hearing (self-reported). They received monetary compensation for their participation. Written informed consent was obtained from every participant before the experiment. This study was approved by the institutional review board at New York University Shanghai.

The sample size was predetermined to be 20 based on previous studies that investigated similar questions of action-induced suppression (Houde et al., 2002; Aliu et al., 2009; Horváth et al., 2012). Using G*power (Faul et al., 2007) to estimate the sample size based on the effect size (d=0.8660) observed in Houde et al. (2002), we found a sample size that was required to have 80% power at an alpha level of 0.05 was 13. Therefore, our sample size is large enough to replicate the action-induced suppression effect. We further calculated the statistical power of the present study using G*power to verify that we had enough power. We found that the present study had 94.84% power with a sample size of 20 at an alpha level of 0.05 based on the effect size (0.847) of the present EEG data.

Materials

Three auditory tokens, each in every sound category – speech sound (a syllable /ba/), non-speech sound (a cough sound), and non-human sound (500 Hz pure tone) were used as auditory probes in the experiment. All stimuli were 400 ms in duration with a sampling frequency of 44.1k Hz and their average (root-mean-square) intensity was normalized to 70 dB SPL using Praat. The auditory syllable (/ba/) was synthesized using the Neospeech web engine (www.neospeech.com) in a male voice, identical to the one used in the target comparison study (Li, Zhu, &Tian, 2020). The cough sound was recorded by a male native Mandarin speaker. The 500 Hz pure tone was generated using MATLAB. The frequency of the tone was selected by considering the usual lower bound of audiometry using pure tones as well as the range of fundamental frequency of human vocal production. The pure tone was included so that we could investigate whether the modulation of CD on non-human sounds differs from human sounds.

Procedures

We first summarize the procedure and its major differences from the target comparison study and then provide details next. The delayed-articulation paradigm was used in the experiment. Participants were required to make a general preparation – preparing to speak in the subsequent articulation task but did not know what to say. We asked participants to produce three types of sounds in the articulation task (syllable, cough, and humming tone). In this case, the general preparation could be truly 'general' – not

constrained by a particular speech category but possibly extending to all sound categories that humans can produce, and hence our hypothesis about the generic inhibitory function of CD can be tested. The auditory probes that were presented during the preparation stage also included the three types of sounds to probe the modulation function of CD during general preparation.

The detailed procedures are as follows. To examine the hypothesis and control confounding variables, four types of trials were included in the experiment: general preparation (*GP*) trials, general preparation with no sound (GP_{NS}) trials, no preparation (*NP*) trials, and passive listening (*PL*) trials. Figure 2A shows examples of four types of trials. A *GP* trial began with a fixation displayed for 500 ms, followed by a general preparation stage with a duration randomly ranging from 1500 ms to 2000 ms with an increment of 100 ms. The general preparation stage was cued by two yellow symbols (#%) presented in the center of the screen. Participants prepared to produce sounds but the symbols did not provide any information about what sound to produce. During the last 400 ms of the general preparation stage, one of the three auditory stimuli (auditory syllable, cough, or pure tone) was presented. After the general preparation stage and a blank period (randomized in a range from 600 to 800 ms), participants were asked to articulate a sound as quickly and accurately as possible according to a visual cue in green that appeared in the center of the screen. Three visual cues, each composed of two green symbols, indicated the sound to produce – visual characters of 'ba' for speaking the syllable /ba/, '<~' for producing cough sound, and '--' cued participants to hum the

first lexical tone (flat tone) in Mandarin Chinese. The reaction time (RT) of the articulation in each trial was recorded as the time interval between the onset of the green visual cue and the onset of participants' vocal responses.

GP_{NS} trials were similar to *GP* trials, except that no sound was presented in the last 400ms of the general preparation stage. The GP_{NS} trials were included in the experiment to ensure preparation in the general preparation stage was independent of auditory probes in *GP* trials. That is, the preparation should occur after the preparatory visual cue and be available during the presentation of the auditory probe in *GP* trials. In the *NP* trials, participants were asked to perform the articulation tasks without any preparation. By comparing RTs in *NP* trials with those in the *GP* trials or GP_{NS} trials, we could quantify general preparation in the *GP* trials or GP_{NS} trials behaviorally. Similarly, comparing RTs in the GP_{NS} trials with those in the *GP* trials could infer whether general preparation occurred independently of the auditory probe.

The *PL* trials were marked by two blue symbols '**'. Similar to the general preparation stage in the *GP* trials, the visual cue was also displayed in a duration randomly selected from 1500 ms to 2000 ms in an increment of 100 ms. An auditory probe was presented during the last 400ms of visual cue presentation. The auditory probes played in the *PL* trials were the same as those in the *GP* trials. However, In the *PL* trials, participants listened to the auditory probes passively without any preparation or articulation task. By comparing EEG responses to the auditory probes in the *PL* trials with those in the *GP* trials, we could examine whether CD generated in the general

preparation stage modulates early auditory responses to the auditory probes.

In summary, a within-subject design with four types of trials (*GP*, GP_{NS}, *NP*, and *PL*) was used in this study. Three auditory probes (syllable, cough, and pure tone) were in the trials of *GP* and *PL*, yielding six conditions in EEG responses. The experiment consisted of six blocks. Each block included 96 trials, with 24 trials for each type of trial. The number for each of the auditory probes was equal and yielded 48 trials separately for the stimulus of syllable, cough, and tone in *GP* and *PL*. The order of trials was randomized. A short break of 1 to 2 minutes was provided between blocks.

Behavioral data analysis

To evaluate the effect of general preparation behaviorally, articulation RTs of the articulation task, the time interval between the onset of the green visual cue and the onset of the vocalization, were compared across different conditions using one-way repeated measures ANOVA to assess the differences among three conditions (*GP*, GP_{NS}, and *NP*). Post-hoc t-tests with Bonferroni correction were carried out for pairwise comparison between conditions using the Pingouin toolbox (Vallat, 2018).

EEG data acquisition and preprocessing

EEG signals were recorded using a 32-channel Brain Products actiCHamp recording system. The 32 electrodes over the scalp were placed based on the 10/20 international electrode system. To monitor ocular activity, the EOG was recorded from two additional electrodes, one placed on 1 cm lateral to the lateral canthus of the left eye, and the other below the right eye. The electrode impedances were kept under 10 kΩ.

The electrode of Cz was used as the online reference. An online low-pass filter with a cutoff at 200 Hz and a notch filter at 50Hz were used. The EEG data were digitized with a sampling frequency of 1000 Hz.

