Psychonomic Bulletin & Review
https://doi.org/10.3758/s13423-023-02446-4

BRIEF REPORT

Amplitude envelope onset characteristics modulate phase locking for speech auditory-motor synchronization

Min Zhu[1] · Fei Chen[1] · Chenxin Shi[1] · Yang Zhang[2]

Accepted: 18 December 2023

Abstract

The spontaneous speech-to-speech synchronization (SSS) test has been shown to be an efective behavioral method to estimate cortical speech auditory-motor coupling strength through phase-locking value (PLV) between auditory input and motor output. This study further investigated how amplitude envelope onset variations of the auditory speech signal may infuence the speech auditory-motor synchronization. Sixty Mandarin-speaking adults listened to a stream of randomly presented syllables at an increasing speed while concurrently whispering in synchrony with the rhythm of the auditory stimuli whose onset consistency was manipulated, consisting of aspirated, unaspirated, and mixed conditions. The participants' PLVs for the three conditions in the SSS test were derived and compared. Results showed that syllable rise time afected the speech auditory-motor synchronization in a bifurcated fashion. Specifcally, PLVs were signifcantly higher in the temporally more consistent conditions (aspirated or unaspirated) than those in the less consistent condition (mixed) for high synchronizers. In contrast, low synchronizerstended to be immune to the onset consistency. Overall, these results validated how syllable onset consistency in the rise time of amplitude envelope may modulate the strength of speech auditory-motor coupling. This study supports the application of the SSS test to examine individual diferences in the integration of perception and production systems, which has implications for those with speech and language disorders that have difculty with processing speech onset characteristics such as rise time.

Keywords Speech auditory-motor synchronization · Phase-locking value · Amplitude rise time · Stimulus onset

Introduction

Humans have an innate ability to detect a regular pulse in an auditory signal, which is known as rhythm processing (Winkler et al., 2009; Zatorre et al., 2007). Along with pitch patterns, rhythmic structures have arguably played a key role in music perception (Winkler et al., 2009) and speech communication (Falk & DallaBella, 2016; Slater et al., 2018) by generating temporal expectancies in listeners. Specifcally, the repetition of prominent events such as musical beats or accented syllables in speech enables listeners to form expectancies about the time of impending events (Large & Jones, 1999). In this regard, temporally expected (as opposed to unexpected) information would be perceived more precisely and efciently owing to temporal regularities that facilitate centralized attention (Falk & Dalla Bella, 2016; Large & Jones, 1999). For example, Jones et al. (2002) demonstrated that listeners judged the pitch of a tone more precisely in a rhythmically regular sequence than in an irregular one. Collectively, in a sequence of spoken words, the accurate tracking of rhythmic regularity could facilitate speech comprehension by isolating continuous sound streams (Andreou et al., 2011), identifying word boundaries (Smith et al., 1989), and providing grammatical structure cues (Gordon et al., 2015). Accumulating evidence has shown that individuals with developmental speech and language disorders such as autism, dyslexia, and apraxia of speech have difculty with temporal patterns and speech onset characteristics that are considered to be essential to the perception and production of stress, intonation, and other aspects of

✉ Fei Chen
chenfeianthony@gmail.com

✉ Yang Zhang
zhanglab@umn.edu

[1] School of Foreign Languages, Hunan University, Changsha, China

[2] Department of Speech-Language-Hearing Sciences and Masonic Institute for the Developing Brain, The University of Minnesota, Twin Cities, MN, USA

language processing including word segmentation (Fiveash et al., 2021; Goswami, 2011; Ladányi et al., 2020).

Interestingly, auditory and motor rhythmic processing are tightly intertwined rather than being independent of each other in both music and speech domains (Assaneo & Poeppel, 2018; Assaneo et al., 2021; Hutchins et al., 2014). Humans are born with the capacity to spontaneously coordinate their motor outputs in time with auditory inputs, such as tapping or dancing to music with a rhythm (Provasi & Bobin-Bègue, 2003; Zatorre et al., 2007), known as auditory-motor synchronization. Bobin-Bègue and Provasi (2008) discovered that children as young as 18 months old could synchronize the rate of fnger tapping with the rhythm of external auditory stimuli (animal noises), and such rhythmic synchronization skills developed with age (Provasi & Bobin-Bègue, 2003). In addition to the music tempo, synchronization between auditory and motor systems is also ubiquitous in the speech domain (Falk & Dalla Bella, 2016; Schmidt-Kassow et al., 2014). For instance, Falk and Dalla Bella (2016) investigated listeners' detection of word changes with aligned or misaligned fnger tapping. They reported an enhancement in accuracy when changes occurred on stressed syllables and motor rhythm was temporally matched with speech rhythm, further suggesting that auditory-motor synchronization enhanced speech perception via efciently reinforcing rhythmic expectancies.

However, extant studies concerning auditory-motor synchronization have mainly emphasized how gross motor movements are entrained by auditory signals, such as fnger tapping and running (Falk & Dalla Bella, 2016; Tryfon et al., 2017; Van Dyck et al., 2021). The synchronization between auditory and vocal production rhythms as well as its infuential factors remained underexplored. Recently, a behavioral task, named spontaneous speech-to-speech synchronization test (SSS test) was introduced to provide an indirect measure of an individual's coupling strength of speech auditory-motor cortices (Assaneo et al., 2019, 2021; Barchet et al., 2022; Kern et al., 2021; Mares et al., 2023). This test has the potential to reveal a bimodal distribution, categorizing participants as either high or low synchronizers based on their level of auditory-motor synchronization. The high synchronizers were shown to better match their continuous speech utterances to the perceived rate, whereas the low synchronizers remained impervious to the external rhythm, and maintained weaker synchronization between their speech output and auditory input. Furthermore, the validity of the SSS test was confrmed at the neural level by Assaneo et al. (2019), who observed a robust functional and structural connectivity of auditory and motor regions in high synchronizers. One limitation of existing studies using the SSS test is that they mainly focused on stress-timed languages such as English (Assaneo et al., 2019) and German (Assaneo et al., 2021; Kern et al., 2021; Rimmele et al., 2022). There is a lack of empirical evidence from otherlanguages with diferent rhythmic structures, as rhythmic typology of languages has been suggested to infuence listeners' perceptual sensitivity (Lidji et al., 2011; Ordin et al., 2019). For instance, it has been observed that native speakers of English (stress-timed language) tend to exhibit greater synchronization strength compared with those of French (syllable-timed language; Lidji et al., 2011). Thus, one question is raised about whether the SSS test would be applicable to a syllable-timed language such as Mandarin Chinese, French, or Italian. To the best of our knowledge, the present study represents the frst attempt to address this question by testing Mandarin-speaking adults.

