Cross-dialectal perception of Vietnamese tones

Exploring the nuances of the hỏi-ngã *merger*

Jeremiah S. Jewell
Advisor: Claire L. Bowern

Department of Linguistics
Yale University

A thesis presented for the degree of
Bachelor of Arts, MMXXII

Acknowledgments

This work did not emerge in a vacuum. I am particularly indebted to

…my brilliant advisor Claire Bowern, who inspires (and puts up with) me.

…giáo sư của em, thầy Phong Văn Quang (cũng cô Võ Thư), vì dạy em gần như tất cả mà em biết về tiếng Việt.

…Linh Lê, for graciously lending his voice, knowledge, and time, without whom this thesis could never have materialized.

…the Vietnamese Student Association at Yale and its members, for helping my study recruit from the broadest range of participants possible.

…all those who've taught me linguistics at Yale, including Claire (again!), Jason Shaw, Veneeta Dayal, Larry Horn, Jim Wood, and Natalie Weber, inter alios.

…my peers- and mentors-cum-friends currently or previously in the ling department and Pama-Nyungan lab, whose camaraderie and guidance has carried me through so much—especially, Nico Kidd, Lena "Lezekiel Venkatrekiel" Venkatraman, Sarah Ba(balabal)inski, Jack M. K. Sullivan, Irene "omg!" Yi, Josh Phillips, Randi Martínez, Martín Fuchs, Andy Zhang, Ushasi Bandyopadhyay, Milena Šereikaitė, and Roslyn Burns.

…other friends, both at Yale and elsewhere, including

- …e13 crew: Anna Trần, Brian Isaacs, & Alexis Teh.
- …poors: Pilar "Bug" Forrest, Carli "Barley" Roush, etc.
- …homies: Noah Johnston, Olivia Dennison, Octavia Perez, and Emma Claire Pinches.
- …miscellaneous: Hanna Ehrlich, MK Speth, Jono Field, Gesa Wortmann, and Matthias Theis, among countless others.

…my parents, sisters, and other relatives.

…the late and great SOPHIE, whose crisp and manic tracks fueled many a sleepless night's work on this thesis.

…chot & grimalkin.

Many thanks to those listed above. It takes a village to raise a thesis.

Contents

Abstract

Cross-dialectal speech perception represents a fertile ground for research into the nature of sound mergers. A typical feature of Southern Vietnamese dialects is the lack of a phonetic distinction between two tones which are contrastive in Northern Vietnamese dialects. The purpose of this thesis is to investigate how Vietnamese speakers with this tonal merger perceive these tones when produced by non-merging speakers in semantically enriched contexts. In so doing, this study also seeks to shed light on the nature of the merger itself and the mechanism(s) by which it developed (or is developing).

To this end, I engage native Vietnamese speakers in a study where they were prompted with audio stimuli consisting of sentences spoken by a non-merging Northern Vietnamese speaker and asked to identify the tone of one of the words in each sentence. The results of the experiment demonstrate that speakers who merge these tones in production generally perform better than chance in accurately identifying the tones when uttered by a non-merging speaker, although they are still outperformed by speakers who natively produce this contrast. My findings also indicate a great deal of variation between merging speakers in their ability to accurately discern the tones in question, ranging from complete competence to none at all. Moreover, the results of this study clearly show that sentential context plays a major role in structuring cross-dialectal perception of Vietnamese tone.

1 Introduction

IETNAMESE, like many other Southeast Asian languages, is tonal. Depending on the dialect, up to six contrastive lexical tones can be identified within Vietnamese. While Northern dialects typically make a six-way tone distinction, Southern dialects are reported to merge two of these tones, resulting in a five-way distinction (Phạm 2003, Thompson 1965, Trần 1967, Vũ 1982). Previous studies have focused on describing cross-dialectal variation in Vietnamese tone *production* (Alves 2007, Bauman et al. 2009, Brunelle 2009a, Phạm 2003, Thompson 1965, Trần 1967, Vũ 1982, Đào and Nguyễn 2017), but less attention has been paid to the cross-dialectal *perception* of tones and tonal contrasts, with the most substantial contributions in this area being those of Brunelle (2009b), Brunelle and Jannedy (2013), Kirby (2010), and Vũ (1981). The current study aims to contribute to a greater understanding of cross-dialectal tone perception in Vietnamese by examining how speakers of Southern Vietnamese perceive the contrast between what are called the *hỏi* and *ngã* tones, a contrast which they do not natively produce.

The most basic question I try to answer in this paper surrounds cross-dialectal perception of the HỎI/C1 and NGÃ/C2 tones. Specifically: can HỎI/C1-NGÃ/C2 merging speakers perceive the HỎI/C1-NGÃ/C2 tonal contrast when produced by a non-merging speaker in sentential environments that ambiguously prime for either of these tones? My treatment will allow for the separation and independent analysis of the effects of not only auditory form, but also syntacto-semantic context in influencing cross-dialectal tone perception in Vietnamese, representing a novel approach not taken in previous attempts at answering this question.

More broadly, this work seeks to uncover hitherto unexamined nuances of the Southern Vietnamese tone merger. Mergers are not all of the same basic type, and the HỎI/C1-NGÃ/C2 merger has not yet been, to my knowledge, categorized with any degree of specificity greater than simply as a "merger". As such, one additional aim of this paper is to provide insight into which *sort* of merger the HỎI/C1-NGÃ/C2 one ought to be considered, as well as whether or not the answer to this question is demonstrably consistent across merging speakers.

Furthermore, this study seeks to produce greater insight into the sociolinguistic dimensions along which the tone merger under question is or is not realized, both in production and perception. Northern Vietnamese is considered a prestige dialect (Thompson 1965: 3), and speakers of other dialects have been noted to introduce non-native, Northern dialect features in their own speech in some contexts (Brunelle and Jannedy 2013: 15). Several sociological factors have been demonstrated to affect language perception, including (but not limited to) age (Drager 2010, Koops et al. 2008), socio-economic status (Hay et al. 2006, Maclagan and Gordon 1996), nationality (Niedzielski 1999), and exposure to a given dialect (Evans and Iverson 2004, Shaw et al. 2018, Sumner and Samuel 2009). As such, it would not be surprising to find that demographic factors like these correlate with speakers of Southern Vietnamese's ability to discern the non-native HỎI/C1-NGÃ/C2 tonal contrast when pronounced by a speaker of a non-merged dialect.

The rest of this thesis is structured as follows. In §2, I review relevant literature on Vietnamese linguistics, lexical processing, cross-dialectal speech perception, and mergers. In §3, I establish in greater detail the research questions this study aims to answer, and in §4, I lay out how I investigate these research questions. Then, in §5, I report on the results of my study. In §6, I discuss these results in the context of related literature and the research questions I pose. Finally, in §7, I provide concluding remarks.

2 Background

In this section, I dedicate space to providing pertinent background information that the crux of my argument will hinge on. The section is split into four parts, with §2.1 opening to an introduction

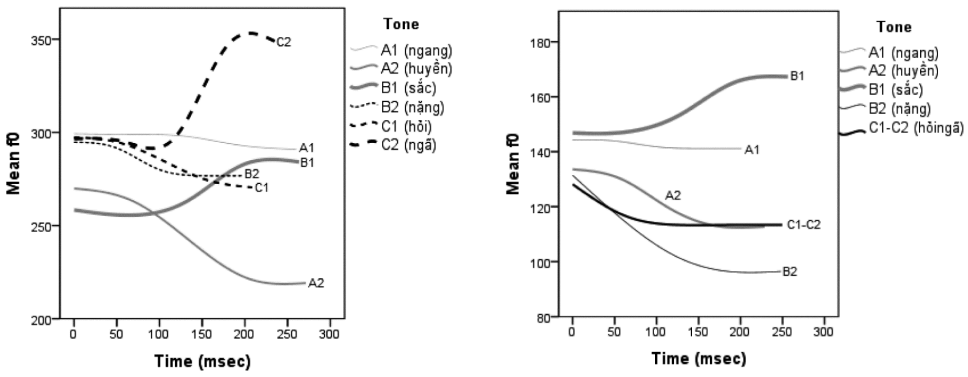


Figure 1. Pitch contours of a Northern (left) and Southern Vietnamese speaker (right). Figures published in Brunelle and Jannedy (2013: 16–17).

on Vietnamese and its tonal typology, §2.2 dealing with theories of speech cognition and lexical processing, §2.3 providing information on the the nuances of cross-dialectal speech perception in particular, and §2.4 presenting an overview on the different kinds of phonological mergers.

2.1 Tone in Vietnamese

Vietnamese distinguishes six tones, each represented orthographically by a distinct diacritic placed by the vowel. The traditional Vietnamese names for these tones are *ngang*, *huyền*, *sắc*, *nặng*, *hỏi*, and *ngã*, though in literature on Vietnamese tones, somewhat of a standard has emerged of referring to them as A1, A2, B1, B2, C1, and C2, respectively (Brunelle and Jannedy 2007, Michaud 2005, Phạm 2003). While pitch contours are not the only cues of Vietnamese tones (see Alves 1995, Phạm 2003, Vũ 1981), fundamental frequency nonetheless does much of the heavy lifting. Typical Northern and Southern Vietnamese pitch contours for each tone are shown in Figure 1.

The tones of Vietnamese are traditionally divided along two axes: **register**, consisting of *high* and *low*, and **contour**, consisting of *level*, *fall/rise*, and *falling-rising* (Cao 2003, Phạm 2003). This is shown in Table 1. This taxonomy captures the most basic phonetic differences between Vietnamese's six tones and is supported both in diachronic analyses of tonogenesis in Vietnamese (Diffloth 1989, Haudricourt 1954, Thurgood 2002) and synchronic accounts of Vietnamese tone sandhi (Nguyễn and Ingram 2008, 2006, Phạm and Phạm 2020). There have been several attempts to describe Vietnamese tones in terms of distinctive features (Alves 1995, Ngô 1984: 110, Thompson 1965: 20), but little consensus has emerged from these endeavors (see Phạm 2003: 8–16 for a more exhaustive overview).

contour →		bằng (even)		trắc (oblique)			
register ↓		bình (level)	**A**	nhập (entering)	**B**	thượng (rising)	**C**
phù (high)	**1**	ngang (flat)	**A1**	sắc (sharp)	**B1**	hỏi (asking)	**C1**
trầm (low)	**2**	huyền (deep)	**A2**	nặng (heavy)	**B2**	ngã (tumbling)	**C2**

Table 1. Typology of Vietnamese tones, with Vietnamese names and English translations (Cao 2003: 82). Tone numbers from Michaud (2005) shown in bold.

2.1.1 Dialectal variation

Vietnamese consists of many different dialects that mostly form north-to-south distributions along the length of the country. In broad terms, dialects are usually grouped into two or three main clusters: Northern, (sometimes) Central, and Southern (Phạm 2003, Vũ 1982). Among these, Central Vietnamese dialects are the most distinct, though also the least widely spoken and most poorly studied (Alves 2007, Thompson 1965). The North-Central dialects, typified by the speech of the Thanh Hoá, Nghệ An, and Quảng Bình provinces, have been argued to represent a fourth major division, due to their unique pronunciation of several consonantal segments, as well as their distinct vocabulary (Alves 2007)—see Table 3. In terms of their geography, Northern dialects center around Hà Nội, while the nuclei of North-Central, Central, and Southern Vietnamese are the cities of Vinh, Huế, and Hồ Chí Minh City, respectively (Alves 2007, Thompson 1965: 78–104). This paper is only capable of offering insights into cross-dialectal perception for Northern and Southern Vietnamese, since no speakers of Central or North-Central Vietnamese varieties were able to be contacted for inclusion in the study. As such, further discussion of features of these dialect groups is not crucial for current purposes.

The phonetic realization of tones varies substantially from dialect to dialect. In both Northern and Southern Vietnamese, NGANG/A1 is mid to high-mid and level, HUYỀN/A2 is low-falling, and SẮC/B1 is high-rising. The other three tones are not quite as similar between Northern and

tone	Northern Viet.		Central Viet.		Southern Viet.	
	voicing	timing	voicing	timing	voicing	timing
NGANG/A1	0		0		0	
HUYỀN/A2	0 (ɦ)	(end)	0		0 (ɦ)	(end)
SẮC/B1	0		0 (ɦ)	(mid)	0	
NẶNG/B2	~ (ʔ)	end	0 (ɦ, ~)	(end)	0 (ɦ, ~)	(end)
HỎI/C1	0 (ɦ)	(mid, end)	~ (ɦ)	end	0 (ɦ)	(mid)
NGÃ/C2	~ (ʔ)	mid	~ (ɦ)	end	0 (ɦ)	(mid)

Table 2. Phonation of tones in open syllables across three main Vietnamese dialects. For voicing, 0 represents modal, ɦ breathy, ~ creaky, and ʔ a glottal constriction. Parentheses represent a commonly occurring variant within a dialect. Adapted from Vũ (1982: 67).

Orth.	NV	NCV	CV & SV
⟨ch⟩	c	c	c
⟨tr⟩		ʈ	ʈ
⟨s⟩	s	ʂ	ʂ
⟨x⟩		s	s
⟨r⟩	z	ɻ	ɻ
⟨d⟩		ɟ	j
⟨gi⟩		z	
⟨v⟩	v	v	

Table 3. Initial consonant correspondences between Northern Vietnamese (NV), North-Central Vietnamese (NCV), and Central and Southern Vietnamese dialects (CV & SV). Adapted from Alves (2007: 3).

Southern dialects. In Northern Vietnamese, NẶNG/B2 is low-falling, while it is low-falling-rising in Southern Vietnamese. In Southern Vietnamese, HỎI/C1 and NGÃ/C2 are merged and are mid-falling and then quick-rising, whereas in Northern Vietnamese HỎI/C1 is low-falling with only a slight final rise, and NGÃ/C2 is mid-falling and then quick-rising, with a glottal constriction between the fall and rise (Bauman et al. 2009, Nguyễn and Edmonson 1998, Phạm 2003: 34–58, Thompson 1965, Trần 1967, Vũ 1981, 1982).

Besides pitch contours, laryngeal features also play an important role in Vietnamese tones, where breathy voice, creaky voice, as well as glottal stops coöccur with some tones (Michaud 2005, Nguyễn and Edmonson 1998). Northern Vietnamese speakers employ a phonation type other than modal voicing for all but two tones: NGANG/A1 and sắc/B1 (Phạm 2003: 46, Vũ 1982). Comparatively little descriptive work has been undertaken on the Southern Vietnamese tones (Bauman et al. 2009, Trần 1967, Vũ 1982); the literature that does exist suggests that Southerners lean more heavily on pitch and less so on laryngeal features than Northern speakers do in producing tonal contrasts (see also Brunelle 2009b, Kirby 2010). A comparison summary of laryngeal state in three main Vietnamese dialect groups for each tone as laid out by Vũ (1982) is reproduced in Table 2.

Tonal differences are not the only distinguishing factors between Northern and Southern Vietnamese. Alongside them, there are also several segmental phonemes—especially syllable-onset consonants—whose pronunciations vary based on dialect. Alves (2007) provides a sketch of this variation which is replicated in Table 3. This cross-dialectal variation in consonant realization is not directly of consequence for this thesis, but it does play a minor role in the structure of the experimental design.

Despite the noted lack of consensus surrounding featural specifications of Vietnamese tones,

tone	Northern Viet.				Southern Viet.			
	high	low	tense	glot.	high	low	tense	glot.
NGANG/A1	+	-	-	-	+	-	-	-
HUYỀN/A2	-	+	-	-	-	+	-	-
SẮC/B1	+	-	+	-	+	-	+	-
NẶNG/B2	-	+	+	+	-	+	+	
HỎI/C1	-	+	+	-	+	+	+	-
NGÃ/C2	+	+	+	+	+	+	+	

Table 4. Rough sketch of distinctive features for Northern and Southern Vietnamese tones.

a rough sketch of the tonemes in terms of their distinctive features will be helpful for the purposes of this paper. In view of this, I provide a tentative classification of the tonemes of Northern and Southern Vietnamese in Table 4. This account is based mostly on the analyses of (Thompson 1965: 20) and Hoàng (1986) (as cited in Phạm 2003: 9), though it does differ from Hoàng (1986) in its value for the [±low] feature for Northern Vietnamese's NGÃ/C2 tone. The presence of two different height features (i.e., [±high] and [±low]) is needed to account for the tones with more complex contours. Thompson (1965) does not provide a very clear definition for tenseness, but Alves (1995) speculates that it "may have had to do with the tenseness of the glottis" (6). The "glot." column in Table 4 stands for glottalic and is used to designate tones that coöccur with glottal constrictions (or creaky voice). This assignment of tonal features is far from perfect (see especially Phạm 2003: 8), but its inclusion in this paper is meant to serve as an illustration or proof of concept more than anything.

2.2 Lexical processing

Similar to my treatment of tonal features described in the preceding section, the granular intricacies of how lexical processing works are not relevant to this thesis, but a general understanding of theories of word recognition is helpful. These theories dealing with of lexical processing generally divide the phenomenon into (at least) two stages (see Tanenhaus and Lucas 1987, Zwitserlood 1989). The first of these involves the partitioning of an auditory signal into segmental phonemes[1] and parallel activation of items in the mental lexicon which contain these interpreted phonemes. The second stage consists of integrating the activated lexical item(s) into the broader context of the utterance in which the word occurs. This stage imposes further restrictions on the probability of a given word being perceived. When performed competently, the composite of these two stages results in the a listener perceiving a chunk of auditory information as a single word. Scholars differ in the terminology they use to refer to the processes described above, but I will call

[1]For purposes of this paper, I will assume the same process applies to suprasegmental features like tone.

the first stage *access* and the latter stage *integration*, while using the word *selection* to talk about the outcome of the marriage of these two stages, where a listener successfully determines the singular word their interlocutor must have meant. For a more complete overview on the theoretical and terminological divergences between researchers within the field, see the aforementioned Tanenhaus and Lucas (1987) and Zwitserlood (1989).

The *access* stage of lexical processing seems obvious enough: acoustic properties of a spoken word naturally will serve as a fundamental component of how listeners interpret it. As such, an English speaker could reasonably be forgiven for mishearing the word *white* as the phonetically and semantically similar *light*, but mistaking it as the phonetically dissimilar but semantically alike *pale* would be much less plausible.

The *integration* stage is where non-auditory content is considered. Differentiating between homophones when they are spoken in isolation is impossible, but doing the very same when they are uttered in the context of a broader sentence is second nature. Nobody would ever misinterpret the sentence *The waiter would place the check on the table by Jim* as the completely nonsensical and ungrammatical *The wader wood place the check awn the table buy gym*,[2] despite the (near) homophony of each of these pairs of words.[3] In essence, when a listener interprets speech, word recognition is not singularly regulated by the acoustic signal of a given word. Rather, auditory forms are examined and made sense of in view of the surrounding sentential contexts in which the words they represent are spoken (cf. Kazanina et al. 2006).