EEG data preprocessing was performed using MNE-python (Gramfort et al., 2014). The continuous EEG data were bandpass filtered (0.1-30Hz). Bad channels were identified visually and repaired using spherical spline interpolation (Perrin et al., 1989). Epochs spanning from -100 to 300 ms related to the onset of the auditory probe were extracted in each trial of *GP* and *PL*. Baseline correction was applied using the 100 ms pre-stimulus period. Epochs with maximum peak-to-peak amplitude exceeding 100μV on any channel were rejected. Epochs contaminated by eyeblink and movement artifacts were rejected manually. The average rejection rate was 20.85%. The EEG data were re-referenced to the average of all electrodes over scalp.

Temporal domain analysis

Event-related potentials (ERPs) were calculated by averaging epochs for each auditory probe and each participant, as well as for three auditory probes combined, yielding four ERP responses (syllable, cough, tone, and three-sound combined) separately in the *PL* and *GP* conditions. The global field power (GFP), calculated as the standard deviation of the ERP responses across all electrodes (Lehmann and Skrandies, 1980) were derived using the EasyEEG toolbox (Yang et al., 2018). The GFP responses reflect an overall power change in all electrodes across time, which avoid potential subjective

bias in selecting electrodes during analysis. Individual N1 and P2 amplitudes were obtained by averaging the 20 ms responses centered at the peak latency of each component in the GFP waveforms using the TTT toolbox (Wang et al., 2019).

First, to demonstrate the overall inhibitory effects of corollary discharge in general preparation, we carried out a paired *t*-test between the *GP* and *PL* conditions in the three-sound combined GFP responses, separately for N1 and P2 components. Next, to investigate the modulation effects on each type of auditory probe, additional three paired *t*-tests were performed, each on the syllable, cough, and tone GFP responses, separately for N1 and P2 components. To better connect with the literature and provide more intuitive results, we also performed the ERP analyses based on the most common representative channel of ERP auditory responses – Cz.

Spatiotemporal analysis

Because the GFP measure is an omnibus index across all electrodes, its statistical power could be limited by noise or lack of signals in any subset of channels. Furthermore, GFP analysis only provides temporal information about the modulation effects. To increase statistical power as well as to further investigate spatial aspects of the modulation effects, the non-parametric spatiotemporal cluster-based permutation test (Maris and Oostenveld, 2007) was performed using the MNE-python toolbox. For each type of sound, the empirical *t* statistics were first obtained via two-tailed paired *t*-tests on the ERP responses between *GP* and *PL* conditions at each time point between -100 to

300ms time-locked to the sound onset and in each electrode. Time points in each electrode with absolute *t*-values exceeding the threshold (alpha = 0.05) were identified. Selected time points in all electrodes with *t*-values of the same sign (positive or negative) were clustered based on spatiotemporal adjacency. The cluster with maximum points was selected separately for the positive and negative sign *t* values, and the empirical statistics were obtained by calculating the sum of the *t* values within a cluster. The same clustering process was repeated 10,000 times after each time shuffling the condition labels. A null distribution was obtained, separately for the positive and negative sign clusters. The *p*-value of each cluster was determined as the proportion of larger *t* values in the null distribution than the empirical statistics.

Time-frequency analysis

To investigate whether the inhibitory function of corollary discharge modulates the amplitude or the timing of perceptual responses, time-frequency analyses were carried out separately on the aspects of power and phase in several frequency bands. Specifically, longer epochs (-2000 to 2000ms time-locked to auditory probe onset) for each sound in *GP* and *PL* conditions were extracted to avoid edge artifacts. Morlet wavelet transform was applied on each of the longer epochs using the function of 'tfr_morlet' in the MNE-python toolbox with the parameter of n_cycles setting to 2 cycles for each frequency in 1 – 3 Hz and frequency/2 for other frequencies (4 – 28 Hz). Power and phase in each frequency at each time point in each electrode were obtained

for every condition. Data between -100 and 300ms were used for further analysis.

For power analysis, the averaged power between -100 and 0ms was used as the baseline. Power values were normalized by dividing the mean of the baseline and converted into a log scale. For phase analysis, inter-trail phase coherence (ITC) was calculated based on the following equation (Tallon-Baudry et al., 1996; Luo and Poeppel, 2007),

$$ITC(t,f) = \left(\frac{\sum_{j=1}^{N} \cos\theta_j(t,f)}{N} \right)^2 + \left(\frac{\sum_{j=1}^{N} \sin\theta_j(t,f)}{N} \right)^2$$

Power and ITC values were further averaged within the following six frequency bands -- the delta (1–3 Hz), the theta (4– 8 Hz), alpha (9–12 Hz), low-beta (13–16 Hz), mid-beta (17–20 Hz) and high-beta (21–28 Hz) band. The nonparametric spatiotemporal cluster-based permutation test was used to assess the significant difference between *GP* and *PL* conditions for each sound, separately for Power and ITC in each frequency band.

The data and codes in the present study are publicly available on the OSF(https://osf.io/au43q/).

Results

Behavioral results

Participants were asked to produce a sound with or without preceding general preparation. A repeated-measure one-way ANOVA on RTs showed a significant main

effect of preparation ($F(2,38) = 37.45$, $p < 0.0001$, partial $\eta^2 = 0.664$). Bonferroni corrected paired *t*-tests revealed that RTs were faster when participants performed articulation task in *GP* than *NP* ($t(19) = 6.060$, $p < 0.0001$, $d = 0.875$). Moreover, RTs in GP_{NS} was also faster than *NP* ($t(19) = 6.970$, $p < 0.0001$, $d = 0.947$). However, no significant difference was observed between *GP* and GP_{NS} ($t(19) = 0.396$, $p = 1$, $d = 0.028$). These results (Fig2. B) replicated the observations in Li et al. (2020) and indicated that participants engaged in general preparation regardless of the existence of an auditory probe, which suggested that CD was available before sound onset and throughout the general preparation stage.

ERP components results based on all channels revealed overall P2 suppression

ERP responses to all auditory probes combined, including GFP waveforms and topographies of N1 and P2 are shown in Figure 2C. Paired *t*-tests revealed that no significant difference between N1 amplitude in *GP* and *PL* conditions ($t(19) = 0.517$, $p = 0.611$, $d = 0.051$), whereas the P2 amplitude in *GP* was significantly suppressed compared to *PL* ($t(19) = 2.528$, $p = 0.020$, $d = 0.416$). These results supported the hypothesis that CD during motor intention exerted an inhibitory function on auditory neural responses.

