Journal of Neurolinguistics 51 (2019) 184–193

Contents lists available at ScienceDirect

Journal of Neurolinguistics

journal homepage: www.elsevier.com/locate/jneuroling

The fundamental phonological unit of Japanese word production: An EEG study using the picture-word interference paradigm

Rinus G. Verdonschot[a,*], Shingo Tokimoto[b], Yayoi Miyaoka[c]

[a] Graduate School of Biomedical & Health Sciences - Hiroshima University, Japan
[b] Foreign Language Studies, Mejiro University, Japan
[c] Department of Liberal Arts Education - Hiroshima University of Economics, Japan

ARTICLE INFO

Keywords:
Language production
Phonology
EEG/ERP
Japanese

ABSTRACT

It has been shown that in Germanic languages (e.g. English, Dutch) phonemes are the primary (or proximate) planning units during the early stages of phonological encoding. Contrastingly, in Chinese and Japanese the phoneme does not seem to play an important role but rather the syllable (Chinese) and mora (Japanese) are essential. However, despite the lack of behavioral evidence, neurocorrelational studies in Chinese suggested that electrophysiological brain responses (i.e. preceding overt responses) may indicate some significance for the phoneme. We investigated this matter in Japanese and our data shows that unlike in Chinese (for which the literature shows mixed effects), in Japanese both the behavioral and neurocorrelational data indicate an important role only for the mora (and not the phoneme) during the early stages of phonological encoding.

The ability to speak represents one of human's distinctive developments during the course of evolution. This ability is quite remarkable considering that through speaking and comprehending the limited sets of speech sounds available in the world's languages, one can convey an almost unlimited amount of information. It comes therefore as no surprise that much research has been devoted to the acquisition, understanding and production of speech. Focusing on the latter, several theoretical models have been created to describe the architecture of the speech production system (e.g. Caramazza, 1997; Dell, 1986; Levelt, Roelofs, & Meyer, 1999). Although specifics differ, most of these models agree that speech production can be divided into two successive steps, specifically: (1) the generation of the meaning of the utterance, followed by (2) the generation of speech sounds (i.e. phonological encoding). The current paper focuses on the latter part, especially whether the unit used in the initial stages of the phonological encoding process is similar or differs between languages.

The most explicit account of how phonological encoding takes place is found in Levelt et al. (1999). For example, upon lexical selection, the production of a word such as *modest* (/ˈmɒdɪst/) has two pieces of stored information which become available, that is, parallel activation of all the phonemes involved (i.e. /m/, /ɒ/, /d/, /ɪ/, /s/, /t/) and activation of its metrical structure (i.e. bisyllabic with stress on the first syllable). Once this information has become available, a so-called *prosodification* (i.e. incremental segment-to-frame association) process initiates. That is, segments (phonemes) are subsequently inserted into their correct position in the metrical frame which ultimately results in the prosodified word /ˈmɒdɪst/. That this is not simply a case of converting a stored representation to a pronunciation can be seen in re-syllabications (e.g. such as the movement of /st/ to the next syllable/word in 'modest aunt' /ˈmɒdɪ.ˈstant/) as they would be difficult to store due to the sheer number of possible combinations.

* Corresponding author. Department of Oral and Maxillofacial Radiology, Graduate School of Biomedical & Health Sciences, Hiroshima University, 1-2-3 Kasumi, Minami-ku, Hiroshima, 734-8553, Japan.
E-mail address: rinusverdonschot@gmail.com (R.G. Verdonschot).

https://doi.org/10.1016/j.jneuroling.2019.02.004
Received 27 September 2018; Received in revised form 14 February 2019; Accepted 15 February 2019
0911-6044/ © 2019 Elsevier Ltd. All rights reserved.

Behavioral evidence advocating an incremental segment-to-frame association process initially came from the implicit priming task (e.g. Meyer, 1990). In this task, participants initially need to learn prompt-response pairs (e.g. 'sea – boat') which will then be used in a cueing paradigm. That is, upon seeing a prompt (e.g. 'sea'), a target word (e.g. 'boat') would have to be produced. In this way, response groups can be controlled which share partial phonological overlap (e.g. boat, bear, ball) or not (e.g. boat, rain, house). While it has been shown in this task that onset overlap is needed to facilitate naming latencies, if the rhyme overlaps but not the onset (e.g. boat, coat, goat), typically no facilitation occurs (Meyer, 1990; but see; Wong, Huang, & Chen, 2012). Similar findings confirming that phonological encoding occurs in an incremental fashion using a segment-to-frame process have been obtained using different paradigms such as masked priming reading aloud (e.g. Forster & Davis, 1991; Kinoshita, 2000) and as well as picture naming (e.g. Meyer & Schriefers, 1991; Schiller, 2008; but see; Lupker, 1982).

Recently, however, it has been suggested (e.g. Chen, Chen, & Dell, 2002; O'Seaghdha, Chen, & Chen, 2010) that the basic phonological unit to fill the metrical frame might not be similar among languages. Specifically, it has been proposed that this unit would be the syllable for Chinese (O'Seaghdha et al., 2010; You, Zhang, & Verdonschot, 2012) and the mora[1] for Japanese (Kureta, Fushimi, & Tatsumi, 2006; Verdonschot et al., 2011). Some of the differences forming the rationale for different phonological units between these languages originate from speech error patterns (e.g. mora switches but not phoneme switches in Japanese; see Kubozono, 1989), the exact speech rhythm a language adheres to (i.e. moraic for Japanese, see Warner & Arai, 2001; but syllabic for Mandarin Chinese; see Lin & Wang, 2007) and other linguistic phenomena (e.g. the absence or presence of re-syllabification). For a comprehensive overview of several key differences between Germanic languages, Mandarin Chinese and Japanese and an elaborate interpretation of how the Levelt et al. (1999) model can handle these differences see Roelofs (2015).

It has been shown for Mandarin Chinese and Japanese that while phonemic onset effects do not appear in the earlier mentioned paradigms (i.e. implicit priming, masked priming, and picture naming), atonal syllable effects (Mandarin Chinese) and mora effects (Japanese) do appear. Additionally, Roelofs (2015) has argued, through the use of computer simulations, that the Levelt et al. (1999) model is able to accommodate these findings by assuming different initial phonological building blocks but keeping the phonological encoding process the same.

