

Mismatch negativity (MMN) as an index of asymmetric processing of consonant duration in fake Mandarin geminates

Yaxuan Meng^{*}, Sandra Kotzor, Chenzi Xu, Hilary S.Z. Wynne, Aditi Lahiri

Faculty of Linguistics, Philology & Phonetics, University of Oxford, Oxford, United Kingdom

ARTICLE INFO

Keywords:

Duration
Fake geminate
Mandarin
MMN
Lexical tone

ABSTRACT

Unlike languages where consonant duration is used contrastively to distinguish word meanings, long consonants in Mandarin Chinese only occur across morpheme boundaries as a result of concatenation and are referred to as fake geminates. To investigate whether Mandarin speakers employ duration contrast to differentiate fake Mandarin geminates and corresponding singletons as well as the underlying pattern of the processing, two auditory oddball tasks were carried out to measure the component of MMN, an index of the automatic detection of deviant stimulus. Mandarin pseudoword pairs which differ only in the duration of the medial consonant ([an1 an1] ~ [an1 nan1] vs. [an2 an2] ~ [an2 nan2]) were used as stimuli. An asymmetric pattern of brain activation was observed where the singleton deviant in the context of geminate words elicited higher MMNs than in the reversed condition. These findings are in line with earlier research suggesting that the singleton is unspecified for a moraic representation, while the geminate is specified. Mandarin speakers can employ duration contrast to distinguish fake geminates and corresponding singletons; furthermore, the processing of fake concatenated geminates in contrast to singletons is similar to that of real geminates and corresponding singletons.

1. Introduction

Speech varies depending on a number of factors, including speakers, gender, and context. Even for the same speaker in the same context, speech changes according to speaking style or rate. One particular aspect of speech which is affected is the duration of sounds. In languages such as Italian, Japanese or Norwegian, duration is used contrastively to distinguish meaning (e.g. consonant duration in Italian [fat:o] 'fact' vs. [fato] 'fate' or Bengali [kana] 'blind' - [kan:a] 'tears'), while in other languages long consonants occur, for instance, across morpheme boundaries as a result of concatenation (e.g. English *greenness* or German *Zahnnerve* 'dental nerve') but are not used contrastively (Kotzor et al., 2016a). The acoustics of short and long consonants (singletons and geminates), both in languages where an underlying contrast exists and in those where it does not, have been studied extensively (cf. Broselow, 1995; Amano and Hirata, 2010; Ridouane, 2010; Kubozono, 2015; Dmitrieva, 2017; Ridouane and Hallé, 2017) and several proposals for the phonological representation of geminates have been put forward in the last four decades (e.g. Kenstowicz and Pyle, 1973; Hayes, 1986; Schein and Steriade, 1986; Kotzor et al., 2017a); however their

phonological representation remains a controversial issue. Despite the breadth of the existing literature, studies investigating the processing of the geminate-singleton contrast are less common and only a small number of studies have examined the effect of duration on processing and the information this can provide on representation (Roberts et al., 2014; Kotzor et al., 2016b; Ehrenhofer et al., 2017). Recent psycholinguistic and neurolinguistic studies (cf. Roberts et al., 2014; Kotzor et al., 2016a, 2016b & 2017a, 2017b) have argued for a privative representation of duration where a monovalent feature is either present in a segment or absent. This proposal is based on experimental evidence from languages with underlying duration contrasts (e.g. Bengali or Swiss German) and have revealed processing asymmetries which seem to indicate underspecified representation of the short consonant.

This paper focuses on the processing of concatenated geminates which has not yet been investigated. In these cases, identical consonants form a geminate across a morpheme or word boundary, e.g. *clean-ness* or *part-time*. These are also frequently referred to as 'fake' or 'derived' geminates and the relevant theoretical literature suggests that these remain as two separate consonants on one level of representation (Hayes, 1986: 327). Thus, fake geminates can be separated by vowel

^{*} Corresponding author. Faculty of Linguistics, Philology and Phonetics, Clarendon Institute, Walton Street, Oxford, OX1 2HG, University of Oxford, United Kingdom.

E-mail address: yaxuan.meng@ling-phil.ox.ac.uk (Y. Meng).

<https://doi.org/10.1016/j.neuropsychologia.2021.108063>

Received 18 December 2020; Received in revised form 9 September 2021; Accepted 12 October 2021

Available online 13 October 2021

0028-3932/© 2021 The Authors.

Published by Elsevier Ltd.

This is an open access article under the CC BY-NC-ND license

(<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

epenthesis while a true geminate cannot, as illustrated in (1) below:

- (1) **Palestinian Arabic epenthesis** (Hayes, 1986 based on Abu-Salim, 1980) Vowel [i] is inserted between C and CC.

(a) /ʔakl/ > [ʔakil]	'food'
(b) /ʔimm/ > [ʔimim], not *[ʔimim]	'mother' <i>epenthesis blocked for a real geminate</i>
(c) /fut-t/ > [futiit]	'enter-1st.P. SG' <i>epenthesis allowed for fake geminate</i>

Many languages which have both underlying and derived geminates do not show significant acoustic differences between these two types in terms of consonant duration: for instance, in languages such as Bengali and Berber, both exist word medially (e.g. Ridouane, 2010; Lahiri and Hankamer, 1988). The critical measure which distinguishes between medial geminates and singletons is the duration of closure which is typically double that of a singleton for both underlying and derived geminates. In this paper, we examine whether the processing of derived concatenated geminates in contrast to singletons in a non-affixing language (Mandarin) is similar to that of real geminates and corresponding singletons.

In production studies with languages such as English and German (e.g. Kotzor et al., 2016a; Oh and Redford, 2012), the duration of fake or derived geminates has also been shown to be reliably approximately twice as long as that of singletons both in absolute measures and in a relative measure compared to the length of the preceding vowel. In sequences such as *keen-ness* or *Zahn-nerv* 'dental nerve', the medial [n] was produced as a geminate consonant and the duration of closure roughly twice that of the [n] in words such as *banner* or *Zahn-arzt* 'dentist' (e.g. 160 ms vs. 80 ms for English; cf. Kotzor et al., 2016a).

