1st International Conference on Tone and Intonation (TAI)
6-9 December 2021, Sonderborg, Denmark

Priming Effects of Tones in Visual Processing of Vietnamese

Rolando Coto-Solano[1], *Diệp Trần*[2]

[1]Dartmouth College, United States
[2]Victoria University of Wellington Te Herenga Waka, Aotearoa New Zealand
rolando.a.coto.solano@dartmouth.edu, diepdhhn@gmail.com

Abstract

This paper presents an experiment to detect the presence of visual tonal priming in Vietnamese. We carried out a lexical decision task with masked priming with 23 Native speakers of Vietnamese. We compared the reaction times of (i) identical primes and targets, which share consonants, vowels and tones, (ii) primes/targets with different tones, which share only the consonants and vowels, and (iii) completely unrelated primes/targets, where no elements are shared. Prime-target pairs with shared consonants and vowels but different tones had significantly faster RTs than unrelated prime-target pairs (645 vs. 674 ms, p<0.05). This means that consonants and vowels produce a facilitation effect. We also found that prime-target pairs with shared consonants and vowels but different tone have significantly slower RTs than identical prime/targets (645 vs. 622 ms, p<0.05). This is evidence that tones provide facilitation separate from that of segments. Several hypotheses can explain this pattern: This could be an effect of visual similarity of tonal diacritics in the prime/targets, or it could be related to the Vietnamese writing system. The explicit and stable tonal marking in Vietnamese writing, which is unavailable in Mandarin, Cantonese, Thai – languages for which tonal priming is elusive – might result in greater tonal metacognition and stronger priming effects.

Index Terms: Tonal priming, psycholinguistics, Vietnamese, tone, phonology, visual lexical decision task, masked priming

1. Introduction

Vietnamese is a tonal language, where a combination of pitch and laryngeal features is used to distinguish meaning between words that share segmental phonemes. For example, the words *ngay* /ŋaj˧/ 'immediately' and ngày /ŋaj˨˩/ 'day' have the same consonants and vowels, but the first word is produced with a mid-level pitch, whereas the second one is pronounced with a mid-low contour. Standard Vietnamese (which corresponds to the northern dialect of Hanoi) has six tones in total [1, 2], which are marked using diacritics in the standard orthography. Table 1 shows a summary of the tones in the Standard (Northern) Vietnamese, their phonetic trajectories and the diacritic used to represent each of them.

Pitch has other functions in spoken languages, including the production of intonation. Intonation is the use of pitch to mark expressive meaning (e.g. surprise) and to distinguish between types of questions and statements (e.g. focus). Most phonological theories make a distinction between tones, which are treated as a part of the lexical representation of words, and intonation, which is treated as a post-lexical phenomenon which is not inherently associated to words, but rather to multiword and phrase-level groupings. Moreover, tones are usually treated as phonemes: They are described using feature systems [3], they can participate in morpho-phonological rules [4], and their pro-

Table 1: *Examples of Vietnamese Tones (Northern Variety)*

Tone	Trajectory	Example
ngang 'flat'	mid level ˧ (33)	*ba* 'three'
huyền 'deep'	low falling, breathy ˨˩ (21, 31)	*bà* 'lady'
sắc 'sharp'	mid rise ˧˥ (35)	*bá* 'governor'
nặng 'heavy'	glottalized mid falling, short: ˧ˀ˨ʔ (3ˀ2ʔ)	*bạ* 'at random'
hỏi 'asking'	low dipping ˧˩˧ (313)	*bả* 'poison'
ngã 'tumbling'	glottalized mid rise ˧ˀ˥ (3ˀ5)	*bã* 'residue'

cessing entails categorical perception [5, 6].

Despite these similarities, there is a large body of psycholinguistic research which suggests that, under certain conditions, tones behave like intonation and not like segments. One such condition is priming. Priming is a psychological phenomenon where exposure to a stimulus influences the response to subsequent stimuli [7]. In *identity priming*, subjects see or hear the same word twice. When they see the word the first time, lexical access processes activate the word. When they are exposed to the word for the second time, the word is already active and the subjects' reaction to the word will be faster because it has already been retrieved. This acceleration effect is called *facilitation*. Facilitation takes place even if the first word, the *prime*, is different from the second word, or *target*, provided there is some level of orthographic and/or phonological overlap. For example, the prime *cat* activates other semantically-unrelated words such as *cold* and leaves them active in memory. This would make a response to the target *cold* slightly faster than it would otherwise be. Facilitation for shared consonants between primes and targets is well documented [8, 9, 10]. Facilitation has also been observed when two words share the same vowels: Primes like *cat* will also activate targets like *bat* because of their shared 'a' vowel [11], even if the activation effect from shared vowels has been documented to be weaker than that of shared consonants [12].