2.3 Cross-dialectal speech perception

When two interlocutors who natively speak different dialects (or languages) communicate, there is a heightened chance of misunderstandings and greater processing cost involved (Brunellière et al. 2011, 2009, Conrey et al. 2005, Lanwermeyer et al. 2016)—such situations are familiar to us all. The root causes of these difficulties in cross-dialectal (or non-native) speech perception ought to do with misalignments between the interlocutors' native dialects (or language), both in phonological contrasts and in phonetic realizations of those contrasts (cf. Buchwald et al. 1994, Gass and Varonis 1991, Rivera-Gaxiola et al. 2000, Sayer 2013). This has been found to be the case for vowels (Lanwermeyer et al. 2016, Näätänen et al. 1997, Shaw et al. 2018), consonants (Chang 2012, Cutler et al. 2006, Ramírez and Simonet 2018, Weber and Cutler 2004), and more recently also tones (Chen et al. 2020, Wu et al. 2017).

While cross-dialectal asymmetries in any linguistic domain (phonology, morphology, se-

[2]Where, of course, *place* has its polysemous nominal meaning of 'location', and *check* and *table* have the verbal meanings of 'review' and 'postpone consideration of', respectively.

[3]For American English speakers with the COT-CAUGHT merger, at least.

mantics, etc.) could easily impede communication generally, the mismatches that would most obviously have effects on speech perception in particular are those in the realms of phonetics and phonology. Moreover, although differences between speakers' native dialects do decrease their mutual intelligibility (even if slightly) and the speakers' overall ease of communication, increased exposure to the other dialect has an attenuative effect on this decrease (Evans and Iverson 2004, Shaw et al. 2018, Sumner and Samuel 2009). In the following subsection, §2.3.1, I highlight examples demonstrating the role native phonology can play in speech perception; subsequently, in §2.3.2, I discuss the counteracting effects of dialect contact. In §2.3.3, I present the findings of earlier studies in the realm of cross-dialectal Vietnamese tone perception.

2.3.1 Native phonology

One of the most prevalent causes of cross-dialectal miscommunication is found in cases of mismatched phoneme contrasts. For instance, Lanwermeyer et al. (2016) describes a phonemic contrast between /o͡a/ and /o͡ʊ/ in Central Bavarian dialects of Southern Germany. However, in the neighboring dialects of the Bavarian-Alemmanic transition zone, this contrast in neutralized to [o͡a] before obstruents. So, whereas in Central Bavarian [ró͡asn̩] ('journeys') and [ró͡ʊsn̩] ('roses') are minimal pairs, the two are homophonous in Bavarian-Alemmanic, both being pronounced [ró͡asn̩]. Lanwermeyer et al.'s study examined the implications of this phonemic asymmetry between the two dialect groups on cross-dialectal speech perception. The authors found that when listening to a native speaker of an /o͡a/-/o͡ʊ/ merging dialect, speakers of contrasting dialects frequently were not able to map [o͡a] stimuli onto their /o͡ʊ/ phoneme category. As such, when presented with the sentence *Was im Garten viel Pflege braucht, sind Rosen* ('What needs a lot of tending to in the garden is the roses'[4]), the auditory signal of the final word—[ró͡asn̩]—is taken only to convey the meaning of 'journeys'. This results in confusion and misunderstanding on the part of /o͡a/-/o͡ʊ/ contrasting speakers, whereas natively merging speakers would presumably understand that [ró͡asn̩] could correspond to both 'journeys' or 'roses'.

Comparable examples perhaps more familiar to native English speakers can be found with the PIN-PEN merger in some dialects of American English. Speakers with this merger often report that confusion can ensue as a result of this merger when talking to speakers who contrast the vowels. This sentiment is aptly summarized in Baranowski (2013), where one merging speaker is quoted as saying, "[p]eople often misunderstand when I use those words so probably… uh they sound more the same than I realize. Pin is something you stick with, and pen is what you write with" (287).

Similar effects have been noted in how bilinguals (mis)perceive contrasts in their second

[4]Translation mine.

(or non-dominant) language. For example, Dutch- and Japanese-English bilinguals, whose native languages lack phonemic /ɛ/-/æ/ and /l/-/ɹ/ distinctions, respectively, are more likely to confuse these phonemes for each other when listening to native English speakers (Cutler et al. 2006; Weber and Cutler 2004). Likewise, Majorcan Catalan draws a phonemic distinction between /ʎ/ and /ʒ/, both of which Majorcan Spanish lacks. Ramírez and Simonet (2018) examined Spanish-Catalan bilinguals' perception of these consonant sounds, finding that Spanish-dominant bilinguals, while performing at a rate better than chance, still were much worse at accurately perceiving [ʎ]-[ʒ] minimal pairs than Catalan-dominant bilinguals.

A principal takeaway regarding the role of native phonology in cross-dialectal speech perception is that one-to-many sound correspondences are more readily decoded by the listener than many-to-one correspondences are. This idea is reflected in Sumner and Samuel (2009), where the author summarizes that, "[g]enerally, research in this area has shown that speakers of merged dialects… are less able to make perceptual discriminations between the merged [phonemes] than speakers of unmerged dialects" (488). These facts allow for the emergence of asymmetries, where an uneven distribution of many-to-one correspondences between two dialects or languages calls into question their *mutual* intelligibility. Such is the case of Portuguese and Spanish, where it is often reported that speakers of the former have greater ease in understanding the latter than the reverse (Jensen 1989).

2.3.2 Contact

Unsurprisingly, familiarity plays a pivotal role in the extent to which a speaker can understand dialects they do not speak natively, and difficulties in speech perception tend to dissipate with greater exposure to a given dialect (Evans and Iverson 2004, Hay et al. 2010, Sumner and Samuel 2009). For example, Sumner and Samuel (2009) examines perception of r-less varieties of New York City English. Many (though not all) speakers of NYC English drop /ɹ/ in coda position, so that words like *dancer* are realized as [dænsə], rather than the typical North American English [dænsɚ]—see especially the now classic Labov (1966). Sumner and Samuel's study investigates the lexical perception of (1) r-less speakers of NYC English, (2) r-ful speakers of NYC English, and (3) r-ful speakers of General American[5] when exposed to stimuli produced by speakers of an r-less NYC variety, and to stimuli produced by speakers of an r-ful General American dialect. They found that r-ful speakers of NYC English, despite having pronunciations of coda-r identical to those of non-NYC, General American English speakers, nonetheless are able to perceive the r-less variety of NYC English much more fluently than the non-New Yorkers. The only viable

[5]Sumner and Samuel (2009) use the term "General American" to refer to "speakers who do not r-drop or exhibit any other regionally marked characteristics" (488).

account of this difference is that it is due to an effect of long-term exposure to r-less NYC English by r-ful NYC speakers that is missing for speakers of General American English.

A perhaps more noteworthy consequence of high levels of dialect contact is its ability to affect not only perception, but even production. One such example is discussed in Bowie (2000), where the author investigates (among other things) the phonetic realization of vowels among American English speakers who were born and raised in the town of Waldorf, Maryland. He distinguishes between what he calls "lifelong Waldorfians" and "Waldorf exiles", with the former group consisting of participants who had spent their whole lives in the town and the latter of those who had spent some amount of time living outside their hometown. Bowie finds that, in the case of some vowels, speakers who had spent time away from the Maryland town had measurably different pronunciations. What's more, these distinctive realizations often corresponded to how a given vowel is produced in the dialect of the area they moved to. In fact, the magnitude of the difference between the pronunciations of lifelong Waldorfians and Waldorf exiles often was found to scale by amount of time spent outside the town. What this all suggests is that living in an area that does not share one's native dialect (and therefore being exposed to a greater deal of that area's dialect) has a gradual effect on language production, in time pulling a speaker's pronunciation more in line with that of their new home.

2.3.3 Cross-dialectal tone perception in Vietnamese

For Vietnamese specifically, only scant research into cross-dialectal speech perception has been done. One of the most significant studies to date is Brunelle (2009b), which compared Northerners and Southerners' perception of Northern Vietnamese tones. The experiment exposed participants to two types of stimuli—natural stimuli elicited from a Northern Vietnamese speaker, and synthetic stimuli, created by altering some phonetic features of the natural stimuli. The natural stimuli consisted of six variants of the syllable /ma/ produced in isolation, one with each of the six tones. To create the synthetic stimuli, the elicited /ma/ syllables were manipulated, having their durations standardized (to 300ms) and their pitch contours adjusted at the onset and offset, with pitch also being adjusted at the midpoint of the syllable to simulate complex pitch contours. For each contour, there were three voicing variants representing common phonation features found in a subset of Northern Vietnamese tones: modal voicing, laryngealization, and a glottal constriction.

The study found that, in natural stimuli perception, Northerners correctly identified every tone at rates above 90%. Southerners, while correctly identifying NGANG/A1, HUYỀN/A2, and SẮC/B1 at similar rates, performed worse with all the other tones, especially HỎI/C1 and NGÃ/C2, which were accurately perceived at rates of just 10% and 63%, respectively. Southerners were

Stimulus tone	Response tone											
	Southerners						Northerners					
	A1	A2	B1	B2	C1	C2	A1	A2	B1	B2	C1	C2
NGANG/A1	**.92**	*.08*	.00	.00	.00	.00	**.97**	.02	.00	.01	.01	.00
HUYỀN/A2	.02	**.97**	.00	.00	.02	.00	.03	**.93**	.02	.01	.01	.01
SẮC/B1	.00	.00	**.95**	.00	.02	.03	.00	.00	**.99**	.00	.01	.00
NẶNG/B2	.00	*.13*	.00	**.85**	.02	.00	.00	.00	.00	**.98**	.02	.01
HỎI/C1	.00	.00	.00	**.87**	*.10*	.03	.00	.01	.00	.02	**.95**	.03
NGÃ/C2	.00	.02	*.27*	.00	*.08*	**.63**	.00	.00	.04	.02	.02	**.93**

Table 5. Tone-wise perception rates for Southern and Northern Vietnamese participants of natural Northern Vietnamese /ma/ stimuli in Brunelle (2009b). Main responses shown in bold; other common responses (>5%) in italics; correct responses in red. Table synthesized from Tables 4 and 5 in Brunelle (2009: 84–85).

fairly consistent in their misclassification of the HỎI/C1 tone stimuli, perceiving them as NẶNG/B2 at a rate of 87%. The NGÃ/C2 tone stimuli were most commonly misperceived as having SẮC/B1 tone (27%), but also not infrequently as having HỎI/C1 (8%). A summary of these findings is shown in Table 5.

The results of the synthesized stimuli perception are much more complex. Brunelle's goal in probing how participants tonally categorize these stimuli was to identify the salient phonetic cues used by speakers of Northern and Southern Vietnamese in tone perception, and so these results are not entirely as relevant to the current study as those of the natural stimuli perception tasks. These findings no doubt warrant a discussion in greater depth than is feasible here, and so I will restrain from highlighting them.

In addition to Brunelle (2009b) and Kirby (2010), other noteworthy findings are presented in Brunelle and Jannedy (2013). The first of these investigated Northern and Southern Vietnamese participants' response times and error rates in determining whether the tone of two audio stimuli (consisting of the syllable /ba/) were the same or different. The results of the study demonstrated that, when listening to stimuli elicited from Northern Vietnamese speakers, Southern Vietnamese speakers were slightly slower, but also more accurate than Northern Vietnamese speakers. However, Southerners both took longer than and performed at rates significantly lower than Northerners in determining whether HỎI/C1-NẶNG/B2 and HỎI/C1-NGÃ/C2 pairs had the same or different tones. These findings suggest that Southerners are more actively attentive to the speech of Northern Vietnamese speakers. With regard to Southerners' worse performance on pairs involving HỎI/C1 tone stimuli, Kirby states that the laryngeal features of these tones in Northern speech are the most likely culprit for the confusion. Since such features are absent from the tones in Southern Vietnamese, it is unsurprising that Southerners would bin stimuli containing

the perceptually salient laryngeal features into their category, inviting confusion between them.

In some ways, Brunelle and Jannedy (2013) replicates the experimental design and findings of Brunelle (2009b). Like the earlier study, this one tests Southern Vietnamese speakers' perceptions of Northern Vietnamese tones when presented stimuli of a single syllable. Unlike the earlier study, though, Brunelle and Jannedy (2013) also investigates Southern Vietnamese speakers' perceptions of Southern Vietnamese tones, as well as whether the dialect of the experimenter and/or the relative order of Northern and Southern Vietnamese stimuli have measurable effects on rates of tone perception. They find that only for the Northern stimuli with HỎI/C1 or NGÃ/C2 tone does the dialect of the experimenter effect the tone response rates.

Another noteworthy result of this experiment is that Southerners' tone perception rates for Northern HỎI/C1 tone stimuli did not match those reported in Brunelle (2009b). In Brunelle (2009b), HỎI/C1 tone stimuli were found to be perceived as NẶNG/B2 by Southern participants at an 87% rate, but Brunelle and Jannedy (2013) places this rate at just 6.7%. The later study finds that HỎI/C1 tone stimuli are perceived as HUYỀN/A2 at a 40.1% rate, as HỎI/C1 at a 25.3% rate, and as NGANG/A1 at a 20.9% rate. Brunelle and Jannedy do not discuss these incongruities between their findings and those of Brunelle (2009b).

Unlike the three studies described in the preceding paragraphs, (Vũ 1981: 166–221) is not exclusively interested in cross-dialectal tone perception for monosyllabic, contextually isolated stimuli. Vũ conducted four experiments testing cross-dialectal perception of tone, three of which involved stimuli that were monosyllabic and context-independent. The results of these experiments in large part mirrored those that would later be found in (Brunelle 2009b, Brunelle and Jannedy 2013, Kirby 2010), with some notable exceptions. For Vũ (1981), Southern participants performed better at correctly identifying the tone of Northern HỎI/C1 stimuli (51.6%). As in Brunelle (2009b), Northern Vietnamese HỎI/C1 stimuli were more likely to be perceived as NẶNG/B2 than as HUYỀN/A2 (22.4% vs. 9.8%). Vũ's results for Southerners' perception of Northern Vietnamese HỎI/C1 stimuli more closely resembled those of Brunelle and Jannedy (2013) than those of Brunelle (2009b).

The more unique experiment performed by Vũ (1981) was to task participants with perceiving the tone of stimuli that consisted of meaningful words, phrases, or sentences, rather than single syllables spoken in isolation. Here, Vũ correctly predicts that, "when lexical meaning can be inferred from meaningful utterances and contribute to the identification of tones… recognition scores for this test would be highest" (168). For this experiment, Northern Vietnamese stimuli were correctly identified by Southern Vietnamese participants at rates of over 95% for each tone. Essentially, Vũ's experiment demonstrates that the integration stage of lexical processing plays

some role in shaping cross-dialectal tone perception in Vietnamese, since we would otherwise expect rates of correct identification of tone not to improve from the experiments that involved monosyllabic, context-independent stimuli.

2.4 Mergers

Discrete phonemic categorization has long been a cornerstone of phonological analysis (Labov et al. 1991: 33). However, there is reason to believe an entirely discrete approach to phonological phenomena might not capture important nuances in the actual mental architecture of language users. Mergers, which are said to have occurred when a phonemic distinction between one sound and another collapses, are one area where these nuances can be observed. In many ways, mergers are more gradient than they are discrete, as they can vary along several axes, including production, perception, phonotactics, and the extent of their productiveness across a speaker's lexicon (Labov et al. 1991, Maguire et al. 2013).

For example, a speaker of American English may merge the KIT and DRESS vowel but only in pre-nasal environments, resulting in the PIN-PEN merger (Austen 2020, Baranowski 2013). Others still (myself among them) may have this merger in certain words but not others (e.g., homophony between *win* and *when*, *Minh* and *men*; but minimal pairs of *tin* and *ten*, *mint* and *meant*).

2.4.1 Asymmetries abound

As discussed in §2.3, asymmetries have been found to exist between speakers' production and perception of mergers. For the PIN-PEN merger, Austen (2020) finds that, among American English speakers who merge the sounds in production, the split is roughly even between those who merge and those who contrast in perception. Another example of an asymmetry is seen in the fact that many speakers of New Zealand English have a merger between the NEAR and SQUARE vowels, but can nonetheless perceive the distinction in these vowels when produced by non-New Zealanders who don't have this merger (Hay et al. 2010). Production-perception asymmetries have also been identified with tonal contrasts; Mok et al. (2013) found that many young Hong Kongers merge two tones of their native Cantonese in their pronunciation, but retain the contrast in perception of others' speech.

In view of these and the production-perception asymmetries described in §2.3, there are a total of four logical possibilities that could theoretically exist for sound mergers: (1), speakers lacking a merger, both in production and perception (Labov et al. 1991: 58); (2), speakers possessing a merger, both in production and perception (Austen 2020); (3), speakers who contrast two phonemes in production but not in perception (Janson and Schulman 1983); and (4), speak-

ers who contrast two phonemes in perception but not in production (Hay et al. 2010, Mok et al. 2013). The results of the prior studies on Vietnamese tone perception mentioned in §2.3.3 seem to suggest that Southern Vietnamese speakers who merge the HỎI/C1 and NGÃ/C2 tones in their own production can nonetheless recover the distinction when produced by non-merging speakers. As such, the HỎI/C1-NGÃ/C2 merger ought to, at least for some speakers, fit squarely in the fourth category listed above. However, there is more nuance here than meets the eye.

2.4.2 In perception

An issue that might at first seem slightly pedantic is the question of what the terms "production" and "perception" actually refer to in the context of sound mergers. The first of these, *production*, represents the perhaps best-known way in which mergers are understood. Mergers in production are cases where a phonemic contrast present in some dialect of a language is lost in another, such that the speakers of the merging dialect do not phonetically distinguish between the two phones in their pronunciation. Relative to production, *perception* is less straightforward. The term is at least somewhat ambiguous—when speaking of mergers in perception, do researchers mean that listeners with such mergers are incapable of perceiving the difference when produced by non-merging speakers? Or, are they rather referring to cases where a speaker reports to believe themself not to produce a contrast between two sounds despite in reality doing so? To me, the more obvious interpretation is the former, and this is consistent with how the term is used by most studies on mergers I've reported on here.

However, the distinction just described is meaningful and testable, as well as being on the radar of some studies. Austen (2020), for example, presented American English speakers *both* with audio data, asking participants to determine which word was said, *and* with a written pair of words, asking participants to determine whether or not the pair rhymed. The first of these was meant to test how speakers heard the stimuli, whereas the second was meant to test how speakers abstractly conceived of their own pronunciation of the target words. Herold (1990) also performed a similar rhyming-task experiment (with COT-CAUGHT stimuli) to test this, and found that, while many speakers' pronunciations and rhyme reports matched, there were also cases where speakers would pronounce the vowels distinctly, but then go on to say they rhymed or vice versa (22–33).

For the purposes of this paper, I will take *merged in perception* to refer to cases involving speakers' lack of an ability to discern minimal pairs produced by non-merging speakers, rather than cases involving speakers' lack of awareness that they themselves pronounce two phonemes distinctly.

