Computational Modeling of Tonal Encoding in Disyllabic Mandarin Word Production

Jiabing Pan (jiabing.pan@connect.polyu.hk)
Department of Chinese and Bilingual Studies, The Hong Kong Polytechnic University
11 Yuk Choi Road, Hung Hom, Hong Kong Special Administrative Region

Xiaocong Chen (xiaocong.chen@polyu.edu.hk)
Department of Chinese and Bilingual Studies, The Hong Kong Polytechnic University
11 Yuk Choi Road, Hung Hom, Hong Kong Special Administrative Region

Caicai Zhang (caicai.zhang@polyu.edu.hk)
Department of Chinese and Bilingual Studies, The Hong Kong Polytechnic University
11 Yuk Choi Road, Hung Hom, Hong Kong Special Administrative Region

Abstract

Approximately half of the world's languages are tonal, and how lexical tone is encoded in spoken word production is still unclear. In Mandarin word production, there are two contrasting views regarding the mechanisms of tonal encoding. The two-stage model assumes that the lexical tone is selected first at the early stage of production, and then integrated with the atonal syllable at the later stage, while other researchers proposed that the lexical tone is retrieved only at the later stage of production. In this study, we performed computational simulations on disyllabic words to uncover the mechanisms underlying the facilitation and interference effects on naming latencies observed in previous primed picture naming studies, which intended to verify the two theoretical accounts of tonal encoding in Mandarin spoken word production. The results supported the two-stage model of tonal encoding in disyllabic Mandarin word production. Increased inhibition between atonal syllables and decreased activation between the tonal frame and the syllable motor program, implying slower tone-to-syllable integration, appear to be the prerequisite for generating the interference effect of tonal overlap without shared syllabic information.

Keywords: Lexical tone; Tonal encoding; Word production; Mandarin Chinese; Computational simulations

Introduction

Speech production involves conceptual preparation, lexical selection, word-form encoding, and articulation (Dell, 1986; Levelt et al., 1999; Roelofs, 1997). According to the WEAVER++ (Word-form Encoding by Activation and Verification++, Levelt et al., 1999) model of speech production, during word-form encoding, speakers retrieve segments and metrical frames of a morpheme in parallel for phonological encoding, followed by prosodification and syllabification, which involves a serial segment-to-frame association process. During phonetic encoding, the articulatory motor program of the syllable is retrieved to prepare for articulation.

Tonal languages account for around half of the languages in the world, spoken by approximately 2 billion people (Liu et al., 2023). For tonal languages like Mandarin Chinese, spoken word production involves a tonal encoding process to plan the suprasegmental information of lexical tones, which systematically distinguish word meanings (M. Y. Chen, 2000; Yip, 2002). Previous studies on Mandarin spoken word production have explored the temporal dynamics of syllabic encoding and segmental encoding (e.g., Cai et al., 2020; J. Y. Chen et al., 2002; Feng et al., 2019; O'Seaghdha et al., 2010; Zhang & Damian, 2019). However, the time course and mechanisms of tonal encoding in Mandarin word production are still controversial.

In Mandarin word production, there are two different models for tonal encoding. The two-stage model proposes two sub-processes for tonal encoding: the tonal framework is selected first at an early stage of production, and then integrated with the atonal syllable at a later stage to prepare for the syllable motor program (Alderete et al., 2019; J. Y. Chen, 1999; Roelofs, 2015). In contrast, other researchers assume that the lexical tone is retrieved only at a later stage of production. Previous studies on Mandarin word production have used the picture-word interference task (e.g., Zeng et al., 2022), the implicit priming task (e.g., J. Y. Chen et al., 2002), or the phonologically primed picture naming task (e.g., Chen & Zhang, 2025) to address this issue. They orthogonally manipulated syllabic and tonal overlap between the target word and the prime/distractor, which led to four phonologically related conditions: the homophonous prime condition (both syllable- and tone-related, S+T+), the syllabic prime condition (syllable-related and tone-unrelated, S+T-), the tonal prime condition (syllable-unrelated and tone-related, S-T+), and the unrelated prime condition (both syllable-