Interestingly, however, is that others did find sub-syllabic effects for Chinese languages. For example, Verdonschot, Lai, Feng, Tamaoka, & Schiller, 2015; using monolingual Mandarin Chinese participants, reported sub-syllabic priming effects (e.g. target: 山 /shan1/ - prime: 沙 /sha1/); though further replication of these findings is needed. Additionally, findings from Cantonese (another Chinese language) using the picture-word interference paradigm (PWI) suggest that sub-syllabic effects readily appear in this language.

The PWI task is a well-known task used for many decades in language production research in various languages (e.g. Glaser & Düngelhoff, 1984; Schriefers, Meyer, & Levelt, 1990; Starreveld & La Heij, 1996; Wong & Chen, 2008; Verdonschot, La Heij, & Schiller, 2010). The objective is to name a picture (e.g. dog) and a distractor word is printed on it which usually has a semantic (e.g. cat) or a phonological relationship (e.g. dark) with the target. Typically, semantically related distractor words slow down naming latencies (relative to unrelated distractors) while phonologically related distractors induce facilitation. There are also auditory variants of the task in which the distractor word is not shown but played via a speaker or headset (e.g. Damian & Bowers, 2009; Wong, Wang, Wong, & Chen, 2018). Stimulus onset asynchrony (SOA) manipulations in which the target picture and the word distractor have a certain time interval between them have provided evidence for a staged model of language production (i.e. semantics before phonology; see Levelt et al., 1999 for an overview). The PWI task is believed to be able to tap into the phonological encoding process and to be informative of the phonological unit. For example, Wong and Chen (2008) obtained significant facilitation effects for the syllable overlap condition (e.g. picture: 箭 /zin3/ 'arrow' with distractor: 氈 /zin1/ 'felt') and also found significant facilitation effects when the rhyme (without tone; e.g. 淺 /cin2/ 'shallow') was shared (though only when they were presented auditorily). No effects were found when only the phonemic onset was shared (滯 /zai6/ 'stagnation'). This suggests that sub-syllabic units might be involved in the initial phonological planning in Cantonese (see Roelofs, 2015). These results were further examined in Wong and Chen (2009) who, using a PWI paradigm, also showed that significant facilitation could be observed when the same initial consonant and vowel were shared between the picture name and the distractor word (e.g. picture: 星 /sing1/, meaning "star" and distractor 食 /sik6/ "food"). The notion that sub-syllabic units are legitimate planning units in Cantonese was further substantiated by the same group of authors through the implicit priming task (i.e. Wong et al., 2012) and seems to be a reliable phenomenon in this language[2]. Note, though, that rhyme effects are somewhat hard to reconcile with the incremental nature of the prosodification process (e.g. Levelt et al., 1999) and Wong and Chen (2008) have proposed that interactive models (e.g. Dell, 1986) might be more suitable to account for this type of result.

Few studies have reported *phoneme* effects for monolingual Mandarin Chinese speakers. For example, Qu, Damian, and Kazanina (2012) investigated the role of the phoneme in Mandarin language production using a behavioral task in combination with EEG (i.e. brain wave) recordings. In their experiment, Mandarin native speakers had to name colored line-drawings of objects using color adjective-noun phrases in which the color and the object name either shared the initial phoneme (e.g. 'red train' or 红 – /hong2/ – 火车 – /huo3che1/) or were phonologically unrelated (e.g. 'blue train' or 蓝 – /lan2/ – 火车 – /huo3che1/). Note that syllable overlap

[1] The mora is a unit which can be V (e.g. あ /a/), CV (e.g. か /ka/), Q (a geminate), N (i.e. ん; a nasal coda) and R (long vowel; e.g. う in ほう) in Japanese. A single consonant (except for ん) can never be a mora in Japanese. For example, the word/ni.hoN/'Japan' consists of 3 moras (i.e. CV.CV.N) but 2 syllables (ni + hoN).

[2] It is unclear whether participants in these Cantonese studies may have been fluent in English as an L2, for details how this might have affected the results, see: Verdonschot, Nakayama, Zhang, Tamaoka, & Schiller, 2013.

was not manipulated. They found that phoneme overlap did not lead to behavioral changes (i.e. reaction times, henceforth RTs, for the overlap condition were not faster) but they did find effects of phoneme (onset) overlap in their EEG data. Specifically, they found significantly more positive event related potentials (ERPs) in posterior regions roughly 200–300 ms after picture onset as well as more negative ERPs in anterior regions about 300–400 ms after picture onset (compared to control conditions). They attributed this earlier posteriorly located effect to phonological encoding processes (see Levelt et al., 1999, p. 15) while the latter negative effect was attributed to monitoring processes by the language production system (see Levelt et al., 1999, p. 33). Importantly, Qu et al. (2012) stated that by merely investigating behavioral effects, any potential involvement of the phoneme in the initial phonological encoding processes of Mandarin Chinese might be overlooked.

Similarly, Yu, Mo, and Mo (2014) also measured ERPs using a picture priming task in which participants continuously needed to name pictures with the relationship between consecutive picture names being manipulated (e.g. for a phonologically related condition a picture of a 'candle' /la4zhu2/ was followed by a picture of a 'tie' /ling3dai4/ so the /l/ onset phoneme was repeated). They found more positive ERPs between 180 ms and 300 ms for phonologically related conditions and more negative ERPs between 350 ms and 450 ms post-picture onset which is in line with the results from Qu et al. (2012).

Both studies, however, did not manipulate phonemic onset versus the syllable as they only investigated the onset effect. Recently, using the same paradigm as Yu et al. (2014), Wang, Wong, Wang, and Chen (2017), though not manipulating the onset, did contrast a full syllable overlap (土豆 'tomato' /tu3dou4/ – 兔子 'rabbit' /tu4zi/) with sub-syllabic overlap (鼻子 'nose' /bi2zi/ – 冰箱 'refrigerator' /bing1xiang1/). They found that only syllable repetition produced significant effects on ERPs. In particular, they observed a widely distributed positivity between 200 ms and 400 ms and a significant anterior positivity between 400 ms and 600 ms. Interestingly, this contrasts earlier findings (from the same lab) showing *behavioral* sub-syllabic effects for both Mandarin and Cantonese. Wang et al. (2017) speculate that this finding might have been caused due to the nature of the task. That is, to avoid muscle artifacts, they used a delayed naming task in which participants need to wait 800 ms before starting to speak, which, in turn, might have diminished any priming effect of repetition. Whether this interpretation is accurate will need to be investigated in future studies which do not employ a delayed procedure.