Given that vowels and consonants provide facilitation in priming, it is plausible that tones would pattern with segments and also facilitate processing. However, there is divergent evidence on whether tone actually does provide facilitation in tonal languages. Visual masked priming experiments in Thai suggest that segmental information "appears to be more important than tone information" [13]. Stroop task experiments also suggest that, if tonal effects do exist, they come "into play at a later stage in lexical processing" [14]. Some authors find facilitation in Mandarin using visual priming lexical decision tasks

10.21437/TAI.2021-17

[15], but most experiments report that tone might be processed "much like the metrical properties of languages such as Dutch or English" ([16], Stroop task), and that it "most likely functions like stress and constitutes part of the metrical frame in Mandarin" ([17], visual priming lexical decision). This lack of facilitation has also been reported during first-language acquisition of Mandarin, where children's performance in recognition of both familiar and novel words suggests that "despite the importance of tones in tone languages, vowels maintain primacy over tones in young children's word recognition [and] findings suggest that early lexical processes are more tightly constrained by variation in vowels than by tones" [18]. Cantonese experiments have also shown contradictory effects: auditory same-different tasks performed by Cutler and Chen [19] did find an effect of tone, whereas Yip's auditory priming results [20] are "not consistent with [the] notion that tone is part of the lexical 'access code' and that tonal contrasts are processed at the early activation phase of word recognition" [21]. In summary, experiments in both visual and auditory modalities, carried out in Thai, Mandarin and Cantonese suggest that tone might not provide facilitation during priming because it is processed later than segments, in a manner similar to prosodic pitch.

Evidence from neuroimaging also exhibits ambivalence in the effects of tone on priming. Some methodologies like electroencephalography (EEG) have shown visual tonal priming in Mandarin [22] and auditory priming in Cantonese [23]. On the other hand, both visual [24] and auditory functional magnetic resonance imaging (fMRI) results for Mandarin suggest that "the cortical circuitry subserving lexical tones differs from that of consonants or rhymes" [25].

There is another psycholinguistic phenomenon where tone behaves differently from segments: phonological awareness. Phonological awareness is the ability to identify different components of a language, such as phonemes [26]. More generally, it entails an explicit knowledge of the phonological units of a language [27]. Phonological awareness is gained through explicit learning, and children gain such awareness as they learn to read [28, 29]. As they progress in reading, children become more capable in tasks such as phoneme segmentation (understanding which phonemes are present in a word), as well as phoneme matching (being able to tell that two words share the same phonemes).

Phonological awareness is strongly correlated with the units that are marked in the writing system; this is referred to as the *psycholinguistic grain size theory* [30]. Children that read with alphabetical systems develop segmental phoneme awareness [31]. On the other hand, children who are speakers of Mandarin and Cantonese gain syllable awareness first because this is the most immediately visible unit in their writing systems. After this, they develop onset-rhyme awareness, in part because this is a usual way in which reading is taught in the Chinese schooling system. Training in writing systems with explicit tonal marking, for example Chinese Pinyin, is significantly correlated with tonal awareness in early readers [32, 33, 34], but this effect appears to dissipate quickly, as both older children and adults show little tonal metacognition [35, 36, 37]. This could be because Pinyin with consistently marked tones is only used early in schooling, and most Pinyin that adults encounter is not marked for tone (e.g. in road signs). Thai, on the other hand, does mark tone, but the pattern is not one-grapheme-to-one-tone. In the Thai writing system some tones are marked with diacritics, but most tones have no explicit marking, and are intuited from the interaction of consonant and vowel graphemes (e.g. a long vowel with a high-class consonant is produced with a rising tone). In this system, where tone is not explicitly marked, tonal awareness is also relatively low [38]. These results, in which the transparency of tonal marking shows correlation with the processing of the tone, has parallels in the processing of consonants and vowels in other languages. There is evidence that, in languages where the grapheme-phoneme correspondence is more transparent (e.g. Italian), the vowels are processed as fast as consonants; on the other hand, in languages where vowels are written with a less transparent grapheme-phoneme correspondence (e.g. English), vowels are processed slower and later than consonants [39].

Difficulties with tonal metacognition can also be observed in the writing of Indigenous languages of the Americas and Africa. Many of these languages use Roman alphabetical systems, but most people who speak these languages learned to read and write in colonial languages like English or Spanish, and only later learned to read and write their own tonal language. As a result, children who eventually become adult readers are not exposed to consistent tonal marking in their early schooling [40, 41]. Mirroring the results in Mandarin, Cantonese and Thai, these adults also report difficulties when performing tasks that involve tonal awareness, even when their segmental phonemic awareness is high [42].

From the priming results (particularly the visual priming research), as well as from metacognition research, it is plausible that speakers of tonal languages will not exhibit visual tonal priming if the writing system of their language does not mark tone explicitly and separately from the segmental phonemes. It is conceivable that if their language does mark tone explicitly (e.g. through diacritics) and separately from the glyphs for consonants and vowels (again, through diacritics), then speakers could benefit from these stable representations and report visual tonal priming. Most languages that use such writing systems are either small or endangered, like Indigenous languages in the Americas and Africa, and learning to read and write in them is an inconsistent process for the people who speak them. However, there is one widely spoken tonal language with explicit and stable tonal marking, and which is the first language in which many children learn to read and write. This language is Vietnamese.

The current Vietnamese writing system, Quốc Ngữ, provides consistent tone marking independent from the graphemes for segments, which makes it different from Mandarin, Cantonese and Thai. The system is learned relatively quickly by children, and the literacy rate for 15-year olds and above is extremely high (96% [43]), which makes it different from Indigenous languages in the Americas and Africa, where very few people are trained to read and write these languages. Because of these characteristics, it was hypothesized that tonal priming effects would be visible in Vietnamese.