When it comes to the potential factors afecting the auditory-motor synchronization, participant-related factors include the participants' musical experience and executive functions, while experiment-related factors are closely linked with the stimulus property itself. For instance, a signifcantly strengthened auditory-motor synchronization was observed in musicians than in nonmusicians (Repp & Doggett, 2007; Rimmele et al., 2022), as well as in participants who scored higher on the attention test than those who scored lower (Tierney & Kraus, 2013). By contrast, a reduction in the auditory-motor cortex coupling was expected inpatients with prefrontal damage who exhibited defcits not only in working memory but also in temporal processing (Perbal et al., 2003).

Relative to participant-related factors, less research has been done on the experiment-related factors concerning the auditory stimulus per se, such as rhythmic presentation rate and amplitude rise time. One magnetoencephalography study measured participants' synchrony between auditory and motor cortices when they listened to syllable sequences with speech rates ranging from 2.5 Hz to 6.5 Hz (Assaneo & Poeppel, 2018). A peak of synchrony occurred at 4.5 Hz,which is right close to the natural rhythm of speech across languages (Ding et al., 2017). Furthermore, it is noteworthy that the rhythmic timing of speech is subject to the rate of change of the amplitude envelope at the syllable onset, referred to as the amplitude rise time or attack (Goswami, 2011; Van Hirtum et al., 2019). Previous research has consistently revealed a sharp perceptual asymmetry in duration and intensity estimation for sounds with diferent amplitude envelopes, which is associated with distinct neural synchronization patterns for the onset and ofset responses (Irsik et al., 2021; Zhang et al., 2016). Specifcally, Irsik et al. (2021) recorded listeners' neural synchronization to sounds with various envelope shapes, revealing a perceptual asymmetry with younger people having heightened sensitivity to ramped shapes while elders showing heightened sensitivity to damped shapes. This means that even an isochronous train of syllables maybe perceived diferently due to the discrepancies in amplitude envelope.

Regarding the role of amplitude rise time, it could provide an important acoustic cue for temporal segmentation

through efective neural phase locking to the speech signal (Goswami, 2011; Hämäläinen et al., 2012). In addition, differences in amplitude envelope rise time serve as cues to distinguish specifc contrasts (Goswami & Leong, 2013; Van Hirtum et al., 2019). Although the efect of amplitude rise time on speech perception has been widely reported, it remains an open question whether and how rise time plays a role in temporal synchrony between perception and production. Moreover, it is unclear whether the extent to which amplitude onset consistency modulates speech auditory-motor synchronization difers in diferent subgroups, high and low synchronizers.

In the SSS test of the current study, we adopted three types of syllable trains in Mandarin (unaspirated stops, aspirated stops, and mixed ones) with acoustically identical rhythm but difering onset consistency to examine the modulatory efects of amplitude envelope onset. Unlike many other languages (e.g., Japanese, French, Russian), the Mandarin consonant system is distinctively featured by aspiration (e.g., /p/-/p^h/, /t/-/t^h/, /k/-/k^h/), with Mandarin aspirated stops exhibiting a substantially longer voicing lag (Chen et al., 2023). The strongly aspirated versus unaspirated distinction identifed in Mandarin stop sounds allows a fexible time window for onset detection, which is ideal to implement desirable conditions systematically for stimulus-related variations in the SSS test. In terms of the correlation between aspiration properties and amplitude rise time, it was suggested that aspirated syllables have longer rise time than unaspirated syllables (Fig. 1). Given the perceptual asymmetry resulting from diferent amplitude envelopes (Irsik et al., 2021; Zhang et al., 2016), listeners would consistently perceive the sounds earlier or later in unaspirated and aspirated syllable trains, respectively, while for the mixed sequence consisting of random unaspirated and aspirated syllables, sounds might not be perceived isochronously even though they are identical in duration. Therefore, in both aspirated and unaspirated conditions (more consistent patterns), listeners would readily track rhythmic regularities of syllables, whereas they would not when hearing the mixed auditory sequence. According to the Dynamic Attending Theory, which proposed that attention and perception are enhanced when an auditory sequence is anticipated at a regular and predictable rhythmic rate (Jones et al., 2002; Large & Jones, 1999), it was expected that participants might exhibit better synchronization in the stimulus conditions that have higher onset consistency (unaspirated or aspirated condition) than that in the mixed condition.