To further test the hypothesis of whether CD has a generic inhibitory function and suppresses all sounds that link to articulatory features even without specific articulatory encoding during the motor intention stage, we examined the modulation effects of CD

on each type of auditory probe. Paired *t*-tests on the GFP response amplitude revealed a similar suppression in P2 component in responses to cough ($t(19) = 2.950$, $p = 0.008$, $d = 0.517$), but not in N1 component ($t(19) = 1.147$, $p = 0.266$, $d = 0.181$). However, the suppression effects were absent in responses to the auditory stimuli of the syllable and tone. These null results could be because of relatively weak suppression effects in the responses to different types of sounds and GFP that summarized over all electrodes cannot provide enough statistical power to detect these weak effects. To be comparable with previous studies and offer more straightforward results, we examined the modulation effects based on the most common representative channel of auditory ERP responses – Cz.

Results of ERP analysis based on the channel of Cz revealed P2 suppression in each type of sound

The results of the representative channel Cz are shown in Figure 3. For ERP responses to the auditory probe of syllable (Fig. 3A), paired t-tests revealed that the amplitude of P2 response in *GP* was reduced relative to that in *PL* ($t(19) = 4.533$, $p = 0.0002$, $d = 0.847$). For ERP responses to cough sound (Fig. 3B), the amplitude of P2 response in *GP* was less than that in *PL* ($t(19) = 3.831$, $p = 0.0011$, $d = 0.653$). For ERP responses to the pure tone (Fig. 3C), P2 suppression in *GP* only survived a one-tailed paired t-test rather than two-tailed ($t(19) = 2.017$, $p = 0.0580$, $d = 0.437$). Additionally, the amplitude of early N1 response in *GP* was enhanced relative to that in *PL* for syllable ($t(19) = 3.872$, $p = 0.0010$, $d = 0.473$). For ERP responses to all sounds average (Fig.

3D), the amplitude of P2 response in GP was suppressed relative to that in PL ($t(19)$ = 4.301, p = 0.0004, d = 0.689), consistent with the GFP results. The representative channel analysis revealed inhibition for all types of sounds. To further test the spatial distribution of the effects, we carried out a spatiotemporal cluster analysis by considering the spatial information in addition to the temporal information to further investigate the hypothesis of the generic inhibitory function of CD.

Results of spatiotemporal cluster-based permutation tests

To collaboratively reveal the modulation effects in the aspects of spatial distributions and temporal characteristics, we carried out spatiotemporal cluster-based permutation tests. The results of spatiotemporal cluster analysis are shown in Figure 4, separately for each type of sound. For syllable, three significant clusters were found. The first significant cluster (p = 0.0497) appeared around time 0 ms (with a range from -40 ms to 51 ms, shown in Fig. 4A of the statistical parametric heatmap). The spatial distribution of this cluster was mostly over parietal regions, as shown in the topography of the statistical map in the first row of Fig. 4B. The nature of the modulation effects was further illustrated by examining the raw ERP topographies (averaged amplitudes across the time interval of the cluster) of *PL* and *GP* conditions (shown in the last two rows in Fig. 4B). Responses in the *PL* condition were around zero, which presumably reflected random processes during a passive task before auditory probe onset. Whereas, responses in the *GP* condition were more negative in the posterior electrodes, which were consistent with neural sources that mediated motor intention and preparatory

processes (Desmurget et al., 2009; Tian and Poeppel, 2010). The more negative ERP in *GP*, compared with random activation in *PL*, resulted in a negative sign of statistics, which reflected the enhancement effects of general preparation (more absolute magnitude of activation but in electrodes of negative polarity) in a significant cluster of electrodes over parietal regions before auditory probe onset.

The other two significant clusters observed in responses to syllable were both around 200ms after the auditory probe onset (Fig. 4A). The cluster that had a central spatial layout had negative statistics (p = 0.0038), whereas the one with a peripheral distribution in electrodes over frontal-temporal regions had positive statistics (p = 0.0149) (Fig. 4B). The adjacent distributions of these two clusters resemble the different polarities in the dipole patterns of ERP topographic responses to the auditory syllable (last two rows in Fig. 4B), collaboratively depicting the suppression effects of *GP* on the neural responses of speech sound. Specifically, the cluster with negative statistics distributed over the central electrodes showed positive ERP values in *GP* and *PL* conditions. Responses in *GP* were less positive than in *PL*. The comparison between *GP* with *PL* hence yielded a significant cluster with negative statistics in this central cluster, reflecting the suppression effects of *GP* on responses to the auditory probe. Similarly, the cluster with positive statistics was caused by less negative ERP in *GP* than *PL* in the peripheral frontal-temporal electrodes, reflecting the inhibition of CD on the response magnitude of ERPs to an auditory syllable. That is, the observed two clusters reflect a significantly smaller magnitude of responses to the auditory syllable in

413 *GP* than *PL*, supporting the suppression effects of CD during *GP* on the neural
414 responses of speech sound.

415 For cough, two significant clusters were observed around 200ms (second column
416 in Fig. 4A). Similar to those in syllable, one located in central regions ($p = 0.0274$) and
417 the other was in peripheral frontal-temporal regions ($p = 0.0485$) (Fig. 4B). These two
418 clusters both reflected absolute amplitude decrement in responses to auditory probe in
419 *GP* than *PL*, separately for two sets of electrodes that had ERP responses in opposite
420 polarities in the P2 component (last two rows in Fig. 4B). Specifically, the central
421 cluster with negative statistics was caused by less positive ERP responses in *GP* (*Mean*
422 = 2.798μV) than *PL* (*Mean* = 3.912μV); whereas the peripheral frontal-temporal cluster
423 with positive statistics was caused by less negative ERP responses in *GP* (*Mean* =
424 -1.571μV) than *PL* (*Mean* = -2.311μV). These results suggested that CD in *GP* also
425 induced suppression effects on the responses to non-speech cough sounds.

426 For tone, only one cluster was found (from 78 ms to 225 ms) in peripheral
427 electrodes of frontal-temporal regions (third column in Fig. 4A). This cluster had
428 positive statistics that were caused by less negative ERP responses in GP (*Mean* =
429 -0.525 μV) than PL (*Mean* = -1.359 μV), similar to the one in syllable and cough (Fig.
430 4B). However, this cluster only survived a one-tailed spatiotemporal cluster
431 permutation test but not a two-tailed test ($p = 0.0689$). These results suggested that CD
432 exerted a weak inhibitory effect on the non-human sound of pure tone. Altogether, the
433 results of spatiotemporal cluster analysis suggested that CD suppressed the neural

responses to all types of auditory probes. The strength of CD was stronger for speech and non-speech sounds than for non-human sound, which suggested that the strength of the inhibition effects was constrained by the established motor-sensory associations – the generic inhibitory function of CD operates in the pathways that link to the auditory features of human sounds; the CD may not suppress the neural responses to pure tones or suppress in a gradient manner based on the distance of pure tones from the range of human voice pitch.