Even taking into account the constraint that *perception* deals only with how speakers interpret the speech of others, there are still nuances to unpack. For instance, what exactly does it mean to ask whether Southern Vietnamese speakers can "recover the distinction" between the tones of Northerners? In talking about speech perception, Herold (1990) draws contrasts between what she refers to as *hearing*, *identifying*, and *sorting*, writing:

> If we wish to distinguish among competing claims about the ability to identify words, on the one hand, and the ability to perceive acoustic differences, on the other, I suggest that we reserve the word *hear* for use in its original sense of "perceive by ear." We can then use *identify* to mean "label with the lexical item intended by the speaker." A third category of behavior, which I call the ability to *sort*, is exhibited by listeners who label a series of auditory stimuli consistently but incorrectly, i.e. with a label which is not that intended by the speaker. (173)

As an illustration, consider the sentence: *Dawn forgot to pack lunch for the trip, but luckily, Don had brought an extra sandwich.* Following Herold's typology, there are four ways in which a speaker with the COT-CAUGHT merger (in production) could be said to interpret the phonetic distinction between *Don* and *Dawn*. The first is not at all, in which case this sentence would likely sound quite unusual to them. The second is that they might *hear* a difference, but have no idea whether Dawn was sharing her meal with Don or vice versa. Thirdly, they could *hear* and also *identify* the contrast, knowing that the sentence's speaker was conveying that it was Don who gave his sandwich to Dawn. Lastly, they could *hear* and *sort* the words, resulting in a reversal of Don and Dawn's roles from the speaker's intended meaning.

What is clear from the results of Brunelle (2009b), Brunelle and Jannedy (2013), Kirby (2010) and Vũ (1981) is that, in Herold's terminology, Southerners are generally able to hear most of the tonal contrasts of Northern Vietnamese, and they can even properly identify most of them (see Table 5). However, the extent to which their findings can be extrapolated to cases of actual communication might be called into question. This sort of disclaimer is levied by Labov et al. (1991) against studies with comparable methodologies. They say:

> These formal situations ask the speakers to perform a metalinguistic act, the *labeling* of linguistic categories… This is quite distinct from the *perception* of features or contrasts, which is a largely unconscious act that enters into the *interpretation* of utterances. (57–58)

The basic criticism here is that in studies like Brunelle (2009b), Brunelle and Jannedy (2013) and Kirby (2010), where researchers subject participants to upwards of hundreds variants of a single syllable produced in isolation, participants are likely to be more acutely attending to differences

they might not pick up on in normal speech. In lexical processing terms, these studies only make room for the access stage (i.e., the auditory signal of a word), while the integration stage of interpreting a word in context cannot play any role.

Instead, Labov et al. (1991) make the case for experimental designs that present participants with more natural and familiar linguistic scenarios, taking their own advice and testing whether listeners can pick up on vowel differences in sentences where the meaning of the sentence hinges on how the listener phonemically interprets that vowel quality. An example of how implementing this approach could work might be to present a group of study participants comprised of NEAR-SQUARE merging speakers of New Zealand English with audio stimuli of a non-merging speaker producing sentences like *Yeah, I think I like this [here/hair]*, where participants alternate between hearing the sentence with *here* and *hair*, gauging the rates at which participants recovered the intended meaning of the sentence.

2.4.3 In production

Although a broad definition of *mergers in production* is comparatively easy to offer, this term also has more fine-grained detailed that shouldn't be overlooked. In pulling ideas from earlier research—especially Herold (1990), Trudgill and Foxcroft (1978)—Labov (1994) delineates between three different "mechanisms of mergers": *by approximation*, *by transfer*, and *by expansion*. Under approximation, phoneme A gradually encroaches on the phonetic space of phoneme B, eventually to such an extent that A and B's phonetic realizations are no longer discernible from one another to the speaker. Nonetheless, this avenue still allows for the preservation of phonemic contrasts, even after loss of any phonetic contrast. Under transfer, phoneme A is reanalyzed by a speaker as being underlyingly phoneme B but (crucially) only in certain lexical items. This avenue does not allow for near-mergers to take hold since the target pronunciation of phoneme A does not shift. Under expansion, the phonemic distinction between A and B is lost entirely and spontaneously at language acquisition, and a speaker utilizes the breadth of the phonetic space once occupied by phonemes A and B individually as the target pronunciation of the merged phoneme (321–323). These three mechanisms of mergers vary with respect to two axes: lexicon and phonetics. Mergers by approximation are phonetically gradual but lexically regular, while mergers by transfer are phonetically discrete but lexically irregular. Mergers by expansion share properties with each, being neither phonetically nor lexically gradual (Herold 1990: 49–50). This typology has been applied to tone mergers at least once, as far as I can tell (Fung and Wong 2011).

Recasting what's been laid out in this section in a more digestible format and bringing it into dialogue with the issues at hand for this study, cross-dialectal phoneme perception involving variably merged sounds can be visualized as a simple matrix, as shown below. Between two inter-

locutors, each could have the two tones under consideration as contrastive or merged, resulting in four possible cells within the matrix:

		Speaker	
		contrasts	merges
Listener	contrasts	i	ii
	merges	iii	iv

Each of these cells makes different predictions as to how the listener should perceive the hỏi/c1 and ngã/c2 tones of the speaker. In situations (i) and (iv), the results seem obvious: the listener hears the tones as contrastive and homophonous, respectively. The likely perception of the listener in (ii) and (iii), however, is less apparent from the outset. Because this paper is concerned with how hỏi/c1-ngã/c2 merging speakers perceive the tones of hỏi/c1-ngã/c2 contrasting speakers, I will focus on (iii) more so than (ii).

If the mechanism of merger is approximation, then the listener in (iii) should be able to recover the hỏi/c1-ngã/c2 distinction either consistently or at least better than chance. If the sort of the merger is that of expansion, then the listener should not be able to recover this distinction. On the other hand, mergers under transfer are at once the easiest to recognize and the hardest to find since they occur mostly sporadically throughout a speaker's lexicon. As a result of this incomplete nature, listeners with merger under transfer should be able to consistently recover the hỏi/c1-ngã/c2 distinction for certain words but not any better than chance for others. Nonetheless, sociological factors could interfere with these generalizations, especially insofar as they might disguise a situation of expansion merger as one of approximation.

3 Research questions

The general aim of this thesis is to contribute to a greater understanding of the nature of the hỏi/c1-ngã/c2 tone merger in Vietnamese. In service of this broader goal, my research seeks to answer the following questions.

1. *Are Vietnamese speakers who merge the* hỏi/c1 *and* ngã/c2 *tones in production able to perceive a phonemic contrast (i.e.,* sort *or* identify *in the Heroldian sense) between these tones when pronounced by non-merging speakers in naturalistic contexts?*

2. *Inasmuch as* hỏi/c1-ngã/c2 *merging speakers are able to perceive a difference in the tones when produced by speakers of a non-merging dialect, how accurate is their identification of these tones?*

3. *Do asymmetries exist in tonal perception?* That is, *Is one of either the* HỎI/C1 *or* NGÃ/C2 *tones more likely to be misperceived by merging speakers than the other is?*

4. *Can interspeaker variation among merging participants be found in perception of the* HỎI/C1 *and* NGÃ/C2 *tones?*

5. *To what extent does the sentential context of a word with* HỎI/C1 *or* NGÃ/C2 *tone play a role in structuring tone perception?*

6. *Are there measurable effects of exposure to Northern Vietnamese dialects on the degree to which merging speakers are able to recover the tonal distinction in perception?*

7. *In term of Labov (1994)'s three-way typological split, what (if any) evidence is found in support of (or against) analyses of the merger as happening via approximation, transfer, or expansion?*

8. *If found to exist, what would the implications of asymmetries in the* HỎI/C1-NGÃ/C2 *merger between production and perception be with regard to speakers' phonemic representations of these tones?*

Answering each of these questions is interesting in its own right, but in so doing I will also keep in mind the deeper question of why it could be that the results show the patterns they do. The results of this study will also be compared to those found in Brunelle (2009b), Brunelle and Jannedy (2013), Kirby (2010), and Vũ (1981), as well as other relevant literature discussed in the preceding sections. A secondary objective of this thesis is to shed an admittedly modest amount of light on the production of tonal contrasts in Southern Vietnamese, given the aforementioned (relative) dearth of descriptive literature on the subject.

4 Methodology

In order to investigate the questions under consideration, I developed and conducted a study hosted by *Phonic.ai*, an online survey platform that allows responses to be collected remotely. The study was composed of an initial portion on biographical information, followed by two main parts. The biographical portion asked respondents a number of basic questions, covering age, gender, place of residence, place of birth, parents' place(s) of birth, years of education with Vietnamese as the primary language of instruction, and highest level of educational attainment. The first portion of the survey's main section was a production task, and the second was a perception task. The order of the production task before the perception task was determined to be crucial, as this prevents the perception task from biasing the participants' pronunciations for the pro-

duction task stimuli. Since the main concerns of my investigation surround how speakers of HỎI/C1-NGÃ/C2 merging dialects perceive the tonal contrasts of non-merging dialects, the inclusion of non-merging speakers in the study was for them to act as a control group. Participants were not individually compensated for their engagement but instead entered into a raffle to receive a $20 Amazon gift card. Survey participants were recruited mostly via social media and email with the assistance of friends, professors, and the Vietnamese Student Association at Yale.

4.1 Production task

The first portion of the survey was the production task. The purpose of this task was to ascertain whether a speaker had contrastive or homophonous pronunciations of the HỎI/C1 and NGÃ/C2 tones.

4.1.1 Stimuli

Participants were prompted with visual stimuli, each consisting of an image or piece of clip art representing a basic object or thing. Because tone is marked in the orthography of Vietnamese, picture stimuli were used to diminish the likelihood of the writing system influencing participants' pronunciation. Participants were asked to record themselves in a quiet room as they name each object by filling in the carrier sentence *Đây là* ___ ('This is ___') using as few words as possible. The stimuli were constructed such that the target word (i.e., the word the object displayed was meant to elicit) should always be placed at the end of the sentence, so as to control for any effects of prosodic position on the realization of tone.

Before the actual production task, participants were given an example stimulus consisting of an image of a wooden chair for practice. They were instructed, *Khi bạn thấy hình ảnh ở dưới bạn nói 'Đây là cái ghế.'* ('When you see the image below, you say 'This is a chair.') After successfully completing this practice round, the participants proceeded to the production task portion of the survey.

The production task included six target stimuli (i.e., stimuli containing a HỎI/C1 or NGÃ/C2 vocabulary item) and four filler stimuli (i.e., stimuli containing vocabulary items with other tones), for a total of ten visual stimuli, presented to each participant in a randomized order. Of the target stimuli, two were minimal pairs, with the only contrasting phoneme being the tone. An additional two were near minimal pairs, differing in initial consonant as well as tone. Examples of what production task questions looked like on the participant's end are shown in Figure 2.[6] A

[6]The prompt translates as "Please press the 'Record' button and then say 'these are ___.' After that, press 'Next' to continue." Unfortunately, *Phonic.ai* does not currently support Vietnamese metalanguage for surveys, and English buttons were the only option.

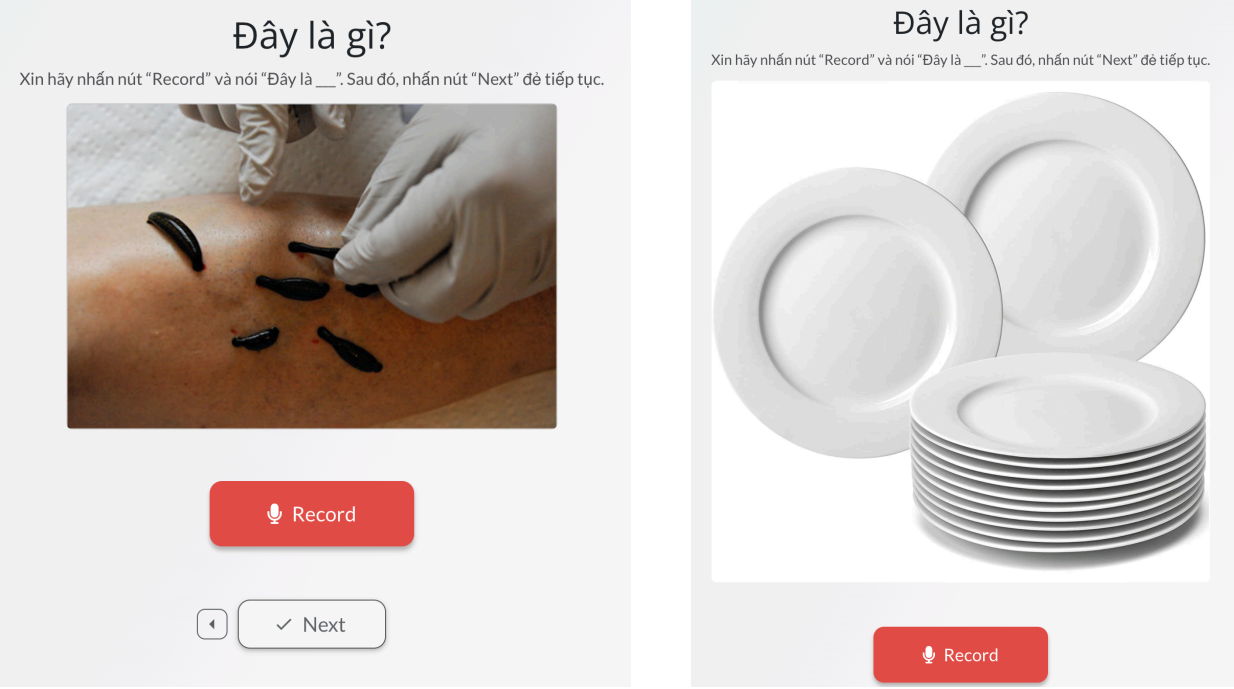


Figure 2. Example of minimal pair stimuli in the production task portion of the survey. Target word for the 'leech' stimulus *đỉa* (HỎI/C1); for the 'plate' stimulus *đĩa* (NGÃ/C2).

repository of all visual stimuli used in the survey can be found in Appendix C.

4.1.2 Analysis

Once participants completed the survey, I am able to retrieve the audio files associated with each of their responses for phonetic analysis. Each participant's production of the target words task will be checked in Praat in order to binarily classify participants as HỎI/C1-NGÃ/C2 merging or non-merging. A more objective method of classifying participants is desired, but this is not feasible given constraints of this thesis.

4.2 Perception task

The perception task was the second of the two major tasks. Where the production task consisted of the elicitation of vocabulary items meant to obtain participants' realizations of tones, the perception task focused on how participants categorized another speaker's tonal contrasts as they heard them.

4.2.1 Stimuli

The stimuli for this task were pairs of sentences, each of which had two versions with only a one-word difference between them. For instance, (1a) and (1b) below show glosses of minimal pair stimuli whose interpretations hinge on the tone of the highlighted words—HỎI/C1 in (1a) and NGÃ/C2 in (1b).

(1) a. *Vợ bị vô sinh nhưng mãi cũng chửa (HỎI/C1) được.*
wife suffer infertile but still be pregnant PASS
'The wife was infertile, but she got pregnant.'

b. *Vợ bị vô sinh nhưng mãi cũng chữa (NGÃ/C2) được.*
wife suffer infertile but still cure PASS
'The wife was infertile, but she got cured.'

Participants were played audio recordings of a male, non-merging, Northern Vietnamese speaker in his early twenties from Hà Nội reading aloud the stimuli twice over, first at a slightly faster pace, and then, after a pause of two seconds at a normal, conversational pace and prosody. The stimuli were recorded in a quiet room. In every stimulus, the target word appeared in sentence-medial environments to control for any effects of prosodic position on the realization of tone. The written transcription of each sentence was displayed as the participants listened to the audio stimuli, save the target word. Participants were instructed to choose between one of six words to fill in the gap. One of the options was the actual word as spoken by the Northern Vietnamese speaker, and the other five were permutations of this word, each with a different tone diacritic. However, not every option represented an actual word, and of those that did, not every one was syntactically possible or semantically plausible, given the matrix sentence—see Table 6. This was done in order to effectively circumscribe which of the six answers the listener deems probable (as modulated by the integration stage of lexical processing), and thus presumably render the tone of the stimulus more likely to be perceived either as the tone the speaker actually uttered, or as that of the stimulus's tonal minimal pair counterpart.

Similar to with the perception task, before coming to the actual production portion, participants were told they would hear an example stimulus of a speaker saying what translates to 'I really like to listen to music, and I'm wanting to learn to play the guitar.'[7] They were instructed to select the word *chơi* ('play') from among the six options, and they subsequently listened to

[7]Vietnamese: *Tôi rất thích nghe nhạc và đang muốn học chơi ghi ta.*

Tone	Word	Meaning	Syn. Possible?	Sem. Plausible?
ngang/A1	*chưa*	'(not) yet'	no	no
huyền/A2	*chừa*	'to spare', 'to quit'	yes	unlikely
sắc/B1	*chứa*	'to contain', 'to conceal'	yes	no
nặng/B2	*chựa*	–	–	–
hỏi/C1	*chửa*	'to be pregnant'	yes	yes
ngã/C2	*chữa*	'to repair', 'to cure'	yes	yes

Table 6. Syntactic and semantic viability of the six choices for the stimulus *Vợ bị vô sinh nhưng mãi cũng ___ được.*

the example stimulus while the sentence was shown on the screen, missing the crucial word *chơi* ('play'). Once they finished the practice round, they continued on to the production task.

For the perception task, twelve stimulus pairs were used, for a total of 24 individual stimuli. These consisted of six target pairs with hỏi/C1-ngã/C2 tonal contrasts and six filler pairs with other tonal contrasts (although five of the filler pairs contained one stimulus of either hỏi/C1 or ngã/C2, paired with a different tone). A complete list of target and filler stimuli with English translations can be found in Appendices A and B, respectively. Figure 3 provides an example of what a perception task question looked like to participants.[8]

Whereas for the production task orthography was thought to potentially bias the results, for the perception task, the six word choices being spelled out was necessary for some stimuli. This was due to the fact that several Vietnamese consonant phonemes have cross-dialectally variable pronunciations, with some dialects merging certain sounds while others do not—see Table 3. As such, the inclusion of orthographic transcriptions was intended to serve to clear up any potential confusion on the part of the speaker caused by cross-dialectally different consonant realization in Vietnamese, which is not the aim of this study.

4.2.2 Analysis

The responses collected in the perception task portion of the study are analyzed by means of a binary logistic regression modeling in R using the *glm* function, as done in (Brunelle and Jannedy 2013: 21–22). The dependent variable is binary, where 1 represents an accurate identification of stimulus tone, and 0 represents an inaccurate identification of stimulus tone. This variable is analyzed with respect to three variables: geography, production of hỏi/C1-ngã/C2 tonal contrast, and tone of stimuli. The first of these variables, geography, groups participants binarily as Northern and Southern. The second, production of contrast, also groups participants binarily, either as merging or non-merging. The final variable, tone of stimuli, sorts the 24 individual stimuli into

[8]The prompt translates as 'Which word is this person saying?'.