In D. Barner, N.R. Bramley, A. Ruggeri and C.M. Walker (Eds.), *Proceedings of the 47th Annual Conference of the Cognitive Science Society*

and tone-unrelated, S-T-). For instance, using the picture-word interference task, Zhou and Zhuang (2000) observed that in the early stimulus onset asynchrony (SOA) condition (57 ms), homophonous distractors with both syllabic and tonal overlap led to shorter naming latencies compared to syllabic and unrelated distractors. However, there was no processing advantage for homophonous distractors compared to syllabic distractors in the late SOA condition (200 ms), suggesting that the tonal encoding starts at an early stage in Mandarin word production. Similarly, Zeng et al. (2022) also found that both tonal and syllabic distractors facilitated naming latencies in the early SOA condition (−100 ms). The above findings revealed an early encoding of lexical tones during Mandarin word production.

In contrast, evidence from EEG studies revealed a mixed picture. While some studies revealed a later tonal encoding, others supported the two-stage model in Mandarin word production. For example, previous studies reported later N200 onset latencies in the Go-NoGo task when Mandarin speakers made phonological decisions on picture names based on the tone than based on the onset consonant (Zhang & Damian, 2009; Zhang & Zhu, 2011). This finding indicated that during covert Mandarin word production, tonal encoding occurred later after the retrieval of the onset consonant. Since many EEG studies found that the encoding of segments occurred after the encoding of atonal syllables in Mandarin (Cai et al., 2020; Feng et al., 2019; Zhang & Damian, 2019), the above finding implied that tonal encoding in Mandarin occurs at a later stage of production, at least after syllabic encoding. However, using a phonologically primed picture-naming task, Chen and Zhang (2025) elucidated the time course of tonal encoding and syllabic encoding in Mandarin and provided evidence for the two-stage model through two EEG experiments. They reported that in both disyllabic and monosyllabic word production, there was a main effect of tonal relatedness in an earlier ERP time window than that of syllabic relatedness, and also an interaction between syllabic and tonal relatedness in later ERP time windows and naming latencies, suggesting an early independent tonal retrieval and a later tone-to-syllable mapping process. Specifically, the tonal encoding started even earlier than syllabic encoding in the stimulus-locked ERPs, with the tonal priming effect emerging in the time window of 180–250 ms for monosyllabic words and 250–320 ms for disyllabic words, while the syllabic priming effect mostly from 320 ms after the picture onset (Chen & Zhang, 2025). These findings supported the two-stage model of tonal encoding in Mandarin word production.

Previous studies showed that both the homophonous and syllabic prime conditions have shorter naming latencies than the unrelated prime condition, with the former mostly eliciting a larger facilitation effect than the latter (J. Y. Chen et al., 2002; Q. Zhang, 2008, but see a different result in Zeng et al., 2022). In contrast, researchers have reported an interference effect (or a trend of interference) under the tonal prime condition compared to the unrelated prime condition (J. Y. Chen et al., 2002; Chen & Zhang, 2025; Q. Zhang, 2008). There are two explanations for the interference effects when there is a tonal overlap but without a syllabic overlap. The first one is related to a competition mechanism whereby the tonal prime may activate a different existing syllabic frame, which will cause stronger competition during the later tone-to-syllable mapping process. Thus, extra efforts are needed to suppress the interfering syllabic frame and revise the syllabic content to match the target syllable, resulting in longer naming latencies (Chen & Zhang, 2025). The second explanation is based on a dummy syllable mechanism, which assumes that speakers prepare the tonal frame in advance with a dummy atonal syllable, but later revise the dummy content to the target syllable (J. Y. Chen et al., 2002; Q. Zhang, 2008). Both mechanisms can account for longer naming latencies under the tonal prime condition compared to the unrelated prime condition. However, these are only tentative explanations, and so far, there has been no empirical evidence to support either of these mechanisms. Therefore, further verification is needed to understand the nature of tonal encoding in Mandarin word production.