There has been some psycholinguistic work on Vietnamese. Pham and Baayen [44] performed visual lexical decision tasks to measure the reaction time to compound words, and found significant random-effects to the tones in the compounds, meaning that different tones can lead to different reaction times. Liu [45] found a form of identity priming: speakers were asked to memorize word pairs and then shown the first word and asked to recall the second word and say it out loud; this process was faster in a condition where all of the second words shared the same first syllable. This advantage is larger in Vietnamese than in Mandarin, and from this Liu [45] argues that this advantage is due to differences in orthography. There are other experiments that also hint at the importance of tone. Brunelle and Jannedy [46] provide evidence that the dialect of the speaker (North versus South Vietnam) influences their perception of tones, and Pham et al. [47, 48] provide evidence that, when tone is considered in the construction of non-word stimuli, this improves the non-words so that they can help in the diagnostic of developmental language disorder.

Given the results reported here, we have two hypotheses: (i) There will be tonal visual priming in Vietnamese, and (ii) this might be related to the writing system in Vietnamese. The next section describes a visual masked lexical decision task testing the first hypothesis, and section 3 will describe the experiment's results, which confirm this hypothesis. Section 4 will discuss the second hypothesis, and how it could be tested against other explanations for tonal priming.

2. Methodology

In order to investigate tonal priming in Vietnamese we carried out a lexical decision task [49] with visual masked priming. In this task, subjects sat in front of a computer screen and looked at sequences of characters. Some of these sequences were real words of Vietnamese, and some of them were possible but unattested nonwords, such as *thỉ* or *po*. The subjects were asked to look at the screen and then press a button to indicate whether the sequence they saw was a real word or not. Internally the experiment showed a sequence of three elements: First, subjects would see a mask with a sequence of hash signs `###` for 500 ms. Then they would see a lowercase prime for 50 ms, which is below the conscious perception threshold for most individuals. After this, they would see an uppercase target for up to 2 seconds. It was to this target that subjects had to respond with whether it was a real word or not. The prime and target would either both be real words (e.g. `kể` 'tell' / `VÌ` 'because'), or they would both be nonwords (e.g. `thỉ` / `NGÚ`).

The experiment had 5 conditions. First, the pairs of primes and targets could be *unrelated*, so that they didn't share consonants, vowels or tone. Table 2 shows both real and nonword pairs where there are no shared elements between prime and target. The second condition is *identity*, where all elements are shared. The other three conditions entail changing one element from the {consonants, vowel, tone} combination. In the condition of *different consonant*, the vowels and tones are the same, but the consonants are different (e.g. `dò` 'move slowly' / `BÒ` 'cow'). In the *different vowel* condition, the consonants and tones are the same, but the vowel is different (e.g. `ngã` 'fall' / `NGŨ` 'five'). Most importantly, in the *different tone* condition, the primes and targets share the same segments, but the tones are different (e.g. `mái` 'roof' / `MÃI` 'forever').

Each condition had a total of 30 items, for a total of 300 items (5 conditions x 30 items x real/nonword). For the *different tone* condition, we included 2 items for each of the fifteen unordered combinations of Vietnamese tones (e.g. *ngang* 'flat' prime vs. *huyền* 'deep' target; *ngang* 'flat' prime vs. *sắc* 'sharp' target, etc.)[1]. For real words, we used the 52.8 million word VNESE Corpus [50] to calculate word frequency and then matched primes and targets for frequency within the same order of magnitude (calculated in words per million). The nonwords were generated using the Wuggy software [51] to ensure that they complied with the phonotactics of Vietnamese; they were also manually verified by the researchers.

The experiment was carried out using the DMDX psycholinguistic program [52] running on an HP 15-AY011NR laptop computer with an Intel i5-6200U, 2.3GHz processor. Trials were separated into 3 blocks and presented at random for each subject. A total of 23 subjects were included in the experiment. These were Native speakers of Vietnamese who were studying at Victoria University of Wellington Te Herenga Waka in Wellington, New Zealand. They were 5 men and 18 women, with an average age of 31±10 years old. We collected additional demographic information about their language histories, most importantly, what region they were from, and whether they spoke the standard (Northern) dialect of Vietnamese, or any of the Central or Southern dialects of the language. In total, 10 participants spoke the northern Hanoi dialect while 13 spoke other dialects.

Once the data was collected, some exclusions were necessary. Only correct answers were kept in the final dataset, and only real words were included in the final statistical model. Any items and subjects with accuracies less than 80% were excluded. This resulted in 4 items and 2 subjects being excluded, reducing the number of items to 146 and of subjects from the original 25 to 23. Responses with unusually short reaction times (less than 2.5 standard deviations) were excluded, given their high probability of being errors. Likewise, unusually long answers (more then 2.5 standard deviations) were excluded. After this filtering, there were a total of 3095 datapoints in the dataset.