Figure 3. Example of a perception task stimulus.

three groups: target stimuli with HỎI/C1 tone, target stimuli with NGÃ/C2 tone, and filler stimuli. Analyses will be performed independently for each tone of stimuli group.

This approach allows accuracy in tone identification to by analyzed in terms of how it is affected both by these variables independently and by the interactions between them. That is, we can test not only for significance of the effects of geography and production of contrast for each tone of the stimulus group, but also for the significance of the effect of these variables together.

To test for differences between the tones of the stimuli, I will perform binary logistic regression analyses where differential rates of accurate tone perception of target-HỎI/C1, target-NGÃ/C2, and filler stimuli are measured in terms of how they are affected by geography, production of contrast, and the interaction between these two.

5 Results

In this section, I present the results of my study, broken down in terms of the responses for the demographics of participants (§5.1), the findings of the production task portion of the survey (§5.2), and finally the perception task portion of the survey (§5.3).

5.1 Demographics

In total, 74 complete responses were collected. Among these completed survey responses, females were overrepresented, accounting for 70% of the total (n=52). All but one of the participants were born in Vietnam and had at least ten years of education with Vietnamese as the primary language of instruction. Of the 74 participants, 61 currently lived in Vietnam, with the remaining thirteen living in the diaspora (mostly in the United States, but also Canada, Australia, and South Korea). In terms of age, most of the participants were between 25 and 39 (59%; n=44), with the vast majority being under 39 (92%; n=68). The respondents were well-educated, overall, with a 45% plurality of them indicating university as their highest level of education completed (n=33), and an additional 31% having earned a graduate degree (n=23), for a total of 76% of respondents having completed a tertiary education (n=56).

When comparing these demographic figures against those who had started the survey but not finished it (which totalled 245 incomplete responses), the statistics are generally comparable. There are, however, two exceptions to this trend. Those who completed the survey were more likely than those who had not completed it to: (1), have at least ten years of education with Vietnamese as the primary language of instruction (χ^2=11.4555; p=0.00071); and (2), have been born in Vietnam, as opposed to in the diaspora (χ^2=6.077; p=0.01369). The most obvious explanation for these differences is that less educated speakers and those born in the Vietnamese diaspora are less likely to be confident in their command of Vietnamese, but there may be alternative explanations. It is not clear whether these facts—especially the latter—could have skewed the results. More specifically, Southern Vietnamese speakers born in the diaspora might reasonably be expected to have lower levels of exposure to Northern Vietnamese dialects, and thus be less likely to accurately distinguish the non-native tonal contrast between HỎI/C1 and NGÃ/C2 that Northerners generally produce. A full summary of the demographic data of study participants can be found in Appendix D.

In the demographic portion of the survey, participants were asked three questions pertaining to geography—the city and province/state of: (1), where they currently reside; (2), where they were born; and (3), where their family is from. The third question was optional, and participants were asked to provide a response if they were born outside of Vietnam. Nonetheless, many participants who were indeed born in Vietnam provided a response to this question. Determining where participants ought to be considered from for the purposes of this study was challenging in some cases, particularly when there were differences between where a participant was born and where they now live. In cases of such mismatches, participants were assigned to a given area in Vietnam if at least two of their responses the three questions listed above indicated that they were associated with that area. If a participant did not provide a response to the third question,

and their responses to the first and second question did not align, then they were associated with the geography in which they were born, rather than where they currently live. There were two cases of participants now living abroad who had been born in Hồ Chí Minh City but indicated that their parents were from Northern Vietnam. Due to these complications, they were not classified in terms of geography. This all is, of course, a rather rough treatment, but in lieu of more detailed information for each participant, it seems the best that can be devised.

The survey received responses from all around the country of Vietnam, though all participants were from areas listed by Alves (2007), Vũ (1982) as belonging to either the Northern or Southern dialect group. Southerners constituted a majority of the respondents, with 50 participants (68%) being from provinces corresponding to the Southern dialect. Northerners, at 22 participants, represented 30% of respondents. A full table of information on which provinces of Vietnam participants were from can be found in Appendix E.

5.2 Production task

The audio data collected in the production task portion of the study were used to classify participants as either HỎI/C1-NGÃ/C2 merging or contrasting. Analyzing the responses to this portion presented complications, as explained below.

First, the responses elicited from the visual stimuli used in the study were not entirely regular. The template participants were instructed to follow when prompted with a picture was "This is ___", and participants were asked to use only a few words in their responses. However, there was still variety. For example, the visual stimulus of a bunch of grapes was meant to elicit the word *nho*, 'grape'. Participants used different classifiers for the grapes, such as *chùm* ('bunch'), *trái* ('fruit'; Southern Vietnamese), or *quả* ('fruit'; Northern Vietnamese), and some further added the quantifier *một* ('one'). A few participants added *xanh* ('green') after *nho* to specify the color of the grapes. The design of the study was meant to standardize the position of the target word at the end of the sentence, but participants adding further specifications often thwarted this attempt at standardization. This was especially the case for the visual stimuli of the roots and peel. The vast majority of participants described the peel as *vỏ cam* ('orange peel') or *vỏ quýt* ('tangerine peel'). A slightly smaller proportion described the roots as *rễ cây* ('tree roots'). As such, the duration of the target words in these responses usually became shortened and their realization of tone harder to analyze. A similar issue is that often participants would use different words to describe the stimuli than what I had intended, usually synonymous with the intended response (e.g., *vắt* instead of *đỉa* for leech). One participant mistook the horse stimulus for a pegasus.

Second, in the best of cases, the study only elicited three tokens of each the HỎI/C1 and

NGÃ/C2 tone for any given speaker. In many cases, because of the issue described in the preceding paragraph, there are fewer tokens, and there are even a few speakers for whom only one viable token of a tone is available, so drawing absolute generalizations for a given speaker is dubious.

Third, the audio files submitted were variable in terms of clarity of speech and quietness of the recording environment. Many participants did record in a quiet space, as instructed, but several did not. Background noises sometimes interfere with pitch—for one participant there is a periodic high-pitched beeping sound that obscures the pitch contours of the tones.

Fourth, the phonetic cues associated with tone in Vietnamese vary from dialect to dialect, even within speakers of a single dialect. So, because of the wide range of participants' geographies (representing a similarly wide range of dialects), there was a considerable amount of interspeaker variation in phonetic realization of tone in the study. Along with the different pitch contours associated with Vietnamese tones (see Figure 1), characteristic non-pitch phonetic features of the tones involve laryngeal state (Michaud 2005, Vũ 1981, 1982). Specifically, breathy voice, creaky voice, and glottal constrictions can coöccur with a number of tones, appearing either at the end or in the middle of syllable nucleus. A sketch of the occurrence of laryngeal features in three main Vietnamese dialect groups is laid out by Vũ (1982) and reproduced in Figure 2.

Furthermore, there were some respondents who had sui generis tonal mergers. For instance, some participants were completely merged, while others seemed to be nearly merged (i.e., small but consistent differences between HỎI/C1 and NGÃ/C2), and others still seemed partially/lexically merged—one participant only produced a glottal constriction in one of the three NGÃ/C2 stimuli, while the other two were indistinguishable from her production of the HỎI/C1 tone. These complications are difficult to concretely diagnose from the limited data I have collected. The case of the partial merger, for example, could be analyzed in at least three ways: (1), the tones are merged in certain lexical items while contrastive in others; (2), there's no phonemic difference between the tones, and the pitch/phonation features associated with the HỎI/C1-NGÃ/C2 tone is broadened (i.e., something like free variation); and (3), the tones are underlyingly contrastive, but merged in quick or informal speech. Each of these analyses is consistent with the data collected.

Because of the factors outlined above, determining based on quantitative phonetic measurements if a participant was HỎI/C1-NGÃ/C2 merging or contrasting was greatly impeded. An initial endeavor to categorize participants along this merging/non-merging axis involved creating TextGrid files for the audio to highlight the target word and running a modified Praat script that extracted various phonetic qualities for the vowels (DiCanio 2012), including the first three

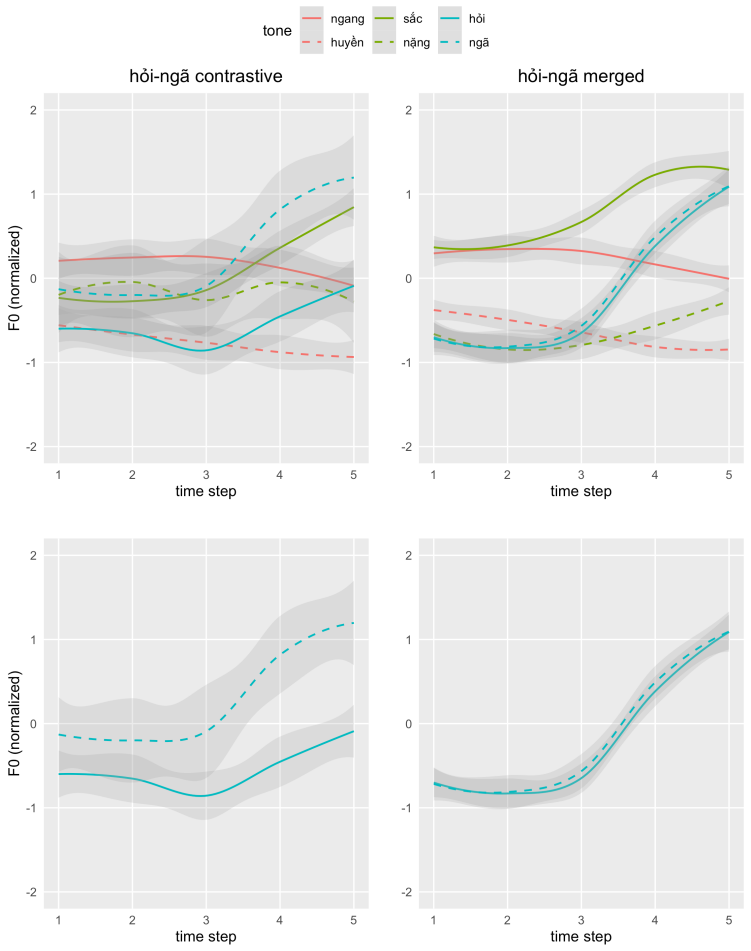


Figure 4. Pitch contours at five normalized timesteps by tone, merging/contrasting, normalized by speaker (Z-score). Top row shows all tones; bottom row shows only HỎI/C1 and NGÃ/C2.

formants, F0, and various measures of the harmonics.[9] There are several acoustic measurements known to correlate with phonation type (see Gordon and Ladefoged 2001 for an overview of these), and it was the hope that some such cue (especially H1-H2), could be measured and demonstrated to correlate with phonation of Vietnamese tones, especially in Northern Vietnamese. After running the script and performing initial analyses, I found that it was frequently impossible to tell from these extracted data alone whether or not a speaker merged HỎI/C1 and NGÃ/C2, due mostly to the limited number of tokens, background noise affecting measurements, etc.

Though certainly a more objective classification procedure is preferable, due to the aforementioned complications and constraints of the current investigation—as well as the fact that the speaking portion of the survey was only meant to inform the analysis of the listening portion, whose results are more pertinent to this thesis—it was determined that this desideratum should

[9]Thanks to Sarah Babinski for help adapting, troubleshooting, and running the script.

be tabled. I opted instead to binarily categorize participants as merging or non-merging by hand. This involved performing a side-by-side comparison of their pronunciations of the HỎI/C1 and NGÃ/C2 tones in Praat, looking at the respective pitch contours, intensity differences, and spectrograms for (a)periodicity in articulation of the target tones. Each speaker was binned as merging or non-merging. Doing this, I arrived at 36 HỎI/C1-NGÃ/C2 merging and 35 HỎI/C1-NGÃ/C2 contrasting participants. Geography was not wholly deterministic of whether or not a speaker merged the tones, but there was a strong association between Southern Vietnamese speakers and merging (35 of 49; 71%), as well as between Northerners and contrasting (19 of 20; 95%). This association was significant (χ^2=25.117; p<0.00001). Southerners' rates of *contrasting* HỎI/C1 and NGÃ/C2 being lower than Northerners' rates of *merging* these tones seems to confirm Brunelle and Jannedy (2013)'s characterization of Northern Vietnamese: "It is the standard variety of the language and benefits from a widespread diffusion in the national media. It is prestigious even outside the North and a certain admixture of [Northern Vietnamese] features by speakers of other varieties is common in formal contexts" (15).

As a test for general validity of my system of by-hand classification of speakers as HỎI/C1-NGÃ/C2 merging or contrasting, I compared the overall pitch contours of participants I had classified as merging to those I'd classified as contrasting. To do so, I extracted the pitch values at five timesteps of participants' production of the target words, excluding the aforementioned cases where the target word was not produced or was produced in a sentence-medial position. Participants' F0 values were normalized by Z-score (weighted by frequency of tones' tokens) and then plotted in R. The results of this are shown in Figure 4, which indicates both a lack of contrast in HỎI/C1-NGÃ/C2 merging participants, as well as a clear distinction between the HỎI/C1 and NGÃ/C2 tones for non-merging participants.

5.3 Perception task

The results of the perception task showed great variability in how participants categorized the tones of the Northern Vietnamese speaker. On aggregate, the tones of the target stimuli were correctly identified at greater-than-chance rates, both by merging and non-merging participants. While both groups performed better than chance, non-merging participants outperformed merging participants in correctly identifying both the HỎI/C1 and NGÃ/C2 tones by 13.2% and 10.2%, respectively. Overall, merging participants perceived the target HỎI/C1 stimuli as HỎI/C1 62.0% of the time and as NGÃ/C2 36.6% of the time. They correctly identified target NGÃ/C2 stimuli at a greater 74.1% rate, perceiving the tone as HỎI/C1 23.6% of the time. Meanwhile, non-merging participants perceived the target HỎI/C1 stimuli as HỎI/C1 75.2% of the time and as NGÃ/C2 21.9% of the time, while they also correctly identified target NGÃ/C2 stimuli as NGÃ/C2 more frequently,

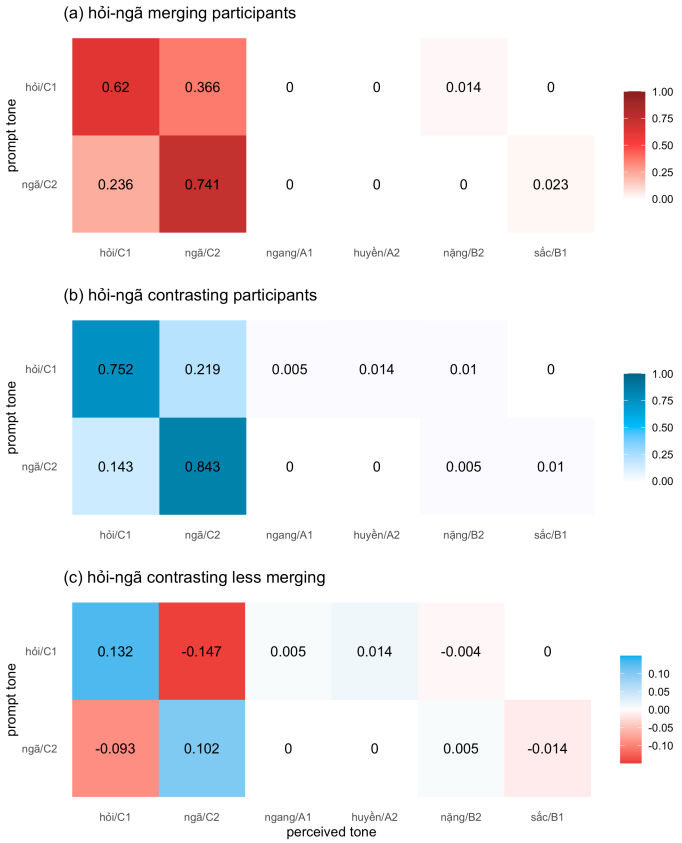


Figure 5. Heatmap of responses for target stimuli by perceived tone, separated by merging vs. non-merging.

at a 84.3% rate, misidentifying them as HỎI/C1 at a 14.3% rate. A summary of the rates of tone identification between merging and contrasting participants for the target stimuli is shown in Figure 5.

This noted difference between rates of accurate tone perception when comparing merging and contrasting participants only becomes clearer when HỎI/C1-NGÃ/C2 contrasting participants are further separated by geography. Southerners who contrast between HỎI/C1 and NGÃ/C2 in their own speech performed at rates comparable to Southerners who merge these tones, while Northerners who contrast generally perform better than both groups of Southerners.[10] This can be seen in Figure 6, which depicts density plots of rates of correct tone identification for target HỎI/C1, target NGÃ/C2, and filler stimuli separated by geography and merging vs. non-merging participants.

The difference between participants' rates of accurate perception for HỎI/C1 and NGÃ/C2 stimuli warrants some interrogation. Overall, 223 of 294 target HỎI/C1 stimuli (75.9%) and 183

[10]The one Northerner who was classified as HỎI/C1-NGÃ/C2 merging was excluded from this portion of the analysis.

