Segment-tone integration in word identification by Dutch-Vietnamese heritage speakers

Jerom Ebenau
Linguistics ResMA, Leiden University
Dr. Y. Chen

26 February 2019
Word count: 29990

TABLE OF CONTENTS

0. ABSTRACT

Much of the literature on heritage language phonology finds heritage speakers to show some influence from their dominant language compared to homeland speakers, but heritage speakers still perform more accurately in their heritage language than do naïve speakers. Yet, research on heritage language phonology is limited compared to that on heritage language syntax and morphology. This is even more so the case for research on heritage speakers' suprasegmental phonology: for instance, very little is known about heritage speakers' perception of lexical tone. The present study used an ABX task to investigate perceptual segment-tone integration in heritage speakers of Vietnamese in the Netherlands, compared to monolingually raised Dutch and Vietnamese speakers in the homeland, respectively. Heritage speakers were found to have a stronger segment-tone integration than the monolingually raised Dutch, whereas the homeland Vietnamese showed a slightly stronger integration than the heritage speakers. Moreover, the groups' integrations were asymmetrical: heritage speakers considered both tones and segments in word identification but had a clear preference for segments; the Dutch controls almost exclusively considered segments and the Vietnamese controls had a slight preference for tone-based word identification. The findings thus conform to previous literature on heritage language phonology: the heritage speakers performed intermediately between monolinguals of their heritage and dominant languages.

Keywords: heritage language, Southern Vietnamese, Dutch, segment-tone integration, word processing

1. INTRODUCTION

This thesis aims to provide insight into a type of language users who, with ongoing globalisation, are likely to become a very prevalent population: heritage speakers (HSs). HSs grow up speaking a minority language at home and, at an early age, start acquiring the dominant language of the area, often to the detriment of the minority (heritage) language (Polinsky & Kagan 2007, see Section 3). The speakers under investigation here were Vietnamese HSs in the Netherlands, who are dominant in Dutch. Vietnamese is a lexical tone language, whereas Dutch is not. Hence the study investigated the HSs' integration of segments and lexical tone in perception, compared to monolingually raised speakers of Vietnamese and Dutch, respectively. That is, to what extent do the groups rely on just one dimension or need both in order to efficiently process words (see Section 2.3)?

This research is relevant for various reasons. First of all, phonological research has generally focused on segmental phonology more so than on suprasegmental phonology. Although there have been numerous studies on lexical tone phonology, many questions remain unanswered. One question concerns the interaction between and integration of segments and tones in perception. Studies on this topic do not always produce findings that agree with each other, often due to the use of different methodologies or paradigms. This highlights that despite substantial research being done already, segment-tone integration should be explored further.

Furthermore, studies on HSs were originally mostly focused on syntax and morphology. Although studies on heritage language phonology have become more common in the past three decades, this does not hold for lexical tone in HSs, a subject that has rarely been researched. Often heritage language phonologies have similarities to the phonologies of both monolinguals of the minority language and of monolinguals of the dominant language. The same tendency was expected in this investigation.

The present study contributes to the literature by combining these two understudied lines of research, providing evidence with perception data from an ABX task (see Section 4) conducted with HSs and monolinguals of their respective languages. Additionally, research on HSs in general provides insights in language acquisition and can have useful implications for language policymaking on community and family levels as well as for language education.

In the following there will be a literature review of studies on lexical tone (Section 2) and HSs (Section 3). Next, the present study's research questions and design are discussed in more detail (Section 4), followed by an overview of the results (Section 5) and the discussion (Section 6). The main conclusions from the study are discussed in Section 7.

2. LEXICAL TONE

Lexical tone (henceforth also referred to as 'tone') can be defined as the modulation of pitch to attribute a different lexical meaning to a word.[1] It is common in most languages and can be found in the Americas, Africa, Asia, Oceania and Europe (Yip 2002, 2007; Maddieson 2013). An example of Vietnamese minimal pairs differing in tone would be the following:

(1)	*ma*	'ghost'	high level tone
	mà	'but, yet'	mid falling tone
	má	'cheek'	rising tone
	mạ	'rice seedling'	low glottalised tone
	mả	'tomb'	low falling/dipping tone
	mã	'code'	high rising and glottalised tone

(adapted from Kirby 2011: 386)

The words in (1) have the same segmental makeup /ma/ but differ in tones (marked by diacritics). More information on tone in Vietnamese is provided in Section 2.4.

Usually two main types of tone languages are distinguished: contour tone languages and register tone languages. Contour tones come in various shapes, as suggested by the above minimal pairs, and are characterised by a particular pitch movement. Contour tones are most commonly found in languages that distinguish many tonal categories. Conversely, register tone languages have tones with a constant pitch level (i.e. level tones). Thus, in these languages, there is a clearer relative pitch range associated with a particular tone than in contour tone languages. Languages that have only few (2-3) tonal categories, usually exclusively have level tones. A language may have up to five level tones and up to three contour tones (e.g. rising, falling, dipping) of the same shape (Yip 2002, 2007; Maddieson 2013; Singh & Fu 2016).

In daily language use, tones are often accompanied by other tones, which can lead to tone sandhi: in a set of subsequent tones, one may influence the other and change its contour and register to facilitate production or perception. Moreover, tones are inevitably accompanied by segments (although on the underlying level, they may be considered to be separate, Goldsmith 1976; Yip 2007), which can also have an effect on tones. For example, both voicing of consonants and pitch rely on vocal fold vibrations. Voiced obstruents have been observed to lower pitch, which consequently lowers tone. Conversely, voiceless obstruents raise pitch and tone (Yip 2007; see also Chen 2011 for a review on these effects as well as a more in-depth investigation on the interaction of segments and F0 in Shanghai Chinese). Breathiness and glottalisation in vowels seem to be correlated with low tone, whereas high vowels are expected to be associated with higher tones (Yip 2002).

Tone processing is one of the main focus points of the present study. In the following, first acquisition of tone contrasts in speakers of tonal languages in general is discussed (Section 2.1). Studies which involved the perception of tones by speakers of non-tonal languages are discussed as well, since the HSs in the present study, apart from speaking Vietnamese, are dominant in a non-tonal language: Dutch (Section 2.2). Next, studies on segment-tone integration in both tonal and non-tonal speakers are discussed (Section 2.3). After this general discussion, there will be a section on Vietnamese tone (Section 2.4) and a brief section on tone in some Dutch dialects (Section 2.5).

2.1. Acquisition of tonal contrasts

Yip (2007) reports that not much is known about how tone is acquired, apart from the assumption that it is acquired very early on. Singh & Fu (2016) likewise argue that too little research on L1 acquisition is based on tone languages. They argue that children might acquire tones more easily than segments as they pay more attention to pitch and other prosodic categories to identify infant-directed speech, their L1, and emotions. Moreover, within tonal languages, there are usually fewer tones than

[1] Yip (2002) reports that other cues, such as duration, amplitude and voice quality are also relevant to tone perception, but also notes that F0 is the only *necessary* cue for tone perception (Yip 2002: 291).

segments, therefore Singh & Fu (2016) speculate that the smaller number of tone categories provides a relatively lighter processing load than that of the segmental categories, making the former easier to acquire. Segments become increasingly apparent in children's perception from 6 to 12 months (Kuhl 1983; Werker & Tees 1984; Werker & Hensch 2015).

Studies using head-turning paradigms, such as Harrison (2000) and Singh & Foong (2012), have suggested that infants acquiring Yoruba and Mandarin already learn to distinguish tones in their respective native languages at an age between six to eight months. For children acquiring Yoruba, Harrison found that these perceived distinctions were limited to specific F0 excursions and ranges. He also found that peers acquiring English were not able to distinguish pitch. Note that Harrison (2000) had only six participants per group and these results may therefore not be generalisable. Still, the results suggest that children acquiring a tone language may consider pitch as contrastive at a very early age. Singh & Foong (2012) moreover found that children acquiring both Mandarin and English initially distinguish pitch in both languages, and then at 9 months no longer distinguish pitch in either language. Only at 11 months, the bilingual infants again attributed contrastiveness to pitch in Mandarin, whereas in English they did not do so any longer, which suggests that the infants had learnt the role of pitch respective to each of their languages.

Singh et al. (2015) found that for toddlers (ages 2;6-3;6) and pre-schoolers (4;5) acquiring Mandarin, both segments and tone are contrastive. Although the participants were bilingual in Mandarin and English, Mandarin was the primary language used at home and in pre-school. The children were shown pairs of images, each containing a familiar target object and an unfamiliar distractor. The target was labelled verbally and in half of the trials, this label was mispronounced with a vowel, consonant, or tone substitution. Toddlers were found to accurately recognise mispronunciations. This was especially the case for tonal mispronunciations: when targets were named with a tonal mispronunciation, the toddlers did not reliably look at the target. Conversely, vocalic and consonantal mispronunciations still led the toddlers to look at the target instead of the distractor. Tone thus seems to be the most salient feature in a word to this group. For the pre-schoolers, tonal mispronunciations did not lead to a different effect than correct pronunciations, whereas vocalic and consonantal mispronunciations did. This suggests that for this older age group, tone has become less important than segments in recognising words. Singh and colleagues argued that perhaps this is due to the older children realising that pitch may fulfil more functions than just that of tone. They furthermore concluded that for both groups, tone and segments seem to be dissociated from each other, i.e. that they might not be strongly integrated. Other experiments investigating attention distribution between tone and segments will be discussed in Section 2.3.

The difference between the early emergence of tonal contrasts and adult-like perception of tone should be stressed. Ciocca & Lui (2003) showed that although at age six, Cantonese-speaking children are able to perceive tonal contrasts, their perception does not become adult-like until age 10. In their experiment, opposing pairs of Cantonese tones were produced on otherwise identical syllables. Participants from different age groups heard sentences containing these target words and were instructed to indicate which word they heard using pictures representing the target and a competitor. A significant improvement among 6-year-olds compared to 4-year-olds was found, as well as for 10-year-olds compared to 6-year-olds, although only for four out of six Cantonese tones. There were no significant differences between 10-year-olds and adults, suggesting that children reach an adult-like perception of tone by this age. Ciocca & Lui also noted that the 4-year-olds already performed above chance level for most tonal contrasts, which shows that although tone perception is not yet accurate at this age, it is present.

In addition to studies investigating when children become sensitive to tone, there have also been studies looking into when this sensitivity is lost. Yeung et al. (2013) tested the perception of Cantonese tones by three groups of 4-month-olds and 9-month-olds acquiring English, Mandarin, or Cantonese, respectively. Two tones were presented on segmentally identical CV syllables: a rising tone and a mid-level tone. The children were presented with the auditory stimuli along with a visual stimulus on a tv screen. When the auditory stimuli played, the time children spent looking at the visual stimulus was measured. Within the English group, only 4-month-olds had different looking times for different tones.

Yeung and colleagues thus concluded that the English 4-month-olds could still discriminate tone but that this differentiation is lost in 9-month-olds. Earlier a similar tendency for children acquiring non-tonal languages to lose sensitivity to tone between the ages of six to nine months was found (Mattock & Burnham 2006; Mattock et al. 2008). In addition to a similar tendency, Liu & Kager (2014) furthermore found that at 18 months, children who learn a non-tonal language may display a greater ability to discriminate tones again. Singh et al. (2014) also found that 18-month-old children acquiring English were still able to recognise tones, whereas at 24 months they no longer were able to. These studies thus suggest that tone can lose its contrastiveness in children acquiring a non-tonal language anywhere between the ages of four months to two years.

For the Mandarin and Cantonese groups in Yeung et al.'s (2013) study, age was not found to be a relevant factor, but the tone used in familiarisation trials was. Only children familiarised with rising tones were able to discriminate tones in the rest of the experiment. Moreover, for these children it was found that the Mandarin group paid more attention when hearing rising tones, which have an equivalent in Mandarin (as opposed to the mid-level tone). For the Cantonese children, there was no preferred tone. Yeung and colleagues concluded that children become sensitive to language-specific tonal patterns as early as 4 months of age. However, children acquiring a tonal language may still have an advantage over children acquiring a non-tonal language when hearing foreign tones, as the English group lost sensitivity at 9 months, whereas the Mandarin group was still sensitive to Cantonese tones at that stage.

The presented studies suggest that the acquisition of tonal contrasts and perception starts at an early age: starting from six months to two years, children are found to attribute contrastiveness to pitch (Harrison 2000; Singh & Foong 2012). However, tone perception does not reach adult accuracy until age 10 (Ciocca & Lui 2003). Sensitivity to tone in children acquiring a non-tonal language may be lost almost as early as it is acquired by children acquiring a tonal language: starting at nine months children acquiring English have been shown to no longer discriminate words based on pitch (Mattock & Burnham 2006; Mattock et al. 2008; Yeung et al. 2013). Although this sensitivity may still be present nine months later, it is not found anymore at 2 years of age (Singh et al. 2014). In children bilingual in a tonal and a non-tonal language, tone perception may receive a smaller role at a later age, when segments become relatively more contrastive (Singh et al. 2015).

The present study focuses on heritage speakers who switched in dominance from a tonal language (Vietnamese) to a non-tonal language (Dutch) around age 4. The studies discussed show that at this age, the acquisition of tone is not yet complete, although tonal contrasts have become part of the speakers' phonologies. Evidence from the present study could thus provide further insights into the development of tone perception in speakers of tonal languages. For instance, the fact that among the HSs this development was interrupted at age 4 can provide information on the time table of tone acquisition.

2.2. Tone perception in non-tonal language speakers

There have also been various studies investigating tone perception in L2 learners of tone languages who have a non-tonal L1. Although the studies from the previous section show that L1 speakers of a non-tonal language lose sensitivity to tone at an early age, other studies show that this sensitivity can later be regained to an extent. Untrained listeners with a non-tonal L1 clearly perceive tone differently compared to speakers of a tonal language. In an fMRI study using pseudowords and hummed sentences, Gandour et al. (2003) found that Mandarin speakers were considerably more accurate (98% correct) than English speakers (61%) in a same-different task testing perception of Mandarin tones. Moreover, the Mandarin group reported finding this task relatively easy, whereas the English group found the task quite difficult.

However, training L2 listeners can make a great difference. Wang & Kuhl (2003) researched how young L1 American English groups could better their perception of Mandarin tones after training. In a pre- and post-test, listeners of 6, 10, 14, and 19 years old had to identify Mandarin tones. In between the two tests was a training phase which lasted two weeks. Regardless of participants participating in the pre- or post-test and regardless of receiving training or not, older participants were generally found to perform better than younger participants. Furthermore, participants who received training were found to perform significantly better in the post-test, whereas no such effect was found for the control group that did not receive training.

Similar to Wang & Kuhl (2003), Francis et al. (2008) trained adult participants between two testing moments. They investigated the perception of Cantonese tones by adult L1 speakers of Mandarin and L1 speakers of English. The groups participated in a tone identification task and a difference rating task where they indicated how (dis)similar pairs of words were. In the identification task both English and Mandarin listeners were found to have improved in the post-test, although neither performed as accurately as Cantonese controls. Mandarin listeners improved most on low falling tone and English listeners improved most on low level and low rising tones. Note how these results relate to those of the children in Yeung et al. (2013) (Section 2.1), where Mandarin 4- and 9-month-olds were found to be more attentive to Cantonese tone than English 9-month-olds. After training, adult English speakers in Francis et al. (2008) were not found to be less accurate than the Mandarin group. Additionally, the difference rating task revealed that originally, English listeners paid relatively more attention to tone height than direction and that this distribution of attention was even more shifted towards height after the training. For Mandarin listeners, there was originally more attention for direction, whereas after training they paid about an equal amount of attention to height and direction. The Cantonese control group paid more attention to height than to direction. From these results, Francis and colleagues concluded that speakers of tonal languages are not necessarily at an advantage in learning the tones of an L2 compared to speakers of non-tonal languages. Instead, speakers may experience advantages in an L2 when they have similar categories in their own language; in Francis et al.'s (2008) case, tones in Mandarin and intonation patterns in English corresponding to some of the tones in Cantonese might have helped achieve higher accuracy for each respective group.

L1 categories are not the only factor that could help L2 learners perceive tone more accurately. There have been various studies on musical ability and the perception of tone, such as Lee & Hung (2008) and Delogu et al. (2010), who tested adult L1 speakers of American English and of Italian, respectively. Both studies tested how well musicians and non-musicians speaking these languages perceive Mandarin tone. Musicians were found to perform more accurately than non-musicians, and in Delogu et al.'s (2010) study, they even performed similarly to advanced learners of Mandarin in tone perception.

The studies discussed here suggest that even with a non-tonal L1, L2 learners may be able to improve their tone perception with appropriate training (Wang & Kuhl 2003) and are not necessarily at a disadvantage compared to other L2 learners who have a tonal L1 (Francis et al. 2008). Even without explicit training in lexical tone perception, musical training can still provide an advantage in perceiving lexical tone (Lee & Hung 2008; Delogu et al. 2010). Without training, however, the difference between these two types of L2 learners (tonal or non-tonal L1) is evident and a lack of experience with a tonal language is found to negatively affect tone perception (Gandour et al. 2003). The Dutch participants in the present study were untrained and unfamiliar with lexical tone and were therefore expected to perform poorly when processing tonal information. The Dutch controls' unfamiliarity with tone helps highlight the sensitivity to tone that Vietnamese heritage speakers' have despite their dominance in Dutch.

2.3. Tone, segments, and attention

Tone and segments are inherently integrated at the surface, as tone cannot appear at the surface without a tone-bearing unit (TBU). However, on the phonological level this may not be the case (Goldsmith 1976). It is thus worthwhile to consider to what extent speakers of a tonal language pay attention to each dimension respectively and to what extent this perception is integrated below the surface. A fair amount of literature on this subject has already been published, mostly focusing on Chinese languages. However, it is noted by Lin & Francis (2014) that results from these studies are sometimes difficult to compare, indicating that more research needs to be done.

Garner (1974) described how the integration of tones and segments in perception may or may not be symmetrical. If segments and tones are symmetrically integrated, speakers' processing and perception of one dimension is influenced by the other dimension and, vice versa, the latter is influenced to the same degree by the former. However, if the dimensions are asymmetrically integrated, speakers' processing of tone might be influenced more by segments than their processing of segments is influenced by tone, or vice versa.

Symmetrical and asymmetrical segment-tone integrations as observed in children and adults are discussed in Subsection 2.3.1 and Subsection 2.3.2, respectively.

2.3.1. Segment-tone integration in infants and children

Wewalaarachchi et al. (2017) conducted an eye-tracking study with Mandarin-English bilingual and Mandarin monolingual 2-year-olds. Participants in both groups did not reliably look at target images when the words corresponding to the images were mispronounced segmentally or tonally, whereas correct pronunciations had no such effect. These results suggest a symmetrical integration, since both mispronouncing the segmental dimension and mispronouncing the tonal dimension hinder word recognition, i.e. there was not one dimension that could consistently lead to word recognition despite the other dimension being mispronounced.

However, the bilinguals were more likely than monolinguals to keep looking at distractor images when there were tonal mispronunciations, indicating that perhaps the segment-tone integration in the bilingual children was not completely symmetrical after all. Note that Singh et al.'s (2015) results differed from this observation (see Section 2.1). In their study, Mandarin-English bilingual children of age 2;5 were quite sensitive to tone. However, Singh and colleagues also found that this sensitivity was lost in bilinguals around age 4;5.

Similarly to Wewalaarachchi et al., Ma et al. (2017) compared Mandarin monolingual 2-year-olds' to 3-year olds' preferential looking across trials with correct pronunciations as well as trials with vocalic or tonal mispronunciations. Both age groups were slower to identify novel words when they were mispronounced either vocalically or tonally. 3-year-olds were only less accurate when hearing vowel mispronunciations, whereas the 2-year-olds were negatively affected by any kind of mispronunciation. In an additional experiment more 3-year-olds were tested, but with familiar words. These 3-year-olds were found to be slowed down by both tonal and vocalic mispronunciations, although more so by the latter than the former. Moreover, the proportion of fixations on the target image was smaller in vowel mispronunciation trials than in correct or tonal mispronunciation trials. Therefore, Ma et al. (2017) concluded that 3-year-old Mandarin learners are more sensitive to vowels than tones, both in learning new words and in recognising familiar words. Moreover, they are less sensitive to tone than 2-year-olds.

Focusing on older children as well as adults, Burnham et al. (2011) conducted experiments with Thai, Cantonese, and (Australian) English listeners. They investigated the relations between awareness of segments vs. tones and age, script, reading ability, level of education, and (non-)tonal language background. Within each language group, children from various grades (kindergarten, year 2, 4, and 6) in school were tested, as well as adults. In the odd-one-out task, participants had to find a deviating word in sets of 3 Thai words. The words could deviate in tone, vowel, or both. For all groups, full mismatch trials were easier than trials with just tonal or vocalic mismatches, indicating that all groups perceive segments and tones in an integrated manner. However, all groups performed more accurately on vocalic mismatch trials than on tonal mismatch trials, indicating that the segment-tone integration in each group is asymmetrical. For Thai and English listeners, this asymmetry became stronger over age, whereas for the Cantonese listeners, the difference became smaller. Overall, adults were found to respond more accurately than children, which means that even around age 12 (when most children are in year 6), children's perception of tones as well as segments is not yet as accurate as adults' perception.

These studies on segment-tone integration in children show that as early as age 2, children acquiring a tonal language may process segments and tones in an integrated way (Burnham et al. 2011; Wewalaarachchi et al. 2017; Ma et al. 2017). The findings from these studies suggest that even if the Vietnamese heritage speakers in the present study started learning and becoming dominant in Dutch around age 4, they are likely to have acquired some degree of segment-tone integration before this age. However, Wewalaarachchi and colleagues' results showed that bilingual children may start to have a segment-tone integration different from that of monolingual peers very early on. Moreover, results from Ma et al. (2017) and Burnham et al. (2011) also suggest that at age 2 and a considerable time after that, this integration is not yet adult-like. If, following the literature on heritage languages (see Section 3), it is assumed that the heritage speakers' early dominance in Dutch inhibits their acquisition of Vietnamese, these studies' results could suggest that the heritage speakers in the present study may not display a segment-tone integration that would be found in adult monolingual speakers of Vietnamese. The present study's experiment could thus provide further information on how segment-tone integration may stabilise in heritage speakers who do not follow a monolingual path of acquisition.

2.3.2. Segment-tone integration in adults

Various studies on segment-tone integration have used Garner's (1974) method to determine (a)symmetrical segment-tone integration, referred to as Garner speeded classification tasks (Garner 1970, 1974, 1976; Lee & Nusbaum 1993; Tong et al. 2008; Lin & Francis 2014). For example, Tong et al. (2008) used these tasks with Mandarin speakers. Participants were asked to classify words based on vowels, consonants, or tone. In the baseline condition, the target dimension (e.g. vowel) changes, whereas the other dimensions (consonants, tone) remain constant (see example 2a; in this example superscript numbers refer to tone). In the orthogonal condition, both the target dimension and one or multiple non-target dimensions change (2b). Therefore, in the orthogonal condition, the participants have to actively ignore changes in the non-target dimensions to give an accurate response about the target dimension. Participants who are slower or less accurate in the orthogonal dimension than in the baseline dimension may be argued to have difficulties in ignoring one or more dimensions to focus solely on a target dimension, and thus have a more integrated perception of the target and distractor dimensions.

(2a) *Baseline Condition*

/ba^2/ vs. /bu^2/ Only the vocalic target dimension varies between a pair of stimuli.

/da^4/ vs. /du^4/ Only the vocalic target dimension varies between a pair of stimuli.

(2b) *Orthogonal Condition*

/ba^2/ vs. /bu^4/ The vocalic target dimension as well as the tonal non-target dimension vary.

/da^4/ vs. /bu^4/ The vocalic target dimension as well as the consonantal non-target dimension vary.