The analysis of the data was performed using a Linear Mixed Effects Model, implemented in the `lme4` package [53] of R [54]. The model had the natural logarithm of the reaction time as the dependent variable. This transformation was necessary to ensure that the model meets the assumption of normality, and it is standard in psycholinguistic research. The model included six independent variables: (i) Experimental *condition* (levels: Unrelated, identity, different consonant, different vowel, different tone), (ii) the *order of presentation* of the items for each subject, (iii) the natural logarithm of the *frequency* of the target item, (iv) the *length of the target* in number of letters (e.g. `mái` = 3), (v) the *number of diacritics* in the target word (e.g. `mái` = 1), and (vi) whether the subject *speaks Standard Vietnamese* or a different dialect of the language. The model also included random intercepts for *item* and *subject*.

[1]Future versions of the experiment need to account for the ordering of the tone pairs. The present experiment included pairs of *ngang* 'flat' primes and *sắc* 'sharp' targets, for example, but not pairs of *sắc* 'sharp' primes and *ngang* 'flat' targets.

Table 2: *Examples for experimental conditions*

		Real words				Nonwords	
Condition	Prime and target share	Prime		Target		Prime	Target
Unrelated	-	kể	'tell'	VÌ	'because'	thỉ	NGÚ
Identity	CVT	dư	'surplus'	DƯ	'surplus'	nũ	NŨ
Different Consonant	VT	dò	'move slowly'	BÒ	'cow'	thẹ	CHẸ
Different Vowel	CT	ngã	'fall'	NGŨ	'five'	po	pi
Different tone	CV	mái	'roof'	MÃI	'forever'	nhia	NHÌA

Table 3: *Results of Linear Mixed-Effects Model*

Variable	Estimate	Significance	
Unrelated/Identity	β = -0.09 ± 0.02	t(135) = -4.6, p<0.00001	***
Unrelated/DiffVowel	β = -0.05 ± 0.02	t(136) = -3.0, p<0.005	**
Unrelated/DiffTone	β = -0.04 ± 0.02	t(135) = -2.2, p<0.05	*
Unrelated/DiffCons	β = -0.04 ± 0.02	t(136) = -2.0, p<0.05	*
Order of presentation	$\beta = -3.4\text{x}10^{-4} \pm 3.9\text{x}10^{-5}$	t(2997) = -8.8, p<0.00001	***
log(Frequency of target)	$\beta = -1.9\text{x}10^{-2} \pm 6.1\text{x}10^{-3}$	t(136) = -3.1, p<0.005	**
Target length	$\beta = -1.3\text{x}10^{-4} \pm 6.4\text{x}10^{-3}$	t(136) = -0.1, p=0.98	
Num. Diacritics in target	$\beta = 5.7\text{x}10^{-3} \pm 8.4\text{x}10^{-3}$	t(136) = 0.7, p=0.50	
Speaks standard Vietnamese	β = 0.04 ± 0.04	t(21) = 1.0, p=0.34	

3. Results

Table 3 shows the estimates and the significance relationships in the LMER model. There were two variables related to the experiment that were significant. First, there is an effect of learning. As subjects make progress through the experiment, their responses become faster ($\beta = -3.4\text{x}10^{-4} \pm 3.9\text{x}10^{-5}$, t(2997) = -8.8, p<0.00001). Second, target words with higher frequencies had faster reaction times ($\beta = -1.9\text{x}10^{-2} \pm 6.1\text{x}10^{-3}$, t(136) = -3.1, p<0.005). No significant effects were found for the length of the target (p=0.98), the number of diacritics in the target (p=0.50) or whether the subject spoke Standard or other dialects of Vietnamese (p=0.34).

Next we will focus on the differences in the experimental conditions. Table 4 shows the means and standard deviation for the reaction time of each condition. The *identity* condition has the shortest reaction time (622 ± 145 ms), and is consistent with other visual masked priming literature. This shorter reaction time of the *identity* condition was significantly faster than the *unrelated* condition (t(135)=-4.6, p<0.00001), which in itself had the longest reaction time (674 ± 148 ms). This confirms that visual identity priming is present in the data. The other conditions were also significantly faster than *unrelated*, and therefore provided some facilitation priming. The *different vowel* condition had an average reaction time of 638 ms (t(136)=-3.0, p<0.005), and the *different consonant* condition had an average reaction time of 646 ms (t(136)=-2.0, p<0.05). Most notably, *different tone* was significantly faster than *unrelated* (645 ± 152 ms; t(135)=-2.2, p<0.05). This means that segments (consonants and vowels) are providing a facilitation effect.

Table 4: *Means for different conditions*

Condition	Prime and target share	RT mean and SD (ms)	Difference with Unrelated
Identity	CVT	622 ± 145	t(135)=-4.6, p<0.00001
DiffVowel	CT	638 ± 143	t(136)=-3.0, p<0.005
DiffTone	CV	645 ± 152	t(135)=-2.2, p<0.05
DiffCons	VT	646 ± 144	t(136)=-2.0, p<0.05
Unrelated	-	674 ± 148	-

In addition to comparing levels of condition with the reference level *unrelated*, we probe the relationships between the different conditions to find any significant differences between them. A Bonferroni-corrected Post-Hoc test was conducted on the linear mixed effects model to examine the difference between *different tone* and the other conditions. The *different tone* condition was significantly slower than *identity* (645 versus 622 ms; p<0.05). *Different tone* remained significantly faster than *unrelated* (645 versus 674 ms; p<0.005), but it was not significantly different from *different consonant* (p=1.0) or *different vowel* (p=1.0).