Computational modeling is a powerful tool for elucidating the underlying mechanisms of human behavior. It can be used to construct, test, and refine theories by comparing the simulation results with corresponding behavioral patterns (Hameau et al., 2021; Palminteri et al., 2017). Previous studies have used the neural network model (or the connectionist model) to simulate human behavior through structures and representations (Farrell & Lewandowsky, 2018). In this study, we built computational models of disyllabic Mandarin word production to elucidate the mechanism of tonal encoding in Mandarin. We also investigated whether models with the competition mechanism and the dummy syllable mechanism could explain the interference effect when there is tonal overlap without shared syllabic information. For the competition mechanism, we predicted that increasing the inhibition rate between atonal syllables under the tonal prime condition would lead to longer naming latencies. For the dummy syllable mechanism, since extra steps are needed for processing the dummy syllable compared with the classic word production model, the network following an interactive activation pattern would experience extra decay due to an additional layer for the dummy syllable, thus we predicted that the model with the dummy syllable mechanism would exhibit an interference effect under the tonal prime condition.

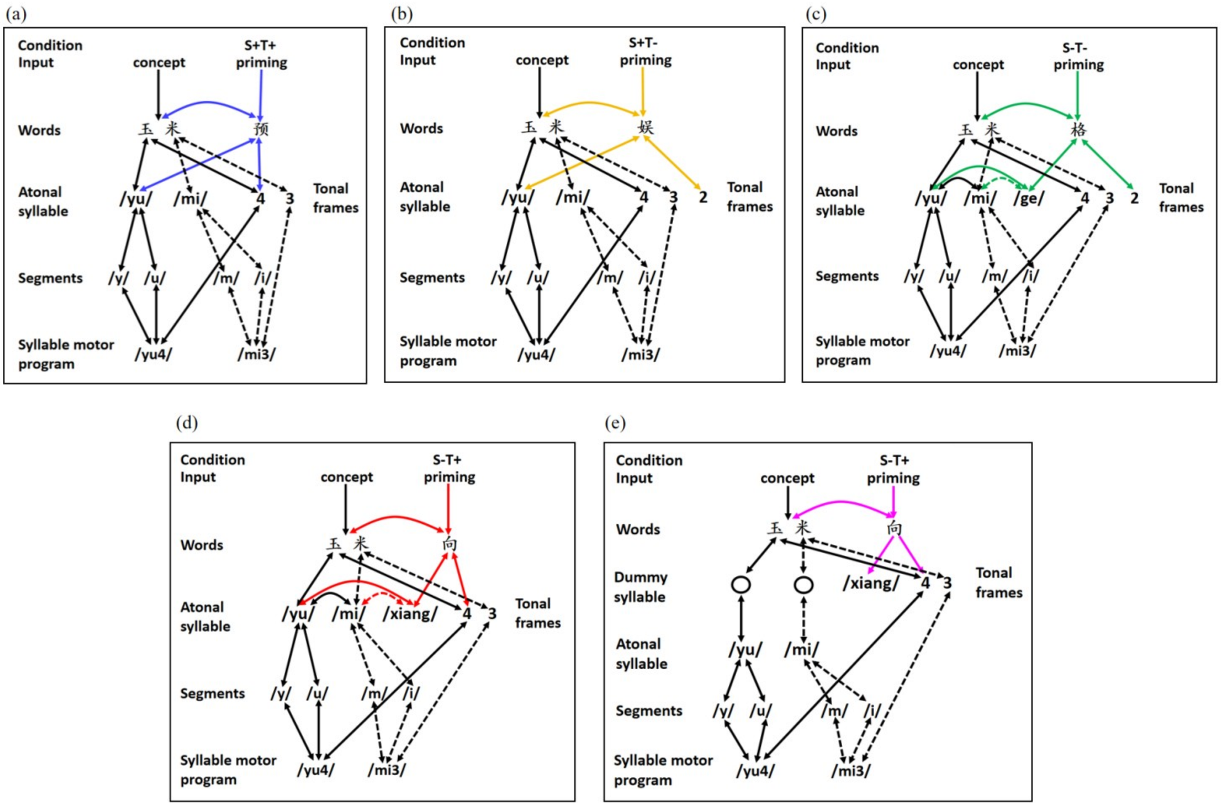

Figure 1: The two-stage model of tonal encoding in disyllabic Mandarin word production under: (a) the homophonous prime condition (S+T+); (b) the syllabic prime condition (S+T-); (c) the unrelated prime condition (S-T-); (d) the tonal prime condition (S-T+) with the competition mechanism; and (e) the tonal prime condition (S-T+) with the dummy syllable mechanism. Arrows between nodes in the network indicate the sequence of activation. For the disyllabic target word, the solid line represents the encoding of the first syllable, and the dotted line represents the encoding of the second syllable.