(stimuli examples from Tong et al. 2008: 696)

Tong et al. (2008) found that reaction times (RTs) were similar for all dimensions in the baseline condition, i.e. the Mandarin listeners were able to identify vowels just as easily as consonants and tones. In the orthogonal conditions, all RTs were longer. Further analysis showed that the consonantal dimension had a greater influence on the tonal dimension than vice versa, affecting both RTs and accuracy of classification. The same held for the influence of vowels on tones and of vowels on consonants. As these latter two combinations showed greater effects, Tong and colleagues concluded that vowels and tones, as well as vowels and consonants, are more integrated than consonants and tones. They also argued that the segmental dimensions influence the tonal dimension more strongly than vice versa and that this might be because consonants and especially vowels are able to distinguish more words in Mandarin than are tones. That is, segments are more informative, because there are more of them. Segments are thus more worthwhile for listeners to pay attention to (Tong et al. 2008: 702-703).

Lin & Francis (2014) used similar Garner speeded classification tasks with Mandarin speakers (as well as a group of English speakers, see below). To see whether language expectation influences segment-tone integration, half of the Mandarin participants were exposed to Mandarin during the task and prior correspondence, whereas the other half were exposed to English. Lin & Francis only analysed the integration between consonants and tones. In both Mandarin groups, tone and consonants mutually influenced each other in speech processing; i.e. when both dimensions varied, RTs were slower than when only one dimension varied. For both groups, the integration was symmetrical. Lin & Francis furthermore concluded that language expectation (i.e. expecting to hear a tonal or non-tonal language) does not influence the integration of consonants and tones.

Cutler & Chen (1997) conducted various other experiments with Cantonese speakers. One was a discrimination task, where disyllabic words could deviate in vowel, onset, tone, or a combination of these parts. For trials with just a tonal mismatch, longer RTs and decreased accuracy were found, whereas no such effects were found for other types of mismatches. In fact, the onset and vowel mismatches improved RTs and accuracy. These results suggest that the Cantonese listeners found it difficult to identify tones as identical or different, whereas it was easier to identify segments as such. The longer RTs caused by tone mismatches were taken by Cutler & Chen as evidence that tones may be processed later than the vowels they appear on. Yip et al. (1998) found similar results for Cantonese participants in a primed word repetition task. In a later experiment with Mandarin participants, Ye & Connine (1999) confirmed Cutler & Chen's (1997) results but also found that in an idiomatic context, tone mismatches are actually detected faster than vowel mismatches. This suggests that in meaningful contexts, as opposed to the isolated stimuli in Cutler & Chen (1997), tone could be more salient than vowels.

Especially relevant to the present study and its methodology is Zou et al.'s (2017) research. They investigated how Dutch learners of Mandarin process segments and tones respectively, as well as how integrated the perception of the two dimensions is. Zou and colleagues tested three main groups of speakers: Dutch controls with no experience with Mandarin, Mandarin controls who had not been in a Dutch environment for more than 3 years, and Dutch learners of Mandarin, who were subdivided into beginner and advanced groups. The participants completed an ABX task in which they heard sets of three disyllabic non-words varying segmentally and/or tonally in their first syllables. This methodology was based on Braun & Johnson's (2011) study, who used the ABX task with Dutch and Mandarin monolinguals only. Zou and colleagues created four conditions: one of the standards could match with X in segments while tone had to be ignored (forced-segment condition), in tone while segments had to be ignored (forced-tone condition), in both dimensions (segment-and-tone condition), or one standard matched with X regarding tone, whereas the other matched segmentally (segment-or-tone condition).

In the forced-segment condition, the groups did not differ in accuracy, although the Dutch controls and beginner learners were faster than the advanced learners. In the forced-tone condition, the Mandarin controls and advanced learners performed better than the Dutch controls and beginner learners, who patterned similarly to each other. In the segment-or-tone condition, Mandarin controls and advanced learners paid less attention to segments (in 62.2% and 69.2% of trials, respectively) than Dutch controls (90.4%) and beginner learners (85.5%) did. Tone was thus more important to participants with more experience with Mandarin than to those with less experience. These results also corresponded to Braun & Johnson's (2011) findings for Dutch and Mandarin monolinguals.

Regarding RTs, both Mandarin controls and advanced learners were found to take longer than the other groups to respond when they had to ignore one dimension to the benefit of another. This would mean that the Mandarin controls and advanced learners have a stronger segment-tone integration than beginner learners and Dutch controls. Similar findings for non-tonal speakers were found by Lin & Francis (2014), whose English listeners showed no segment-tone integration. Cutler & Chen (1997), however, found that listeners in both their Cantonese and Dutch groups experienced difficulties processing tones.

Based on the advanced learners' performance compared to the Mandarin controls, Zou and colleagues conclude that tone and its integration with segments is learnable for L1 speakers of a non-tonal language and that it can be phonologised. Lastly, all groups were slower in the forced-tone condition than in the forced-segment condition. This means that even if a group's processing of the two dimensions was integrated, this integration was asymmetrical and the segmental dimension was more important than the tonal dimension.

In brief, due to the use of different paradigms and methodologies, results across studies on segment-tone integration may sometimes be difficult to compare. Although most studies find some degree of integration in most speakers regardless of language background, some find the integration is asymmetrical (Cutler & Chen 1997; Yip et al. 1998; Ye & Connine 1999; Tong et al. 2008; Braun & Johnson 2011; Burnham et al. 2011; Singh et al. 2015; Wewalaarachchi et al. 2017; Zou et al. 2017; see also Lee & Nusbaum 1993), whereas proof for a symmetrical integration also exists (Lee & Nusbaum 1993; Burnham et al. 2011; Lin & Francis 2014; Ma et al. 2017; Wewalaarachchi et al. 2017). Furthermore, most studies find that asymmetrical integration usually benefits perception of segments (Cutler & Chen 1997; Yip et al. 1998; Tong et al. 2008; Braun & Johnson 2011; Burnham et al. 2011; Singh et al. 2015; Wewalaarachchi et al. 2017), while some find it benefits tone perception instead (Ye & Connine 1999; Singh et al. 2015; Ma et al. 2017). Segment-tone integration may be absent in non-tonal speakers (Lin & Francis 2014; Zou et al. 2017), but if it is present, it usually favours segments more strongly than in tonal speakers (Lee & Nusbaum 1993; Braun & Johnson 2011). Variability in results also persists in neurolinguistic studies, with event-related potential (ERP) studies like Hu et al. (2012) and Tong et al. (2014). Hu and colleagues find an asymmetrical integration benefiting segments in Mandarin adults, while Tong and colleagues find a fairly symmetrical integration in Cantonese-speaking 7- to 8-year-olds. Moreover, Gandour et al. (2003) find that speakers of Mandarin process both segments and tone in the left hemisphere.

It is to be expected that different paradigms yield different perspectives on segment-tone integration. With evidence from various paradigms being available, it is now important that further research helps to strengthen this evidence with results obtained through the same paradigms. The present study uses a methodology similar to Zou et al.'s (2017), although with Vietnamese and Dutch speakers instead of Mandarin and Dutch speakers. This allows for an almost direct comparison to Zou and colleagues' results and thus could help support their evidence. The present study will furthermore contribute to research on segment-tone integration by investigating heritage speakers, a group of speakers that has not yet been specifically considered in this line of research (see Section 3.2.2).

2.4. Tone in Vietnamese

As briefly touched upon in the introduction, Vietnamese is a tone language. Generally, it is suggested that there are minimally four and up to six Vietnamese tones depending on dialect:[2] the two tones relevant to this study are *sắc*, with mid-high rising pitch, and in writing marked with an accent *aigu* on the vowel (referred to here as 'rising tone'); and *nặng*, with low dropping pitch, rising to mid-high pitch in some circumstances, sometimes accompanied by glottalisation, and in writing marked with a dot below the vowel it is produced on (referred to here as 'falling tone'). The other tones are *ngang*, with mid-high level pitch; *huyền*, with low pitch and marked with a grave accent; *hỏi*, with mid-low dropping or dipping (i.e. rising to high after dropping) pitch and marked by a 'hook' accent; *ngã*, with a high rising pitch and a sometimes glottalised voice quality of the vowel it is produced on, marked with a tilde. *Ngang* is not marked in writing and therefore every vowel without a tone diacritic has *ngang* high level tone (Thompson 1965: 16; Mai 1967: 20-21; Phạm & McLeod 2016).[3] Figures 1 and 2 (created using a Praat script by Elvira García 2018) show F0 contours of two stimuli used in the present study respectively carrying the rising *sắc* and falling *nặng* tones on the first syllables and both carrying high level *ngang* on the second syllable.

[2] Some authors, however, argue that there are up to eight tones in (Northern) Vietnamese, cf. Pham (2003). In this analysis, *sắc* and *nặng* tone have alternative versions that appear in closed syllables and are classified as two additional tones.

[3] Note that vowels may also have diacritics to designate different vowel qualities: circumflex (*a* [a] vs. *â* [ʌ]; *e* [ɛ] vs. *ê* [e]; *o* [ɔ] vs. *ô* [o]), breve accent (*a* [a] vs. *ă* [ɐ]) and a horn (*o* [ɔ] vs. *ơ* [ɤ]; *u* [u] vs. *ư* [ɯ]) are used for this.

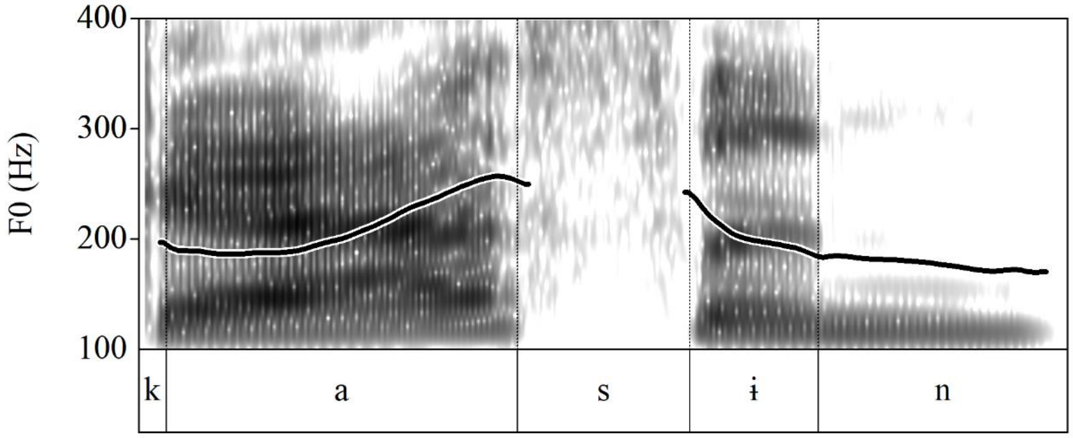


Figure 1: Spectrogram with F0 contour for the non-word cá xin *(rising* sắc*, neutral* ngang*) stimulus used in the present study, produced by a male speaker.*

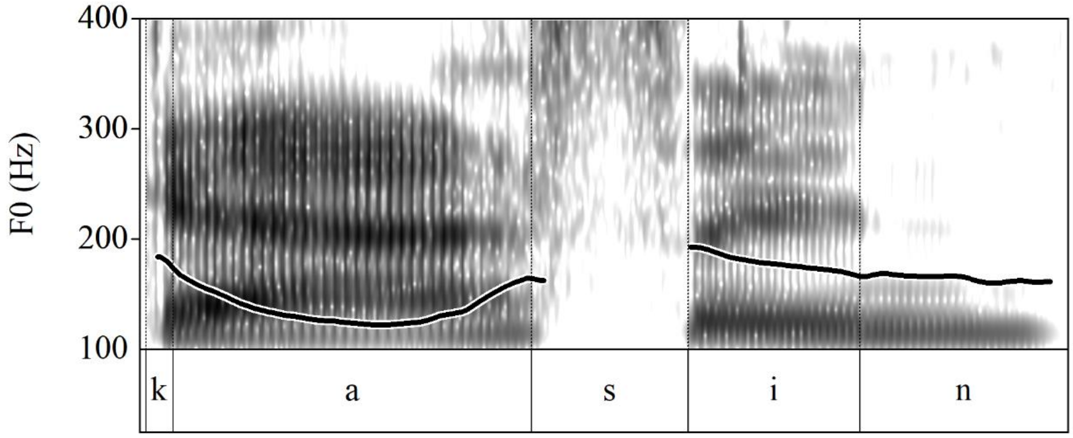


Figure 2: Spectrogram with F0 contour for the non-word cạ xin *(falling* nặng*, neutral* ngang*) stimulus used in the present study, produced by a male speaker.*

Note that Brunelle (2009) argues it is better to refer to Vietnamese tones using Michaud's (2004) alphanumerical labels, as referring to tones just by the shape of their contour (e.g. 'rising' and 'falling' used in the present study) can create confusion with contours differing in shape across dialects. However, these labels seem just as difficult to remember as are Vietnamese tone names, hence in this thesis, from this point on, the tones will always be referred to as [contour shape] [Vietnamese tone name]. This way, readers who are unfamiliar with Vietnamese tones are given an indication of what the tones may look like (while being advised to remember that tone contours may vary across dialects), while those more familiar will know which specific tone is being referred to regardless of dialect.

Southern Vietnamese is the dialect of the participants in the present study and is spoken in Hồ Chí Minh City and other Southern regions in Vietnam. In Southern Vietnamese, the dipping *hỏi* and *ngã* tones are not contrastive and are both pronounced as dipping *hỏi*. This results in Southern Vietnamese having five tones, as opposed to the standard dialect of Northern Vietnamese, which has six (Thompson 1965; Brunelle & Jannedy 2013). In the literature, Vietnamese tones are generally classified by pitch register, i.e. high or low, and by shape, i.e. level, rising, falling, or dipping. A classification of (Northern) Vietnamese tones adapted from Pham (2003) is provided in Table 1:

	'Even' (level) tones	**'Non-even' (contour) tones**	
		rise/fall	*curve*
High/unmarked	ngang	sắc	hỏi
Low/marked	huyền	nặng	ngã

Table 1: (Northern) Vietnamese tones classified by height and shape. Note that in Southern Vietnamese, hỏi and ngã are both pronounced as hỏi. Adapted from Pham (2003: 23, 31).

In recent experimental studies, the perception of Vietnamese tones by various groups of speakers has been investigated and it has been shown that different cues are relevant for different dialects. Brunelle (2009) had Northern and Southern Vietnamese listeners identify tones in Northern speech as well as in resynthesised stimuli. Additionally, participants were asked to rate the quality of each tone compared to tones in their respective dialects. Regarding the natural and duration-resynthesised stimuli, all listeners thought most of them sounded natural, although Southern listeners gave worse ratings for the falling *nặng* and dipping *ngã* tones. Southern participants were also less accurate than Northern participants in identifying the Northern tones: dipping *hỏi* was often confused with falling *nặng*, *nặng* was sometimes identified as low-level *huyền,* and dipping *ngã* was often perceived as rising *sắc.* Brunelle noted that these Northern Vietnamese tones are confusable for Southern Vietnamese listeners, because in Northern Vietnamese they are distinguished through different phonation types, i.e. glottalisation and laryngealisation, which are not relevant cues in Southern Vietnamese. For the resynthesised stimuli, Southern listeners relied considerably less on phonation type than Northern listeners. In general, they judged rising contours as rising *sắc*, low contours were judged as low level *huyền,* and other contours as mid-high level *ngang*. However, when there was laryngealisation or glottalisation, lower contours were judged as falling *nặng* instead of low level *huyền.* Brunelle noted that this contrasts with other works on tone perception in Southern Vietnamese, where it is shown that perception in this dialect is independent of voice quality. He explains that this is likely due to the Southerners' familiarity with Northern Vietnamese (e.g. through media) and an awareness of the role phonation types play in this variety. Lastly, Southern listeners mostly identified more complex resynthesised stimuli (with a dipping contour) as dipping *ngã* and *hỏi* tones, but glottalised stimuli were again judged as falling *nặng* tones. Through further analysis, Brunelle showed that for Southern Vietnamese listeners, pitch direction as well as contour are the most important perceptual cues.

Kirby (2010) argued that Brunelle's (2009) use of an identification task allowed listeners to think too much before responding, which makes it difficult to understand how listeners perceive tone in a pre-linguistic mode of processing. Kirby's solution was the use of a speeded AX discrimination task. Northern and Southern Vietnamese participants heard pairs of words with Northern Vietnamese tones and had to decide whether they were the same. Southern Vietnamese speakers were less accurate and slower when distinguishing falling *nặng* and dipping *hỏi* tones, as well as the two dipping *hỏi* and *ngã* tones. Kirby argued that these tones are distinguished using voice quality in Northern Vietnamese and Southern listeners might be aware of this cue because of familiarity with Northern Vietnamese, but they may not necessarily associate it with a particular tone. Hence, the Southern listeners confused the tones that make use of phonation type as a distinguishing cue. Yet, overall participants had low error rates, regardless of dialect.

Pham et al. (2018) conducted a non-word repetition (NWR) task with children bilingual in English and Vietnamese in California. These children were in kindergarten or in their first or second year of school (5;8 to 8;6 years old), speaking Vietnamese at home and English at school. Unfortunately, Pham et al. (2018) do not provide enough information to establish whether the children were HSs with a profile similar to that of the HSs in the present study. It is also unclear which Vietnamese variety the children spoke. Regardless, the study provides some interesting results. In NWR tasks, participants are presented with auditory stimuli of non-words obeying the phonotactics of the language under investigation and are asked to repeat them. The way in which the participants repeat the stimuli can give indications of their perception skills and the way they conceptualise phonological categories within the target language.

Pham et al.'s (2018) results showed that, overall, tone production was most accurate and remained fairly constant over age, whereas segment production improved over time. This stronger sensitivity to and production of tones conforms to some of the previous literature (Burnham et al. 2011; Singh et al. 2015) but contrasts in that other results seem to show a lesser sensitivity to tone in general compared to segments (Burnham et al. 2011; Wewalaarachchi et al. 2017) or a tendency for tone sensitivity to decrease with age, with segments becoming the easiest to process instead (Singh et al. 2015; Ma et al. 2017). Pham et al. (2018) did note, however, that no detailed acoustic analyses were performed and that only a subset of Vietnamese tones were used. Moreover, the same tone was used successively until non-words of another length were played, which also carried different tones. The children may thus have had an advantage in hearing the same tone repeatedly.

Lastly, Nguyen & Macken (2008) investigated tone production in L1 American English L2 learners of Northern Vietnamese. They noted that English learners are usually found to solely rely on pitch levels to produce Vietnamese tones and ignore other cues such as duration, glottalisation, and intensity. In the study, two beginner learners, two intermediate learners, and two advanced learners were tested in free conversation and picture description tasks. The researchers mentioned that, in a preliminary test, their participants were generally able to accurately identify tones, which is necessary for the speakers to accurately produce them. There was considerable variation in the data, but for speakers across all levels, there were difficulties in correctly producing dipping *hỏi* and rising *sắc* tones in non-emphasised syllables. The researchers suggested that dipping *hỏi* might have been difficult to produce correctly, because the learners were not aware of its reduced form in non-emphasised speech, where it does not rise after falling. In emphasised syllables, dipping *hỏi* was usually produced quite accurately. Five out of six participants were found to make errors in producing rising *sắc* when this tone was preceded by dipping *hỏi*, which caused the rising *sắc* to become falling or generally lower than its usual mid-high rising form. Overall, dipping *ngã* was pronounced most accurately. Nguyen & Macken suggested that this is because of the tone's glottalisation marking the transition from a falling to rising contour. So far, no studies seem to focus specifically on learners perceiving Vietnamese tones instead of producing them, therefore it is difficult to estimate how the Dutch listeners in the present study might perceive Vietnamese tones.

Overall, the studies in this section suggest that for Southern Vietnamese, the dialect under investigation in the present study, the primary perceptual cues for tone are pitch height and contour. In fact, when Southern Vietnamese speakers have to rely on glottalisation or laryngealisation in Northern speech, they have significant difficulties telling tones apart (Brunelle 2009; Kirby 2010; cf. Pham 2001; Brunelle & Jannedy 2013). L2 learners of Vietnamese have been found to generally have issues producing dipping *hỏi* and rising *sắc* tones, although their perception of Vietnamese tone seems to be fairly accurate after as little as 7 weeks of learning the language (Nguyen & Macken 2008). The two tones used in the present study, rising *sắc* and falling *nặng*, usually differ in pitch height and contour and are thus expected to be easy to distinguish for Southern Vietnamese listeners in the present study, but may cause more problems for the untrained Dutch listeners. Pham et al. (2018) showed that children acquiring Vietnamese and English may be able to accurately produce Vietnamese tones starting around age six, but not much is known about Vietnamese children's tones before this age. Since the heritage speakers in the present study became dominant in Dutch before age 6, it could be suspected that, despite rising *sắc* and falling *nặng* being fairly distinctive, this early dominance in a non-tonal language affects the heritage speakers' accuracy in tone perception. This would provide new evidence on tone development in speakers bilingual in Vietnamese and a non-tonal language (see Sections 2.1 and 2.3.1).

2.5. Tone in Dutch

In most dialects of Dutch, there is no use of lexical tone. Instead of a tonal language, Dutch is an intonational language, where pitch only changes meaning at the syntactic (e.g. questions) or pragmatic (e.g. sarcasm, excitement) level. As for other speakers of non-tonal languages discussed in Section 2.2, Dutch L1 speakers may have difficulties in perceiving and producing tones. For instance, Zou et al. (2017) showed that Dutch listeners without experience with a tonal language recognise words based mostly on segments, whereas in speakers of a tonal language like Mandarin, words are recognised based on both segments and tone (see Section 2.3.2).

However, for some dialects of Dutch in the province of Limburg, it is argued that there are lexical tones. Many of the attested tonal minimal pairs are pairs of singular and plural nouns (3), but there are also pairs that have completely different meanings (4). There is a two-tone contrast, in which the absence an H-tone constitutes one category ('Accent 1') and its presence constitutes another ('Accent 2') (Gussenhoven 2000; Fournier 2008).

(3) *knien*^{I} 'rabbits'
knien^{II} 'rabbit'

(4) *haas*^{I} 'hare'
haas^{II} 'glove'
(examples adapted from Fournier 2008: 20)

Gussenhoven (2000) argued that tone in Limburgian and other nearby Franconian dialects developed because a morphological distinction of plurality would otherwise be lost. Before tone was used (as in example 3), this distinction was made using vowel length. However, the vowel length distinction was neutralised over time and to maintain a difference in plurality, tone emerged. For an alternative hypothesis of tonogenesis in Limburgian, see Boersma (2018), who furthermore countered Gussenhoven's explanation. Regardless, a detailed discussion Limburgian tonogenesis is beyond the scope of the present study and both accounts agree that Limburgian has tone.

Limburgian tones' phonetic realisations may differ across dialects within the continuum (Gussenhoven & Peters 2008). Fournier (2008) found that Limburgian lexical tone in Roermond and Venlo dialects appears in specific prosodic contexts and that, in these contexts, speakers of the dialects are able to distinguish tonal contrasts as accurately as they distinguish segmental contrasts. However, she also noted that the tonal contrast seems to be reduced in younger speakers and that it might eventually disappear from the dialects.

Köhnlein (2016) argued that tone in Franconian dialects such as Limburgian is not, in fact, lexical tone. Rather, he described it as an effect caused by contrastive foot structure. Regardless of whether tones exist in Limburgian dialects, they are not reported in other dialects of Dutch. The present study will thus avoid speakers of Limburgian dialects to represent Dutch speakers without experience with a tonal language.

3. HERITAGE SPEAKERS

Heritage speakers (HSs) form a special group of language users in linguistic research. At home, HSs acquire a minority language of the society they live in as an L1, which is called their heritage language (HL). But before the HL is fully acquired, the HSs switch to a different language, the one that is dominant in society (Polinsky & Kagan 2007: 369-370). HSs are worth investigating because they constitute a group that provides insights on language acquisition as well as language change due to contact in individuals and communities. Moreover, these speakers are sometimes caught somewhere in between the 'nativeness' and, consequently, also in between the identities associated with each of their respective languages, which may lead to negative experiences (Lam 2006). Research on HSs can thus not only benefit linguistic theories of acquisition and language contact, it can also benefit HSs themselves through creating more understanding in language education and policymaking, for instance.

