Univerzita Karlova vPraze
Filozofická fakulta

Fonetický ústav

Disertační práce

Mgr. Ondřej Slówik

Rozdíly realizace tónů hanojského a saigonského dialektu vietnamštiny mezi čteným a polospontánním mluveným projevem

Tone realization differences in Hanoian and Saigonese dialects between reading and semi-spontaneous speech

V Praze, 2018 vedoucí práce: doc. PhDr. Jan Volín, Ph.D.

Prohlašuji, že jsem disertační práci vypracoval samostatně a že jsem uvedl všechny použité prameny a literaturu.
Souhlasím se zapůjčením disertační práce ke studijním účelům.

Prague, 2018

I declare that the following PhD thesis is my own work for which I used only the sources and literature mentioned.
I have no objections to the PhD thesis being borrowed and used for study purposes.

Acknowledgements

Firstly, I would like to express my sincere gratitude to my supervisor, doc. PhDr. Jan Volín Ph.D., for the continuous support of my PhD study and related research. Even one rather famous Vietnamese fortuneteller confirmed that without his patience, motivation, and knowledge, this thesis could not have been written.

Besides my supervisor, I would like to thank Tomáš Bořil and Pavel Šturm from the Institute of Phonetics in Prague for their insightful assistance in resolving technical difficulties concerning data processing. I also thank the late Ivo Vasiljev who provided me with many useful comments with regard to the Vietnamese language, and whose multilingual erudition was the chief driving force of my linguistic development. Unfortunately, his newly reincarnated self is not old enough yet to assess the quality of my recent research.

Last but not the least, my utmost gratitude goes to my family: my parents for supporting me throughout all the stages of the education system, my wife for generally putting up with me especially during the finishing moments before the submission and my daughter for sleeping after lunch with obedient regularity.

Abstract

The chief objective of this dissertation is the description of tone realization differences in Hanoian and Saigonese dialects based on a representative sample of recorded material, with special focus on read monologue and semi-spontaneous conversational speech. The research discusses mainly issues of tone production but it is complemented by a section on tone perception in form of a perception test.

The theoretical background in Section 2.1. describes the topic of tonality and tonal languages in general. Section 2.2. is devoted to the description of the Vietnamese language and attention is specifically paid to tonal inventories of both researched dialects. Tonogenesis is mentioned on a general level as well as in the Vietnamese language in particular.

Chapter 3 introduces the research methodology, namely the speaker selection, speech material preparation and recording, data extraction and preparation for the analyses and the perception test.

Chapter 4 is divided into three sections. Section 4.1. speaks about tone realizations in isolation and carefully preselected context. Its goal is to investigate the behaviour of tonal contours influenced by as few variables as possible. The results should be comparable to the findings of previously conducted studies. Section 4.2. strives to assess data from a quantitative point of view and yield data closely reflecting linguistic reality. Section 4.3. constitutes a perception test the purpose of which is to investigate the ability of tone distinction across the dialects when the tones are stripped of all contextual cues.

The results of the analyses indicate substantial differences in tonal contours between the Hanoian and Saigonese dialects. They also suggest that F_0 is the sole discrimination cue neither for Hanoian nor Saigonese tones. Voice quality, duration and, based on the findings of 4.2., especially context are phenomena influencing tone discrimination to a great extent. In connected speech possibly even more than the contour of F_0. The perception test revealed a surprisingly low success rate of tone discrimination for both dialects.

Keywords: Vietnamese, tones, intonation, phonetics, tone production, tone perception

Abstrakt

Hlavním cílem této disertace je popis rozdílů v realizacích tónů mezi hanojským a saigonským dialektem se zaměřením na situaci ve čteném a polospontánním mluveném projevu. Výzkum se zabývá zejména produkcí, avšak je doplněn i oddílem, který řeší problematiku percepce tónů s využitím percepčního testu.

Oddíl 2.1. popisuje tonalitu a tónové jazyky v obecném smyslu. V oddíle 2.2. se popisuje vietnamský jazyk s důrazem na tónové inventáře obou zkoumaných dialektů. Zmiňuje se též vznik a vývoj tónů obecně i v rámci vietnamštiny.

Třetí kapitola představuje metodu výzkumu, zejména výběr mluvčích, přípravu materiálu a nahrávek, extrakce dat a přípravu analýz a percepčního testu.

Čtvrtá kapitola je rozdělena do tří oddílů. Oddíl 4.1. hovoří o realizacích tónů v izolaci a ve speciálně zvoleném kontextu. Jeho účelem je popsat chování tónů co nejméně ovlivněných okolními proměnnými. Výsledky této analýzy by měly být srovnatelné s již publikovanými studiemi na stejné téma. V oddíle 4.2. je na data nahlédnuto z kvantitativní perspektivy a výsledky této analýzy by měly věrněji reflektovat jazykovou realitu. Oddíl 4.3. představuje percepční test, jehož účelem je zmapovat schopnost rozlišování tónů mimo přirozený kontext.

Výsledky analýz indikují zásadní rozdíly mezi konturami tónů hanojského a saigonského dialektu. Ukazuje se, že základní frekvence F_0 není jediným vodítkem při rozlišování ani hanojských, ani saigonských tónů. Fonační modifikace, trvání a zejména kontext taktéž přispívají k úspěšnému rozlišování jednotlivých tónů. Percepční test ukázal překvapivě nízkou úspěšnost při určování tónů jak u respondentů z Hanoje, tak z Ho Či Minova Města.

Klíčová slova: vietnamština, tóny, intonace, fonetika, produkce tónů, percepce tónů

Contents

Abbreviations

VN	Vietnam
HN	Hanoi
SG	Saigon
HCMC	Ho Chi Minh City
DRV	Democratic Republic of Vietnam
F_0	fundamental frequency
T	tone
C	consonant
V	vowel
Vc	vocalic core
TBU	tone bearing unit
H	high
M	medium
L	low
DL	difference limen
ms	millisecond
Hz	Hertz
dB	decibel
TG	text grid

PT pitch tier

1. Introduction

It can be argued that, given the geographical location of the Czech Republic, tonal languages represent a concept from a very distant part of the world with scarce diplomatic and economic ties. Such assumption is, in fact, rather removed from reality. Although the population of speakers of best known and most documented tonal languages such as Mandarin, Cantonese or Thai is quite small in the Czech Republic compared to other European countries, there is one ethnicity of tonal language speakers living in there in abundance. The official number of Vietnamese people living in the Czech Republic was 59 534[1] individuals as of September, 2017, which means that they constitute the third most numerous ethnic minority after the Slovaks and the Ukrainians, and the most numerous non-European minority in the Czech Republic. They were granted the status of a national minority in 2013, which entitles them to requesting funds at the Czech Ministry of Culture for propagation of Vietnamese culture, free counseling services in Vietnamese, and their children are entitled to elective courses of Vietnamese language at elementary schools. Research on the Vietnamese language and its phonetic features in particular is crucial for better understanding of how the Vietnamese acquire the Czech language be it as a second language learnt in adulthood or as a mother tongue in case of the rising numbers of bilingual Vietnamese children growing up in the Czech cultural environment.

However, it must be noted that lexical tone and tonal languages in general constitute phenomena that do not stand in the center of attention of the Czech academic circles. This is

[1] Data retrieved from the database of the Czech Statistical Office (www.czso.cz).

true particularly for the Vietnamese language as the only monograph addressing this issue is *Vietnamese Phonetics* authored by Slavická (2008). Its purpose is pedagogical rather than academic and, although it is a very valuable source and a useful tool in the process of acquisition of Vietnamese as a second language, it does not address the topic of tonality in much detail.

On the other hand, the Czechoslovak interest in Asian studies dates back to the pre-WWII era when Jaroslav Průšek was granted a research scholarship at the university in Beijing in 1932. Průšek subsequently became the head of the newly established department of Chinese studies at Charles University in Prague in 1945. The department of Vietnamese studies was established 1960 due to the political pressure caused by the effort to create close diplomatic and economic relations with DRV (Democratic Republic of Vietnam). Throughout the history of the two departments, many academic works have been written but very few have dealt with linguistic issues. This tendency is likely to be caused by the study motivation and scope of interest of the researchers. Most people decide to engage in Asian studies because of their interest in Asian history, literature or politics. These topics can be satisfactorily accessed and researched even with just basic command of the local language, whereas when the language itself serves as the research subject, a higher degree of proficiency is required.

The aspects of Vietnamese phonetics in general as well as Vietnamese tones in particular have been given more attention in the international and academic community and in Vietnam itelf. There are, however, certain methodological issues that need to be addressed and that triggered the origin of this thesis.

The problem of the canonical works on Vietnamese phonetics (Thompson, 1965; Đoàn, 1977; Vũ, 1982; Gordina & Bystrov 1984) is their obsoleteness. Up until 1990s worldwide

and well into 2000s in Vietnam, the use of computers and digital recording devices in linguistics was scarce and the data gathered by the researchers then is difficult to compare to the data we are gathering nowadays. This situation was caused simply by the state of technological development throughout the time. For this reason, researchers were forced to rely prevalently on their hearing skills and intuition, which are tools still used in our contemporary linguistic research but no longer considered particularly reliable.

Only with the works of Nguyễn and Edmondson (1997) and later Phạm (2003) and Brunelle (2003) we can begin to talk about the rise of modern phonetic analysis using modern recording devices, recording methodology and more dependable data analysis by means of computer software. Certain objections can, nonetheless, be made towards the nature of their data. They tend to operate with small sets of speech material gathered from a limited number of speakers (usually no more than four). The recordings prevalently consist of words uttered in isolation or within artificial sentences (e.g. *Say the word "X" once*). When the authors use isolated sentences that are constructed to conform to the research purpose, they hardly feel natural in terms of collocation and context. Hence we could pose the question whether such material corresponds with real-life natural speech and whether the conclusions drawn from the gathered data represent reality accurately.

With the exception of Brunelle, who has authored several articles dealing with cross-dialect issues (comparison of Northern and Southern dialects), other prominent researchers of Vietnamese tonality have dealt predominantly with the Northern dialect and other dialects escaped their academic attention. The reason for this tendency lies in the fact that the Northern (Hanoian) is considered standard and it is also documented in greatest detail on other levels not just in terms of tonality. Moreover, two most developed departments of linguistics in Vietnam belonging to the National University are located in Hanoi and in the

Ho Chi Minh City so doing research in other regions than these two is more complicated in terms of logistics, recording conditions and data storage. It is necessary to point out that there is also a significant variation in sociolinguistic consistency of the speakers. In the past 60 years, most of Vietnam has experienced turbulent population shifts. It is not difficult to find speakers born and raised in Hanoi, whose parents were also born there or at least spent there their whole life after leaving their homeland in the nearby countryside. However, in the Ho Chi Minh City and other main cities in Southern and Central Vietnam, many people are descendants of individuals coming from the North after 1954 or 1976, which is why their accent cannot be regarded as authentically local. Consequently, selecting suitable recording subjects needs more effort and caution than in Hanoi. Researchers must be capable of accurate accent assessment otherwise the data might be compromised and might not describe the linguistic reality.

For the reasons listed above, the two dialects selected for the analysis in this thesis are the dialect of Hanoi, as it is considered standard and its speakers are considerably easy to assess, and the Ho Chi Minh City dialect where more attention had to be paid to the speaker selection in terms of family origin but it is still the dialect of the largest city in Vietnam as well as the Vietnamese economic hub. As opposed to the dialect of Hanoi that has been described in numerous academic and pedagogic works in great detail, a canonical version of the HCMC (Saigonese[2]) dialect has not been clearly defined, which is another aspect making it necessary to approach speaker selection with utmost caution. In order to achieve highest clarity of the recorded material, it would have been optimal to compare the speech of inhabitants of two villages removed from civilization with lack of population shift. Unfortunately, it would be very bold to draw any general conclusion from the comparison of

[2] The difference between the terms “Ho Chi Minh City” and “Saigon” is further explained in section 2.2.3. As this thesis deals with linguistics and not politics, the terms can be used interchangeably especially because the adjective “Saigonese” already exists and it suitably complements the adjective “Hanoian”.

two isolated niches with limited population. Hanoi and the Ho Chi Minh City, on the other hand, represent economic and cultural centers of their regions with population amounting to millions of individuals.

Building up on the facts mentioned above, the aim of this thesis is to present the most detailed description of Hanoian and Saigonese tones in the Czech academic context. Furthermore, the thesis is meant to attempt to improve some of the methodological strategies in researching Vietnamese tonality. The number of recorded subjects is set to 12 individuals for each dialect, which is above the average in this field of research. Furthermore, the analyzed material consists of not only isolated syllables but also reading and semi-spontaneous speech. For more details see section 3.2.

The beginning of the thesis, namely section 2.1. strives to introduce the topic of intonation (Cruttenden 1997; Gusenhoven 2004) with focus on lexical tone in various languages across the globe (Yip 2002). The next section, 2.2., is devoted to the description of the Vietnamese language with emphasis on tonality, tonal development and dialects. More space is reserved for detailed description of the dialects spoken in Hanoi and Saigon, their segmental as well as suprasegmental structure. Chapter 3. introduces the method of speaker selection, material recording and material processing, which are key factors preparing the ground for a reliable analysis.

The empirical research part of the thesis consists of three separate tasks. The first task analyses at the tones in isolation and in high/low left context in terms of pitch contours and phonation types. Its aim is to assess to what degree pitch contours of the individual tones differ across the two dialects and to what extent phonation types might be a decisive factor for tonal identification. The second part of the first task is aimed at determining the truth

value of the claims made by Brunelle (2009) attesting to the presence of progressive coarticulation among the Vietnamese tones despite not playing any phonological role. (see 2.2.2.6. and 2.2.3.4.)

