Applied Psycholinguistics (2020), **41**, 771–795
doi:10.1017/S0142716420000223

CAMBRIDGE
UNIVERSITY PRESS

ORIGINAL ARTICLE

A cross-language study on feedforward and feedback control of voice intensity in Chinese–English bilinguals

Xiao Cai, Yulong Yin and Qingfang Zhang*

Renmin University of China
*Corresponding author. Email: qingfang.zhang@ruc.edu.cn

(Received 03 July 2019; revised 14 April 2020; accepted 20 April 2020; first published online 10 July 2020)

Abstract

Speech production requires the combined efforts of feedforward and feedback control, but it remains unclear whether the relative weighting of feedforward and feedback control is organized differently between the first language (L1) and the second language (L2). In the present study, a group of Chinese–English bilinguals named pictures in their L1 and L2, while being exposed to multitalker noise. Experiment 1 compared feedforward control between L1 and L2 speech production by examining intensity increases in response to a masking noise (90 dB SPL). Experiment 2 compared feedback control between L1 and L2 speech production by examining intensity increases in response to a weak (30 dB SPL) or strong noise (60 dB SPL). We also examined a potential relationship between L2 fluency and the relative weighting of feedforward and feedback systems. The results indicated that L2 speech production relies less on feedforward control relative to L1, exhibiting attenuated Lombard effects to the masking noise. In contrast, L2 speech production relies more on feedback control than L1, producing larger Lombard effects to the weak and strong noise. The relative weighting of feedforward and feedback control is dynamically changed as second language learning progresses.

Keywords: L2 speech motor control; feedforward control; feedback control; Lombard effect; voice intensity

Introduction

An issue receiving intense attention in speech production is how the brain plans linguistic processing prior to overt speech production (Levelt et al., 1999). Although articulation is seen as a lower-level motor output (Indefrey & Levelt, 2004), speaking is a highly complex sensorimotor task that requires the combined efforts of feedforward and feedback control systems (Guenther et al., 2006). To date, how these two subsystems work to ensure successful communication remains poorly understood.

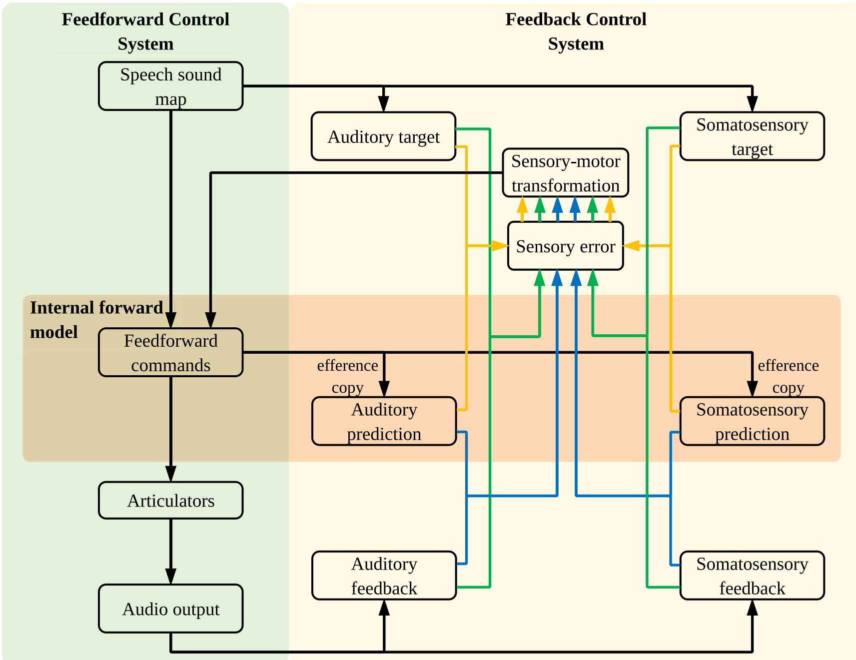

Figure 1. A schematic diagram of the processes involved in speech motor control. The model includes an internal forward model (pink box) that generates auditory prediction based on a copy of planned feedforward commands (efference copy). Auditory feedback control compares actual auditory feedback with auditory prediction and auditory target, indicated by blue and green arrows, respectively. A special case of feedforward control eliminates the involvement of auditory feedback by comparing auditory prediction with auditory target (indicated by yellow arrows). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article).

Terminology and general principles of speech motor control

Several speech motor control models have been formulated; we integrated Directions into Velocities of Articulators (DIVA; Guenther, 2006) and State Feedback Control (Houde & Nagarajan, 2011) to describe feedforward and feedback control (see Figure 1 for details). Speech production begins with a unit in the "speech sound map," which can be a phoneme, syllable, or phrase. As schematized in Figure 1, feedforward control reads out previously learned motor commands for speech sounds and further issues them to articulators. This mechanism emphasizes its independence from the sensory feedback associated with articulation (Guenther, 2016). Therefore, feedforward control enables the rapidity of speech, but lacks the ability to monitor errors in speech output (Parrell et al., 2019). Because we live in time-varying and unpredictable surroundings, feedforward control alone cannot ensure effective speech.

Unlike feedforward control, feedback control relies on sensory feedback to maintain speech (Guenther, 2016; Kearney & Guenther, 2019). The auditory feedback control system compares actual auditory feedback with intended auditory feedback, and in case of any mismatch, auditory errors are transformed into corrective

https://doi.org/10.1017/S0142716420000223 Published online by Cambridge University Press

commands that decrease the perceived errors. This is similar to a somatosensory feedback control mechanism (Guenther, 2006, 2016; Hickok et al., 2011). Within this framework, there are two coexisting routes to generate intended auditory feedback (Tian & Poeppel, 2012). Firstly, the activation of speech sound map leads to the activation of auditory target, which defines the desired auditory feedback that should arise when a speaker correctly produces the sound (Guenther, 2016; Tourville & Guenther, 2011). Secondly, an internal forward model utilizes an efference copy of feedforward commands to internally estimate the current state of vocal tract dynamics and generate auditory prediction (Hickok, 2012; Tian & Poeppel, 2010, 2012, 2013). The feedback control system is indispensable in speech motor control, allowing speakers to regulate movements and interact well with the environment in presence of external perturbations (Bays & Wolpert, 2006).

Parrell et al. (2019) proposed a special case of feedforward control by eliminating auditory feedback involvement. An auditory prediction (realized using an internal forward model) is the possible outcome of articulatory movement before auditory feedback is received. This prediction is based on previously established causal associations between motor commands and auditory output. This is also why speakers feel that they can "hear" the speech internally when they imagine speaking without moving any articulators (Tian & Poeppel, 2012, 2015). Critically, the motor-based auditory predictions can be directly compared with auditory targets to verify the correctness of planned feedforward commands (Hickok, 2012). If auditory predictions fail to match auditory targets, the feedforward control system transforms error signals into corrective motor commands (Parrell et al., 2019).

Speech motor control in bilinguals

Current models detail the organization of feedforward and feedback control exclusively in the first language (L1; see Parrell et al., 2019 for a review), while research into the second language (L2) has not yet fully considered this issue. More recently, researchers noticed that speech motor control in bilinguals may vary by language type (L1 vs. L2; Liu & Tian, 2018; Mitsuya et al., 2011; Simmonds et al., 2011a, 2011b). Here we adopt Grosjean's (2010) succinct definition of bilinguals as people who use two languages in their daily life. Of note, there is still insufficient theoretical and empirical information on L2 speech motor control, which highlights the need for further research in this field.

For L1 speech production, a basic idea is that the feedforward and feedback control subsystems cooperate with each other (Parrell et al., 2019); thus, it is important to understand the relative weighting of these systems in speech motor control (Guenther, 2016; Guenther et al., 2006). Researchers have emphasized a transition from feedback-dominant to feedforward-dominant, driven by production experiences (Guenther, 2006; Guenther & Vladusich, 2012; Liu et al., 2010c; Scheerer et al., 2013). Speakers' initial attempts to produce speech result in errors, and production relies heavily on feedback control. With sufficient practice, feedforward commands can result in the same sensory consequences without errors, and production principally relies on feedforward control (Guenther, 2006; Guenther & Vladusich, 2012). However, L1 and L2 production experiences are inherently different (Mitsuya et al., 2011). L1 speech motor learning begins in infancy

(Tourville & Guenther, 2011), but within the broad population of bilinguals, the L2 acquisition age is widely varied. Some bilinguals acquire L2 from birth, some around puberty, and others during adulthood (Woumans et al., 2015). In most cases, bilinguals are exposed to an L2 after their L1 has already been established. It is therefore possible that feedforward and feedback control are weighted differently for bilinguals' two language systems.