Method

Two different models of tonal encoding were built. One was based on the two-stage model (Figure 1) and the other was based on the assumption that the tone is only retrieved at the later stage of speech production (referred to as the late-retrieval model, Figure 2). We examined which model could better simulate the effects in the phonologically primed picture-naming task. For the disyllabic target word (e.g., 玉米, /yu4 mi3/, 'corn', the numbers '4' and '3' indicates the falling tone and the falling-rising tone in Mandarin, respectively), there were four different monosyllabic characters as the primes, leading to the homophonous (S+T+, e.g., 预, /yu4/, 'in advance'), syllabic (S+T-, e.g., 娱, /yu2/, 'entertain', the number '2' indicates the rising tone in Mandarin), tonal (S-T+, e.g., 向, /xiang4/, 'towards'), and unrelated (S-T-, e.g., 格, /ge2/, 'grid') prime conditions. According to most current Chinese speech production models like the proximate units principle (O'Seaghdha et al., 2010), during Mandarin spoken word production, atonal syllables are retrieved first as the functional units for phonological encoding and further decomposed into segments/phonemes. Then segments are attached to the tonal frame to prepare for the syllable motor programs during the phonetic encoding (O'Seaghdha et al., 2010; Roelofs, 2015). Thus, each model included external input for the prime and the target, and layers for words, atonal syllables, tonal frames, segments, and syllable motor programs. For the two-stage model, the word activation propagated to both the atonal syllables and the tonal frame in parallel at an early stage and then spread to segments, which were further combined with the tonal frame into syllable motor programs. In contrast, for the late-retrieval model, the word activation only spread to atonal syllables during phonological encoding, while an additional tonal input was applied to the tonal frame only at a later stage of production. Additionally, since previous studies in Mandarin speech production have presented a cascadedness pattern of the

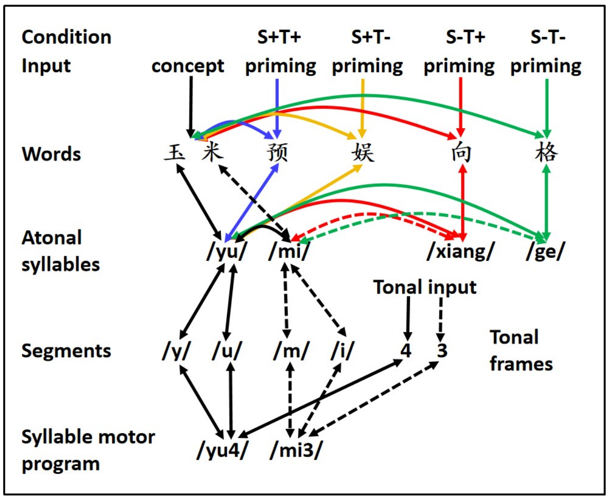

Figure 2: The late-retrieval model of tonal encoding in disyllabic Mandarin word production. Arrows between nodes in the network indicate the sequence of activation. For the disyllabic target word, the solid line represents the encoding of the first syllable, and the dotted line represents the encoding of the second syllable.

activation from semantics to phonology during lexical access, which indicated feedback from the phonological level to the semantic level (e.g., Bao et al., 2023), we adopted an interactive activation pattern that allowed activation to spread bidirectionally in the model (see Chen & Mirman, 2012; Dell, 1986, 1988).

Activation spread in the network through bidirectional excitatory links, allowing feedback between adjacent layers. Nodes in the word layer were connected by bidirectional inhibitory links to implement lexical competition. Additionally, there were bidirectional inhibitory links between atonal syllables to implement syllabic competition under the tonal-related and unrelated prime conditions. These inhibitory weights between words or atonal syllables were scaled through a sigmoid function of unit activation, so that the weakly activated nodes implement very little inhibition on other nodes, while strongly activated nodes implement strong inhibition on other nodes, which allowed a series of candidates to be initially activated in parallel but still forced the model to settle to a single active representation ultimately (Chen & Mirman, 2012). All the units followed an interactive activation and competition function in which positive input drove the unit activation toward its maximum, and negative input drove the unit activation toward its minimum (Chen & Mirman, 2012; McClelland & Rumelhart, 1981). The net input to unit *i* was

$$net_i = extinput_i + \sum_j w_{ij} a_j$$

where w_{ij} was the connection weight from unit *j* to unit *i*, a_j was the activation of unit *j*, and $extinput_i$ was external input to unit *i*. The activation of every unit was updated as follows:

If $net_i > 0$, $\Delta a_i = (max - a_i)net_i - decay(a_i - rest)$.
Otherwise, $\Delta a_i = (a_i - min)net_i - decay(a_i - rest)$.

where *max* was the maximum unit activation level, *min* was the minimum unit activation level, *rest* was the resting activation level, and *decay* was for bringing the unit activation back to the resting activation level.