The second task encompasses a quantitative analysis of a text reading and a semi-spontaneous speech. In order to proceed with this task, it was necessary to devise a method to measure pitch values of all the analyzed syllables automatically. After applying the method to both the read-out text and the semi-spontaneous speech, it is possible to compare the two with each other as well as across the dialects.

The third task uses a perception test to determine the ability of Hanoian and Saigonese speakers to discriminate tones across dialects and with respect to stress patterns. Respondents selected in Hanoi as well as Saigon are asked to listen to Hanoian and Saigonese syllables uttered in three levels of sentential stress and identify them. The results should provide deeper understanding of the issue of tone perception in Vietnamese.

2. Theoretical background

The purpose of this chapter is to introduce key ideas and theories connected to prosody of tonal languages in general with particular focus on tonal production, perception and tonogenesis in section 2.1. Section 2.2. is devoted to characterization of the Vietnamese language with emphasis on dialects and tonal system.

2.1. Tone language

According to Yip (2002: 17), up to 60-70 per cent of the world's languages are tonal. A language is classified as a 'tone language' if the pitch of the word can change the meaning of the word. In standard Vietnamese, for example, the syllable [ma] can be uttered with six different pitch contours to convey six different meanings.

Tone	*Meanin*
ma	ghost
mà	but
mã	horse
mả	tomb
má	cheek
mạ	rice

Table 2.1. *Meanings of the Vietnamese syllable [ma] in all tonal variants*

However, languages like Vietnamese or Cantonese with broad tonal registers and tones firmly embedded into syllabic cores constitute a relatively small sub-group within the large pool of tonal languages despite being considered stereotypical examples of tonal languages. Most tonal languages have smaller tonal registers and contain substantial amounts of atonal

syllables. Pike (1948) introduced two terms: 'register tone languages' and 'contour tone languages', the first type describing languages containing only level tones and the second type was a label for languages containing contour tones.

Languages such as Cantonese or Vietnamese are often described as monosyllabic, which is partially caused by orthography. Cantonese uses idiographic characters where each character normally represents one syllable and although Vietnamese uses Latin alphabet, it is actually only a transcription of an idiographic writing system based on ancient Chinese script working on very similar grounds as pin-yin. In the modern Vietnamese script (chữ quốc ngữ - national language script), every syllable is written separately and contains a tonal diacritical sign, hence the monosyllabic classification. However, there are many syllables that cannot exist on their own and they must be combined with another syllable to carry full lexical meaning, e.g. *địa* (land, earth) must be combined with other syllables like *địa lý* (geography) in order to become grammatically acceptable; the syllable *hoá* stands for the element –ize or –ization, e.g. *Tây hoá* (westernize/-ization). Neither the syllable *địa* nor *hoá* would be understood as a full lexical unit by native speakers of Vietnamese. Moreover, there are lexemes consisting of two to three syllables carrying full lexical meaning in themselves but their combination creates a new meaning, e.g. *người* (human being), *bán* (sell), *hàng* (goods) are three independent lexical units that can be combined into one - *người bán hàng* meaning 'a vendor'. The abovementioned examples illustrate that classifying Vietnamese as a monosyllabic language does not seem to be entirely watertight. This is, nevertheless, not the concern of this dissertation.

Polysyllabic languages can either place the tone on more syllables in one word, Yip (2002: 2) mentions the example of *yùòrí* (penis) and *yúórì* (name) in Dagaare spoken in Ghana, or there can be only one tone-carrying syllable in the whole lexeme such as in the verb *ku-*

lombéz-a (to request) in the Chizigula language used in Tanzania. Moreover, the tone is not fixed to one syllable but can be shifted based on inflection *ku-lombez-éz-a* (to request for) or *ku-lombez-ez-án-a* (to request for each other). According to Yip (2002: 18), most polysyllabic tonal languages have more limited tonal registers and they usually contrast only two to three tones that tend to be level. As it was stated before, Vietnamese contains no atonal syllables so there are no instances similar to the Chizigula example. It might be argued that there are examples of lexemes similar to Dagaare such as *truyền thông* (media) and *truyền thống* (tradition) but that would depend on the whether we pronounced Vietnamese for a truly monosyllabic language or admitted the existence of certain polysyllabic features.

2.1.1. Tone production

In order to understand the production of lexical tones, we have clarify three important terms: fundamental frequency (F_0), pitch and tone. F_0 is an acoustic term referring to the number of pulses per second contained in the speech signal. Each pulse is produced by the single vibration of the vocal folds. The frequency of the pulses is measured in Hertz (Hz), one Hertz being one cycle per second. Gussenhoven (2004: 2) claimed that if the pulses occur more than 40 times per second, the human ear perceives it as a continuous event. This brings us to pitch, which is a perceptual term determining the way the listener hears the signal: the faster the vibration of vocal chords, the higher pitch is perceived. Finally, tone for the purposes of this work is a phonological category distinguishing meaning based on the change in pitch. However, Ohala (1978: 6) admitted the existence of "the distinctive use of other phonetic parameters besides pitch, for example, duration, voice quality, manner of tone offset, and vowel quality." As demonstrated in Phạm (2003), pitch is not the sole perception cue for differentiating tones in Vietnamese. In certain tones, phonation types

such as breathiness and creakiness of the voice can play a more significant role than pitch contour.

The organ responsible for production of F_0 perceived by our auditory organs as pitch is the larynx (see fig. 1). The larynx is composed of two cartilage rings, the cricoid and the thyroid. Two arytenoid cartilages are located on the top of the rear rim of the cricoid cartilage and connected to the thyroid cartilage by vocal folds. There is a space between the two vocal folds called the glotttal opening or glottis allowing the air to flow from the lungs to the oral cavity. When the vocal folds are brought together, the glottal opening narrows increasing the pressure of the passing air and resulting in closing the glottal opening due to Bernoulli's Law. Pressure build-up from lungs behind the closure releases a burst of air, which leads to the ensuing decrease in sub-glottal pressure and the cycle can start again. Takefuta et al. (1971) conducted a study on 24 male and 24 female speakers of American English and established the average F_0 to 127 Hz for men and 186 for women. Chen (1974) looked at Mandarin Chinese and concluded that average F_0 for Chinese males was 108 Hz and 184 for Chinese females. Johns-Lewis (1986) researched average F_0 of English speakers in conversation, reading and acting. In conversation, the value for female speakers was 182 Hz, which is a figure similar to Takefuta's findings. Johns-Lewis' results for male speakers in conversation were somewhat lower, 101 Hz, than Takefuta's. The average F0 was growing up to 142 Hz for men and 239 Hz for women when they were acting.

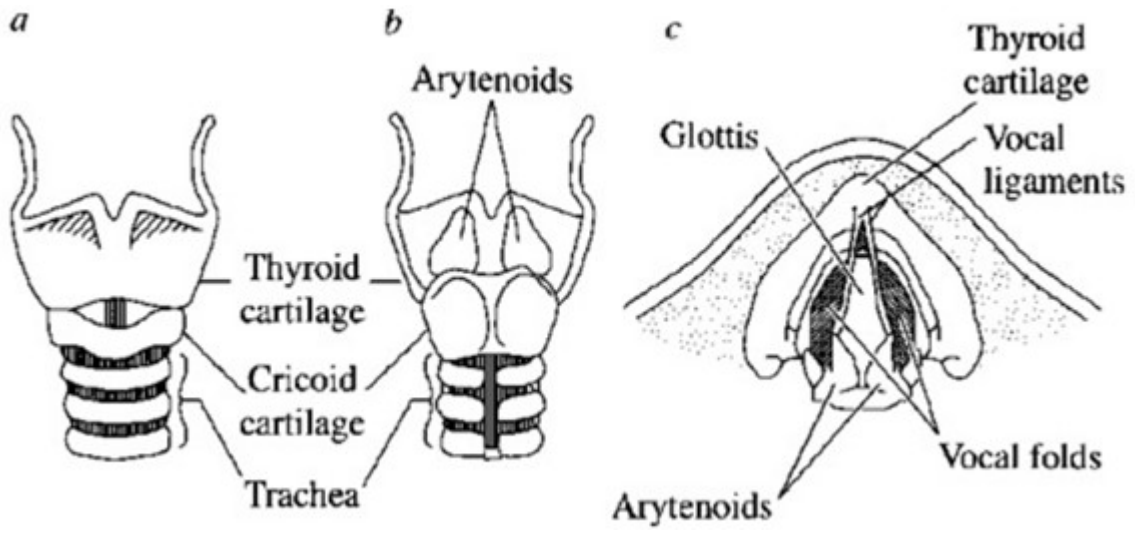


Figure 2.1. *Image representation of the larynx according to O'Grady et al. (1997). Image a) portrays the front, b) shows the back and c) is a horizontal cut viewed from above.*

Manipulation of F_0 based on the setup of the larynx results into changes in pitch perception that can gain various linguistic functions. In non-tonal languages, these functions are prevalently pragmatic whereas in tonal languages, they become lexical. Going back to Pike's (1948) dichotomy, production of register tones is easier than production of contour tones when perceived from the articulatory point of view. Sundberg (1973) claimed that rising tones are most difficult to produce as they require most energy for their realization. Tones with complex contour are also difficult because they require a very precise control of the articulatory tract.

2.1.2. Physiological aspects affecting pitch perception

There are two physiological aspects of speech production affecting pitch perception. Although their effects do not seem to influence tone perception in Vietnamese and they have

not become phonologized like in other tonal languages, it is still necessary to get acquainted with the two phenomena at this point.

The first is called declination and it has been observed in both tonal and non-tonal languages. “Most languages exhibit a gradual fall in pitch from the beginning to the end of an utterance. (…) This need not apply to questions, however. In many tone languages this results in successive tones becoming phonetically lower and lower in pitch until, at the end of the phrase, the high tones could be phonetically as low or even lower than the low tones at the beginning of the phrase. This is called ‘downdrift’ in African languages but it is evident in non-tone languages as well.” (Ohala 1978: 31)

Ohala then introduced three hypotheses for downdrift origin. The first, postulated by Maeda (1975) suggested progressive lowering of the larynx during individual breath groups due to the decrease of lung volume. As there is a correlation between larynx height and pitch, the movement of larynx should cause a gradual lowering of pitch. This theory was disproved by Ewan (1976) who demonstrated that larynx moved upwards during the production of speech. Another possibility for lowering of the pitch can be seen in the reduction of sub-glottal pressure as the lung volume decreases. The difficulty of this hypothesis dwells in the fact that the variation of sub-glottal pressure is much smaller than the magnitude of the downdrift. Ohala presented a third hypothesis as the most plausible. He claimed that the downdrift was caused by “changes in vocal cord tension, and that is not an ‘automatic’ effect at all, but purposeful (…) because the gradual pitch decrement in utterances serves a useful linguistic purpose in signalling clause and sentence boundaries.” (Ohala 1978: 32)

The second aspect is called the peak delay and it occurs because of the time gap before the impulses from the brain reach the muscles operating the vocal folds and command them to

move. As can be seen in Xu (1999) the effects of these time gaps can be observed most clearly on rising tones where the tonal peak is shifted from the syllabic core to the end or even to the onset of the next syllable. However, the same phenomenon influences low or different contour tones as well. It is known as tonal coarticulation and it is phonologized in certain tonal languages. In Vietnamese, Brunelle (2009a) investigated coarticulation in Northern and Southern Vietnamese and discovered a certain degree of it on bi-directional scale but not prominent enough to influence perception. However, in his more recent research (Brunelle et al. 2016), he claimed that the rising tone in Hanoi Vietnamese before a high level tone can be in some situations perceived as the falling tone exactly due to peak delay. He further noted that this phenomenon could result in tonal sandhi in the future but there was no phonological effect on the contemporary language. (see 2.2.2.6. and 2.2.3.4)

2.1.3 Tonogenesis

When dealing with tone languages, we cannot forget to mention some of the theories explaining the origin of lexical tones. We will focus on three theories based on different perspectives. The first, postulated by Hombert et al. (1979), is a phonetic theory that seems to be the most widely acknowledged by the scientific audience. The second theory could be called biomechanical (Everett et al. 2015) and the third, advocated by Dediu and Ladd (2007), is the genetic theory. Although the latter two theories seem to be received with suspicion by the scientific community, they should also be mentioned as they are based on valid research methodologies.

Hombert et al. authored a now canonical paper on tonogenesis based on the findings of Matisoff (1973) and Ohala (1973) in terms of the general nature of tone as well as, among

others, on two studies by Maspéro (1912) and Hadricourt (1954) that are also considered canonical in the field of Vietnamese tonology.

Studying Vietnamese, Maspéro noted the correlation between initial consonants and tone. Hombert et al. tested Maspéro's findings supporting his hypothesis with empirical data. They confirmed that voicing distinction in prevocalic position can affect the F_0 of the following vowel, specifically, that F_0 of a vowel following a voiced consonant is significantly lower than F_0 of a vowel after a voiceless consonant. Fig. 2.2. shows that although the difference is most apparent at the beginning of the vocalic onset, it is still clearly detectable 100 ms into the vowel.