Motor movements used to produce native sounds are highly overlearned and automatic, requiring much less online sensory monitoring (Simmonds et al., 2011a, 2011b). However, evidence shows that L2 sounds are produced with larger variability (Chen et al., 2001; Ng et al., 2008; Wang & van Heuven, 2006), implying that L2 feedforward commands are less familiar and more variable (Mitsuya et al., 2011). Thus, we hypothesized that, compared with L1, L2 production relies on feedback control to a greater extent and on feedforward control to a lesser extent. This hypothesis is supported by two early studies reporting that bilinguals speak more slowly and have more hesitation or sound repetitions in L2 relative to L1 under a delayed auditory feedback condition (Mackay, 1970; Van Borsel et al., 2005). The underlying logic is that an increased weighting of feedback control increases the disturbing influence of incoming perturbed auditory feedback (Guenther, 2006).

The past several decades have seen an unprecedented upsurge in the number of bilinguals; however, for most bilinguals, speaking a second language is a challenging task (Bergmann et al., 2015). Typical disfluency markers include pauses, syllable repetition, and self-corrections (Götz, 2013; Kormos, 2006). Growing evidence shows that speakers are considerably less fluent in L2 compared with their L1. For example, Wiese (1984) reported that L2 speech contains two to three times as many hesitations as L1 speech. Hincks (2008) also found slower speech rates in L2 than in L1. It is well known that feedforward control is crucial for fluent speech, while excessive reliance on feedback control causes a time-lag problem because of the delay inherent in processing auditory feedback and launching corrective commands (Civier, 2010; Civier et al., 2010; Perkell, 2012). Thus, it is reasonable to hypothesize that poorer L2 fluency is correlated with heavier weighting of feedback control, and, accordingly, better L1 fluency is associated with heavier weighting of feedforward control.

This fluency-related hypothesis is supported only by indirect evidence from patients with speech motor disorders. Guenther (2016) found that patients with speech motor disorders usually have impaired feedforward control. For example, apraxia of speech, a speech motor planning and programming disorder, is most often associated with damage to the left inferior frontal gyrus, anterior insula, and/or ventral precentral gyrus. According to the DIVA model, damage to these areas affects the speech sound map and the feedforward commands for articulating speech sounds. Stuttering is also a disorder that disrupts speech fluency, but the mechanism remains controversial. Several researchers believe stuttering is a result of abnormal auditory-motor transformation in the feedback control system (Cai et al., 2012; Loucks et al., 2012). Other researchers suggest that stuttering results from a general auditory prediction deficit (Daliri & Max, 2015a, 2015b) and a heavy weight on feedback control (Civier et al., 2010; Tourville et al., 2008).

Feedforward and feedback control of voice intensity

The current study aimed to address whether the relative weighting of feedforward and feedback control varies between L1 and L2. In previous bilingualism research, investigators either compared same-language L1 and L2 speakers or compared L1 and L2 in the same bilinguals (Bergmann et al., 2015). The difficulty for intraspeaker comparisons lies in interpreting whether the observed differences are caused by language status or language differences. To avoid this confusion, we selected voice intensity to dissociate the role of language status because this attribute has few well-known language-specific phonological features.

There is a considerable amount of research describing how speakers control pitch (Chang et al., 2013; Chen et al., 2012), formant (Cai et al., 2012; Mitsuya et al., 2011), and intensity (Bauer et al., 2006; Heinks-Maldonado & Houde, 2005; Liu et al., 2007). Typically, auditory perturbations induce compensatory behaviors that change speech parameters in the opposite direction (Behroozmand et al., 2015; Chang et al., 2013). To date, previous studies have provided evidence that pitch and formant control may differ across languages. In tonal languages, such as Chinese, pitch plays a key role in differentiating meanings, while in nontonal languages, such as English, pitch only conveys stress and intonation information (Chen et al., 2012; Liu et al., 2010b; Ning et al., 2015; Ning et al., 2014). Languages also differ in the number, location, and relative proximity of vowels, thus the requirements for formant control also vary across languages (Mitsuya et al., 2011). Uniquely, voice intensity is a basic and low-level sound attribute (Tian et al., 2018) that is not highly effective for encoding linguistic contrasts (Liu et al., 2007). To date, there is no direct evidence suggesting that voice intensity is sensitive to the different languages native speakers use.

Studies have shown that online intensity control is similar to pitch and formant control (Bauer et al., 2006; Heinks-Maldonado & Houde, 2005; Liu et al., 2007). For example, Bauer and colleagues found that during vowel production, individuals demonstrated a compensatory response to unexpected intensity perturbations (200 ms, ±1, 3 vs. 6 dB SPL; see also Heinks-Maldonado & Houde, 2005). Furthermore, Liu et al. (2007) observed that Mandarin speakers also compensated for intensity perturbations (200 ms, ± 3 dB SPL) during Mandarin production. These studies imply that intensity control works to monitor and stabilize voice intensity around a desired level. In this line of research, it is assumed that speakers who rely more on feedforward control will produce speech based more on stored feedforward commands, and hence more stable vocal output. Also, speakers who rely more on feedback control will produce speech based more on auditory feedback to correct for errors, and hence are more affected by perturbations and produce a larger compensatory response (Guenther, 2006).[1] These studies addressed intensity control through real-time manipulation of speakers' original auditory feedback.

Noise experiments offer another line of intensity control research. Lombard (1911) was the first to find that speakers unconsciously increase their voice intensity to compensate for reduced audibility in a noisy environment. This phenomenon is known as the Lombard effect and has been documented in many studies (Lin et al., 2015; Patel & Schell, 2008). Noise experiments typically instruct participants to produce speech while a constant noise is added to their feedback (Bauer et al., 2006). Because speaking is a goal-oriented behavior developed to facilitate communication,

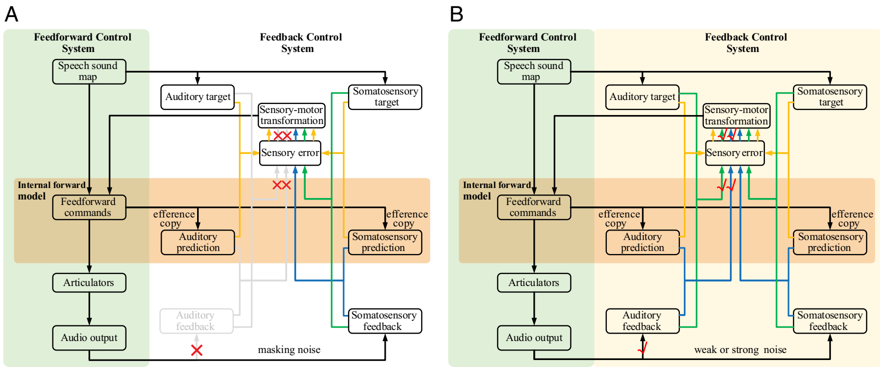

Figure 2. (A) A model for determining the voice intensity increases under a masking noise. The red crosses indicate that auditory feedback is not audible for feedback control. (B) A model for determining the voice intensity increases under a weak or strong noise. The red ticks indicate that auditory feedback is less audible but still available for feedback control. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article).

speakers usually automatically increase voice intensity to improve the signal-to-noise ratio (Liu et al., 2007). Previous noise studies demonstrated that intensity control in a noisy environment differs from that in online perturbation paradigms (i.e., to monitor and stabilize), instead it works to regulate intensity around a loudness level to make speakers heard over the noise. In other words, intensity control in online perturbation paradigms functions to monitor and stabilize, while intensity control in noisy environments functions to overcome noise (Chang-Yit et al., 1975).

Studies have addressed feedforward control by observing how speakers adapt motor commands when auditory feedback is perturbed over a long period (Ballard et al., 2018; Lametti et al., 2014). However, speech adaptation paradigms only reveal the updating of feedforward control, rather than feedforward control in and of itself. In the present study, we innovatively employed a noise-masking paradigm to investigate feedforward control of voice intensity. This paradigm is based on the premise that a loud masking noise would effectively eliminate the auditory feedback for controlling speech movements (Christoffels et al., 2007; Houde et al., 2002; Lin et al., 2015; Maas et al., 2015; Terband et al., 2015). As schematized in Figure 2A, a masking noise disrupts comparisons involving actual auditory feedback. Although it is impossible to create an experimental condition without any feedback[2] (Kent et al., 2000; Maas et al., 2015), it is reasonable to expect a much heavier reliance on feedforward control in the absence of auditory feedback (Guenther, 2006, 2016; Guenther & Vladusich, 2012).