The production of the disyllabic word 玉米 (/yu4 mi3/) followed a serial encoding pattern (see Roelofs, 1997), whereby the excitation rate between the word and atonal syllables was first increased for /yu4/ and decreased for /mi3/ for encoding the first syllable (with the excitation rate for /mi3/ being half of that for /yu4/), and later vice versa for encoding the second syllable after the activation of the first syllable motor program reached a specific threshold (set as 0.7 in this study, see Chen & Mirman, 2012). Naming responses were produced when both the activation of the two syllable motor programs reached that threshold. The number of cycles needed for the second syllable motor program to reach this threshold constituted the model's reaction time (akin to the naming latency). After validating the two-stage model for tonal encoding in Mandarin, we further built another two models to investigate the prerequisite for the interference effect under the tonal prime condition based on the competition mechanism (Figure 1d) and the dummy syllable mechanism (Figure 1e), respectively. The model parameters are shown in Table 1. The activation of the second syllable motor program changing over time was plotted as curves (Figure 3).

Table 1: Parameters of the computational simulations

Parameter	Value
External input	0.001
Tonal input	0.001
Excitation rate between layers	0.1
Inhibition rate between words/syllables	0.1
Decay rate of node activation	0.01
Maximum unit activation	1
Minimum unit activation	0
Resting activation	0
Threshold for word production	0.7

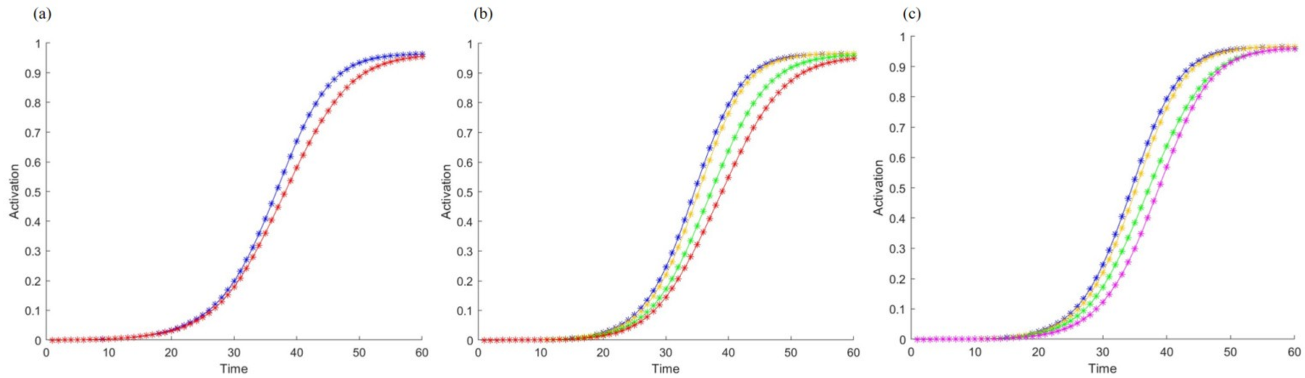

Figure 3: The simulation results of: (a) the late-retrieval model; (b) the two-stage model with the competition mechanism; and (c) the two-stage model with the dummy syllable mechanism, when the excitation rate between the tonal frame and the syllable motor program was 0.07. The color of the curves represents the corresponding condition: Blue: homophonous prime (S+T+); Yellow: syllabic prime (S+T-); Green: unrelated prime (S-T-); Red: tonal prime (S-T+) with the competition mechanism; Violet: tonal prime (S-T+) with the dummy syllable mechanism.