HSs are often compared to either speakers from the homeland or to (older) speakers from the same community as the HSs but who are monolingual in the HL or did not learn a different language until adulthood. There is a crucial difference between these two groups: the former, called 'homeland speakers' (HMs), speak a variety of the HL that usually is not influenced by a different dominant language. Moreover, this variety is not necessarily a variety the HSs are familiar with: Southern Vietnamese HSs in the U.S. may for instance not be familiar with the Northern Vietnamese variety that is the standard language many HMs in Vietnam are familiar with, even if these HMs speak a different dialect themselves (Polinsky & Kagan 2007).

Speakers in the other group that HSs are compared to, form the baseline for HSs (hence they are called 'baseline speakers': BSs): they are the parents of HSs or are part of the community in some other way and often provide the only input the HSs base their HL on. The baseline does not necessarily correspond to the homeland variety, but it is the standard most HSs are acquiring before switching to the dominant language (Polinsky & Kagan 2007: 372). In the present study, Southern Vietnamese HSs in the Netherlands are compared to Southern Vietnamese HMs and not BSs, because most BSs of Vietnamese in the Netherlands seem to be bilingual and quite proficient in Dutch after living in the Netherlands for decades, which could have led to L1 attrition (cf. Schmid 2013, who discussed the unclear effects of various extralinguistic variables on L1 attrition). A comparison to the homeland thus seems more appropriate (see Sections 3.3 and 4.2.1).

Note that in the literature, both HMs and BSs are often referred to simply as native speakers, in contrast with HSs. This terminology has problematic implications: first of all, it implies that HSs are not native speakers. However, if a native speaker is defined simply as an L1 speaker, they are. Additionally, there is the 'native speaker fallacy': it is not always clear what a native speaker is and when this notion is even useful (Faez 2011; Rothman & Treffers-Daller 2014 focused specifically on HSs being native speakers). Because of the above reasons, in this study these groups will be referred to as HMs and BSs (or simply monolinguals, when applicable).

Some of the characteristics of HSs and HLs are discussed in the sections below. In Section 3.1, general characteristics of HLs are briefly discussed and in Section 3.2 there is a focus on the phonology and phonetics of HLs. In Section 3.3, a brief impression of Vietnamese HSs is provided, as this is the population of interest in the present study.

3.1. General characteristics of heritage languages

One of the most characteristic traits of HS groups is the variation in proficiency found within them. Polinsky & Kagan (2007) discussed a continuum similar to that posed for creole languages (Bailey 1973, 1974; Bickerton 1973, 1975). Basilectal HSs show radical changes compared to BSs and have very low proficiency, whereas mesolectal HSs speak more similarly to BSs, and acrolectal speakers are as close to BSs as possible (Polinsky & Kagan 2007: 371). Factors that influence HSs' proficiency in the HL include the similarity of the dominant language to the HL and age of acquisition (AoA) of the dominant language. Parents have some influence on HSs' proficiency through the type and amount of input the HSs get in the HL: e.g. whether they used to live in a region where the HL was the dominant language or not, whether the parents switch to the dominant language at home as soon as the HSs do, whether the parents allow codeswitching at home, and what kind of attitudes the parents display towards the HL and their culture. Similarly, the HSs' attitude towards and general involvement with the community in which the HL is spoken also has a strong influence on their proficiency (Polinsky & Kagan 2007; Chang 2016). Parents and other community members may furthermore influence HSs' proficiency through the variability in the input: when HSs are exposed to a larger number of different varieties of the HL, they are likely to have a better understanding of variation in their HL (Polinsky & Kagan 2007; Montrul 2018). When reading about the characteristics reported below, the reader is advised to keep in mind that HS groups may be quite heterogeneous and that the characteristics described may be present to varying degrees (i.e. more present in basilectal HSs and less so in acrolectal HSs).

Polinsky & Kagan (2007) discussed studies that showed various common observations in HS. They reported that many HSs have a decreased speech rate in the HL compared to monolinguals. They often have a lower lexical proficiency and in many cases the vocabulary may be limited to language used within the household. A smaller vocabulary has moreover been found to be correlated to weaknesses in morphosyntax. These weaknesses are the most noticeable in HLs. HSs may over-regularise and simplify morphological inflectional systems such as case, gender, verb agreement, or mood. In general, their nominal morphology is weaker than that of verbs. Passives and relative clauses are also difficult for HSs to interpret or produce, especially when they require overt case marking. Issues with morphology are most clearly the case for irregular or infrequent forms. Infrequent forms may not only be simplified but also fossilised (Polinsky & Kagan 2007; Montrul 2018).

As a result of reduced morphological paradigms, HSs often also have to rely on rigid word order to account for the meaningful relationships between words lost through the reduced use of morphology. Other syntactic strategies that prove to be difficult are pro-drop and similar dependencies, such as the use of reflexives. Still, syntax seems to be easier for HSs than morphology (Montrul 2018).

3.2. Phonology and phonetics of heritage languages

Considering the above tendencies and the way many HSs are exposed to their languages, it is often argued that HSs generally perform best in aural comprehension (Polinsky & Kagan 2007; Montrul 2016). Researchers even used to go so far as to argue that HSs have a '(near-)native' phonology, in both perception and production. Consequently, less research has been done on HL phonology than on HL morphology or syntax (Polinsky & Kagan 2007: 378; Lukyanchenko & Gor 2011: 415; Kim 2015: 107; Montrul 2016: 82). However, over the past decades, it has been shown that HSs may actually differ substantially from HMs and BSs in phonology as well. Although HSs may be closer to these groups than L2 speakers of their HL, they are nevertheless often found to form a separate group. This holds for acoustic measures but also for more global impressions by both researchers and monolingual BSs or HMs listening to the HSs (Polinsky & Kagan 2007; Montrul 2018). HL phonology may be the part of the grammar that is assumed to be retained best compared to the baseline, but it is not immune to changes from the baseline (Benmamoun et al. 2013; Montrul 2016, 2018). This will become evident from discussions on HLs' segmental and suprasegmental phonologies in Sections 3.2.1 and 3.2.2, respectively.

3.2.1. Segmental information

Most studies on HL phonology have focused on HSs' production and perception of segments, usually compared to BSs or HMs, and L2 learners. For instance, a series of studies by Au, Knightly, Jun, and Oh investigated the production and perception by HSs of Korean and Spanish in the U.S. (Au et al. 2002; Knightly et al. 2003; Oh et al. 2003). Au et al. (2002) found that HSs who overheard Spanish during childhood produced Spanish voiced and voiceless stops more accurately than L2 learners. Moreover, their accents were rated better by Spanish monolinguals. It is not clear, however, how or whether these HSs differed from the study's Spanish monolingual controls. Knightly et al. (2003) subsequently found similar results with a HS group of overhearers of Spanish. Moreover, they found that the overhearers produced significantly different voiced consonants in Spanish compared to monolingual controls, at times even patterning with the L2 speakers.

For childhood overhearers and HSs of Korean, Oh et al. (2003) found that these two groups of HSs did not differ significantly from monolingual controls in perceiving a three-way VOT contrast in Korean stop consonants, whereas L2 learners did differ from the other groups. In production, however, both L2 learners and childhood overhearers were found to perform less accurately than the other two groups. Additionally, in accent ratings by a group of monolingual Korean speakers, HSs were rated lower than monolingual controls but better than L2 learners. A tendency for HSs to perform in the middle between L2 learners and monolinguals (BSs and HMs) is thus suggested by these studies. More recent studies have put forth similar findings, comparing monolinguals, L2 learners (or monolinguals of the dominant language), and HSs of Russian (Lukyanchenko & Gor 2011) and Mexican Spanish (Shea 2017). Shea (2017) included variables related to dominance and proficiency in her analysis, such as the AoA of English, use of Spanish inside and outside of the home, fluency, and vocabulary. She found that higher dominance and proficiency in the Spanish HL in HSs correlate to stronger similarity to HMs and a stronger difference from English L2 speakers in Spanish vowel production.

Chang (2016) found that HSs of Korean may perceive unreleased stops in Korean and English equally or even more accurately than monolinguals of either language. In Korean, HSs were found to perform the same as Korean HMs and both performed better than English monolinguals. In English, the HSs outperformed the HMs who, in turn, outperformed the English monolinguals. Chang's (2016) results suggest that not only may HSs perform similarly to HMs in the HL but also similarly to or better than monolinguals of their dominant language, due to positive transfer from the HL.

To sum up, the above studies on segmental phonology and phonetics in HLs suggest that at least within the HL, HSs often have an advantage over L2 learners in producing and perceiving segmental contrasts (Au et al. 2002; Knightly et al. 2003; Oh et al. 2003; Lukyanchenko & Gor 2011; Shea 2017). Sometimes the HSs perform similarly to HMs or BSs of their HL (Lukyanchenko & Gor 2011; Chang 2016) and in cases of positive transfer, they may even outperform both HMs and monolinguals of the dominant language in various circumstances (Chang 2016). Still, within the HL, HSs may also differ significantly from HMs or BSs, and in some circumstances pattern more closely with L2 learners (childhood overhearers in Knightly et al. 2003; Oh et al. 2003). HSs may thus show changes in phonology compared to HMs or BSs, in contrast to the similarities that used to be expected.

3.2.2. Suprasegmental information

Within the relatively limited number of studies on HL phonology, studies on HL prosody, including lexical tone, form an even smaller group. This is odd because, for instance in Stangen et al. (2015), global foreign accent in HSs of Turkish in Germany was mentioned by monolinguals of German and Turkish to be due mainly to prosody (in 26.8% and 28% of cases according to German and Turkish raters, respectively). Studies on HL prosody therefore seem highly relevant.

Previous research reveals that HSs may use intonation contours from both the HL and the dominant language regardless of which language they are speaking, and redistribute the meanings of different contours (Turkish-German HSs in Queen 2001). This mixing of intonation contours across languages was also found for HSs of Icelandic in Canada (Dehé 2018). These HSs produced both polar question contours used by HMs, and contours used by monolinguals of English, irrespective of which language they were speaking. Robles-Puente (2014) found that Mexican Spanish HSs in the U.S. patterned more like L1 English speakers than like Spanish monolinguals, showing more variability in vowel length. The HSs also had intonational contours intermediate between the two other groups. Kim (2015) also found that Spanish HSs in the U.S. produce lexical stress intermediate between HMs and L2 speakers, with the main giveaway being vowel length, as in Robles-Puente's findings. Yet, HSs may also pattern more with HMs than with speakers of the dominant language, for instance in producing boundary tones and vocatives (Robles-Puente 2014), and in some cases, lexical stress (Kim 2015). In perception as well, HSs may judge intonational contours (Mexican-Spanish HSs in the U.S. in Hoot 2012) and lexical stress in similar ways as HMs (Kim 2015).

The number of studies focusing specifically on lexical tone in HSs seems very limited. Up until the writing of the present study, only So (2000), Đào (2013), Yang (2015), Chang & Yao (2016), Soo & Monahan (2017), Lam (2018), and Kan & Schmid (2019) seem to have dealt with this particular subject and most focus exclusively on Chinese heritage languages. Đào (2013) focuses on Vietnamese and will therefore be discussed in Section 3.3.

So (2000) conducted production and perception experiments with three groups of Cantonese bilinguals in Canada: English-dominant HSs in Cantonese-speaking families, Cantonese-dominant speakers who moved from Hong Kong to Canada as teenagers (referred to here as late HSs), and Cantonese-dominant speakers who moved to Canada as adults (HMs). In the production task, the participants read out carrier syllables with all six Cantonese tones in isolation and in carrier phrases. The isolated productions were analysed for F0, whereas the productions in context were checked for duration. In the F0 analysis, So found that the HSs' tonal space (i.e. the range of F0 used in tone production) was smaller than that of the HMs, whereas the group of late HSs did not differ from the HMs. Consequently, the HSs' level and contour tones were less distinct than and sometimes different in shape from the HMs'. For the late HSs, only Tone 5 (a rising tone) was less distinct than among the HMs and also differed in shape. In the durational analysis, no significant differences between the three groups were found.

In the perception task, participants were instructed to identify the six Cantonese tones in the same carrier syllables. HSs were less accurate than late HSs and HMs, whereas the late HSs only differed from the HMs in identifying Tone 5. Both HS groups found tones that were similar in shape most difficult to identify correctly. So (2000) suggested that the HSs and late HSs thus have difficulties both producing and perceiving tone in the HL, compared to HMs.

Yang (2015) compared English-dominant HSs of Mandarin in America to L2 learners and Mandarin HMs studying in America. Both the L2 learners and the HSs took Mandarin courses at university level. In a perception task, Yang investigated the acoustic cues of the starting and end points of Mandarin tone as perceived by L2 learners and HSs. The participants listened to sentences in which a target word *tao* was embedded in different tonal contexts and was therefore subject to tone sandhi. *Tao* itself was assigned 81 different synthesised tones. The participants had to decide whether a synthesised *tao* tone was T1 (high level), T2 (low rising), T3 (dipping), or T4 (high falling).

HSs showed more agreement with each other and had more stable areas for each tone than L2 learners, who showed much more variation. The HSs' categorisation of tones was also more similar to the HMs'. However, they were generally less certain than HMs about which tones fall within one category and which fall within another. This is not only shown through the fact that HSs show less agreement with each other than HMs, but also because the same synthesised tone was sometimes categorised into up to three different categories. Yang suggested that for the identification of tones, HSs rely mostly on tones' register and not so much on their contour (this could correspond to Đào's (2013) findings for production by Vietnamese HSs in Australia, see Section 3.3). Moreover, the pitch range where HSs were able to accurately recognise tones was smaller than the range HMs could accurately perceive. For instance, it was easier for the HSs to identify tones with a high-pitch ending than a low-pitch ending. Lastly, HSs differed from HMs in that their perception of target tones was not strongly influenced by the surrounding tonal context (i.e. tone sandhi). This is similar to the L2 learners' perception, who showed even less influence of tonal context in perception.

In the production task, the participants read out a paragraph in Mandarin and Yang analysed the tones produced in target syllables. Regarding the F0 at onset and offset of tones, HSs were found to show considerable overlap between T1 and T4 on one hand, and T2 and T3 on the other hand. T2 had a lot of overlap in general, leading Yang to conclude that it has almost no stable area in the HSs' phonology. The HSs' made the clearest distinctions between tones at the tone onsets. Compared to L2 learners, HSs performed much closer to HMs and they had a larger pitch range to produce tones. Yet, Yang's results clearly showed that the HSs differ from HMs in production and perception.

Chang & Yao (2016) investigated the F0 contour, range, duration, turning points, and variability in the production of tones by Mandarin HSs, HMs, and L2 learners of Mandarin in the U.S. To account for the high level of variability in experience with the HL within the HS group, the HSs were subdivided into high-exposure (HE) and low-exposure (LE) groups. HMs and HE speakers were found to produce shorter high level T1s in monosyllabic words than the other groups and HSs overall produced shorter dipping T3s than L2 speakers as well. In contrast to Yang (2015), it was not found that L2 speakers or HSs had a narrower F0 range than HMs. Regarding the timing of turning points in low rising T2 and dipping T3, HSs and HMs were found to perform similarly, whereas L2 learners' turning points in T3 were consistently earlier than usual. Still, there was considerable variation among the HSs in T3 turning points compared to HMs.

In connected speech, HSs were found to shorten tones more than L2 speakers but less than HMs. Similarly, L2 speakers were the least likely to reduce T3 in non-phrase-final contexts, where reduction is expected. HSs were more likely to do this, with the HE subgroup even reducing T3 marginally more often than HMs. Chang & Yao argued that this may be because HSs generally do not hear Mandarin words in isolation, where T3 is not reduced. After all, they do not necessarily have formal education in the language. The L2 speakers and HMs did enjoy formal education in Mandarin and were thus more familiar with unreduced T3 in emphasised classroom speech.

The production data were rated by Mandarin HM listeners on intelligibility and additionally the listeners had to identify tokens as being produced by HMs, HSs, or L2 learners. The HMs' isolated tones were identified more accurately and faster than those produced by HSs and L2 speakers. There were no substantial differences between the latter two groups. When tones were identified accurately, the listeners had to rate the tones' "goodness". Results showed that HMs had the best tones, followed by HSs' tones which were, in turn, better than L2 speakers' tones. In connected speech, HSs' tones were identified more accurately (but not faster) than those produced by L2 speakers. Moreover, they were not less accurately identified than HMs' tones. Within the HS group, HE speakers' tones were slightly easier to identify than LE speakers' tones and the former were also perceived as better tones.

When the listeners had to identify tokens as produced by HMs, HSs, or L2 speakers, the HSs' tokens were consistently the most difficult to classify and listeners were generally less confident about their decision about HSs' identity than the identities of HMs or L2 speakers. Among HSs, HE speakers were more difficult to classify than LE speakers.

Soo & Monahan (2017), like So (2000), investigated tone in Cantonese-Canadian HSs, using an AX discrimination task. In this task, the syllable /ji/ was played twice with a match or mismatch in tone. All six Cantonese tones were used. It was found that the HSs did not differ from an HM control group; both groups were better at telling apart tones that were dissimilar than tones that were similar.

Soo & Monahan argued that this similarity between HSs and HMs may have been due to short intervals between stimuli, which could have encouraged participants to listen on a phonetic level instead of a phonological level. Therefore, in a second experiment, Soo & Monahan used medium-distance repetition priming (MDRP) during a lexical decision task to make sure the results reflected the participants' phonology. CV syllables were either followed by an identical syllable, or followed by a syllable differing in consonant, vowel, or tone, but the repeated word or deviating word only appeared 8-20 trials later. If this latter (non-)word shows priming because of the former, then the words must be similar in the participants' phonology. HSs were found to show fewer priming effects for the identical word pairs than HMs, suggesting that the former do not retain phonological information as well as the latter. Still, for the pairs differing in tone, HSs were primed more than HMs, indicating that the HSs are not disturbed as much by a tonal mismatch as HMs and that they retain segmental information better than tonal information. In addition to the discrimination task, participants were asked to produce all Cantonese tones on different vowels. Even after acoustical analysis, Soo & Monahan found that there were no considerable differences between HSs' and HMs' productions of tones.

Lam (2018) also tested Cantonese-Canadian HSs' perception of tones in their HL compared to that of HMs. The participants had to identify which tone they heard in Cantonese words (isolated as well as in context) that were natural recordings or edited to be rid of segmental or of tonal information. In general, HMs were found to identify tones more accurately. Especially when segmental information was filtered out, HSs were less accurate in identifying the words they heard, i.e. they found it difficult to identify words using just tones. When words were offered in a semantic context, however, the groups were more similar. When tonal information was neutralised, the two groups were very comparable as well. Lam argued that these results show that HSs and HMs rely on tone to different extents: both can rely on just segments, but HMs feel more comfortable than HSs relying on just tone. For both groups, the same tone pairs were confusable, meaning that although HSs made more errors than HMs, they did make *similar* errors. Lastly, targets that were presented in a context that was semantically unlikely (with one of the distractor words depicted being more semantically likely), HSs were found to choose distractors more often than HMs. This means that HSs sometimes relied more on semantic information than tonal information than did HMs, which, Lam argued, shows that the two groups attend to different types of information in listening to Cantonese.

Kan & Schmid (2019) did an ABX task with Cantonese HSs (aged 5-12) and HM controls. They investigated whether the HSs accurately acquire two Cantonese tone pairs that could be mapped to intonational categories in English (the dominant language). Cantonese high level Tone 1 was expected to be mapped to English flat pitch, whereas low falling Tone 4 would be more similar to a statement intonation. The other pair of tones, mid rising Tone 2 and low rising Tone 5, are more similar in shape and could both be mapped to English question intonation. Due to the similarity between the tones in the second pair, Kan & Schmid expected HSs to be less accurate in distinguishing them than Tones 1 and 4. Tones 1 and 4 were expected to be quite discriminable, just like the segmental control pairs that were included. Both groups had members that performed 100% accurately regardless of tone pair, but there was more variation among the HSs. Overall, both groups were better able to tell apart Tones 1 and 4 than Tones 2 and 5. The HSs were usually less accurate than the HMs in all conditions (similar tones, distinct tones, and segmental controls). Among HSs, participants were most accurate in the segmental controls, followed by the distinct tone pairs and lastly by the similar pairs. Kan & Schmid suggested that the HSs may have become less sensitive to pitch due to their dominance in English. For the segmental control condition, it was suggested that the HSs were less accurate because of their lower proficiency in Cantonese compared to the HMs.

The results from the presented studies again suggest that in many cases, HSs may perceive or produce prosodic cues differently from both L2 learners on the one hand, and HMs and BSs on the other. Rather, HSs show an intermediate pattern (Yang 2015; Chang & Yao 2016; Soo & Monahan 2017; Lam 2018; Kan & Schmid 2019). The few studies on tone in HSs show that in production, HSs may have a reduced pitch range compared to HMs and tones may be different in shape, less distinct, and less stable (So 2000; Yang 2015). HSs' tones may show more variability and, although they are perceived as more accurate than L2 speakers' tones by monolingual judges, they are sometimes difficult to identify (Chang & Yao 2016). Yet, sometimes HSs' and HMs' tone productions may seem similar (Soo & Monahan 2017). In perception, HSs may show more variability than HMs in recognising tones (Yang 2015; Kan & Schmid 2019) and may be found to have a reduced sensitivity to tone (Soo & Monahan 2017; Kan & Schmid 2019). None of these studies focused specifically on segment-tone integration, but nevertheless they provide clues about the distribution of attention between segments and tones among HSs. HSs may hold on to segmental information more than to tonal information in word representation, compared to HMs (Soo & Monahan 2017). Likewise, HSs may be more dependent on segmental information than HMs, experiencing difficulties identifying words based solely on tonal information but having fewer problems identifying words based solely on segmental information (Lam 2018). Considering the different ways HSs and HMs pay attention to the two dimensions in these studies, it is worthwhile to investigate further not only how HSs distribute attention between segments and tones, but also how *integrated* these dimensions are in HSs' word processing. This is how the current study aims to contribute.

3.3. Vietnamese heritage speakers

After the Vietnam War, many Vietnamese, especially Vietnamese from the Southern regions, fled their country. This resulted in a diaspora not only in nearby Asian countries but also countries further away such as the U.S., France, and the Netherlands. Originally, an estimated 6,200 Vietnamese refugees came to the Netherlands directly following the war (Van Der Hoeven & De Kort 1983). 88% of these refugees came from the Southern regions. It is estimated that, as of January 1 2018, there are 22,741 people with a Vietnamese background living in the Netherlands. Of these people, 9,071 (40%) are estimated to be part of the second generation (the first generation that is likely to consist of mostly HSs). A majority of 7,023 (77%) of the second-generation Vietnamese are reported to have two Vietnamese parents (*Centraal Bureau voor de Statistiek* 2018).

Vietnamese are among the least densely populated immigrant communities in the Netherlands, possibly due to an initial distrust in fellow refugees and due to efforts from the Dutch government to help Vietnamese integrate into Dutch society (Kleinen 1988; De Valk et al. 2001). However, in previous research, I found that nowadays first-generation speakers (i.e. the Vietnamese who came to Vietnam after the war) usually report that they feel like they belong to a Vietnamese community within the Netherlands and that they have Vietnamese friends who they see regularly. There are Vietnamese Buddhist temples and cultural associations within the Netherlands. These provide a place for community members to host events and also to maintain their culture outside of the homeland through, for instance, language classes for later generations (Ebenau 2017). More information on the HSs in the present study is provided in Section 4.2.1.

Lam (2006) reported that most American-born Vietnamese HSs understand very little Vietnamese. Many experience some formal instruction in the language within local communities but do not keep up with it when they get older. Due to their limited proficiency, many Vietnamese HSs in the U.S. feel insecure about their HL and are reluctant to speak it. Lam described the conflict for these speakers between assuming an American identity and doing away with 'foreignness' on the one hand, and maintaining their Vietnamese identity on the other hand. Within the community, HSs may be considered by others as 'too white', whereas outside of the community, they are still not considered 'fully' American.