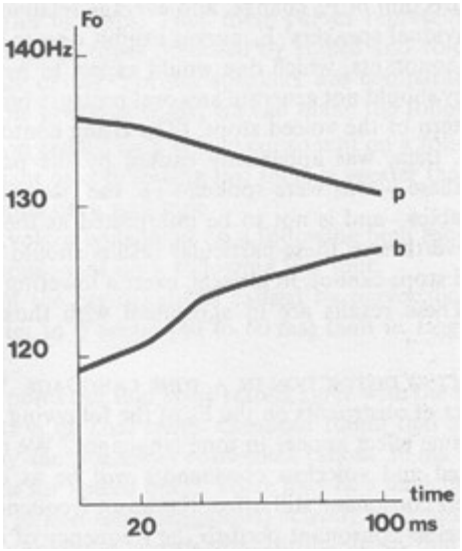


Figure 2.2. *Average F_0 values of vowels following English stops gathered from five speakers (Hombert et al. 1979)*

Hombert's research team suggested the following hypothesis: "During a voiced stop, oral pressure gradually builds up, thus decreasing the pressure drop across the vocal cords – which in turn decreases the F_0. Upon the release of the stop, the pressure drop returns to normal, producing an initially low and rising F_0 contour after voiced stops. In case of

voiceless stops (particularly aspirated ones), the airflow past the vocal cords is supposedly very high upon release, creating a higher-than-normal Bernoulli force – which will draw the vocal chords together more rapidly, and thus increase the rate of their vibration at vowel onset. As the airflow returns to normal, the F_0 will too. Thus, after voiceless stops, the F_0 contour will be initially high and falling." (Hombert et al. 1979: 42)

Besides the effect of voicing on tonal development, Hombert et al. researched the loss of syllable-final glottal stops. Hadricourt (1954) claimed that the loss of final glottal stops in the 6th century AD led to the emergence of a high tone in Vietnamese. Hadricourt's claim was put to test on Arabic and it was discovered that "[h] produces a drop in F_0 (varying from 25 to 50 Hz) on the preceding vowel, while [ʔ] produces a rise in F_0 (from 9 to 48 Hz). It was also shown that these two curves became significantly different at least 70 ms before vowel offset." (Hombert et al. 1979: 51)

Although it has been shown (Peterson & Barney 1952; Maddieson 1997) that higher vowels have an intrinsically higher fundamental frequency than low vowels, which means that they influence F_0 in a similar way as voiced/voiceless prevocalic consonants, it seems that the emergence of tones by means of merging vowels is very rare. Hombert et al. attempted to explain this phenomenon by claiming that "our auditory system seems to be more 'efficient' at detecting dynamic changes in F_0, rather than differences in level (of the same magnitude) between two F_0 signals, this may account for the difference in perceptual saliency of the two phenomena." (Hombert et al. 1979:53)

Besides the linguistic theories of lexical tone origin mentioned above, there are others that perceive the emergence of tones from rather unorthodox angles. Everett at al. (2015) saw climate as the main cause of the origin of lexical tones. "The biomechanical properties of the vocal folds are influenced directly by hydration levels. For instance, dehydration of the

vocal folds results in decreased amplitude of vocal fold vibration. (…) Increased hydration is associated with heightened vocal fold viscosity and facility of phonation. In contrast, dehydration of the vocal folds is associated with increased phonation threshold pressure and increased perceived phonation effort." (Everett et al. 2015: 1)

The authors hypothesized further that because lexical tones (complex ones in particular) require precise manipulation of F_0, they should be disfavoured in arid climatic conditions. They clarified their claim by stating that "languages with complex tone can be spoken in any geographic context – for instance, Cantonese speakers are more than capable of communicating in the Siberian tundra – but it seems less likely that such languages would develop their complex tonality in areas with typically frigid and/or desiccated ambient air." (Everett et al. 2015: 2)

If we take a look at the map of tone language distribution as presented by Maddieson (2013) in Fig. 2.3, the hypothesis seems quite plausible as most of languages containing complex tonal systems gather around the equator where the climate is hot and humid whereas areas with frigid climate accommodate very few tone languages and hardly any of them contain complex contour tones.

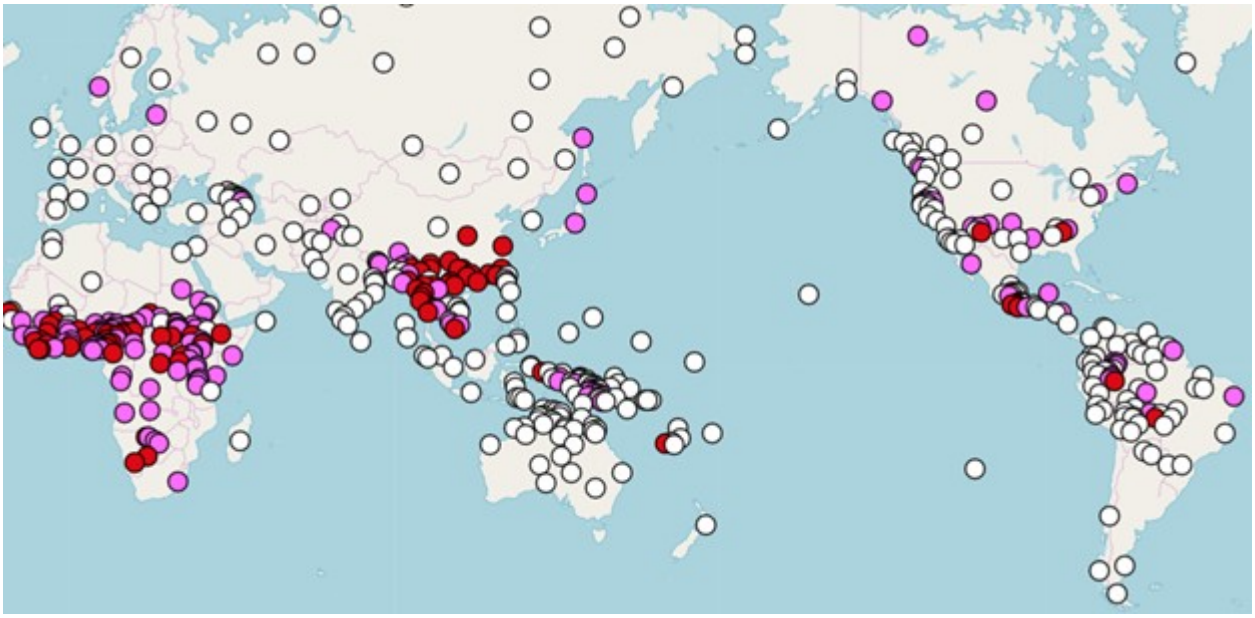


Figure 2.3. *Map of languages composed by Maddieson (2013). White points represent non-tonal languages, pink points tonal languages with simple tone system and red points stand for languages with complex tone system.*

The last perspective regarding tonogenesis mentioned in this thesis is the genetic one advocated by Dediu and Ladd (2007) who devoted their attention to determining the correlation between the occurrence of lexical tones and two haplogroups of brain growth and development related genes ASPM (Abnormal Spindle-like Microcephaly-Associated Protein) and Microcephalin. Their results can be seen in Fig. 2.4., and they indicate that speakers of tone languages manifest low frequency of ASPM and partially even of Microcephalin.

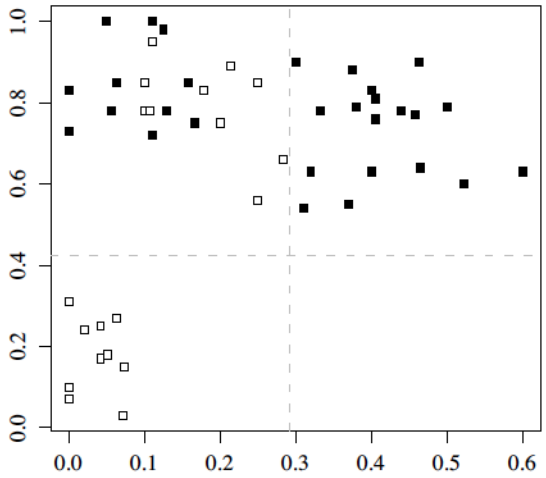


Figure 2.4. *The horizontal axis represents the frequency of ASPM, the vertical axis represents the frequency of MCPH. Filled squares represent non-tonal languages and open squares tonal languages. (Dediu & Ladd 2007)*

The research of Dediu and Ladd, however, should be regarded with caution as the correlation might be just random. Their research has not been cited by any big names in the field of tone language research.

2.1.4. Tone marking

There are three major enclaves of tone languages: Africa, the Americas and Asia. Due to the fact that native speakers do not deem it necessary to overtly mark tones in writing, most tone marking systems were devised for phonetic transcription by scholars (often foreign) researching and describing the languages in question. As each linguistic region has its idiosyncratic descriptive needs and academic habits, the tone notation systems vary significantly. In African languages, acute accent *á* marks a rising or high tone, grave accent *à* represents a falling or low tone, level accent *ā* represents a level or a mid tone and fall-rise or rise-fall tones are marked *ǎ* and *â*. Most African tonal languages do not exhibit rich tonal inventories, hence they employ just some of the accents often in opposition to tonally unmarked graphemes *a*.

The Asian tradition is more elaborate than the previous two. Vietnamese similarly to Mandarin, Cantonese and other prominent tonal languages adopted the system known as "Chao tone letters" introduced by Chao (1930), his "letters" being, in fact, numbers. The numbers divide the speech pitch into five levels with 1 being the lowest and 5 the highest level. "Five levels seems to be the maximum used by any language, with anything beyond four being exceedingly rare. Each syllable is given zero to three digits, usually written after the segmental transcription. Zero digits means the syllable has no phonological tone of its own. Most syllables are given two digits, one for the starting pitch and one for the ending pitch. This is true even for level tones, unless the syllable is very short, in which case only one digit is usually used. Three digits are used for tones which change direction in the middle of the syllable." (Yip 2002: 20) It has been established in Vietnamese that glottalization or creakiness is a phonological cue for tonal distinction, therefore a glottal stop tends to be used to indicate this (ma3ʔ5). Diagrams of tonal shape drawn next to a vertical line to illustrate voice range can accompany or replace the numeric notation.

In the Americas they use a numeric system similar to what we work with in Asian languages but the digits are reversed. The system also works on the scale from 1 to 5 but, as opposed to Asian languages, 1 being the highest *ma1* and 5 being the lowest *ma5*. Complex tones use multiple numbers with hyphens to reflect the contour or direction of the tone (e.g. ma3-1 for rising or ma3-5 for falling).

The numeric system is very helpful in linguistic analysis of Asian languages. Had there been a universal set of accents like in the African notation style, it would have to be very complex and it would require a vast inventory of diacritical signs. Asian tonal languages with large numbers of speakers like Mandarin or Cantonese do not even deem it necessary to convey

precise phonemic information in writing. In the past, Vietnamese also used to be written in characters originating from ancient Chinese and there was no overt tone marking. However, Western missionaries in the 17th century devised a transliteration method of the characters into the Latin script including diacritical marks for 5 of the 6 tones long before the invention of pinyin in mainland China. As the Vietnamese script is used for every day communication and pinyin also serves various non-scientific purposes, employing the numeric system for tone notation would render it too complicated for layman usage and noting tones by means of diacritical signs makes it more user friendly.

Moreover, the numeric notation faces one major data-related problem. Tonal F_0 is not fixed but varies from speaker to speaker and even in one speaker based on gender, age, mood, position within an utterance, speech rate etc. It is therefore rather problematic to establish whether a high level tone should be marked 44 or 55 or whether a mid falling tone resembles most the pattern 31, 32 or 21. Nowadays, it would be technologically possible to gather a large quantity of data with large gender and age variation, measure F_0, normalize it and calculate the numeric distances. However, the numeric notation for Vietnamese as well as Mandarin and other languages was decided on in the past when it was determined mainly based on subjective opinions of the researchers.

Another difficulty with the numeric system does not concern Vietnamese as tones within one utterance do not seem to influence each other enough to lead to any phonological changes. In Mandarin, on the other hand, tones can be deleted or altered based on their environment and therefore the numbers can also change.

2.1.5. Autosegmental representation of tones

As opposed to the tone marking methods discussed in the section 2.1.4., autosegmental phonology represents tones on a tier separated from the syllable turning them into autonomous segments. The advantage of the autosegmental representation is that it can be utilized to represent intonation even in non-tonal languages employing only a limited number of symbols. Fig. 2.5. illustrates three types of relationships between tones (T) and tone bearing units (TBUs) marked as vocalic core (Vc).

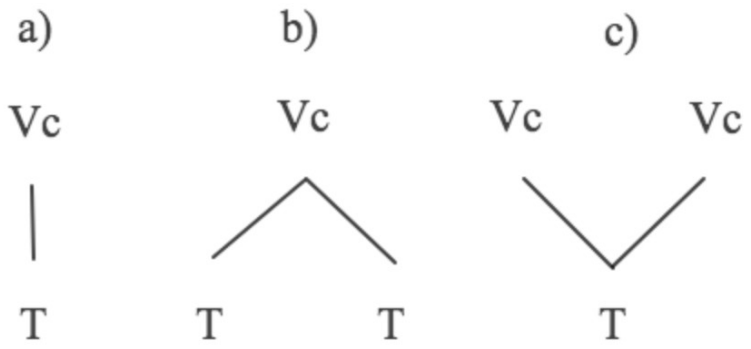


Figure 2.5. *Relationships between tones (T) and TBUs (Vc); a) represents the one-to-one case where one one tone is linked to one TBU, b) is a manifestation of a contour tone because two or more tonal segments are linked to one TBU and c) shows multiple association where one tone runs over two or more syllables.*

Symbols used for autosegmental representation are based on the work of Pierrehumbert (1980). Level tones are marked by a single letter according to tone height and contour tones are marked by a sequence of letters H and L based on the shape of the contour.