Results

For the late-retrieval model, the motor program of the second syllable first exceeded the threshold of 0.7 after 41 cycles under both the homophonous and syllabic prime conditions and after 43 cycles under both the tonal and unrelated prime conditions (Figure 3a, where the yellow curve overlaps with the blue one, and the green curve overlaps with the red one), indicating the same effect size of homophonous and syllabic priming and no tonal priming effect, which was inconsistent with empirical findings. For the two-stage model, we successfully generated facilitation effects under both homophonous and syllabic prime conditions. Specifically, the former elicited relatively larger facilitation effects than the latter, reflected as 38 cycles for the second syllable motor program to exceed the threshold under the homophonous prime condition and 39 cycles under the syllabic prime condition, compared to 42 cycles under the unrelated prime condition (Figure 3b~3c).

For the two-stage model with the competition mechanism, if increasing the inhibition rate between atonal syllables from 0.1 to 0.5 for the tonal prime condition, we successfully generated the interference effect, reflected as 43 cycles for the second syllable motor program to exceed the threshold under the tonal prime condition compared to 42 cycles under the unrelated prime condition. For the model with the dummy syllable mechanism, although it had an additional dummy syllable layer, which would cause extra activation decay, the model did not exhibit an interference effect as predicted but a facilitation effect instead, reflected as 41 cycles for the second syllable motor program to exceed the threshold under the tonal prime condition compared to 42 cycles under the unrelated prime condition. However, if decreasing the excitation rate between the tonal frame and the syllable motor program from 0.1 to 0.07 for the tonal prime condition, we obtained the interference effect in the two-stage model with either the competition mechanism (Figure. 3b, after 44 cycles) or the dummy syllable mechanism (Figure. 3c, after 43 cycles), compared to 42 cycles under the unrelated prime condition. These results suggested that the decreased activation between the tonal frame and the syllable motor program, which implies a slower tone-to-syllable integration process, is the key to generating the interference effect of tonal priming, while the role of the dummy syllable may not be essential. Furthermore, we adjusted the excitation rate between the tonal frame and the syllable motor program by 0.01 each time to identify the cutoff when the facilitation effect would turn to a null effect before generating the interference effect. Specifically, the cutoff of the excitation rate for this change was 0.09 for the two-stage model with the competition mechanism and 0.08 for the model with the dummy syllable mechanism to obtain a null effect.

Discussion

This study investigated the mechanisms of tonal encoding in disyllabic Mandarin word production through computational simulations. The late-retrieval model revealed no differences in naming latencies between the homophonous and syllabic prime conditions or between the tonal and unrelated prime conditions, which was incompatible with the empirical findings (J. Y. Chen et al., 2002; Chen & Zhang, 2025; Q. Zhang, 2008). In contrast, for the two-stage model with either the competition mechanism or the dummy syllable mechanism, we obtained a relatively larger facilitation effect under the homophonous prime condition than the syllabic prime condition, consistent with previous evidence (J. Y. Chen et al., 2002; Chen & Zhang, 2025; Q. Zhang, 2008). These results support the two-stage model of tonal encoding in Mandarin word production, while both the competition mechanism and the dummy syllable mechanism appear appropriate. We further examined the prerequisite for generating the interference effect under the tonal prime condition in the two-stage model. It was found that by increasing the inhibition rate between atonal syllables under the competition mechanism, or by decreasing the excitation rate between the tonal frame and the syllable motor program under either the competition mechanism or the dummy syllable mechanism, the facilitation effect under the tonal prime condition gradually changed to the interference effect compared to the unrelated prime condition. For the model with the competition mechanism, the increased inhibition rate between atonal syllables may cause stronger competition in the later tone-to-syllable integration process. Speakers need to dissociate the tone and the competing syllabic frame of the prime, suppress the interfering syllabic frame, and then revise the syllabic content to the target syllable, thus leading to slower tone-to-syllable integration (Chen & Zhang, 2025). Additionally, since the connection strength determines the speed in the rise of the endpoint nodes' activation (Hameau et al., 2021), the weaker connection between the tonal frame and the syllable motor program (as modulated by the excitation rate) in our simulations means that the activation of the syllable motor program would rise more slowly, which also implies slower tone-to-syllable integration. Therefore, for the two-stage model, both the competition mechanism and the dummy syllable mechanism can successfully simulate the interference effect of tonal overlap without shared syllabic information. The prerequisite for generating this interference effect does not depend on the differences between these two mechanisms, but essentially lies in a slower tone-to-syllable integration process, reflected as either the increased inhibition between atonal syllables or the decreased activation between the tonal frame and the syllable motor program.