One thing which stands out from this summary is that while work investigating the neural underpinnings of the phonological unit of language production is steadily progressing in Chinese and European languages, Japanese is thoroughly under investigated. For instance, as far as we know, no study exists using Japanese which concurrently records EEG and investigates the phonological unit of language production. That is surprising as Japanese is particularly helpful to the debate on phonological encoding units during prosodification as it is one of the few mora-timed languages in the world in which clear predictions can be made and reliable results have been obtained. For example, Kureta et al. (2006) using an implicit priming paradigm showed that only mora overlap (e.g. かつら /katsura/ 'wig', 歌舞伎 /kabuki/ 'kabuki theatre' and 鞄/kaban/ 'bag') but not phoneme overlap (e.g. かつら /katsura/, くじら /kujira/ 'whale', and 古墳 /kofun/'mound') elicited facilitation effects. Additionally, this pattern has also been found using the masked priming paradigm (Verdonschot et al., 2011) as well as the phonological Stroop task (Verdonschot & Kinoshita, 2018).

Such results indicate a major role for the mora but not for phonemic segments. Particularly, considering the conflicting EEG/ERP results on other languages in the literature (e.g. Chinese) indicating that the phonemic segments may (or may not) play a role during the initial construction of phonology, it is essential to add Japanese to this debate. Accordingly, this paper reports data from a Japanese picture-word interference paradigm for which concurrent EEG/ERP is recorded. The main point of interest is whether, when only the onset is overlapping between the to-be-named picture (e.g. /kutsu/ 'shoe') and a (nonword) distractor (e.g. けや/keya/), behavioral and/or ERP effects will emerge or not (compared to control). If the phoneme plays a role at some point during the prosodification process in Japanese, then EEG/ERP effects might appear even in the absence of behavioral effects (similar to the Chinese data in Qu et al., 2012 and Yu et al., 2014). With respect to the mora, we expect both significant behavioral and EEG/ERP effects for distractors which share a *mora* with the picture name (e.g. くや /kuya/) versus the control as this is the fundamental phonological unit used during prosodification in Japanese. In fact, such a result would be one of the first instances in the literature showing full congruency between behavioral and brain data when the proposed phonological unit overlaps (and when it does not).

1. Experiment – PWI task using C and CV non-word distractors with concurrent EEG recording

1.1. Method

1.1.1. Participants

Twenty-four participants (Mean age: 20.5 ± 1.3; nineteen males) studying at Hiroshima University of Economics were paid 2000 Japanese Yen (18.5 USD) for participation. All were native speakers and native readers of Japanese and had normal or corrected-to-normal vision.

1.1.2. Design

The experiment used a picture word interference (PWI) naming task and manipulated: (1) distractor type (i.e. onset vs. mora), and (2) overlap (i.e. overlap vs control). The dependent variables were response latency and accuracy.

1.1.3. Materials

The stimuli consisted of 63 pictures which, whenever available, were taken from the Snodgrass and Vanderwart (1980) picture set; though, occasionally some pictures had to be created when the intended words were not present (e.g. 'kimono', 'torii gate'). Care was taken that these pictures matched well with the pictorial characteristics of the Snodgrass & Vandewart set (e.g. black and white

line drawings). Prior to the experiment a booklet was shown which had the pictures and their correct names in a suitable Japanese script displayed for learning purposes (see the Appendix for an overview). After that, another booklet was presented with the pictures (different order) without the name written down to assess whether participants mastered all picture names before the experiment proper started. Corrections were given when needed. For each picture, four hiragana non-words were created which had a relationship with the picture (e.g. /kugi/ 'nail'), that is: (1) onset overlap (e.g. けぴ /kepi/), (2) onset control (e.g. めぴ /mepi/), (3) mora overlap, (e.g. くぴ /kupi/), and (4) mora control (e.g. むぴ /mupi/). Nonwords were chosen to avoid any lexical effects from the distractors as well as to create a suitably balanced set (Lupker, 1982). See the Appendix for an overview of all stimuli used. Eight practice pictures (not used in the experiment proper) to practice the task, as well as warmups for each block were given to train participants to the task.

1.1.4. Apparatus and procedure

Participants were first shown a blinking interval (marked by a pair of eyes on the screen) which had a random duration from 1200 to 1900 ms in which they were allowed to blink their eyes. Then a fixation screen (+) appeared for 1000 ms after which the picture target appeared (simultaneously with the distractor word printed on it) which they had to name. After naming, the picture disappeared from the screen and the experimenter judged each trial whether the trial was correct or whether a voicekey error (e.g. coughing, accidental triggering) or a genuine error (e.g. wrong word) had occurred. After this, the trial started again. E-Prime 2.0 was used for the presentation of the stimuli and collection of responses (see Spapé, Verdonschot, van Dantzig, & van Steenbergen, 2014). The experiment contained 252 trials (63 pictures x 4 conditions [i.e. C, C-control, CV, CV-control) and was divided into four blocks. The order of trials within each experiment was pseudo-randomised using Mix (Van Casteren & Davis, 2006) with three restraints (1) pictures of the same semantic category (e.g. tiger, cat): could not directly follow each other (2) same condition stimuli (e.g. 'nail' /kugi/ – /kepi/ and 'carrot' /nin.jin/ – /nubo/; both C-overlap), could not directly follow each other and (3) same onset pictures could not directly follow each other (e.g. 'nail' /kugi/ and 'ticket' /kippu/).

1.1.5. EEG recording and preprocesses

A portable 16-channel EEG-System (V-AMP: Brain Products, München, D) was used for data acquisition with a sampling rate of 500 Hz. An EEG-actiCap (Brain Products, München, D) which is an electrode cap consisting of high-quality Ag/AgCl active electrodes was applied prior to the experiment. EEG activity was recorded using the international 10–20 system (Klem, Lueders, Jasper, & Elger, 1999) on positions F3, Fz, F4, T3, C3, Cz, C4, T4, T5, P3, Pz, P4, T6 with a reference electrode on the right ear and the ground was set on Fp2. If the impedance of an electrode exceeded 5 kΩ, then extra gel was put in the electrode or it was excluded from further analyses. Electrodes were applied to canthi of the right eye (HEOG) and above and below the left eye (VEOGu and VEOGl) for post-hoc eye blink removal procedures.