Level		
high	-	H
mid	-	M
low	-	L
Contour		

rise	-	LH
fall	-	HL
fall-rise	-	HLH
rise-fall	-	LHL

Table 2.2. *Symbols used for autosegmental representation of tones*

2.1.6. Tonal contours

"True contour tones seem to be added to tonal inventories only in languages with large numbers of tonal contrasts. Two-tone languages usually contrast two level tones, rather than a rise and a fall." (Yip 2002: 24-25) Many languages contrast three level tones, some, such as Grebo and other Kru languages or Jianyang, contrast four level tones, and there have been reports on languages contrasting 5 level tones (Hei-Miao, Gaoba Dong, Dan or Trique) but it has been argued that most of them in fact contrast merely 4 tones (Yip 2002: 27). The claims concerning the 5-level contrast, however, led to establishment of the 5-digit numeric tonal notation scale.

As mentioned above, contrasting more than three level tones is rather rare but many tonal languages have a tonal inventory of 5 or more tones. In these languages, a small number of level tones tends to be supplemented by contours. The two simplest contour tones are a fall and a rise. According to Zhang (2000), falls are more common than rises. Two frequently occurring complex contours are a rise-fall (convex tone) and fall-rise (concave/dipping tone). The existence of tones with complex contours implies the existence of rises and/or falls within the researched language.

Standard Vietnamese contains one high level tone, one rising tone, two falls (one accompanied by laryngalization) and two fall-rise tones (one interrupted by heavy glottalization), which conforms to the theory that level tones trigger the origin of simple

contour tones and simple contour tones lead to the appearance of complex contour tones in case the tone register needs to be extended.

When discussing contour tones, a question arises whether they are individual units or just sequences of level tones. In polysyllabic languages, vowel deletion may lead to the origin of a contour tone, which would favour the hypothesis that complex contour tones, in fact, might be a sequence of two simpler tones. Polysyllabic tonal languages occur mainly in Africa but they are rare in Asia; even there can be found examples supporting the same hypothesis mainly because they contain sandhi or phonologized coarticulation. Vietnamese is neither a polysyllabic language nor it contains tonal sandhi and, therefore, might serve as evidence for contours as units. However, a fall-rise tone with distinct central glottalization in prominent syllables of the Hanoian dialect could be perceived as two separate tones (glottalized fall and level) on the same vowel.

2.1.7. Consonant types, vowel quality and phonation

In natural speech, it is impossible to isolate tones from their TBUs. There are languages where tones do not affect the TBUs but there are also many tone languages where tones interact with consonant types, vowel quality or phonation.

Consonantal interaction with tones can be clearly demonstrated by the association of voiced obstruents with lower tones (see Hombert et al. 1979). Synchronically, this phenomenon is extremely common in African tonal languages. In Ewe, according to Smith (1968), a language spoken south-eastern Ghana, almost all voiced consonants have this effect. In Vietnamese, we can observe this interaction only diachronically (see section 2.2.4.) but it is present nonetheless. Voiced consonants can also cause lowering of neighbouring tones, in

which case they are labelled depressor consonants. However, it has been mentioned above that there are no phonological changes of Vietnamese tones based on their interaction with the exception of reduplication. (see 2.2.2.5)

It has been shown (Peterson & Barney 1952; Maddieson 1997) that higher vowels have an intrinsically higher fundamental frequency than low vowels, which should imply that high vowels should be found with higher tones but there does not seem to be any manifestation of this phenomenon in Vietnamese.

"In some languages, different tones are associated with different laryngeal properties, such as breathiness or glottalization." (Yip 2002: 31) Breathy vowels can create contrast to plain vowels and they tend to have a lower pitch as explained earlier in this section when discussing the consonantal interaction. Breathiness seems to be one of the cues for discriminating tones in Vietnamese. Apart from breathiness, glottalization can be a crucial component of low tones in particular. This applies also to the Hanoian as well as Saigonese dialect of Vietnamese.

2.1.8. Intonation and tone

Lexical tone can be categorized as a subcategory of intonation. Tone is a segmental manifestation of intonation where the segments constitute lexical items. It is therefore necessary to devote this chapter to the relationship between tone and intonation in general.
Ladd (1997) defined intonation as "the use of suprasegmental phonetic features to convey 'postlexical' or sentence-level pragmatic meanings in a linguistically structured way." Yip (2002) argued that the phrase 'pragmatic meanings' is not sufficient because fails to include syntactic and semantic information.

It might seem logical to assume that tone languages reserve pitch changes entirely for the purposes of lexical contrast but, in fact, they also employ intonation in order to convey pragmatic meaning. Vietnamese is rather impoverished in this respect and, therefore, it must seek lexical means of expressing pragmatic meaning (see 2.2.2.7.), which is a consequence of its abnormally wide tonal register. However, according to Brunelle (2012) it cannot be claimed that Vietnamese has lost the ability to express pragmatic meaning by means of sentential pitch change completely.

In various tonal languages, sentential intonation might be more limited than in non-tonal languages but it exists and Yip (2002: 260-261) lists four mechanisms of its use. "First, the entire pitch register may be moved up or down, so that all tones are higher or lower in pitch than usual. Second, the pitch range can be widened or narrowed, with highs and lows either moving further apart or closer together. Third, boundary tones may be inserted at domain edges, and these may then surface either on the closest head or on the edgemost syllable. Fourth, downstep may apply within some prosodic domain, such as the phonological phrase or intonational phrase, but at domain boundaries downstep may be suspended and register reset. These mechanisms may be associated with particular syntactic, semantic, or pragmatic environments, such as questions vs. statements, focus or emphasis, certainty vs. suggestion, and so on."

Chao Yuen Ren (pin-yin: *Zhào Yuán Rèn*) pointed out in his book, *A Grammar of Spoken Chinese*, that there is no cross-linguistic guarantee that any particular function is associated with any particular intonation pattern. Bringing attention back to Vietnamese, in many European languages yes-no questions are marked by rising intonation towards the end of the utterance. In Vietnamese, they are usually expressed by a sentence-final syllable with a high

level tone (không, chưa) or a falling tone (à). Furthermore, Chao (1968) compared the relationship of lexical tones and intonation to "small ripples riding on larger waves".

2.1.8.1. Intonation in non-tone languages

Intonation in non-tonal languages can be described as "the ensemble of pitch variations in the course of an utterance" (Noteboom 1999: 2) or "the product of a conflation of different prosodic systems of pitch contrast" (Crystal 1969: 6). Although lexical meaning in non-tone languages is not influenced by pitch change, it does not mean that that all lexemes are uttered with the same pitch. Different languages possess different sets of intonation melodies with respect to their semantic and pragmatic needs. English, according to Cruttenden (1986), contains seven different melodies.

Awareness in regard to intonation melodies is crucial for native speakers of non-tone languages attempting to learn a tonal language. There are languages in Africa such as Yoruba that contrasts a number of level tones (Yip 2002) but Asian tone languages like Mandarin or Vietnamese usually contrast at least one rising and one falling tone. In English, one of the functions of falling intonation is indication of a statement and rising intonation often indicates a yes-no question. Realizing this, even the most tone-deaf student of a tone language with good command of English is able to gain a helpful learning aid.

2.1.8.2. Intonation in tone languages

A common strategy to avoid possible clashes between intonation and lexical tones is using particles most of them being sentence-final; sentence-internal particles also exist but they are not as frequent. There seems to be a positive correlation between the number of tones and the number of particles. Chao (1968) listed twenty-eight such particles in Mandarin,

Law (1990) talked about 35 to 40 particles in Cantonese and claimed that almost all Cantonese utterances end with a particle. Vietnamese has more than 20 sentence-final particles and a number of additional sentence-internal particles.

Addition of a phrase-level tone can be another mechanism that substitutes intonation in tone languages. Yip (2002: 275) mentioned that "In Vietnamese, questions usually end in a particle, but all the question particles have high tone (e.g. *sao*, *ai*, *nhé*, *không*, and many others), so that the sentence overall rises at the end. The same rising pattern is found in one type of interrogative with no final particle, suggesting that the high pitch is a question intonation superimposed on inherently toneless particles, or on the last syllable of a sentence with no particle, and is thus a type of boundary tone."

However, Yip's claim seems to be misleading if we look at other question particles like *gì* (what) or *nào* (which) that are at least as frequent as those listed by Yip and they carry a falling tone. Moreover, although Vietnamese yes-no questions often end with the particle *không* carrying a high level tone, it seems a matter of syntactic structure rather than intonation. Moreover, there is the particle *à* also serving as a yes-no question marker although it carries a falling tone.

In order for the tones to retain their relative pitch contours, the whole pitch registers can be adjusted – lowered, raised or expanded. Taiwanese (Peng 1997) can be seen as an example of register adjustment by lowering. Overall pitch raising and expansion for emphatic reasons can be found in Mandarin (Xu 1999b).

2.1.8.3. Tone languages, stress languages and accent languages

There are three language types categorized based on intonation features. Tone languages where the pitch (often combined with various types of phonation) alters lexical meanings are discussed in substantial depth throughout the section 2.1. Tone, however, is closely related

to another integral part of intonation – stress. According to the definition by Hayes (1995: 8), "stress is the linguistic manifestation of rhythmic structure. That is, in stress languages, every utterance has a rhythmic structure which serves as an organizing framework for that utterance's phonological and phonetic realization." "Accent language" can be seen as a subcategory of tone language because "the majority of such languages have lexical tones, and what makes them special is that these languages have a small number of contrasting tones, sparsely distributed or absent on some words." (Yip 2002: 257)

2.1.8.4. Stress

Acoustic and perceptual properties of lexical tones change based on duration or prominence (Hermes 2006), which are values directly linked to stress. Gussenhoven (2004) presented a set of criteria helpful in detecting the differences between stressed and unstressed syllables.

a) *Vowels in stressed syllables have a fairly even intensity distribution across the frequency spectrum, while vowels in unstressed syllables tend to have lower intensities for the higher frequencies, thus displaying a downward slope towards the higher end of the spectrum.*
b) *Due to a reduced effort to pronounce vowels in unstressed syllables, their quality is likely to be more centralized and less rounded (more schwa-like) than of unstressed syllables.*
c) *Consonants and vowels in stressed syllables tend to be longer than those in unstressed syllables.* (Gussenhoven 2004: 14-15)

2.1.9. Tone perception

In section 2.1.1., we introduced the three crucial terms: F_0, pitch and tone, and we discussed their mutual relation. Tone perception is dependent on pitch perception and therefore on decoding the changes in F_0. The key question to ask when dealing with tone perception is how large the F_0 changes must be in order for the human ear to detect them.

According to Klatt (1973), the minimal detectable difference for sounds (difference limen) with a level F_0 was around 0.3 Hz and 2 Hz for slope F_0. Klatt also demonstrated that discrimination of pitch is easier on segments without formant change (simple vowels opposed to diphthongs). Moreover, he introduced the idea that vocalic cores are the principal carriers of the tone and that pitch is absent or at least unstable on consonantal onsets. All of these findings proved very valuable in devising the research methodology of this thesis.

Considering Klatt's claim that the human ear is sensitive enough to notice differences of 0.3 Hz, we must realize that he was using steady synthetic vowels and not natural speech. Harris and Umeda (1987) developed Klatt's research thanks to the progress of speech data processing technology, tested the difference limens (DLs) of natural speech and discovered that the subjects exhibited DLs for natural sentences that were 20 times larger (5 to 16 Hz) than those for steady synthetic vowels. Therefore, they drew the conclusion that subjects were considerably less sensitive to change of F_0 in natural sentences in comparison with synthetic stimuli. Nevertheless, it should be noted that subjects of the initial experiment with synthetic vowels were highly trained individuals whereas the subjects of this experiment were completely linguistically untrained.

In Pollack's (1952) experiment, listeners were presented with sets of level tones and asked to label them numerically. The results showed that the subjects were able to label up to 5 tones rather reliably but after the number of tones reached six and above, the performance became substantially more erratic. The experiment could supply the answer as to why it seems to be more difficult to master Vietnamese pronunciation with six tones in its standard dialect than Mandarin with merely four. On the other hand, Pollack's findings might only be a manifestation of what Miller (1956) demonstrated in his famous article, *The Magical*

Number Seven Plus Minus Two: Some Limits on Our Capacity for Processing Information, i.e. that the number of objects an average human being can hold in working memory is 7±2.

In terms of contour tones perception, there is a temporal limitation. Greenberg and Zee (1979) discovered that if the syllable duration is less than 40-65ms, the tone will be perceived as level even if the F_0 changes. Furthermore, they established that the segment should be at least 120ms long in order for the contour to fully show. Hermes (2006) assessed in his model any tone under 100ms as a level tone.

Hermes (2006) further revealed that the interval of the pitch contour covering the syllable onset and the first 20-30 ms of the vowel contributes very little to the perception of the syllabic tones. Therefore, perception of tones is mainly determined by the part of the contour covering the syllable rhyme 20-30 ms after the vowel onset. Such observation conforms to research on tonal languages. Hermes also mentioned that the vocalic core of is the most important area for the process of tone discrimination.

2.1.10. Tone identification

In natural speech, F_0 is always accompanied by other features that can serve as auxiliary cues for tone identification such as duration, amplitude or voice quality. Although F_0 tends to be the primary cue, it likely is not the only one as proven by the research on Mandarin by Fu and Zeng (2000) who found out that even after stripping the tone-bearing segments of all F_0 information, the recognition rate was around 70%.