Our experimental design represented eforts to advance our understanding of speech auditory-motor synchronization by taking advantage of Mandarin aspirated and unaspirated stops with diferent onset rise times. Given the importance of temporal encoding of speech signals in speech and language acquisition, fndings of the present study may have clinical relevance concerning the role of auditory-motor synchronization in a better understanding of the relationship between language and brain to improve diagnosis and intervention (Schmidt-Kassow et al., 2014; Tierney & Kraus, 2014). For instance, the auditory-motor mapping technique has been applied to efectively facilitate speech production in minimally verbal autistic children (Wan et al., 2011; Yan et al.,

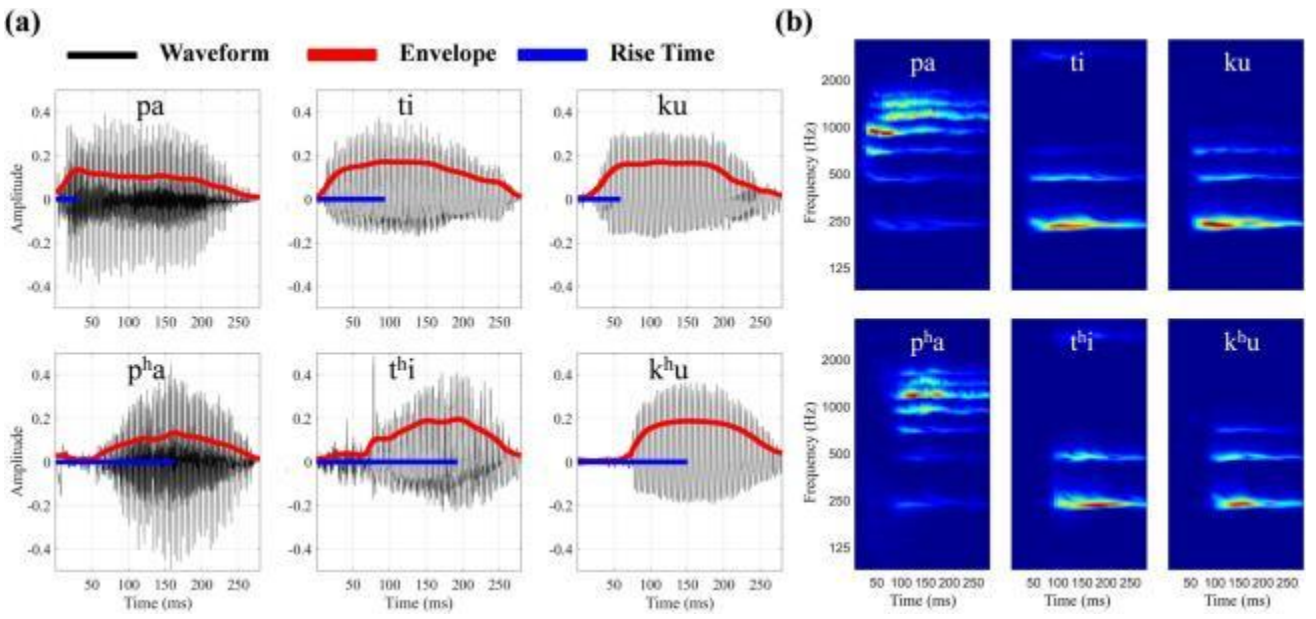

Fig. 1 Waveforms **(a)** and spectrograms **(b)** of aspirated and unaspirated Mandarin stops using /pa/, /p^ha/, /ti/, /t^hi/, /ku/, /k^hu/ as examples. (Color fgure online)

2021). Moreover, it has been shown that individuals with dyslexia are accompanied by defcits in rise time discrimination and reduced neural synchronization (Goswami, 2011; Van Hirtum et al., 2019). In this regard, our preliminary study reported here on the neurotypical adult population could provide the foundational work to examine the generalization of the SSS test to a syllable-timed language, and further analyze how speech onset characteristics may infuence auditory-motor coupling at the group and individual levels.

Method

Participants

A total of 60 Mandarin-speaking college students, recruited from the Hunan University (30 female, 30 male; mean age = 22.02 years, *SD* = 2.25) participated in this study. Half of the participants had amateur musical backgrounds, with an average of 4.97 years of experience (*SD* = 3.37) in either vocal or instrumental music. All of them spoke standard Mandarin without any other dialectal accents and had no self-reported history of neurological defcits or speech, language, or hearing disorders. In compliance with the ethical protocol approved by the Human Research Ethics Committee of Hunan University, written informed consent was signed by all participants before proceeding with any of the experimental procedures. They were compensated for their participation.

Materials

A total of 20 Mandarin monosyllables were employed as auditory stimuli in the SSS test. They were recorded fve times by a female Mandarin native speaker in a sound-attenuated booth with a professional condenser microphone at a 44.1 kHz sampling rate and 16-bit resolution. One token per syllable was carefully chosen for a similar intensity and duration to maintain naturalness. For the sake of exploring the efect of syllable onset characteristics, stimuli were assigned to three conditions depending on the features of consonant onsets: aspirated, unaspirated, and mixed sounds (half aspirated and half unaspirated) with 10 syllables each (Table 1). Amplitude energies of partial aspirated and unaspirated stops were depicted in Fig. 1. Rise time tracks the time taken for the onset amplitude envelope to reach its highest amplitude. For example, rise times were 163 ms, 193 ms, and 150 ms for aspirated /pʰa/, /tʰi/, /kʰu/, and 31 ms, 94 ms, and 59 ms for corresponding unaspirated /pa/, /ti/, /ku/, respectively. For the current study, the target stimuli were restricted to stops and were normalized to the same duration, pitch, and intensity. They all carried the Mandarin level tone (Tone 1) to avoid confounding factors from lexical tone variations.

Table 1 The list of Mandarin syllables used as the auditory input in three stimulus conditions

Aspirated onset	Unaspirated onset	Mixed onset
/pʰa/	/pa/	/pʰa/
/pʰi/	/pi/	/pi/
/pʰu/	/pu/	/pʰu/
/pʰo/	/po/	/po/
/tʰa/	/ta/	/tʰa/
/tʰi/	/ti/	/ti/
/tʰu/	/tu/	/tʰi/
/kʰa/	/ka/	/tu/
/kʰu/	/ku/	/kʰɤ/
/kʰɤ/	/kɤ/	/ka/

Each auditory sequence of three stimulus conditions (aspirated, unaspirated, and mixed sounds) lasted 70 s, with a progressively accelerated speed ranging from 4.3 to 4.7 Hz (mean = 4.5 Hz), using steps of 0.1 Hz. As an example, at 4.3 Hz, each syllable lasted around 233 ms, while at 4.4 Hz, each syllable lasted about 227 ms. Each rate was maintained for 60 syllables (10 stimuli of each condition repeated six times would be combined pseudorandomly) except for the last one, which remained constant until the end of the audio. In each auditory sequence, all syllables were pseudorandomly concatenated in Praat without a gap in between. The only restriction was the avoidance of consecutive repetition of the same syllable. The experiment design was based on the approached outlined in Assaneo et al. (2019) and Kern et al. (2021).