This picture also holds to an extent for Vietnamese HSs in the Netherlands. In most cases, the HSs learn Vietnamese as an L1. Around age 4, they start going to school and start learning Dutch. Within the home, families seem to vary in whether they only speak Vietnamese or a mix of Dutch and Vietnamese. One HS in Ebenau (2017) reported that he and his family had not spoken Vietnamese with each other for over ten years, but this seemed to be an exceptional situation. In a number of cases, Vietnamese HSs in the Netherlands are the only ones in their families who speak Dutch and consequently have to interpret for their parents from an early age on. However, there are also numerous first-generation baseline speakers who are quite proficient in Dutch and use it daily in both personal and professional contexts (Kleinen 1998; Ebenau 2017).

Parents of HSs in the Netherlands do not seem to report harsh opinions of their children or other HSs the way Lam (2006) reports, but they do sometimes mention that they fear their children get out of touch with their Vietnamese identity and that they do not speak (enough) Vietnamese. Some parents argue that this is due to Vietnamese youth in the Netherlands not interacting with each other (Ebenau 2017).

There do not seem to be many studies on Vietnamese HSs' proficiency in their HL. Only one study seems to have focused specifically on tone in Vietnamese HSs. Đào (2013) conducted a study with four groups: Vietnamese BSs who came to Australia after age 20, Vietnamese HSs who were born in Australia, and Vietnamese HMs living in Hồ Chí Minh City and Cần Thơ in Vietnam. The last group consisted of an older and younger subgroup, with the younger group all living in Hồ Chí Minh City. The groups all spoke Southern Vietnamese. Participants were asked to read out syllables carrying Southern Vietnamese tones imbedded in a carrier sentence. HSs often made errors in tone production, with tones belonging to the same register or pitch range being particularly difficult (Đào 2013: 28-29). Furthermore, only the HSs showed deviating patterns in contour shapes, whereas the other groups matched descriptions in the literature. For instance, the dipping tone (*hỏi/ngã*) became merely rising in HSs, whereas in the corresponding young HM group it was clearly dipping. Additionally, HSs' tones were different in length compared to all other groups. Both HSs and BSs in Australia were found to have wider F0 ranges than the groups in the homeland.

Arguing that his results could be explained by a loss of distinctions in the HL, Đào stressed that language teachers should pay extra attention when it comes to teaching HSs about tone in the HL. However, one should note that these productions are most of all evidence of difference in phonetics and a perception task should reveal more about the way Vietnamese HSs categorise and characterise tones in their phonologies. The present study cannot provide such evidence, as it only tests two tones in the HSs' phonologies, but it should give a more direct view into what role tone has in their HL, especially in relation to segments. For a direct comparison with Đào (2013), a study with a methodology similar to Yang's (2015) or Soo & Monahan's (2017) would be more appropriate, as it could reveal how Vietnamese HSs perceive each respective tone in the HL.

4. THE PRESENT STUDY

The present study provides data from an ABX task conducted with an experimental group of heritage speakers (HSs) of Vietnamese in the Netherlands and two control groups of monolingually raised Vietnamese and Dutch speakers in Vietnam and the Netherlands, respectively. In the following, the study's research questions and hypothesis are explained in reference to the literature review above (Section 4.1), followed by the study's methodology (Section 4.2).

4.1. Research question and hypothesis

The discussions in Sections 2 and 3 show that more research is needed on the acquisition of tone in both production and perception, on HSs' phonologies, and specifically, on lexical tones in heritage speakers' phonologies. The present study aims to contribute to the literature by investigating segment-tone integration in HSs, which has not been focused on before and could give information on the developmental path of tone and its integration with segments. The following research question is posed: *to what extent are tone and segments perceptually integrated in the phonologies of Vietnamese heritage speakers in the Netherlands?* As most of the research on segment-tone integration seems to focus primarily on monolinguals or late bilinguals of tonal and non-tonal languages, the study furthermore asks: *in what way does this integration or balance differ from the balance in monolingually raised (i) speakers in Vietnam and (ii) Dutch speakers without experience with a tonal language?*

Research on segment-tone integration in various types of speakers shows that, generally, speakers from any background may show some degree of sensitivity to tone. However, only tone language speakers consistently show integration of tones and segments in speech processing, whereas in speakers of non-tonal languages, this integration may be absent (Lin & Francis 2014). Speakers of a tonal language may show a symmetrical integration of tones and segments (Lin & Francis 2014; Tong et al. 2014; Wewalaarachchi et al. 2017), but in most of the studies discussed, both these speakers, bilingual speakers, as well as speakers of non-tonal languages more attention to segments (Yip et al. 1998; Tong et al. 2008; Braun & Johnson 2011; Burnham et al. 2011; Hu et al. 2012; Singh et al. 2015; Ma et al. 2017; Wewalaarachchi et al. 2017; Zou et al. 2017). Still, speakers with experience with a tonal language usually pay relatively more attention to tones than do speakers without this experience (Braun & Johnson 2011; Zou et al. 2017).

Studies on HSs' phonologies suggest that this group has both perception and production benefits over second language learners of their heritage language (HL), but they may also differ from monolinguals, be it quite noticeably or mostly on a fine-grained acoustic level (see Section 3.2). Studies on HSs' tone perception (So 2000; Yang 2015; Soo & Monahan 2017; Lam 2018; Kan & Schmid 2019) suggest that this group may show reduced sensitivity to tone compared to monolinguals of the HL, but also that they outperform L2 listeners.

Based on the above, the current study hypothesises that the HSs in the present study show less segment-tone integration than the homeland speakers (HMs) of Vietnamese but more so than Dutch speakers who have no experience with a tonal language. Moreover, apart from a reduced segment-tone integration, a general preference for word-identification based on segments is expected in all groups, but it is expected to be stronger in HSs than in HMs, and less so than in the naïve Dutch group.

4.2. Methodology

The present study's methodology was kept similar to Zou et al.'s (2017), with the intent of making the two studies more easily comparable. However, there are some differences, as will be described below. Most obviously, the present study was conducted with participants from different backgrounds (Section 4.2.1). The materials and procedure are fairly consistent with Zou and colleagues' (Sections 4.2.2 and 4.2.3 respectively). In the analysis (Section 4.2.4), some changes from Zou and colleagues' method were necessary.

4.2.1. Participants

Three groups of speakers participated in the study:

(i) 20 monolingually raised Dutch speakers (age M = 22.25, SD = 2.49, 12 women, 8 men). Originally two more participants were part of this group, but they were not included in the analysis due to their sociolinguistic background not matching the rest of the group.

(ii) 20 Dutch-Vietnamese heritage speakers (age M = 24.65, SD = 4.33, 10 women, 10 men).

(iii) 35 Vietnamese homeland speakers (age M = 20.34, SD = 1.49, 20 women, 15 men). Originally 48 HMs participated, but some could not be included in the analysis due to technical issues (1) or their sociolinguistic background not matching the rest of the group (12).

The Dutch group (NLs) did not have any experience with tonal languages, but it is difficult to find fully monolingual Dutch speakers as English is taught obligatorily early on in school, in addition to French, German, Latin, and/or Ancient Greek in high school. However, all NLs reported being raised monolingually at home and mostly using Dutch in their day-to-day lives. Most of the NLs grew up in the North or South Holland provinces, none ever lived in the Limburg province (where some dialects are argued to use lexical tone, see Section 2.5). Like the NLs, the HSs also knew other languages than their HL Vietnamese and their dominant language Dutch due to obligatory (English, French, German, etc.) language classes in school.

The HSs and HMs, although living in different countries, have origins in the same region of Hồ Chí Minh City and the general regions of Southern and South-Western Vietnam (regions indicated in Vietnamese by *miền Nam* and *miền Tây*). For the HSs, it was required that at least one of the participants' parents was from Southern Vietnam and that the participant self-reported speaking Southern Vietnamese.

Many of the HSs reported Dutch as their mother tongue, arguing that they feel more comfortable or proficient in this language. However, all HSs indicated speaking Vietnamese before any other language, followed by Dutch. All HSs grew up with Vietnamese in the household, although in some cases there was also early influence from Dutch due to elder siblings going to school already, Dutch TV, or, in one case, exposure to Dutch-speaking babysitters. However, all HSs reported Vietnamese as the main language of communication within the home during childhood. Many participants in this group indicated that nowadays they speak both Vietnamese and Dutch at home.

There was a lot of variation in the HSs' involvement in Vietnamese communities within the Netherlands, some reporting 75% of their friends having a Vietnamese background and meeting other Vietnamese weekly, while others reported not having any Vietnamese friends and rarely meeting other members from the community. Literacy in Vietnamese varied as well, with the HSs consistently self-reporting equal, if not better Dutch literacy (as well as speaking and listening) skills. None of the HSs had ever enjoyed formal education in Vietnamese at the time of testing.

Similarly to the HSs, the HMs were required to have origins in Southern Vietnam. Due to there being more participants in this group, stricter requirements could be maintained: both of the participants' parents had to be from Southern Vietnam and the participants themselves had to have grown up in this region as well. The HMs were all exposed to English to some degree, just like in the NL and HS groups, but many reported not feeling comfortable or proficient in languages other than Vietnamese. All HMs were literate in Vietnamese and had attended or were still attending formal education in Vietnamese. All but one were students at Tôn Đức Thắng University in Hồ Chí Minh City.

HMs were asked to participate instead of baseline speakers (BSs), as many Vietnamese BSs in the Netherlands are proficient in Dutch and seem to use it at least as much as Vietnamese, which could possibly lead to L1 attrition.

All subjects reported normal hearing and articulatory skills. Moreover, apart from one HS participant, all participants were attending or had attended courses at a university or university of applied sciences (called WO or HBO, respectively, in the Netherlands).

4.2.2. Materials

The participants completed an ABX task (cf. Liberman et al. 1957) adapted from Zou et al. (2017). The task used nine sets of three similar non-words played to the participants. The participants had to decide whether the first non-word (standard A) or the second non-word (standard B) was more similar to the third (the target X). There were four conditions in which participants may match X to one standard rather than the other based on the non-words' tones or on segments, as described in (i-iv) below. Each description also contains an example where X matches A (both in bold). Recall that tones in Vietnamese are indicated by diacritics; in the cases below rising *sắc* tone is marked in writing by an accent *aigu* above the vowel; falling *nặng* tone by a dot below the vowel.

(i) X is more similar to one standard in segmental content and differs from both standards in tonal content: forced-segment condition.
(A) **cá xin**
(B) tá phin
(X) **cạ xin**

(ii) X is more similar to one standard in tonal content and differs from both standards in segmental content: forced-tone condition.
(A) **cá xin**
(B) cạ xin
(X) **tá phin**

(iii) X is similar to one standard in both tonal and segmental content, and differs from the other standard in both tonal and segmental content: segment-and-tone condition.
(A) **cá xin**
(B) tạ phin
(X) **cá xin**

(iv) X is similar to one standard in tonal content but not segmental content, while being similar to the other in segmental content but not tonal content: segment-or-tone condition. Note that in this condition there is no correct or incorrect answer: the participants' choices simply indicate a preference for segment-based or tone-based word identification.
(A) **cá xin**
(B) **tạ phin**
(X) **cạ xin**

The use of these conditions, as Zou et al. (2017) explain, allows to see whether participants can process tones and segments phonologically and moreover, to what extent they pay more attention to one dimension rather than to the other. The segment-and-tone condition provides a baseline for reaction times (RTs) and accuracy, as participants can use both the segmental and tonal dimension to make their decision and there is no mismatch between the correct standard and the target. Longer RTs, corresponding to more processing effort, are expected for the other three conditions, as in these conditions the participants can only rely on one dimension. By comparing the baseline RTs and accuracies to those in other conditions, it can be revealed to what extent each group experiences difficulties processing just one dimension while ignoring another (forced-segment and forced-tone conditions) and having to choose between the two dimensions (segment-or-tone condition).

Non-words are used in the task, as this allows for a fairer comparison across groups: there is not one particular group that is advantaged because their L1 is used for the stimuli, while another group does not speak this language. Nevertheless, the non-words have to conform to the phonotactics of both Dutch and Vietnamese, so that the stimuli help elicit a linguistic mode of processing for each group. Consequently, they ideally have to consist of the following components shared between the two languages (cf. Pham et al. 2018): (i) phonemes; (ii) phoneme sequences, and (iii) suprasegmental patterns. Clearly lexical tone is not shared by both languages, but this is exactly a factor of interest in the present study.

As in Zou et al. (2017), disyllabic non-words were created, as the ABX task is fairly cognitively demanding. Longer non-words might thus make the task considerably more difficult. The non-words had the following structure built up from consonants (C) and vowels (V): $C_1V_1C_2V_2C_3$. This structure was chosen instead of the CVCV structure in Zou et al. (2017) as it seemed to conform more to Dutch syllable structure: although CVCV syllables are quite common in Dutch, it was thought this mostly concerns words ending in [ə], a vowel that was not used in the current study.

The syllable-initial consonants that have equivalents present in both Southern Vietnamese and Dutch were /t, k, ʔ, m, n, f, v, s, j, h, l/. /v/ was excluded as some Southern Vietnamese speakers may realise it as [j] instead of [v] and this segment could thus lead to confounds across groups. The syllable-final consonants that were shared were /m, n, ŋ, j, w/, but as final /ŋ/ has a limited word-final distribution in isolated Dutch words, this consonant was not used. Both languages also share /p, t, k/ word-finally, but in Vietnamese these are unreleased whereas they are not in Dutch, so they were avoided. The shared vowels are /i, e, a, u, o, ɛ, ɔ/ (cf. Thompson 1965: 93-97). No diphthongs were used, hence the final semivowels /j, w/ were also avoided. Within each set of non-words, segmental differences only ever concerned the syllable-initial consonants (C_1 and C_2), meaning that the vowels and the word-final nasal remained constant throughout each set.

The non-words were accompanied by rising *sắc* or falling *nặng* tone on the first syllable and with high level *ngang* tone on all second syllables. Rising *sắc* and falling *nặng* were used in contrast to each other, as these tones have (partially) opposite contours and are therefore easier to distinguish than more similar tone pairs (cf. Tsao 2008). The rising *sắc* and falling *nặng* tones were used on the first syllable, as Braun & Johnson (2011) find that in this position pitch becomes less relevant for Dutch listeners. On the second syllables, these tones could be interpreted as question or statement intonation, respectively, in which case it is no longer possible to argue lexical tone perception is being tested.

The list of non-words created was checked with a Vietnamese HM for how natural they seemed. They were also checked by the author and two other Dutch speakers for naturalness, although the necessary likeness to Vietnamese made it difficult to make the non-words sound truly Dutch. Next, four Southern Vietnamese HMs (exchange students in the Netherlands) were paid to help record the non-words in a sound-booth in the Phonetics Laboratory at Leiden University. The recordings were made in Adobe Audition C6 version 5.0.2 (build 7) using a Sennheiser MKH 416T condenser microphone.

It was necessary to specifically record Southern Vietnamese speakers, as dialects of Vietnamese may differ considerably in both segments and tones (see Phạm & McLeod 2016 and Section 2.4). Moreover, most of the HSs in the Netherlands also have Southern Vietnamese origins and are thus most familiar with this dialect (see Section 3.3). The recording sessions started with obtaining informed consent and providing both oral and written instructions. After the speakers filled out a sociolinguistic survey, the non-words were presented in *quốc ngữ*, the Vietnamese orthography, in randomised order on a Dell OptiPlex 3040 monitor. Each non-word (9 pairs × 2 tone options = 36 non-words) was pronounced approximately 12 times by each speaker. The experimenter controlled the speed with which the speakers could move on to the next non-word.

After the recording, the individual non-words were extracted in Praat (Boersma & Weenink 2019) and the best repetition of each non-word by each speaker was selected by the experimenter. As the experimenter's proficiency in Vietnamese was limited, the selected tokens were then judged by three Southern Vietnamese HMs not participating in the study, who rated them on scales ranging from 1 ('*Not at all*') to 4 ('*Very much*') regarding (i) how natural the non-words sounded and (ii) how much the speaker seemed to speak in a Southern Vietnamese accent. Tokens that were rated as 1 or 2 on one or both of the scales (i.e. not very natural or Southern Vietnamese) by two or more raters, were replaced with a different token in a second round. In the second round, all 26 replacements were rated as sounding like (almost) natural words. The majority of the non-words that were originally rated as sounding different from Southern Vietnamese were rated better in the second round (12 tokens) and the ones that were not rated better, did not receive a worse score than in the first round (9 tokens).

One speaker's productions were excluded from the experiments as over half of the selected tokens were judged as not Southern Vietnamese and this speaker spoke both Northern as well as Southern Vietnamese, growing up in the South with a Northern Vietnamese mother. Consequently, the productions of three speakers were used, the same number of speakers helping in Zou et al. (2017). To account for the remaining 9 tokens that were rated as not sounding completely like Southern Vietnamese (all associated with one female speaker of the remaining three speakers), the speaker of each target (X) was included in the analysis as a control variable (see Section 4.2.4).

Figures 1 and 2 (repeated from Section 2.4) show the F0 contour in two stimuli, representing rising *sắc* (Figure 1) and falling *nặng* (Figure 2), respectively, in each first syllable, as well as high level *ngang* in each second syllable:

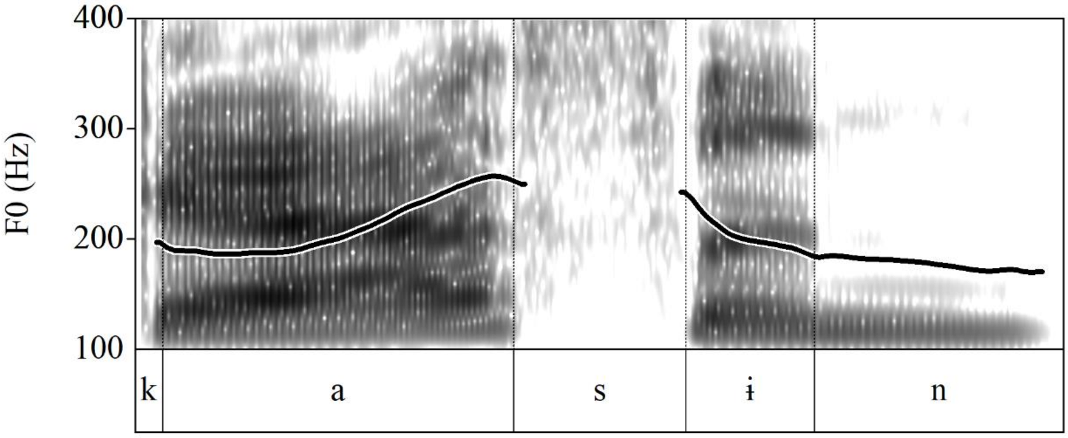


Figure 1 (repeated): Spectrogram with F0 contour for the non-word cá xin *(rising* sắc*, neutral* ngang*)* stimulus produced by the male speaker.

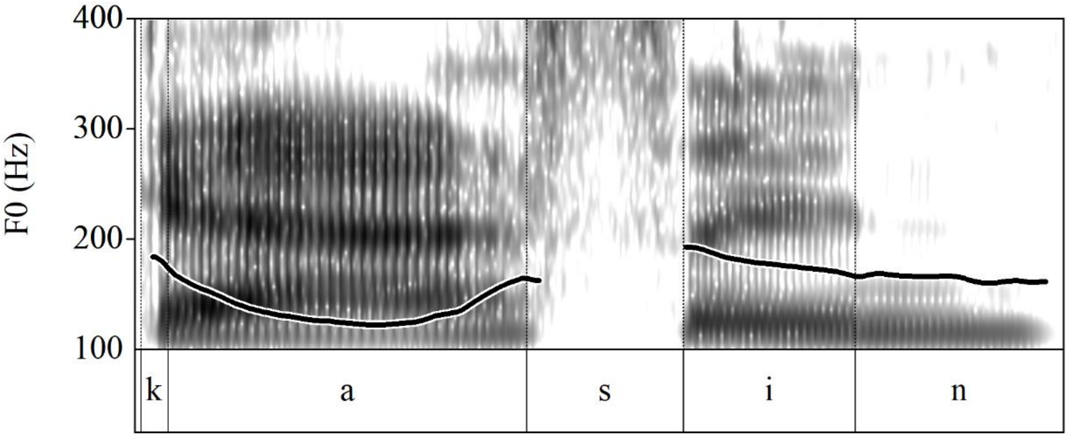


Figure 2 (repeated): Spectrogram with F0 contour for the non-word cạ xin *(falling* nặng*, neutral* ngang*) stimulus produced by the male speaker.*

Among the productions used, the male speaker's average change in F0 in the first syllables with falling *nặng* tone was 49.4 Hz (SD = 23.6 Hz), a smaller change than in the first syllables with rising *sắc* tone (M = 86.4 Hz; SD = 15.9 Hz). The first female speaker's average F0 changes in falling *nặng* tone syllables (M = 57.0 Hz; SD = 19.8 Hz) likewise were smaller than those in rising *sắc* tone syllables (M = 73.6 Hz; SD = 15.2 Hz), but the second female speaker's F0 changes were larger in falling *nặng* syllables (M = 107.9 Hz; SD = 30.9 Hz) than in rising *sắc* syllables (M = 95.3 Hz; SD = 33.0 Hz).

In the second syllables, which always carried mid-high level *ngang* tone, all speakers produced higher mean F0 following a rising *sắc* tone syllable (male speaker M = 180.4 Hz; SD = 4.9 Hz; first female speaker M = 229.9 Hz; SD = 4.7 Hz; second female speaker M = 218.1 Hz; SD = 5.6 Hz) than in syllables following a falling *nặng* tone syllable (male speaker M = 173.6 Hz; SD = 4.1 Hz; first female speaker M = 220.3 Hz; SD = 5.5 Hz; second female speaker M = 209.5 Hz; SD = 3.9 Hz).

Within each trial, each non-word (A, B, and X) was pronounced by a different speaker. This use of different speakers for each non-word in a trial is intended to mimic real-life phonetic variation and to encourage participants to use more than phonetic (i.e. phonological) perception to understand whether different speakers are saying the same non-word (cf. Lukyanchenko & Gor 2011; Zou et al. 2017). Trials were assigned an order of the three speakers following a Latin Squares design. The trials were then randomised for each participant. In total, there were 5 practice trials (all in the segment-and-tone condition) and 288 experimental trials (9 different sets of non-words × congruency of X with A or B × 4 different combinations within each set (i.e. A = non-word 1 with rising *sắc* tone, A = non-word 1 with falling *nặng* tone, A = non-word 2 with rising *sắc* tone, A = non-word 2 with falling *nặng* tone and vice versa for B) × 4 conditions).

4.2.3. Procedure

The groups in the Netherlands (NLs and HSs) were tested in a sound-booth at the Phonetics Laboratory at Leiden University, or, in the case of a number of HSs, in an otherwise quiet room. The HMs in Hồ Chí Minh City, Vietnam, were tested in a quiet room on the campus of Tôn Đức Thắng University and in one case, at home. Before participating in the experiment, participants signed an informed consent form and took part in a sociolinguistic survey which focused on language use, education, and attitudes (based on surveys for bilingual and HS communities by Dai & Zhang (2008), Chang et al. (2011), Birdsong et al. (2012) and Moro (2016)). Interactions between the participants and the experimenter were in Dutch for the HSs and NLs, and in Vietnamese for the HMs. Participants were not told the purpose of the study, apart from the focus being on the perception of non-words. HSs and HMs were aware the study focused on Vietnamese, since the calls for participation specified the need for Southern Vietnamese participants.

The experiment was run using E-prime (version 2.0.10.356) on an HP Thinbook 14-bp085nd 14-inch laptop with Bose Soundlink headphones attached. For NLs, on-screen instructions were written in Dutch; for HSs, the instructions were provided in both Dutch and Vietnamese; for HMs, the instructions were provided in Vietnamese only. The instructions informed participants of the expected way to answer and encouraged them to answer as quickly and accurately as possible.

After the instructions, the participants were presented with the 5 practice trials to see whether they understood the task. If they had no more questions, the participants could move on to the experimental trials and the experimenter would leave the sound-booth or, during experiments at participants' home or at Tôn Đức Thắng University, sit a bit further away from the participants to prevent nervousness.