An obvious difficulty with distinguishing tones based on F_0 might be that the F_0 of a high tone uttered by a male speaker might, in fact, be lower than the F_0 of a low tone uttered by a

female speaker. Tone identification without linguistic context is therefore seemingly impossible. However, Abramson (1975) in his early work on Thai demonstrated that native speakers of tonal languages can identify stimuli deprived of linguistic context rather well. He discovered that confusion occurs most frequently with tones of similar shapes. Rising tones are rarely confused with falling tones.

Cutler and Chen (1997) carried out an experiment asking a group of native speakers of Cantonese to judge whether the presented stimuli were words or non-words. The non-words differed from words in various aspects – onset consonant, vocalic core, coda, tone or a combination of some of them. They found out that the highest error rate appeared when the difference was tonal, alteration of the vocalic core also had a high error rate. In their second experiment, they asked the subjects to determine whether a pair of syllables was identical or different. As in the first experiment, if the only difference was the tone, the results were less accurate and the response time was slower. Cutler and Chen offered an explanation that the tonal information comes late in the syllable. The vocalic core carries the tone and its contour cannot be fully identified until late in the vowel. Hearers thus identify tones more cautiously and not as quickly as the other aspects. Their findings together with the findings of Hermes (2006) serve as a basis for the decision to start measuring the tonal contour 15 ms into the vowel duration (for more details see chapter 3).

The paragraphs above dealt with how tone identification could be influenced by the lack of context but there are many situations in tonal languages where abundance of context can affect or even alter tone perception. Peng (2000) looked at the third-tone sandhi rule in Mandarin. If there are two low dipping tones (214) one after another, the first changes from the standard fall-rise contour to a high-rising tone (35), e.g. *hello* 你好 (nǐ + hǎo » ní hǎo). His experiments yielded surprising results since although the realizations of the sandhi-

affected tones had phonetic properties almost exactly like the rising tone, the listeners categorized them very convincingly as the low dipping tone, hence another proof that tone perception is not based solely on F_0. Sections 2.2.2.6 and 2.2.3.4. describes the effects of context on Vietnamese tones but, unlike this change in Mandarin, no context interactions of tones have been phonologized in Vietnamese.

As already hinted above in section 2.1.2., there are no context-induced phonological changes of tones in the Vietnamese language. However, Brunelle (2009a) dealt with the issue of tonal coarticulation and concluded that Vietnamese really exhibits non-phonological features of coarticulation. Fifteen years prior to that, Xu (1994) researched coarticulation in Mandarin with the aid of perception tests where he used stimuli naturally occurring in speech context, the same stimuli in isolation as well as with artificially inverted context. The perception of the stimuli changed significantly. For instance, a rising syllable (LH) between low (L) and high (H) syllables in the preceding and following context would be perceived as rising in the natural environment, high level in isolation and falling in the artificially reversed context. Xu's experiment clearly manifested the importance of context in tone recognition. At the same time, it showed that native speakers hardly ever make mistakes in tone discrimination in a natural language.

2.1.11. Tone language acquisition

Using infants and toddlers as subjects for linguistic research is rather problematic as they have a very short attention span and communication with them is generally challenging. Adding the fact that most of the works on early language acquisition have been carried out on English or other non-tonal languages, it all indicates that not many scientific studies have been written on the early stages of tone language acquisition in children.

Harrison (2000) carried out a research on Yoruba, one of the official languages in Nigeria, with three lexical tones, high, mid and low. He selected two groups of infants aged six to eight months. In the first step, he presented the subjects with pairs of a synthetic syllable [ki] with pitch difference of 10 Hz, 20 Hz or 40 Hz and attempted to investigate whether the infants could detect any of the differences. The English infants did not manage to detect any differences whatsoever whereas the Yoruba infants displayed a level of consistency in differentiating between the stimuli 20 Hz and 40 Hz apart. The difference of 10 Hz was left unnoticed even by the Yoruba infants. In the second step, he set the differential of the stimuli to 20 Hz; the English infants were completely unsuccessful telling the stimuli apart and even the Yoruba speakers were consistently accurate only in the pitch-range of 190-210 Hz. For the purposes of this thesis, the key information of Harrison's work is that language environment seems to affect the ability to detect lexical tone in infants as early as in 6 to 8 months of age.

What renders tone production in small children more complicated is that the main tone articulators are the muscles controlling vocal chords and larynx whereas in terms of segments, the most important articulators are oral and nasal together with accurate control of tongue movement. Moreover, acquisition of tones is a gradual continuous process and therefore the child can accurately produce syllables with certain tones, whereas stay erratic with other tones. It might also be possible that the child produces some lexemes very accurately while not being aware of the tonal system yet.

Li and Thompson (1977) carried out a study on tonal acquisition in Mandarin later developed by Clumeck (1980). Li and Thompson tested the order of tonal acquisition in children with age span of 1.5 to 3 years and discovered that the high level tone (55) is

acquired first followed by the high falling tone (41), then the high rising tone (35) and finally the low dipping tone (214). Clumeck looked at the accuracy with which his subjects, 1 year 10 months to 3 years 5 months, were able to produce the four tones in Mandarin. The high level tone and the high falling tone were produced with accuracy over 95% whereas the rising tone was merely 61% accurate and the accuracy of the low dipping tone was around 74%.

Clumeck assumed that the low accuracy of the rising tone might be caused by the fact that rising tones are generally more difficult to produce due to the increased amount of energy necessary for their production. The case of the low dipping tone is more complex as there are two influential factors at play. The first tone in a sequence of two low dipping tones changes to the rising tone according to the sandhi rule in the adult speech whereas children tend to disregard the sandhi rule at first. In adult speech, the dipping tone is generally rarely pronounced in its full contour due to context, hence the children do not get exposed to it as much as to the other tones. Li and Thompson noticed that the oldest subjects of their research, who were about 3 years old, applied the third tone sandhi rule but still very inconsistently. Unfortunately, the age of their research subjects did not extend past 3 years of age and therefore they were not able to establish the threshold when the sandhi rule begins to be followed consistently.

Based on the works mentioned in this section, it is quite apparent that growing up in the linguistic environment of a tonal language influences both language production and perception. Speakers of tonal languages are more sensitive to changes in pitch but at the same time it takes them longer to master the production of their mother language effectively.

Deutsch et al. (2009) focused her research on absolute pitch among students of an American music conservatory. In their hypothesis, they argued that the speakers of tonal languages acquire the absolute pitch more frequently and easily than speakers of non-tonal languages because they basically obtain this skill as a by-product of their first language acquisition. They mentioned the well-known example from Mandarin where the syllable *mā* with a high level tone means "mother" and the same syllable *mǎ* with a low dipping tone means "horse". They claimed that the same labelling strategy that is used for lexical items is also used for musical tones such as E or F#.

Four groups of test subjects were selected all of them being musical students. There were three groups of Chinese- and Vietnamese-Americans categorized based on their fluency in Mandarin and Vietnamese. The first group was native-like, the second group was labelled semi-fluent and the third bore the name non-fluent speakers of a tonal language i.e. individuals of Chinese or Vietnamese origin growing up in a tone language environment without the ability to speak it. The fourth group was a control group comprised of Caucasian English speakers.

Deutsch et al. observed, that the absolute pitch was indeed present more frequently among the speakers of tone languages and there was a positive correlation between absolute pitch and tone language fluency. The results of the Asian group non-fluent in the tonal language did not differ from the Caucasian group by a statistically significant margin. Therefore, Deutsch et al. managed to prove that absolute pitch is dependent on linguistic background rather than ethnicity. Furthermore, the results indicated that musical training improves the ability of absolute pitch independently from language and also that the process of absolute pitch acquisition is similar to second language acquisition and it can be done by meticulous training.

Although the work of Deutsch et al. came from the field of musicology and it was not focused linguistically, the results managed to answer the ultimate question of people interested in learning a tone language as a second language: do I need to have a musical ear in order to learn a tone language? The answer based on Deutsch et al. as well as my own experience with learning as well as teaching tone languages would be: yes, but it is not absolutely necessary and a lot can be achieved through adequate exposure and drill.

2.2. The Vietnamese language

Vietnamese is a tonal language belonging to the Austroasiatic language family. In the past, it used to be further classified as the Mon-Khmer group and Việt-Mường subgroup (Peiros 1998). However, Sidewell (2009) introduced the hypothesis where he claims that it is unnecessary to differentiate between Austroasiatic and Mon-Khmer rendering the labels synonymous. Khmer and Vietnamese are the only widely spoken Austroasiatic languages serving as official national languages. There are approximately 16 million Khmer speakers and around 86 million Vietnamese speakers including nearly 60 000 living in the Czech Republic.

Typologically, Čermák (2004) classified Vietnamese as a polysynthetic language because it combines full lexical items to form other lexemes e.g. *người bán hàng* (shop assistant) = *người* (person) *bán* (sell) *hàng* (goods). Full lexical items can be also used as grammar markers e.g. *mới* (new) in *Anh ây mới về nhà*. (He has *just* gone home.). Based on isolating features such as the use of particles *đã* and *sẽ* functioning as temporal markers indicating past and future, or the means of expressing nominal plural by adding the elements like *các* and *những* before nouns to put them in plural, Skalička (2004) classified Vietnamese as an

isolating language. Typological classification of languages has been very popular in the Czech academic world. Thanks to the functionalist legacy of the Prague Linguistic Circle, it employs a very well devised methodology and it constitutes a compelling supplement or possibly even an alternative to the genealogical classification. On the other hand, the fame of typological classification of languages rarely reaches outside the scope of the functionalist tradition. Moreover, languages are organic entities and most of them contain features of multiple language types.

It is estimated (Trần Trí Dõi 2011) that about 60% of the Vietnamese vocabulary are borrowings from Chinese although the languages are genealogically rather unrelated and typologically different. Very often there are two words for one concept, one of which is considered purely Vietnamese and the other Sino-Vietnamese. Sino-Vietnamese words tend to be used in higher registers (science, academia, art, poetry, ritual and religious language) whereas the purely Vietnamese words are reserved for everyday conversation. E.g., the Vietnamese words for *wind* and *water* are *gió* and *nước*, the Sino-Vietnamese *phong* and *thuỷ*. *Phong thuỷ* is the Sino-Vietnamese transliteration of the Chinese characters 堪輿 (fēng shuǐ) and the Western world knows it as Feng-shui. Vietnamese speakers would never use the words *phong* or *thuỷ* when talking about weather. Analogically, using the syllables *gió nước* to denote Feng-shui would not be understood by native speakers.

There is a saying in Vietnamese, *phong ba bão táp không bằng ngữ pháp Việt Nam*, meaning that Vietnamese grammar is worse than a typhoon. The Vietnamese use it every time they want to emphasize how complicated and complex Vietnamese grammar is. The understanding of complexity and complicatedness can differ substantially depending on linguistic and cultural background. It is true that Vietnamese grammar employs a lot of features that are difficult to grasp without prior understanding of the Vietnamese culture. For

instance, personal pronouns are in fact nouns denoting family relationships that became grammaticalized. Vietnamese people address each other: sister, brother, father, uncle, aunt… based on age and social status despite not being related in any way. The meaning of the verb "to go" is conveyed by a set of verbs differing only by the direction of the movement. Translation of the sentence "we are going to the mountains" – *chúng ta lên núi* uses the verb *lên* meaning "to go up" whereas the verb *xuống* in *chúng ta xuống thung lũng* – "we are going to the valley" means "to go down". There are verbs *vào* and *ra* denoting inward and outward movement as well as movement to the south and to the north. On the other hand, speakers of Indo-European languages especially of those heavily inflected will find Vietnamese grammar rather impoverished with almost non-existent morphology in terms of prefixes or suffixes, inflection, conjugation and even expressing temporality. Functions of all these missing categories are taken over by word order and various particles and elements. Memorizing paradigms and drilling their usage in speech is therefore not the main obstacle to overcome in acquisition of the Vietnamese language. However, Vietnamese is extremely rich on the lexical level with numerous cases of strict context-based partial synonymy and abundant idiomatic expressions used on daily bases across all registers.

Contrary to morphology, Vietnamese phonetics is highly complex. In terms of consonants, there are many features considered problematic to acquire by non-native speakers. There is the phonological distinction between /t/ and /th/, the velar nasal /ŋ/ in syllable-final as well as syllable-initial position, and the voiceless velar fricative /x/. Syllable-finally, we can find two noteworthy consonantal phenomena: a) voiceless plosives /p/, /t/, /k/, /c/ are unreleased lacking the explosion and therefore difficult to identify, which in some dialects leads to homophony. Moreover, these syllabic codas dramatically affect tonality as they only allow for 2 tonal variants to occur (see 2.2.2.4.); b) rounded vowels /u/; /o/; /ɔ/ trigger a place of articulation shift forward in velar finals /k/ and /ŋ/ so they are released as velars but finish as

bilabials, hence they are called labio-velar by Phạm (2003). The vocalic system is extraordinarily rich with 9 single vowels differentiated by quality and 2 more by quantity. In addition, there are 2 semi-vowels and 3 diphthongs. The pinnacle of the Vietnamese phonetic system is the tonality. The standard dialect distinguishes 6 tones, the Saigonese distinguishes 5 and even the tonally simplest dialects of the Central Vietnam distinguish 4 tones. There has not been found any evidence for tonal sandhi or tonal reduction (Phạm, 2003), which are features typical of many tonal languages although Brunelle (2016) suggested a possible emergence of tonal sandhi in Vietnam. Moreover, Brunelle published two studies on tonal coarticulation. The first was the work on coarticulation in Northern Vietnamese (2003) and later he compared tonal coarticulation in Northern and Southern Vietnamese (2009a). His research suggests that a certain degree of bidirectional coarticulation is present in both dialects.