Procedures

Following Kern et al. (2021), an explicit version of the SSS behavioral test was adopted to indirectly estimate listeners' coupling strength between speech auditory and motor cortices (Fig. 2; Assaneo et al., 2019, 2021; Rimmele et al., 2022). Participants were explicitly instructed to wear headphones and align their rhythm of speech output (mouth symbol) with a heard random stimulus sequence (ear symbol) as depicted in Fig. 2. All of them need to complete three blocks difering in auditory sequence (aspirated; unaspirated; mixed) with the presentation order counterbalanced across participants. Each block consisted of two runs of an identical procedure with three steps: volume adjustment, practice trials, and a formal test. More specifcally, at the stage of volume adjustment, subjects were required to adjust the headphone volume of background babbles (a sequence of random syllables played backward) to an appropriate level while concurrently whispering the syllable /ma/ with an intended level tone until they could no longer hear their

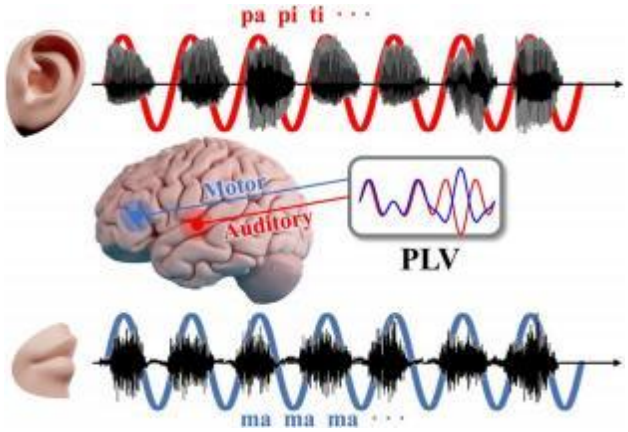

Fig. 2 An experimental paradigm of the SSS test. Red and blue lines represent envelopes of auditory (top) and produced (bottom) speech signals, respectively. (Color fgure online)

own sounds, which might cause auditory interference. Then, a practice trial would be performed in which participants would try to whisper /ma/ at the same rate for 10 s after listening to a 10-s sequence of syllable /ma/ at 4.5 Hz. Finally, in the formal test, listeners were presented with an auditory sequence with random syllables at an accelerated rate (4.3 to 4.7 Hz) while simultaneously whispering the syllable /ma/ in synchrony with the audio. Then, participants need to repeat the whole process for each block. Their speech productions would be recorded and saved automatically. Stimulus presentation and recording were conducted in MATLAB R2018a. Participants completed the SSS test individually in a soundproof room.

Data analysis

Speech auditory-motor coupling strength was quantifed by the phase locking value (PLV; Assaneo & Poeppel, 2018; Assaneo et al., 2019, 2021; Barchet et al., 2022; Kern et al., 2021), which was calculated based on the synchronization of envelopes between auditory and produced speech signals (Fig. 2) via the following formula:

$$\mathrm{PLV}(\ ) = \frac{1}{T}\sum_{t=1}^{T} e^{i(\theta_1(\ ,t)-\theta_2(\ ,t))}, \quad (1)$$

where f represents the frequency, t represents the discretized time, $T$is the total number of time points, and θ_1 and θ_2 are the phases of the audio and the recorded signals, respectively.

Signals were processed with NSL (Neural Systems Laboratory) Auditory Model toolbox in MATLAB R2018a, according to the following steps: extracting envelopes (Hilbert transform), resampling (100 Hz), bandpass fltering (3.3–5.5 Hz), computing phases (Hilbert transform). PLV was calculated with a time window of 5s and an overlap of 2s (Assaneo et al., 2019, 2021). For each run of the SSS test, the mean PLV across windows was provided as the synchronization measurement. Participants' specifc performance for each stimulus condition was obtained by averaging PLVs across two runs of the corresponding block, and linear regression analyses confrmed consistent performance across the two runs within each block for all participants.

Results

Since previous studies have indicated that the synchronization measurement follows a bimodal distribution (Assaneo et al., 2019; Rimmele et al., 2022), a *k*-means clustering with two clusters from the package *factoextra* (Kassambara & Mundt, 2017) in R was applied to the mean PLVs obtained from three stimulus conditions. Thus, our cohort was divided into distinct populations of high and low synchronizers (Highs = 25, Lows = 35). Furthermore, PLVs in high and low clusters conformed to the normal distribution through the function of Shapiro.test embedded in R (High: $W = 0.93, p = .07$; Low: $W = 0.97, p = .53$).

Statistical analyses and data visualization in this study were carried out using R 3.6.1 (R Core Team, 2020). Linear mixed-efects models (LMMs) were implemented using the R package *lme4* (Bates et al., 2014). Main efects and interactions were assessed via Type II Wald chi-square tests through the package *car* (Fox et al., 2012). Pairwise comparisons were conducted using the package *emmeans* (Lenth et al., 2019).