This contrast, however, has not yet been investigated in non-affixing languages without underlying geminates such as Mandarin, which exhibits compounding but also contains largely monosyllabic words (e.g. 电脑 *dian4nao3* 'computer') and has a restricted set of onset and coda consonants. The study of Mandarin is particularly informative because of the extremely limited number of duration contrasts in this language. The questions we ask are, (1) will the duration contrast be used to differentiate fake singleton/geminate nonwords by Mandarin speakers whose native language has no real geminates and (2) what is the underlying pattern in the processing of fake singleton/geminate nonwords? To investigate these questions, we used a mismatch negativity (MMN) paradigm to examine the differences in the automatic processing of long consonants in a concatenated compounding context compared to single consonants.

1.1. Representation and processing of consonant duration

The underlying representation of the phonemic singleton/geminate contrast has been extensively debated and a number of theories concerning its representation have been proposed (Chomsky and Halle, 1968; Hyman, 1985; Kenstowicz and Kisseberth, 1979; Kenstowicz and Pyle, 1973; McCarthy, 1982). For instance, Chomsky and Halle (1968) propose that the contrast is distinguished with the binary feature [\pm LONG]: geminates are represented as [$+$ LONG], while singletons are [$-$ LONG]. Further approaches proposing a difference in the structural representation were also put forward: geminates are either represented by a mora or specified in timing units (two X-slots) while singletons have no weight or only one X-slot (Hyman, 1985; McCarthy, 1982). In order to distinguish between these different proposals, experimental studies have been conducted using both behavioural and neurobiological methods (e.g. Roberts et al., 2014; Kotzor et al., 2016b, 2017a, 2017b).

Kotzor et al. (2016b), examined medial geminates in Bengali in a series of cross-modal lexical decision tasks using both fragment form priming and semantic priming with two different types of mispronunciation: singletons (e.g., [big:æn] 'science' > *[bigæn]) and geminates

(e.g., [kʰoma] 'pardon' > *[kʰom:a]). Their aim was to investigate whether these mispronounced versions would still activate the real word they were based on and lead to priming of a related target. Medial geminates are common in Bengali and most consonants occur as both geminates and singletons (e.g., [kana] 'blind' ~ [kan:a] 'tears'). These medial geminate consonants are usually treated as phonologically heterosyllabic and formed by the coda of one syllable and the onset of the following syllable (Kotzor et al., 2016b, 2017a; Lahiri and Marslen-Wilson, 1992). If the singleton/geminate contrast is distinguished by a binary feature such as [\pm LONG], we would expect that neither singleton nor geminate mispronunciations to lead to facilitation of the target as both are fully specified for length and therefore mismatch in their featural information. A mismatch of features occurs, according to the Featurally Underspecified Lexicon (FUL; Lahiri and Reetz, 2002, 2010), when extracted features contrast with the stored representation where a different feature is fully specified (e.g. if [$-$ SHORT] were to be extracted from the signal and [$+$ LONG] was specified in the representation of a geminate consonant). This leads to a lack of activation. When the featural information extracted from the speech signal matches the stored representation, this results in a simple match and leads to activation of the lexical entry. However, unlike other theoretical approaches, FUL proposes that not all features are necessarily stored and specified fully in the mental lexicon and this results in a third option: no-mismatch. When the extracted features are mapped onto a stored representation with an underspecified feature, they do not match the underlying representation but do not mismatch either, resulting in a no-mismatch. Listeners reject underlying mismatching lexical candidates while they accept matching or no-mismatching ones, despite the lack of a complete match.

If, however, the contrasts are represented on the basis of structural difference or with a privative feature (e.g. [$+$ LONG] for geminates only), an asymmetric pattern would be expected as the mispronounced geminates comprise all the featural information to activate the singletons, while the opposite mispronounced singleton does not. Kotzor et al. (2016b) results consistently showed that real singleton words and mispronounced geminate fragments primed singleton targets equally well, while the real geminates resulted in significantly greater facilitation of the visual targets than the mispronounced singletons, suggesting an asymmetric processing pattern. Therefore, it seems that the lexical representation is accessed only if the consonant duration is sufficient to map onto the weight/length specification of the real word (Kotzor et al., 2016b). In line with the FUL model (Lahiri & Reetz, 2002, 2010), a short consonant duration (singleton) is unspecified for a moraic representation while the long consonant duration (geminate) is specified.

Further neurophysiological studies (e.g. Roberts et al., 2014; Kotzor et al., 2017a) also confirmed this asymmetric processing pattern. Roberts et al. (2014) used a cross-modal semantic priming paradigm with EEG which showed activation patterns in line with the behavioural experiments in Kotzor et al. (2017a, 2017b), with similar N400 attenuation for real words and singletons which were mispronounced as geminates but no N400 reduction for original geminates where the medial vowel was shortened to a singleton. The same asymmetric pattern was also found in pre-attentive processing in a standard oddball paradigm indicated by a difference in the mismatch negativity (MMN) component (Kotzor et al., 2017a). As the MMN represents pre-attentive processing, the asymmetry is thus evident in both early and late stages of accessing the lexical representations from the mental lexicon (Kotzor et al., 2017a, 2017b).

Therefore, the main difference between the singleton/geminate contrast lies on the structural level as the consonants share the same set of features but differ in duration (Ehrenhofer et al., 2017; Kotzor et al., 2017a). A geminate word cannot be accessed until both the structural and featural information is available as the phonetic duration of the singleton/geminate contrast is mapped onto both individual phonemes and syllable structures (Ehrenhofer et al., 2017). Meanwhile, the asymmetric processing of consonant duration suggests the

underspecified representation of singleton/geminate contrast (Lahiri and Reetz, 2010). To examine whether Mandarin speakers use the duration contrast to differentiate singleton and fake-geminate nonwords and investigate the consequences for underlying mechanisms of processing, the current study investigates the phonetic singleton/geminate contrast in Mandarin Chinese using an MMN paradigm.