The results so far indicate that $\textit{identity}_{622\text{ms}}$ is significantly faster than $\textit{DifferentTone}_{645\text{ms}}$, which is in itself significantly faster than $\textit{unrelated}_{674\text{ms}}$. This means that, when we look at reaction time, *identity* < *DifferentTone* < *unrelated*. First, the difference between the unrelated condition, where there are no segments shared between prime and target, and the *different tone* condition, where the segments are shared but the tone is different (e.g. `mái`/`MÃI`), results in a significant facilitatory effect. Second, going from *different tone*, where only the CV component is shared, to the *identity* condition, where CVT are shared, offers additional facilitation for reaction time. This implies that the tone component of the word is providing visual priming. Therefore, this constitutes evidence of tonal priming in Vietnamese.

4. Discussion

The results above provide evidence that tonal priming exists when processing written Vietnamese, and therefore provides confirmation for the first hypothesis. This implies that tones are not always secondary to lexical access and that they can be processed early like segmental phonemes. The main question we will tackle in the discussion is related to the second hypothesis: Why is there tonal facilitation in Vietnamese priming, when it is so difficult to find it in Mandarin, Thai or Cantonese? Does this have to do with the writing system of Vietnamese?

One explanation for the presence of tonal priming is that this could be a mere visual/orthographic effect [55], related to the shapes of the graphemes in Quốc Ngữ. The tonal diacritics don't have a lower/uppercase difference. For example, the acute accent diacritic for the *sắc* 'sharp' tone (contour 35) has the same graphic shape in the lower case (e.g. `bá` 'embrace') and in the upper case (e.g. `BÁ`). Because there isn't a lower/uppercase difference, the subjects are seeing an identical tonal grapheme twice, once in the prime and once in the target. This visual pattern is not possible in Chinese characters, where there are no constant visual elements between two characters with the same tone. Pinyin could potentially produce this effect of visual repetition of diacritics, but since Pinyin is frequently written without tonal diacritics, the visual input would be inconsistent. As for Thai, its writing system does have two diacritics that have consistent tonal values (*mai chattawa* 'rising' and *mai tri* 'high'), but most tones have to be extrapolated from the combination of consonant onset class, vowel length and consonant coda type. Because Thai and Mandarin cannot produce an effect of visual similarity and Vietnamese can, this could be an explanation for the results of the experiment. Therefore, the observed effect might be less linguistic and more related to the visual modality. Moreover, this could also explain the difference between Vietnamese and Indigenous languages. Languages with Roman-based writing systems do have diacritic shapes that remain constant (e.g. the acute diacritic marks the high tone in the word *ù* and *Ù* 'house' in Bribri, a language from Costa Rica). However, the use of the diacritics is not consistent in everyday writing, so even if the shapes remain the same, they wouldn't allow for stable matching between the tonal grapheme and the tone. The constant and stable input that Quốc Ngữ provides to Vietnamese speakers would allow for stronger visual matching.

A second possible explanation would align with the hypothesis described above: That the facilitation effect is due to increased tonal awareness, mainly driven by the explicit marking of tone in the writing system of Vietnamese. Tonal priming could have a phonological component, similar to how rhymes, consonants and vowels produce priming. Therefore, constant exposure to consistent tonal diacritics might strengthen pathways for speakers to retrieve words through tone, creating phonological priming as it creates tonal metacognition. This would be in line with the *psycholinguistic grain size theory* [30], and would explain the differences observed between Vietnamese and other languages where tonal priming is difficult to find. Mandarin and Thai wouldn't show priming because the lack of stable diacritics wouldn't enable speakers to gain enough tonal awareness. In the case of Indigenous and African languages, the instabilities of newly created orthographies and the fact that people learn to read and write in non-tonal languages first would again mean that the input available to readers is not stable enough to create tonal awareness or tonal priming.

One way to test these two hypothesis (visual effect versus tonal awareness as the driver of facilitation) would be to repeat the experiment in the auditory modality, to verify that the priming effects extend there as well. A separate, second test would be necessary to examine the sources of potential priming effects, and to address the question: Is tonal priming an effect of continued education in Vietnamese? The experiment should be adapted for different age groups, particularly children. This could help us look for correlations between tonal facilitation and longer exposure to the Vietnamese writing system, as well as greater proficiency in reading and writing the language. Such an experiment could clarify the role of the Vietnamese writing system in aiding tonal priming.

The experiment described in Section 2 also needs refinement to confirm its results. An important modification would entail changing the relationship between prime and target, so that there is only one element in common between them. In the experiment carried out here, we are studying priming as the subtraction of the effect of *identity* minus the effect of *different tone*. Doing it this way, we can study whether tones provide any facilitation in addition to the facilitation already provided by the consonants and vowels. However, in a future experiment, the prime and target pairs could be made so that they share their tone but have different consonants and vowels. For example, `thơi` 'poetry' and `chi` 'limb' share the level *ngang* 'flat' tone (level: 33), but not their segments. Likewise, `chú` 'uncle' and `đế` 'sole' share the *sắc* 'sharp' tone (contour: 35), but have different consonants and vowels. Such an experiment would allow a clearer view of tonal lexical activation and would help confirm the results presented here.