Each trial started with a 1000 ms silence. Then, for each non-word within a trial, the letters A, B, and X, respectively, appeared in the middle of a white screen along with the auditory stimuli, so that participants could keep track of what part of a trial they were listening to. There were 600 ms between A and B, and another 900 ms between B and X (cf. Braun & Johnson, Zou et al. 2017). As soon as the X stimulus was presented aurally and visually, participants were able to answer by pressing on keys on the laptop's keyboard labelled A and B with stickers. The answer keys corresponded to the 'A' and 'K' keys on a QWERTY keyboard and thus were on an equal height and allowed for enough space for both hands to rest comfortably on the keyboard. The participants' RTs were measured from X stimulus onset and when they answered their response was registered. If the participants did not answer within 7 seconds after the onset of the X stimulus, the next trial would start (preceded, again, by a 1000 ms silence).

The 288 experimental trials were randomly presented in four blocks of 72 trials. Between blocks, participants were allowed to take a short break and were able to move on with the next block whenever they decided to. After completing the experiment, the participants received a compensation for their participation.

4.2.4. Analysis

The analysis focused on the two types of data collected: (i) response type, operationalised as 1 (correct; or, in the segment-or-tone condition: classification along the segmental dimension) or 0 (incorrect; classification along the tonal dimension in the segment-or-tone condition); and (ii) RTs, which were logarithmically transformed, as non-transformed RTs are rarely normally distributed.

Two models were thus built using R (R Core Team 2019), one for each type of data. Recall that in Section 4.1 it was hypothesised that HSs have a less integrated perception of segments and tones than HMs, but a more integrated perception of these dimensions than the NLs. Moreover, all groups are expected to be most sensitive to the segmental dimension, but with NLs relying on this dimension more than HSs and they, in turn, more than HMs. This hypothesis therefore points to two effects on response type and RTs:

- An interaction effect of group with condition: as the various conditions in the experiment reveal each participant's sensitivity to the tonal and segmental dimensions, respectively, and as this sensitivity is expected to vary per group, certain conditions may be easier or more difficult for specific groups.
- A main effect of condition: regardless of group, all participants are expected to be at least somewhat more sensitive to the segmental dimension than to the tonal dimension. The forced-segment condition might thus lead to overall higher accuracies and shorter RTs than the forced-tone condition.

Additionally, three control variables and their interaction effects were taken into account:

- The target X's tone: included because there is an expected interaction effect as described in the next point.
- Interaction effect of group with X's tone: Yeung et al. (2013) in Section 2.1 found that Mandarin children exposed to Cantonese tones sometimes paid more attention to tones that had an equivalent in their L1, which was also suggested for adults by Francis et al. (2008). This connects to Zou et al.'s (2017) suggestion that Dutch listeners may have more ease identifying targets with a falling tone on the first syllable and neutral tone on the second, which they argued could be mapped to a Dutch intonation contour (H* L L%). A high-tone first syllable would cause more problems, however, and thus make identification more difficult (Zou et al. 2017: 1026). Although Zou and colleagues found no such effect, the interaction was still taken into account here, considering the evidence from Yeung et al. (2013). The NLs may be found to perform better when X has falling *nặng* tone on its first syllable than when it has rising *sắc* tone.
- The speaker who produced the X token in a trial: as mentioned in Section 4.2.2, some of the recorded stimuli were not convincingly rated as sounding like Southern Vietnamese for one speaker. Subtle phonetic differences may thus affect word identification.
- Interaction effect of group and X's speaker: it could be expected that the Vietnamese-speaking groups respond differently to phonetic differences between X's speakers than do the NLs, as the former groups are assumed to at least be somewhat familiar to dialectal variation in Vietnamese, whereas this variation might be harder to recognise and process phonologically for the NLs.
- Response button, i.e. whether the correct answer is associated with A or B (for the segment-or-tone condition: whether the segmental answer is associated with A or B): participants may respond differently to the same trio of non-words based on whether the matching standard is

temporally closer or further away from the target X (cf. Macmillan & Creelman 2005; Braun & Johnson 2011; Zou et al. 2017).

- Interaction effect of condition and response button: if there is a main effect of condition, it may be found that, in conditions that are difficult for participants, there is a preference for A or B standards based on their temporal distance to X (cf. Braun & Johnson 2011; Zou et al. 2017).
- Three-way interaction of group, condition, and response button: if there is an interaction of group and condition, it may also be found that, in conditions that are particularly hard for a specific group, there is a preference for A or B standards based on their temporal distance to X (cf. Braun & Johnson 2011; Zou et al. 2017).

For response type, a mixed effects logistic regression model was constructed, and for RTs a linear mixed effects model was built, both using the lme4 package (Bates et al. 2015). The models both had random intercepts by participant and by non-word pair. The independent variables listed above were added bottom-up in a stepwise manner, considering for each variable whether its effect contributed to the respective model. Finally, random slopes were added to both models, but this led to the models failing to converge or getting overfitted, i.e. there was not enough data to support the models with random slopes. Therefore, random-slopes were excluded from the analysis.

For RTs, z-scores were generated for all trials. Zou et al. (2017) did the outlier trimming after the model has been constructed. Given that there were some extreme outliers, I opted to do so before the analysis and removed data points with an absolute z-score larger than three standard deviations.

After construction of the models, the models' marginal R^2 (proportion of variance accounted for by the models' fixed effects) and conditional R^2 (proportion of variance accounted for the models' fixed and random effects combined) were calculated using the *MuMIn* package (Barton 2019; cf. Nakagawa & Schielzeth 2013). Next, post-hoc, Bonferroni-corrected pairwise t-tests (from the *stats* package, R Core Team 2019) were conducted to further interpret the found effects. In Zou et al. (2017), post-hoc analysis was done using the *multcomp* R package (Hothorn et al. 2008), but as the models in the present study turned out to be quite complex, there was consequently too little data to use this same method of post-hoc analysis. Note, additionally, that a t-test cannot be used with the raw binary response type data, as t-tests assume continuous or ordinal data. Therefore, for the response type post-hoc, t-tests were to be performed on by-subject means, which are continuous. For instance, to see the effects of condition, each subject's average score per condition was calculated and would be used in the t-test. However, these means were rarely normally distributed and indeed show a considerable amount of variation, thus violating another assumption of the t-test and increasing the chances of false positives, making the conclusions drawn from the tests dubious at best.[4] For these reasons, no statistical post-hoc tests could be performed on the response type data (see Section 6.5 for a discussion of limitations of the present study). Instead, in Sections 5 and 6, percentages of correct trials will be discussed without further statistical tests, with a clear warning to the reader that these results can only be taken as a general tendency and that no strong conclusions should be drawn about the response type data for now.

5. RESULTS

There were 21.456 trials in total for the 75 participants included in the analysis. Of these trials, 115 (0,5%) were excluded from the analysis as participants were not able to answer in time. A total of 21.341 trials were left. Recall that for the RT model, the trials were furthermore z-trimmed to remove extreme outliers, excluding an additional 99 trials (0,5% of the trials still included in the response type analysis). The two models were thus constructed with marginally different sample sizes.

In Table 2, degrees of freedom, χ^2-values, and *p*-values are reported for each fixed effect added to the models. Note that the effects were added using stepwise modelling and the table reflects the order in which they were added; i.e. the experimental fixed effect of group was added first, then

[4] See Appendix B for various histograms revealing the non-normal distributions of the by-subject means.

condition and so on, with the Speaker X × Group interaction being added last. At the bottom of the table, the marginal and conditional R^2 of each model are given.

	Response type			Logarithmically transformed reaction times		
	df	χ^2	*p*	*df*	χ^2	*p*
Experimental variables						
Main effects						
Group	2	37.864	< .001	2	7.2714	0.026
Condition	3	1028.4	< .001	3	1604	< .001
Interaction						
Group × Condition	6	604.94	< .001	6	596.61	< .001
Control variables						
Main effects						
Response button	1	143.23	< .001	1	39.42	< .001
Tone X	1	7.4623	0.006	1	1.1792	0.278
Speaker X	2	6.9287	0.031	2	5.3048	0.070
Interactions						
Response button × Condition	3	32.054	< .001	3	29.735	< .001
Response button × Condition × Group	8	53.301	< .001	8	25.138	0.001
Tone X × Group	2	0.8155	0.665	3	1.629	0.653
Speaker X × Group	4	2.3322	0.675	6	16.372	0.012
R^2 values						
Marginal R^2		0.119			0.100	
Conditional R^2		0.158			0.329	

Table 2: Degrees of freedom (df), χ^2-values, and p*-values for the models' fixed effects. Effects were added to the models in a stepwise manner, going from the top of the table (Group) to the bottom (Speaker X × Group). Below the fixed effects the models' R^2 values are provided.*

Table 2 shows that, apart from main effects of group ($\chi^2(2) = 37.864$, $p < .001$ for response type, $\chi^2(2) = 7.2714$, $p < .05$ for RTs) and condition ($\chi^2(3) = 1028.4$, $p < .001$ for response type, $\chi^2(3) = 1604$, $p < .001$ for RTs), there was an interaction of group and condition in both models ($\chi^2(6) = 604.94$, p < .001 for response type, $\chi^2(6) = 596.61$, p < .001 for RTs). For response type, Figure 3 shows that the groups are most similar in the segment-and-tone condition (HMs: 83.0% of responses correct; HSs: 84.2%; NLs: 92.3%), although the HMs and HSs are closer to each other than to the NLs. This clustering of the Vietnamese groups persists throughout all conditions, although in the forced-tone condition the HMs (68.1%) are not that much closer to the HSs (72.6%) than to the NLs (59.5%). The forced-tone condition is also the only condition where the NLs have a lower score than the other two groups. Within the Vietnamese groups, the HMs consistently had fewer correct answers than the HSs, regardless of condition.

For RTs, as evident from the graphs in Figure 4, the post-hoc t-tests show that the groups differ in each condition, ($p < .001$), except in the segment-and-tone condition, where the HSs and NLs do not differ significantly from each other ($p = 1$). In conditions with a partial mismatch, the Vietnamese groups consistently have longer RTs than the NLs, except in the forced-tone condition.

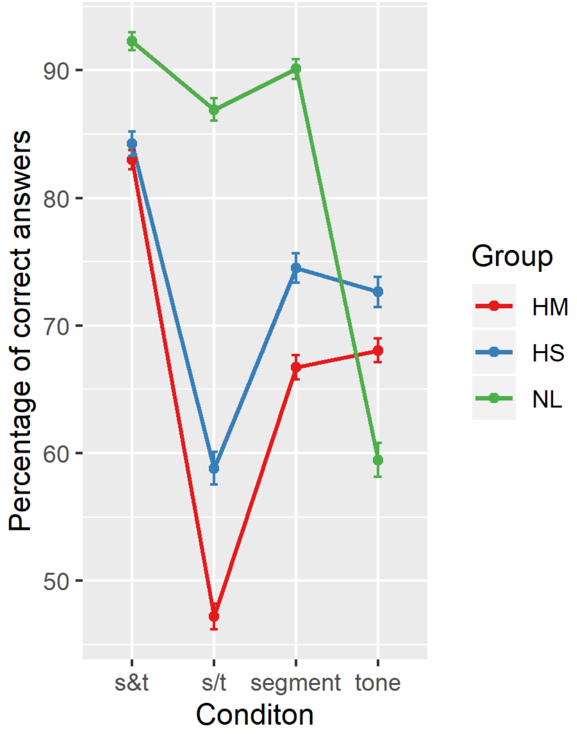


Figure 3: Group percentages of correct trials per condition (and percentages of segment-based choices in segment-or-tone condition). S&t = segment-and-tone condition; s/t = segment-or-tone condition; segment = forced-segment condition; tone = forced-tone condition.

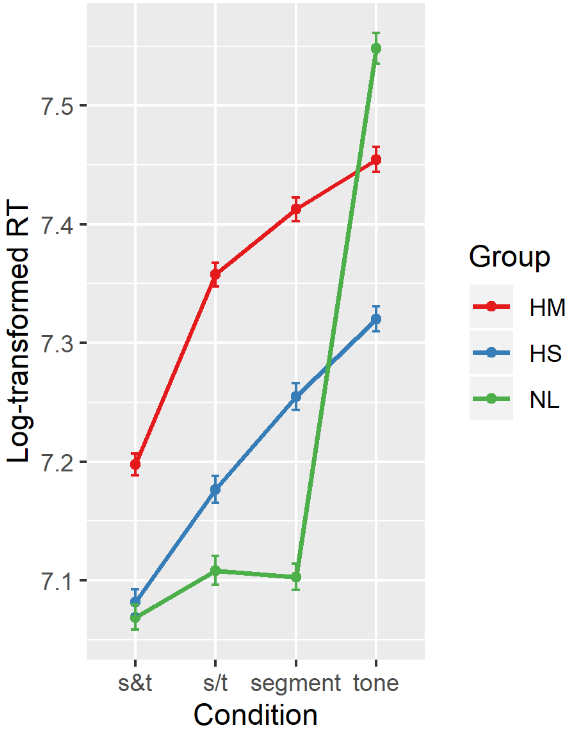


Figure 4: Group mean log-RTs per condition. S&t = segment-and-tone condition; s/t = segment-or-tone condition; segment = forced-segment condition; tone = forced-tone condition.

For the forced-segment and forced-tone conditions, difference scores for the three groups were calculated and plotted in Figure 5, in line with Zou et al. (2017). The difference scores are calculated from each participant's percentage of correct answers for the forced-segment and forced-tone conditions, respectively, where the forced-tone score is subtracted from the forced-segment score. There is some overlap between all three groups, but clearly more so between the HMs and HSs than between either of those groups and the NLs. The NL scores are considerably higher, whereas the HMs and HSs score closer to zero.

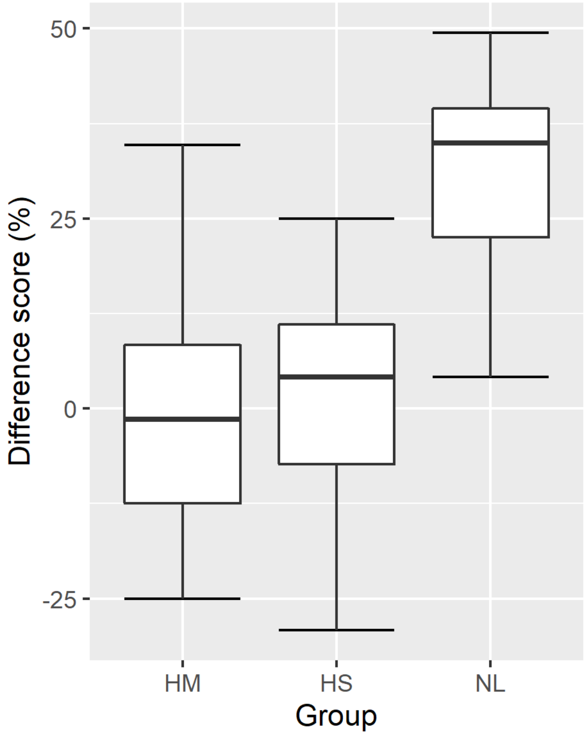


Figure 5: Boxplots of difference scores per group. The by-participants scores were calculated by subtracting their percentages of correct responses in the forced-tone condition from those in the forced-segment condition. The groups' respective most prevalent scores (i.e. medians) are indicated by the black lines in the middle of the boxes. The two parts of the boxes each represent 25% of the scores closest to the median and the whiskers extending from the boxes represent each represent 25% of the scores farthest away from the median.

Lastly for the interaction effect of group and condition, Figures 6 and 7 below can be used to interpret within-group differences across conditions. Regarding accuracy (Figure 6), HMs show a different tendency than the other groups. Like the others, they were most accurate in the segment-and-tone condition (83.0% correct), but in contrast to the other groups, they were then most accurate in the forced-tone condition (68.1%) and slightly less accurate in the forced-segment condition (66.7%), though the difference between these conditions seems very marginal. In the segment-or-tone condition, they chose along the segmental dimension in 47.2% of the trials. Both the HSs and NLs are relatively most accurate in the segment-and-tone condition (HSs: 84.2%; NLs: 92.3%), then in the forced-segment condition (HSs: 74.5%; NLs: 90.1%), and lastly in the forced-tone condition (HSs: 72.6%; NLs: 59.5%). Note, that for the HSs, as for the HMs, the scores in the forced-segment and forced-tone conditions are fairly similar. In the segment-or-tone condition, the HSs choose along the segmental dimension in 58.8% of the trials and the NLs do so in 86.9% of the trials.

Regarding RTs (Figure 7), both HMs and HSs relatively have the shortest RTs in the segment-and-tone condition, next in the segment-or-tone condition, then in the forced-segment condition, and they are slowest in the forced-tone condition (for the HMs, forced-segment vs. forced-tone yields $p < .05$, all other comparisons between conditions yield $p < .001$; for the HSs: all comparisons yield $p < .001$). For the NLs, only the RTs in the forced-tone condition are significantly longer than those in other conditions ($p < .001$ for each comparison), the other conditions do not differ significantly from each other ($p > .05$ for all comparisons).

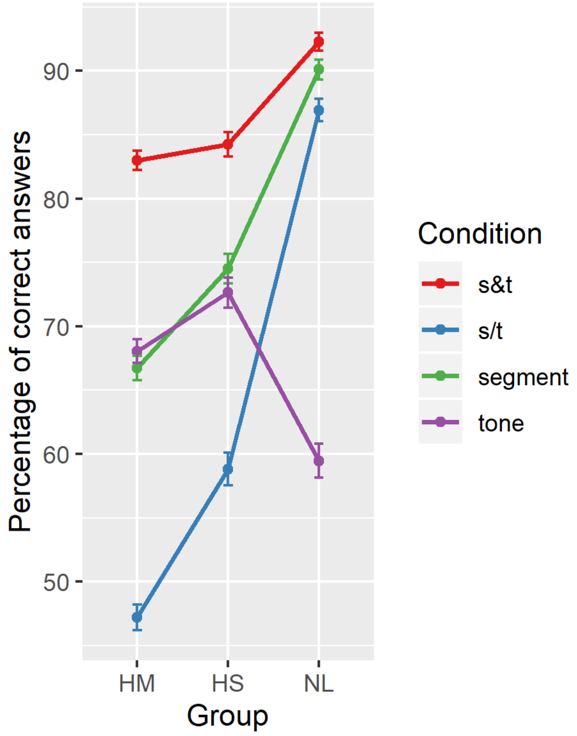


Figure 6: Percentages of correct trials in each condition per group (and percentages of segment-based choices in segment-or-tone condition). S&t = segment-and-tone condition; s/t = segment-or-tone condition; segment = forced-segment condition; tone = forced-tone condition.

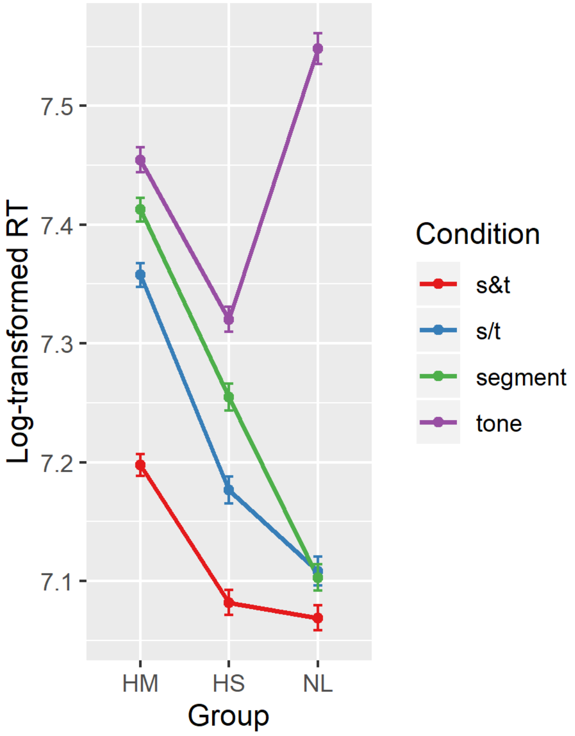


Figure 7: Mean log-RT in each condition per group. S&t = segment-and-tone condition; s/t = segment-or-tone condition; segment = forced-segment condition; tone = forced-tone condition.

Moving on to the effects from the control variables in Table 2, the main effect of response button on response type ($\chi^2(1) = 143.23$, p < .001) shows that trials where X matches with the A standard (or in the case of the segment-or-tone condition, with segmental matches with A) led to lower overall accuracy (68.8%) than trials where X matches with B (75.7%; the two types of trials are henceforth referred to as A-trials and B-trials, respectively). This effect can be seen for the sample of participants as a whole in Figure 8. Likewise, the effect was present in the RT model ($\chi^2(1) = 39.42$, p < .001), where A-trials usually led to longer RTs than B-trials (see Figure 9).

Additionally, there was an interaction effect of response button and condition in the response type data ($\chi^2(3) = 32.054$, $p < .001$), which seems to overlap almost completely with the main effect of response button somehow. The interaction effect could be due to the lesser difference between A- and B-trials in the forced-segment condition (73.5% vs. 76.7% respectively), or the similarity between the segment-or-tone and forced-tone conditions in the B-trials (66.4% vs. 69.8% respectively). Regardless, these tendencies do not seem to differ from the main effect enough to warrant extensive discussion in Section 6.

Considering the same interaction effect on RTs ($\chi^2(3) = 29.735$, $p < .001$), however, more interesting observations can be made. A- and B-trials differ significantly in RTs in all conditions (*p* < .01), except in the forced-tone condition (*p* = .24). Figure 9 shows that in this condition, B-trials caused slightly longer RTs than A-trials (although not significantly so), whereas in the other conditions the reverse is true (and these latter differences are statistically significant). This effect will be discussed in Section 6.4.

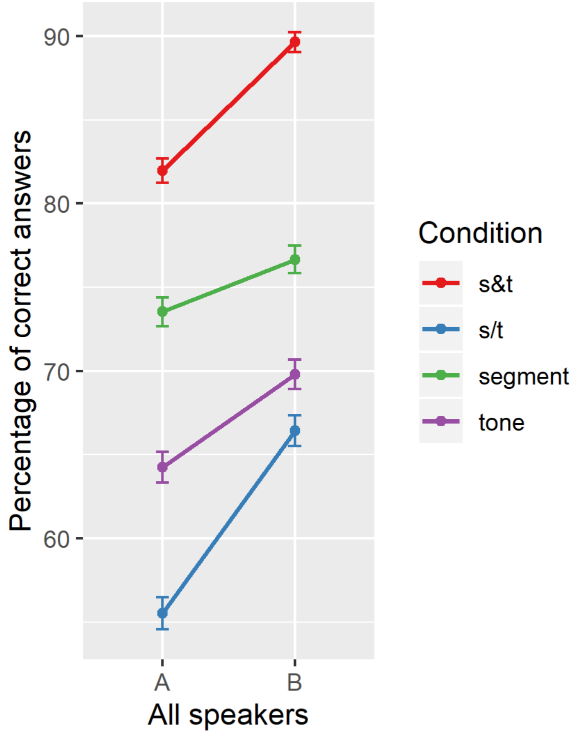


Figure 8: Percentages of correct answers in A-trials vs. B-trials in each condition (and percentages of segment-based choices in segment-or-tone condition). S&t = segment-and-tone condition; s/t = segment-or-tone condition; segment = forced-segment condition, tone = forced-tone condition.

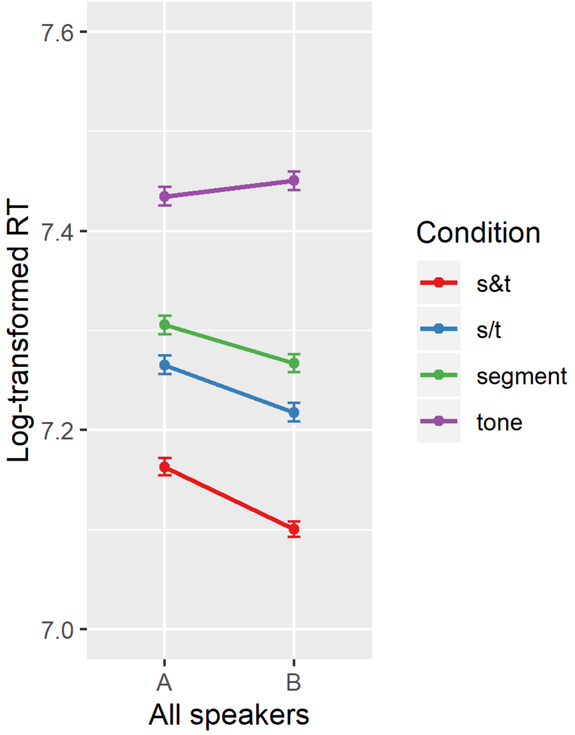


Figure 9: Mean log-RT in A-trials vs. B-trials in each condition. S&t = segment-and-tone condition; s/t = segment-or-tone condition; segment = forced-segment condition, tone = forced-tone condition.