1.2. Fake geminates in Mandarin

Although consonant duration is used to differentiate word meanings in roughly half of the world's languages (Kotzor et al., 2016b; Roberts et al., 2014), Mandarin is not one of them (Huang and Lin, 2016). Mandarin is a monosyllabic language with a maximum syllable structure of consonant + glide + vowel/syllabic consonant + vowel/nasal and the minimal structure simply being a vowel (Lin, 2007). Usually, Mandarin syllables end with vowels and the only consonant that can occur in the coda position is either [n] or [ŋ]. Meanwhile, only the nasals [m] or [n] can occur in the onset position, resulting in the only consonant which occurs in both coda and onset position being [n] (Duanmu, 2007; Lin, 2007). Compared to languages such as English or German, Mandarin native listeners would only have been exposed to [n:] vs. [n] singleton/geminate contrast in running speech. That is, the potential for fake geminates in Mandarin is very restricted as it only occurs through concatenation of words with identical nasals across syllables in phrases or disyllabic compounds such as 电脑 *dian4nao3* 'computer' (i.e. identical coda and following onset). As a result, Mandarin speakers do not use consonant duration contrastively to differentiate word meanings. Acoustically, these sequences of identical nasals, when compared to single nasals, show a duration contrast similar to that of geminate-singleton duration differences found in languages with underlying geminates (see Fig. 1; cf. Ridouane, 2010). One can also observe the differences in the tonal melody for Tone 1 and Tone 2. The rise for Tone 2 continues through the articulation of the sonorant medial

nasal and for the geminate articulation in [yun2 nan2], the peak comes towards the end of the [n:].

We conducted an MMN study to examine the processing consequences of consonant length in Mandarin speakers. As mentioned above, earlier research has shown asymmetric MMN patterns in languages where singletons and geminates contrast, such as Swiss German and Bengali. However, Mandarin does not use consonant length contrastively to distinguish between two words. Nevertheless, longer closure durations, similar to real geminate consonants are detected when two identical consonants are adjacent (see Fig. 1). Thus, native speakers of Mandarin do come across longer vs. shorter durations in running speech. Would such singleton-geminate differences cause asymmetric effects for native Mandarin speakers as has been found for languages with an underlying contrast? The MMN is a negative deflection which usually peaks at around 100–250 ms after the onset of a stimulus with a distribution at frontocentral sites (Näätänen et al., 2004). This component reflects automatic or pre-attentive detection of changes in the acoustic signal, such as the infrequent stimuli (*deviants*) embedded in a sequence of frequent auditory stimuli (*standards*). The amplitude of the MMN component is correlated with the degree of difference between standard and deviant and can reflect the ability of perceivers to detect the stimuli variance (Näätänen et al., 1992). Previous studies compared the perception of vowels or consonants which vary in duration using the MMN component, as MMN is sensitive to changes of acoustic features such as duration (Kirmse et al., 2008; Leppänen and Lyytinen, 1997).

In the present study, Mandarin pseudoword pairs which only differ in the duration of the medial consonant ([an1 an1] ~ [an1 nan1]) were presented in an oddball paradigm. The numbers in the examples refer to tones. As discussed above, the MMN provides a pre-attentive measure for phonological representation of sound features in the human brain (Näätänen and Alho, 1997). Besides the manipulation of real-word stimuli, pseudowords have also been widely used in MMN studies to investigate the representation of phonetic/acoustic information (Kotzor

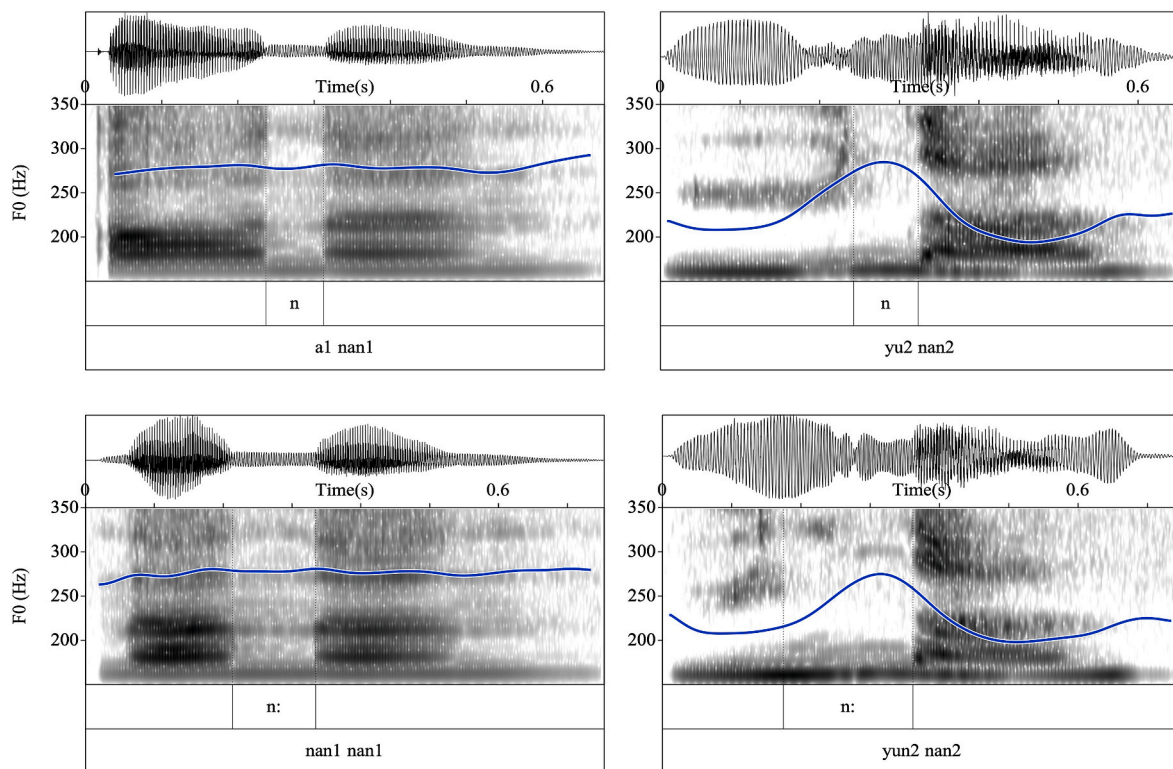


Fig. 1. Oscillograms (above) and spectrograms (below) of Mandarin fake geminates and corresponding singletons with Tone 1 and Tone 2 (duration of medial nasal: 75.78 vs. 120.57 ms; 83.45 vs. 233.34 ms). Left column: 阿囡 *a1nan1* 'nickname for daughter', 囡囡 *nan1nan1* 'nickname for children'; right column: 鱼腩 *yu2nan2* 'fish belly', 云南 *yun2nan2* 'a province in China'.