The acquired EEG data were processed offline using EEGLAB (Delorme & Makeig, 2004). The preprocessing proceeded as follows. (1) The EEG data were downsampled to 250 Hz. (2) The data were high pass-filtered at 1 Hz to minimize slow drifts. (3) Line noise was removed using the CleanLine plug-in within EEGLAB. (4) An artifact subspace reconstruction was used to remove high-amplitude artifacts from the EEG data (Mullen et al., 2015). (5) The EEG data were then re-referenced to a common average reference. (6) The data were decomposed using adaptive mixture independent component analysis (AMICA) (Palmer, Kreutz-Delgado, Rao, & Makeig, 2007). (7) A best-fitting single equivalent current dipole was calculated for each independent component (IC) to match the scalp projection of each IC source using a standardized three shell boundary element head model. The electrode locations corresponding to the extended 10–20 system were aligned with a standard brain model (Montreal Neurological Institute). (8) ICs were excluded from further analysis for which the equivalent dipole model was located outside the brain or explained less than 85% of variance of the corresponding IC scalp map. Furthermore, ICs were removed that were judged by their frequency spectra to be related to eye blinks, horizontal and vertical eye movements, and electromyogram. (9) The data were segmented into time epochs relative to the presentation of the picture targets from −1 to 4 seconds around the onsets of the pictures.

2. Results

2.1. Behavioral results

For the analysis we used mixed effects modelling, treating subjects and stimuli as crossed random effect factors. RTs were logarithmically transformed in order to meet the distributional assumption of the model. Voicekey errors were excluded from further analysis (3.2% of the data) and we used a ± 2.5 SD cutoff for outliers (2.8% excluded). In total 5600 data points were used for analysis. We used the *lme4* package (Version 1.1–17, Bates, Mächler, Bolker, & Walker, 2015), implemented in R Version 3.5.0 (R Core Team, 2018). Degrees of freedom (Satterthwaite's approximation) and *p*-values were estimated using the *lmerTest* package (Version 3.0–1; Kuznetsova, Brockhoff, & Christensen, 2017). In line with the recommendation to keep the random effect structure maximal (Barr, Levy, Scheepers, & Tily, 2013), the initial model included random slopes on participants and stimuli; the final model we report was selected using a backward stepwise model selection procedure. The final formula was logRT ~ Trial + Condition + (1 | Participant) + (1 | Item). We found there was a main effect of Trial (Estimated coefficient: 0.0007, $t = 5.5$, $p < .001$) indicating that participants got slightly slower during the course of the experiment. Multiple comparisons using the *multcomp* package (Hothorn, Bretz, & Westfall, 2009) showed that there was a significant 42 ms effect when comparing the CV (mora) condition versus its control (Estimated coefficient: 0.0619, $z = 9.8$, $p < .001$) but no significant effect for the C (onset) condition when comparing it via its control (Estimated coefficient: 0.0149, $z = -2.3$, $p = .09$). In fact, if anything, it took longer to name the picture when a C

Table 1
Behavioral RTs (in ms; SD between parentheses) and accuracy (in % error).

Condition	Overlap	%E	Control	%E
C (onset)	767 (173)	1.4	754 (162)	1.5
CV (mora)	702 (140)	1	744 (150)	1.7

overlapping distractor was superimposed (see Table 1). Concerning the error analysis, in total there were 1.4% errors which were roughly equally distributed between conditions. An analysis of errors rates using a logistic link function did not reveal any differences between conditions (all *t*'s < 1.1).

2.2. Electrophysiological results

Event-related potentials (ERPs) were calculated for the epochs that were followed by the correct naming of the picture targets, with the baseline set as 100 ms before the onset of the picture. To test the significance of the condition effects, three-way ANOVAs were carried out for the mean amplitudes of every 100 ms time-window from 200 to 800 ms latency in six regions of interest (ROIs) using the C vs. CV contrast, the presence or the absence of the overlap, and the ROI as within factors. The grouping of the thirteen electrodes into the six ROIs was as follows. Frontal: F3, Fz, F4; left centro-temporal: C3, T3; right centro-temporal: C4, T4; centro-parietal: Cz, Pz; left temporo-parietal: P3, T5; right temporo-parietal: P4, T6. The *F*-values for the significant main effects and the significant interactions are summarized in Table 2 together with their significance levels.

We found significant interactions of C(V) x Overlap and of C(V) x Overlap x ROI, and consequently, we examined the significance of the contrasts between CV-overlap and CV-control and between C-overlap and C-control. To examine the significance, non-parametric random permutation statistics were calculated using the STUDY command structure of EEGLAB. In the current study, 2000 random permutations were computed and compared to *t*-values for the mean condition differences.

The mean topography of ERPs from 400 to 500 ms for CV-overlap and CV-control are presented in Fig. 1A, and the mean ERPs of the two conditions at the representative electrodes (i.e. F3, Fz, and Cz) are shown in Fig. 1B.

We observed a significant positive deflection for CV-overlap against CV-control in the time-window of 380–490 ms (Fig. 1B). We found no time window in which the contrast between the two conditions was significant for the HEOG, VEOGu, and VEOGl electrodes.

Contrastingly, there were no significant contrasts found between C-overlap and C-control in any time-window between 0 and 1000 ms at all electrodes using the cluster-based permutation test. The mean topography of ERPs from 200 to 300 ms and from 300 to 400 ms for C-overlap and C-control are presented in Fig. 2A and B, and the ERPs of the two conditions at Fz and Pz are presented in Fig. 2C and D.

3. Discussion

To our knowledge, the current study is the first to compare the ERP effects of mora versus phoneme segments in Japanese spoken word production. To this end, a picture-word interference paradigm was run with concurrent recordings of electric brain potentials. In this paradigm, participants needed to name a picture while ignoring a superimposed distractor word.

Behaviorally we found significant effects when the CV was overlapping between picture name (e.g. /kippu/ 'ticket') and mora non-word distractors (e.g. /kiho/ vs /niho/) but not for onset distractors (e.g. /kaho/ vs. /naho/). This pattern was closely reflected in the EEG results. While there were no significant effects for onset distractors (e.g. picture of a carrot or /ninjin/ in Japanese with distractors ぬぼ /nubo/ vs むぼ /mubo/), there was a significant positive deflection roughly within the 400–500 ms time range for the CV related distractors (e.g. にぼ /nibo/vs. みぼ /mibo/). This is in line with Wang et al. (2017) who, using sequential picture naming, also observed an anterior positivity between 400 and 600 ms for the predicted unit (i.e. syllable) in Cantonese. Our findings resonate well with the time frame in which prosodification has been proposed to occur after all the phonological code has been retrieved. According to Indefrey and Levelt (2004) prosodification (which is the incremental generation of the phonological word given the phonological unit and the metrical/tonal frame; see Levelt et al., 1999, p. 22) should start approximately after 330 ms post stimulus onset and proceeds roughly at a pace of 25 ms per segment (for phonemes). That means that at this point segments (phonemes) and

Table 2
Summary of F-values by ANOVAs on mean ERP amplitudes for every 100 ms time-windows from 200 to 800 ms latency (C(V): C vs. CV contrast; Overlap: overlap vs. no overlap (control); ROI: region of interest. The degrees of freedom for the effects are given in parentheses).