Figure 3 depicts high and low synchronizers' performances of speech auditory-motor synchronization, as refected by PLVs across three conditions. High and low synchronizers achieved mean PLVs of 0.53, 0.32 for unaspirated sounds, 0.52, 0.32 for aspirated sounds, and 0.45, 0.30 for mixed sounds, respectively. An LMM was ftted with "Cluster (high vs. low)," "Stimulus condition (unaspirated vs. aspirated vs. mixed)," and their interaction as the fxed efects, "PLV" as the dependent variable, and "Subject" as a random efect. To control for potential infuences of musical experience on the PLVs, the duration of musical training received by the participants was scaled and included as covariates in the LMM model. The LMM results showed signifcant main efects of "Cluster," $\chi^2(1) = 129.87$, $p < .001$, and "Stimulus condition," $\chi^2(2) = 21.06$, $p < .001$, as well as a signifcant interaction between "Cluster" and "Stimulus condition," $\chi^2(2) = 9.45$, $p < .01$. Then, Tukey-adjusted post hoc pairwise comparisons were conducted and manifested a signifcant disparity between high and low synchronizers for all three conditions ($ps < .001$). As for the efects of stimulus condition on PLVs in each cluster, the

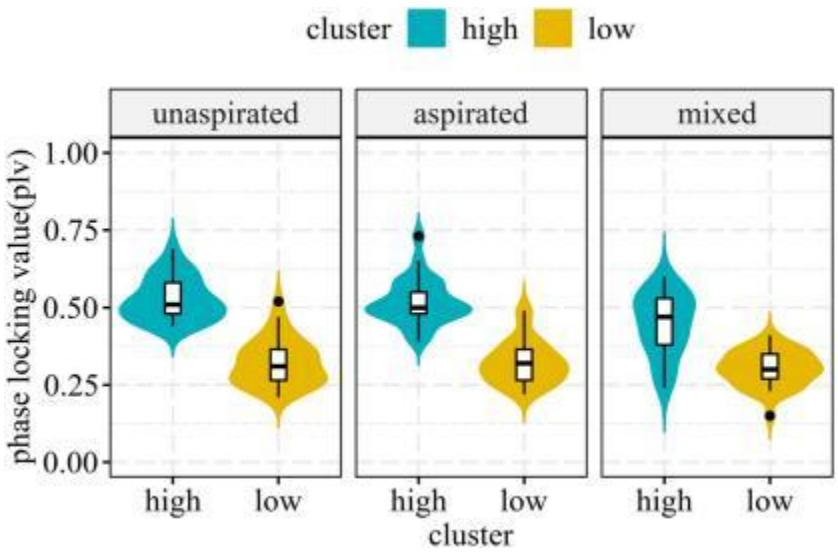

Fig. 3 Mean phase locking values as a function of stimulus condition (unaspirated vs. aspirated vs. mixed) by high and low synchronizers

PLVs of high synchronizers in the aspirated condition were signifcantly higher than those in the mixed condition [β = 0.07, SE = 0.02, t = 4.24, p < .001]; Moreover, they exerted signifcantly higher PLVs in the unaspirated condition than those in the mixed condition [β = 0.08, SE = 0.02, t = 4.75, p < .001], while performing comparably for unaspirated and aspirated sounds [β = 0.009, SE = 0.02, t = 0.51, p = .99]. However, in low synchronizers, none of the signifcant discrepancies were discerned among diferent stimulus conditions (ps > 0.05).

Together, the statistical results suggested that the stimulus onset consistency did afect participants' speech auditory-motor synchronization; That is, participants were sensitive to the amplitude rise time and exhibited higher synchronization when the onset of the syllables was more consistent. Furthermore, high synchronizers would be more susceptible to onset perturbations since they displayed a higher alignment in aspirated and unaspirated conditions than in the mixed condition. Low synchronizers, on the other hand, performed similarly in all three conditions.

Discussion and conclusion

Using the SSS test, this study employed three types of Mandarin syllable sequences to investigate the efect of amplitude envelope rise time on Mandarin speakers' speech auditory-motor synchronization. The current study complements previous neuroimaging fndings by providing behavioral evidence on rhythmic integration between speech perception and production, which involves highly overlapping neural networks, including the auditory cortex, motor and premotor cortex, as well as Broca's area (Skipper et al., 2017). Additional compelling evidence is found in a study by Fadiga et al. (2002), using transcranial magnetic stimulation. This study demonstrated that the speech motor system remained active during a speech perception task, even when there was no requirement to articulate the presented words.

A key finding of the current study is the significant modulatory efects of speech onset rise time on speech auditory-motor synchronization at the group level. The fndings support the hypothesis that auditory sequences with higher onset consistency elicit better synchronization than the mixed syllable sequence. It sounds a little intriguing given prior research indicating that speech comprehension tends to be better for natural nonisochronous speech compared with isochronous sounds (Aubanel et al., 2016). However, it is important to note that the mixed condition in our study does not refect the temporal patterns found in natural speech. Unlike natural speech, the auditory stimuli in the mixed sequence maintain an isochronous pace at each rate, which markedly diverges from the cadence of natural speech. Additionally, syllables in the mixed condition were pseudorandomly combined without gaps in between for all sequences, making them semantically incomprehensible for participants.

Concerning the accounts of efect of onset consistency, as presented in Fig. 1, unaspirated and aspirated stops exhibit substantial diferences in the timing of energy peaks (Van Hirtum et al., 2019), with aspirated syllables lagging almost 100 ms behind unaspirated ones, the disparity of which would further infuence listeners' judgment of syllable duration. As a result, when mixed syllables were concatenated in the sequence, the speech envelope of the auditory sequence would have temporal variations among syllables, compared with pure sequences with fairly consistent onsets. The

jittered temporal profle in the mixed sequence would disturb participants in establishing stable rhythms. As indicated by Dynamic Attending Theory, regular temporal patterns would generate expectancies of future events, thus increasing the attention to expected events (Falk & Dalla Bella, 2016; Jones et al., 2002). Consistency in speech onset rise time might strengthen rhythmic expectancies in the integration of perception and production systems. However, the absence of a signifcant diference between the aspirated (with a smoother onset) and unaspirated (with a steeper onset) conditions potentially suggests that the sharpness of syllable onset may not play a decisive role in synchronization. This is corroborated by Mares et al. (2023) who demonstrated that the improved performance of low synchronizers in the tone condition, as compared with the speech condition, could be attributed to the repeatability of tonal stimuli, rather than the sharpness of tonal onsets. They observed that the advantage in the tonal condition persisted even when tonal stimuli featured a smooth transition similar to speech stimuli.