In phonologically primed naming tasks, there might be a trade-off between the facilitation and inhibition effects on tonal encoding in different processing stages. Previous studies manipulating syllable frequency in spoken word production also revealed opposite effects. They proposed a hypothesis suggesting that whether the effect is facilitatory or inhibitory arises from a trade-off between the relative disadvantage in lexical access and the advantage in constructing the phonological output of syllables (Hutzler et al., 2005; Perea & Carreiras, 1998). Similarly, we assume a trade-off mechanism of the phonologically priming effects during tonal encoding. In our simulations, the model with the competition mechanism obtained a null effect under the tonal prime condition when the excitation rate between the tonal frame and the syllable motor program decreased to 0.09, while the cutoff point for the model with the dummy syllable mechanism was 0.08. These results indicated that the reduced activation spreading between the tonal frame and the syllable motor program may have equal strengths with the facilitation effect arising from the tonal overlap. This processing advantage from priming and the disadvantage in the tone-to-syllable integration process may cancel each other out in naming latencies. Thus, we assume that for the tonal prime condition, the reduced activation spreading between the lexical tone and the syllable motor program surpassed the facilitation caused by the tonal overlap, leading to the interference effect upon articulation.

This study addressed the controversy about whether the tonal encoding in disyllabic Mandarin word production follows a late-retrieval or two-stage encoding, and contributed to a broader understanding of the two-stage tonal encoding model by highlighting the dynamic interaction between tonal and syllabic representations. Specifically, this study examined the two-stage model with either the competition mechanism or the dummy syllable mechanism from a computational perspective. Since it would be difficult for behavioral or even neuroimaging studies to uncover whether spoken word production involves dummy syllables or not, computational modeling offers a feasible approach to look into the specific internal processes of speech production. A key limitation of the current model lies in the unresolved relationship between phonology and semantics in disyllabic words. Many disyllabic words in Mandarin consist of stand-alone morphemes, which would activate their homophones and phonological or semantic neighbors under the interactive activation framework. For example, the disyllabic word "玉米" (/yu4 mi3/, "corn") can be decomposed into "玉" (/yu4/, "jade") and "米" (/mi3/, "rice"), both of which are stand-alone morphemes with their own semantic representations. In contrast, words like "葡萄" (/pu2 tao2/, "grape") consist of characters with a higher transitional probability that are less likely to activate independent meanings. Future research could examine how transitional probability modulates the activation of phonological and semantic representations during lexical access. Additionally, the current model should be further validated through different tonal languages. Moreover, investigating the interplay between linguistic and cognitive factors (e.g., working memory, attention, etc.) and how frequencies of representational units (e.g., atonal syllables and segments) affect the results would provide a more comprehensive understanding of tonal encoding mechanisms.

Acknowledgments

This study was supported by the conference grant for research postgraduate students at the Hong Kong Polytechnic University.

References

Alderete, J., Chan, Q., & Yeung, H. H. (2019). Tone slips in Cantonese: Evidence for early phonological encoding. *Cognition, 191*, Article 103952.

Bao, L., Qian, Z., & Zhang, Q. (2023). The multiple phonological activation in Chinese spoken word production: An ERP study supporting cascaded model. *Behavioural Brain Research*, *451*, 114523.

Cai, X., Yin, Y., & Zhang, Q. (2020). The roles of syllables and phonemes during phonological encoding in Chinese spoken word production: A topographic ERP study. *Neuropsychologia, 140*, Article 107382.

Chen, J. Y. (1999). The representation and processing of tone in Mandarin Chinese: Evidence from slips of the tongue. *Applied Psycholinguistics, 20*(2), 289–301.

Chen, J. Y., Chen, T. M., & Dell, G. S. (2002). Word-form encoding in Mandarin Chinese as assessed by the implicit priming task. *Journal of Memory and Language, 46*(4), 751–781.

Chen, M. Y. (2000). *Tone sandhi: Patterns across Chinese dialects*. Cambridge: Cambridge University Press.