	200–300 ms	300–400 ms	400–500 ms	500–600 ms	600–700 ms	700–800 ms
C(V) (1, 23)		3.07^{+}	35.68^{***}		3.38^{+}	
Overlap (1, 23)			106.27^{***}			
ROI (5, 115)	54.95^{***}	43.16^{***}	38.56^{***}	38.70^{***}	8.49^{***}	
C(V) x Overlap (1, 23)			43.15^{***}			
C(V) x ROI (5, 115)			56.02^{***}		2.06^{+}	
Overlap x ROI (5, 115)			73.23^{***}			
C(V) x Overlap x ROI (5, 115)			58.67^{***}	2.64^{*}	2.81^{*}	

$^{+}p < .1$, $^{*}p < .05$, $^{***}p < .001$.

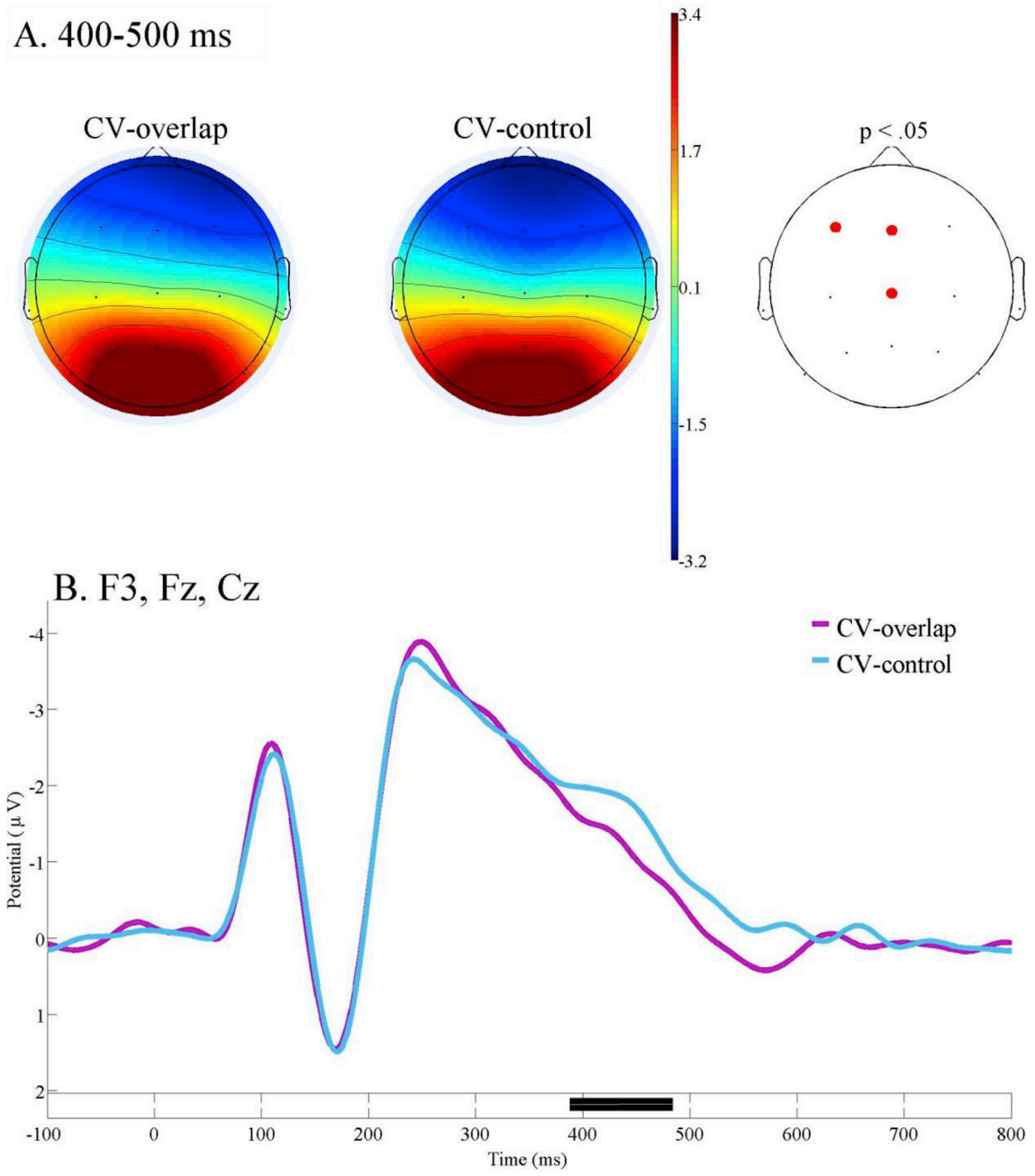

Fig. 1. ERPs of CV-overlap and CV-control time-locked to the onsets of picture targets. (A) Mean topography of ERPs from 400 to 500 ms for the two conditions, and electrodes in red for which significant differences were found by the cluster-based permutation test (Maris & Oostenveld, 2007) ($p < .05$). (B) Mean ERPs of the two conditions at F3, Fz, and Cz. Negativity is plotted upward, and the time windows in which significant differences were found by cluster-based permutation test are indicated by black on the time axis in B. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

metrical frame information is being combined to produce the correct syllable motor program.

Port, Dalby, and O'Dell (1987) have shown that the duration of Japanese words also increases in constant increments depending on the number of morae. However, at this point it is difficult to speculate how long a *mora* would take to be prosodified and whether the timing values of Indefrey and Levelt (2004) are applicable to Japanese phonological encoding[3]. Still, following Indefrey and Levelt (2004), it seems that the observed positive EEG deflection in our study is situated at time window which is assumed to reflect

[3] note: and perhaps even whether such exact pinpointing is reliable as RTs vary in most psycholinguistic experiments.