The fact that syllable rise time afected Mandarin speakers' performance in the SSS test is not surprising, as previous research has shown that participants altered their neural entrainment to stimuli when the rise time of the envelope was changed (Van Hirtum et al., 2019),particularly depending on their ages that afect thresholds of stimulus onset detection (Irsik et al., 2021). In addition, the interaction between onset rise time and speech rhythmic synchronization also supported the view that precise detection of envelope rise time is critical in rhythmic timing by marking the beginning of the syllables (Goswami, 2011; Goswami & Leong, 2013; Hämäläinen et al., 2012).

Furthermore, the signifcant disparities of PLVs among our Mandarin-speaking subjects confrmed a similar pattern found in previous studies that tested participants who spoke stress-timed languages: high synchronizers could whisper simultaneously with the auditory speech rhythm, whereas their low counterparts did not show an interaction between the produced and perceived rhythms (Assaneo et al., 2019, 2021; Kern et al., 2021). Indeed, the diferences in auditory-motor synchronization strength among participants, as observed behaviorally, fnd support in the neural results presented in Assaneo et al. (2019). Their study revealed increased neural entrainment, especially in frontal areas, during passive speech listening in high synchronizers. This localized pattern has been shown to correlate with precise microstructural properties in the white matter pathways linking frontal and auditory areas (Blecher et al., 2016). Hence, these neural attributes might strengthen their sensitivity to rate changes of envelopes, which is benefcial in detecting syllables in speech streams. One of the factors contributing to this discrepancy between high and low synchronizers might lie in the musical experience, which could facilitate rhythmic processing and transfer to the speech domain owing to overlapped neural circuits (Mares et al., 2023; Rimmele et al., 2022; Slater & Kraus, 2016). We have found a moderate positive correlation (r = .53, p <. 001) between participants' synchronization strength (PLVs across three conditions) and their years of musical experience (see supplementary material). In the present study, 22 out of the 25 high synchronizers had prior experience with amateur musical training, with an average duration of 4.77 years. However, unlike earlier studies conducted by Assaneo et al. (2019, 2021, Kern et al., 2021), there was a lack of participants who attained extremely high PLVs (larger than 0.75) in the present study. On the one hand, this discrepancy was presumably due to the diferent musicianship screening criteria applied in the previous and current investigations. On the other hand, the diference could also stem from the distinct rhythmic typology of languages, with native speakers of stress-timed languages demonstrating greater entrainment advantages compared with those of syllable-timed languages (Lidji et al., 2011).

Regarding asymmetric performance across three onset conditions, our data indicated that high synchronizers were more sensitive to the fuctuations of auditory stimulus onset in the SSS test. In other words, high synchronizers had the capacity to detect and adapt their productions fexibly to onset changes of syllable rise time. By contrast, low synchronizers were more resistant to rise time perturbation across conditions, which meant that they were unable to adjust their contemporaneous syllable production rate according to the perceived rhythm. Our results are compatible with the previous fnding that high synchronizers' speech rhythmicity was considerably reduced in the no-rhythm condition (listening to white noise) compared with the rhythm condition, whereas low synchronizers remained constant in both scenarios (Assaneo et al., 2019). A similar case was also found with Mares et al. (2023) who revealed that low synchronizers exhibited impaired auditory-motor synchronization when the auditory stimulus involved speech with varied syllables in a sequence.

In summary, the present study used Mandarin speech materials in neurotypical adult Mandarin speakers and confrmed the generalizability of the SSS test in a syllable-timed language. Moreover, amplitude rise time was identifed as a signifcant factor infuencing PLVs for rhythmic synchronization. In this regard, it is crucial to maintain consistent syllable onsets when directly comparing rhythmic synchronization across participants. Some limitations of the study are acknowledged. Firstly, since the overall year and frequency of musical training were not controlled, and professional musicians were not included, it is untenable to determine whether perturbations across stimulus conditions in high synchronizers are subject to musical profciency. Secondly, our study only tested adult participants. It is worth noting that spontaneous speech synchronization tests

could potentially shed light on word learning and speech development in general, as previous research revealed that the connections between the auditory and frontal regions are crucial in the early stages of word acquisition (Assaneo et al., 2019). Future studies are needed to determine which specifc language abilities or cognitive skills are associated with auditory-motor synchronization in both typical and atypical populations (Fiveash et al., 2021; Goswami, 2011; Ladányi et al., 2020), including studies involving brain imaging and neuromodulation techniques as well as clinical screening and intervention for individuals with speech and language disorders.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.3758/s13423-023-02446-4.

Acknowledgements The authors would like to thank Dr. M. Florencia Assaneo for kindly providing the MATLAB function for PLV calculation. Furthermore, Y. Zhang was additionally supported by the Brain Imaging Grant and SEED Grant from the College of Liberal Arts, University of Minnesota.

Funding This study was supported by grants from National Science Foundation of Hunan Province (Grant No. 2023JJ40107), and the Postgraduate Scientifc Research Innovation Project of Hunan Province (CX20230401).

Data availability The datasets generated and analyzed during the current study are available in the Open Science Framework repository (https://osf.io/z43ny/).

Code availability All data analysis scripts are available in the Open Science Framework repository (https://osf.io/z43ny/).

Declarations

Conflicts of interest The authors have no competing interests to declare that are relevant to the content of this article.

Ethics approval This study was performed in line with the principles of the Declaration of Helsinki. Approval of the research was granted by the Human Research Ethics Committee of Hunan University.

Consent to participate All participants gave written informed consent prior to the study.

Consent for publication All participants gave written informed consent for publication.

References

Andreou,L.-V., Kashino,M., & Chait,M. (2011). The role of temporal regularity in auditory segregation. *Hearing Research, 280*(1/2), 228–235.