Feedforward control incorporates a mechanism that allows speakers to make vocal adjustments independent of auditory feedback (Hickok, 2012; Parrell et al., 2019). In face of a loud masking noise, speakers evaluate the disturbance from noise signals before they speak. Considering the adverse environment, speakers retrieve the predetermined commands but will not issue them directly to the articulators to avoid obvious errors. Instead, speakers generate an auditory prediction of voice intensity based on feedforward commands and background noise. Then, speakers internally compare auditory prediction and auditory target, which further activate

https://doi.org/10.1017/S0142716420000223 Published online by Cambridge University Press

auditory error representing the noise-induced decrease in audibility. At this time, speakers launch a corrective command based on established auditory-motor transformations to surpass the masking noise. We thus predicted that speakers who rely more on feedforward control would adjust motor plans based more on predicted loss in audibility, and hence produce a larger Lombard effect than those who rely less on feedforward control.

The premise of feedback control involves speakers' perception of their auditory feedback. We thus applied noise signals where participants could hear their voice over the noise, but their voice would be less audible than what they expected to hear. The purpose of added noise was to partially mask air-conducted auditory feedback, and thus reduce the signal-to-noise ratio. As exhibited in Figure 2B, a noise that is not intense enough to mask the original auditory feedback activates feedback control comparisons. Although feedforward control also plays a role in vocal adjustments to noise signals, it was reasonable to expect increased weighting of feedback control to correct motor commands based on perceived auditory errors (Guenther, 2006; Guenther & Vladusich, 2012). We thus predicted that speakers who relied more on feedback control would adjust motor plans based more on perceived loss in audibility, and hence produce a larger Lombard effect than those who relied less on feedback control.

The current study

We designed two noise experiments to test whether the relative reliance on feedforward and feedback control was affected by language type and L2 fluency in Chinese–English bilinguals. In Experiment 1, we addressed the weighting of feedforward control by observing how bilinguals react to a masking noise (90 dB SPL multitalker noise) during L1 and L2 spoken word production. We predicted that L1 relies more on feedforward control compared with L2, so the Lombard effect would be larger in L1. We also predicted a correlation between L2 fluency and the reliance on feedforward control for L2 speakers, where more fluent bilinguals would exhibit larger Lombard effects.

In Experiment 2, we addressed the weighting of feedback control by observing how bilinguals react to a weak noise (30 dB SPL multitalker noise) and a strong noise (60 dB SPL multitalker noise) during L1 and L2 spoken word production. Although both strong noise and masking noise have a high volume, they differ in that speakers' auditory feedback is still available for feedback control under strong noise but is not available under masking noise. We predicted that L2 relies more on feedback control compared with L1, so the Lombard effect would be larger in L2. In addition, we also expected a correlation between L2 fluency and the reliance on feedback control for L2 speakers, where less fluent bilinguals would be accompanied by larger Lombard effects.

Experiment 1

Methods

Participants

Experiment 1 was completed by 24 Chinese–English bilinguals from Renmin University of China. All participants were right-handed, free of any neurological disease, and self-reported to have normal hearing. On enrolling in the study, participants were instructed to name pictures in both L1 and L2 at their habitual volume while wearing a headphone. The multitalker noise with random varying levels (30 dB, 60 dB, and 90 dB SPL) was sent to the headphone, and participants needed to judge whether they could perceive their auditory feedback under each noise level. All participants reported that they could hear their voice under 30 dB and 60 dB noise but not under 90 dB noise. This screening test was performed to ensure noise manipulation validity in the current study.

The bilinguals' (11 males) mean age was 20.3 years (SD = 2.2, range 18–28). Note that bilinguals can be classified based on age of L2 acquisition. The cut point for early and late bilinguals is learning L2 between birth and eight years, or at eight years or older (Birdsong & Molis, 2001). In the current study, we only included bilinguals who reported receiving their schooling in Chinese and being exposed to English after age 8 (see also Epstein et al., 1996). The mean age of L2 acquisition for the 24 Chinese–English bilinguals was 9.7 years (SD = 1.1, range 9–13).

Stimuli

Twenty black-and-white simple line pictures (including 15 targets and 5 practice items) were selected from a database created by Zhang and Yang (2003). The practice items were used to familiarize participants with the experimental procedure and were not employed in the formal experiment. All pictures referred to common objects and had good indexes in visual complexity, familiarity, and image agreement. All pictures had monosyllabic names in both Chinese and English (e.g., Chinese: "猫" /mao/; English: cat). We employed a 90 dB SPL multitalker noise to mask participants' auditory feedback (Patel & Schell, 2008).

Design

The experiment adopted a 2 (language: L1 and L2) × 2 (noise condition: quiet and masking noise) within-subjects and within-items design. Within a block, participants named 15 target pictures consecutively in each experimental condition, for a total of 15 trials. The order of blocks (L1-quiet, L1-masking noise, L2-quiet, L2-masking noise) was randomized. Participants finished three blocks for each experimental condition, generating a total of 180 trials. The order of items was randomized within L1 blocks but pseudo-randomized in L2 blocks to ensure that a target would not follow a target with the same initial phoneme (e.g., ball, book) to avoid a phonological facilitation effect. A new order was generated for each participant and for each block.

Apparatus

The auditory experiment was conducted in a soundproof room and controlled by E-Prime Professional Software (version 2.0; Psychology Software Tools). Naming latencies were recorded from target presentation using a voice-key, connected to the computer using a PST Serial Response Box. This multitalker noise was calibrated through an audiometer (SMART SENSOR AS804) and presented to participants

https://doi.org/10.1017/S0142716420000223 Published online by Cambridge University Press

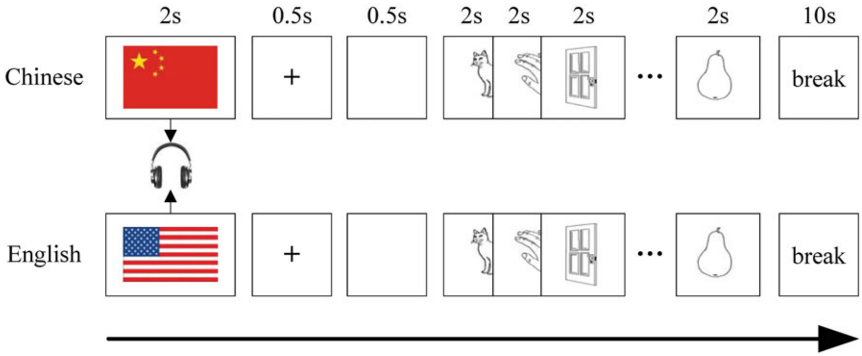

Figure 3. A schematic diagram of the sequence for a block (upper panel: L1 picture naming; lower panel: L2 picture naming).

using a supra-aural headphone (Bose QuietComfort35 II). The participants' speech was collected and recorded with an external condenser microphone (SHURE SM58S) connected to a YAMAHA Steinberg CI1 soundcard. The microphone was fixed on a short microphone holder standing on the desk and secured at 10 cm from the subjects' mouth. The target words were extracted and saved as separate WAV files. The recorded speech signals were analyzed with the Praat speech analysis software (version 6.0.43; Boersma & Weenink, 2013). The syllabic boundaries of all words were labeled by hand and the vocal cycles were hand-checked for errors such as a miss or double mark. A custom-written Praat script was used to extract the mean intensity of each syllable.

Procedure

Participants were tested individually. First, participants were asked to familiarize themselves with target pictures by viewing each target for 2000 ms with the picture name printed below. After the learning phase, participants received a picture-naming test without concurrently presented names. When the experimenter determined that participants named all pictures correctly in both L1 and L2, the practice blocks were administered. In the practice phase, participants finished one block composed of five practice pictures for each experimental condition. The practice blocks procedure was identical to the experimental blocks procedure, except for the number of pictures. When the experimenter determined that participants understood the naming task instructions, the experimental blocks were administered.

Figure 3 is a schematic representation of the sequence for a block. At the beginning, a flag was presented for 2 seconds to cue the target language in the block. Meanwhile, the noise signals were played continuously in the masking noise condition but remained silent in the quiet condition. Then a fixation point (+) appeared in the middle of the screen for 500 ms, followed by a blank screen. Next, 15 pictures were presented on the screen, 2 seconds apart. Participants were asked to name the picture as quickly and accurately as possible. We stopped playing

https://doi.org/10.1017/S0142716420000223 Published online by Cambridge University Press

noise signals after participants finished naming 15 pictures. Finally, a break lasting 10 seconds concluded each block.