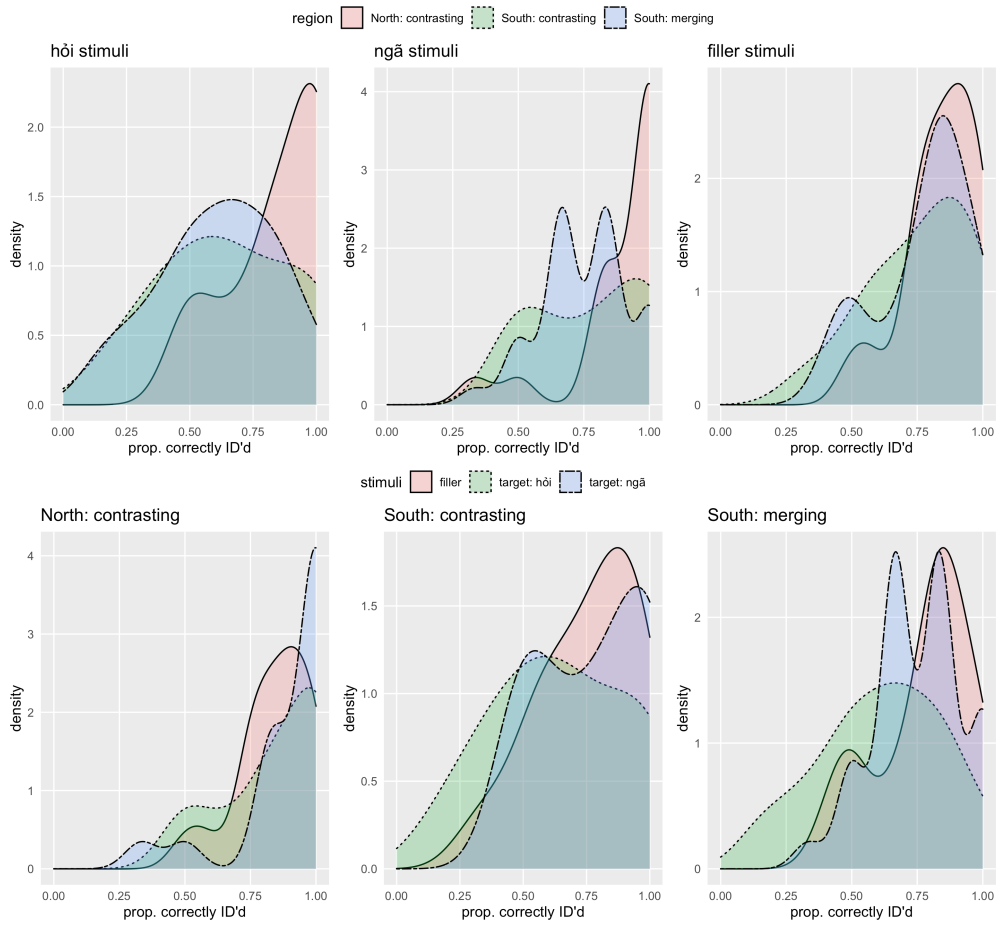


Figure 6. Density plots of correct tone identification for perceptual task stimuli. Top row separated by type stimulus type; bottom row by dialect region and (non-)merging.

of 294 target NGÃ/C2 stimuli (62.2%) were correctly assigned tone by Southerners, representing a statistically significant difference (χ^2=12.7321; p=0.00036). This suggests that Southerners may be more attuned to the non-native laryngeal features of Northern speech, using them to distinguish NGÃ/C2 tone. The fact that Southerners do not typically produce any laryngeal features in their own realizations of the NGÃ/C2 tone likewise lends credence to the idea that, upon hearing a Northerner produce a HỎI/C1 tone stimulus without any such laryngeal features, Southern Vietnamese speakers are more likely to inaccurately categorize it as NGÃ/C2.

When participants misperceived tone, they generally did so without much regularity across the stimuli (i.e., there was no participant who consistently misidentified one target tone as the other). Summing across all target stimuli, in the vast majority of instances, a misperceived HỎI/C1 stimulus was interpreted as NGÃ/C2 and vice versa. However, there was a great deal of variation between individual stimuli. In particular, *mủ* (HỎI/C1), *chửa* (HỎI/C1), and *rữa* (NGÃ/C2) were more likely to be misperceived than other stimuli.

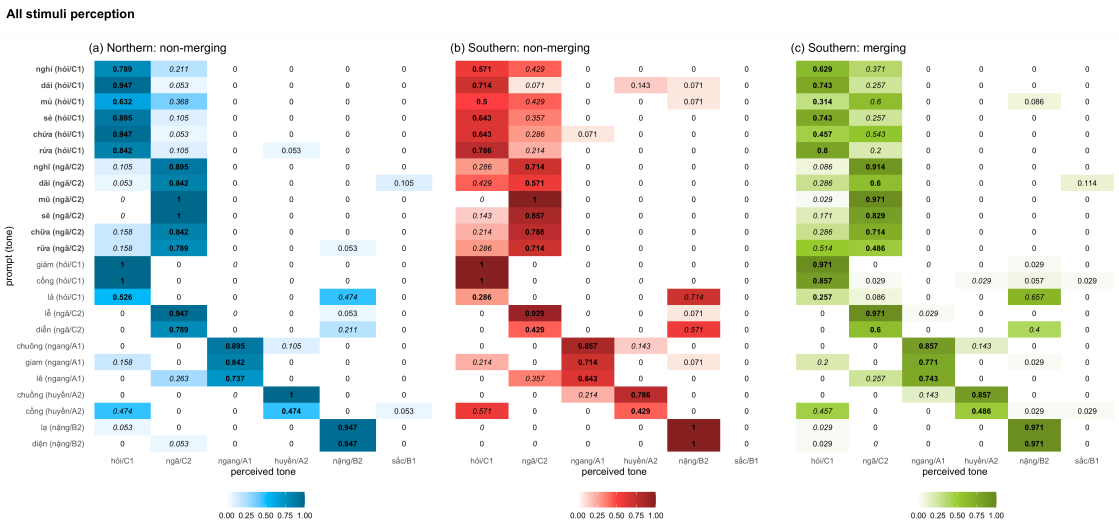


Figure 7. Heatmap of responses for each stimuli by perceived tone, separated by geography and merging vs. non-merging. Prompts appearing in boldface along *y*-axis are target stimuli, plain are filler. Response rates shown in bold represent the correct tone for the stimulus; those shown in italics are the partner tone. n_a=19, n_b=14, n_c=35.

More specifically, *chửa* (HỎI/C1) was perceived as having NGÃ/C2 tone by 46.9% of Southern participants, while *rữa* (NGÃ/C2) was classified as HỎI/C1 by 44.9% of Southern participants. These stimuli were apparently less difficult for non-merging participants, however, who correctly identified their tones at rates of 94.7% and 78.9%, respectively. Of all the target stimuli, *mủ* (HỎI/C1) was the most likely to be misperceived by Northerners, 36.8% of whom believed it had NGÃ/C2 tone. This rate for Southerners was also high, at 55.1%.

In many cases, perception of filler stimuli also seemed to vary from one stimulus to another. The *lả* (HỎI/C1) stimulus was a minimal pair to the *lạ* (NẶNG) stimulus, and *lả* (HỎI/C1) was confused for *lạ* (NẶNG) by a majority of Southern participants (67.3%) and nearly half of Northern participants (47.4%). The *cồng* (HUYỀN) stimulus was roughly as likely to be perceived as having HỎI/C1 tone as it was to be correctly perceived as HUYỀN tone, both by Southern participants (49.0% vs. 46.9%) and Northern participants (47.4%, each). The two filler stimuli with NẶNG tone were overwhelmingly correctly identified by both merging and non-merging participants, with only two participants (one Southern and one Northern) misidentifying each stimulus.

Why it is that these differences can be found in the rates of accurate perception between stimuli consisting of the same tonal pairs is difficult to say. There are a couple possible explanations. One is that the stimuli themselves were spoken in a somewhat ambiguous way. This seems unlikely, however, since each of the stimuli were recorded and played for participants twice. Another explanation is that the frequencies of the target words used in stimulus pairs were different

enough that the perceptual results were biased in favor of the more frequent word. Similarly, there may have been some words that are more common in Northern Vietnamese dialects than in Southern Vietnamese ones. Potential effects of (cross-dialectal) frequency are a promising route, but nonetheless beyond the scope of this paper.

One especially striking finding is that, for a substantial majority of all stimuli, the tone that a stimulus is most commonly mistaken for is that of its minimal pair which was used to create its partner stimulus—e.g., *lả* (HỎI/C1) being mistaken for *lạ* (NẶNG) in the stimulus *Có nhiều người lả/lạ đi trong phòng khám* ('There are a lot of strange/exhausted people going into the clinic.') This can be seen most clearly in contrasting target stimuli against cases of filler stimuli which had one of the same tones as the target stimuli: either HỎI/C1 or NGÃ/C2. For Southern merging participants, target HỎI/C1 stimuli were misperceived as NGÃ/C2 and vice versa at a 30.0% rate, overall. For Southern non-mergers, this figure was 26.2%, and for Northern non-mergers, it was 11.4%. However, the filler stimuli with one of either HỎI/C1 or NGÃ/C2 tone (of which there were five) were misperceived as having NGÃ/C2 or HỎI/C1 (respectively) at a rate of just 1.7% for Southern merging participants, 3.0% for Southern non-merging participants, and 1.8% for Northern non-merging participants. This translates to HỎI/C1 being misperceived as NGÃ/C2 or NGÃ/C2 being misperceived as HỎI/C1 roughly 20 times more frequently in target stimuli than in filler stimuli across all participants. This strongly implies that not only phonetics cues, but also syntactic and semantic cues effectively circumscribe the words speakers deem to be plausible candidates in lexical processing.

This is summarized in Table 7, which shows the rates at which Southern merging, Southern contrasting, and Northern contrasting participants: (1), accurately recovered the tone of the stimulus, (2) inaccurately recovered the tone of the stimulus as the tone of the partner stimulus, and (3), inaccurately recovered the tone of the stimulus as some other tone. On aggregate, the perception of a stimulus's tone as being different from its actual tone and that of its partner stimulus was very similar across merging and non-merging, Southern and Northern participants alike, for both target and filler stimuli.

Although on average Southern participants were able to accurately identify these tones when spoken by a HỎI/C1-NGÃ/C2 contrasting Northerner at rates greater than chance, this was not the case for every Southern participant. Recall Herold (1990)'s typology of *hearing*, *sorting*, and *identifying*. Adopting her terminology, we can say that speakers who are able to accurately choose the tone of target stimuli can *identify* a Northerner's HỎI/C1 and NGÃ/C2 tones. Likewise, speakers who consistently choose HỎI/C1 tone for NGÃ/C2 target stimuli and vice versa could be said to *sort* the tones. Those who can distinguish the HỎI/C1 and NGÃ/C2 tones but inconsistently apply target stimuli to either category would be considered able to *hear* the difference between

participant		stimulus	perceived tone		
			correct tone	incorrect tone	
				partner tone	other tone
North non-merging		target-C1	84.2%	14.9%	0.9%
		target-C2	89.5%	7.9%	2.6%
		filler	84.2%	14.9%	0.9%
South	non-merging	target-C1	64.3%	29.8%	6.0%
		target-C2	77.4%	22.6%	0.0%
		filler	75.6%	23.2%	1.2%
	merging	target-C1	61.4%	37.1%	1.4%
		target-C2	75.2%	22.9%	1.9%
		filler	77.6%	19.5%	2.9%

Table 7. Rates of tone perception for HỎI/C1-NGÃ/C2 contrasting and merging participants, separated by geography and by type of stimuli. 'Partner tone' refers to the tone of each stimulus's minimal pair stimulus.

the tones. Lastly, we could say that participants who consistently apply only one label of either HỎI/C1 or NGÃ/C2 to all target stimuli are *merged in perception*.

Of the 35 HỎI/C1-NGÃ/C2 merging Southern Vietnamese participants in my sample, there are three speakers who could be classified as HỎI/C1-NGÃ/C2 *identifying*. Although none consistently *sorted* the tones, one additional merging Southern participant could be said to *hear* the tones, as he correctly identified the tone of three target stimulus pairs and flipped the tones of the other three pairs, marking the HỎI/C1 as NGÃ/C2 and vice versa. No merging Southern participant always merged the tones in their perception to a single tone, but two did consistently merge stimulus pairs, both doing so in five cases to the NGÃ/C2 tone and once to the HỎI/C1 tone. However, these two participants, along with a few others, were merged for all of the stimuli pairs (i.e., both target and filler stimuli). This could indicate one of two things about these participants: (1), they were just bad at distinguishing between the tones of Northern Vietnamese speakers overall; or (2), they were listening only to the first stimulus in each pair that the survey randomly dealt them and replicating this answer in the second stimulus in each pair. Without more detailed information than my survey offered, it is difficult to say which of these explanations is the right one. Most participants' responses to the perception portion of the study fell somewhere between the extremes of exclusively identifying, sorting, and merging in perception, often correctly reproducing the tonal contrast for some stimulus pairs but not others.

It's worth considering in greater detail those study participants who merged HỎI/C1 and NGÃ/C2 in production but were able to accurately identify these tones in perception. At first glance, this presents a challenge for a theory of such speakers' mental representations of these

tone	Northern Viet.				Southern Viet.			
	high	low	tense	glot.	high	low	tense	glot.
NGANG/A1	+	-	-	-	+	-	-	-
HUYỀN/A2	-	+	-	-	-	+	-	-
SẮC/B1	+	-	+	-	+	-	+	-
NẶNG/B2	-	+	+	+	-	+	+	
HỎI/C1	-	+	+	-	+	+	+	-
NGÃ/C2	+	+	+	+	+	+	+	

Table 4. Rough sketch of distinctive features for Northern and Southern Vietnamese tones.

sounds. What underlying structure in the mind could give rise to this sort of situation? Given these speakers' phonetic realizations of tonal contrast, a five-way tonemic split is tempting since we see no evidence for an alternative. However, data from the perceptual side points toward a six-way tonemic analysis. The latter seems preferable, since it could account for both the production and perception sides of the problem.

Returning to the tentative outline of distinctive features for Northern and Southern Vietnamese tonemes presented in §2.1.1 (with Table 4 recopied here), we can begin to better understand the phenomenon of merged-in-production-but-not-perception participants. This pattern could be explained by the featural specifications I've proposed for Southern Vietnamese speakers. Specifically, while both the HỎI/C1 and NGÃ/C2 tonemes are specified as [+high], [+low], and [+tense], the former is further specified as underlyingly [-glottalic], whereas the latter is underspecified for this feature. As such, in production the NGÃ/C2 toneme would be realized identically to the HỎI/C1 toneme if we are to reasonably assume that a negative specification is the unmarked value for the [±glottalic] feature in Southern Vietnamese.[11] Happily, my featural specifications laid out here could also account for the tendency noted in Brunelle (2009b) for the Northern Vietnamese realization of HỎI/C1 tone to be perceived as NẶNG/B2 by Southern speakers when produced in isolation.

In view of the fact that accurate identification rates for the HỎI/C1-NGÃ/C2 tonal contrast vary massively from one merging speaker to another, the exact tonemic inventory and featural specifications thereof ought to vary as well. For speakers who merge not only in production but also perception, for instance, it is likely the case that a five-way toneme contrast (lacking the underlying NGÃ/C2 toneme) is reflected in their abstract mental representations of tones as well.

The precise details of this analysis remain unclear, especially since this study is not capable of offering more explicit insight into speakers' internal categorizations of tonal contrasts.

[11]The treatment of Southern Vietnamese's NẶNG/B2 toneme as underspecified for the [±glottalic] feature reflects Vũ (1982)'s findings that this tone in Southern Vietnamese is variably produced with modal and creaky voice.

One promising diagnostic could be to investigate processes involving tone sandhi in Vietnamese reduplication. Phạm and Phạm (2020) examined this question, finding that Southern Vietnamese speakers' assignment of tone in reduplicants with NGÃ/C2 bases patterned identically to their assignment of tone for HỎI/C1 bases in a majority of cases, contra what a six-toneme analysis might predict. Further research would be required to resolve this apparent tension between my tentative analysis and Phạm and Phạm's findings.

5.4 Statistical analysis

The results of the binary logistic regression modeling technique described in §4.2.2 are presented here. Analyses showed that which dialect region participants were from is a significant factor for rates of correct identification of target-HỎI/C1 stimuli (p=0.00003), target-NGÃ/C2 stimuli (p=0.00289), and filler stimuli (p=0.0248). Whether participants merged or contrasted HỎI/C1 and NGÃ/C2 in their production of the tones was also a significant factor for target-HỎI/C1 stimuli (p=0.00200) and target-NGÃ/C2 stimuli (p=0.0234), but not filler stimuli (p=0.303). The interaction of participant region and production of the merger was also significant for target-HỎI/C1 (p=0.00155), target-NGÃ/C2 (p=0.0234), and filler stimuli (p=0.0334). In the case of the region:production effect, Northern contrasting speakers were the odd group out, performing better than both Southern contrasting speakers and Southern merging speakers in accurately identifying tone. These results are summarized in Table 5.

The next set of analyses were meant to gauge whether rates of accurate tone identification differed between the tones of the stimuli, particularly between target-HỎI/C1, target-NGÃ/C2, and filler stimuli. These tests were run separately for Northerners, Southerners, mergers, non-mergers, and the three combinations thereof (excluding Northern mergers, since there was only one of these in the participant pool). The results showed that there were, in fact, statistically significant differences in how well participants were able to accurately identify tone across the different tones of the stimuli. Specifically, for the groups that had such discrepancies, it was the target-HỎI/C1 stimuli that were misperceived at higher rates than either the target-NGÃ/C2

factor(s)	target-c1	target-c2	filler
region	***	**	*
production	**	*	
region:production	**	*	*

Table 5. Degrees of statistical significance for rates of accurate perception for stimulus tone groupings as affected by factor. Three asterisks *** denote effect is significant at p<0.001; ** p<0.01; * p<0.05.

factor	region	prod.	region:production		
			n:con.	s:con.	s:mer.
target-c1	***	***			***
target-c2					
filler					

Table 6. Degrees of statistical significance for rates of accurate perception with tone of stimulus as a factor. Tests performed separately by region, production of merger, and interaction of these two. Three asterisks *** denote effect is significant at p<0.001; ** p<0.01; * p<0.05.

or the filler stimuli. This was true for Southerners (p<0.00001) and mergers (p<0.00001), and therefore also for Southern mergers (p<0.00001). It was not the case for Northerners, Southern non-mergers, and therefore all non-mergers, with of these groups accurately identifying tone at comparable rates between target-HỎI/C1, target-NGÃ/C2, and filler stimuli.

6 Discussion

Generally, the results of my study indicated a high degree of variation in Southern Vietnamese speakers' ability to correctly identify non-native tonal contrasts of Northern Vietnamese speech. In aggregate, speakers who merge the HỎI/C1 and NGÃ/C2 tones in production displayed some level of competence in distinguishing these tones in perception, although they were better at distinguishing non-HỎI/C1-NGÃ/C2 pairs and were outperformed by HỎI/C1-NGÃ/C2 contrasting speakers of Northern Vietnamese. Interestingly, whether or not a speaker merged HỎI/C1 and NGÃ/C2 was not as important a factor in determining their ability to accurately discern these tones as was their geography. Non-merging participants from provinces in the Southern dialect region of Vietnam performed generally at similar rates to merging participants from these provinces.