To provide direct visualization of individual-level data and comparisons among sounds, we extracted each participant's data using the group-level clusters as a spatial-temporal filter. For each sound, the results of the cluster with consistent suppression patterns across sounds were presented in Figure 4C. For the sum ERP responses in the suppression cluster of syllable sound, paired t-tests revealed that the amplitude of P2 response in *GP* was reduced relative to that in *PL* ($t(19) = 7.269$, $p < 0.0001$, $d = 1.114$). For the sum ERP responses to cough sound, the amplitude of P2 response in *GP* was less than that in *PL* ($t(19) = 4.437$, $p = 0.0002$, $d = 0.835$). For the sum ERP responses to the pure tone, P2 response in *GP* was significantly suppressed than that in *PL* ($t(19) = 5.240$, $p < 0.0001$, $d = 1.234$). These results indicated that the suppression effect of CD was found in each kind of sound, consistent with the results of the spatiotemporal cluster permutation test. To compare the suppression effect of GP across auditory probes, a repeated-measure one-way ANOVA was performed on the differences between the sum of ERP data in the cluster of *PL* and *GP* across three

types of auditory probes. The results showed a significant effect of sound ($F(2,38)$ = 7.980, p = 0.001, partial η^2 = 0.296). Bonferroni corrected paired t-tests revealed that the suppression effect in syllable sound was larger than cough sound and tone. (syllable vs. cough: $t(19) = 3.419$, $p = 0.009$, $d = 0.804$; syllable vs. tone: $t(19) = 3.461$, $p = 0.008$, d = 1.131). These results were consistent with the ERP cluster results as well as the component analysis results (Fig. 4) that showed smaller inhibitory effects in tones compared with the other two types of sound.

Results of time-frequency analysis

To further investigate how CD influenced auditory processes – whether suppressed the response magnitude or disrupted the timing, we carried out time-frequency analysis using spatiotemporal cluster-based permutation tests, separately for response power and phase. Because the three sounds included in this study had different modulation rates (the cough sound had sharper acoustic onset and hence had relative more energy in the theta band compared with syllable and tone sounds), we first carried out the time-frequency analysis to explore the modulation effects in separate delta (1-3 Hz) and theta (4-8 Hz) bands. Next, for a fair comparison with more statistical power, we pooled the two frequency bands together and performed the time-frequency analysis in one lower-frequency band (1-8 Hz) that included the most speech processes for all types of sounds (Giraud and Poeppel, 2012). Because similar results were obtained in separate and combined frequency bands, we elaborated on the results of one lower

frequency band.

As shown in Fig. 5, syllable and pure tone were suppressed in the delta frequency (1-3 Hz) band for both power (for syllable, $p = 0.0082$; tone, $p = 0.0112$) and ITC (for syllable, $p = 0.0055$; tone, $p = 0.0003$), whereas inhibition to auditory responses to cough sound was mostly in the theta frequency (4-8 Hz) band (for power, $p=0.0041$; for ITC, $p=0.0046$). Spectrum analysis of the three acoustic stimuli revealed that the modulation spectrum of cough sound had a wider distribution of 1-8 Hz, compared with auditory syllable of 1-5 Hz and pure tone of 1-3 Hz. The inhibitory effects on different sounds in corresponding frequency bands indicated that the suppression presumably concentrated in the frequency bands that tracked the acoustic signals.

The results of ITC and power in the lower-frequency band (1-8 Hz) exhibited consistent patterns across all types of auditory probes (Fig. 6), similar to the results in the separate frequency bands. Specifically, for ITC results (Fig. 6A), two significant clusters that were distinct in spatial and temporal dimensions were found. The first significant cluster (in yellow) had significantly higher ITC values in *GP* than those in *PL* (for syllable, $p = 0.0002$; cough, $p = 0.0197$; tone, $p = 0.0249$). This cluster in responses to each type of auditory probe occurred at -100 ms (the earliest time included in the analysis) and lasted until 100 ms after stimulus onset (for syllable, 200 ms). Significant electrodes were mostly located in parietal regions, and some extended to frontal regions. The characteristics of this cluster – occurrence before auditory stimuli, posterior spatial distribution, and more consistent phase coherence in *GP* than *PL* –

collaboratively suggested that general preparation for actions increased the timing consistency of neural processing across each instance of preparation.

On the contrary, the second significant cluster (in green) had significantly lower ITC values in *GP* than those in *PL* (for syllable, $p = 0.0499$; cough, $p = 0.0016$; tone, $p = 0.0232$). Moreover, this cluster was apparent in the period of 100 to 300 ms after sound onset and had a central distribution. These temporal and spatial features of this cluster resembled the configuration of the auditory P2 component. The less consistent phase coherence in *GP* than *PL* in a response component to all auditory probes suggested that CD in general preparation decreased the timing consistency of auditory processing.

For results of power (Fig. 6B), only one significant cluster was observed after sound onset (for syllable, $p = 0.0185$; cough, $p = 0.0069$; tone, $p = 0.0475$). This cluster indicated less power of neural signals in *GP* than those in *PL*. The decrement in power was sparse in tone and more prominent for cough and syllable, consistent with the ERP results. These results suggest that CD during general preparation dampened response power ubiquitously for all auditory stimuli, but the quantity of the power decrease may depend on the established associations between the features in articulation and its auditory consequences. No consistent differences were observed in other frequency bands either for ITC or power. Taken together, these results suggested that the generic inhibition functions of CD manifested in the modulation of both power and timing of perceptual processes in low-frequency bands. Modulation on process timing applies

equally to each type of auditory probe, whereas modulation on process power may depend on the degree of overlaps between features in articulatory and auditory domains.