Assaneo, M. F., & Poeppel, D. (2018). The coupling between auditory and motor cortices is rate-restricted: Evidence for an intrinsic speech-motor rhythm. *Science Advances*, *4*(2), Article eaao3842. https://doi.org/10.1126/sciadv.aao3842

Assaneo, M. F., Rimmele,J. M., Sanz Perl, Y., & Poeppel, D. (2021). Speaking rhythmically can shape hearing. *Nature Human Behaviour, 5*(1), 71–82.

Assaneo, M. F., Ripollés, P., Orpella, J., Lin, W. M., de Diego-Balaguer, R., & Poeppel, D. (2019). Spontaneous synchronization to speech reveals neural mechanisms facilitating language learning. *Nature Neuroscience, 22*(4), 627–632.

Aubanel, V., Davis, C., & Kim, J. (2016). Exploring the role of brain oscillations in speech perception in noise: Intelligibility of isochronously retimed speech. *Frontiers in Human Neuroscience, 10*. https://doi.org/10.3389/fnhum.2016.00430

Barchet, A. V., Henry, M. J., Pelof, C., & Rimmele, J. M. (2022). Auditory–motor synchronization and perception suggest partially distinct time scales in speech and music [Preprint]. *PsyArXiv.* https://doi.org/10.31234/osf.io/7nh98

Bates, D., Mächler, M., Bolker, B. M., & Walker, S. C. (2014). Fitting linear mixed-efects models using lme4. *Journal of Statistical Software*, 67(1).1–48. https://doi.org/10.18637/jss.v067.i01

Blecher, T., Tal, I., & Ben-Shachar, M. (2016). White matter microstructural properties correlate with sensorimotor synchronization abilities. *NeuroImage, 138*, 1–12.

Bobin-Bègue, A., & Provasi, J. (2008). Régulation rythmique avant 4 ans: Efet d'un tempo auditif sur le tempo moteur [Rhythmic regulation before 4 years: Efect of an auditory tempo on the motor tempo]. *Année Psychologique.* https://doi.org/10.4074/s000350330800403x

Chen, F., Xia, Q., Feng, Y., Wang, L., & Peng, G. (2023). Learning challenging L2 sounds via computer-assisted training: Audiovisual training with an airfow model. *Journal of Computer Assisted Learning, 39*(1), 34–48.

Ding, N., Patel, A. D., Chen, L., Butler, H., Luo, C., & Poeppel, D. (2017). Temporal modulations in speech and music. *Neuroscience & Biobehavioral Reviews, 81*, 181–187.

Fadiga, L., Craighero, L., Buccino, G., & Rizzolatti, G. (2002). Speech listening specifically modulates the excitability of tongue muscles: A TMS study. *European Journal of Neuroscience, 15*(2), 399–402.

Falk, S., & Dalla Bella, S. (2016). It is better when expected: Aligning speech and motor rhythms enhances verbal processing. *Language, Cognition and Neuroscience, 31*(5), 699–708.

Fiveash, A., Bedoin, N., Gordon, R. L., & Tillmann, B. (2021). Processing rhythm in speech and music: Shared mechanisms and implications for developmental speech and language disorders. *Neuropsychology*, *35*(8), Article 8. https://doi.org/10.1037/neu0000766

Fox, J., Weisberg, S., Adler, D., Bates, D., Baud-Bovy, G., Ellison, S., Firth, D., Friendly, M., Gorjanc, G., & Graves, S. (2012). *Package 'car.'* R Foundation for Statistical Computing, *16*. Retrieved August, 2023, from https://cran.r-project.org/web/packages/car/index.html

Gordon, R. L., Shivers, C. M., Wieland, E. A., Kotz, S. A., Yoder, P. J., & Devin McAuley, J. (2015). Musical rhythm discrimination explains individual diferences in grammar skills in children. *Developmental Science, 18*(4), 635–644.

Goswami, U. (2011). A temporal sampling framework for developmental dyslexia. *Trends in Cognitive Sciences, 15*(1), 3–10.

Goswami, U., & Leong, V. (2013). Speech rhythm and temporal structure: Converging perspectives? *Laboratory Phonology, 4*(1), 67–92.

Hämäläinen, J. A., Rupp, A., Soltész, F., Szücs, D., & Goswami, U. (2012). Reduced phase locking to slow amplitude modulation in adults with dyslexia: An MEG study. *NeuroImage, 59*(3), 2952–2961.

Hutchins, S., Larrouy-Maestri, P., & Peretz, I. (2014). Singing ability is rooted in vocal-motor control of pitch. *Attention, Perception, & Psychophysics, 76*(8), 2522–2530.

Irsik, V. C., Almanaseer, A., Johnsrude, I. S., & Herrmann, B. (2021). Cortical responses to the amplitude envelopes of sounds change with age. *Journal of Neuroscience, 41*(23), 5045–5055.

Jones,M. R., Moynihan, H., MacKenzie,N., & Puente,J. (2002). Temporal aspects of stimulus-driven attending in dynamic arrays. *Psychological Science, 13*(4), 7.

Kassambara, A., & Mundt, F. (2017). Package 'factoextra.' *Extract and Visualize the Results of Multivariate Data Analyses*, *76*(2). Retrieved August, 2023, from https://CRAN.R-project.org/package=factoextra

Kern, P., Assaneo, M. F., Endres, D., Poeppel, D., & Rimmele, J. M. (2021). Preferred auditory temporal processing regimes and auditory–motor synchronization. *Psychonomic Bulletin & Review, 28*(6), 1860–1873.

Ladányi, E., Persici, V., Fiveash, A., Tillmann, B., & Gordon, R. L. (2020). Is atypical rhythm a risk factor for developmental speech and language disorders? *Wiley Interdisciplinary Reviews: Cognitive Science*, *11*(5), Article 5. https://doi.org/10.1002/wcs.1528

Large,E. W., & Jones,M. R. (1999). The dynamics of attending: How people track time-varying events. *Psychological Review, 106*, 119–159.