2.2.1. Dialects in Vietnam

The geographical shape of Vietnam is the cause of a rich variety of dialects. Lately, due to the gradual improvement of infrastructure, dialectal idiosyncrasies have slowly become leveled-up or they are vanishing completely. Nonetheless, centuries of geographical isolation due to obstacles like mountain ranges or rivers caused areas within close proximity to use dialects that are seemingly mutually incomprehensible. Although Gordina & Bystrov (1984) differentiated merely two major dialects, Northern and Southern, majority of authors (Đoàn, 1977; Vũ, 1982; Phạm, 2003) recognize three, adding the dialect of Central Vietnam to the two mentioned above. Thompson (1965) classified Vietnamese dialects as Tonkinese (Northern), Annamese (Central) and Cochinchinese (Southern)[3]; however, he created further

[3] Tonkin, Annam and Cochinchina were the names of Northern, Central and Southern regions of Vietnam used by the French in the colonial era before 1954.

sub-categories for the cities of Hà Nội, Sài Gòn (HCMC)[4], Đà Nẵng, Huế and Vinh. Such approach seems the most sensible firstly because city language tends to differ from the language of the surrounding rural areas and secondly because the more detailed the description, the more probable it is to capture distinctive features of the individual dialects.

The label "North Vietnamese dialect" still contains an extremely broad spectrum of varieties. For instance, the dialects of Hà Nội, Ninh Bình or Lào Cai all within the group of North Vietnamese dialects express an array of differences although they also share a number of common features, the most important of which is the full inventory of six tones. The "South Vietnamese dialect" is even a more controversial category as South Vietnam is divided in half by the Mekong delta and the area below the Mekong river located to the South West of HCMC is known in Vietnam as *miền Tây* meaning "Western region". Due to the geographical conditions set by the Mekong river, the Western region of Vietnam is linguistically quite different from the area around HCMC. Up until 2010 when the cable bridge to Cần Thơ was built, the only means of crossing the delta was by boat and together with the migration policies substantially limiting population mobility that were fully lifted only after the Đổi Mới economic reforms in 1986, these were very favourable conditions for the dialects below the Mekong river to develop to a large extent independently of the South dialects in HCMC and its vicinity. In regard to the topic of migration it should be noted that two rather massive migration waves from the North to the South occurred in 1954 and after 1975 (see 2.2.3.) but they did not in fact affect the Western region of Vietnam very significantly.

[4] Sài Gòn is the name of the river at the bank of which Nguyễn Hữu Cảnh established the Citadel of Gia Định in 1698 that was later also renamed Sài Gòn. As of 1929, population of Sài Gòn was mere 130 000 people and it was only in 1955 when the township of Sài Gòn was merged with neigbouring Chợ Lớn to become the capital city of the newly established Republic of Vietnam. Although the city of Sài Gòn was renamed Ho Chi Minh City in 1976 to commemorate the victory of the Democratic Republic of Vietnam and the subsequent reunification, the regional name Sài Gòn is still in use with very little political connotation. On the other hand, Ho Chi Minh City is the official name that must be listed in all kinds of legal, administrative and diplomatic context.

The map in Fig. 2.7. does not attempt to divide the country based on dialects but on regions with common geographical features. The orange colour indicates the Hoàng Liên Sơn mountain range in North Vietnam (with the highest Indochinese mountain Phan Xi Păng in the Lào Cai (4) province) surrounding the Red River delta with the capital city of Hà Nội all marked in red. North Central Coast (green) stretches down south past the imperial city of Hue to the Hải Vân Pass dividing it from the South Central Coast (blue). Above the lowlands of the South Central Coast, there are Central Highlands (pink) that are not as high as the Hoàng Liên Sơn mountain range in the North but they were historically still impenetrable enough to constitute a barrier that gave rise to dialect diversity. The South Eastern area (yellow) consists of the lowlands surrounding the Ho Chi Minh City neighbouring in the west with the Mekong Delta area (light green). Although the purpose of the map is to divide Vietnam geographically, it also provides a rather accurate overview of the regional dialects as their origin is strongly conditioned by geographical features.

2.2.2. Hanoian dialect

The Hanoian dialect is nowadays most often implied when talking about the standard Vietnamese language. The reasons for recognizing this dialect as the standard are historical and political rather than linguistic. The main and possibly only relevant linguistic reason dwells in the fact that Hanoian Vietnamese possesses the full register of 6 tones and therefore does not manifest any signs of tonal homophony as opposed to most of the other dialects. On the downside, there is a lot of homophony in Hanoian syllable-initial consonants.

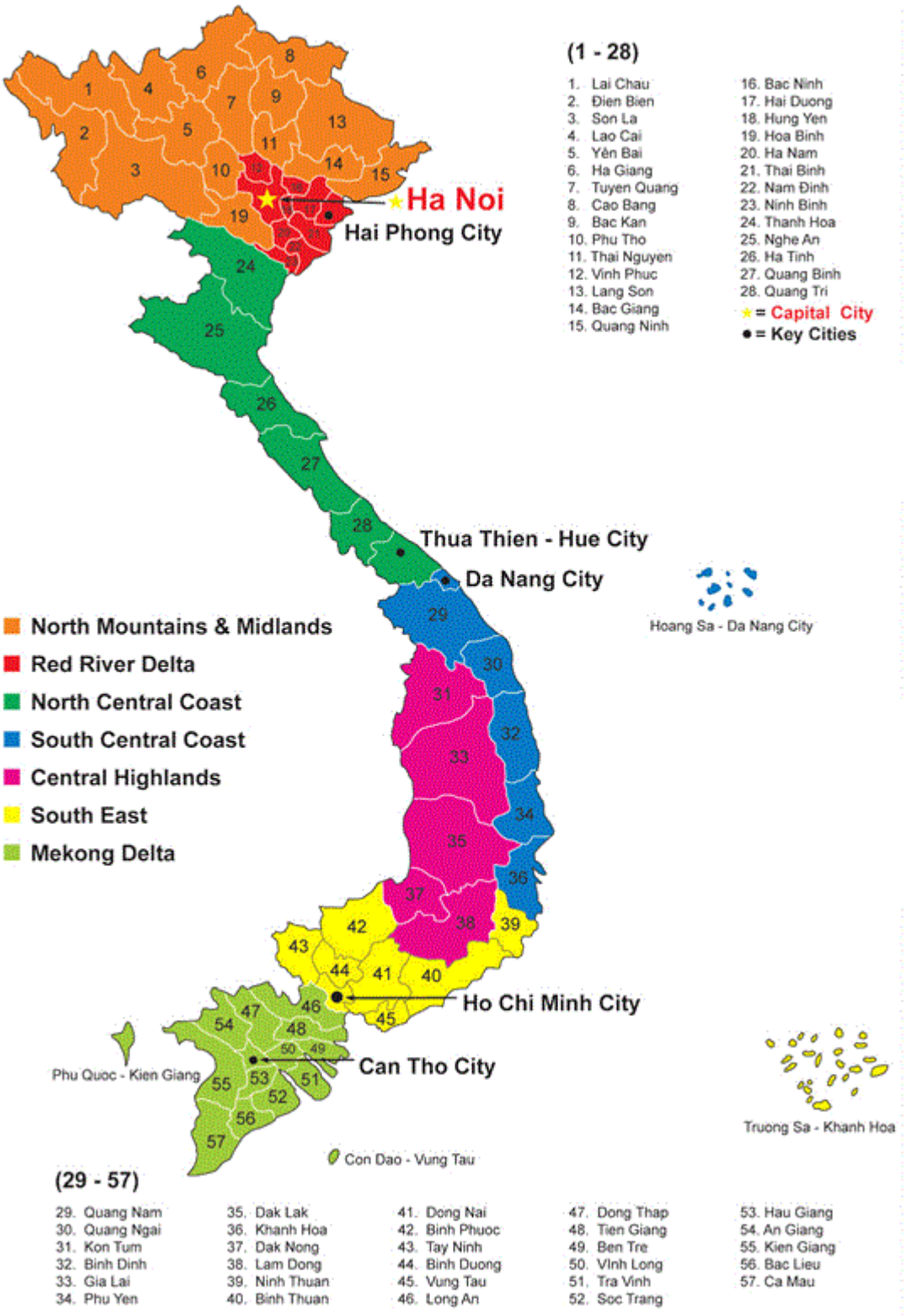


Figure 2.7. *A map of Vietnam listing all of its 57 provinces divided by colours according to the geographical affiliation (see text).*

Yet as it was already mentioned, the reasons for higher status of the Hanoian dialect were mostly historical and political. The first king of the Nguyễn dynasty, Gia Long, moved the royal court to Huế in 1802 when he managed to unify Vietnam after a period of political turmoil into a state with borders very similar to the present situation. He was from the South so he did not have a very close relationship to Hanoi but ruling from the South would have been inconvenient as well due to the length of the country. Hence, he proclaimed Huế the capital. The last king of the Nguyễn dynasty, Bảo Đại, abdicated in 1945 (although he stayed active in the South for another decade before seeking asylum in France) and Hồ Chí Minh founded the Democratic Republic of Vietnam in the North with Hanoi as its capital city in the same year. Since Hanoi became the centre of politics as well as education of the DRV, it was only logical for the local dialect to become the standard. In the South, the legacy of Huế still remained but after establishing the State Vietnam in 1949, Bảo Đại moved the capital to Saigon and following the Geneva Agreement in 1954 when he was forced to abdicate and flee to France, Saigon became the capital of the newly established Republic of Vietnam in 1955 with its own radio channels. The Saigonese dialect became the standard of the South. This setting, however, lasted merely two decades until the final victory of DRV and the subsequent reunification in 1976 under the new name "Socialist Republic of Vietnam". To emphasize the victory, Saigon was renamed the Ho Chi Minh City and Hanoian Vietnamese was imposed on the Southerners as the new standard. This, of course, affected the daily linguistic behaviour of the population very marginally but the Hanoian dialect became the language of the media and show business. The strict language policy in media was loosened but the aura of the dialect spoken in Hanoi being the most prestigious still remains among the people throughout the country.

Contemporary authors (Brunelle, Phạm 2003, Kirby 2011, Kiều 2012) use the Hanoian dialect as the chief representative of the North Vietnamese dialects due to the fact that it has

been described most clearly and thoroughly whereas the only comprehensive study of the Saigonese dialect was carried out by Thompson (1965) and Saigonese tones were also addressed by Vũ (1982) and Nguyễn & Edmondson (1997). There are other underlying reasons for choosing the Hanoian dialect as the research topic; mainly the easy access to suitable recording subjects as most people there speak an uncompromised variation of the dialect. Vast majority of individuals born and raised in Hanoi have parents who were also born in Hanoi or in its vicinity whereas people from the Ho Chi Minh City very often trace their immediate family roots to other places (see 2.2.3.).

2.2.2.1. Consonants

There are neither syllabic consonants nor consonant clusters in Vietnamese and thus a consonant can only occur syllable-initially or syllable-finally. Furthermore, it is always followed or preceded by a vowel.

According to Thompson (1965), there are 18 initial consonants in the standard dialect: /tʰ/; /t/; /tʃ/; /k/; /b/; /d/; /m/; /n/; /ŋ/; /ɲ/; /f/; /s/; /x/; /h/; /v/; /f/; /ɣ/ (sometimes realized allophonically as /g/); /l/ (initial /p/ appears in a limited number of lexical borrowings but only rarely and so /p/ is not considered a "native" sound). Kirby (2011) reclassified the semi-vowel /u̯/ (Đoàn 1977) to a labial approximant /w/ and lists it under consonants. Furthermore, he mentions the glottal stop that appears very frequently but it is technically not a phoneme. It occurs before syllables lacking initial consonants and according to the traditional classification it occurs at the end of the tone *nặng* and inside the vocalic core in *ngã*. On the other hand, if we consider the findings of Phạm (2003), then the articulatory features classified as glottal stop in the earlier studies become mere manifestations of

phonation types. Tab. 2.3. lists initial consonants in Northern Vietnamese according to Kirby (2011).

	Labial	Labio-dental	Dental	Alveolar	Palatal	Velar	Glottal
Plosive	ɓ		t t^h	ɗ	tɕ	k	ʔ
Nasal	m		n		ɲ	ŋ	
Fricative		f v		s z		x ɣ	h
Approximant	w						
Lateral approximant			l				

Table 2.3. *List of initial consonants in Northern Vietnamese (Kirby 2011).*

The choice of final consonants in the Hanoian dialect is limited to 8: four nasals /m/; /n/; /ŋ/; /ɲ/ and four voiceless plosives /p/; /t/; /k/; /c/. Articulation of voiceless plosives does not involve any audible explosion, which renders their perceptual distinction problematic. Furthermore, syllables with final plosives only occur with tones sắc and nặng (see 2.2.2.4.) and the tonal vowel appears to be perceptually shorter due to the abrupt change of pitch before the final consonant. Kirby (2011) distinguished final approximants /w, j/ that Đoàn (1977) classified as semi-vowels /u̯, i̯/. This thesis decided to include these segments into the vocalic core because not doing so would disrupt the continuity of the lexical tone's contour. During the articulatory process of the velar finals /ŋ/ and /k/ preceded by rounded vowels /u/; /o/; /ɔ/, the place of articulation shifts forward and they become labio-velar. They are, however, merely allophones to the velars in the other vocalic environments and they are not phonologically distinctive. Tab. 2.4. introduces the list of final consonants according to Phạm (2003).

	labial	alveolar	palatal	labio-velar	velar
Obstruents	p	t	c	kp	k
Nasals	m	n	ɲ	ŋm	ŋ
Glides	w	j			

Table 2.4. *List of final consonants in Northern Vietnamese (Phạm 2003)*

As opposed to the Saigonese dialect where the speakers struggle with homophony in the final plosives, the Hanoian speakers seem to be able to distinguish the finals without greater effort. This can be manifested looking at the written language. People from the South very often make spelling errors in syllables closed by plosives but the Hanoians hardly ever do. On the other hand, speakers of the Hanoian dialect make spelling mistakes when they have to distinguish between graphemes x|s; r|d|gi and ch|tr pronounced as [s]; [z]; and [tʃ] whereas in the South the variation of pronunciation is differentiated. Therefore, the words *da* (skin), *gia* (sino-viet. inner) and *ra* (go out) are all homophonic in Hanoi but not so in HCMC. Similar situation can be observed in *trán* (forehead) and *chán* (boring), *se* (almost dry) and *xe* (vehicle).