Similar to Fig. 4C, individual data of the sum of power was presented in the last row of Fig. 6B and the sum of ITC in each significant cluster was presented in the top and bottom row of Fig. 6A separately. First, all paired t-tests between GP and PL on each measure were significant (all ps < 0.05), consistent with the results of the time-frequency cluster analysis. To compare the suppression effect of GP across auditory probes, repeated measures one-way ANOVA was performed on the difference between *PL* and *GP,* separately for ITC and power. All results showed a significant effect of sound (the first ITC cluster: $F(2,38) = 11.97$, $p = 0.0004$, partial $\eta^2 = 0.387$; the second ITC cluster: $F(2,38) = 4.993$, $p = 0.012$, partial $\eta^2 = 0.208$; power: $F(2,38) = 4.929$, $p = 0.013$, partial $\eta^2 = 0.206$). Bonferroni corrected paired *t*-tests for the first ITC cluster revealed that the enhancement effect in the first ITC cluster in syllable sound was larger than cough sound and tone (syllable vs. cough: $t(19) = 3.948$, $p = 0.003$, $d = 1.057$; syllable vs. tone: $t(19) = 3.563$, $p = 0.006$, $d = 0.939$). For the second ITC cluster, the post-hoc paired t-tests revealed that the suppression effect in the second ITC cluster in syllable sound was smaller than cough sound ($t(19) = 2.766$, $p = 0.037$, $d = 0.702$). The paired t-tests on power revealed that the suppression effect in the power cluster in tone was significant smaller than syllable and cough sound (tone vs. syllable: $t(19) = 2.947$, $p = 0.025$, $d = 0.821$; tone vs. cough: $t(19) = 3.089$, $p = 0.018$, $d = 0.963$). These

results suggest that the smaller inhibitory effects on tones compared with the other two types of sound were more consistent in the modulation of power.

We also carried out a spectrotemporal cluster analysis in the middle of the preparation stage (0.5 – 1.1 s after visual cue onset, a period without possible contamination of visual fixation and subsequent auditory probes). The results showed a similar power decrease in the lower frequency band in both GPns and GP conditions compared to the PL, suggesting the availability of motor signals in the early stage of motor intention.

Discussion

We investigated the function of the motor signal generated in the early stage of motor intention. With a delayed articulation paradigm including three different types of sounds to produce in the articulation task, we found that the motor signal during motor intention contained no specific information about the sound and suppressed later auditory neural responses to all types of sounds, including speech (syllable /ba/), non-speech (cough), and non-human sound (pure tone). The inhibitory effects were stronger for sounds that humans can produce than non-human sounds. Moreover, we found that the inhibitory modulation of CD was mediated by dampening response amplitude and adding temporal variance to sensory processes. These results suggest a generic inhibitory function of CD that is implemented in the form of modulations on neural response magnitude and timing.

We observed suppression of auditory responses caused by motor signals in the

stage of motor intention (Fig. 2). These results are consistent with our previous findings (Li et al., 2020) and suggest that motor signals can transmit to sensory regions in the earliest stage of action. In addition, CD suppressed the neural responses to auditory probes in general preparation, when participants did not know any specific information about actions or consequences of actions. This finding indicates that the inhibitory CD is generated early in the motor intention stage, consistent with the observations that suppression effects were absent when the action is involuntarily triggered without movement intention (Timm et al., 2014). This early onset of motor signals, complementary to commonly observed suppression at the time of action (Blakemore et al., 1998; Ross et al., 2001; Aliu et al., 2009), serves the computational purpose of monitoring throughout the time course of action (Eliades and Wang, 2008; Tian and Poeppel, 2014).

More importantly, the early available CD takes a generic form of inhibition, as the inhibition function modulates all types of sounds (Fig. 3 & 4 & 6). The generic inhibitory function of CD found in the study was consistent with the previous findings that both speech sounds and non-human sounds (pure tone) were suppressed during speech production (Houde et al., 2002). This non-specific form of prediction may provide the probability of self-induced sensory consequence without the demand for specific representation and hence establishing the agency in motor intention (Blakemore and Decety, 2001; Desmurget et al., 2009). Moreover, the observed generic inhibitory function mediates the presupposition of a theoretical mechanism that motor

signals increase the signal-to-noise ratio of perceptual responses (Reznik and Mukamel, 2019).

Furthermore, we found that the intensity of suppression effects was associated with the distance between feedback sounds and the sounds that humans can produce. Specifically, the strength of inhibition was stronger for the auditory stimuli of syllables and cough than pure tones (Fig. 3 & 4 & 6). These results are consistent with previous studies in which the strength of suppression effects correlated with the action-perception association established via learning – suppression was strongest for the tones with the frequency that paired with action during training, whereas the suppression strength decreased in neurons with auditory receptive fields of adjacent frequencies (Schneider et al., 2018). In the present study, the 500 Hz pure tone was off the normal pitch voice that humans' vocal folds usually produce. The less suppression of general preparation on the non-human sound of pure tone could cause by connection strength differences in different associations between motor and auditory areas. The associations between motor and auditory systems for sounds that humans can produce, including speech and non-speech sounds, are strengthened via everyday pronunciation. Whereas the motor system only links to the auditory features of non-human sounds that overlap with features of sounds that humans can produce, but less or none to the auditory features that humans cannot produce. Via these available links, the CD transmits and modulates auditory processes, but less strength in the links yields less suppression for non-human sounds, even in the generic inhibitory function of CD

during general preparation. That is, the motor signal of CD during the intention stage in human articulation does not contain specific information about the sounds that humans can produce, but the CD may be still constrained by the established action-perception associations and has less influence on the auditory processes of non-human sounds.

We analyzed EEG signals both in the temporal domain (ERP) and time-frequency domain (power and ITC). Each of these analyses reveals phase-locked and non-phase-locked aspects of EEG data. Specifically, ERP was obtained by averaging epochs that were time-locked to the sound onset. This ERP analysis in the temporal domain amplified the SNR of signals that phase-locked to the events. Whereas, induced power indicates the response strength of non-phase locked signals in a certain frequency band, and ITC quantifies phase consistency across trials in the time-frequency domain. The combination of power and ITC yields the effects in ERP. Using these three complementary measures, we found the generic inhibitory function of CD was implemented both in the modulations of response power and timing. As shown in Fig 6, suppression effects of general preparation were observed both in power and phase coherence for every type of sound probe in the low-frequency band (1~8Hz). These results of spectral-temporal analyses are consistent with ERP results (Fig. 3&4) , and demonstrate that the neural modulation mechanisms of CD on sensory processing are dampening response amplitude and increasing temporal variance.

We observed that the inhibition effects manifested in both ITC and power, but with different modulation patterns (Fig. 6). The dissociation between power and phase hints

at potential processes of generic inhibition modulation of CD -- CD may influence the timing of processing for all sensory features over auditory cortices; then based on the strength of established connections between motor and sensory features, the detailed inhibition was realized by manipulating the rate of responses and hence the response power. Moreover, 'adding noise' could be more 'economic' than precisely manipulating neural sensitivity. Increasing temporal variance in the neural phase decreases the probability of alignment between external stimuli and the high excitability state of the neural phase (Schroeder and Lakatos, 2009; Giraud and Poeppel, 2012). When no content information is available during general preparation, the temporal manipulation on the neural phase primarily mediates the suppression effects over vast neural populations. When motor signals become concrete, especially when action is executed, the modulation on response amplitude dominates the suppression effects precisely on a specific auditory target (Houde et al., 2002; Baess et al., 2011).