One more aspect of this hypothesis has to do with the general role of education and tonal priming. The hypothesis presented here depends on continued exposure to a writing system with explicit tonal marking, which is present in Vietnamese but not in Mandarin, Cantonese or Thai. However, it also depends on stable tonal marking, which is not always present in the writing of Indigenous languages. This opens the possibility of a new experiment: Speakers of Indigenous languages who didn't learn to read and write using tones appear to have low tonal metacognition. But what would happen to children who did learn to read and write in their Indigenous language, and where tone might have had a more stable representation? Would their patterns match those of Vietnamese, where tonal priming is clearly visible? Studying this will be part of our future work.

5. Conclusions

We conducted a visual lexical decision task experiment that confirmed the presence of visual tonal priming in Vietnamese. This is important because this type of priming is not always found in other tonal languages such as Mandarin, Cantonese and Thai. Also relevant is the contrast this study has between Vietnamese and other tonal languages with Roman-based writing systems (i.e. Indigenous languages of the Americas and Africa), where tone is marked but priming is also difficult to find.

We have also outlined two hypotheses to explain why this pattern is present in Vietnamese. This could be an effect of visual similarity of the diacritics across prime-target orthography styles, or it could be due to the influence of the writing system of Vietnamese, which provides the speakers with stable and explicit mappings between graphemes and tones. Finally, we have laid out a series of future experiments to distinguish between these two hypotheses, and to further clarify the role of literacy in the development of tonal metacognition and priming.

6. Acknowledgements

The authors would like to thank all our participants for sharing their time and their knowledge of Vietnamese. We would also like to thank Prof. Samantha Wray, Prof. Sasha Calhoun, Prof. Paul Warren, Prof. Miriam Meyerhoff, Prof. Marcela Alfaro-Córdoba, Prof. Laura McPherson, Prof. James Stanford and audiences at Victoria University of Wellington Te Herenga Waka (henceforth VUW), Dartmouth College, and the TAI 2020 conference for their comments on previous iterations of this project. We also wish to thank Janet Attril, Matthew Vink, and the personnel at the Language Learning Center at VUW for their assistance with the project. Finally, we wish to acknowledge the financial support of VUW (Faculty Research Grant 216287) and the Royal Society of New Zealand Te Apārangi (18-VUW-184), without which the project could not have been possible.

7. References

[1] J. Edmondson and N. V. Lợi, "Tones and Voice Quality in Modern Northern Vietnamese: Instrumental Case Studies," *Mon-Khmer Studies*, vol. 28, no. 35, pp. 1–18, 1997.

[2] N. Đ. Hoà, *Vietnamese: Tiếng Việt Không Son Phấn*. John Benjamins Publishing Company. Amsterdam/Philadelphia, 1997.

[3] Z. Bao *et al.*, *The Structure of Tone*. Oxford University Press, 1999.

[4] M. Yip, *Tone*. Cambridge University Press, 2002.

[5] G. Peng, H.-Y. Zheng, T. Gong, R.-X. Yang, J.-P. Kong, and W. S.-Y. Wang, "The Influence of Language Experience on Categori-

cal Perception of Pitch Contours," *Journal of Phonetics*, vol. 38, no. 4, pp. 616–624, 2010.

[6] J. Xi, L. Zhang, H. Shu, Y. Zhang, and P. Li, "Categorical Perception of Lexical Tones in Chinese Revealed by Mismatch Negativity," *Neuroscience*, vol. 170, no. 1, pp. 223–231, 2010.

[7] K. I. Forster and C. Davis, "Repetition Priming and Frequency Attenuation in Lexical Access," *Journal of Experimental Psychology: Learning, Memory, and Cognition*, vol. 10, no. 4, p. 680, 1984.

[8] J. L. McClelland and D. E. Rumelhart, "An Interactive Activation Model of Context Effects in Letter Perception: I. An Account of Basic Findings," *Psychological Review*, vol. 88, no. 5, p. 375, 1981.

[9] L. Ferrand and J. Grainger, "Phonology and Orthography in Visual Word Recognition: Evidence from Masked Non-Word Priming," *The Quarterly Journal of Experimental Psychology Section A*, vol. 45, no. 3, pp. 353–372, 1992.

[10] L. J. Evett and G. W. Humphreys, "The Use of Abstract Graphemic Information in Lexical Access," *The Quarterly Journal of Experimental Psychology*, vol. 33, no. 4, pp. 325–350, 1981.

[11] L. M. Slowiaczek, J. M. McQueen, E. G. Soltano, and M. Lynch, "Phonological Representations in Prelexical Speech Processing: Evidence from Form-Based Priming," *Journal of Memory and Language*, vol. 43, no. 3, pp. 530–560, 2000.

[12] B. New and T. Nazzi, "The Time Course of Consonant and Vowel Processing During Word Recognition," *Language, Cognition and Neuroscience*, vol. 29, no. 2, pp. 147–157, 2014.