For Southern Vietnamese participants, stimuli with HỎI/C1 tone were more likely to be misperceived than those with NGÃ/C2 tone, suggesting that Southern Vietnamese speakers, who do not natively laryngealize the NGÃ/C2 tone, are sensitive to this laryngealization when produced by Northern speakers. This is consistent with the findings of Kirby (2010), who reports that,

> [D]espite only a limited familiarity with cues to voice quality, [Southern Vietnamese] listeners do show some sensitivity to these cues… [T]he presence of an unfamiliar cue like voice quality may actually aid [Southern Vietnamese] listeners when discriminating between two syllables where one is laryngealized and one is not. (3755)

Another tone that heavily features laryngealization in Northern Vietnamese is NẶNG/B2. Similar to with NGÃ/C2, filler stimuli with NẶNG/B2 tone used in my experiment were correctly

identified by Southern Vietnamese participants much more frequently than were their respective partner stimuli (98.0% vs. 40.8%). Overall, in cases where words with NẶNG/B2 tone were contextually viable, Southern participants were more likely to perceive HỎI/C1 stimuli as NẶNG/B2 than they were to perceive NGÃ/C2 stimuli as NẶNG/B2 (67.3% vs. 44.9%). This was also true for Northern participants (47.4% vs. 21.1%). It is not clear whether this is a robust generalization, since there was only one stimuli pair for each HỎI/C1-NẶNG/B2 and NGÃ/C2-NẶNG/B2; however, this finding would confirm the results of Brunelle (2009b), which posits that the HỎI/C1-as-NẶNG/B2 confusion is a product of the similarity of these tones' cross-dialectal pitch contours, where Northern Vietnamese HỎI/C1 and Southern Vietnamese NẶNG/B2 are both falling-rising. A trend that becomes clear when comparing the results of this study to those of earlier research is that sentential context plays a significant role in structuring cross-dialectal communication in Vietnamese. Brunelle (2009b) found that Southern Vietnamese speakers identified the HỎI/C1 tone of a Northern Vietnamese speaker as NẶNG/B2 87% of the time when spoken in isolation. My study indicated that, when placed in a sentence where only HỎI/C1 or NGÃ/C2 are viable, Southerners correctly identify HỎI/C1 tone a majority of the time, only misidentifying the tone as NẶNG/B2 for five out of these 294 stimuli (1.7%). As discussed in the previous paragraph, however, sentences where HỎI/C1 and NẶNG/B2 are both viable result in much higher degrees of misperception of HỎI/C1 for NẶNG/B2. This evidence indisputably affirms the hypothesis that cross-dialectal communication in Vietnamese proceeds not merely on the grounds of the *access* stage of lexical processing (involving the activation of lexical items based on phonetics), but that the *integration* stage (involving activation based on the syntactic/semantic environment) plays a major role in shaping lexical processing.

This fits well with one of the main findings of (Vũ 1981: 166-221), namely, that when lexically primed to hear a single tone, Southern Vietnamese listeners correctly identify that tone in the speech of Northern Vietnamese speakers at rates well above when they are not lexically primed to perceive tone. Where my study diverges from Vũ's, though, is that mine allows us to more precisely disentangle, measure, and compare the degree to which the access and integration stages of lexical processing each influence cross-dialectal speech perception in Vietnamese. This is because the sentential contexts of my experimental stimuli ambiguously prime participants to perceive either of two tones. In comparison with the results of Vũ (1981), which finds Southern Vietnamese participants to accurately identify Northern Vietnamese tones at rates of over 95%, my study places this figure at 73%, implying that ambiguous syntacto-semantic priming is weaker than unambiguous priming in its effect on lexical processing.

In terms of their production of tone, most Southern Vietnamese study participants (35 of 49) were found to merge HỎI/C1 and NGÃ/C2. Of the non-merging Southern participants, several only

produced irregular contrasts between the two tones—either only pronouncing a distinction for some words but not others, or assigning the tonal contours of HỎI/C1 or NGÃ/C2 sporadically in their realization of tone across both HỎI/C1 or NGÃ/C2 stimuli. With regard to the merger typology articulated in Labov (1994), the first of these sorts of irregularities seems to support an analysis of the tones as merged under *transfer* (i.e., lexically irregular, phonetically discrete), whereas the second could be merged under *expansion* (i.e., a collapse of a phonemic contrast whereby the breadth of the phonetic space previously used to produce both phonemes independently is used in realizing the now merged phoneme). Several speakers produced minor but consistent differences in their realizations of the tones, a fact which could lend itself to an analysis of merger under *approximation* (i.e., lexically regular, phonetically gradual). Either way, these results seem to confirm the idea that speakers of Southern Vietnamese sometimes make use of features of Northern Vietnamese, likely due to its status as a prestigious variety (Brunelle and Jannedy 2013), albeit without much regularity.

On the perceptual side of things, the noted variation in HỎI/C1-NGÃ/C2 merging speakers' perception of these tones when produced by a non-merging, Northern speaker indicates a range of cross-dialectal competence in tone perception across study participants. In terms of Herold (1990)'s distinction between *hearing*, *sorting*, and *identifying* phonemic contrasts, there was a number of speakers who were merged in production but fit into both the *hearing* and *identifying* categories. The first of these groups were consistently able to tell the HỎI/C1 and NGÃ/C2 stimuli apart, but only sometimes correctly classified the tones of these pairs (i.e, they would sometimes assign HỎI/C1 tone to NGÃ/C2 stimuli or vice versa). The second were similarly able to discern the HỎI/C1 stimuli from the NGÃ/C2, and were furthermore consistently accurate in their categorization of the tones. Herold (1990)'s final category of *sorting* (whereby a listener consistently hears a contrast and incorrectly perceives *A* as *B* and *B* as *A*) was not applicable to any speakers in the study. A handful of merging participants could also be said to be entirely merged in production, in that they always applied the same tone to HỎI/C1-NGÃ/C2 stimulus pairs (though this could be due to a flaw in the experimental design; see §4.2). The majority of merging participants could not be neatly binned into groups described these four labels, and instead their perception was commonly characterized by a mix of merging, sorting, and identifying the target stimulus pairs.

6.1 Answering *Research questions*

The results of the study indicate the following with respect to each to the questions posed in §3:

1. *Are Vietnamese speakers who merge the* HỎI/C1 *and* NGÃ/C2 *tones in production able to perceive a phonemic contrast (i.e.,* sort *or* identify *in the Heroldian sense) between these tones when pronounced by non-merging speakers in naturalistic contexts?*

Yes. As a group, the Vietnamese speakers who participated in this study and have the HỎI/C1-NGÃ/C2 merger in production were able to, at rates better than chance, discriminate between and correctly identify the tones of a non-merging, Northern Vietnamese speaker reading sentences where syntactics and semantics alone could not adequately cue the listener in on whether a HỎI/C1-NGÃ/C2 competitor word had been uttered.

2. *Inasmuch as* HỎI/C1-NGÃ/C2 *merging speakers are able to perceive a difference in the tones when produced by speakers of a non-merging dialect, how accurate is their identification of these tones?*

Somewhat. In aggregate, merging participants were found to accurately identify tone for roughly two-thirds of all target stimuli (68.1%). Despite performing better than chance, however, merging participants were still worse at accurately perceiving the difference between HỎI/C1 and NGÃ/C2 stimuli than non-merging participants were, who accurately identified the tone of nearly four-fifths of all target stimuli (79.8%). The magnitude of this perceptual difference grew when taking participant geography into account; comparing the perception task results of HỎI/C1-NGÃ/C2 merging Southerns, non-merging Southerners, and non-merging Northerners, we see that both groups of Southerners performed at similar rates of 68.3% and 70.8%, respectively, while for non-merging Northerners this rate was 86.8%.

3. *Do asymmetries exist in tonal perception?* That is, *Is one of either the* HỎI/C1 *or* NGÃ/C2 *tones more likely to be misperceived by merging speakers than the other is?*

Yes. Target stimuli containing a HỎI/C1 tone were more likely to be misidentified by Southern merging participants as NGÃ/C2 than vice versa, with perceptual accuracy rates of 61.4% and 75.2%, respectively. This was not true for Northern speakers who contrasted HỎI/C1 and NGÃ/C2 in production, who had similar accuracy rates of 84.2% and 89.5% for these tones.

4. *Can interspeaker variation among merging participants be found in perception of the* HỎI/C1 *and* NGÃ/C2 *tones?*

Yes. On the whole, merging speakers were able to accurately perceive the HỎI/C1-NGÃ/C2 tonal contrast better than chance. However, this was not the case for all merging participants, and the rates of correct identification varied greatly across study participants. Six of the 74 participants accurately identified the tone of every stimuli, while five actually performed worse than chance.

5. *To what extent does the sentential context of a word with* HỎI/C1 *or* NGÃ/C2 *tone play a role in structuring tone perception?*

A great extent. Of the 1,776 total responses to individual perceptual stimuli, just 33 (1.9%) of these corresponded to tones that were neither the actual tone or its "partner" tone (i.e., the incor-

rect tone that nonetheless corresponded to a syntactically and semantically viable word for the sentential context).

6. *Are there measurable effects of exposure to Northern Vietnamese dialects on the degree to which merging speakers are able to recover the tonal distinction in perception?*

Not enough information. It was anticipated that heritage Vietnamese speakers and those living in the diaspora would constitute a main demographic for this study. Such speakers would likely have less exposure to Northern Vietnamese than speakers of the same dialect background having been raised and currently living in Vietnam. As such, in comparing rates of accurate tonal perception between diasporic/heritage Vietnamese and those in the country, it would have been possible to test the effect of contact on tone perception. However, only one of the 74 respondents was born outside of Vietnam, and a similarly meager thirteen indicated that they currently reside abroad.

7. *In term of Labov (1994)'s three-way typological split, what (if any) evidence is found in support of (or against) analyses of the merger as happening via approximation, transfer, or expansion?*

Some tentative evidence exists. While most participants were easy to categorize as HỎI/C1-NGÃ/C2 merging or contrasting, some presented challenges. A handful of speakers produced small but consistent differences between the words with HỎI/C1 and NGÃ/C2 tones. Three participants (all Southern) were inconsistent in their use of laryngeal features for the two tones. One of these three produced both *đỉa* (HỎI/C1) and *rễ* (NGÃ/C2) with the same pitch contour and a glottal constriction, while producing *đĩa* (NGÃ/C2) and *rổ* (HỎI/C1) with a different pitch contour and no laryngealization. For another speaker, *đỉa* (HỎI/C1) and *đĩa* (NGÃ/C2) were homophonous, but *lỗ* (NGÃ/C2) was pronounced with a clear glottal constriction.

These limited data seem to indicate that there is not a single unifying mechanism across all merging speakers. For the first group (who produced small but consistent differences between HỎI/C1 and NGÃ/C2), these speakers' pronunciations could be in line with an analysis of their HỎI/C1 and NGÃ/C2 tones as merged under approximation. The second group (who were inconsistent in their use of laryngeal features) could be analyzed as merged under either transfer or expansion. To say with any degree of certainty, though, would require more thorough data on how these participants produce tones, instead of maximally six tokens, which is what the production portion of this study provided.

8. *If found to exist, what would the implications of asymmetries in the* HỎI/C1-NGÃ/C2 *merger between production and perception be with regard to speakers' phonemic representations of these tones?*

Two underlying representations. Speakers who merge in production but not in perception could reasonably be said to have two separate tonemic representations for the HỎI/C1 and NGÃ/C2 tones. A single representation for both of these tones would fail to account for the fact that speakers are able to recover the tonal contrast when produced by a non-merging speaker. Given the cross-speaker range in accuracy when identifying the HỎI/C1-NGÃ/C2 tonal contrasts, it is hypothesized that the phonetic cues associated with the NGÃ/C2 toneme should also reflect this cross-speaker range from highly to somewhat to not-at-all specified.

6.2 Caveats & considerations

There are some limitations to this study that I will outline here. Firstly, issues of sampling bias may have been present. Beyond the fact that access to the internet was a prerequisite for participation, there were also statistically significantly higher rates of survey completion found in participants with at least ten years of education in Vietnamese compared to those without, as well as between participants born in Vietnam and those born abroad. Only one participant who completed the survey had under ten years of schooling in Vietnamese, and an additional one was born outside the country of Vietnam. As such, the findings and generalizations herein discussed may not be applicable to less educated or heritage speakers of Southern Vietnamese, or to those without internet access.

The other main issue for this study was the nature of the data collection. In some ways, the fact that the study participants was accessible online was preferable, as this allowed for a higher volume of participation. Conducting the experiment in-person would have made it possible to control and standardize the setting and recording equipment of the responses, but would necessarily restrict the sample to bilingual Vietnamese-English speakers living in the United States since it would not have been feasible to hire a trained native speaker of Vietnamese to conduct the experiment or collect responses in-person in Vietnam.

One oversight in experimental design was that no perception task stimuli included the sắc/B1 tone, which Brunelle (2009b) found Southern Vietnamese speakers commonly misperceived Northern Vietnamese NGÃ/C2 stimuli as being. Similarly, although only HỎI/C1-NGÃ/C2 pairs were considered target stimuli, of the fourteen possible filler tonal pairs, only six were used (A1-A2, A1-C1, A1-C2, A2-C1, B2-C1, and B2-C2), meaning that participants were not tested on their perception of eight pairwise tonal contrasts (A1-B1, A1-B2, A2-B1, A2-B2, A2-C2, B1-B2, B1-C1, and B1-C2). With the exception of the sắc/B1-NGÃ/C2 tone pair, it is not suspected that including any other pair would have led to especially interesting results since none of these pairs led Brunelle (2009b) to especially interesting findings.

Additionally, word frequency is expected to play some role in structuring speech perception, with more commonly occurring words being more likely to be selected in lexical processing than less commonly occurring ones. Given this, the noted discrepancies between the rates of accurate tone identification across the experiment's stimuli could tentatively be accounted for by the relative frequencies (both in Vietnamese overall and cross-dialectally) of each word in a stimulus pair. However, there is not yet an obvious way of accounting for effects of word frequency since no good source detailing word-frequency in Vietnamese exists yet, to my knowledge (Brunelle and Jannedy 2013: 20). Further evidence would be needed in (dis)confirming this hypotheses.

Finally, responses to the production task portion of the survey did not paint as clear a picture of whether a given participant ought to be considered HỎI/C1-NGÃ/C2 merging or contrasting. The maximum number of elicited words with each of these tones was three for each participant, and many speakers did not consistently respond with the expected target word. The number of stimuli was kept low in order to allow participants to complete the survey in a reasonably short amount of time (around 15 minutes was the goal), which in turn was meant to increase overall participation. Whether it would have been preferable to include more stimuli at the potential expense of a high participation rate it's difficult to say, but the answer is not important.

7 Conclusion

This study presents evidence that speakers of Southern Vietnamese show a high degree of variation in their capacity to accurately perceive the non-native HỎI/C1-NGÃ/C2 contrast of Northern Vietnamese. Nonetheless, Southern Vietnamese participants on average perform better than chance at identifying the HỎI/C1 and NGÃ/C2 tones as produced by a Northern Vietnamese speaker. Overall, words with HỎI/C1 tone were more likely to be misperceived as having NGÃ/C2 tones than vice versa by Southern Vietnamese speakers, which suggests that merging speakers of Southern Vietnamese are attuned to the salient laryngeal features of the NGÃ/C2 tone in Northern Vietnamese speech that do not occur in their own native dialect.

Furthermore, by comparing the results of this study to the prior findings of Brunelle (2009b) and Kirby (2010), it becomes clear that sentential context influences cross-dialectal tone perception in Vietnamese to a great extent. What else is apparent is that there does not seem to be a single mechanism by which the HỎI/C1-NGÃ/C2 merger can be said to have occurred in Southern Vietnamese speakers—this study identifies cases in which the merger could be analyzed as having taken place *by approximation*, *by transfer*, and *by expansion*, alike (Herold 1990, Labov 1994). Similarly, adopting terms from the typology of perception put forth in Herold (1990), results from

the perception task portion of this study indicate that some speakers who merge the HỎI/C1 and NGÃ/C2 tones in production are able to *identify* these tones in Northern Vietnamese speech, while others are only able to *hear* the difference between them, and others still can do neither, and have these tones entirely merged in perception. Most, however, lie somewhere between these three extremes.

This study adds to a growing amount of literature on cross-dialectal communication in Vietnamese (Brunelle 2009b, Brunelle and Jannedy 2007, 2013, Kirby 2010). Although not the primary focus of this paper, with audio data from 49 Southern Vietnamese participants, my study represents—to my knowledge—the largest descriptive account dealing with acoustic properties of the tones and tonal contrasts of the relatively poorly studied Southern Vietnamese dialects, to date. My findings also substantiate previous claims that Southern Vietnamese speakers do occasionally incorporate dialect features of Northern Vietnamese into their own speech, at least with regard to tonal contrasts (Brunelle and Jannedy 2013: 15).

Though still quite preliminary, this study also demonstrates that asymmetries between production and cross-dialectal perception of tone in Vietnamese could theoretically be accounted for in terms of cross-dialectally variably-specified distinctive features. Further research in this arena is recommended.

References

Alves, M. (1995). Tonal Features and the Development of Vietnamese Tones. *Working Papers in Linguistics: Department of University of Hawaii at Manoa*, 27:1–13.

Alves, M. J. (2007). A look at North-Central Vietnamese. In Wayland, R., Hartmann, J., and Sidwell, P., editors, *SEALS XII: papers from the 12th annual meeting of the Southeast Asian Linguistics Society (2002)*, pages 1–7. Pacific Linguistics, Research School of Pacific and Asian Studies, The Australian National University, Canberra. OCLC: 929725400.

Austen, M. (2020). Production and perception of the PIN-PEN merger. *Journal of Linguistic Geography*, 8:115–126.

Baranowski, M. (2013). On the role of social factors in the loss of phonemic distinctions. *English Language and Linguistics*, 17(2):271–295.

Bauman, J., Blodgett, A., Rytting, C. A., and Shamoo, J. (2009). The ups and downs of Vietnamese tones: A description of native speaker and adult learner tone systems for Northern and Southern Vietnamese.

Bowie, D. (2000). *The Effect of Geographic Mobility on the Retention of a Local Dialect.* Doctoral, University of Pennsylvania, Philadelphia.

Brunelle, M. (2009a). Northern and Southern Vietnamese Tone Coarticulation: A comparative case study. *Journal of the Southeast Asian Linguistics Soceity*, 1:49–62.

Brunelle, M. (2009b). Tone perception in Northern and Southern Vietnamese. *Journal of Phonetics*, 37:79–96.

Brunelle, M. and Jannedy, S. (2007). Social Effects on the Perception of Vietnamese Tones. Saarbrücken.

Brunelle, M. and Jannedy, S. (2013). The cross-dialectal perception of Vietnamese tones: Indexicality and convergence. In Hole, D. and Löbel, E., editors, *Linguistics of Vietnamese: An international survey*, volume 253 of *Trends in Linguistics: Studies and Monographs*, pages 9–34. De Gruyter Mouton, Berlin.

Brunellière, A., Dufour, S., and Nguyễn, N. (2011). Regional differences in the listener's phonemic inventory affect semantic processing: A mismatch negativity (MMN) study. *Brain and Language*, 117(1):45–51.

Brunellière, A., Dufour, S., Nguyễn, N., and Frauenfelder, U. H. (2009). Behavioral and electro-

physiological evidence for the impact of regional variation on phoneme perception. *Cognition*, 111(3):390–396.

Buchwald, J. S., Guthrie, D., Schwafel, J., Erwin, R. J., and van Lancker, D. (1994). Influence of Language Structure on Brain-Behavior Development. *Brain and Language*, 46:607–619.

Cao, X. H. (2003). Hai cách miệu tả hệ thống thanh điệu tiếng Việt. In Cao, X. H., editor, *Tiếng Việt: Mấy vấn đề ngữ âm, ngữ pháp, ngữ nghĩa*, pages 79–87. Nhà Xuất Bản Giáo Dục, Hồ Chí Minh City.