Next, there was a three-way interaction effect of response button, group, and condition on response type ($\chi^2(8) = 53.301$, $p < .001$). Figure 10 shows that the HMs responded slightly differently to A- and B-trials mostly in the segment-and-tone (80.1% correct in A-trials vs. 85.9% in B-trials) and segment-or-tone conditions (choosing along the segmental dimension in 42.4% of A-trials and 52.0% of B-trials). On the other hand, in the forced-segment (65.1% correct in A-trials vs. 68.4% in B-trials) and forced-tone conditions (67.6% correct in A-trials vs. 68.5% in B-trials) preferences for one standard seem to be even less present.

A three-way interaction effect of response button, group, and condition was also found for the RTs ($\chi^2(8) = 25.138$, $p < .01$): as also suggested in Figure 11, HMs showed a significant difference in RT between A- and B-trials only in the segment-and-tone (*p* < .001) condition, where they were faster in B-trials. In the other conditions, A- and B-trials did not show a significant difference (*p* > .05).

The HSs were generally more accurate in B-trials as can be seen in Figure 10 (segment-and-tone: 75.7% in A-trials vs. 92.8% in B-trials; forced-segment: 72.9% vs. 76.1%, forced-tone: 67.9% vs. 77.4%), although this effect is less noticeable in the forced-segment condition. Note that, whereas the HSs were slightly more accurate in the forced-segment condition than in the forced-tone condition in A-trials, this effect is reversed in B-trials. In the segment-or-tone condition they chose along the segmental dimension less often in A-trials (50.0%) than in B-trials (67.6%). These results mirror the RT data (see Figure 11): RTs were usually shorter in B-trials, although this was only significant in the segment-and-tone (p < .001) and segment-or-tone (p < .01) conditions. The effect was not statistically significant in the forced-segment and forced-tone conditions (p > .05).

The NLs were generally slightly more accurate in B-trials (see Figure 10), with this preference being clearest in the forced-tone condition (segment-and-tone: 91.5% correct in A-trials vs. 93.0% in B-trials; forced-segment: 88.7% vs. 91.5%; forced-tone: 54.6% vs. 64.3%). In the segment-or-tone condition the NLs chose along the segmental dimension less often in A-trials (83.7%) than in B-trials (90.1%). Considering RTs, NLs only showed significant differences between A- and B-trials in the forced-segment (p < .05) and segment-or-tone conditions (p < .01), where B-trials yielded shorter RTs (see Figure 11). Note that, although differences in the remaining conditions were not statistically significant (p > .05), NLs were slightly faster in A-trials than in B-trials in the forced-tone condition.

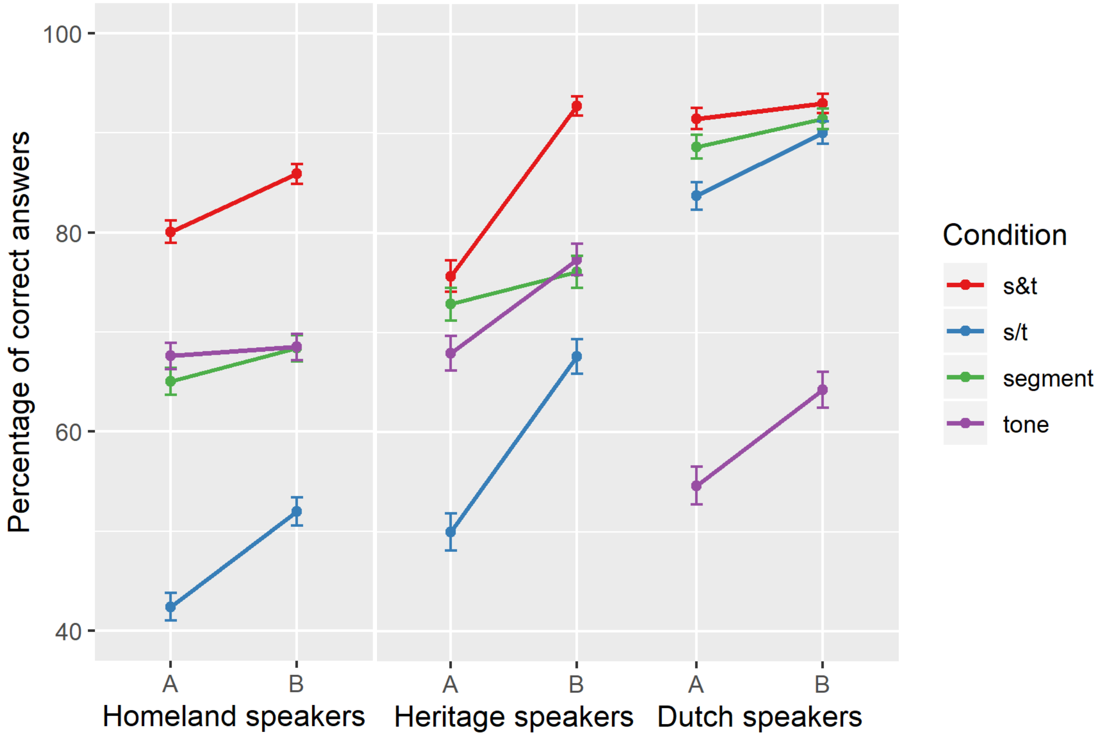


Figure 10: The groups' percentages of correct A- and B-trials for each condition (and percentages of segment-based choices in segment-or-tone condition). S&t = segment-and-tone condition; s/t = segment-or-tone condition; segment = forced-segment condition, tone = forced-tone condition.

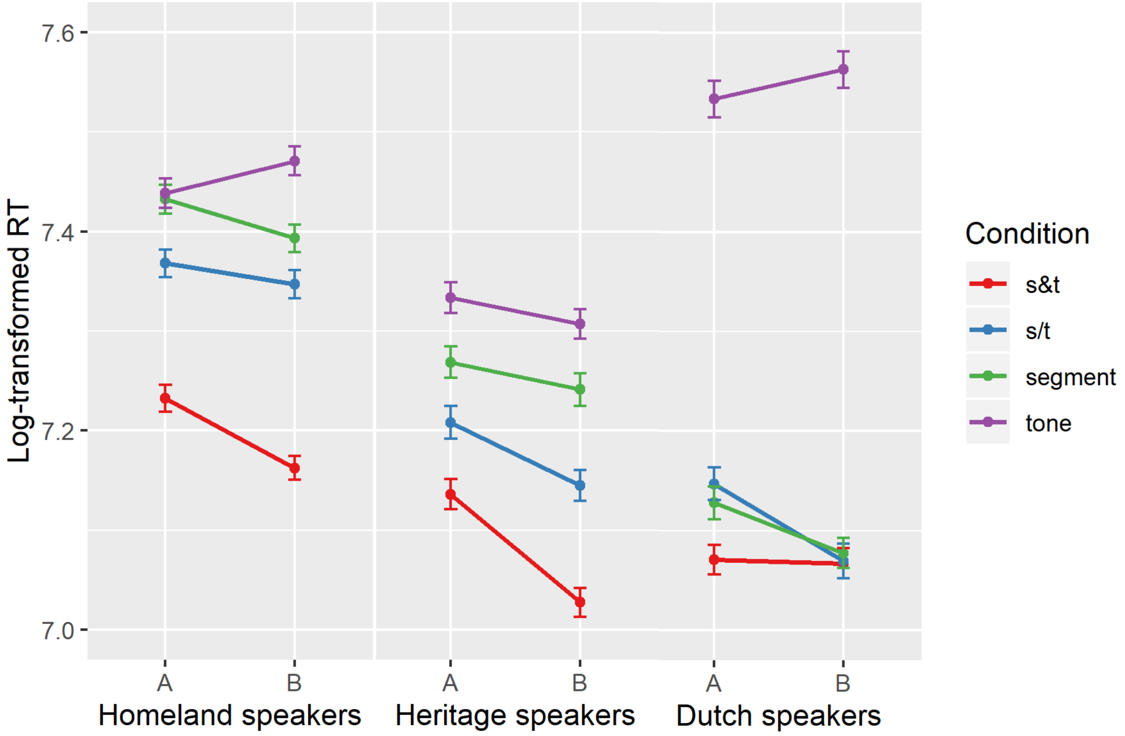


Figure 11: The groups' mean log-RTs in A- and B-trials for each condition. S&t = segment-and-tone condition; s/t = segment-or-tone condition; segment = forced-segment condition, tone = forced-tone condition.

Continuing with a different control variable from Table 2, there was a main effect of the target X's tone on response type only ($\chi^2(1) = 7.4623$, $p < .01$). Figure 12 shows that overall, participants regardless of group were marginally more accurate in rising *sắc* tone trials (74.5% correct) than in falling *nặng* tone trials (72.9% correct).

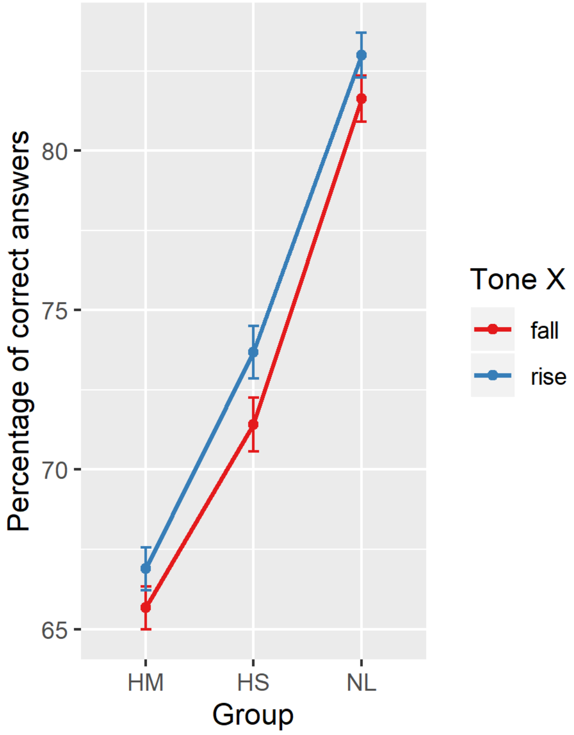


Figure 12: Percentages of correct responses in trials with a falling nặng *tone on X vs. rising tone on X, per group.*

The last effect on response type is X's speaker ($\chi^2(2) = 6.9287$, $p < .05$). Figure 13 shows that trials where the male speaker pronounced the target overall yielded the lowest accuracy (64.2% correct on average), followed by the targets pronounced by the first female speaker (71.4%). Trials with the second female speaker pronouncing the targets yielded the highest accuracies (81.1%).

Similarly, there was an interaction of X's speaker and group in the RT model ($\chi^2(6) = 16.372$, $p < .05$). As visible in Figure 14, HMs show different RTs for all speakers (between the female speakers $p < .01$; $p < .001$ for comparisons between the male speaker and the female speakers). For the HSs, only the second female speaker elicited significantly shorter RTs than the other speakers ($p < .01$ compared to the first female speaker, $p < .001$ compared to the male speaker). There was no statistically significant difference between the first female speaker and the male speaker ($p = 0.41$). Among the NLs, all speakers elicited different RTs (all speaker comparisons: $p < .001$), with the second female speaker being the easiest to process, followed by the first female speaker, and with the male speaker causing the longest RTs. Although the male speaker elicited the longest RTs for all groups, among the NLs this was even less proportionate compared to the other speakers.

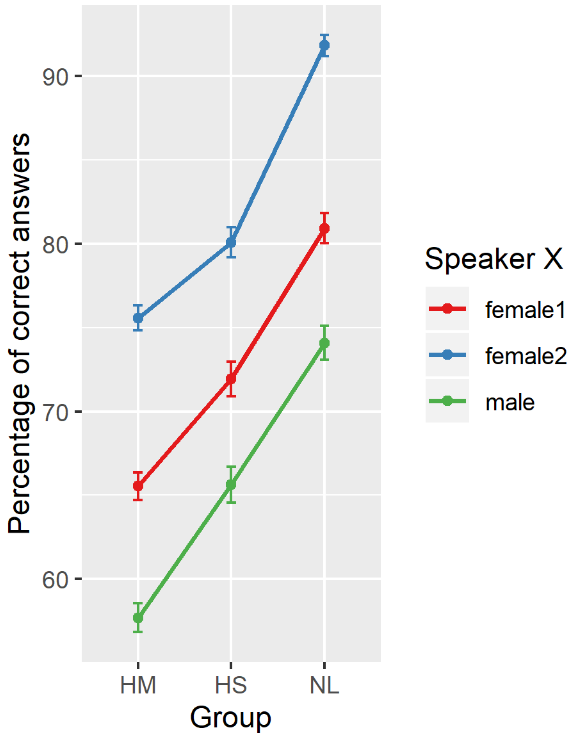


Figure 13: Percentages of correct responses elicited by the three speakers when they produced X, per group.

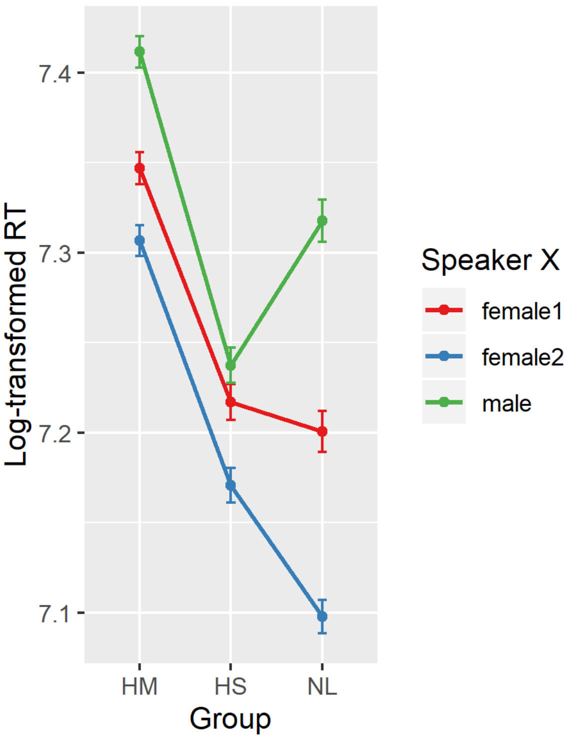


Figure 14: Mean log-RTs elicited by the three speakers when they produced X, per group.

6. DISCUSSION

The findings from Section 5 will now be discussed in more detail in relation to the study's research questions and with reference to the previous literature, in particular, to Zou et al. (2017). In Section 6.1, each group's performance in the forced-segment and forced-tone conditions will be discussed, highlighting the groups' ability to ignore or process the segmental and tonal dimensions, respectively. In Section 6.2, the groups' performance in the segment-or-tone condition is used to infer about the groups' sensitivity to one dimension over the other. Section 6.3 deals with performance within each group across conditions to highlight (a)symmetry in each group's perceptual segment-tone integration. In Section 6.4, the effects from control variables (response button, X's tone, X's speaker) are discussed more in-depth. Lastly, Section 6.5 deals with the limitations of the study.

6.1. Processing segmental and tonal dimensions

In Section 5, the three groups (HMs, HSs, NLs) were found to consistently differ from each other in average correct-incorrect response type in the forced-segment and forced-tone conditions. In the forced-segment condition, the HMs answered correctly least often (66.7% of the trials), whereas the HSs performed better (74.5%), and the NLs performed quite well (90.1%). These results corresponded with the RTs for each group, with HMs being the slowest, followed by HSs, and lastly, NLs. These findings on the NLs are consistent with Zou et al. (2017), who also found that naïve Dutch listeners performed better than Mandarin monolinguals in the forced-segment condition, likely due to being able to ignore the tonal dimension more easily. The HSs and HMs on the other hand, both speak Vietnamese and seem to have more difficulties ignoring tone. The fact that the HSs performed more accurately and faster than the HMs furthermore suggests that the HSs may find it easier to ignore tonal information. After all, in their dominant language, tone does not play an important role in word identification and thus becomes irrelevant (the same was argued by Kan & Schmid 2019 and cf. Soo & Monahan 2017). None of the groups were considerably more accurate in A- or B-trials in the forced-segment condition, but the NLs were significantly faster in B-trials in this condition, whereas this effect was not significant for the other groups. This will be discussed in relation to results from the forced-tone condition below.

In the forced-tone condition, the groups again performed differently from each other. Here, the NLs were the least accurate (59.5% of trials correct), followed by the HMs (68.1%), and the HSs performed best (72.6%). Likewise, the HSs were fastest, followed by the HMs, and lastly, the NLs. It was unexpected that the HMs had a slightly lower accuracy than the HSs, as the HMs speak Vietnamese considerably more often than the HSs and hence were assumed to rely on the tonal dimension more often as well. Moreover, evidence from other studies also suggests HMs may have better tonal perception than HSs (So 2000; Yang 2015; Lam 2018). However, it should be noted that, except in the forced-tone condition, the HMs were consistently the least accurate and slowest group in the present study. They even differed significantly from the other groups in the segment-and-tone condition, which should be easiest for all groups and where only marginal group differences were expected. Perhaps a different sample of HMs would have performed more similarly to the other groups (see Section 6.5). As expected, NLs performed least accurately and slowest in this condition, which corresponds to Zou et al.'s (2017) findings.

There was additionally a clear interaction effect of response button, group, and condition for the HSs and NLs in the forced-tone condition: these groups performed with noticeably higher accuracy in B-trials. This suggests that for these groups the task in this condition was more difficult when the correct standard was further removed from the target and that perhaps tonal information is more difficult for them to store in their working memory than is segmental information. After all, the difference between A- and B-trials in the forced-segment condition was not very big for the HSs and NLs. Based on the poor performance of monolingual Dutch listeners in their equivalent forced-tone condition, Zou et al. (2017) also suggested that this group does not retain tonal information as well as Mandarin speakers.

Within the forced-tone condition, HSs also responded faster in B-trials than in A-trials (but this effect was not statistically significant). These RTs conform to the response type data and suggest that the HSs find tonal information easier to accurately and quickly process when temporal distance between the standard and target is short. The NLs, on the other hand, were slower in forced-tone B-trials than in A-trials (although not significantly), contrasting with the response type data. This suggests that tonal information is more easily processed when there is a large temporal distance. Perhaps since tonal information is usually not relevant in the NLs' L1, it takes longer for them to process this information. The larger temporal distance between A and X gives the NLs more time to process this information, consequently allowing them to match the two faster when it is time to make a decision. The distance between B and X is shorter, not giving the NLs enough time to process B's tonal information before it is time to match it to X. Taking this into account, the NLs' significantly slower performance in A-trials in the forced-segment and segment-or-tone conditions (see Section 6.2) and their equally short RTs in A- and B-trials in the segment-and-tone condition (Section 6.3) could also be explained. In the former two conditions, tonal information needs to be suppressed. Therefore, if tonal information is not processed on time when B matches X, it becomes easier to ignore it and make a decision based solely on segments. When A matches X, tonal information is processed before X appears and the NLs have to put in more effort to ignore this information. This would suggest that the NLs do have some segment-tone integration, but that this integration is only at play when there is enough time to process tone. Although the NLs are not the group of main interest in the present study, this interpretation could be worthwhile to investigate further.

For the HSs, there was one additional effect of response button on accuracy, however marginal. In both the forced-segment and forced-tone conditions, there was a slight improvement in accuracy in B-trials over A-trials. In the forced-tone condition, however, this improvement was big enough to yield a slightly higher accuracy than in the forced-segment condition, whereas in A-trials accuracy was higher in the forced-segment condition. The higher forced-tone scores in B-trials could suggest that the HSs initially process tonal information marginally more accurately than segmental information, but that they also lose the former more easily over a longer time span (in A-trials), which led to a higher overall accuracy in processing and *retaining* segmental information. This tendency could be interesting for future research, but it should be stressed that it was only minimal and moreover, the HSs were much more convincingly faster in the forced-segment condition than in the forced-tone condition, regardless of A- or B-trials, which suggests that segments are still easiest to process (see Sections 6.2 and 6.3).

Lastly, the three groups' difference scores (see Figure 5) suggest that the HMs and HSs, who overlap considerably around zero, have a quite similar way of processing segments and tones. Note that these scores around zero are similar to those found in Zou et al. (2017) for Mandarin controls and for advanced Dutch learners of Mandarin, the groups that Zou and colleagues argued were more sensitive to tone than naïve Dutch speakers. The NLs had high difference scores, indicating that they were considerably less accurate in the forced-tone condition than in the forced-segment condition and find tone much more difficult to process than segments.

6.2. Sensitivity to segmental or tonal dimensions

In the segment-or-tone condition, the groups all behave differently from each other, although the HMs and HSs still pattern together to some extent: the HMs match standards to X along the segmental dimension in 47.2% of the trials, suggesting a slight preference for word recognition based on the tonal dimension. The HSs choose along the segmental dimension in 58.8% of the segment-or-tone trials, indicating that they prefer this dimension in word recognition rather than the tonal dimension preferred by the HMs. Lastly, the NLs identify non-words along the segmental dimension in the vast majority of segment-or-tone trials (86.9%).

Firstly, focusing on the HSs as the current study's group of interest, it could be stated that their preference for word identification along the segmental dimension confirms the hypothesis that it is stronger than the HMs' (absent) preference for the segmental dimension and weaker than that of the NLs. Where HMs have a preference for classification along the tonal dimension, HSs, arguably due to their dominance in Dutch, differ considerably from this tendency. Instead, they recognise non-words mostly based on segmental content. Still, their tonal L1 background is noticeable in their relatively stronger preference for identification along the tonal dimension compared to the NLs. Apparently, tone does play a considerable role in Vietnamese HSs' word identification and it is sometimes easier for them to classify nonwords tonally than to do so segmentally. For the NLs without a tonal language background, this choice is much more predominantly based on segments, as was also found by Zou et al. (2017). These results all conform fairly well to previous research, where HMs are relatively more sensitive to tone than HSs (So 2000; Yang 2015; Soo & Monahan 2017; Lam 2018; Kan & Schmid 2019), but HSs are still more sensitive to tone than L2 learners of their HL (Yang 2015).

The results show one direct contrast to Zou and colleagues' results: the HMs in the current study show a preference for the tonal dimension in word recognition, whereas the Mandarin controls from Zou et al. (2017) showed a preference for the segmental dimension in 62.2% of the corresponding trials. Indeed, other studies also find a preference for segment-based word identification in speakers of Mandarin (Tong et al. 2008; Braun & Johnson 2011), Cantonese (Cutler & Chen 1997; Yip et al. 1998;

Burnham et al. 2011), and Thai (Burnham et al. 2011). Possibly the HMs in the present study, with their marginal preference for tone-based word identification, are more similar to the Mandarin speakers in Lin & Francis' (2014) study, who show a fairly symmetrical segment-tone integration. Ye & Connine (1999) did find Mandarin speakers to identify tone more easily than segments, but only in idiomatic (i.e. predictable) sentences, which is not the case here. Pham et al. (2018) found that Vietnamese-English bilingual children produced tones more accurately than segments and thus might be more sensitive to tone, like the HMs in the present study. However, Pham and colleagues admit that their analysis of the children's productions was not based on detailed acoustical data and that they might differ from adult productions on a more fine-grained level.