We did not find suppression in the N1 component that was observed in our previous study (Li et al., 2020). The absence of N1 suppression could be due to the different nature of motor signals induced by important experimental differences between the two studies. In the present study, the CD is more general due to the inclusion of three types of sounds. The uncertainty of what types of sound to produce makes that even the categorical information could not be established during preparation. Whereas in the previous study, the CD contained specific information in a sound category because the subsequent articulation task was only about syllables.

Our previous studies suggest that the more concrete and detailed prediction about the sound from the motor signals, the earlier the modulation effects occurred (Tian and Poeppel, 2013, 2015; Tian et al., 2018). The more abstract 'prediction' rather than 'specific' prediction of a particular type of sound may make the modulation effects in the current study in the later perceptual component because the component of P2 is more relative to abstract categorical coding (Bidelman et al., 2013; Mankel et al., 2020).

The observed N1 enhancement for syllables could be the result of motor intention interacting with speech sounds. We observed the ITC increases caused by motor intention around the onset of the auditory probe and extending to the period that overlapped with N1 latency. Previous studies have demonstrated that the phase at the theta range automatically synchronized with subsequent perceptual responses as early as in the motor planning stage (Tomassini et al., 2017). The observed increased consistency in phase probably reflects the interaction of motor preparation and auditory stimuli, as the motor intention may facilitate the onset of auditory processing, especially for speech sounds. This facilitation could even be as early as in the subcortical pathway, as the studies in vision and eye movement suggest that the CD signals can be available in the colliculus and thalamus (Cavanaugh et al., 2020).

The coexistence of generic inhibitory effects at the latency of P2 and mixed effects at the latency of N1 could be the results of our specific paradigms in the combination of the recording methods used in this study. We designed this study by exploiting the

modulation effects of the action on auditory perception. However, the EEG recordings with low spatial resolution could not clearly separate the neural sources of motor preparation and auditory processes, especially at the sound onset. Future studies using methods that have both high temporal and spatial resolutions, such as intracranial EEG would offer further evidence distinguishing the sources of CD and its modulation effects in the auditory cortices. Moreover, we used the auditory stimuli with a male voice. Separating participants into two gender groups would provide further evidence investigating the gradient suppression effects based on the distance of the auditory stimuli from the predictive auditory consequences, just like our observations of less suppression for pure tones. However, the random recruitment of participants did not give us enough power to test this interesting point. Future studies can explore the gradient modulation effects in the direction of gender differences.

Using the delayed articulation paradigm, we observed that corollary discharge can be available in motor intention and take a generic form of modulation function to suppress all types of sounds. The generic inhibition function was constrained by the strength of associations between motor and auditory features and realized by adjusting the amplitude and timing of neural responses. By dissecting the motor-to-sensory transformation signals in functional and temporal dimensions, our results suggest a functional granularity of corollary discharge that mediates the dynamics of motor-to-sensory transformation to fulfill distinct computations in sensorimotor integration and motor control.

686

687

Reference

689 Aliu SO, Houde JF, Nagarajan SS (2009) Motor-induced suppression of the auditory
690 cortex. J Cogn Neurosci 21:791-802.
691 Baess P, Horváth J, Jacobsen T, Schröger E (2011) Selective suppression of self-
692 initiated sounds in an auditory stream: An ERP study. Psychophysiology
693 48:1276-1283.
694 Bidelman GM, Moreno S, Alain C (2013) Tracing the emergence of categorical speech
695 perception in the human auditory system. Neuroimage 79:201-212.
696 Blakemore S-J, Decety J (2001) From the perception of action to the understanding of
697 intention. Nat Rev Neurosci 2:561-567.
698 Blakemore S-J, Wolpert DM, Frith CD (1998) Central cancellation of self-produced
699 tickle sensation. Nat Neurosci 1:635-640.
700 Cavanaugh J, McAlonan K, Wurtz RH (2020) Organization of corollary discharge
701 neurons in monkey medial dorsal thalamus. J Neurosci 40:6367-6378.
702 Crapse TB, Sommer MA (2008) Corollary discharge across the animal kingdom. Nat
703 Rev Neurosci 9:587-600.
704 Desmurget M, Reilly KT, Richard N, Szathmari A, Mottolese C, Sirigu A (2009)
705 Movement intention after parietal cortex stimulation in humans. Science
706 324:811-813.
707 Eliades SJ, Wang X (2008) Neural substrates of vocalization feedback monitoring in
708 primate auditory cortex. Nature 453:1102-1106.
709 Faul F, Erdfelder E, Lang A-G, Buchner A (2007) G* Power 3: A flexible statistical
710 power analysis program for the social, behavioral, and biomedical sciences.
711 Behav Res Methods 39:175-191.
712 Giraud A-L, Poeppel D (2012) Cortical oscillations and speech processing: emerging
713 computational principles and operations. Nat Neurosci 15:511-517.
714 Gramfort A, Luessi M, Larson E, Engemann DA, Strohmeier D, Brodbeck C,
715 Parkkonen L, Hämäläinen MS (2014) MNE software for processing MEG and
716 EEG data. Neuroimage 86:446-460.
717 Grill-Spector K, Henson R, Martin A (2006) Repetition and the brain: neural models of
718 stimulus-specific effects. Trends Cogn Sci 10:14-23.
719 Guenther FH (1995) Speech sound acquisition, coarticulation, and rate effects in a
720 neural network model of speech production. Psychol Rev 102:594.
721 Hickok G (2012) Computational neuroanatomy of speech production. Nat Rev
722 Neurosci 13:135-145.
723 Horváth J, Maess B, Baess P, Tóth A (2012) Action–sound coincidences suppress
724 evoked responses of the human auditory cortex in EEG and MEG. J Cogn
725 Neurosci 24:1919-1931.