L2 fluency test

We included two English speaking tests to measure participants' L2 fluency. Previous research typically addressed L2 fluency by measuring temporal features, such as speech rate, duration and rate of hesitations, and filled and silent pauses (Hilton, 2014; Kormos, 2006; Segalowitz, 2010). The current study measured L2 fluency using speech rate indicated by the length of time required to complete the speaking tasks, with shorter time indexing more fluent L2 speech and longer time indexing less fluent L2 speech.

The order of speaking tasks was as follows. First, participants completed a rapid automated naming task in which four 7 × 7 item grid stimulus displays were created for computer presentation. Each grid consisted of one of the following randomly ordered stimulus types: letters (g, k, m, r), objects (pictures depicting a dog, chair, bed, or key), colors (boxes colored red, blue, yellow, or green), and digits (2, 4, 6, 9). Participants were instructed to sequentially name aloud each item in the grid from the top left item to the bottom right item as quickly as possible without errors. This was repeated for each stimulus grid (letters, objects, colors, and digits), and the task order was counterbalanced across participants. The experimenter manually pressed the mouse to start and end the timing procedure for each grid. The final rapid naming duration was the average of each grid's duration for letter, object, color, and digit-naming tasks.

Next, participants completed a passage reading task in which four English passages were extracted from *New Concept English Two*. They were instructed to sequentially read aloud each passage at their habitual speed. The experimenter manually pressed the mouse to start and end the timing procedure for each passage reading. The final duration for passage reading was the average of the four passage reading durations. The English fluency tests were performed using E-prime Professional Software.

Results

Two participants were excluded; one could not tolerate the loud masking noise and quit the experiment, and the other's voice intensity in the quiet and masking noise conditions differed by more than two standard deviations from the group mean (see Lametti et al., 2014 for a similar data removal procedure). The remaining data from 22 participants were included in the subsequent analyses. Table 1 presents the mean picture-naming reaction time, error percentages, and mean intensity by language and noise condition.

We used the *lmer* program of the *lme4* package (Baayen et al., 2008; Bates, 2005; Bates et al., 2014) in R software (R Core Team, 2015) to estimate fixed and random effects. The data (i.e., response time, the percentage of error responses, and mean intensity) were analyzed using a linear mixed-effects model with language and noise condition as the fixed factors and participants and items as the random factors. Models used restricted maximum likelihood estimation to find the optimal

Table 1. Mean picture-naming response time (RT, in ms), percentage of errors (PE, %), mean intensity (MI, in dB), and standard deviations (SD, in parenthesis) as a function of language and noise condition in Experiment 1

	L1			L2		
	RT (SD)	PE (SD)	MI (SD)	RT (SD)	PE (SD)	MI (SD)
Quiet	688 (114)	0.51 (0.83)	48.50 (4.16)	706 (96)	0.40 (0.86)	48.46 (4.42)
Masking noise	635 (97)	0.91 (1.09)	58.80 (3.77)	661 (80)	1.41 (1.07)	58.13 (3.93)

parameter estimation of the best-fitting model to the observed data. The best-fitting model was defined as the most complex model that significantly improved the variance estimation over previous models. Model fitting included three steps: specifying a model (i.e., null model) that included only random factors (participants and items); enriching the null model by adding fixed factors (i.e., language and noise condition) one by one, and then by adding the two-way interactions between the two factors; and comparing the newly established model to the previous model using the chi-square test. If adding a fixed factor or an interaction to the existing model did not significantly improve the variance estimation ($p > 0.05$), then the current model was designated the best-fitting model.

Behavioral results

Data from incorrect responses (0.81%), naming latencies longer than 1500 ms or shorter than 200 ms (2.25%), and latencies deviating two standard deviations from each participant's mean (6.14%) were removed from the behavioral analyses. For response time, the best-fitting model only included the factors of language and noise condition (see Table 2). Adding the two-way interaction between language and noise condition did not significantly improve the fit, $\chi2\,(1, 3596) = 1.15, p = 0.28$. A parallel analysis was conducted on the errors, but a binomial family was used because of the binary nature of the response. The best-fitting model only included the factor of noise condition; adding language, $\chi2\,(1, 3928) = 0.51, p = 0.48$, and the interaction between language and noise condition, $\chi2$ (1, 3928) = 1.22, $p = 0.54$, did not significantly improve the fit.[3] Table 2 displays parameter estimates for the fixed effects for response time, percentage of errors, and mean intensity.

Acoustic analysis

Only data from incorrect responses (0.81%) were removed from the acoustic analyses. Figure 4A illustrates the mean intensity score distribution from 22 participants in each experimental condition. For mean intensity, the best-fitting model included the noise condition and the interaction between language and noise condition (see Table 2). Adding the factor of language did not significantly improve the fit, $\chi2\,(1, 3928) = 2.53, p = 0.11$. The two-way interaction is interesting; simple analyses indicated that masking noise increased speakers' voice intensity relative to the quiet condition in both L1 ($\beta = 10.05$, $t = 83.10$, $p < 0.001$) and L2 word production ($\beta = 9.94$, $t = 73.25$, $p < 0.001$), but the intensity increase was larger in L1 than L2 (see Figure 4B). As shown in Figure 4C, simple analyses in the other direction

https://doi.org/10.1017/S0142716420000223 Published online by Cambridge University Press

Table 2. LMM estimates of fixed effects for picture-naming response time (RT), percentage of errors (PE), and mean intensity (MI) in Experiment 1

	Measure											
	RT				PE				MI			
Fixed effects	*β*	*SE*	*t*	*p*	*β*	*SE*	*z*	*p*	*β*	*SE*	*t*	*p*
(Intercept)	687.76	21.90	31.36	<0.001	5.39	0.33	16.13	<0.001	48.55	0.85	57.23	<0.001
Language2	22.59	3.46	6.54	<0.001	–	–	–	–	–	–	–	–
Noise2	−49.19	3.46	−14.23	<0.001	−0.95	0.39	-2.40	0.02	102.93	3.31	31.10	<0.001
Language2:Noise2	–	–	–	–	–	–	–	–	−0.52	0.14	−3.71	<0.001

Note: Language2, L2; Noise2, masking noise.

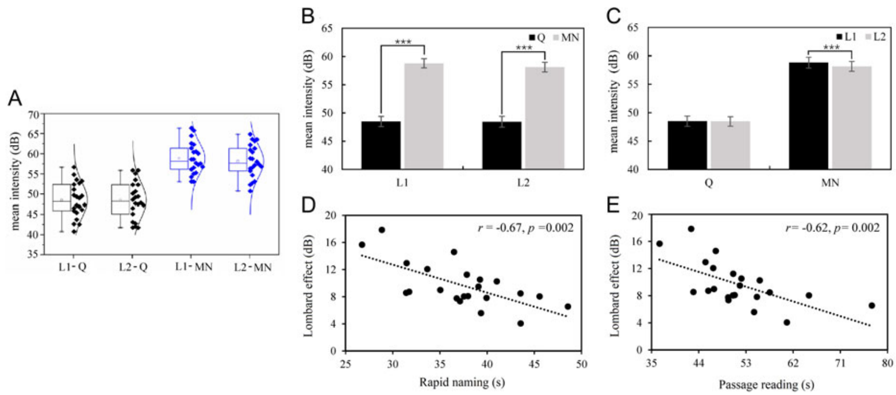

Figure 4. Results in Experiment 1. (A) Box plots illustrating the distribution of average mean intensity scores of 22 participants in each experimental condition. Box definitions: middle line is the median, top and bottom of boxes are 75th and 25th percentiles, and square is the mean. (B) Column charts of the mean intensity (mean and standard error) in the L1 and L2 speech production as a function of noise condition. (C) Column charts of the mean intensity (mean and standard error) in the quiet (Q) and masking noise (MN) conditions as a function of language. Asterisks indicate the significant effects. (D) The scatterplot for the correlation between rapid naming and the Lombard effect. (E) The scatterplot for the correlation between passage reading and the Lombard effect. Here, the Lombard effect is defined as the difference between the mean intensity in L2-quiet and L2-masking noise conditions.

indicated that in the quiet condition, the mean intensity did not differ between L1 and L2 word production ($\beta = -0.13$, $t = -1.06$, $p = 0.29$), but in the masking noise condition, the mean intensity was significantly higher in L1 than L2 word production ($\beta = -0.52$, $t = -4.51$, $p < 0.001$).