et al., 2020; Ylinen et al., 2009). In the present study, the duration contrast (or gemination) is a consequence of concatenated phonemes independent of specific lexical representations. Therefore, pseudowords were used to avoid any potential confound of word/morpheme frequency and semantics. To examine the potential influence that lexical tone may have on the processing of geminates, we used two pairs of pseudowords, differing in tones. There are four tones in Mandarin — the high-level tone (Tone 1), high-rising tone (Tone 2), low-dipping tone (Tone 3), and high-falling tone (Tone 4). The primary cue to lexical tone identity is the fundamental frequency (F0) contour that each tone is associated with a particular F0 contour/value (Chao, 1968; Massaro et al., 1985). For our fake geminates, we chose two contours, Tone 1 (level-level) and Tone 2 (rising-rising). Since Tone 1 is a high-level tone, the sequence of two syllables with Tone 1 shows no obvious tonal perturbation but has a continuous level high tone. However, since the Tone 2 syllable has a rising contour, the F0 rise is on the nasal. In a sequence of [an2 an2], the rise ends on the middle of the coda [n] of the first syllable. In the case of a fake geminate situation [an2 an2], where there is a single [n:] release, the contour rises all the way through the nasal and peaks at the end of the consonant (Xu, 1998). The question we asked is whether the differences in the tonal perturbations would have concomitant reflexes in the MMN.

Based on the previous studies discussed above, two competing predictions were made: (1) if Mandarin speakers are purely sensitive to the acoustic difference between singleton and geminate pre-attentively, then a higher MMN amplitude is expected for the stimulus pair /an an/[an nan] (/standard/[deviant]) than for /an nan/[an an] since deviants with longer duration have been found to elicit MMNs with higher amplitudes (Jacobsen and Schröger, 2003; Näätänen et al., 1992); (2) if instead the duration difference between geminates and singletons is processed in the same way as in languages with an underlying duration contrast, then the opposite activation pattern would be expected: the stimulus pair /an nan/[an an] would elicit larger MMNs than /an an/[an nan]. This prediction is made on the basis that the acoustic information in the deviant [an an] mismatches with the phonological representation generated by the standard [an nan] since this requires an additional mora for which the signal does not provide sufficient durational information (mismatch; see Fig. 2). The acoustic signal of the deviant [an nan], however, does not cause the same problem since the phonological representation of singleton [an an] is underspecified (no-mismatch) for duration.

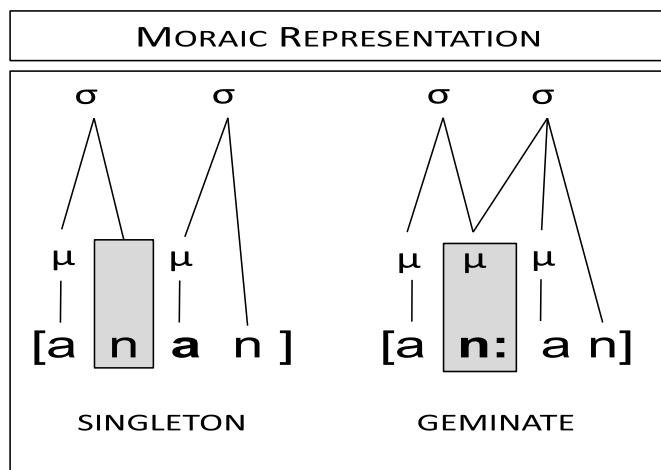


Fig. 2. Moraic representation of Mandarin singleton ([an]) and fake geminate ([an nan]) nonwords with syllable structures.

2. Methods

2.1. Participants

Twenty-one participants (11F/10M, mean age = 23.86 years) recruited at the University of Oxford took part in the study. They were all native Mandarin speakers who lived in China until adulthood and were residing in Oxford (UK) at the time of testing. All participants had normal or corrected-to-normal vision and self-reported as right-handed (a modified version of the Edinburgh Handedness Inventory was also used to confirm handedness; Oldfield, 1971). No history of neurological disorders or hearing deficits was reported. The experiment was approved by the Central University Research Ethics Committee (CUREC) and written informed consent was obtained from all subjects prior to formal experiment.

2.2. Stimuli

To explore geminates resulting from sequences of identical consonants via morpheme concatenation, syllables ending in [n] which precede another syllable that begins with a [n] were considered [n#n], since the alveolar nasal [n] is the only consonant that can occur both in the onset and coda of a syllable in Standard Mandarin. The compound [an#nan] was compared to a corresponding non-geminating condition where the second syllable begins with a vowel [an#an].

This study selected two pairs of disyllabic sequences [an an] and [an nan] with different tones as experimental stimuli (see Fig. 3), all of which are nonwords consisting of phonetically legitimate syllables in Mandarin. In one pair both syllables have high tone (Tone 1) while in the other both have rising tone (Tone 2). Multiple repetitions of the stimuli were recorded by a female native speaker of Mandarin in a sound-attenuated recording room at a sampling rate of 44.1 kHz, and in mono. To prevent coarticulatory effects on the vowel induced by the onset nasal, stimuli were created by means of cross-splicing. In each pair of [an an] and [an nan], the initial vowel [a] and the second syllable [an] in the singleton [an an] were spliced to the geminate [an nan] respectively using Praat (Boersma and Weenink, 2020), replacing the original initial vowel and the rhyme of the second syllable in [an nan]. Segmentation was performed manually on the basis of oscillograms and spectrograms. The spliced stimuli were examined auditorily by native speakers to ensure that no audible discontinuities had resulted from the manipulation. By doing so, the pitch (F0) contours and intensity in most parts except for the medial nasal between the two stimuli were also controlled.

The cross-splicing minimized the acoustic differences between the stimuli and each pair differed only in the duration of the medial nasal [n], reducing the likelihood that other acoustic variables lead to any observed MMN differences. The waveforms and the spectrograms of the stimuli are shown in Fig. 3.

All stimuli were also controlled for length (see Fig. 3): the initial [a] vowels were approximately 150 ms (average: 150.6 ms) and the second [an] sequences approximately 322 ms (average: 322.5 ms). The duration of the medial singletons was, on average, 76.7 ms while the average duration of the medial geminates was 156.5 ms, and thus approximately twice the duration of the singletons. The total duration of each word was approximately 550 ms for [an an] and 630 ms for [an nan]. The maximum F0 for stimuli [an1 an1], [an1 nan1], [an2 an2], and [an2 nan2] are 256.98Hz, 256.96Hz, 224.83Hz, and 243.70Hz, respectively. The intensity of all stimuli was also equalized to 75 dB using Praat.