Most of the studies on segment-tone integration discussed above focused on Chinese languages, whereas the present study focused on Vietnamese. The contrast could therefore possibly be due to language-specific patterns. However, as Tong et al. (2008) also argue, a preference for segments in word recognition makes sense as they are logically more distinctive: the languages in those studies, as well as Vietnamese in the present study, have more segmental contrasts than tonal contrasts. It would thus be expected that tonal speakers in general still prefer segments over tones in word recognition, as the former simply give more information about word identity. Further research on the preference for identification along the tonal rather than the segmental dimension in Vietnamese monolinguals might thus be necessary.

Considering the RT data in the segment-or-tone condition, the HMs are found to be the slowest, followed by the HSs, and finally the NLs. The RTs in this condition could indicate that the HMs and HSs have more difficulties than the NLs when ignoring one dimension to benefit another (see also Section 6.1 and the continued discussion in Section 6.3).

Lastly, for all groups, there was an effect of response button in this condition: there were relatively more choices along the segmental dimension when B matched X segmentally and A matched X tonally, than when A matched X segmentally and B matched X tonally. This suggests that the participants generally retained and matched information more easily when this information was close to the target: when B matches X segmentally, a segmental choice was most likely; when B matches X tonally, a tonal choice became (relatively) more likely. Moreover, when B matched X segmentally, the responses were faster than when the segmental match was with A (although for HMs this effect was not significant), reaffirming that matches, regardless of whether they be tonal or segmental, are easier to process when they are temporally closer.

6.3. Integration of segmental and tonal dimensions

Comparing the segment-and-tone condition to the forced-segment and forced-tone conditions reveals each group's respective segment-tone integration in speech processing. Considering the segment-and-tone condition first, the overall highest scores were found compared to other conditions. As in the forced-segment condition, the NLs were the most accurate (92.3%), followed by the HSs (84.2%), and lastly, the HMs (83.0%). It is clear that the HMs and HSs pattern together more closely with each other than with the NLs. However, the HSs patterned with the NLs with regards to RT. Both of these groups were faster than the HMs.

In the segment-and-tone condition, the HMs and HSs were faster than they were in any other condition, followed by the segment-or-tone condition, then the forced-segment condition, and finally the forced-tone condition. Thus, regardless of which dimension is suppressed, it always leads to some processing difficulty for the two Vietnamese-speaking groups. This means that the Vietnamese groups have an integrated perception of tones and segments. Note, however, that the HSs had considerably shorter RTs than the HMs in the forced-segment and forced-tone conditions, whereas RTs were more similar in the segment-and-tone condition. This could point to the HSs having a looser integration of these two dimensions. This result was expected and provides the first direct insight into HSs' perceptual segment-tone integration. However, note again that the HMs were consistently slower than the other groups in most conditions, which means this evidence is not yet conclusive.

In addition to this observation, it is clear that despite both dimensions being important for the HMs and HSs, the segmental dimension is the easiest to progress for both of these groups, i.e. the integration of tones and segments is asymmetrical, with a stronger weight attached to segments. These results conform to the results found for the Mandarin controls and Dutch learners of Mandarin familiar with tone in Zou et al. (2017), as well as Mandarin speakers in Tong et al. (2008), Braun & Johnson (2011), and Lin & Francis (2014) and Cantonese and Thai speakers in Yip et al. (1998) and Burnham et al. (2011), respectively.

However, this argument based on RTs can only be supported by response type data for the HSs, not for the HMs. The HSs performed very accurately in the segment-and-tone condition, where they can rely on both dimensions, but worse in the conditions where they can only rely on one dimension, with slightly higher scores in the forced-segment condition than in the forced-tone condition. For the HMs, the response type data also show lower accuracies in the forced-segment and forced-tone conditions compared to the segment-and-tone condition. Yet they performed slightly more accurately in the forced-tone condition than in the forced-segment condition, corresponding to their preference for tone-based word identification in the segment-or-tone condition. Thus, although the HMs, like the HSs, process segments faster and thus more *easily*, they do seem to process tone slightly more *accurately* than segments.

It is important to consider that the response type data, too, suggest a looser integration in HSs than in HMs. In these data, it is clearer that the two groups perform very similarly when there is no tonal or segmental mismatch (segment-and-tone condition), whereas the HSs clearly outperform the HMs in the forced-segment and forced-tone conditions. This indicates the HSs can ignore one dimension to accurately process the other more easily than the HMs.

Less importantly, there was an effect of response button for the HMs and HSs in the segment-and-tone condition: B-trials were more likely to get correct and fast answers than A-trials. As discussed, this effect might point to a general effect of participants finding it easier to retain and process information over a short temporal distance (see Section 6.4).

The NLs differ from the other groups in that they do not seem to have such a strong preference for A or B in the segment-and-tone condition. Additionally, for the NLs there are no significant differences between the segment-and-tone, segment-or-tone, and forced-segment conditions in RTs. The only condition causing significantly longer RTs is the forced-tone condition. This condition being the only one causing processing difficulties suggests that the NLs' perception of segments and tone might not be integrated to a meaningful extent. Instead, it indicates that the NLs are not sensitive to tone in word recognition and that they rely on the segmental dimension only, having great difficulty identifying non-words based on tones alone. These observations conform to findings on Dutch speakers without experience with a tonal language in Braun & Johnson (2011) and Zou et al. (2017), and similarly naïve English speakers in Burnham et al. (2011) and Lin & Francis (2014): these studies did not find considerable segment-tone integration in naïve non-tonal language speakers either.

The absence of sensitivity to tone in speakers of a non-tonal language can also be supported by the response type scores from the NLs. The NLs performed about equally well in the segment-and-tone and forced-segment conditions, where they can rely on the segmental dimension for word identification. The scores from the forced-tone condition are considerably lower, again indicating that the NLs do not perform worse because they have to ignore one dimension, but because they have more ease with the segmental dimension than with the tonal dimension.

Note, however, that the discussion in Section 6.1 suggests that NLs might have some segment-tone integration after all: in the segment-and-tone condition there was no effect of response button for this group, but in the forced-segment and segment-or-tone conditions, they were faster in B-trials than in A-trials. This was suggested to be due to the B-trials not allowing the NLs enough time to process tonal information and thus making it easier for them to ignore this information to the benefit of making a segment-based decision. In A-trials, where the NLs had enough time to process tone, they could not ignore this information and consequently took longer to make a segment-based decision.

6.4. Effects from control variables

As seen in Section 5, there were various effects from control variables, both on response type and on RTs. First of all, apart from the interaction effects mentioned in Sections 6.1, 6.2, and 6.3, there was a main effect of response button for both types of data: participants overall were less accurate and slower in A-trials than in B-trials, as was the case in Zou et al. (2017). As Zou and colleagues mention, this effect is common in ABX tasks (cf. Macmillan & Creelman 2005: 235). As the participants have to store the phonological information of both A and B standards in their working memory to listen which one matches with X best, it is possible that the participants lose more information on the initial A standard than on the second B standard, and consequently become less accurate and slower in responding whether A matches X or not.

The interaction effect of response button and condition mainly becomes clear in the RT data: for all conditions, except the forced-tone condition, responses were faster in B-trials. For the forced-tone condition, there was no statistically significant difference between A- and B-trials. As discussed, the preference for the B standard in the other conditions is expected. The faster responses in A-trials in the forced-tone condition, although not statistically significant, are thus remarkable. It is even more remarkable that only the HMs and NLs seem to cause this effect (see Figure 11, but note that this effect is only significant for the NLs). For the NLs, an explanation for this effect was already provided above: the tone dimension is likely difficult for them to progress due to their non-tonal L1, causing the NLs to respond more slowly to B standards as there is less time between these standards and X, than between A standards and X. This makes it difficult for the NLs to process tone in time to quickly match B with X. For the HMs, however, this explanation does not hold. For them, the faster responses in A could be more easily explained if the HMs had a preference for word identification along the segmental dimension. Then they might have had a more difficult time processing the tonal information when segmental information was lacking. That is, they would have had the same problem as the NLs. However, the results from the segment-or-tone condition suggest the HMs in fact prefer tone-based classification, hence this explanation does not hold. The HMs behaved in ways difficult to explain in more situations in the present study (see Section 6.1). Further research might thus help explain the unexpected results in this condition as well as other conditions (see also Section 6.5).

The main effect of X's tone on response type was not expected. Rather, in Section 4.2.4 it was stated that an interaction effect for the NLs was expected, but this was not found. Overall, participants were marginally more likely to give a correct response when X had a rising *sắc* tone than when it had a falling *nặng* tone. Perhaps this is due to the roles associated with markedness that high pitch fulfils in many languages. For instance, in stress languages high or rising pitch is often associated with stress (Gay 1978; Rietveld & Van Heuven 2013) and at least for Dutch speakers, it has been found that high pitch can lead to more prominence in perception (Streefkerk 2002: 88). Indeed, when some of the NLs were asked after participation what they suspected was the focus of the present study, they often referred to the forced-tone ABX sets as word sets that seemed to test "something about stress". The HSs might also perform more accurately with rising *sắc* tone targets because of high pitch causing prominence in Dutch, their dominant language. However, evidence from studies on Vietnamese, such as those by Brunelle & Jannedy (2013) and Kirby (2010), suggest that the rising *sắc* and falling *nặng* tones used in the present study are not very confusable for Southern Vietnamese listeners. At least for the HMs, and possibly for the HSs, the higher accuracy in rising *sắc* tone trials thus seems inconsistent with previous studies. One factor that could be of influence, is the frequency at which the two tones occur in Vietnamese. Based on an analysis of the Corpora of Vietnamese Texts (CVT; Pham et al. 2008), rising *sắc* seems to be more common (22.0% of all tones) than falling *nặng* (15.6%). This difference in frequency should be taken into account in future research.

Considering the last control variable, the effects from X's speakers are unexpected as well. As described in Section 5, the second female speaker's target tokens were associated with the highest accuracies. This speaker was originally judged by separate HM raters as sounding unlike Southern Vietnamese (see Section 4.2.2) and could thus have led to more confusion among HMs and HSs than among NLs. This clearly does not appear to be the case. The same tendency is found for the interaction effect of X's speakers with group on RTs: for all groups, the second female speaker elicited the shortest

RTs. It is therefore assumed that the inclusion of this speaker did not necessarily have a problematic effect on perception by either of the Vietnamese-speaking groups. It is strange, however, that the male speaker's tokens led to RTs that were disproportionately longer among NLs, but not among HMs or HSs. Arguably this effect could be due to this speaker being the only male speaker in the recordings. There is some evidence for men having generally smaller pitch ranges than women (e.g. Haan & Van Heuven 1999, but see this source and Simpson 2009 for discussions of the contrary case and how the idea of women's wider pitch ranges may be linked to stereotypes), which could lead naïve listeners such as the NLs to encounter more problems distinguishing tones. Yet, the male speaker in the present study had an average pitch excursion (in rising and falling tones) of 67.9 Hz, which is wider than that of the first female speaker (65.3 Hz). Only the second female speaker (whose productions led to the best performances for all groups) had a considerably larger average excursion of 101.6 Hz. Since the male speaker and the first female speaker are not that far apart in average excursions and since the male's pitch excursion is slightly wider on average, this range cannot be the only reason for the lower performance in trials with the male's productions. In future research, a pilot study should make sure this effect is not present. Time and financial constraints did not allow for using this precaution during the present study.

6.5. Limitations of the present study

There are a number of limitations to the present study. First, it should be reiterated that the response type data could not undergo a proper post-hoc analysis due to (i) the small sample size of the study in relation to the complexity of the model, making a post-hoc in the *multcomp* package impossible; and (ii) this data being unsuitable for other post-hoc tests like pairwise t-tests, since the raw data is binary and the transformed data (using by-subject means) is not normally distributed; (iii) the mixed effects logistic regression model created (using the *lme4* package) cannot be fed into many other types of post-hoc functions in R that could have replaced a post-hoc analysis in *multcomp*, such as TukeyHSD() from the pre-installed *stats* package. The lack of a post-hoc analysis for this model means that the discussion of the response type data is only provisional and needs to be backed up by further research. This future research should ideally have a larger sample size and prevent the need for many control variables, leading to a less complex model. Additionally, possible participants could be tested on their overall performance in a different task to make sure they meet a predetermined standard so that within-group responses are more homogeneous. However, it is important to bear in mind that this manipulation of the sample to prevent heterogeneity could give a warped representation of reality and should be done with caution, if at all.

The limited sample size in this study also meant that it was not possible to differentiate among HSs. As noted in Section 3.1, HSs are known to differ widely in proficiency in the HL and hence some HSs in the present study may also have performed considerably differently from others. Ideally, similar studies should have a large group of HSs that can be split up into subgroups based on proficiency or, as in Chang & Yao (2016), based on exposure to the HL from the community.

Next, as noted before, the HMs consistently performed less accurately and more slowly than the other groups. This was even the case in the segment-and-tone condition, where no considerable differences between groups were expected. One reason for this could be the fact that the experimenter was an L2 speaker of Vietnamese but did introduce the ABX task in Vietnamese. Results from Brunelle & Jannedy (2013) suggest that Vietnamese listeners' perception of Vietnamese tone may be influenced by the Vietnamese dialect they hear from experimenters. The experimenter undoubtedly had a noticeable foreign accent in Vietnamese and despite some familiarity with Southern Vietnamese, had influences from Northern Vietnamese as well. Considering Brunelle & Jannedy's findings, it is not unlikely that this may have had some effect on the HMs' performance. Therefore, ideally in future research, an L1 Southern Vietnamese speaker should conduct the experiment with the HMs, while the experiments with the HSs and NLs are conducted in their dominant language, Dutch. This ensures participants are introduced to the task in their respective dominant language. Additionally, to prevent this strong difference across groups as well as the discussed heterogeneity among participants across and within groups, the experiment could be preceded by a training task to help secure a good baseline performance.

Recall also that previous research found various variables that can have an impact on tone perception, such as experience with musical instruments (Lee & Hung 2008; Delogu et al. 2010), literacy, and education level (Burnham et al. 2011). Although data on some of these variables were collected through the sociolinguistic questionnaire in the present study, further subsetting groups, which each consisted of 35 participants at most, would reduce statistical power and thus make it harder to draw solid conclusions.

Furthermore, an intrinsic difficulty of researching segment-tone integration is that it is difficult to quantify how different two tones are compared to two segments. For example, do the stimuli *cá xin* and *cạ xin* differ in tone as much as *cá xin* and *tá phin* differ in segments? What makes either of these pairs more or less similar? If the tonal differences are somehow larger or smaller than the segmental differences, this would also have a confounding effect on the results and discussion above.

Next, it should be noted that the present study, although giving some indication of each group's respective sensitivity to and integration of tones and segments, only used two Vietnamese tones in the stimuli, whereas Southern Vietnamese has five tones in total. The results from this study consequently do have limitations for generalisation to tone perception in Vietnamese HSs, or monolingually raised Vietnamese and Dutch speakers in general. Insights regarding Vietnamese HSs' tone perception therefore remain limited.

Lastly, another worthwhile direction in future studies could be to consider different language pairs or dynamics in HSs: this study, like most of the studies on bilinguals of tonal languages mentioned in Section 2, deals with bilinguals who are dominant in a non-tonal Indo-European language and their performance in a tonal language. It would be interesting to see how HSs of two tonal languages with different tonal inventories - such as the Cantonese L1 speakers listening to Thai in Burnham et al. (2011) - compare to monolinguals of each respective language, or how these groups differ when the bilinguals are dominant in a tonal language and are HSs of a non-tonal language.

7. CONCLUSION

The aim of this thesis was to find out whether heritage speakers (HSs) of Southern Vietnamese in the Netherlands have an integrated perception of segments and tones, and how their integration compares to that in monolingually raised speakers of respectively Vietnamese (HMs) and Dutch (NLs). When language users have an integrated perception of segments and tones, they rely on both dimensions in word recognition and experience difficulties recognising words when they can only rely on one of these dimensions.

It was expected that all groups show some segment-tone integration, though this integration was assumed to be strongest in the HMs and weakest in the NLs, with the HSs' integration being somewhere in between. Moreover, for all groups, it was expected that their integration was asymmetrical: segments were assumed to be more important in word identification than tones.

These hypotheses were partly confirmed in an ABX task that manipulated tones and segments across conditions. The data showed that indeed all groups show some segment-tone integration in speech processing; HMs and HSs clearly had a stronger segment-tone integration than NLs, who showed very minimal integration. The HSs seemed to have a weaker integration than the HMs as well, as the former performed better than the HMs in conditions where participants can rely on only one dimension. These findings therefore conform to the more general literature on heritage language sound systems: when compared to HMs and to speakers unfamiliar with the heritage language, HSs tend to produce and perceive sounds in a way that is intermediate between the two other groups.

This tendency is also found for the groups' preferences for each dimension: the HMs, contrary to many similar groups in previous research, showed a slight preference for tone-based word recognition; the HSs had a preference for segment-based word identification and in the NLs this preference was very strong. The HSs are thus again 'in between' the other two groups.

Although limitations such as limited sample size and high variability in the response type data make it difficult to draw strong conclusions, the present study does contribute to the field of heritage linguistics by providing additional data on tone perception in HSs. Moreover, it is the first study to present a direct view into segment-tone integration in HSs' perception; whereas other studies on tone in HSs primarily focus on the accuracy with which they perceive different tones, the present study focused on how much attention HSs pay to tones compared to segments. The findings from the present study conform to the existing literature on heritage language speech representation and processing and additionally, contribute to research on segment-tone integration in general with data from Southern Vietnamese speakers. The monolingual HMs in the present study differ somewhat from many of the Mandarin- and Cantonese-speaking participants in previous studies, in that they do not show a preference for segment-based word identification, but rather a (slight) preference for tone-based word identification. This shows the need for more research, including research on different speaker populations.

In conclusion, the present study should prompt future research to investigate segment-tone integration and word processing in heritage speakers as well as in speakers with different language backgrounds. These projects should aim to research larger samples from populations that have not yet been extensively investigated, using paradigms that make the new data comparable to existing research. It would be worthwhile to examine whether other samples would show intermediate, asymmetrical segment-tone integration favouring segment perception in HSs of tone languages, or an asymmetrical integration benefiting tone perception in monolinguals of these languages, as found in the present study.

REFERENCES

Au, Terry Kit-Fong, Leah M. Knightly, Sun-Ah Jun & Janet S. Oh. 2002. Overhearing a language during childhood. *Psychological science* 13(3). 238-243.

Bailey, Charles-James N. 1973. *Variation and linguistic theory*. Arlington, VA: Center for Applied Linguistics.

Bailey, Charles-James N. 1974. Some suggestions for greater consensus in creole terminology. In David DeCamp & Ian Hancock (eds.), *Pidgins and creoles: Current trends and prospects,* 88-91. Washington, DC: Georgetown University Press.

Barton, Kamil. 2019. *MuMIn: Multi-Model Inference*, version 1.43.6.

Bates, Douglas, Martin Maechler, Ben Bolker & Steve Walker. 2015. Fitting linear mixed-effects models using lme4 [version 1.1-21]. *Journal of Statistical Software* 67(1). 1-48.

Benmamoun, Elabbas, Silvina Montrul & Maria Polinsky. 2013. Heritage languages and their speakers: Opportunities and challenges for linguistics. *Theoretical Linguistics* 39(3-4). 129-181.

Bickerton, Derek. 1973. The nature of a creole continuum. *Language* 49(3). 640-669.

Bickerton, Derek. 1975. *Dynamics of a creole continuum*. Cambridge, MA: Cambridge University Press.

Birdsong, David, Libby M. Gertken & Mark Amengual. 2012. *Bilingual language profile: An easy-to-use instrument to assess bilingualism*. Austin, TX: COERLL, University of Texas. https://sites.la.utexas.edu/bilingual/. (19 February 2019.)

Boersma, Paul. 2018. The history of the Franconian tone contrast. In Wolfgang Kehrein, Björn Köhnlein, Paul Boersma & Marc van Oostendorp (eds.), *Segmental structure and tone*, 27-98. Boston: Walter De Gruyter.

Boersma, Paul & David Weenink. 2019. Praat: Doing phonetics by computer, version 6.0.46. http://www.praat.org.

Braun, Bettina & Elizabeth K. Johnson. 2011. Question or tone 2? How language experience and linguistic function guide pitch processing. *Journal of Phonetics* 39(4). 585-594.

Brunelle, Marc. 2009. Tone perception in Northern and Southern Vietnamese. *Journal of Phonetics* 37(1). 79-96.

Brunelle, Marc & Stefanie Jannedy. 2013. The cross-dialectal perception of Vietnamese tones: Indexicality and convergence. In Daniel Hole & Elisabeth Löbel (eds.), *The linguistics of Vietnamese: An international survey*, 9-34. Berlin/Boston: De Gruyter Mouton.

Burnham, Denis, Jeesun Kim, Chris Davis, Valter Ciocca, Colin Schoknecht, Benjawan Kasisopa & Sudaporn Luksaneeyanawin. 2011. Are tones phones? *Journal of Experimental Child Psychology* 108(4). 693-712.

Centraal Bureau voor de Statistiek (CBS). 2018. *Bevolking; generatie, geslacht, leeftijd en migratieachtergrond, 1 januari* [Population; generation, sex, age and migration background, January 1]. https://opendata.cbs.nl/statline/#/CBS/nl/dataset/37325/table?ts=1548256007635. (January 23 2019.)

Chang, Charles B. 2016. Bilingual perceptual benefits of experience with a heritage language. *Bilingualism: Language and Cognition* 19(4). 791-809.

Chang, Charles B., Yao Yao, Erin F. Haynes & Russell Rhodes. 2011. Production of phonetic and phonological contrast by heritage speakers of Mandarin. *The Journal of the Acoustical Society of America* 129(6). 3964-3980.

Chang, Charles B. & Yao Yao. 2016. Toward an understanding of heritage prosody: Acoustic and perceptual properties of tone produced by heritage, native, and second language speakers of Mandarin. *Heritage Language Journal* 13(2). 134-160.

Chen, Yiya. 2011. How does phonology guide phonetics in segment-*f0* interaction? *Journal of Phonetics* 39. 612-625.

Ciocca, Valter & Jessica Lui. 2003. The development of the perception of Cantonese lexical tones. *Journal of Multilingual Communication Disorders* 1(2). 141-147.

Cutler, Anne & Hsuan-Chih Chen. 1997. Lexical tone in Cantonese spoken-word processing. *Perception & Psychophysics* 59(2). 165-179.

Dai, Jin-Huei Enya & Lihua Zhang. 2008. What are the CHL learners inheriting? Habitus of the CHL learners. In Agnes Weiyun He & Yun Xiao (eds.), *Chinese as a heritage language: Fostering rooted world citizenry,* 37-51. Honolulu, HI: National Foreign Language Resource Center, University of Hawaii.

Đào Mục Đích. 2013. *Phân tích một vài đặc điểm thanh điệu tiếng Việt của người Úc trẻ gốc Việt (ứng dụng cho việc dạy Tiếng Việt)* [Analysing some characteristics of the Vietnamese tones produced by young Vietnamese Australian people (for the teaching of the Vietnamese language)]. *Science & Technology Development* 16(3). 26-33.

De Valk, Helga A.G., Ingrid Esveldt, Kène Henkens & Aart C. Liefbroer (see Valk).

Dehé, Nicole. 2018. The intonation of polar questions in North American ("heritage") Icelandic. *Journal of Germanic Linguistics* 30(3). 213-259.

Delogu, Franco, Giulia Lampis & Marta Olivetti Belardinelli. 2010. From melody to lexical tone: Musical ability enhances specific aspects of foreign language perception. *European Journal of Cognitive Psychology* 22(1). 46-61.

Ebenau, Jerom. 2017. *Overextended labialisation of coda consonants: The case of Vietnamese heritage speakers in the Netherlands*. Leiden: Unpublished Bachelor's thesis.

Elvira García, Wendy. 2018. Create pictures with tiers v.4.5.: Praat script. http://stel.ub.edu/labfon/en/praat-scripts. (25 October 2019.)

Faez, Farahnaz. 2011. Reconceptualizing the native/nonnative speaker dichotomy. *Journal of Language, Identity & Education* 10(4). 231-249.

Fournier, Rachel Agnès. 2008*. Perception of the tone contrast in East Limburgian dialects*. Utrecht: LOT publications.