Many speakers of Northern dialects experience difficulties distinguishing /l/ and /n/ in the syllable-initial position although the two consonants are phonologically distinctive. The speakers occasionally utter /l/ when they mean /n/ or vice versa. For example, the word *lo* means “to be afraid” and the word *no* means “full, unable to eat any more”. Deciphering the correct meaning must be done contextually and in regard to the particular speaker, i.e., we realize that the speaker suffers from this speech impediment (called *nói ngọng* in Vietnamese) and adjust our perception accordingly. It is a socially undesirable phenomenon and inhabitants of Hanoi often claim that it affects only the uneducated peasants from the rural areas in the vicinity of Hanoi travelling to the capital city seeking work and not the people born and raised in Hanoi. However, rather than a regional phenomenon, it seems to be conditioned by access to education. In Hanoi, teachers fight this problem from pre-school

levels and universities offer classes aimed at eliminating it. People growing up in the rural areas around the capital city often struggle to acquire education matching the standard of Hanoi. It is noteworthy that in the Central and Southern regions of Vietnam, people do not seem to exhibit this impediment at all.

2.2.2.2. Vowels

Hữu Quỳnh & Vương Lộc (1980: 38) claim that the number of Hanoian Vietnamese vowels is 11. They distinguish front vowels (/i/; /e/; /ɛ/), central unrounded vowels (/ă/; /ɤ̆/; /ɑ/; /ɤ/; /ɯ/) and back rounded vowels (/u/; /o/; /ɔ/). From the presence of /ă/; /ɤ̆/ x /ɑ/; /ɤ/ and /o/; /ɔ/ x /e/; /ɛ/ it can be deduced that Vietnamese distinguishes vowels based on quantity as well as quality. Only vowels can be the centre of a syllabic peak in Vietnamese. In some syllables the peak is preceded by the pre-tonal semi-vowel /u̯/ that is classified as the approximant /w/ in more recent works (Kirby 2011). For the purposes of this thesis the approximant /w/ was included in the vocalic core as it quite clearly bears a part of the lexical tone's contour. Hanoian Vietnamese accommodates three centring diphthongs (/i͡ɤ̯/; /u͡ɤ̯/; /ɯ͡ɤ̯/). Combinations of vowels with final approximants /w/ and /j/ (Kirby's classification) or semi-vowels /u̯, i̯/ (Đoàn's classification) are not considered diphthongs but single vowels despite the fact that the lexical tone clearly manifests itself across the whole segment. There are no vocalic limitations to tones, any vowel can bear any tone. Fig. 2.8. shows the vowel quadrilateral with North Vietnamese single vowels and diphthongs according to Kirby (2011).

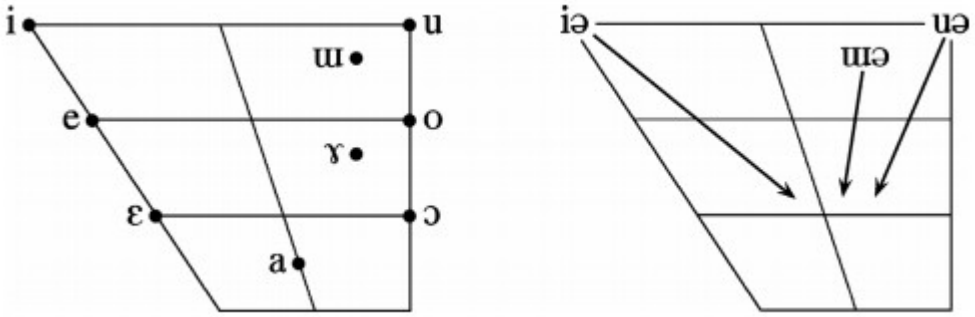


Figure 2.8. *Positions of Northern Vietnamese vowels and diphthongs (Kirby 2011), two short vowels /ă/ and /ɤ̆/ are not listed. Although Kirby decided to substitute the symbol "ɤ" (used by Đoàn and others) with "ə" in diphthongs, he did not mean to redefine their quality.*

2.2.2.3. Syllable

The only two obligatory elements in a Vietnamese syllable are the vowel serving as the syllabic peak and the tone. Onset, pre-tonal semi-vowel and coda are optional elements. The number of syllables containing all the elements at once is limited. Figures 2.9. and 2.10. below show a scheme of Vietnamese syllables suggested by Hữu Quỳnh & Vương Lộc (1980: 41) and two concrete realizations of such syllables.

	TONE		
	RHYME		
Onset	pre-vowel	Peak	coda

Figure 2.9. *General scheme of a Vietnamese syllable.*

nặ **ng** **(B** **2)**	**h** **ỏi** **(** **C** **1)**

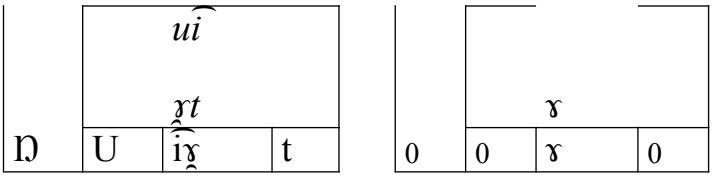


Figure 2.10. *Realizations of syllables nguyệt (moon) and ở (to be, to live).*

Phạm (2003) used a different syllabic scheme:

(C(w)) V (G, N or T)

C stands for any legal consonant optionally accompanied by the element classified either as a semi-vowel (Đoàn 1977) or a labial approximant (Kirby 2011). Phạm pointed out that "since the focus of this research is on the final segments, the initial consonant is simplified as C." (Phạm 2003: 3) She began measuring F_0 immediately after the end of the initial consonant but this study, influenced by findings of Hermes (2006), begins F_0 measurements 15ms after the onset of the vocalic core. Contrary to her, this thesis considers the element labelled (w) a part of the vocalic core because its signal intensity is stronger than the one of the preceding consonant and, moreover, there is a clear tonal contour present. As for the final segments, G stands for an off-glide, N for nasals and T for voiceless plosives. For the purposes of this thesis, both G and N are counted as part of the vocalic core. Phạm further discussed the important issue of what the domain of tone was. She arrived at the same conclusion as the one by Hữu Quỳnh & Vương Lộc (1980) portrayed in fig. (2.9.) that the most suitable unit for research of tones is what they call "rhyme" i.e. vocalic core plus sonorous coda if present. Syllable is a unit too large to be the domain of tone mainly because of the F_0 fluctuates in the transition between the consonant and the vocalic core whereas the vocalic peak is not broad enough and using it as the domain would fail to capture the tone contour in its entirety.

2.2.2.4. Tones

Traditional description of modern standard Vietnamese tonal inventory mentions six tones. Their labels in the Vietnamese language are very helpful because they are in fact carriers of the tones in question but an alternative numeral labelling had to be devised for people who lack the knowledge of Vietnamese. In older studies (Thompson 1965, Hữu Quỳnh & Vương Lộc 1980) as well as in pedagogical texts (Slavická 2008, Healy 2004), Vietnamese tones are distinguished solely on the basis of pitch contour. In modern studies (Nguyễn & Edmondson 1997, Phạm 2003, Brunelle 2003, 2009a, 2009b, 2012, Kirby 2011), however, it was proposed that the tones are also distinguished based on voice quality, specifically features like breathiness and creakiness.

When the Vietnamese themselves talk about the tones they use lexical labels (and numeric marking devised by Đoàn): *ngang* (1), *huyền* (2), *ngã* (3), *hỏi* (4), *sắc* (5), *nặng* (6). We can notice that the syllables used for tonal labelling are in fact realizations of the tones they describe. Moreover, most of the labels have a primary lexical meaning to some extent relevant to the tonal contour. *Ngang* means horizontal, *ngã* – to fall down, *hỏi* – to ask, *sắc* – sharp and *nặng* – heavy. *Huyền* is nowadays used mainly as the label of the second tone (and a female first name) but there is also a secondary Sino-Vietnamese meaning of "gloomy" or "dark".

The traditional approach distinguishing tones merely by pitch contour uses upper-index numbers that follow the syllable and represent its pitch contour: *ngang* **[ŋɑŋ^{44}]**; *huyền* **[hu̯i̯^{21}]**; *ngã* **[ŋa^{3ˀ5}]**; *hỏi* **[hɔi̯^{312}]**; *sắc* **[sa̯k^{35}]** *nặng* **[na̯ŋ^{21ˀ}]**. It should be pointed out that there has not been any consensus reached on the numbering (and glottal stop marking) and so the labels can differ with individual authors.

Ngô (1984) devised a binary classification of Vietnamese tones based on strictly pitch-related categories: *+-concave, +-contour, +-high*.

	ngang	huyền	sắc	nặng	hỏi	ngã
concave	-	-	-	-	+	+
contour	-	-	+	+	+	+
high	+	-	+	-	+	-

Table 2.5. *Binary classification of Vietnamese tones according to Ngô (1984)*

It should be noted that *huyền* is classified as a low level tone whereas most other works consider it a low falling tone. *Hỏi* is marked as high and *ngã* as low but most other researchers consensually classify them the other way around as can be seen in Tab. 2.5. by Hoang (1986). Phạm (2003) stated that although *ngã* is phonetically high, phonologically it is low based on its reduplication rules (see 2.2.2.5.). Fig. 2.11. shows the same table organized into a binary branching model.

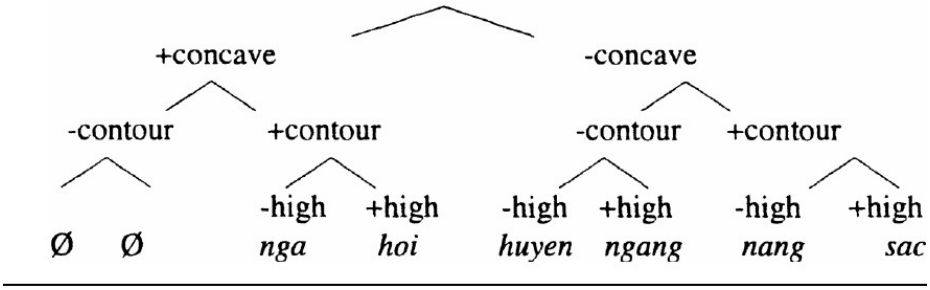


Figure 2.11. *Binary branching model of Vietnamese tones according to Ngô (1984) cited from Phạm (2003)*

Hoàng (1986) presented a similar classification using different categorization: *+-phonation, +-falls into one register* (whether the tone contour is realized within one or two registers), *+-low*. Moreover, she noted that the tones *sắc* and *nặng* in "checked syllables", were manifesting properties different from the open syllables and hence he classified them separately.

Tone	Number	Phonation	One Register	Low
Ngang	1	-	+	-
Huyền	2	+	+	+
Ngã	3	+	-	-
Hỏi	4	-	-	+
Sắc	5	-	-	-
Nặng	6	+	-	+
Sắc1	5'	+	+	-
Nặng1	6'	+	+	+

Table 2.6. *Binary classification of Vietnamese tones according to Hoàng (1986)*

As opposed to open syllables and syllables with final nasals that can carry any of the six tones, checked syllables, i.e. syllables closed by voiceless stops, can only carry *sắc* or *nặng*. Older studies (Đoàn 1977, Thompson 1965, Ngô 1984) considered what Hoàng labelled *sắc*, *nặng* and *sắc1, nặng1* allotones in complementary distribution and paid no scientific attention to the dissimilarities between them. However, researchers have recently agreed on giving the checked syllables specific treatment as well as on acknowledging the importance of voice quality for the process of tonal discrimination. Didactic texts, nonetheless, still strictly follow the pitch-based, 6-tone classification, which is understandable as a certain degree of simplification is vital for making learning processes more effective. The purpose of language education, after all, is to acquire a tool of communication rather than then to comprehend the system in its full complexity.

Phạm (2003) developed Hoàng's work and further challenged the traditional approach by claiming that pitch contour was merely one of the distinctive features and that there were tones impossible to distinguish based plainly on pitch contour. She presented a hypothesis that the realizations of *sắc* and *nặng* in syllables closed by plosives should not be treated as variants but as separate tones. Furthermore, she decided to investigate spectral

characteristics of phonation types such as creakiness and breathiness within Vietnamese syllables and relation to tonal recognition.