[13] H. Winskel and M. Perea, "Does Tonal Information Affect the Early Stages of Visual-Word Processing in Thai?" *Quarterly Journal of Experimental Psychology*, vol. 67, no. 2, pp. 209–219, 2014.

[14] H. Winskel, T. Ratitamkul, and A. Charoensit, "The Role of Tone and Segmental Information in Visual-Word Recognition in Thai," *Quarterly Journal of Experimental Psychology*, vol. 70, no. 7, pp. 1282–1291, 2017.

[15] X. Zhou and J. Zhuang, "Lexical Tone in the Speech Production of Chinese Words," in *Sixth International Conference on Spoken Language Processing*, 2000.

[16] C. Li, C. Y. Lin, M. Wang, and N. Jiang, "The Activation of Segmental and Tonal Information in Visual Word Recognition," *Psychonomic Bulletin & Review*, vol. 20, no. 4, pp. 773–779, 2013.

[17] J.-Y. Chen, T.-M. Chen, and G. S. Dell, "Word-Form Encoding in Mandarin Chinese as Assessed by the Implicit Priming Task," *Journal of Memory and Language*, vol. 46, no. 4, pp. 751–781, 2002.

[18] W. Ma, P. Zhou, L. Singh, and L. Gao, "Spoken Word Recognition in Young Tone Language Learners: Age-Dependent Effects of Segmental and Suprasegmental Variation," *Cognition*, vol. 159, pp. 139–155, 2017.

[19] A. Cutler and H.-C. Chen, "Lexical Tone in Cantonese Spoken-Word Processing," *Perception & Psychophysics*, vol. 59, no. 2, pp. 165–179, 1997.

[20] M. C. Yip, "Phonological Priming in Cantonese Spoken-Word Processing," *Psychologia*, vol. 44, no. 3, pp. 223–229, 2001.

[21] C.-Y. Lee, "Does Horse Activate Mother? Processing Lexical Tone in Form Priming," *Language and Speech*, vol. 50, no. 1, pp. 101–123, 2007, pMID: 17518105. [Online]. Available: https://doi.org/10.1177/00238309070500010501

[22] X.-D. Wang, A.-P. Liu, Y.-Y. Wu, and P. Wang, "Rapid Extraction of Lexical Tone Phonology in Chinese Characters: A Visual Mismatch Negativity Study," *PLoS One*, vol. 8, no. 2, p. e56778, 2013.

[23] X. Tong, C. McBride, and D. Burnham, "Cues for Lexical Tone Perception in Children: Acoustic Correlates and Phonetic Context Effects," *Journal of Speech, Language, and Hearing Research*, vol. 57, no. 5, pp. 1589–1605, 2014.

[24] L. Liu, D. Peng, G. Ding, Z. Jin, L. Zhang, K. Li, and C. Chen, "Dissociation in the Neural Basis Underlying Chinese Tone and Vowel Production," *Neuroimage*, vol. 29, no. 2, pp. 515–523, 2006.

[25] X. Li, J. T. Gandour, T. Talavage, D. Wong, A. Hoffa, M. Lowe, and M. Dzemidzic, "Hemispheric Asymmetries in Phonological Processing of Tones vs. Segmental Units," *Neuroreport*, vol. 21, no. 10, p. 690, 2010.

[26] J. E. Gombert and J. É. Gombert, *Metalinguistic Development*. University of Chicago Press, 1992.

[27] J. K. Torgesen, R. K. Wagner, and C. A. Rashotte, "Longitudinal Studies of Phonological Processing and Reading," *Journal of Learning Disabilities*, vol. 27, no. 5, pp. 276–286, 1994.

[28] C. McBride-Chang, E. Bialystok, K. K. Chong, and Y. Li, "Levels of Phonological Awareness in Three Cultures," *Journal of Experimental Child Psychology*, vol. 89, no. 2, pp. 93–111, 2004.

[29] P. H. Seymour, M. Aro, J. M. Erskine, and C. with COST Action A8 Network, "Foundation Literacy Acquisition in European Orthographies," *British Journal of Psychology*, vol. 94, no. 2, pp. 143–174, 2003.

[30] J. C. Ziegler and U. Goswami, "Reading Acquisition, Developmental Dyslexia, and Skilled Reading Across Languages: A Psycholinguistic Grain Size Theory," *Psychological Bulletin*, vol. 131, no. 1, p. 3, 2005.

[31] H. Wimmer and U. Goswami, "The Influence of Orthographic Consistency on Reading Development: Word Recognition in English and German Children," *Cognition*, vol. 51, no. 1, pp. 91–103, 1994.

[32] W. T. Siok and P. Fletcher, "The Role of Phonological Awareness and Visual-Orthographic Skills in Chinese Reading Acquisition," *Developmental Psychology*, vol. 37, no. 6, p. 886, 2001.

[33] H. Shu, H. Peng, and C. McBride-Chang, "Phonological Awareness in Young Chinese Children," *Developmental Science*, vol. 11, no. 1, pp. 171–181, 2008.