Lenth, R., Singmann, H., Love, J., Buerkner, P., & Herve, M. (2019). Package 'emmeans.' *American Statistician*. Retrieved August, 2023, from https://cran.r-project.org/web/packages/emmeans/index.html

Lidji, P., Palmer, C., Peretz, I., & Morningstar, M. (2011). Listeners feel the beat: Entrainment to English and French speech rhythms. *Psychonomic Bulletin & Review, 18*(6), 1035–1041.

Mares, C., Echavarría Solana, R., & Assaneo, M. F. (2023). Auditory–motor synchronization varies among individuals and is critically shaped by acoustic features. *Communications Biology, 6*(1), 658.

Ordin, M., Polyanskaya, L., Gómez, D. M., & Samuel, A. G. (2019). The role of native language and the fundamental design of the auditory system in detecting rhythm changes. *Journal of Speech, Language, and Hearing Research, 62*(4), 835–852.

Perbal, S., Couillet,J., Azouvi, P., & Pouthas, V. (2003). Relationships between time estimation, memory, attention, and processing speed in patients with severe traumatic brain injury. *Neuropsychologia, 41*(12), 1599–1610.

Provasi, J., & Bobin-Bègue, A. (2003). Spontaneous motor tempo and rhythmical synchronisation in $2^{1/2}$-and 4-year-old children. *International Journal of Behavioral Development, 27*(3), 220–231.

Repp, B. H., & Doggett, R. (2007). Tapping to a very slow beat: A comparison of musicians and nonmusicians. *Music Perception, 24*(4), 367–376.

Rimmele, J. M., Kern, P., Lubinus, C., Frieler, K., Poeppel, D., & Assaneo, M. F. (2022). Musical sophistication and speech auditory–motor coupling: Easy tests for quick answers. *Frontiers in Neuroscience*, *15*, Article 764342. https://doi.org/10.3389/fnins.2021.764342

R Core Team. (2020). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.

Schmidt-Kassow, M., Zink, N., Mock, J., Thiel, C., Vogt, L., Abel, C., & Kaiser, J. (2014). Treadmill walking during vocabulary encoding improves verbal long-term memory. *Behavioral and Brain Functions, 10*(1), 1–9.

Skipper, J. I., Devlin, J. T., & Lametti, D. R. (2017). The hearing ear is always found close to the speaking tongue: Review of the role of the motor system in speech perception. *Brain and Language, 164*, 77–105.

Slater, J., & Kraus, N. (2016). The role of rhythm in perceiving speech in noise: A comparison of percussionists, vocalists and non-musicians. *Cognitive Processing, 17*(1), 79–87.

Slater, J., Kraus, N., Carr, K. W., Tierney, A., Azem, A., & Ashley, R. (2018). Speech-in-noise perception is linked to rhythm production skills in adult percussionists and non-musicians. *Language, Cognition and Neuroscience*, *33*(6), Article 6. https://doi.org/10.1080/23273798.2017.1411960

Smith, M. R., Cutler, A., Butterfeld, S., & Nimmo-Smith, I. (1989). The perception of rhythm and word boundaries in noise-masked speech. *Journal of Speech, Language, and Hearing Research, 32*(4), 912–920.

Tierney, A., & Kraus, N. (2014). Auditory–motor entrainment and phonological skills: Precise auditory timing hypothesis (PATH). *Frontiers in Human Neuroscience, 8*, 949. Retrieved March, 2023, from https://www.frontiersin.org/article/10.3389/fnhum.2014.00949

Tierney, A. T., & Kraus, N. (2013). The ability to tap to a beat relates to cognitive, linguistic, and perceptual skills. *Brain and Language, 124*(3), 225–231.

Tryfon, A., Foster, N. E., Ouimet, T., Doyle-Thomas, K., Anagnostou, E., Sharda, M., & Hyde, K. L. (2017). Auditory–motor rhythm synchronization in children with autism spectrum disorder. *Research in Autism Spectrum Disorders, 35*, 51–61.

Van Dyck, E., Buhmann, J., & Lorenzoni, V. (2021). Instructed versus spontaneous entrainment of running cadence to music tempo. *Annals of the New York Academy of Sciences, 1489*(1), 91–102.

Van Hirtum, T., Ghesquière, P., & Wouters, J. (2019). Atypical neural processing of rise time by adults with dyslexia. *Cortex, 113*, 128–140.

Wan, C. Y., Bazen, L., Baars, R., Libenson, A., Zipse, L., Zuk, J., Norton, A., & Schlaug, G. (2011). Auditory–motor mapping training as an intervention to facilitate speech output in non-verbal children with autism: A proof of concept study. *PLOS ONE*, *6*(9), Article e25505. https://doi.org/10.1371/journal.pone.0025505

Winkler, I., Háden, G. P., Ladinig, O., Sziller, I., & Honing, H. (2009). Newborn infants detect the beat in music. *Proceedings of the National Academy of Sciences, 106*(7), 2468–2471.

Yan, J., Chen, F., Gao, X., & Peng, G. (2021). Auditory–motor mapping training facilitates speech and word learning in tone language–speaking children with autism: An early efcacy study. *Journal of Speech, Language, and Hearing Research, 64*(12), 4664–4681.

Zatorre, R. J., Chen, J. L., & Penhune, V. B. (2007). When the brain plays music: Auditory–motor interactions in music perception and production. *Nature Reviews Neuroscience, 8*(7), 547–558.

Zhang, Y., Cheng, B., Koerner, T. K., Schlauch, R. S., Tanaka, K., Kawakatsu, M., Nemoto, I., & Imada, T. (2016). Perceptual temporal asymmetry associated with distinct on and of responses to time-varying sounds with rising versus falling intensity: A magnetoencephalography study. *Brain Sciences*, *6*(3), Article 3. https://doi.org/10.3390/brainsci6030027

Open practices statement All experimental materials, raw data and analysis scripts are available in the Open Science Framework repository (https://osf.io/z43ny/), and none of the studies was preregistered.