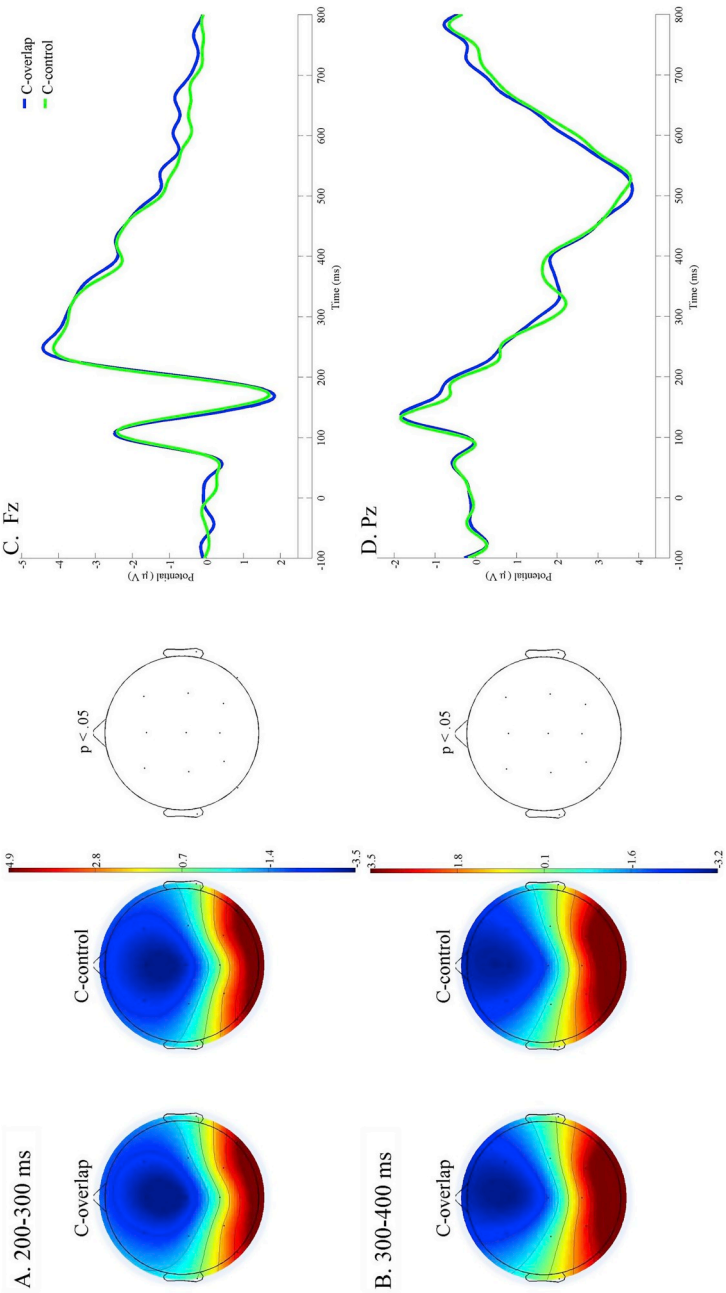

Fig. 2. ERPs of C-overlap and C-control time-locked to the onsets of picture targets. Mean topography of ERPs from 200 to 300 ms (A) and that from 300 to 400 ms (B) for the two conditions. (C) Mean ERPs of the two conditions at Fz (C) and P3 (D). Negativity is plotted upward.

the neuronal underpinnings of the prosodification process (see Levelt et al., 1999, p. 22; and Roelofs, 2015, p. 31; for Japanese).

It seems therefore that in Japanese the phoneme does not play a major role during the prosodification process. Earlier studies using Chinese (i.e. Mandarin) in spite of the absence of behavioral evidence did find neurocorrelational evidence (e.g. Qu et al., 2012; Wang et al., 2017; Yu et al., 2014) indicating a role for the phoneme during prosodification despite of lack of behavioral evidence. However, the current study did not obtain similar results for Japanese, therefore indicating that, even during the recording of real-time brain data which precedes overt responses, the phoneme does not play a significant role in Japanese.

A possible restriction of our study might be the limited number of channels on the apparatus which was used to record EEG (i.e. BrainVision V-Amp 16 from Brain Products). However, this seems unlikely, as it is suitable to detect key ERP components (e.g. *P300*: e.g. Amaral, Simões, & Castelo-Branco, 2015; e.g. Balconi, Crivelli, & Vanutelli, 2018; *N400*: Usai, O'Neil, & Newman, 2018; *P600*: e.g. Lancier, 2016). Additionally, it has been shown that electrodes in larger channel systems show a very high correlation between closely spaced channels due to the fact that the electrocortical source signals are typically not generated immediately below the electrode. Given the fact that our setup was able to detect differences in the CV condition, it is reasonable to assume that the absence of effects in the C condition does reflect a genuine outcome and is not simply due to having a sparse number of channels (otherwise the CV condition would have shown a similar result).

In conclusion, despite of the mixed evidence between behavioral and neurocorrelational data in Chinese, the case for Japanese seems more clear-cut. Both behavioral and ERP data indicate a significant role for moras, not phonemes, during the early stages of phonological encoding (i.e. prosodification).

Acknowledgements

This research is supported by a JSPS Grant-in-Aid for Scientific Research (C) (18K00857) to YM and a JSPS Grant-in-Aid for Scientific Research (C) (17K02748) to RV. The authors would like to thank Michael P. Mansbridge for his insightful comments on this manuscript.