2.3. 3 Experimental procedure

Two pseudoword pairs were presented to participants during the experiment. Each pseudoword pair was presented in two conditions, once with the singleton word as deviant and the geminate word as standard, and once with the roles reversed (see Table 1). As a result of

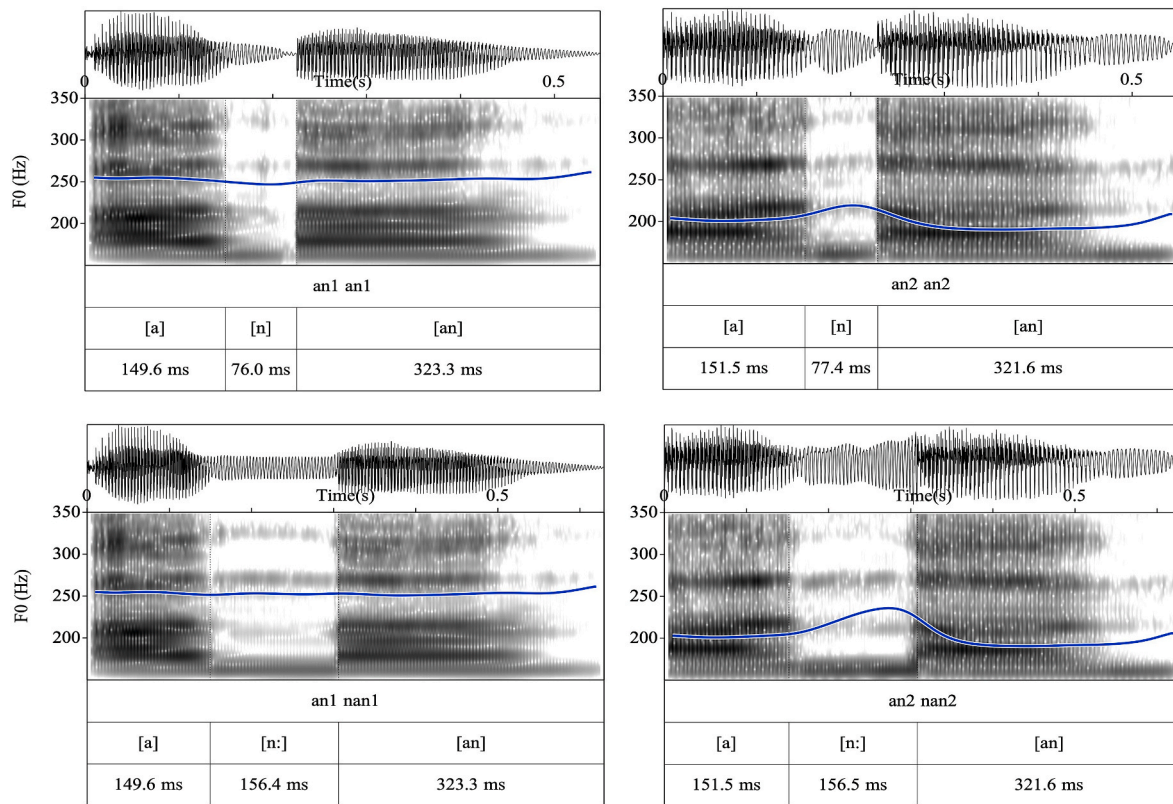


Fig. 3. Sample oscillograms (top half) and spectrograms (bottom half) of the nonword stimuli. The F0 contours indicate the different tonal melody: Tone 1 (level high) and Tone 2 (contour tone).

Table 1
Task design in MMN tasks.

Experiment 1 [an1 an1] ~ [an1 nan1]		Experiment 2 [an2 an2] ~ [an2 nan2]	
Standard	Deviant	Standard	Deviant
[an1 an1]	[an1 nan1]	[an2 an2]	[an2 nan2]
[an1 nan1]	[an1 an1]	[an2 nan2]	[an2 an2]

this reversed design, four oddball blocks were presented to each participant with the order of blocks counterbalanced across participants. Within each block, the deviant occurred pseudo-randomly among standards with a probability of 15% with at least two standard stimuli between deviants. A total of 610 stimuli, beginning with ten consecutive standard stimuli, were presented in each block. To eliminate the influence of rhythmic patterns established by the temporal characteristics of the acoustic stimuli, the inter-stimulus interval (ISI) between standard and deviant varied randomly between 350 ms and 650 ms. During the experiment, participants were seated in a sound-attenuated, shielded EEG booth at a comfortable distance from a screen and watched a nature documentary without sound while the stimuli were presented through headphones (Sennheiser PX200). The volume of the auditory stimuli was kept constant across all participants. Participants were instructed to ignore the auditory stimuli when watching the documentary. The total duration of the whole experiment was approximately 90 min and participants were given the option of short breaks after every two blocks.

2.4. EEG recordings

EEG recordings were made using a Biosemi ActiveTwo amplifier with 64 sintered Ag/AgCl pin electrodes placed in a 10-10 montage and online referenced to the mastoids. EOG activity was measured using four facial electrodes (IO1, IO2, LO1, LO2). All electrode offsets (in an active-

electrode system this is comparable to impedance) were kept below 30 mV and signals were digitised at 2048 Hz.

2.5. Data analysis

EEG data were analysed offline using EEGLAB 14.1.2b. All continuous data was digitally filtered offline in 0.3–30 Hz range using a Finite Impulse Response filter (FIR filter). Bad channels and artifacts were detected and removed automatically with the Artifact Subspace Reconstruction (ASR) method as implemented in the Clean Raw Data plug-in. Data were subsequently re-referenced to the linked mastoids for all analyses except for mastoid amplitudes. By conducting an Independent Components Analysis (ICA, [Delorme and Makeig, 2004](#)), ICA components that may represent eye blinking, lateral eye movement, muscle activity or channel noise were detected and excluded from further analysis. Furthermore, epochs were created from –100 to 800 ms with the time window from –100 to 0 ms used as a baseline. An additional artifact detection was carried out to ensure that trials were rejected if they exceeded an amplitude of 100 μ V. Any participant with a trial-acceptance rate lower than 70% was excluded, which led to the exclusion of five participants from further analysis ([Table 2](#)). Finally, the first ten responses of each block, and the first two standards after each deviant were not included in the grand average. For the difference waves, the deviant-minus-standard calculation was carried out for each participant and condition, and the standard was subtracted from physically identical stimuli (e.g., subtracting standard [an1 an1] from

Table 2
The average number of epochs (standard deviation) for each condition.