Correlation analysis between the Lombard effect and L2 fluency

To test whether more fluent L2 speakers rely more on feedforward control than less fluent L2 speakers, we examined the relationship between the Lombard effect in L2 spoken word production and the fluency performance in L2 rapid naming and passage reading. Data from 22 participants in Experiment 1 were entered into the Pearson's correlation analysis. Here, the Lombard effect was defined as the difference between the mean intensity in the L2-quiet and L2-masking noise conditions. In addition, L2 fluency was measured by two English production tasks and defined as the average durations for rapid naming and passage reading, respectively. The results indicated that the Lombard effect was negatively correlated with the duration for L2 rapid naming task ($r = -0.67$, 95% CI [–0.36, –0.83], $p = 0.002$) and the duration for L2 passage reading task ($r = -0.62$, 95% CI [–0.42, –0.80], $p = 0.002$). This suggests that the more fluently bilinguals speak in their L2, the larger Lombard effect they exhibit in L2 speech production (see Figures 4D and 4E).

Discussion

To the best of our knowledge, this was the first cross-language study to compare feedforward control in a group of Chinese–English bilinguals using a masking noise.

https://doi.org/10.1017/S0142716420000223 Published online by Cambridge University Press

A 90 dB SPL multitalker noise virtually eliminated speakers' auditory feedback, thus the mechanism of feedback-based motor correction had little role to play, whereas the predictive feedforward control dominated the speech motor control. Notably, we observed that the Lombard effect elicited by a masking noise was larger in L1 than L2 word production. In addition, correlation analyses showed that the Lombard effect in L2 word production was larger in more fluent L2 speakers than less fluent ones. Thus, the results support our two hypotheses that bilinguals' feedforward control is influenced by language and related to L2 fluency, where a heavier weighting was assigned to feedforward control in the L1 production system and more fluent L2 speakers than to the L2 production system and less fluent L2 speakers.

In Experiment 2, we adjusted the levels of multitalker noise signal from 90 dB to either 60 dB or 30 dB to enable the involvement of auditory feedback in speech motor control. By measuring the magnitude of the Lombard effect in response to a noise that was not as loud as the masking noise, we examine bilinguals' relative reliance on feedback control in L1 and L2 spoken word production.

Experiment 2

Method

Participants

Participants in Experiment 2 were the same as those in Experiment 1. The order of the two experiments was counterbalanced between participants, with half of the participants completing Experiment 2 after Experiment 1 and the other half completing Experiment 1 after Experiment 2. The interval between the two experiments was about 15 minutes (5 minutes' short break plus 10 minutes' L2 fluency tests). This within-subjects practice not only maximized the sensitivity to compare between experiments but also confirmed that result differences were unrelated to individual differences.

Stimuli

The picture stimuli were the same as those in Experiment 1. Previous works suggest that the magnitude of the Lombard effect is influenced by noise level. For example, Patel and Schell (2008) manipulated noise condition by using quiet, 60 dB, and 90 dB multitalker noise and observed more voice intensity increases when the background noise was 90 dB compared to 60 dB. Following their practice, we also included a 60 dB multitalker noise and added a 30 dB multitalker noise to further investigate how proportional changes in noise level affect vocal adjustments in voice intensity. To differentiate from the masking noise in Experiment 1, we called the 60 dB multitalker noise the strong noise and the 30 dB multitalker noise the weak noise.

Design

Experiment 2 adopted a 2 (language: L1 and L2) × 3 (noise condition: quiet, weak noise, and strong noise) within-subjects and within-items design. Within a block, participants named 15 target pictures consecutively in each experimental condition,

https://doi.org/10.1017/S0142716420000223 Published online by Cambridge University Press

Table 3. Mean picture-naming response time (RT, in ms), percentage of errors (PE, %), mean intensity (MI, in dB), and standard deviations (SD, in parenthesis) as a function of language and noise condition in Experiment 2

	L1			L2		
	RT (SD)	PE (SD)	MI (SD)	RT (SD)	PE (SD)	MI (SD)
Quiet	673 (98)	0.51 (0.93)	48.90 (3.85)	703 (63)	0.71 (1.04)	48.77 (4.13)
Weak noise	627 (92)	0.40 (0.86)	54.87 (3.97)	655 (63)	1.11 (1.30)	55.24 (3.70)
Strong noise	634 (70)	0.91 (1.09)	56.07 (4.32)	656 (66)	1.62 (1.20)	57.33 (3.74)

for a total of 15 trials. The order of blocks (L1-quiet, L1-weak noise, L1-strong noise, L2-quiet, L2-weak noise, L2-strong noise) was randomized. Participants finished three blocks for each experimental condition, generating a total of 270 trials.

Apparatus and procedure
Identical to Experiment 1.

Results
Data from 22 participants were entered into the final analyses using the same LMM estimation. Table 3 presents the mean picture-naming response time, percentage of errors, and mean intensity by language, and noise condition.

Behavioral results
Data from incorrect responses (0.88%), naming latencies longer than 1500 ms or shorter than 200 ms (2.19%), and latencies deviating two standard deviations from each participant's mean (5.49%) were removed from all analyses. For response time, the best-fitting model only included the factors of language and noise condition (see Table 4). Adding the two-way interaction between language and noise condition did not significantly improve the fit, $\chi2$ (2, 5432) = 2.93, $p = 0.23$. For the percentage of errors, the best-fitting model only included the factor of language; adding noise condition, $\chi2$ (2, 5888) = 5.21, $p = 0.07$, and the interaction between language and noise condition, $\chi2$ (2, 5888) = 5.93, $p = 0.20$, did not significantly improve the fit.[4] Table 4 displays parameter estimates for fixed effects for response time, percentage of errors, and mean intensity.

Acoustic analysis
Only data from incorrect responses (0.88%) were removed from the acoustic analyses. Figure 5A illustrates average mean intensity score distribution of the 22 speakers in each experimental condition. For mean intensity, the best-fitting model included language and noise condition as well as their interaction (see Table 4). To examine the two-way interaction, simple analyses indicated that both weak noise (L1: $\beta = 5.95$, $t = 52.29$, $p < 0.001$; L2: $\beta = 6.40$, $t = 53.46$, $p < 0.001$) and strong noise (L1: $\beta = 6.40$, $t = 53.46$, $p < 0.001$; L2: $\beta = 6.40$, $t = 53.46$, $p < 0.001$) increased speakers' voice intensity relative to the quiet condition, but

https://doi.org/10.1017/S0142716420000223 Published online by Cambridge University Press

Table 4. LMM estimates of fixed effects for picture-naming response time (RT), percentage of errors (PE), and mean intensity (MI) in Experiment 2

	Measure											
	RT				PE				MI			
Fixed effects	*β*	*SE*	*t*	*p*	*β*	*SE*	*z*	*p*	*β*	*SE*	*t*	*p*
(Intercept)	673.34	18.12	37.17	<0.001	5.20	0.28	18.82	<0.001	48.98	0.85	57.32	<0.001
Language2	28.85	2.77	10.42	<0.001	−0.64	0.29	−2.22	0.03	−0.18	0.12	−1.41	0.16
Noise2	−46.49	3.40	−13.68	<0.001	–	–	–	–	5.95	0.13	47.56	<0.001
Noise3	−41.98	3.41	−12.31	<0.001	–	–	–	–	7.12	0.12	57.09	<0.001
Language2:Noise2	–	–	–	–	–	–	–	–	0.46	0.18	2.58	0.01
Language2:Noise3	–	–	–	–	–	–	–	–	1.39	0.18	7.88	<0.001

Note: Language2, L2; Noise2, weak noise, Noise3, strong noise.

https://doi.org/10.1017/S0142716420000223 Published online by Cambridge University Press