Appendix. Experimental stimuli

Picture	Name in Japanese	C	trans	C-con	trans	CV	trans	CV-con	trans
nail	釘	けぴ	kepi	めぴ	mepi	くぴ	kupi	むぴ	mupi
spinning top	こま	きぜ	kize	みぜ	mize	こぜ	koze	もぜ	moze
shoe	靴	けや	keya	めや	meya	くや	kuya	むや	muya
jellyfish	クラゲ	けば	keba	れば	reba	くば	kuba	るば	ruba
kimono	着物	こは	koha	そは	soha	きは	kiha	しは	siha
ticket	切符	かほ	kaho	なほ	naho	きほ	kiho	にほ	niho
kendama toy	けん玉	くざ	kuza	るざ	ruza	けざ	keza	れざ	reza
compass	コンパス	くた	kuta	すた	suta	こた	kota	そた	sota
crayon	クレヨン	かぺ	kape	らぺ	rape	くぺ	kupe	るぺ	rupe
monkey	猿	しぐ	sigu	にぐ	nigu	さぐ	sagu	なぐ	nagu
suzu (bell)	鈴	さに	sani	らに	rani	すに	suni	るに	runi
cicada	セミ	そだ	soda	ほだ	hoda	せだ	seda	へだ	heda
water mill	水車	しへ	sihe	みへ	mihe	すへ	suhe	むへ	muhe
bird house	巣箱	さろ	saro	なろ	naro	すろ	suro	ぬろ	nuro
water mellon	スイカ	せぽ	sepo	ねぽ	nepo	すぽ	supo	ぬぽ	nupo
drinking bottle	水筒	そび	sobi	もび	mobi	すび	subi	むび	mubi
vacuum cleaner	掃除機	さべ	sabe	らべ	rabe	そべ	sobe	ろべ	robe
slipper	スリッパ	せほ	seho	けほ	keho	すほ	suho	くほ	kuho
octopus	タコ	とせ	tose	ろせ	rose	たせ	tase	らせ	rase
bamboo	竹	てぞ	tezo	へぞ	hezo	たぞ	tazo	はぞ	hazo
tiger	虎	たぜ	taze	させ	saze	とぜ	toze	そぜ	soze
torii gate	鳥居	てぶ	tepu	せぶ	sepu	とぶ	topu	そぶ	sopu
television	テレビ	とぬ	tonu	そぬ	sonu	てぬ	tenu	せぬ	senu
clock	時計	てざ	teza	めざ	meza	とざ	toza	もざ	moza
glove	手袋	とな	tona	もな	mona	てな	tena	めな	mena
playing cards	トランプ	てぱ	tepa	ねぱ	nepa	とぱ	topa	のぱ	nopa
onion	玉ねぎ	とぺ	tope	そぺ	sope	たぺ	tape	さぺ	sape
cat	猫	にぞ	nizo	しぞ	sizo	ねぞ	nezo	せぞ	sezo
eggplant	ナス	ねひ	nehi	れひ	rehi	なひ	nahi	らひ	rahi
rainbow	虹	のへ	nohe	こへ	kohe	にへ	nihe	きへ	kihe
mouse	ネズミ	のさ	nosa	ろさ	rosa	ねさ	nesa	れさ	resa
notebook	ノート	ぬが	nuga	くが	kuga	のが	noga	こが	koga
chicken	鶏	なゆ	nayu	らゆ	rayu	にゆ	niyu	りゆ	riyu
carrot	ニンジン	ぬぼ	nubo	むぼ	mubo	にぼ	nibo	みぼ	mibo
necktie	ネクタイ	にも	nimo	りも	rimo	ねも	nemo	れも	remo
chopsticks	箸	へぶ	hepu	けぶ	kepu	はぶ	hapu	かぶ	kapu
snake	ヘビ	ほざ	hoza	のざ	noza	へざ	heza	ねざ	neza
book	本	へぎ	hegi	せぎ	segi	ほぎ	hogi	そぎ	sogi
scissors	ハサミ	への	heno	ての	teno	はの	hano	たの	tano

broom	ほうき	ひむ	himu	しむ	simu	ほむ	homu	そむ	somu
sheep	羊	ほせ	hose	こせ	kose	ひせ	hise	きせ	kise
sunflower	ひまわり	はべ	habe	まべ	mabe	ひべ	hibe	みべ	mibe
airplane	飛行機	ほて	hote	ろて	rote	ひて	hite	りて	rite
ashtray	灰皿	へご	hego	せご	sego	はご	hago	さご	sago
insect	虫	まそ	maso	なそ	naso	むそ	muso	ぬそ	nuso
window	窓	もぬ	monu	ほぬ	honu	まぬ	manu	はぬ	hanu
ear	耳	めば	meba	ねば	neba	みば	miba	にば	niba
glasses	メガネ	むよ	muyo	くよ	kuyo	めよ	meyo	けよ	keyo
centipede	ムカデ	みぺ	mipe	しぺ	sipe	むぺ	mupe	すぺ	supe
business card	名刺	もひ	mohi	こひ	kohi	めひ	mehi	けひ	kehi
scarf	マフラー	みぶ	mipu	しぷ	sipu	まぶ	mapu	さぷ	sapu
roll sushi	巻き寿司	めぎ	megi	れぎ	regi	まぎ	magi	らぎ	ragi
eye drops	目薬	むの	muno	るの	runo	めの	meno	れの	reno
squirrel	りす	れぐ	regu	へぐ	hegu	りぐ	rigu	ひぐ	higu
register	レジ	ろみ	romi	ほみ	homi	れみ	remi	へみ	hemi
donkey	ロバ	りだ	rida	みだ	mida	ろだ	roda	もだ	moda
apple	りんご	れぬ	renu	へぬ	henu	りぬ	rinu	ひぬ	hinu
camel	ラクダ	ろぴ	ropi	そぴ	sopi	らぴ	rapi	さぴ	sapi
lemon	レモン	ると	ruto	むと	muto	れと	reto	めと	meto
candle	ロウソク	らね	rane	さね	sane	ろね	rone	そね	sone
rocket	ロケット	りぽ	ripo	にぽ	nipo	ろぽ	ropo	のぽ	nopo
robot	ロボット	れは	reha	めは	meha	ろは	roha	もは	moha

C = first consonant overlap, trans = transliteration into English, C-con = control condition for C, CV = mora overlap, CV-con = control condition for CV.

References

Amaral, C. P., Simões, M. A., & Castelo-Branco, M. S. (2015). Neural signals evoked by stimuli of increasing social scene complexity are detectable at the single-trial level and right lateralized. *PLoS One, 10*(3), e0121970.

Balconi, M., Crivelli, D., & Vanutelli, M. E. (2018). Neurophysiology of intra-and inter-species emotional interactions. Personality trait effect, P300 and N300 ERPs measures. *Neuropsychological Trends,* (23), 97–114.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language, 68*(3), 255–278.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48.

Caramazza, A. (1997). How many levels of processing are there in lexical access? *Cognitive Neuropsychology, 14*(1), 177–208.

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*, 751–781.

Damian, M. F., & Bowers, J. S. (2009). Assessing the role of orthography in speech perception and production: Evidence from picture–word interference tasks. *European Journal of Cognitive Psychology, 21*(4), 581–598.

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

Delorme, A., & Makeig, S. (2004). EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *Journal of Neuroscience Methods, 134*, 9–21.

Forster, K. I., & Davis, C. (1991). The density constraint on form-priming in the naming task: Interference effects from a masked prime. *Journal of Memory and Language, 30*(1), 1–25.

Glaser, W. R., & Düngelhoff, F. J. (1984). The time course of picture-word interference. *Journal of Experimental Psychology: Human Perception and Performance, 10*(5), 640.

Hothorn, T., Bretz, F., & Westfall, P. (2009). *Multcomp: Simultaneous inference for general linear hypotheses (R package version 1.4-8).* Retrieved from http://CRAN.Rproject.org/package=multcomp.

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

Kinoshita, S. (2000). The left-to-right nature of the masked onset priming effect in naming. *Psychonomic Bulletin & Review, 7*, 133–141.