	[an1 an1]	[an1 nan1]	[an2 an2]	[an2 nan2]
Standard	272.50 (35.56)	268.69 (34.01)	279.81 (31.19)	268.13 (37.95)
Deviant	73.88 (10.28)	74.13 (8.71)	77.00 (9.19)	73.56 (8.47)

deviant [an1 an1]; Jacobsen and Schröger, 2003; Kirmse et al., 2008).

3. Results

The mean amplitude of the MMN was determined using the ERPLAB Toolbox (Lopez-Calderon and Luck, 2014). Based on visual inspection of the grand-average waveform, the mean peak amplitude of MMN was determined for each participant and condition as the peak amplitude within 200–300 ms after the onset of medial consonant at Fz. For each condition, the peak amplitude at Fz was confirmed using one-tailed T-tests against zero (see Table 3). According to the previous study (Näätänen et al., 1992), the MMN is typically maximal over fronto-central electrode sites. Thus, MMN analyses were restricted to twelve frontocentral electrodes AF3, AFz, AF4, F3, Fz, F4, FC3, FCz, FC4, C3, Cz, C4. For each experiment, repeated measures ANOVA with *Tone* (Tone 1, Tone 2), *Duration* (singleton, geminate), *Laterality* (left, middle, right) and *Gradient* (AF-, F-, FC-, C- line) as within-subject variables were carried out for mean amplitude and peak latency, respectively. For all analyses, degrees of freedom were adjusted according to the method of Greenhouse–Geisser.

3.1. Mismatch negativity (MMN)

Repeated ANOVAs were conducted for the MMN amplitudes and the main effect of *Duration* was significant, $F(1, 15) = 6.50, p = 0.022$. The mean amplitude of singleton words was significantly more negative than the geminate words (refer to Fig. 4). In addition, the main effect of *Gradient* also reached significance, $F(3, 45) = 4.35, p = 0.029$. However, the interaction of *Tone*, *Laterality* and *Gradient* was also significant, $F(6, 90) = 3.34, p = 0.018$. Post-hoc analyses were carried out for the interaction and the results showed that for word pairs with Tone 1, the activation was more negative at AF-, Fz-, and FCz- than at Cz- within left hemisphere, $t_1(13) = -3.37, p_1 = 0.025$; $t_2(13) = -3.62, p_2 = 0.015$; $t_1(13) = -3.90, p_1 = 0.009$. In addition, within middle and right hemispheres, the activations of words with Tone 1 were more negative than words with Tone 2 at anterior-frontal, and frontal electrodes (middle hemisphere: $t_1(15) = -2.32, p_1 = 0.035$; $t_2(15) = -3.09, p_2 = 0.007$; right hemisphere: $t_1(15) = -2.59, p_1 = 0.021$; $t_2(15) = -2.98, p_2 = 0.009$).

4. Discussion

This current study investigated how the consonant duration is exploited by Mandarin speakers to differentiate fake singleton/geminate nonwords, to see whether an asymmetric processing pattern could be found using an oddball paradigm. As the singleton/geminate distinction in Mandarin only occurs in disyllables or phrases, two disyllabic Mandarin pseudowords differing only in the duration of the medial consonant ([an1 an1] ~ [an1 nan1]) were constructed. In addition, to determine the potential influence of lexical tone on the representation of geminates, the same word pair with different lexical tones were also used ([an2 an2] ~ [an2 nan2]). A crossed design with each pseudoword being presented as both standard and deviant was adopted and the MMN difference was obtained by subtracting responses to physically identical stimuli (Jacobsen and Schröger, 2003; Näätänen et al., 1992). Fully in

line with the prediction that the representation of the phonemic difference between singletons and geminates is accessed from the lexicon, the results show a pattern of asymmetric activation, with singleton deviants in the context of geminate stimuli (/an nan/[an an]) eliciting higher MMNs than in the reverse condition (/an an/[an nan]).

These outcomes are similar to those for geminate-singleton pairs in Swiss German and Bengali (Ehrenhofer et al., 2017; Kotzor et al., 2017a, 2017b). The present results indicate that this duration contrast is represented with different structural specifications regardless of whether speakers use the contrast to differentiate words. Therefore, it appears that duration is a gradable property which can be used to identify phoneme categories. Phonetic duration of voice onset time (VOT), for instance, is used to distinguish phonologically voiced and voiceless consonants in English (e.g./d/vs./t/), although essentially the phonetic cue which maps onto an underlying distinction between voiced and voiceless stops is voicing (i.e. [+VOICE] [-VOICE] (Ehrenhofer et al., 2017). Unlike the VOT contrast where the consonant is represented with a certain feature either present or absent, the singleton and geminate consonants share the same set of features and only differ in their duration. What makes consonant duration different is that there is a difference in phonetic duration at the melodic level and, simultaneously, it cues a structural distinction in syllabification. Thus, a geminate word cannot be accessed until both the structural and featural information is available as the phonetic duration is not only mapped to individual phonemes but also determines syllable structures (Ehrenhofer et al., 2017; Kotzor et al., 2017a, 2017b).

In terms of the difference in tones, although the main effect of the lexical tone did not reach statistical significance, there was nevertheless a difference between the stimulus pairs. As mentioned above, there are four lexical tones in Mandarin Chinese, namely a high-level tone (Tone 1), a high-rising tone (Tone 2), a low-dipping tone (Tone 3), and a high-falling tone (Tone 4). Numerous studies have examined the phonetic distinction between these four tones and indicated that both the F0 height and contour are the primary acoustic cues for differentiation (Chao, 1968; Jongman et al., 2006; Massaro et al., 1985). As we mentioned earlier, Tone 1 exhibits a high and relatively flat contour over most of the duration, while Tone 2 shows a rise with the onset occurring in the middle region of the F0 range and approaching the F0 height of Tone 1 at the end. The sequence of high tones forms a plateau and between the syllables there is no pitch movement (Xu and Wang, 2001). However, when there are two contour tones in a sequence, then there is a movement from low-high to another low-high and across syllables there will be a stronger pitch perturbation. Clearly the pitch difference has an effect on the MMN, but it is not a representational issue and thus the difference is not significant. Future work with different tones may show greater asymmetries across some tonal conditions than others. Our goal here was to examine the duration contrast keeping the tones constant. Therefore, the results were in line with our expectation that the asymmetric processing pattern between fake geminate and corresponding singleton was found for stimulus pairs with both tones, as the amplitude of MMN triggered by the word pair with Tone 1 is higher than the pair with Tone 2.