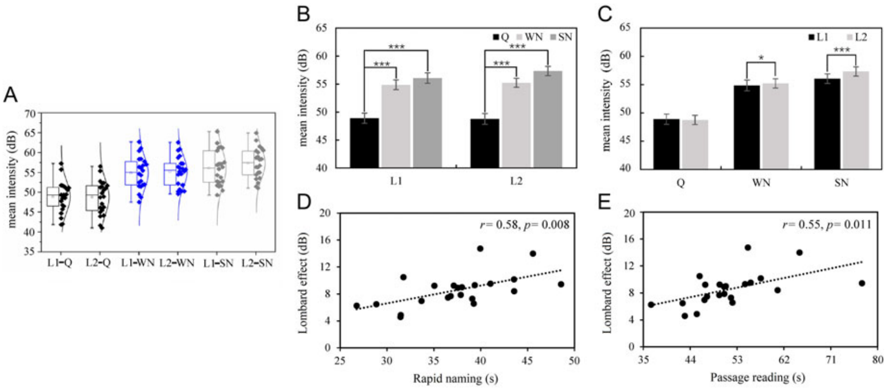

Figure 5. Results in Experiment 2. (A) Box plots illustrating the distribution of average mean intensity scores of 22 participants in each experimental condition. Box definitions: middle line is the median, top and bottom of boxes are 75th and 25th percentiles, and square is the mean. (B) Column charts of the mean intensity (mean and standard error) in the L1 and L2 speech production as a function of noise condition. (C) Column charts of the mean intensity (mean and standard error) in the quiet (Q), weak noise (WN) and strong noise (SN) conditions as a function of language. Asterisks indicate the significant effects. (D) The scatterplot for the correlation between rapid naming and the Lombard effect. (E) The scatterplot for the correlation between passage reading and the Lombard effect. Here, the Lombard effect is defined as the difference between the mean intensity in L2-quiet and L2-strong noise conditions.

the intensity increase was larger in L2 than L1 (see Figure 5B). As shown in Figure 5C, simple analyses in the other direction also indicated that the mean intensity was not different between L1 and L2 word production in the quiet condition ($\beta = -0.15$, $t = -1.24$, $p = 0.22$), but the mean intensity was significantly higher in L2 than L1 word production in the weak noise condition ($\beta = 0.27$, $t = 2.24$, $p = 0.03$), and the strong noise condition ($\beta = 1.18$, $t = 10.64$, $p < 0.001$).[5]

Correlation analysis between the Lombard effect and L2 fluency

To test whether less fluent L2 speakers rely more on feedback control than more fluent L2 speakers, we examined the relationship between the Lombard effect in L2 spoken word production and the fluency performance in L2 rapid naming and passage reading. Data from 22 participants in Experiment 2 were entered into the Pearson's correlation analysis. The Lombard effect was defined as the difference between the mean intensity in the L2-quiet and L2-strong noise conditions. The results indicated that the L2 Lombard effect was positively correlated with the duration for L2 rapid naming task ($r = 0.58$, 95% CI [0.30, 0.81], $p = 0.005$) and the duration for L2 passage reading task ($r = 0.55$, 95% CI [0.31, 0.81], $p = 0.009$). This suggests that the less fluently bilinguals speak in their L2, the larger Lombard effect they exhibit in L2 speech production (see Figure 5D and 5E).

Discussion

In Experiment 2, we observed that the Lombard effect elicited by a weak or strong noise was larger in L2 than L1 word production. In addition, correlation analyses suggest that the Lombard effect in L2 word production was larger in less fluent L2

https://doi.org/10.1017/S0142716420000223 Published online by Cambridge University Press

speakers than more fluent L2 speakers. Thus, the results lend support to the hypotheses that bilinguals' feedback control is also affected by language and related to L2 fluency, where a heavier weighting is assigned to feedback control in the L2 production system and in less fluent L2 speakers compared to the L1 production system and more fluent L2 speakers.

Our findings suggest that the mean intensity did not differ between L1 and L2 word production under the quiet condition in either Experiment 1 or Experiment 2. The similar characteristic of voice intensity is of great importance in the L1-L2 contrast of two different languages; we thus suggest that the observed vocal changes indeed resulted from the experimental manipulations rather than the languages.

General discussion

The purpose of the two experiments was to systematically determine the relative weighting of feedforward and feedback control in bilinguals' L1 and L2 speech production, and to evaluate whether individual differences in L2 fluency are related to the organisation of feedforward and feedback control in the L2 speech motor system. We manipulated the noise level mixed with the auditory feedback that participants received while speaking. When the noise intensity (90 dB multitalker noise) exceeds a masking threshold where participants could not perceive their original auditory feedback, bilinguals showed a larger Lombard effect in L1 than in L2 word production. In addition, as L2 fluency increased, the Lombard effect in L2 word production also increased. In contrast, when the noise intensity (30 dB or 60 dB multitalker noise) was below the masking threshold but hampered speech intelligibility, the same bilinguals showed increased Lombard effect in L2 word production compared to L1 word production. In addition, as L2 fluency decreased, the Lombard effect in L2 word production increased. The overall results indicate that compared to L1, L2 speech motor control relies on feedforward control to a lesser extent but relies on feedback control to a greater extent. Also, the correlation findings provide initial evidence in second language learners that L2 speech rapidity is related to higher weighting of feedforward control but lower weighting of feedback control.

Feedforward control between L1 and L2

We investigated bilinguals' feedforward control using a masking noise in Experiment 1, and, for the first time, we observed that bilinguals exhibited a larger Lombard effect in L1 word production than L2 word production, reflecting that L1 speech motor execution relies more on feedforward control compared to L2. A previous study provided neuro-imaging evidence to differentiate native and novel speech production in terms of feedforward control (Moser et al., 2009). According to the DIVA model, the left inferior frontal gyrus and anterior insula are important brain regions involved in feedforward control (Guenther, 2016; Kearney & Guenther, 2019). Damage to these areas typically cause a disorder in motor speech planning. In Moser et al.'s study (2009), 30 normal adults completed a speech production task consisting of two types of three-syllable nonwords: English (native) syllables and non-English (novel) syllables. The authors found that when novel syllable production was compared to native syllable production, greater activations were observed in an

https://doi.org/10.1017/S0142716420000223 Published online by Cambridge University Press

extensive neural network including the left inferior frontal gyrus and anterior insula. Of close relevance, they speculated that increased activity in motor speech networks may directly reflect unfamiliarity with the motor commands necessary for target sounds; that is, a difference in feedforward control. In our study, we cannot explain the neural mechanism of a feedforward deficit (Alm, 2004, 2005; Kearney & Guenther, 2019) or its exact nature (Civier, 2010). Although L2 words were not novel speech sounds, they were not as familiar as the L1 counterparts to bilingual speakers (as reflected by longer naming latencies). Thus, it is not surprising that a difference in feedforward control was found between L1 and L2 speech motor control in Experiment 1.

For bilinguals, L1 is an overlearned language. The feedforward commands that store detailed instructions for how to move the articulators to achieve a linguistic goal, should be directly read out from "mental syllabary" without effort, with its mechanism similar to singing a familiar song from memory (Civier et al., 2010). Speakers in a highly automatic language have established accurate auditory-motor bidirectional mappings; not only can they predict auditory consequences based on an efference copy of the motor commands and environmental influence but they can also issue motor commands based on the intended auditory consequences. By accurately measuring the level of masking noise, native speakers adjust their voice intensity more to make themselves heard. However, for second language learners, articulation of L2 words is less rehearsed (Parker Jones et al., 2012) due to factors such as the age of acquisition, the amount of exposure, and the involvement in daily life (Abutalebi et al., 2001), so they are less likely to generate long-term representations of L2 words that are as accurate as their L1 counterparts in the "mental syllabary." Thus, when L2 speakers face a loud masking noise, they show smaller intensity adjustments to compensate for the inaudibility.

Theoretical frameworks contend that feedforward control can be accomplished quickly by preventing additional processing of sensory feedback (Guenther, 2016; Perkell, 2012). Thus, it is reasonable to associate speed of speech with the relative weighting of feedforward control. Many patient studies also found that brain damage related to feedforward control causes significant motor impairment (Kearney & Guenther, 2019). In Experiment 1, we found a negative correlation between L2 fluency and the Lombard effect in L2 speech production, suggesting that more fluent L2 speakers have superior feedforward control ability. This finding provided additional evidence for a fluency-related hypothesis in normal L2 speakers. Speech control models in native language assume that feedforward control weighting is increased as language acquisition progresses (Tourville & Guenther, 2011). Recent research findings highlight L2 fluency as a reliable predictor of L2 proficiency (De Jong et al., 2012), thus our study also shows that feedforward control is increased as second language learning progresses. Although L2 speech production is inferior in feedforward control compared with L1, we should be optimistic about the difference because, with increasing L2 proficiency, speech control may develop on a continuum, biasing away from feedback control and toward feedforward control, allowing for more native-like speech production.