[34] C. Y. Lin, M. Wang, and H. Shu, "The Processing of Lexical Tones by Young Chinese Children," *Journal of Child Language*, vol. 40, no. 4, pp. 885–899, 2013.

[35] X. Chen, Y.-M. Ku, E. Koyama, R. C. Anderson, and W. Li, "Development of Phonological Awareness in Bilingual Chinese Children," *Journal of Psycholinguistic Research*, vol. 37, no. 6, pp. 405–418, 2008.

[36] M. Wang, C. A. Perfetti, and Y. Liu, "Alphabetic Readers Quickly Acquire Orthographic Structure in Learning to Read Chinese," *Scientific Studies of Reading*, vol. 7, no. 2, pp. 183–208, 2003.

[37] Y. Liu, M. Wang, C. A. Perfetti, B. Brubaker, S. Wu, and B. MacWhinney, "Learning a Tonal Language by Attending to the Tone: An In Vivo Experiment," *Language Learning*, vol. 61, no. 4, pp. 1119–1141, 2011.

[38] H. Winskel and K. Iemwanthong, "Reading and Spelling Acquisition in Thai Children," *Reading and Writing*, vol. 23, no. 9, pp. 1021–1053, 2010.

[39] B. New, V. Araújo, and T. Nazzi, "Differential Processing of Consonants and Vowels in Lexical Access through Reading," *Psychological Science*, vol. 19, no. 12, pp. 1223–1227, 2008.

[40] M. R. Kadyamusuma and M. Kadenge, "The Potential Effect of Non-Tone Marking on Homograph Processing in Shona: A Preliminary Exploration," *Per Linguam: a Journal of Language Learning= Per Linguam: Tydskrif vir Taalaanleer*, vol. 29, no. 2, pp. 85–94, 2013.

[41] F. E. Saynes-Vazquez, "Zapotec Language Shift and Reversal in Juchitán, Mexico," Ph.D. dissertation, The University of Arizona, 2002.

[42] R. A. Coto Solano, "Tonal Reduction and Literacy in Me̲'pha̲a̲ Vátháá," Ph.D. dissertation, The University of Arizona, 2017.

[43] UNESCO Institute for Statistics. (2022) Vietnam: Education and Literacy. http://uis.unesco.org/en/country/vn. [Online]. Available: http://uis.unesco.org/en/country/vn

[44] H. Pham and H. Baayen, “Vietnamese Compounds Show an Anti-Frequency Effect in Visual Lexical Decision,” *Language, Cognition and Neuroscience*, vol. 30, no. 9, pp. 1077–1095, 2015.

[45] N. Liu, “Implicit Priming Effects in Chinese Word Recall: The Role of Orthography and Tones in the Mental Lexicon,” *International Journal of Computational Linguistics & Chinese Language Processing: Special Issue on Processing Lexical Tones in Natural Speech*, vol. 18, no. 3, 2013.

[46] M. Brunelle and S. Jannedy, “The Cross-Dialectal Perception of Vietnamese Tones: Indexicality and Convergence,” *Linguistics of Vietnamese, Amsterdam: John Benjamins*, 2013.

[47] G. Pham, K. D. Ebert, K. T. Dinh, and Q. Dam, “Nonword Repetition Stimuli for Vietnamese-Speaking Children,” *Behavior Research Methods*, vol. 50, no. 4, pp. 1311–1326, 2018.

[48] G. T. Pham, S. Pruitt-Lord, C. E. Snow, Y. H. T. Nguyen, B. Phạm, T. B. T. Dao, N. B. T. Tran, L. T. Pham, H. T. Hoang, and Q. D. Dam, “Identifying Developmental Language Disorder in Vietnamese Children,” *Journal of Speech, Language, and Hearing Research*, vol. 62, no. 5, pp. 1452–1467, 2019.

[49] K. Forster, “Form-Priming with Masked Primes: The Best Match Hypothesis,” in *Attention and Performance 12: The Psychology of Reading)*, M. Coltheart, Ed. Lawrence Erlbaum Associates, Inc, 1987.

[50] L. N. Phong, T. A. Lưu, and N. Văn Hai, *Natural Language Processing - VNESE Corpus*. Nagaoka University of Technology, 2012, http://viet.jnlp.org/download-du-lieu-tu-vung-corpus.

[51] E. Keuleers and M. Brysbaert, “Wuggy: A Multilingual Pseudoword Generator,” *Behavior Research Methods*, vol. 42, no. 3, pp. 627–633, 2010.

[52] K. I. Forster and J. C. Forster, “DMDX: A Windows Display Program with Millisecond Accuracy,” *Behavior Research Methods, Instruments, & Computers*, vol. 35, no. 1, pp. 116–124, 2003.

[53] D. Bates, M. Mächler, B. Bolker, and S. Walker, “Fitting Linear Mixed-Effects Models Using lme4,” *Journal of Statistical Software*, vol. 67, no. 1, pp. 1–48, 2015.

[54] R Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2021. [Online]. Available: https://www.R-project.org/

[55] S. Kinoshita, L. Yu, R. G. Verdonschot, and D. Norris, “Letter Identity and Visual Similarity in the Processing of Diacritic Letters,” *Memory & Cognition*, vol. 49, no. 4, pp. 815–825, 2021.