In conclusion, the singleton-geminate contrast offers an opportunity to examine the phonetics–phonology interface in speech processing. The data in our study shows that although there is no underlying duration contrast in Mandarin, Mandarin speakers nevertheless use durational cues to differentiate between singleton and fake-geminate nonwords. Similar to processing patterns found in languages with ‘real’ geminates such as Bengali or Swiss German, the duration contrast is represented asymmetrically in Mandarin with a geminate consonant specified for its length with a mora while a singleton is not. Underspecification exists at both melodic and structural levels and thus the processing of the singleton/geminate contrast requires listeners to mediate between the melodic and structural tiers of representation. The moraic representation of geminates triggers a processing asymmetry where the singleton duration mismatches with the underlying specified geminate, while the

Table 3

Mean amplitude of MMN peak at Fz for all conditions and for the mean amplitude difference from zero (one-tailed).

	Mean peak amplitude (μV)	t
[an1 an1]	−5.56	−5.40**
[an1 nan1]	−2.60	−4.41*
[an2 an2]	−2.96	−4.42**
[an2 nan2]	−1.74	−3.52*

Note. ** $p < 0.001$, * $p < 0.05$.

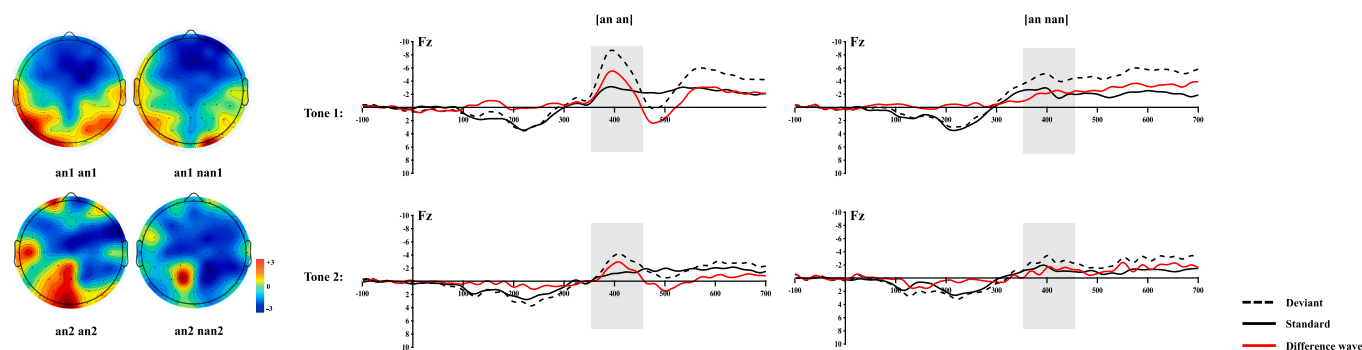


Fig. 4. Deviant-minus-standard difference waves for singleton and geminate conditions within 200–300 ms after the onset of the medial consonant (350–450 ms after the onset of stimuli). Upper row: nonword pair with Tone 1; Lower row: nonword pair with Tone 2. Maps display the topographic distribution of the mean amplitude in the MMN analysis window.

geminate duration does not mismatch with the underlying underspecified singletons, shown by a clear difference in MMN amplitude between the conditions.

Credit author statement

YM: data collection, formal analysis, methodology, investigation, conceptualization, writing–original draft, and writing–review and editing. SK: methodology, conceptualization, and writing–review and editing. CX: data collection and writing–original draft. HW: writing–review and editing; AL: conceptualization, funding acquisition, methodology, project administration, supervision, writing–original draft, and writing–review and editing. All authors contributed to the article and approved the submitted version.

Acknowledgements

This research was funded by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (Grant agreement number: 695481 awarded to Aditi Lahiri).