Feedback control between L1 and L2

We investigated bilinguals' feedback control under weak noise and strong noise conditions. Contrary to Experiment 1, we observed that bilinguals exhibited larger Lombard effects in L2 word production than in L1 word production, and the effect was magnified in the strong noise condition relative to the weak noise condition. These contrasting findings are interesting because both experiments introduced noise to interfere with the perception of auditory feedback, and the only striking difference was that neither the weak noise nor the strong noise were loud enough to eliminate the auditory feedback needed for feedback control. Thus, the difference in noise levels (weak and strong noise vs. masking noise) was not only quantitative but also qualitive. Notably, in Experiment 2, the strong noise decreased the signal-to-noise ratio to a greater extent than the weak noise, but both noise levels elicited the same result patterns despite a difference in magnitude. Thus, the difference in noise levels (weak noise vs. strong noise) was only quantitative, not qualitive. By controlling for other factors, we suggest that L2 speech motor execution relies more on feedback control compared to L1.

The finding of language-specific feedback control echoes an early study by Mackay (1970), who employed a delayed auditory feedback technique to interfere with normal speech production and found that artificial disfluency was more serious in L2 speech production for both German–English bilinguals and English–German bilinguals. These findings provided direct evidence that the feedback control difference was unrelated to the language but related to language status. Future studies should investigate the influence of masking noise, weak noise, and strong noise on speech motor control in a group of English–Chinese bilinguals.

In addition, Simmonds et al.'s (2011b) brain imaging study differentiated L1 and L2 speech production in terms of feedback control. According to the DIVA model, the auditory and somatosensory association cortices are important brain regions involved in feedback control (Guenther, 2016; Kearney & Guenther, 2019). A perturbation to speakers' auditory feedback typically results in increased neural activities in these areas (Tourville et al., 2008; Toyomura et al., 2007). In Simmonds et al.'s (2011b) study, bilinguals produced overt propositional speech (i.e., defined visually presented pictures) in both their L1 and L2. The results provided reliable evidence of increased activations for L2 relative to L1 within the temporoparietal cortex. Of close relevance, they attributed the increased temporoparietal cortex activity to more taxing sensory monitoring of any discrepancies between the predicted and actual sensory outcomes in L2 production. Thus, it is not surprising that our study found a difference in feedback control between L1 and L2 production.

Previous research has shown that reliance on feedback control is a dynamic process in nature that ranges from heavily to rarely dependent through vocal development (Civier et al., 2010; Scheerer et al., 2013; Schmidt & Lee, 2005). The transition is modulated by practice or experience (Guenther et al., 2006). For L1 speakers, the brain has already internalized the relationships between speech movements and the desired auditory feedback during the process of language acquisition; thus, the additional information provided by auditory feedback becomes redundant. However, for L2 speakers, the mapping between motor commands and their sensory consequences is less reliable, as evidenced by larger vocal variability (Chen et al., 2001; Ng et al., 2008; Wang & van Heuven, 2006). Thus, auditory feedback is still required to retune and strengthen the motor-sensory transformations. Growing evidence also suggests that L2 speech output needs more careful monitoring to avoid errors (Ganushchak & Schiller, 2009; Parker Jones et al., 2012). Overall, the feedback subsystem may have a more prominent role to play in L2 speech motor control.

Overreliance on feedback control may introduce disfluency problems because a feedback-based strategy is relatively slow to detect and correct errors (Parrell et al., 2019; Perkell, 2012). Thus, it is reasonable to associate disfluency with the relative weighting of feedback control. Civier and colleagues (2010) also found that people who stutter may suffer from a motor strategy that weights too much toward auditory feedback control, leading to a higher probability of triggering a repetition, resulting in more stuttering. In Experiment 2, we found a positive correlation between L2 fluency and the Lombard effect in L2 speech production, indicating that less fluent L2 speakers are more dependent on feedback control. The findings in Experiments 1 and 2 are important complement to each other because we showed that increasing efficiency of L2 speech motor control is related to a bias away from feedback control and toward feedforward control in the same group of bilinguals. A large body of literature indicates that reliance on feedback control decreases as language acquisition progresses (Liu et al., 2010a; Scheerer et al., 2013; Tourville & Guenther, 2011). Concerning the close relationship between L2 fluency and L2 proficiency, our study also suggests that feedback control plays a less prominent role as second language learning progresses. This finding suggests that differences between native speakers and L2 learners are not always ever-lasting; it is possible that L2 learners can reach native-like efficiency of speech motor control.

Conclusion

In summary, our findings suggest that voice intensity control in bilinguals' speech production requires a joint effort of feedforward and feedback subsystems, and the relative weighting of feedforward and feedback control depends on whether bilinguals are producing words in L1 or L2. The correlation analyses suggest a close relationship between L2 fluency and the organization of feedforward and feedback control. Although more work is needed to establish these finding in different populations with improved methodologies, this study opens a potential new line of research into bilinguals' speech motor control.

Acknowledgments. This work was supported by the Key Project of Beijing Social Science Foundation in China (16YYA006) to Qingfang Zhang, the Fundamental Research Funds for the Central Universities, and the Research Funds of Renmin University of China (18XNLG28) to Qingfang Zhang.

Notes

1. Researchers investigate the relative weighting of feedback control by measuring the magnitude of compensatory response, with larger response indexing heavier reliance on feedback control (Scheerer & Jones, 2014). However, the reliance on feedforward control cannot be addressed directly but inversely speculated by measuring the reliance on feedback control, with higher weighting of feedback control indexing lower weighting of feedforward control.

2. It needs to be highlighted that somatosensory feedback control is still at work under masking noise because auditory feedback is not the sole input perceived by sensory system (Lametti et al., 2012). In

https://doi.org/10.1017/S0142716420000223 Published online by Cambridge University Press

addition, the perception of auditory feedback is indeed a very complex process, involving air conduction and more peripheral bone conduction (Howell & Powell, 1984). Researchers usually minimize the influence of bone-conducted auditory feedback using loud noise (Christoffels et al., 2007), whispered speech (Houde & Jordan, 2002; Zheng et al., 2010), or acoustic calibration that provides the feedback with a sound pressure level gain of 10 dB relative to participants' vocal output (Ballard et al., 2018; Chen et al., 2015). In the current experiment, a multitalker noise was presented at a high volume to effectively mask air-conducted feedback and partially mask bone-conduced feedback.

3. The results indicated that there was no interaction between language and noise condition, reflecting that the influence of noise on cognitive processing before articulation was not different between L1 and L2.

4. The results indicated that there was no interaction between language and noise condition, reflecting that the influence of noise on cognitive processing before articulation was not different between L1 and L2.

5. The experiment was performed with 30 dB and 60 dB white noise as well, and that the same pattern of results was obtained.

References

Abutalebi, J., Cappa, S. F., & Perani, D. (2001). The bilingual brain as revealed by functional neuroimaging. *Bilingualism: Language and Cognition*, **4**, 179–190.

Alm, P. A. (2004). Stuttering and the basal ganglia circuits: A critical review of possible relations. *Journal of Communication Disorders*, **37**, 325–369.

Alm, P. A. (2005). *On the causal mechanisms of stuttering.* Lund University.

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, **59**, 390–412.

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.

Bates, D. M. (2005). Fitting linear mixed models in R. *R News*, **5**, 27–30.

Bates, D. M., Maechler, M., Bolker, B., & Walker, S. (2014). lme4: Linear mixed-effects models using Eigen and S4. *R package version 1*, 1–7.

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**, 2363–2371.

Bays, P. M., & Wolpert, D. M. (2006). Computational principles of sensorimotor control that minimize uncertainty and variability. *Journal of Physiology*, **578**, 387–396.

Behroozmand, R., Shebek, R., Hansen, D. R., Oya, H., Robin, D. A., Howard III, M. A., & Greenlee, J. D. (2015). Sensory-motor networks involved in speech production and motor control: An fMRI study. *NeuroImage*, **109**, 418–428.

Bergmann, C., Sprenger, S. A., & Schmid, M. S. (2015). The impact of language co-activation on L1 and L2 speech fluency. *Acta Psychologica*, **161**, 25–35.

Birdsong, D., & Molis, M. (2001). On the evidence for maturational constraints in second-language acquisition. *Journal of Memory and Language*, **44**, 235–249.

Boersma, P., & Weenink, D. (2013). Praat: Doing Phonetics by Computer [Computer program]. http://www.praat.org

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

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**, 2653–2658.

Chang-Yit, R., Pick, H. L., & Siegel, G. M. (1975). Reliability of sidetone amplification effect in vocal intensity. *Journal of Communication Disorders*, **8**, 317–324.