Francis, Alexander L., Valter Ciocca, Lian Ma & Kimberly Fenn. 2008. Perceptual learning of Cantonese lexical tones by tone and non-tone language speakers. *Journal of Phonetics* 36(2). 268-294.

Gandour, Jack, Mario Dzemidzic, Donald Wong, Mark Lowe, Yunxia Tong, Li Hsieh, Nakarin Satthamnuwong & Joseph Lurito. 2003. Temporal integration of speech prosody is shaped by language experience: An fMRI study. *Brain and Language* 84(3). 318-336.

Garner, Wendell R. 1970. The stimulus in information processing. *American Psychologist* 25. 350-358.

Garner, Wendell R. 1974. *The processing of information and structure.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Garner, Wendell R. 1976. Interaction of stimulus dimensions in concept and choice processes. *Cognitive Psychology* 8(1). 98-123.

Gay, Thomas. 1978. Physiological and acoustic correlates of perceived stress. *Language and Speech* 21(4). 347-353.

Goldsmith, John. 1976. *Autosegmental phonology*. London: Doctoral dissertation, MIT Press.

Gussenhoven, Carlos. 2000. On the origin and development of the Central Franconian tone contrast. In Aditi Lahiri (ed.), *Trends in linguistics: Analogy, levelling, markedness: Principles of change in phonology and morphology,* 215-260. Berlin/New York: Mouton de Gruyter.

Gussenhoven, Carlos & Jörg Peters. 2008. De tonen van het Limburgs [The tones of Limburgian]. *Nederlandse Taalkunde* 13. 88-115.

Haan, Judith & Vincent J. van Heuven. 1999. Male vs. female pitch range in Dutch questions. *14th International Congress of Phonetic Sciences* [ICPhS-14]. 1581-1584.

Harrison, Phil. 2000. Acquiring the phonology of lexical tone in infancy. *Lingua* 110(8). 581-616.

Hoeven, Erik van der & Henk de Kort. 1983. *Over Vietnamezen in Nederland: Een beschrijving van 720 Vietnamese vluchtelingen* [On Vietnamese in the Netherlands: A description of 720 Vietnamese refugees]. Den Haag: Coördinatiecommissie wetenschappelijk onderzoek kinderbescherming.

Hoot, Bradley. 2012. *Presentational focus in heritage and monolingual Spanish.* Chicago, IL: Dissertation, University of Illinois.

Hothorn, Torsten, Frank Bretz & Peter Westfall. 2008. Simultaneous inference in general parametric models. *Biometrical Journal* 50(3). 346-363.

Hu, Jiehui, Shan Gao, Weiyi Ma & Dezhong Yao. 2012. Dissociation of tone and vowel processing in Mandarin idioms. *Psychophysiology* 49(9). 1179-1190.

Kan, Rachel T.Y. & Monika S. Schmid. 2019. Development of tonal discrimination in young heritage speakers of Cantonese. *Journal of Phonetics* 73. 40-54.

Kim, Ji-Young. 2015. Perception and production of Spanish lexical stress by Spanish heritage speakers and English L2 learners of Spanish. *Selected Proceedings of the 6th Conference on Laboratory Approaches to Romance Phonology* [LARP 6]. 106-128.

Kirby, James. 2010. Dialect experience in Vietnamese tone perception. *The Journal of the Acoustical Society of America* 127(6). 3749-3757.

Kirby, James P. 2011. Vietnamese (Hanoi Vietnamese). *Journal of the International Phonetic Association* 41(3). 381-392.

Kleinen, John. 1988. *Vietnamezen in Nederland* [Vietnamese in the Netherlands]. Rijswijk: Ministerie van Welzijn, Volksgezondheid en Cultuur.

Knightly, Leah M., Sun-Ah Jun, Janet S. Oh & Terry Kit-Fong Au. 2003. Production benefits of childhood overhearing. *The Journal of the Acoustic Society of America* 114(1). 465–474.

Köhnlein, Björn. 2016. Contrastive foot structure in Franconian tone-accent dialects. *Phonology* 33(1). 87-123.

Kuhl, Patricia K. 1983. Perception of auditory equivalence classes for speech in early infancy. *Infant Behavior and Development* 6(2-3). 263-285.

Lam, Beevi Mariam. 2006. The cultural politics of Vietnamese language pedagogy. *Journal of Southeast Asian Language Teaching* 12(2). 1-19.

Lam, Wai Man. 2018. *Perception of lexical tones by homeland and heritage speakers of Cantonese.* Vancouver: Dissertation, University of British Columbia.

Lee, Lisa & Howard C. Nusbaum. 1993. Processing interactions between segmental and suprasegmental information in native speakers of English and Mandarin Chinese. *Perception & Psychophysics* 53(2). 157-165.

Lee, Chao-Yang & Tsun-Hui Hung. 2008. Identification of Mandarin tones by English-speaking musicians and nonmusicians. *The Journal of the Acoustical Society of America* 124(5). 3235-3248.

Liberman, Alvin M., Katherine Safford Harris, Howard S. Hoffman & Belver C. Griffith. 1957. The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology* 54(5). 358-368.

Lin, Mengxi & Alexander L. Francis. 2014. Effects of language experience and expectations on attention to consonants and tones in English and Mandarin Chinese. *The Journal of the Acoustical Society of America* 136(5). 2827-2838.

Liu, Liquan & René Kager. 2014. Perception of tones by infants learning a non-tone language. *Cognition* 133(2). 385-394.

Lukyanchenko, Anna & Kira Gor. 2011. Perceptual correlates of phonological representations in heritage speakers and L2 learners. *Proceedings of the 35th Annual Boston University Conference on Language Development* [BUCLD 35] vol. 2. 414-426.

Ma, Weiyi, Peng Zhou, Leher Singh & Liqun Gao. 2017. Spoken word recognition in young tone language learners: Age-dependent effects of segmental and suprasegmental variation. *Cognition* 159. 139-155.

Macmillan, Neil A. & C. Douglas Creelman. 2005. *Detection theory: A user's guide*, 2nd edn. Mahwah, N.J.: Lawrence Erlbaum Associates.

Maddieson, Ian. 2013. Tone. In Matthew S. Dryer & Martin Haspelmath (eds.), *The world atlas of language structures online.* Leipzig: Max Planck Institute for Evolutionary Anthropology. http://wals.info/chapter/13. (29 January 2019.)

Mai, A. Tran Huong. 1967. Tones and intonation in South Vietnamese. *Pacific Linguistics: Series A: Occasional Papers* 9. 19-34.

Mattock, Karen & Denis Burnham. 2006. Chinese and English infants' tone perception: Evidence for perceptual reorganization. *Infancy* 10(3). 241-265.

Mattock, Karen, Monika Molnar, Linda Polka & Denis Burnham. 2008. The developmental course of lexical tone perception in the first year of life. *Cognition* 106(3). 1367-1381.

Michaud, Alexis. 2004. Final consonants and glottalization: New perspectives from Hanoi Vietnamese. *Phonetica* 61(2). 119-146.

Montrul, Silvina. 2016. *The acquisition of heritage languages.* Cambridge: Cambridge University Press.

Montrul, Silvina. 2018. The Bottleneck Hypothesis extends to heritage language acquisition. In Jacee Cho, Michael Iverson, Tiffany Judy, Tania Leal & Elena Shimanskaya (eds.), *Meaning and structure in Second Language Acquisition: In honor of Roumyana Slabakova*, 149-178. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Moro, Francesca Romana. 2016. *Dynamics of Ambon Malay: Comparing Ambon and the Netherlands.* Utrecht: LOT Publications.

Nakagawa, Shinichi & Holger Schielzeth. 2013. A general and simple method for obtaining R2 from generalized linear mixed-effects models. *Methods in ecology and evolution* 4(2). 133-142.

Nguyen, Hanh Thi & Marlys A. Macken. 2008. Factors affecting the production of Vietnamese tones: A study of American learners. *Studies in Second Language Acquisition* 30(1). 49-77.

Oh, Janet S., Sun-Ah Jun, Leah M. Knightly & Terry Kit-Fong Au. 2003. Holding on to childhood language memory. *Cognition* 86(3). B53-B64.

Pham, Hoa T. 2001. *Vietnamese tone: Tone is not pitch.* Ottawa, ON: Dissertation, National Library of Canada.

Pham, Andrea Hoa. 2003. *Vietnamese tone: A new analysis.* New York: Routledge.

Phạm, Ben & Sharynne McLeod. 2016. Consonants, vowels and tones across Vietnamese dialects. *International Journal of Speech-Language Pathology* 18(2). 122-134.

Pham, Giang, Kathryn Kohnert & Edward Carney. 2008. Corpora of Vietnamese texts: Lexical effects of intended audience and publication place. *Behavior research methods* 40(1). 154-163.

Pham, Giang, Kerry Danahy Ebert, Kristine Thuy Dinh & Quynh Dam. 2018. Non-word repetition stimuli for Vietnamese-speaking children. *Behavior Research Methods* 50(4). 1311-1326.

Polinsky, Maria & Olga Kagan. 2007. Heritage languages: In the 'wild' and in the classroom. *Language and Linguistics Compass* 1(5). 368-395.

Queen, Robin M. 2001. Bilingual intonation patterns: Evidence of language change from Turkish-German bilingual children. *Language in Society* 30(1). 55-80.

R Core Team. 2019. *R: A language and environment for statistical computing.* R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Rietveld, Antonius C.M. & Viencent J. Van Heuven. 2013. *Algemene fonetiek* [General phonetics], 3rd edn. Bussum: Uitgeverij Coutinho.

Robles-Puente, Sergio. 2014. *Prosody in contact: Spanish in Los Angeles.* Los Angeles, CA: Dissertation, University of Southern California.

Rothman, Jason & Jeanine Treffers-Daller. 2014. A prolegomenon to the construct of the native speaker: Heritage speaker bilinguals are natives too! *Applied linguistics* 35(1). 93-98.

Schmid, Monika S. 2013. First language attrition. *Linguistic Approaches to Bilingualism* 3(1). 94-115.

Shea, Christine. 2017. Dominance, proficiency, and Spanish heritage speakers' production of English and Spanish vowels. *Studies in Second Language Acquisition*. 1-27.

Simpson, Adrian P. 2009. Phonetic differences between male and female speech. *Language and Linguistics Compass* 3(2). 621-640.

Singh, Leher & Joanne Foong. 2012. Influences of lexical tone and pitch on word recognition in bilingual infants. *Cognition* 124(2). 128-142.

Singh, Leher, Tam Jun Hui, Calista Chan & Roberta Michnick Golinkoff. 2014. Influences of vowel and tone variation on emergent word knowledge: A cross-linguistic investigation. *Developmental Science* 17(1). 94-109.

Singh, Leher, Hwee Hwee Goh & Thilanga D. Wewalaarachchi. 2015. Spoken word recognition in early childhood: Comparative effects of vowel, consonant and lexical tone variation. *Cognition* 142. 1-11.

Singh, Leher & Charlene S. L. Fu. 2016. A new view of language development: The acquisition of lexical tone. *Child Development* 87(3). 834-854.

So, Kwok Lai Connie. 2000. *Tonal production and perception patterns of Canadian raised Cantonese speakers.* Burnaby: Master's thesis, Simon Fraser University.

Soo, Rachel & Philip J. Monahan. 2017. Language exposure modulate the role of tone in perception and long-term memory: Evidence from Cantonese native and heritage speakers. *Proceedings of the forty-third annual meeting of the Berkeley Linguistics Society* 2, 47-54*.* Berkeley, CA: Berkeley Linguistics Society.

Stangen, Ilse, Tanja Kupisch, Anna Lia Proietti Ergün & Marina Zielke. 2015. Foreign accent in heritage speakers of Turkish in Germany. In Hagen Peukert (ed.), *Transfer effects in multilingual language development*, 87-108. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Streefkerk, Barbertje M. 2002. *Prominence: Acoustic and lexical/syntactic correlates*. Utrecht: LOT Publications.

Thompson, Laurence C. 1965. *A Vietnamese grammar.* Seattle: University of Washington Press.

Tong, Yunxia, Alexander L. Francis & Jackson T. Gandour. 2008. Processing dependencies between segmental and suprasegmental features in Mandarin Chinese. *Language and Cognitive Processes* 23(5). 689-708.

Tong, Xiuhong, Catherine McBride, Chia-Ying Lee, Juan Zhang, Lan Shuai, Urs Maurer & Kevin K.H. Chung. 2014. Segmental and suprasegmental features in speech perception in Cantonese-speaking second graders: An ERP study. *Psychophysiology* 51(11). 1158-1168.

Tsao, Feng-Ming. 2008. The effect of acoustical similarity on lexical-tone perception of one-year-old Mandarin-learning infants. *Chinese Journal of Psychology* 50(2). 111-124.

Valk, Helga A.G. de, Ingrid Esveldt, Kène Henkens & Aart C. Liefbroer. 2001. *Oude en nieuwe allochtonen in Nederland: een demografisch profiel* [Old and new migrants in the Netherlands: a demographic profile]. Den Haag: Wetenschappelijke Raad voor het Regeringsbeleid.

Van Der Hoeven, Erik & Henk de Kort (see Hoeven).

Wang, Yue & Patricia K. Kuhl. 2003. Evaluating the "Critical Period" Hypothesis: Perceptual learning of Mandarin tones in American adults and American children at 6, 10 and 14 years of age. *15th International Congress of Phonetic Sciences* [ICPhS-15]. 1537-1540.

Werker, Janet F. & Richard C. Tees. 1984. Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. *Infant behavior and development* 7(1). 49-63.

Werker, Janet F. & Takao K. Hensch. 2015. Critical periods in speech perception: New directions. *Annual review of psychology* 66. 173-196.

Wewalaarachchi, Thilanga D., Liang Hui Wong & Leher Singh. 2017. Vowels, consonants, and lexical tones: Sensitivity to phonological variation in monolingual Mandarin and bilingual English-Mandarin toddlers. *Journal of Experimental Child Psychology* 159. 16-33.

Yang, Bei. 2015. *Perception and production of Mandarin tones by native speakers and L2 learners.* Berlin: Springer.

Ye, Yun & Cynthia M. Connine. 1999. Processing spoken Chinese: The role of tone information. *Language and Cognitive Processes* 14(5-6). 609-630.

Yeung, H. Henny, Ke Heng Chen & Janet F. Werker. 2013. When does native language input affect phonetic perception? The precocious case of lexical tone. *Journal of Memory and Language* 68(2). 123-139.

Yip, Michael C.W., Po-Yee Leun & Hsuan-Chih Chen. 1998. Phonological similarity effects in Cantonese spoken-word processing. In Robert H. Mannell & Jordi Robert-Ribes (eds.), *Fifth International Conference on Spoken Language Processing* [ICSLP-1998], paper 0661. Sydney: Australian Speech Science and Technology Association.

Yip, Moira. 2002. *Tone*. Cambridge: Cambridge University Press.

Yip, Moira. 2007. Tone. In Paul De Lacy (ed.), *The Cambridge Handbook of Phonology*, 229-252. Cambridge: Cambridge University Press.

Zou, Ting, Yiya Chen & Johanneke Caspers. 2017. The developmental trajectories of attention distribution and segment-tone integration in Dutch learners of Mandarin tones. *Bilingualism: Language and Cognition* 20(5). 1017-1029.

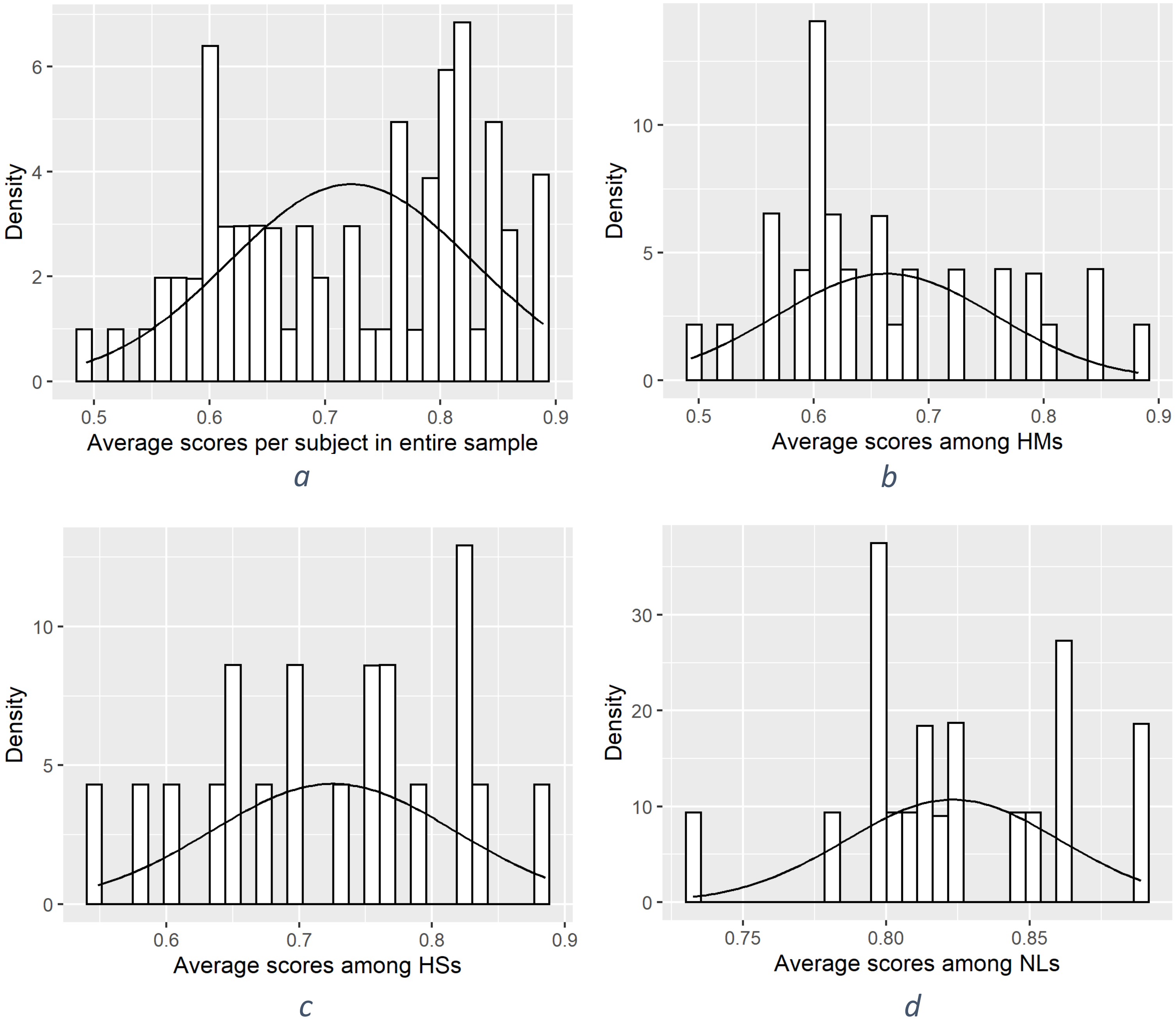
APPENDIX A: NON-WORDS USED AS STIMULI

The non-word pairs are provided here as IPA transcriptions (without tone):

[ka sin]	[ta fin]
[fa lun]	[sa run]
[ti lon]	[ki ron]
[fi mun]	[hi nun]
[fu kam]	[su tam]
[ju kom]	[lu tom]
[mo kim]	[no tim]
[nu fam]	[mu ham]
[ko tan]	[to kan]

APPENDIX B: HISTOGRAMS FROM RESPONSE TYPE DATA

Histograms of by-participant percentages of correct trials (as decimals between 0 and 1) in the entire sample (a), among the HMs (b), the HSs (c), and the NLs (d); in the segment-and-tone condition (e), the segment-or-tone condition (f), the forced-segment condition (g), and the forced-tone condition (h). Note that Density on the y-axis stands for the number of participants with the same mean. Bars may align with a decimal number as not all participants had the same number of trials included in the analysis (due to missing answers etc.).

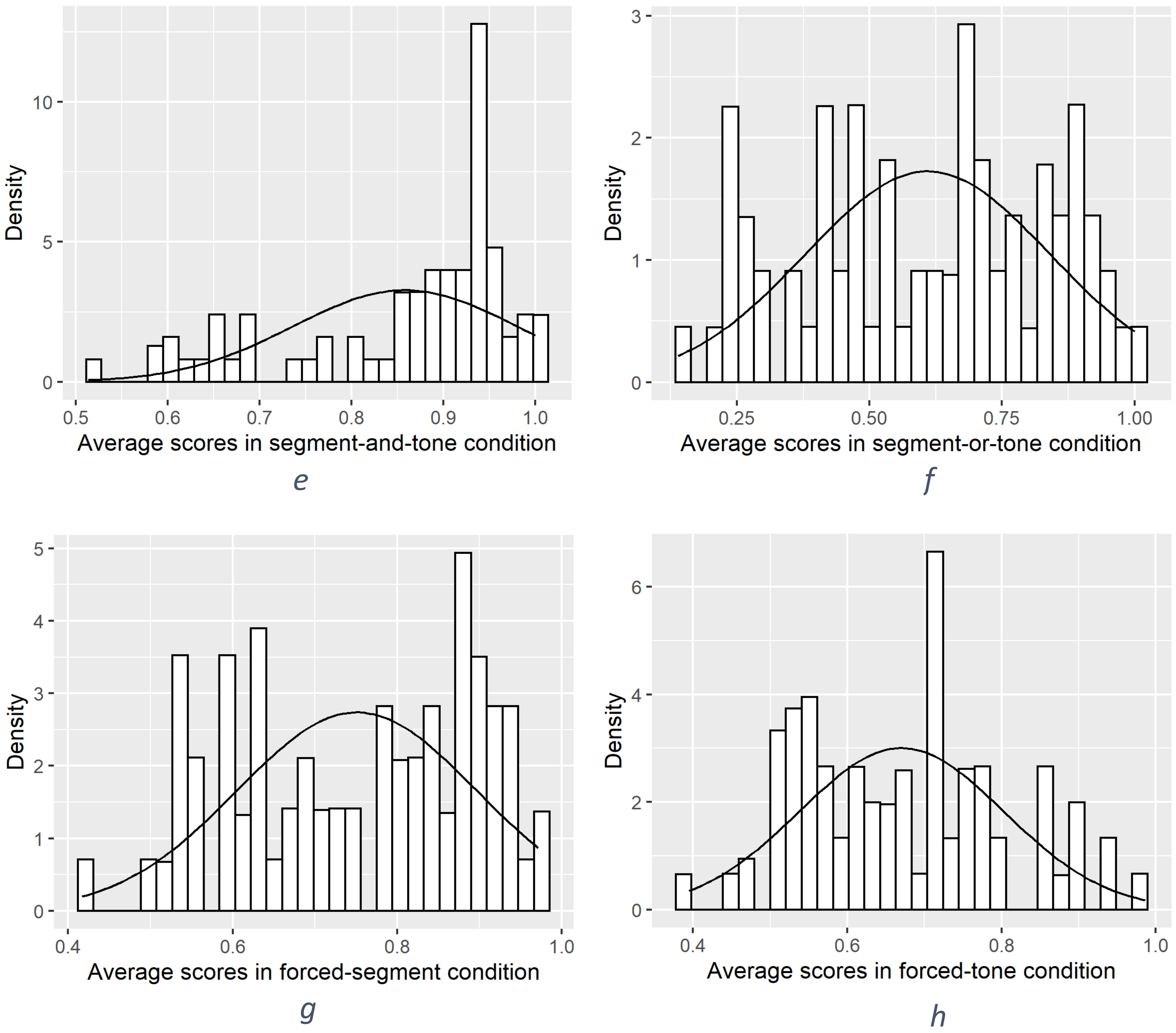


a

b

c

d


Density
10
5
0
0.5
0.6
0.7
0.8
0.9
1.0
Average scores in segment-and-tone condition
e
Density
3
2
1
0
0.25
0.50
0.75
1.00
Average scores in segment-or-tone condition
f
Density
5
4
3
2
1
0
0.4
0.6
0.8
1.0
Average scores in forced-segment condition
g
Density
6
4
2
0
0.4
0.6
0.8
1.0
Average scores in forced-tone condition
h