Chen, Q., & Mirman, D. (2012). Competition and cooperation among similar representations: Toward a unified account of facilitative and inhibitory effects of lexical neighbors. *Psychological Review, 119*(2), 417-430.

Chen, X., & Zhang, C. (2025). Setting the "tone" first and then integrating it into the syllable: An EEG investigation of the time course of lexical tone and syllable encoding in Mandarin word production. *Journal of Memory and Language*, *140*, 104575.

Chen, X., Zhang, C., Chen, Y., Politzer-Ahles, S., Zeng, Y., & Zhang, J. (2022). Encoding category-level and context-specific phonological information at different stages: An EEG study of Mandarin third-tone sandhi word production. *Neuropsychologia, 175*, Article 108367.

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review, 93*(3), 283–321.

Dell, G. S. (1988). The retrieval of phonological forms in production: Tests of predictions from a connectionist model. *Journal of memory and language*, *27*(2), 124-142.

Farrell, S., & Lewandowsky, S. (2018). *Computational modeling of cognition and behavior*. Cambridge University Press.

Feng, C., Yue, Y., & Zhang, Q. (2019). Syllables are retrieved before segments in the spoken production of Mandarin Chinese: An ERP study. *Scientific Reports, 9*(1), 11773.

Hameau, S., Biedermann, B., Robidoux, S., & Nickels, L. (2021). Effects of phonological neighbourhood density and frequency in picture naming. *Journal of Memory and Language, 120*, Article 104248.

Hutzler, F., Conrad, M., & Jacobs, A. M. (2005). Effects of syllable-frequency in lexical decision and naming: An eye-movement study. *Brain and Language, 92*(2), 138-152.

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences, 22*(1), 1–38.

Liu, Y., Zhao, Z., Xu, M., Yu, H., Zhu, Y., Zhang, J., ... & Wu, J. (2023). Decoding and synthesizing tonal language speech from brain activity. *Science Advances*, *9*(23), eadh0478.

McClelland, J. L., & Rumelhart, D. E. (1981). An interactive activation model of context effects in letter perception: I. An account of basic findings. *Psychological Review, 88*(5), 375-407.

O'Seaghdha, P. G., Chen, J. Y., & Chen, T. M. (2010). Proximate units in word production: Phonological encoding begins with syllables in Mandarin Chinese but with segments in English. *Cognition, 115*(2), 282–302.

Palminteri, S., Wyart, V., & Koechlin, E. (2017). The importance of falsification in computational cognitive modeling. *Trends in Cognitive Sciences, 21*(6), 425-433.

Perea, M., & Carreiras, M. (1998). Effects of syllable frequency and syllable neighborhood frequency in visual word recognition. *Journal of Experimental Psychology: Human Perception and Performance, 24*(1), 134-144.

Roelofs, A. (1997). The WEAVER model of word-form encoding in speech production. *Cognition, 64*(3), 249-284.

Roelofs, A. (2015). Modeling of phonological encoding in spoken word production: From Germanic languages to Mandarin Chinese and Japanese. *Japanese Psychological Research, 57*(1), 22–37.

Yip, M. (2002). *Tone*. Cambridge: Cambridge University Press.

Zeng, Y., Fiorentino, R., & Zhang, J. (2022). Experimental investigation of Mandarin lexical tone production. In *Proceedings ExLing 2022: 13th International Conference of Experimental Linguistics* (pp. 177–180).

Zhang, Q. (2008). Phonological encoding in monosyllabic and bisyllabic Mandarin word production: Implicit priming paradigm study. *Acta Psychologica Sinica, 40*(03), 253–262.

Zhang, Q., & Damian, M. F. (2009). The time course of segment and tone encoding in Chinese spoken production: An event-related potential study. *Neuroscience, 163*(1), 252–265.

Zhang, Q., & Damian, M. F. (2019). Syllables constitute proximate units for Mandarin speakers: Electrophysiological evidence from a masked priming task. *Psychophysiology, 56*(4), 1–15.

Zhang, Q., & Zhu, X. (2011). The temporal and spatial features of segmental and suprasegmental encoding during implicit picture naming: An event-related potential study. *Neuropsychologia, 49*(14), 3813–3825.
Zhou, X., & Zhuang, J. (2000). Lexical tone in the speech production of Chinese words. *6th International Conference on Spoken Language Processing, ICSLP 2000*, Beijing, China.