Chang, Y.-H. (2012). *Variability in cross-dialectal production and perception of contrasting phonemes: The case of the alveolar-retroflex contrast in Beijing and Taiwan Mandarin*. Ph.D, University of Illinois at Urbana-Champaign, Urbana, Illinois.

Chen, J., Best, C. T., and Antoniou, M. (2020). Native phonological and phonetic influences in perceptual assimilation of monosyllabic Thai lexical tones by Mandarin and Vietnamese listeners. *Journal of Phonetics*, 83.

Conrey, B., Potts, G. F., and Niedzielski, N. A. (2005). Effects of dialect on merger perception: ERP and behavioral correlates. *Brain and Language*, 95(3):435–449.

Cutler, A., Weber, A., and Otake, T. (2006). Asymmetric mapping from phonetic to lexical representations in second-language listening. *Journal of Phonetics*, 34(2):269–284.

DiCanio, C. (2012). General vowel acoustics script for Praat.

Diffloth, G. (1989). Proto-Austroasiatic creaky voice. *Mon-Khmer Studies*, 15:139–154.

Drager, K. (2010). Speaker Age and Vowel Perception. *Language and Speech*, 54(1):99–121.

Evans, B. G. and Iverson, P. (2004). Vowel normalization for accent: An investigation of best exemplar locations in northern and southern British English sentences. *The Journal of the Acoustical Society of America*, 115(1):352–361.

Fung, R. S. Y. and Wong, C. S. P. (2011). Acoustic Analysis of the New Rising Tone in Hong Kong Cantonese. pages 715–718, Hong Kong.

Gass, S. M. and Varonis, E. M. (1991). Miscommunication in Nonnative Speaker Discourse. In Coupland, N., Giles, H., and Wiemann, J. M., editors, "*Miscommunication*" *and Problematic Talk*. Sage, Newbury Park, California.

Gordon, M. and Ladefoged, P. (2001). Phonation types: a cross-linguistic overview. *Journal of Phonetics*, 29(4):383–406.

Haudricourt, A.-G. (1954). De l'origine des tons en vietnamien. *Journal Asiatique*, (242):69–82.

Hay, J., Drager, K., and Warren, P. (2010). Short-term Exposure to One Dialect Affects Processing of Another. *Language and Speech*, 53(4):447–471.

Hay, J., Warren, P., and Drager, K. (2006). Factors influencing speech perception in the context of a merger-in-progress. *Journal of Phonetics*, 34:458–484.

Herold, R. (1990). *Mechanisms of merger: The implementation and distribution of the low back merger in eastern Pennsylvania*. Ph.D, University of Pennsylvania, Philadelphia.

Janson, T. and Schulman, R. (1983). Non-distinctive features and their use. *Journal of Linguistics*, 19:321–336.

Jensen, J. B. (1989). On the Mutual Intelligibility of Spanish and Portuguese. *Hispania*, 72(4):848–852.

Kazanina, N., Phillips, C., and Idsardi, W. (2006). The influence of meaning on the perception of speech sounds. *Proceedings of the National Academy of Sciences*, 103(30):11381–11386.

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

Koops, C., Gentry, E., and Pantos, A. (2008). The effect of perceived speaker age on the perception of PIN and PEN vowels in Houston, Texas. *University of Pennsylvania Working Papers in Linguistics*, 14(2):93–101.

Labov, W. (1966). *The Social Stratification of English in New York City*. Center for Applied Linguistics, Washington, DC.

Labov, W. (1994). *Principles of Linguistic Change*, volume 1 of *Language in Society*. Blackwell, Malden, Mass., digital print edition.

Labov, W., Karen, M., and Miller, C. (1991). Near-mergers and the suspension of phonemic contrast. *Language Variation and Change*, 3:33–74.

Lanwermeyer, M., Henrich, K., Rocholl, M. J., Schnell, H. T., Werth, A., Herrgen, J., and Schmidt, J. E. (2016). Dialect Variation Influences the Phonological and Lexical-Semantic Word Processing in Sentences. Electrophysiological Evidence from a Cross-Dialectal Comprehension Study. *Frontiers in Psychology*, 7.

Maclagan, M. A. and Gordon, E. (1996). Out of the AIR and into the EAR: Another view of the New Zealand diphthong merger. *Language Variation and Change*, 8:125–147.

Maguire, W., Clark, L., and Watson, K. (2013). Introduction: what are mergers and can they be reversed? *English Language and Linguistics*, 17(2):229–239.

Michaud, A. (2005). Final Consonants and Glottalization: New Perspectives from Hanoi Vietnamese. *Phonetica*, 61:119–146.

Mok, P. P. K., Zuo, D., and Wong, P. W. Y. (2013). Production and perception of a sound change in progress: Tone merging in Hong Kong Cantonese. *Language Variation and Change*, 25(3):341–370.

Nguyễn, A.-T. T. and Ingram, J. C. L. (2008). Stress and tone Sandhi in Vietnamese reduplications. *Mon-Khmer Studies*, 37:15–39.

Nguyễn, T. and Ingram, J. (2006). Reduplication and word stress in Vietnamese. In Warren, P. and Watson, C. I., editors, *Proceedings of the 11th Australian International Conference on Speech Science & Technology*, pages 187–192. University of Auckland, Auckland.

Nguyễn, V. L. and Edmonson, J. A. (1998). Tones and voice quality in modern northern Vietnamese: Instrumental case studies. *Mon-Khmer Studies*, 28:1–18.

Ngô, T. N. (1984). *The Syllabeme and Patterns of Word Formation in Vietnamese*. Ph.D, New York University, New York.

Niedzielski, N. (1999). The Effect of Social Information on the Perception of Sociolinguistic Variables. *Journal of Language and Social Psychology*, 18(1):62–85.

Näätänen, R., Lehtokoski, A., Lennes, M., Cheour, M., Huotilainen, M., Iivonen, A., Vainio, M., Alku, P., Ilmoniemi, R. J., Luuk, A., Allik, J., Sinkkonen, J., and Alho, K. (1997). Language-specific phoneme representations revealed by electric and magnetic brain responses. *Nature*, 385(6615):432–434.

Phạm, A. H. (2003). *Vietnamese tone: a new analysis*. Outstanding dissertations in linguistics. Routledge, New York; London.

Phạm, A. H. and Phạm, A. A. (2020). Productive Reduplication in Southern Vietnamese. page 10.

Ramírez, M. and Simonet, M. (2018). Language dominance and the perception of the Majorcan Catalan /ʎ/–/ʒ/ contrast: Asymmetrical phonological representations. *International Journal of Bilingualism*, 22(6):638–652.

Rivera-Gaxiola, M., Csibra, G., Johnson, M., and Karmiloff-Smith, A. (2000). Electrophysiological correlates of cross-linguistic speech perception in native English speakers. *Behavioural Brain Research*, 111(1-2):13–23.

Sayer, I. M. (2013). Misunderstanding and language comprehension. In *Procedia - Social and Behavioral Sciences*, volume 70, pages 738 – 748, Antalya, Turkey.

Shaw, J. A., Best, C. T., Docherty, G., Evans, B. G., Foulkes, P., Hay, J., and Mulak, K. (2018). Resilience of English vowel perception across regional accent variation. *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, 9(1):11.

Sumner, M. and Samuel, A. G. (2009). The effect of experience on the perception and representation of dialect variants. *Journal of Memory and Language*, 60:487–501.

Tanenhaus, M. K. and Lucas, M. M. (1987). Context effects in lexical processing. *Cognition*, 25:213–234.

Thompson, L. C. (1965). *A Vietnamese Grammar*. University of Washington Press, Seattle, Wash.

Thurgood, G. (2002). Vietnamese and tonogenesis: Revising the model and the analysis. *Diachronica*, 19(2):333–363.

Trudgill, P. and Foxcroft, T. (1978). On the sociolinguistics of vocalic mergers: Transfer and approximation in East Anglia. In Trudgill, P., editor, *Sociolinguistic Patterns in British English*. University Park Press, Baltimore.

Trần, H. M. (1967). Tones and Intonation in South Vietnamese. In Nguyễn, D. L., Trần, H. M., and Dellinger, D., editors, *Series A—Occasional papers #9*, Papers in Southeast Aisan Linguistics no. 1, pages 19–27. Linguistics Circle of Canberra, Canberra.

Vũ, P. T. (1981). *The Acoustic and Perceptual Nature of Tone in Vietnamese*. Ph.D, Australian National University, Canberra.

Vũ, P. T. (1982). Phonetic Properties of Vietnamese Tones Across Dialects. In Bradley, D., editor, *Papers in Southeast Asian Linguistics*, pages 55–75. Australian National University, Sydney.

Weber, A. and Cutler, A. (2004). Lexical competition in non-native spoken-word recognition. *Journal of Memory and Language*, 50(1):1–25.

Wu, J., Chen, Y., Van Heuven, V. J., and Schiller, N. O. (2017). Interlingual two-to-one mapping of tonal categories. *Bilingualism: Language and Cognition*, 20(4):813–833.

Zwitserlood, P. (1989). The locus of the effects of sentential-semantic context in spoken-word processing. *Cognition*, 32:25–64.

Đào, M.-D. and Nguyễn, A.-T. T. (2017). Vietnamese Tones Produced by Australian Vietnamese Speakers. *Heritage Language Journal*, 14(3):224–247.

Appendices

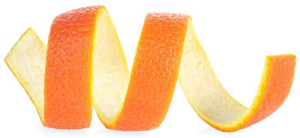
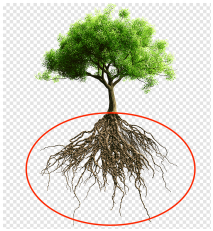
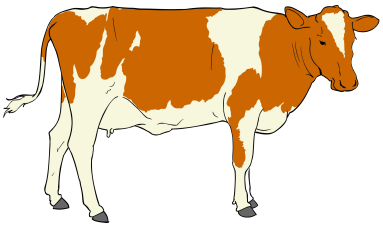
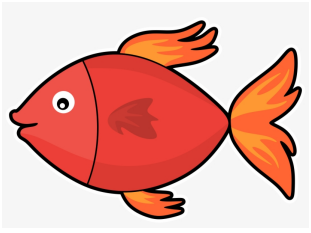
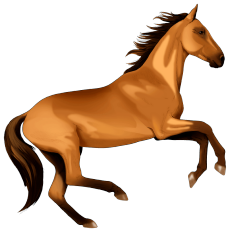
Appendix A: Target Stimuli

	Vietnamese stimuli	English translation
1.	*nghỉ/nghĩ*	'to rest'/'to think'
	Có nhiều tiếng ồn quá. Tôi không ___ được.	'There's too much noise. I can't ___.'
2.	*dải/dãi*	'ribbon'/'saliva'
	Cô gái trẻ có ___ vướng trên tóc.	'The young girl has ___ stuck in her hair.'
3.	*mủ/mũ*	'pimple'/'hat'
	Có một cái ___ trên đầu.	'There's a ___ on your head.'
4.	*sẻ/sẽ*	'sparrow'/'will'
	Nghe tiếng chim ___ làm bạn vui.	'Listening to the sound of a [sparrow makes/bird will make] you happy.'
5.	*chửa/chữa*	'to be pregnant'/'to cure'
	Vợ bị vô sinh nhưng mãi cũng ___ được.	'The wife was infertile, but she got [pregnant/cured].'
6.	*rửa/rữa*	'to wash'/'to decompose'
	Táo này ___ hết rồi.	'The apples were all ___-ed.'

Appendix B: Filler Stimuli

	Vietnamese stimuli	English translation
7.	*chuông/chuồng*	'bell'/'coop'
	Mình không có ___ đủ lớn.	'I don't have loud/big enough ___-s.'
8.	*giam/giảm*	'to imprison'/'to reduce'
	Chính sách mới quyết tâm ___ số trẻ vị thành niên chơi ma tuý.	'The new policy aims to ___ the number of youths abusing drugs.'
9.	*lê/lễ*	'pear'/'festival'
	Nhà tôi có ___ hôm nay.	'My house has ___ today.'
10.	*cồng/cổng*	'gong'/'gate'
	Làng tôi có cái ___ to lớn lắm.	'My village has a huge ___.
11.	*lạ/lả*	'strange'/'exhausted'
	Có nhiều người ___ đi trong phòng khám.	'There's a lot of ___ people going into the clinic.'
12.	*diện/diễn*	'wear (expensive things)'/'act'
	Hôm nay có chuyện gì mà chị ___ quá vậy?	'What's the occasion that you're ___-ing like that today?'

Appendix C: Visual Stimuli

Image	Intended response	Image	Intended response
	*đỉa* (HỎI/C1) 'leech(es)'		*đĩa* (NGÃ/C2) 'plate(s)'
	*rổ* (HỎI/C1) 'basket'		*lỗ* (NGÃ/C2) 'hole'
	*vỏ* (HỎI/C1) 'peel'		*rễ* (NGÃ/C2) 'roots'
	*nho* (NGANG/A1) 'grape(s)'		*bò* (HUYỀN/A2) 'cow'
	*cá* (SẮC/B1) 'fish'		*ngựa* (NẶNG/B2) 'horse'

Appendix D: Participant demographics

demographic category	bin	complete (n=74)		incomplete (n=245)[12]	
		count	prop.	count	prop.
age	18–24	24	0.32	53	0.27
	25–39	44	0.59	121	0.61
	40–54	4	0.05	20	0.10
	55 or over	2	0.03	3	0.02
gender	female	52	0.70	128	0.67
	male	21	0.28	61	0.32
	other	1	0.01	3	0.02
educational attainment	not H.S. grad	2	0.03	8	0.05
	H.S. grad	16	0.22	35	0.23
	uni. grad	33	0.45	86	0.55
	grad school	23	0.31	26	0.17
years of Vietnamese education	0	0	0.00	3	0.02
	1–4	1	0.01	11	0.07
	5–9	0	0.00	12	0.08
	10 or more	73	0.99	129	0.83
born	in Vietnam	73	0.99	144	0.89
	in diaspora	1	0.01	17	0.11
reside	in Vietnam	61	0.82	138	0.83
	in diaspora	13	0.18	29	0.17

Appendix E: Geography of respondents

Northern Dialect Provinces	n 22	Southern Dialect Provinces	n 50
Hà Nội	10	Trà Vinh	11
Phú Thọ	4	Đồng Tháp	7
Nam Định	2	Hồ Chí Minh City	7
Thanh Hoá	2	Bà Rịa-Vũng Tàu	5
Hà Giang	1	Long An	3
Hải Dương	1	An Giang	2
Hải Phòng	1	Cần Thơ	2
Tuyên Quang	1	Kiên Giang	2
		Tiền Giang	2
		Bến Tre	1
		Bình Định	1
		Bình Phước	1
		Bình Thuận	1
		Đồng Nai	1
		Lâm Đồng	1
		Ninh Thuận	1
		Quảng Nam	1
		Vĩnh Long	1

[12]Some incomplete surveys recorded no or only partial demographic data. Proportion column reflects participants that entered data for that demographic category.