Klem, G. H., Lueders, H. O., Jasper, H. H., & Elger, C. (1999). The ten-twenty electrode system of the international federation. *Electroencephalography and Clinical Neurophysiology, 52*(3), 3–6.

Kubozono, H. (1989). The mora and syllable structure in Japanese: Evidence from speech errors. *Language and Speech, 32*(3), 249–278.

Kureta, Y., Fushimi, T., & Tatsumi, I. F. (2006). The functional unit of phonological encoding: Evidence for moraic representation in native Japanese speakers. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32*, 1102–1119.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software, 82*(13), 1–26.

Lancier, S. (2016). *Spatial memories in place recognition* (Doctoral dissertation, Eberhard Karls Universität Tübingen).

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

Lin, H., & Wang, Q. (2007). Mandarin rhythm: An acoustic study. *Journal of Chinese Language and Computing, 17*(3), 127–140.

Lupker, S. J. (1982). The role of phonetic and orthographic similarity in picture–word interference. *Canadian Journal of Psychology, 36*(3), 349.

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG- and MEG-data. *Journal of Neuroscience Methods, 164*, 177–190.

Meyer, A. S. (1990). The time course of phonological encoding in language production: The encoding of successive syllables of a word. *Journal of Memory and Language, 29*(5), 524–545.

Meyer, A. S., & Schriefers, H. (1991). Phonological facilitation in picture-word interference experiments: Effects of stimulus onset asynchrony and types of interfering stimuli. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17*(6), 1146.

Mullen, T. R., Kothe, C. A. E., Chi, Y. M., Ojeda, A., Kerth, T., Makeig, S., et al. (2015). Real-time neuroimaging and cognitive monitoring using wearable dry EEG. *IEEE Transactions on Biomedical Engineering, 62*, 2553–2567.

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.

Palmer, J. A., Kreutz-Delgado, K., Rao, B. D., & Makeig, S. (2007). Modeling and estimation of dependent subspaces with non-radially symmetric and skewed densities. In M. E. Davies, C. C. James, S. A. Abdallah, & M. D. Plumbley (Eds.). *International conference on independent component analysis and signal separation, lecture notes in computer science book series (LNCS, Vol. 4666)* (pp. 97–104). Berlin; Heidelberg: Springer.

Port, R. F., Dalby, J., & O'Dell, M. (1987). Evidence for mora timing in Japanese. *Journal of the Acoustical Society of America, 81*(5), 1574–1585.

Qu, Q., Damian, M. F., & Kazanina, N. (2012). Sound-sized segments are significant for Mandarin speakers. *Proceedings of the National Academy of Sciences, 109*(35), 14265–14270.
R Core Team. (2018). *R: a language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.
Roelofs, A. (2015). Modeling of phonological encoding in spoken word production: From Germanic languages to Mandarin Chinese and Japanese. *Japanese Psychological Research, 57*, 22–37.
Schiller, N. O. (2008). The masked onset priming effect in picture naming. *Cognition, 106*, 952–962.
Schriefers, H., Meyer, A. S., & Levelt, W. J. (1990). Exploring the time course of lexical access in language production: Picture-word interference studies. *Journal of Memory and Language, 29*(1), 86–102.
Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning & Memory, 6*(2), 174.
Spapé, M., Verdonschot, R., van Dantzig, S., & van Steenbergen, H. (2014). The E-Primer: An introduction to creating psychological experiments. *E-Prime*. Amsterdam, NL: Leiden University Press.
Starreveld, P. A., & La Heij, W. (1996). Time-course analysis of semantic and orthographic context effects in picture naming. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 22*(4), 896.
Usai, F., O'Neil, K. G., & Newman, A. J. (2018). Design and empirical validation of effectiveness of LANGA, an online game-based platform for second language learning. *IEEE Transactions on Learning Technologies, 11*(1), 107–114.
Van Casteren, M., & Davis, M. H. (2006). Mix, a program for pseudorandomization. *Behavior Research Methods, 38*, 584–589.
Verdonschot, R. G., & Kinoshita, S. (2018). Mora or more? The phonological unit of Japanese word production in the Stroop color naming task. *Memory & Cognition, 46*, 410–425.
Verdonschot, R. G., Kiyama, S., Tamaoka, K., Kinoshita, S., La Heij, W. L., & Schiller, N. O. (2011). The functional unit of Japanese word naming: Evidence from masked priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*, 1458–1473.
Verdonschot, R. G., La Heij, W., & Schiller, N. O. (2010). Semantic context effects when naming Japanese kanji, but not Chinese hanzi. *Cognition, 115*, 512–518.
Verdonschot, R. G., Lai, J., Feng, C., Tamaoka, K., & Schiller, N. O. (2015). Constructing initial phonology in Mandarin Chinese: Syllabic or sub-syllabic? A masked priming investigation. *Japanese Psychological Research, 57*, 61–68.
Verdonschot, R. G., Nakayama, M., Zhang, Q.-F., Tamaoka, K., & Schiller, N. O. (2013). The proximate phonological unit of Chinese-English Bilinguals: Proficiency matters. *PLoS One*e61454.
Wang, J., Wong, A. W. K., Wang, S., & Chen, H. C. (2017). Primary phonological planning units in spoken word production are language-specific: Evidence from an ERP study. *Scientific Reports, 7*(1), 5815.
Warner, N., & Arai, T. (2001). Japanese mora-timing: A review. *Phonetica, 58*(1–2), 1–25.
Wong, A. W. K., & Chen, H. C. (2008). Processing segmental and prosodic information in Cantonese word production. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 34*(5), 1172.
Wong, A. W. K., & Chen, H. C. (2009). What are effective phonological units in Cantonese spoken word planning? *Psychonomic Bulletin & Review, 16*(5), 888–892.
Wong, A. W. K., Huang, J., & Chen, H. C. (2012). Phonological units in spoken word production: Insights from Cantonese. *PLoS One, 7*(11), e48776.
Wong, A. W. K., Wang, J., Wong, S. S., & Chen, H. C. (2018). Syllable retrieval precedes sub-syllabic encoding in Cantonese spoken word production. *PLoS One, 13*(11), e0207617.
You, W.-P., Zhang, Q.-F., & Verdonschot, R. G. (2012). Masked syllable priming effects in word and picture naming in Chinese. *PLoS One, 7*(10).
Yu, M., Mo, C., & Mo, L. (2014). The role of phoneme in Mandarin Chinese production: Evidence from ERPs. *PLoS One, 9*(9), e106486.