726 Houde JF, Nagarajan SS (2011) Speech production as state feedback control. Front
727 Hum Neurosci 5:82.
728 Houde JF, Nagarajan SS, Sekihara K, Merzenich MM (2002) Modulation of the
729 auditory cortex during speech: an MEG study. J Cogn Neurosci 14:1125-1138.
730 Kawato M (1999) Internal models for motor control and trajectory planning. Curr Opin
731 Neurobiol 9:718-727.
732 Kilteni K, Andersson BJ, Houborg C, Ehrsson HH (2018) Motor imagery involves
733 predicting the sensory consequences of the imagined movement. Nat Commun
734 9:1-9.
735 Lehmann D, Skrandies W (1980) Reference-free identification of components of
736 checkerboard-evoked multichannel potential fields. Electroencephalogr Clin
737 Neurophysiol 48:609-621.
738 Li S, Zhu H, Tian X (2020) Corollary discharge versus efference copy: distinct neural
739 signals in speech preparation differentially modulate auditory responses. Cereb
740 Cortex 30:5806-5820.
741 Lu T, Liang L, Wang X (2001) Temporal and rate representations of time-varying
742 signals in the auditory cortex of awake primates. Nat Neurosci 4:1131-1138.
743 Luo H, Poeppel D (2007) Phase patterns of neuronal responses reliably discriminate
744 speech in human auditory cortex. Neuron 54:1001-1010.
745 Mankel K, Barber J, Bidelman GM (2020) Auditory categorical processing for speech
746 is modulated by inherent musical listening skills. Neuroreport 31:162.
747 Maris E, Oostenveld R (2007) Nonparametric statistical testing of EEG-and MEG-data.
748 J Neurosci Methods 164:177-190.
749 Miall RC, Wolpert DM (1996) Forward models for physiological motor control. Neural
750 Netw 9:1265-1279.
751 Perrin F, Pernier J, Bertrand O, Echallier JF (1989) Spherical splines for scalp potential
752 and current density mapping. Electroencephalogr Clin Neurophysiol
753 72:184-187.
754 Press C, Thomas E, Yon D (2022) Cancelling cancellation? Sensorimotor control,
755 agency, and prediction.
756 Reznik D, Mukamel R (2019) Motor output, neural states and auditory perception.
757 Neurosci Biobehav Rev 96:116-126.
758 Ross J, Morrone MC, Goldberg ME, Burr DC (2001) Changes in visual perception at
759 the time of saccades. Trends Neurosci 24:113-121.
760 Schneider DM, Sundararajan J, Mooney R (2018) A cortical filter that learns to
761 suppress the acoustic consequences of movement. Nature 561:391-395.
762 Schroeder CE, Lakatos P (2009) Low-frequency neuronal oscillations as instruments of
763 sensory selection. Trends Neurosci 32:9-18.
764 Schubotz RI (2007) Prediction of external events with our motor system: towards a new
765 framework. Trends Cogn Sci 11:211-218.
766 Sommer MA, Wurtz RH (2006) Influence of the thalamus on spatial visual processing
767 in frontal cortex. Nature 444:374-377.

768 Sperry RW (1950) Neural basis of the spontaneous optokinetic response produced by
769 visual inversion. J Comp Physiol Psychol 43:482.
770 Tallon-Baudry C, Bertrand O, Delpuech C, Pernier J (1996) Stimulus specificity of
771 phase-locked and non-phase-locked 40 Hz visual responses in human. J
772 Neurosci 16:4240-4249.
773 Teng X, Ma M, Yang J, Blohm S, Cai Q, Tian X (2020) Constrained structure of ancient
774 Chinese poetry facilitates speech content grouping. Curr Biol 30:1299-1305.
775 e1297.
776 Tian X, Poeppel D (2010) Mental imagery of speech and movement implicates the
777 dynamics of internal forward models. Front Psychol 1:166.
778 Tian X, Poeppel D (2013) The effect of imagination on stimulation: the functional
779 specificity of efference copies in speech processing. J Cogn Neurosci
780 25:1020-1036.
781 Tian X, Poeppel D (2014) Dynamics of self-monitoring and error detection in speech
782 production: evidence from mental imagery and MEG. J Cogn Neurosci
783 27:352-364.
784 Tian X, Poeppel D (2015) Dynamics of self-monitoring and error detection in speech
785 production: evidence from mental imagery and MEG. J Cogn Neurosci
786 27:352-364.
787 Tian X, Ding N, Teng X, Bai F, Poeppel D (2018) Imagined speech influences
788 perceived loudness of sound. Nat Hum Behav 2:225-234.
789 Timm J, SanMiguel I, Keil J, Schröger E, Schönwiesner M (2014) Motor intention
790 determines sensory attenuation of brain responses to self-initiated sounds. J
791 Cogn Neurosci 26:1481-1489.
792 Tomassini A, Ambrogioni L, Medendorp WP, Maris E (2017) Theta oscillations locked
793 to intended actions rhythmically modulate perception. Elife 6:e25618.
794 Vallat R (2018) Pingouin: statistics in Python. J open source softw 3:1026.
795 von Helmholtz HLF (1910) Handbuch der physiologischen Optik. v. 2, 1910: L. Voss.
796 von Holst E, Mittelstaedt H (1950) Das reafferenzprinzip. Naturwissenschaften
797 37:464-476.
798 Wang X, Zhu H, Tian X (2019) Revealing the temporal dynamics in non-invasive
799 electrophysiological recordings with topography-based analyses.
800 bioRxiv:779546.
801 Wolpert DM, Ghahramani Z (2000) Computational principles of movement
802 neuroscience. Nat Neurosci 3:1212-1217.
803 Yang J, Zhu H, Tian X (2018) Group-level multivariate analysis in EasyEEG toolbox:
804 examining the temporal dynamics using topographic responses. Front Neurosci
805 12:468.

806

Figure Legends

Figure 1. A schematic of the hypothesis about the generic inhibitory functions of corollary discharge (CD). CD can be available as early as in the motor intention stage and throughout the entire course of action. CD inhibits sensory processing (demonstrate as in blue). The inhibition function of CD could be non-specific (generic) to all sounds at the beginning of the action course (motor intention) and becomes stronger and specific to the sensory consequences of actions. The generic inhibition function of CD may also depend on the established associations between actions and their sensory consequences – the strength of generic inhibition on the sensory process may depend on feature overlaps between feedback stimuli and the sensory representation that can be induced by actions performed by an agent. Specific in the auditory domain, the generic inhibition function of CD may suppress non-human sound less than for speech and non-speech sounds that humans can produce, as indicated by the shorter blue bar in non-human sound.