Chen, Y., Robb, M. P., Gilbert, H. R., & Lerman, J. W. (2001). Vowel production by Mandarin speakers of English. *Clinical Linguistics & Phonetics*, **15**, 427–440.

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

Christoffels, I. K., Formisano, E., & Schiller, N. O. (2007). Neural correlates of verbal feedback processing: An fMRI study employing overt speech. *Hum Brain Mapping*, **28**, 868–879.

Civier, O. (2010). *Computational modeling of the neural substrates of stuttering and induced fluency* (Unpublished doctoral dissertation, Boston University).

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**, 246–279.

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.

De Jong, N. H., Steinel, M. P., Florijn, A. F., Schoonen, R., & Hulstijn, J. H. (2012). Facets of speaking proficiency. *Studies in Second Language Acquisition*, **34**, 5–34.

Epstein, S., Flynn, S., & Martohardjono, G. (1996). Second language acquisition: Theoretical and experimental issues in contemporary research. *Behavioral and Brain Sciences*, **19**, 677–714.

Ganushchak, L. Y., & Schiller, N. O. (2009). Speaking one's second language under time pressure: An ERP study on verbal self-monitoring in German–Dutch bilinguals. *Psychophysiology*, **46**, 410–419.

Götz, S. (2013). *Fluency in native and nonnative English speech.* John Benjamins.

Grosjean, F. (2010). *Bilingual.* Harvard University Press.

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

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

Guenther, F. H., & Vladusich, T. (2012). A neural theory of speech acquisition and Production. *Journal of Neurolinguistics*, **25**, 408–422.

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**, 280–301.

Heinks-Maldonado, T. H., & Houde, J. F. (2005). Compensatory responses to brief perturbations of speech amplitude. *Acoustics Research Letters Online*, **6**, 131–137.

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

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

Hilton, H. (2014). Oral fluency and spoken proficiency: Considerations for research and testing. In Pascale Leclercq, Amanda Edmonds, & Heather Hilton (Eds.), *Measuring L2 proficiency: Perspectives from SLA* (pp. 27–53). Multilingual Matters.

Hincks, R. (2008). Presenting in English or Swedish: Differences in speaking rate. *Proceedings of FONETIK 2008* (pp. 21–24). Department of Linguistics, Gothenburg University.

Houde, J. F., & Jordan, M. I. (2002). Sensorimotor adaptation of speech I: Compensation and adaptation. *Journal of Speech Language and Hearing Research*, **45**, 295–310.

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**, 1125–1138.

Howell, P., & Powell, D. J. (1984). Hearing your voice through bone and air: Implications for explanations of stuttering behaviour from studies of normal speakers. *Journal of Fluency Disorders*, **9**, 247–264.

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

Kearney, E., & Guenther, F. H. (2019). Articulating: The neural mechanisms of speech production. *Language, Cognition and Neuroscience*, **34**, 1214–1229.

Kent, R. D., Kent, J. F., Weismer, G., & Duffy, J. R. (2000). What dysarthrias can tell us about the neural control of speech. *Journal of Phonetics*, **28**, 273–302.

Kormos, J. (2006). *Speech production and second language acquisition.* Lawrence Erlbaum Associates.

https://doi.org/10.1017/S0142716420000223 Published online by Cambridge University Press

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**, 9351–9358.

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**, 1325–1336.

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

Lin, I-F., Mochida, T., Asada, K., Ayaya, S., Kumagaya, S. I., & Kato, M. (2015). Atypical delayed auditory feedback effect and Lombard effect on speech production in high-functioning adults with autism spectrum disorder. *Frontiers in Human Neuroscience*, **9**, 510.

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**, 1042–1046.

Liu, H., Zhang, Q., Xu, Y., & Larson, C. R. (2007). Compensatory responses to loudness-shifted voice feedback during production of Mandarin speech. *Journal of the Acoustical Society of America*, **122**, 2405–2412.

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**, 3739–3746.

Liu, P., Chen, Z., Larson, C. R., Huang, D. & Liu, H. (2010c). Auditory feedback control of voice fundamental frequency in school children. *Journal of the Acoustical Society of America*, **128**, 1306–1312.

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**, 11910.

Lombard, E. (1911). The sign of the elevation of the voice [in French: Le signe de l'elevation de la voix], *Annales des Maladies de l'Oreille, du Larynx, du Nez et du Pharynx*, *37*, 101–119, English translation: http://paul.sobriquet.net/wp-content/uploads/2007/02/lombard-1911-p-h-mason-2006.pdf

Loucks, T., Chon, H., & Han, W. (2012). Audiovocal integration in adults who stutter. *International Journal of Language & Communication Disorders*, **47**, 451–456.

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**, 185–200.

Mackay, D. G. (1970). How does language familiarity influence stuttering under delayed auditory feedback? *Perceptual and Motor Skills*, **30**, 655–669.

Mitsuya, T., MacDonald, E. N., Purcell, D. W., & Munhall, K. G. (2011). A cross-language study of compensation in response to real-time formant perturbation. *Journal of the Acoustical Society of America*, **130**, 2978–2986.

Moser, D., Fridriksson, J., Bonhilha, L., Healy, E. W., Baylis, G., Baker, J. M., & Rorden, C. (2009). Neural recruitment for the production of native and novel speech sounds. *NeuroImage*, **46**, 549–557.

Ng, M. L., Chen, Y., & Sadaka, J. (2008). Vowel features in Turkish accented English. *International Journal of Speech-Language Pathology*, **10**, 404–413.

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.

Parker Jones, Ō., Green, D. W., Grogan, A., Pliatsikas, C., Filippopolitis, K., Ali, N., . . . & Seghier, M. L. (2012). Where, when and why brain activation differs for bilinguals and monolinguals during picture naming and reading aloud. *Cerebral Cortex*, **22**, 892–902.

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**, 1456–1481.

Patel, R., & Schell, K. W. (2008). The influence of linguistic content on the Lombard effect. *Journal of Speech, Language, and Hearing Research*, **51**, 209–220.

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

R Core Team. (2015). R: A language and environment for statistical computing. R foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/

https://doi.org/10.1017/S0142716420000223 Published online by Cambridge University Press

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**, 3793–3806.

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

Schmidt, R. A. & Lee, T. D. (2005). *Motor control and learning: A behavioral emphasis.* Human Kinetics Publishers.

Segalowitz, N. (2010). *Cognitive bases of second language fluency.* Routledge.

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

Simmonds, A. J., Wise, R. J., Dhanjal, N. S., & Leech, R. (2011b). A comparison of sensory-motor activity during speech in first and second languages. *Journal of Neurophysiology*, **106**, 470–478.

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. In *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. (2013). The effect of imagination on stimulation: The functional specificity of efference copies in speech processing. *Journal of Cognitive Neuroscience*, **25**, 1020–1036.

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**, 352–364.

Tian, X., Ding, N., Teng, X., Bai, F., & Poeppel, D. (2018). Imagined speech influences perceived loudness of sound. *Nature Human Behaviour*, **2**, 225–234.

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

Tourville, J. A., Reilly, K. J., & Guenther, F. H. (2008). Neural mechanisms underlying auditory feedback control of speech. *NeuroImage*, **39**, 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**, 499–503.

Van Borsel, J., Sunaert, R., & Engelen, S. (2005). Speech disruption under delayed auditory feedback in multilingual speakers. *Journal of Fluency Disorders*, **30**, 201–217.

Wang, H., & van Heuven, V. J. (2006). Acoustic analysis of English vowels produced by Chinese, Dutch and American speakers, *Linguistics in Netherlands*, **23**, 237–248.

Wiese, R. (1984). Language production in foreign and native languages: Same or different? In H. W. Dechert, D. Möhle, & M. Raupach (Eds.), *Second Language Productions* (pp. 11–25). Gunter Narr Verlag.

Woumans, E., Santens, P., Sieben, A., Versijpt, J., Stevens, M., & Duyck, W. (2015). Bilingualism delays clinical manifestation of Alzheimer's disease. *Bilingualism: Language and Cognition*, **18**, 568–574.

Zhang, Q., & Yang, Y. (2003). The determiners of picture-naming latency. *Acta Psychologica Sinica*, **35**, 447–454.

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**, 1770–1781.

Cite this article: Cai, X., Yin, Y., and Zhang, Q. (2020). A cross-language study on feedforward and feedback control of voice intensity in Chinese–English bilinguals. *Applied Psycholinguistics* **41**, 771–795. https://doi.org/10.1017/S0142716420000223