References

- Abu-Salim, I.M., 1980. Epenthesis and geminate consonants in Palestinian Arabic. *Studies in the Linguistic Sciences Urbana*, 110 (2), 1–11.
- Amano, S., Hirata, Y., 2010. Perception and production boundaries between single and geminate stops in Japanese. *J. Acoust. Soc. Am.* 128 (4), 2049–2058.
- Boersma, P., Weenink, D., 2020. Praat: Doing Phonetics by Computer, version 6.1. 16. Phonetic Sciences, University of Amsterdam, Amsterdam.
- Broselow, Ellen, 1995. Skeletal positions and moras. In: Goldsmith, J.A. (Ed.), *The Handbook of Phonological Theory*. Blackwell, Oxford, pp. 175–205.
- Chao, Y.R., 1968. *A Grammar of Spoken Chinese*. University of California Press, Berkeley, CA.
- Chomsky, N., Halle, M., 1968. *The Sound Pattern of English*. Harper & Row, New York.
- Delorme, A., Makeig, S., 2004. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J. Neurosci. Methods* 134 (1), 9–21.
- Dmitrieva, O., 2017. Production of geminate consonants in Russian: implication for typology. In: Kubozono, H. (Ed.), *The Phonetics and Phonology of Geminate Consonants*. Oxford University Press, Oxford, pp. 34–65.
- Duanmu, S., 2007. *The Phonology of Standard Chinese*. Oxford University Press, Oxford, UK.
- Ehrenhofer, L., Roberts, A.C., Kotzor, S., Wetterlin, A., Lahiri, A., 2017. Asymmetric processing of consonant duration in Swiss German. In: Kubozono, H. (Ed.), *The Phonetics and Phonology of Geminate Consonants*. Oxford University Press, Oxford, pp. 204–229.
- Hayes, B., 1986. Inalterability in CV phonology. *Language* 62 (2), 321–351.
- Huang, H.H., Lin, Y.H., 2016. When unnecessary repairs become necessary: the case of nasal insertion in Standard Mandarin loanwords. In: Albright, A., Fullwood, F.A. (Eds.), *Proceedings of the 2014 Annual Meeting on Phonology*. Linguistic Society of America.
- Hyman, L., 1985. *A Theory of Phonological Weight*. De Gruyter Mouton.
- Jacobsen, T., Schröger, E., 2003. Measuring duration mismatch negativity. *Clin. Neurophysiol.* 114 (6), 1133–1143.
- Jongman, A., Wang, Y., Moore, C.B., Sereno, J.A., 2006. Perception and production of Mandarin Chinese tones. In: Li, P., Tan, L.H., Bates, E., Tzeng, O.J.L. (Eds.), *The Handbook of East Asian Psycholinguistics*, Chinese, vol. 1. Cambridge University Press, Cambridge, pp. 209–217.
- Kenstowicz, M., Kisseberth, C., 1979. *Generative Phonology: Description and Theory*. Academic Press, New York.
- Kenstowicz, M., Pyle, C., 1973. On the phonological integrity of geminate clusters. In: Kenstowicz, Michael, Kisseberth, Charles W. (Eds.), *Issues in Phonological Theory: Proceedings of the Urbana Conference on Phonology*, vol. 973. The Hague & Paris: Mouton, pp. 27–43.
- Kirmse, U., Ylinen, S., Tervaniemi, M., Vainio, M., Schröger, E., Jacobsen, T., 2008. Modulation of the mismatch negativity (MMN) to vowel duration changes in native speakers of Finnish and German as a result of language experience. *Int. J. Psychophysiol.* 67 (2), 131–143.
- Kotzor, S., Molineaux, B.J., Banks, E., Lahiri, A., 2016a. “Fake” gemination in suffixed words and compounds in English and German. *J. Acoust. Soc. Am.* 140 (1), 356.
- Kotzor, S., Wetterlin, A., Roberts, A.C., Lahiri, A., 2016b. Processing of phonemic consonant length: semantic and fragment priming evidence from Bengali. *Lang. Speech* 59 (1), 83–112.
- Kotzor, S., Wetterlin, A., Lahiri, A., 2017a. Bengali geminates: processing and representation. In: Kubozono, H. (Ed.), *The Phonetics and Phonology of Geminate Consonants*. Oxford University Press, Oxford, pp. 187–203.
- Kotzor, S., Wetterlin, A., Lahiri, A., 2017b. Symmetry or asymmetry: evidence for underspecification in the mental lexicon. In: Lahiri, A., Kotzor, S. (Eds.), *The Speech Processing Lexicon: Neurocognitive and Behavioural Approaches*. De Gruyter Mouton, Berlin, pp. 85–106.
- Kotzor, S., Zhou, B., Lahiri, A., 2020. (A) Symmetry in vowel features in verbs and pseudoverbs: ERP evidence. *Neuropsychologia* 143, 107474.
- Kubozono, H. (Ed.), 2015. *Handbook of Japanese Phonetics and Phonology*, vol. 2. De Gruyter Mouton, Boston.
- Lahiri, A., Hankamer, J., 1988. The timing of geminate consonants. *J. Phonetics* 16 (3), 327–338.
- Lahiri, A., Marslen-Wilson, W.D., 1992. Lexical processing and phonological representation. In: Docherty, G.J., Ladd, D.R. (Eds.), *Papers in Laboratory Phonology II*. Cambridge University Press, Cambridge, pp. 229–254.
- Lahiri, A., Reetz, H., 2002. Underspecified Recognition, vol. 7. *Laboratory Phonology*, pp. 637–675.
- Lahiri, A., Reetz, H., 2010. Distinctive features: phonological underspecification in representation and processing. *J. Phonetics* 38 (1), 44–59.
- Leppänen, R., P.H.T., Lyytinen, H., 1997. Auditory event-related potentials in the study of developmental language-related disorders. *Audiology and Neurotology* 2 (5), 308–340.
- Lin, Y.H., 2007. *The Sounds of Chinese*, vol. 1. Cambridge University Press.
- Lopez-Calderon, J., Luck, S.J., 2014. ERPLAB: an open-source toolbox for the analysis of event-related potentials. *Front. Hum. Neurosci.* 8, 213.
- Massaro, D.W., Cohen, M.M., Tseng, C.Y., 1985. The evaluation and integration of pitch height and pitch contour in lexical tone perception in Mandarin Chinese. *J. Chin. Ling.* 13, 267–289.
- McCarthy, J., 1982. *Formal Problems in Semitic Phonology and Morphology*. Routledge.
- Näätänen, R., Alho, K., 1997. Mismatch negativity—the measure for central sound representation accuracy. *Audiology and Neurotology* 2 (5), 341–353.
- Näätänen, R., Pakarinen, S., Rinne, T., Takegata, R., 2004. The mismatch negativity (MMN): towards the optimal paradigm. *Clin. Neurophysiol.* 115 (1), 140–144.
- Näätänen, R., Teder, W., Alho, K., Lavikainen, J., 1992. Auditory attention and selective input modulation: a topographical ERP study. *Neuroreport* 3 (6), 493–496.
- Oh, G.E., Redford, M.A., 2012. The production and phonetic representation of fake geminates in English. *J. Phonetics* 40, 82–91.
- Oldfield, R.C., 1971. The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia* 9 (1), 97–113.
- Ridouane, R., 2010. Geminates in the junction of phonetics and phonology. In: Vallee, Nathalie (Ed.), Fougere, Cécile, Barbara Kühnert, Mariapaola D’Imperio, *Laboratory Phonology*. De Gruyter Mouton, Berlin, pp. 61–90.
- Ridouane, R., Hallé, P.A., 2017. Word-initial geminates: from production to perception. In: Kubozono, H. (Ed.), *The Phonetics and Phonology of Geminate Consonants*. Oxford University Press, Oxford, pp. 66–84.

- Roberts, A.C., Kotzor, S., Wetterlin, A., Lahiri, A., 2014. Asymmetric processing of durational differences-electrophysiological investigations in Bengali. *Neuropsychologia* 58, 88–98.
- Schein, B., Steriade, D., 1986. On geminates. *Ling. Inq.* 17 (4), 691–744.
- Xu, Y., 1998. Consistency of tone-syllable alignment across different syllable structures and speaking rates. *Phonetica* 55 (4), 179–203.
- Xu, Y., Wang, Q.E., 2001. Pitch targets and their realization: evidence from Mandarin Chinese. *Speech Commun.* 33 (4), 319–337.
- Ylinen, S., Strelnikov, K., Huotilainen, M., Näätänen, R., 2009. Effects of prosodic familiarity on the automatic processing of words in the human brain. *Int. J. Psychophysiol.* 73 (3), 362–368.