Appendix F: Perception task full results: target stimuli

	Demographic info.							Stimulus											
Participant no.	C1-C2 are	Region	Abroad?	Age	Gender	Education level	Years Viet. edu.	nghi-C1	nghi-C2	dai-C1	dai-C2	mu-C1	mu-C2	se-C1	se-C2	chưa-C1	chưa-C2	rưa-C1	rưa-C2
2	M.	S.	U.S.	18-24	m	Uni.	1-4	C2	C2	C1	B1	B2	C2	C2	C2	C2	C2	C1	C1
23	C.		Can.	≥ 55	m	Grad.	≥ 10	C2	C2	C1	C1	C2	C2	C1	C2	C1	C2	C1	C1
24	C.	S.	U.S.	25-39	f	Uni.	≥ 10	C2	C1	B2	C1	C1	C2	C2	C1	C2	C2	C2	C2
25	M.	S.		25-39	m	Grad.	≥ 10	C1	C2	C2	C1	C2	C1	C1	C2	C2	C1	C1	C2
33	M.	S.		25-39	f	Grad.	≥ 10	C2	C2	C1	B1	C2	C2	C1	C1	C1	C2	C2	C2
49	C.	C.		18-24	f	H.S.	≥ 10	C2	C1	C1	C1	C2	C2	C2	C2	C1	C2	C2	C1
50	C.	S.		18-24	m	H.S.	≥ 10	C1	C2	C1	C2	C1	C2	C1	C2	C1	C2	C1	C2
61	C.	S.		18-24	f	H.S.	≥ 10	C1	C2	C2	C2	C1	C2	C1	C2	C2	C2	C1	C2
62	M.	S.		18-24	m	Uni.	≥ 10	C2	C2	C1	C2	C2	C2	C2	C2	C1	C2	C1	C2
63	M.	S.		18-24	f	Uni.	≥ 10	C2	C2	C2	C2	C1	C2	C1	C2	C2	C2	C1	C2
71	M.	S.		25-39	f	Uni.	≥ 10	C2	C2	C2	C2	C2	C2	C2	C2	C2	C2	C1	C1
72	M.	N.	U.S.	18-24	m	Uni.	≥ 10	C1	C1	C1	B1	C1	C2	C1	C2	C2	C2	C1	C1
74	C.	C.		25-39	f	Grad.	≥ 10	C2	C2	C1	C2	C2	C2	C1	C2	C1	C2	C2	C2
75	M.	S.		40-54	m	Grad.	≥ 10	C1	C2	C1	B1	C1	C2	C1	C2	C2	C1	C1	C2
78	C.	S.		25-39	f	Uni.	≥ 10	C1	C2	C1	C1	B2	C2	C1	C2	C1	C1	C1	C2
80		S.		25-39	m	Grad.	≥ 10	C1	C2	C1	C2	C2	C2	C1	C2	C1	C2	C1	C1
84	C.	N.		18-24	x	Uni.	≥ 10	C1	C2	C2	C2	C2	C2	C1	C2	C1	C1	C1	C2
90	C.	N.		18-24	m	H.S.	≥ 10	C2	C2	C1	C2	C2	C2	C1	C2	C1	C2	C2	C2
91	C.	N.		18-24	f	Uni.	≥ 10	C1	C2	C1	C2	C1	C2	C1	C2	C1	C2	C1	B2
104	C.	N.		25-39	f	Uni.	≥ 10	C1	C1	C1	C1	C2	C2	C2	C2	C2	C2	C1	C1
105	C.	N.		25-39	f	Uni.	≥ 10	C1	C2	C1	C2	C1	C2	C1	C2	C1	C2	C1	C2
108	M.	S.		40-54	f	Grad.	≥ 10	C2	C2	C2	C1	C1	C2	C1	C2	C2	C2	C2	C2
110	C.	S.		18-24	m	Uni.	≥ 10	C1	C2	A2	C1	C2	C2	C1	C2	C2	C1	C1	C1
113	M.	S.		25-39	f	Grad.	≥ 10	C1	C2	C1	C1	C1	C2	C2	C2	C1	C2	C2	C2
115		N.		40-54	f	H.S.	≥ 10	C1	C2	C1	C2	C2	C2	C1	C2	C1	C2	C1	C2
121	C.	S.		25-39	f	Grad.	≥ 10	C1	C2	C1	C2	C1	C2	C1	C1	C1	C1	C1	C1
133	M.	S.		25-39	f	Uni.	≥ 10	C1	C2	C1	C2	C2	C2	C1	C2	C1	C2	C1	C1
142	C.	S.	Aus.	25-39	f	Grad.	≥ 10	C1	C2	C1	C2	C1	C2	C1	C2	C1	C2	C1	C2
144	M.	S.		18-24	f	H.S.	≥ 10	C2	C2	C1	C1	C2	C2	C2	C1	C2	C2	C1	C1
147	M.	S.		25-39	f	Uni.	≥ 10	C2	C2	C1	C2	C2	C2	C1	C2	C1	C2	C1	C1
156	M.	S.		25-39	f	Uni.	≥ 10	C2	C2	C2	C1	C1	C2	C1	C2	C2	C1	C1	C2
166	C.	N.	Can.	≥ 55	m	Grad.	≥ 10	C1	C2	C1	C2	C2	C2	C1	C2	C1	C2	C1	C2
185	M.	S.		18-24	m	H.S.	≥ 10	C2	C1	C1	C2	B2	C2	C1	C1	C2	C2	C2	C1
198	C.	S.		40-54	f	Grad.	≥ 10	C1	C2	C1	C1	C2	C2	C2	C2	A1	C2	C1	C1
199	C.	N.		25-39	f	Uni.	≥ 10	C1	C2	C1	C2	C2	C2	C1	C2	C1	C2	C1	C2
200	C.	S.		25-39	f	Grad.	≥ 10	C2	C1	C1	C2	C2	C2	C1	C2	C2	C2	C1	C2
203	C.	S.		18-24	f	H.S.	≥ 10	C2	C1	A2	C2	C1	C2	C1	C2	C1	C2	C1	C2
204	M.	S.		18-24	f	H.S.	≥ 10	C1	C2	C1	C2	C1	C2	C1	C2	C1	C2	C1	C2
213	M.	S.		25-39	f	Uni.	≥ 10	C1	C1	C2	C1	C2	C2	C1	C1	C2	C2	C1	C1
214	C.	N.		25-39	f	Uni.	≥ 10	C1	C2	C1	C2	C1	C2	C1	C2	C1	C2	C1	C2
215	C.	N.		25-39	f	Uni.	≥ 10	C2	C2	C1	C2	C1	C2	C1	C2	C1	C2	C1	C1
216	C.	N.		18-24	f	H.S.	≥ 10	C1	C2	C1	C2	C1	C2	C1	C2	C1	C1	C1	C2
219	M.	S.		25-39	m	Uni.	≥ 10	C1	C2	C1	B1	C2	C2	C1	C2	C1	C2	C1	C2
220	M.	S.		25-39	f	Uni.	≥ 10	C2	C2	C2	C2	C2	C2	C1	C2	C1	C2	C1	C1
221	M.	S.		18-24	m	<H.S.	≥ 10	C1	C1	C2	C2	C2	C2	C2	C2	C2	C2	C2	C2
227	M.	S.		25-39	f	Uni.	≥ 10	C1	C2	C1	C2	C2	C2	C2	C2	C1	C2	C1	C1
236	C.	N.		25-39	f	Uni.	≥ 10	C1	C2	C1	C2	C1	C2	C1	C2	C1	C2	C1	C2
238	M.	S.	U.S.	25-39	f	Grad.	≥ 10	C1	C2	C1	C2	C2	C2	C1	C2	C2	C2	C2	C1
240	M.	S.		25-39	f	Grad.	≥ 10	C2	C2	C2	C1	C2	C2	C2	C2	C2	C2	C1	C1
241	M.	S.		18-24	f	<H.S.	≥ 10	C1	C2	C1	C2	C2	C2	C1	C2	C1	C1	C1	C1
242	C.	C.		25-39	f	Uni.	≥ 10	C2	C2	C1	B1	C1	C2	C1	C2	C1	C2	C1	C2
248	M.	S.		25-39	f	H.S.	≥ 10	C1	C2	C1	C2	C2	C2	C1	C2	C2	C1	C1	C1
249	M.	S.	U.S.	25-39	f	Grad.	≥ 10	C1	C2	C1	C1	C2	C2	C1	C2	C2	C1	C1	C2
250	M.	S.		18-24	m	H.S.	≥ 10	C1	C2	C1	C2	C2	C2	C1	C2	C1	C2	C1	C2
258	C.	N.		25-39	f	Uni.	≥ 10	C1	C2	C1	C2	C1	C2	C1	C2	C1	C2	C1	C2
260	M.	C.	U.S.	25-39	f	Uni.	≥ 10	C1	C2	C1	C2	C1	C2	C1	C2	C1	C2	C1	C2
262	C.	S.		25-39	m	Grad.	≥ 10	C1	C2	C1	C2	C1	C2	C2	C2	C1	C2	C1	C2
265	M.	C.		25-39	f	Grad.	≥ 10	C1	C2	C1	C2	C1	C2	C1	C2	C1	C2	C1	C2
268	C.		U.S.	25-39	f	Uni.	≥ 10	C1	C2	C1	C2	C2	C2	C1	C2	C1	C2	C2	C2
269		N.		25-39	f	Grad.	≥ 10	C1	C2	C1	C2	C1	C2	C1	C2	C2	C2	C1	C2
274	M.	S.		18-24	m	H.S.	≥ 10	C1	C2	C1	C2	C1	C2	C1	C2	C1	C1	C2	C1
275	C.	S.	S.K.	25-39	f	H.S.	≥ 10	C2	C2	C1	C2	C2	C2	C1	C2	C1	C2	C1	C2
277	M.	S.		18-24	m	H.S.	≥ 10	C1	C2	C1	C1	C2	C2	C2	C2	C1	C1	C1	C1
278	M.	S.		25-39	f	Grad.	≥ 10	C1	C2	C1	C1	C2	C2	C1	C2	C2	C1	C1	C2
279	C.	N.		25-39	f	Grad.	≥ 10	C1	C2	C1	C2	C1	C2	C1	C2	C1	C2	C1	C2
285	C.	N.	U.S.	25-39	f	Grad.	≥ 10	C1	C2	C1	C2	C1	C2	C1	C2	C1	C2	C1	C2
292	M.	S.		25-39	m	Uni.	≥ 10	C1	C2	C1	C2	C1	C2	C1	C1	C2	C2	C1	C1
296	M.	C.		25-39	f	Uni.	≥ 10	C1	C2	C1	C2	C1	C2	C1	C2	C2	C2	C1	C1
308	C.	N.		18-24	f	Uni.	≥ 10	C1	C2	C1	C2	C1	C2	C1	C2	C1	C2	C1	C2
309	M.	C.		18-24	f	H.S.	≥ 10	C1	C2	C1	C2	C2	C2	C1	C1	C2	C2	C1	C1
311	C.	S.		25-39	f	Grad.	≥ 10	C2	C2	C1	C1	C2	C2	C2	C2	C1	C2	C2	C2
312	C.	N.	U.S.	18-24	f	Uni.	≥ 10	C1	C2	C1	C2	C1	C2	C1	C2	C1	C2	A2	C2
317	M.	S.		25-39	m	Uni.	≥ 10	C2	C2	C1	C2	B2	C2	C1	C2	C1	C1	C1	C2
318	C.	N.		25-39	m	Uni.	≥ 10	C1	C1	C1	B1	C2	C2	C2	C2	C1	C1	C1	C1

Appendix G: Perception task full results: filler stimuli

	Demographic info.							Stimulus											
Participant no.	C1-C2 are	Region	Abroad?	Age	Gender	Education level	Years Viet. edu.	chuông-A1	chuông-A2	giam-A1	giam-C1	lê-A1	lê-C2	công-A2	công-C1	la-B2	la-C1	diên-B2	diên-C2
2	M.	S.	U.S.	18-24	m	Uni.	1-4	A1	A1	A1	B2	C2	C2	B2	B2	B2	B2	C1	C2
23	C.		Can.	≥ 55	m	Grad.	≥ 10	A2	A2	A1	C1	C2	C2	C1	C1	B2	B2	B2	B2
24	C.	S.	U.S.	25-39	f	Uni.	≥ 10	A2	A1	C1	C1	C2	B2	C1	C1	B2	B2	B2	B2
25	M.	S.		25-39	m	Grad.	≥ 10	A2	A2	C1	C1	C2	C2	C1	C1	B2	B2	B2	B2
33	M.	S.		25-39	f	Grad.	≥ 10	A1	A2	A1	C1	A1	C2	A2	C1	B2	B2	B2	B2
49	C.	C.		18-24	f	H.S.	≥ 10	A1	A2	C1	C1	A1	C2	C1	C1	B2	B2	B2	B2
50	C.	S.		18-24	m	H.S.	≥ 10	A1	A2	A1	C1	A1	C2	A2	C1	B2	B2	B2	B2
61	C.	S.		18-24	f	H.S.	≥ 10	A1	A2	A1	C1	A1	C2	C1	C1	B2	C1	B2	C2
62	M.	S.		18-24	m	Uni.	≥ 10	A1	A2	A1	C1	A1	C2	A2	C2	B2	C2	B2	C2
63	M.	S.		18-24	f	Uni.	≥ 10	A1	A2	A1	C1	A1	C2	A2	C1	B2	B2	B2	C2
71	M.	S.		25-39	f	Uni.	≥ 10	A1	A1	C1	C1	C2	C2	B1	B1	B2	B2	B2	B2
72	M.	N.	U.S.	18-24	m	Uni.	≥ 10	A1	A2	A1	C1	C2	C2	C1	C1	B2	C1	B2	B2
74	C.	C.		25-39	f	Grad.	≥ 10	A1	A2	C1	C1	A1	C2	C1	C1	B2	C1	B2	C2
75	M.	S.		40-54	m	Grad.	≥ 10	A1	A2	A1	C1	A1	C2	A2	C1	B2	C1	B2	C2
78	C.	S.		25-39	f	Uni.	≥ 10	A1	A2	A1	C1	A1	C2	A2	C1	B2	B2	B2	C2
80		S.		25-39	m	Grad.	≥ 10	A1	A2	A1	C1	A1	C2	A2	C1	B2	B2	C2	C2
84	C.	N.		18-24	x	Uni.	≥ 10	A1	A2	A1	C1	A1	B2	C1	C1	B2	B2	B2	C2
90	C.	N.		18-24	m	H.S.	≥ 10	A1	A2	A1	C1	A1	C2	A2	C1	B2	B2	C2	C2
91	C.	N.		18-24	f	Uni.	≥ 10	A1	A2	A1	C1	C2	C2	C1	C1	B2	B2	B2	C2
104	C.	N.		25-39	f	Uni.	≥ 10	A2	A2	C1	C1	C2	C2	C1	C1	B2	B2	B2	B2
105	C.	N.		25-39	f	Uni.	≥ 10	A1	A2	A1	C1	A1	C2	A2	C1	B2	C1	B2	C2
108	M.	S.		40-54	f	Grad.	≥ 10	A1	A2	A1	C1	A1	C2	C1	C1	B2	B2	B2	C2
110	C.	S.		18-24	m	Uni.	≥ 10	A1	A2	A1	C1	A1	C2	A2	C1	B2	C1	B2	C2
113	M.	S.		25-39	f	Grad.	≥ 10	A1	A2	A1	C1	C2	C2	C1	C1	B2	B2	B2	B2
115		N.		40-54	f	H.S.	≥ 10	A1	A2	A1	C1	A1	C2	C1	C1	B2	B2	C2	C2
121	C.	S.		25-39	f	Grad.	≥ 10	A1	A1	A1	C1	C2	C2	A2	C1	B2	C1	B2	B2
133	M.	S.		25-39	f	Uni.	≥ 10	A1	A2	A1	C1	A1	C2	A2	C1	B2	B2	B2	B2
142	C.	S.	Aus.	25-39	f	Grad.	≥ 10	A1	A2	A1	C1	A1	C2	A2	C1	B2	C1	B2	C2
144	M.	S.		18-24	f	H.S.	≥ 10	A1	A1	C1	C1	A1	A1	C1	C1	B2	B2	B2	B2
147	M.	S.		25-39	f	Uni.	≥ 10	A1	A2	B2	C1	A1	C2	C1	C1	B2	C1	B2	C2
156	M.	S.		25-39	f	Uni.	≥ 10	A1	A2	A1	C1	A1	C2	A2	C1	B2	C2	B2	C2
166	C.	N.	Can.	≥ 55	m	Grad.	≥ 10	A1	A2	A1	C1	A1	C2	C1	C1	B2	B2	B2	C2
185	M.	S.		18-24	m	H.S.	≥ 10	A1	A2	A1	C1	A1	C2	C1	B2	B2	B2	B2	C2
198	C.	S.		40-54	f	Grad.	≥ 10	A2	A2	A1	C1	C2	C2	C1	C1	B2	B2	B2	B2
199	C.	N.		25-39	f	Uni.	≥ 10	A1	A2	A1	C1	C2	C2	A2	C1	C1	C1	B2	C2
200	C.	S.		25-39	f	Grad.	≥ 10	A1	A2	B2	C1	C2	C2	C1	C1	B2	B2	B2	B2
203	C.	S.		18-24	f	H.S.	≥ 10	A1	A2	A1	C1	A1	C2	A2	C1	B2	B2	B2	C2
204	M.	S.		18-24	f	H.S.	≥ 10	A1	A2	A1	C1	A1	C2	A2	C1	B2	C1	B2	C2
213	M.	S.		25-39	f	Uni.	≥ 10	A1	A2	A1	C1	C2	C2	C1	C1	B2	B2	B2	C2
214	C.	N.		25-39	f	Uni.	≥ 10	A1	A2	A1	C1	A1	C2	A2	C1	B2	C1	B2	C2
215	C.	N.		25-39	f	Uni.	≥ 10	A1	A2	A1	C1	A1	C2	A2	C1	B2	B2	B2	C2
216	C.	N.		18-24	f	H.S.	≥ 10	A1	A2	A1	C1	A1	C2	C1	C1	B2	C1	B2	C2
219	M.	S.		25-39	m	Uni.	≥ 10	A1	A2	A1	C1	A1	C2	C1	C1	B2	C1	B2	C2
220	M.	S.		25-39	f	Uni.	≥ 10	A1	A2	A1	C1	A1	C2	A2	C1	B2	B2	B2	B2
221	M.	S.		18-24	m	<H.S.	≥ 10	A2	A2	C1	C1	C2	C2	C1	C1	B2	B2	B2	B2
227	M.	S.		25-39	f	Uni.	≥ 10	A1	A2	A1	C1	C2	C2	A2	C1	B2	B2	B2	C2
236	C.	N.		25-39	f	Uni.	≥ 10	A1	A2	A1	C1	A1	C2	A2	C1	B2	B2	B2	C2
238	M.	S.	U.S.	25-39	f	Grad.	≥ 10	A2	A2	C1	C1	A1	C2	A2	C1	B2	B2	B2	B2
240	M.	S.		25-39	f	Grad.	≥ 10	A2	A1	C1	C1	C2	C2	A2	A2	B2	C1	B2	B2
241	M.	S.		18-24	f	<H.S.	≥ 10	A1	A2	A1	C1	A1	C2	A2	C1	B2	B2	B2	C2
242	C.	C.		25-39	f	Uni.	≥ 10	A1	A2	A1	C1	C2	C2	C1	C1	B2	C1	B2	B2
248	M.	S.		25-39	f	H.S.	≥ 10	A1	A2	A1	C1	A1	C2	C1	C1	B2	C2	B2	C2
249	M.	S.	U.S.	25-39	f	Grad.	≥ 10	A1	A2	A1	C1	A1	C2	A2	C1	B2	C1	B2	C2
250	M.	S.		18-24	m	H.S.	≥ 10	A1	A2	A1	C1	A1	C2	A2	C1	C1	B2	B2	C2
258	C.	N.		25-39	f	Uni.	≥ 10	A1	A2	A1	C1	A1	C2	C1	C1	B2	C1	B2	C2
260	M.	C.	U.S.	25-39	f	Uni.	≥ 10	A1	A2	A1	C1	A1	C2	A2	C1	B2	C1	B2	C2
262	C.	S.		25-39	m	Grad.	≥ 10	A1	A2	A1	C1	A1	C2	C1	C1	B2	B2	B2	C2
265	M.	C.		25-39	f	Grad.	≥ 10	A1	A2	A1	C1	A1	C2	C1	C1	B2	C1	B2	C2
268	C.		U.S.	25-39	f	Uni.	≥ 10	A1	A2	A1	C1	A1	C2	A2	C1	B2	B2	B2	C2
269		N.		25-39	f	Grad.	≥ 10	A1	A2	A1	C1	A1	C2	A2	C1	B2	C1	B2	C2
274	M.	S.		18-24	m	H.S.	≥ 10	A1	A2	A1	C1	A1	C2	C1	C1	B2	B2	B2	C2
275	C.	S.	S.K.	25-39	f	H.S.	≥ 10	A1	A2	A1	C1	A1	C2	C1	C1	B2	B2	B2	B2
277	M.	S.		18-24	m	H.S.	≥ 10	A1	A1	C1	C1	C2	C2	C1	C1	B2	B2	B2	B2
278	M.	S.		25-39	f	Grad.	≥ 10	A1	A2	A1	C1	A1	C2	C1	C1	B2	B2	B2	B2
279	C.	N.		25-39	f	Grad.	≥ 10	A1	A2	A1	C1	A1	C2	A2	C1	B2	C1	B2	C2
285	C.	N.	U.S.	25-39	f	Grad.	≥ 10	A1	A2	C1	C1	A1	C2	A2	C1	B2	B2	B2	B2
292	M.	S.		25-39	m	Uni.	≥ 10	A1	A2	A1	C1	A1	C2	C1	C1	B2	B2	B2	C2
296	M.	C.		25-39	f	Uni.	≥ 10	A1	A2	A1	C1	A1	C2	A2	C1	B2	C1	B2	C2
308	C.	N.		18-24	f	Uni.	≥ 10	A1	A2	A1	C1	A1	C2	B1	C1	B2	C1	B2	C2
309	M.	C.		18-24	f	H.S.	≥ 10	A2	A2	A1	C1	A1	C2	C1	C1	B2	B2	B2	B2
311	C.	S.		25-39	f	Grad.	≥ 10	A1	A1	C1	C1	C2	C2	C1	C1	B2	B2	B2	B2
312	C.	N.	U.S.	18-24	f	Uni.	≥ 10	A1	A2	A1	C1	A1	C2	A2	C1	B2	C1	B2	C2
317	M.	S.		25-39	m	Uni.	≥ 10	A1	A2	A1	C1	A1	C2	A2	C1	B2	B2	B2	B2
318	C.	N.		25-39	m	Uni.	≥ 10	A2	A2	A1	C1	C2	C2	C1	C1	B2	B2	B2	B2