Figure 2. Experimental paradigms, behavioral and ERP results. (A) Illustration of four types of trials. In *GP* trials, participants were asked to prepare to speak when two meaningless symbols were on screen. The symbols did not provide any information about what participants were going to say, and hence they generally prepare the action of speaking. An auditory probe (randomly selected from a syllable /ba/, a cough sound, and a 500 Hz pure tone) was presented at the end of the preparation stage to probe the

modulatory effect of CD on auditory processes. When a green visual cue appeared, participants were asked to articulate accordingly. The visual cue 'ba' is used as an example for illustration purposes; two other visual cues were included for producing cough and humming tone. GP_{NS} trials were identical to *GP* trials except that no auditory probe was presented during the preparation stage. In *NP* trials, participants performed the articulation task without preceding preparation. GP_{NS} and *NP* trials were used to control and quantify the general preparation. In the *PL* trials, participants were asked to passively listen to the auditory probes that were identical to those in *GP* trials. No preparation or articulation task was required in the *PL* trials. The *PL* trials were used to compare with auditory responses in *GP* trials to quantify the neural modulation effects of preparation. (See Methods for details.) (B) Mean RTs across three conditions with individual data. Participants articulated faster in *GP* and GP_{NS} conditions than in *NP* condition, but no difference between *GP* and GP_{NS} conditions, suggesting that the performance of general preparation was independent of the auditory probes. Error bars indicate ±SEM. ***$p < 0.001$. (C) Grand average GFP waveforms and topographic responses for three types of auditory probes combined. Auditory N1 and P2 components were observed in each condition. Yellow and blue represent *GP* and *PL* conditions, respectively. Individual waveform responses are superimposed on the plot. (D) Mean N1 and P2 amplitudes in two conditions with individual data. The magnitude of the P2 component was significantly smaller in *GP* than that in *PL*, suggesting that general preparation suppressed the auditory responses. Error bars indicate ±SEM. *p <

849 0.05.

850

851 **Figure 3.** Grand average ERP responses to auditory probes in the representative
852 channel of Cz. The waveform responses are in the left column, and the N1/P2
853 component responses are in the right column in each panel. (A), (B), (C), and (D) depict
854 responses to syllable, cough, tone, and the average across three types of sounds,
855 respectively. Individual data are superimposed on each plot. Yellow and blue represent
856 *GP* and *PL* conditions, respectively. Error bars indicate ±SEM. **p < 0.01, ***p <
857 0.001.

858

859 **Figure 4.** The results of spatiotemporal analysis on ERP responses. Each column
860 indicates the results for syllable, cough, and tone, respectively. (A) The results of the
861 spatiotemporal analysis. The x-axis represents time relative to the auditory probe onset
862 at 0 ms, and the y-axis represents each of the 32 electrodes. The grayscale in the
863 background represents *t* values comparing the ERP responses between *GP* and *PL*
864 conditions (*GP* minus *PL*). Yellow and green indicate significant clusters with positive
865 and negative *t* values, respectively. (B) Topographic representation of the significant
866 spatiotemporal clusters in (A) and the raw ERP topographies that derived the
867 significant cluster results. Each topography in the first row represents averaged *t* values
868 across the time interval of each significant cluster in (A), indicated by corresponding
869 color dashed lines. The black squares on the topographies indicate the significant

electrodes in the cluster. The second and third rows are the topographies of averaged ERP responses across the corresponding time interval of the cluster in the *PL* and *GP,* respectively. The black squares on the ERP topographies label the same electrodes in the corresponding significant cluster above. Considering the polarity of ERP responses, the clusters observed in typical latency of auditory responses (100 – 200 ms after stimuli onset) showed inhibition effects of general preparation on all types of auditory probes -- the ERP amplitudes in *GP* were smaller than those in *PL* (less positive or less negative in electrodes of positive or negative ERP responses, respectively). (C) The summarized results of three ERP clusters with individual data extracted using the group-level clusters as a spatial-temporal filter. To better compare the suppression effect of GP in each sound, the sum of ERP data in a similar cluster of frontal-temporal distribution was presented for each sound. Error bars indicate ±SEM. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$.

Figure 5. Results of spatiotemporal cluster analysis in the delta frequency band (1-3Hz) and the theta frequency band (4-8Hz) for each type of auditory probe, separately for ITC and power. Each column indicates the results for syllable, cough, and tone, respectively. The grayscale images represent *t* values in each of the 32 electrodes across time, obtained by comparing the ITC or power between *GP* and *PL* conditions (*GP* minus *PL*). The yellow and green indicate clusters with positive and negative *t* values and hence enhancement and suppression effects, respectively. Topographies of

averaged *t* values are plotted every 50 ms from -100 to 300 ms when the significant clusters were observed. Significant electrodes in each cluster are marked with black squares on each topography. (A) Results in the delta frequency band. The ITC results are presented in the top row. For syllable, two significant clusters were found. Topographies of the first clusters in yellow, spanning from -100 to 200 ms, are shown at the top of the spatiotemporal plots. Significant electrodes in this cluster were mostly located in parietal regions and some extended to frontal regions. Topographies of the second cluster in green, spanning from 100 to 300 ms, are shown at the bottom of the spatiotemporal plots. Significant electrodes in this cluster were located in central regions. For cough and tone, one significant cluster in green was found, similar to the second cluster in the results of syllable. Power results are presented in the bottom row. For syllable, one significant cluster was found. For tone, two significant clusters were found. No significant cluster was found in the result of cough. (B) Results in the theta frequency band. The top row shows ITC results in the theta band. For syllable and tone, one significant cluster in yellow was found. For cough, two significant clusters were found, which is similar to the results of syllable in the delta band. The bottom row shows the power results in the theta band. For syllable, one significant cluster in yellow was found. For cough, one significant cluster in green was found. For tone, no significant cluster was found.

Figure 6. Results of spatiotemporal cluster analysis in one lower frequency band

(1-8Hz) for each type of auditory probe, separately for ITC and power. Each column indicates the results for syllable, cough, and tone, respectively. The grayscale images represent *t* values in each of the 32 electrodes across time, obtained by comparing the ITC or power between *GP* and *PL* conditions (*GP* minus *PL*). The yellow and green indicate clusters with positive and negative *t* values and hence enhancement and suppression effects, respectively. Topographies of averaged *t* values are plotted every 50 ms from -100 to 300 ms when the significant clusters were observed. Significant electrodes in each cluster are marked with black squares on each topography. (A) ITC results. For each auditory probe, two significant clusters were found. Topographies of the first clusters in yellow, spanning from -100 to 100 ms (for syllable, to 200 ms), are shown at the top of the spatiotemporal plots. Significant electrodes in this cluster were mostly located in parietal regions and some extended to frontal regions. Topographies of the second cluster in green, spanning from 100 to 300 ms, are shown at the bottom of the spatiotemporal plots. Significant electrodes in this cluster were located in central regions. The summarized results of two ITC clusters of each sound with individual data superimposed are presented at the top and bottom near each cluster separately. (B) Power results. For each auditory probe, one significant cluster was found. The clusters were observed from 100 to 300ms after sound onset in central regions. The summarized results of the power cluster with individual data superimposed are presented at the bottom near each cluster. Error bars indicate ±SEM. ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$.