Phạm recorded 9 subjects (three males and six females, age span 17-49, living in Toronto, originally from Northern Vietnam) to gather desired speech material consisting of a set of CV and CVC. The vowel /a/ was the only sound used as V. Initial consonants included the nasal /m/, stops /t/, /k/ and fricatives /s/, /z/. Velar nasal and stop were selected among final consonants. The syllables were recorded in fixed elicitation frame: *chữ SYL phải ngay* (the word SYL must be straight). The author claimed that there was no indication for tonal sandhi or coarticulation and, therefore, the item in the preceding context had no impact on SYL and the item in the following context was deliberately chosen to begin with /f/ in order to create a clear phonetic boundary. Phạm decided to measure the F_0 manually every 30ms with the first measurement immediately after the "burst of the initial segment". (Phạm 2003: 36) She decided to incorporated nasal finals into the syllabic core because "the second part of the tone was realized during the nasal. Although voicing was not strong in the nasal portion of the spectrogram, F_0 was still measurable. If this part was ignored, the tone lost its second part and became unusually short. In such cases there was not enough information about the contour of the tone and the tone became unrecognizable." (Phạm 2003:37) The first step of Phạm's classification was to evaluate tones based only on F_0 employing the dichotomy marked/unmarked as illustrated by tab. 2.7. below.

	Even	***Non-Even***	***Non-Even***	***Non-Even***
		rise/fall	*rise/fall*	*Curve*
Unmarked	ngang	sắc1	sắc2	hỏi
Marked	huyền	nặng1	nặng2	Ngã

Table 2.7. *Classification of Hanoian tones based on F0 based on Phạm (2003)*

Subsequently, spectrograms were subjected to visual inspection in order to assess phonation types. Phạm looked for signs of modal, breathy and creaky voice.

ngang	modal voice (periodic, regular glottal pulses and moderate amplitude)
huyền	breathy voice (regular glottal pulses and reduced amplitude)
sắc 1	modal voice (periodic, regular glottal pulses and moderate amplitude)
nặng 1	glottal stop or creaky portion close to the end of the tone (irregular widely spaced pulses, sometimes interrupted by one or two irregular pulses; reduced amplitude; complete closure of the vocal folds results in glottal stop and incomplete closure results in creakiness)
hỏi	breathy after approximately 40 ms, breathiest from 70 to 120 ms (regular pulses and reduced amplitude)
ngã	Glottal stop or creakiness in the middle of the tone (irregular widely spaced pulses from about 70 to 130 ms, and gaps in the spectrogram)
sắc 2	modal voice (periodic, regular glottal pulses and moderate amplitude), but very short because of the final stop, ending before 120 ms
nặng 2	Some breathiness after 60 ms (regular pulses and reduced amplitude), but also very short because of the final stop, ending before 120 ms. The F0 of *nặng 2* is very close to that of *huyền*.

Table 2.8. *Phonation types manifested in North Vietnamese tones according to Phạm (2003:46)*

By combining the information from Tab. 2.7 and Tab. 2.8, Phạm (2003) came up with classification reflecting both F_0 contour as well as the findings regarding phonation types.

Tone	**ngang**	**huyền**	**sắc1**	**nặng1**	**hỏi**	**ngã**	**sắc2**	**nặng2**
Contour	level	level	rise	fall	curve	curve	rise	fall
Phonation	modal	breathy	modal	creaky	(breathy)	creaky	modal(obst)	breathy(obst)
Height	H	L	H	L	L	L	H	L

Table 2.9. *Classification of North Vietnamese tones reflecting F0 as well as phonation types according to Phạm (2003). The parentheses in hỏi signalize that breathiness is not present in all realizations of the tone or that it is only present in one section of the realization. Phonation of sắc2 and nặng2 is marked (obst) because the syllables finish in voiceless obstruents.*

There were a few places where Phạm disagreed with the traditional approach on tonal classification advocated by Đoàn (1977). Despite admitting herself that "phonetically, it may fall a little" (Phạm 2003: 71), she classified *huyền* as a level tone and put it in opposition of *ngang* that is higher and uttered with modal voice. Furthermore, she claimed that "breathiness predicts lowness" (Phạm 2003: 59), which can be manifested by the fact that all breathy tones in Vietnamese are low. She classified *ngã* as a low tone because although *ngã* ends and often also begins higher than *sắc*, in her study, its average F_0 was lower than the F_0 of *sắc*.

Marchand (2004) introduced an alphanumeric classification grouping the tones into pairs based on shared features and the notion of markedness introduced by Phạm (see tab. 2.7.). *Ngang* was labelled A1 and *huyền* A2 where A1 belongs to the high register and A2 to the low one, both of them lacking contour (although A2 has been classified as falling by some). *Sắc* bears the label B1 and *nặng* B2. Tones B1 and B2 also belong to opposite registers. B1 has a rising contour while B2 is heavily glottalized at the end resulting in rapid decrease in F_0 in the final section of the syllable. *Hỏi* and *ngã* were marked C1 and C2 based on their complex contour.

Drawing inspiration from Phạm (2003) and Marchand (2004), Brunelle (2009a) formulated the following classification:

> The first tone is called ngang (or A1). It is level, a little higher than the mid-range and has a modal voice quality. The tone sắc (or B1) is also modal but its pitch starts at mid-range and rises rapidly. The tone huyền (or A2), which can be modal or breathy, starts relatively low and falls smoothly. Nặng (or B2) is also falling, but typically shorter than the other tones and ends on a glottal stop or at the very least a strong glottalization. The last two tones have a light mid-laryngealization (or creakiness). Ngã (or C2) starts on a fall, is interrupted by a

glottalization that range from strong laryngealization to a full glottal stop and ends on a dramatic rise. The tone hỏi (or C1), on the other hand, falls dramatically until it reaches a turning point where it is accompanied by a slight laryngealization (breathy voice has also been reported) and then rises slightly, at least in very formal speech. In colloquial Hanoi speech, hỏi (C1) has lost its final rise. It is now a low falling tone that is shorter than its formal counterpart and ends on a mild laryngealization. Unfortunately, the variant is understudied because it is difficult to elicit in recording sessions. (Brunelle, 2009a)

label	Name	characteristics	contour	Diacritics
A1	*Ngang*	level, higher	33	A
A2	*huyền*	falling, breathy	21	À
B1	*sắc*	rising, tense	35	Á
B2	*nặng*	falling, glottalized	31ʔ	ạ
C1	*hỏi*	fall-rise	313	ả
C2	*Ngã*	fall-rise, broken	3ʔ5	Ã

Table 2.10. *Lexical tones in Hanoian Vietnamese based on Brunelle (2009a)*

Brunelle's classification is very detailed. As opposed to Phạm, who categorized huyền (A2) as a low level tones, he classified both huyền (A2) and nặng (B2) as falling tones. He also mentioned the tendency to drop the final rise in hỏi (C1) in colloquial Hanoi speech, which is a phenomenon that might lead to greater confusion of huyền (A2) and hỏi (C1) in the future. Brunnelle also accepted the traditional classification of Vietnamese tones distinguishing only six of them. Phạm's argument for adding two additional tones into the classification was not reflected until Kirby (2011).

Kirby (2011) graphically represented F_0 contours of the six tones occurring in open or sonorant-closed syllables and then added the categories D1 and D2 for what Phạm (2003) called *sắc2* and *nặng2*. Fig. 2.12. clearly shows that the contours of especially B1 and D1

are apparently different and they deserve being treated separately as two different tones. Moreover, Kirby's representation illustrates that the duration of B2 and D2 is significantly shorter than duration of the other tones. However, this representation measures only the F_0 curve but the creaky or glottalized final section of the tone where F_0 is distorted or missing completely should still be included into the duration of the tone as it is one of the cues necessary for tone recognition.

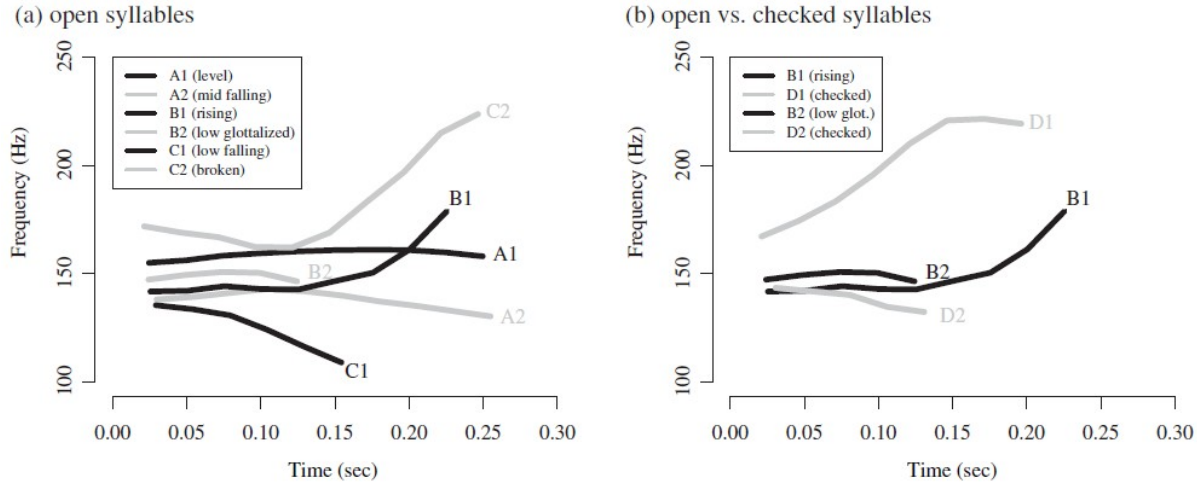


Figure 2.12. *F_0 contours for a male speaker of Hanoian Vietnamese. Section (a): six tones in open and sonorant-closed syllables. Section (b) compares contours of sắc and nặng in open and checked syllables (Kirby, 2011).*

2.2.2.5. Reduplication and borrowing

Unlike many tone languages, Vietnamese has no phonological processes like tonal assimilation or sandhi and, therefore, reduplication and borrowing are the two valuable processes that offer an insight into the phonology of Vietnamese tones. They are also two key factors indicating tone markedness. Phạm (2003) offered the following markedness-based categorization of Vietnamese tones:

ngang-huyền<sắc-nặng<hỏi-ngã

We can see the tones ordered in pairs from the least marked tone (*ngang*) to the most marked (*ngã*). *Ngang* is the tone with modal voice and least complicated contour, *huyền* is breathy but still level (provided we accept the low-level classification over the falling contour), *ngã* is on the other side of the spectrum as the most marked tone with the most complex contour and prominent creakiness. Moreover, *ngã* is merged with *hỏi* in many regional dialects. According to Phạm, markedness is supported by distribution (although she did not offer any specific figures), i.e. *ngang* occurs most frequently whereas *ngã* most rarely; reduplication patterns and borrowing mechanisms.

Reduplication in the Vietnamese language is a process when a base syllable already bearing a full lexical meaning (usually adjectival or adverbial) is accompanied by a reduplicant syllable in order to increase stylistic register and/or intensity of the lexeme. The reduplicant syllable begins with the same phoneme as the base and Phạm (2003) claimed that marked tones always reduplicate as marked tones, e.g. lạnh-cold (nặng) into lạnh lẽo-very cold (nặng, ngã); whereas unmarked tones reduplicate as umarked tones, e.g. láu-cunning (sắc) láu lỉnh-very cunning (sắc, sắc).

It is useful to differentiate between two types of borrowing processes in Vietnamese: from Chinese and from other, mostly non-tonal, languages. Borrowing from Chinese was extremely productive in the past, and, as a consequence around 60% of Vietnamese vocabulary is of Chinese origin (Trần Trí Dõi 2011). In times prior to WWII, Chinese characters were used in official written communication in Vietnam. All existing Chinese characters have a Vietnamese reading similarly to Korean and especially Japanese where kanji is still an integral part of the writing system. The Chinese characters 汉字 (pinyin: hàn zì) literarily meaning “Han” (largest ethnic group in China) and “character” can be read as “Han ja” in Korean, “Kan ji” in Japanese or “Hán tự” in Vietnamese. New concepts and ideas used to come to Vietnam mainly from China and mostly in written form hence the

Vietnamese would simply adopt the Vietnamese reading of a Chinese lexeme. For instance, the word “socialism”, 社会主义 (shè huì zhǔ yì) in Chinese, was borrowed into Vietnamese as xã hội chủ nghĩa. This type of lexical borrowing is interesting from the perspective of tracking phonemic and tonal changes between Chinese and Vietnamese but does not reveal much in respect to the phonology of Vietnamese tone as such.

Since the mid-19th Century, Vietnamese speakers have been subjected to significant influence of two non-tonal Indo-European languages. First, it was the French during the colonial era ending in 1954, and later on English in the 1960s and 70s in the South. Since the 1990s the influence of English has been omnipresent in all strata of the Vietnamese society. It is still rather early to draw any conclusions but it seems that the most recent borrowings from English are being adopted in the original spelling and there is no effort to adjust them to the Vietnamese orthography. In regard to French borrowings from the colonial era, it is a very different situation as can be seen in tab. 2.11. below.

Vietnamese	French	English
cà phê	café	Coffee
ăng ten	antenne	Antenna
ba lô	ballot	Backpack
băng công	balcon	Balcony
bê tông	béton	Concrete
bi da	billard	Billiards
cao su	caoutchouc	Rubber
Ga	gare	train station
Gôn	golf	Golf
ni long	nylon	Nylon
pa tê	paté	Pate
Phanh	frein	Brake
ra đi ô	radio	Radio
sơ mi	chemise	Shirt
tắc xi	taxi	Taxi
xăng đan	sandale	Sandals
xi măng	cimant	Cement

Table 2.11. *Examples of French borrowings in Vietnamese*.

Similarly to Chinese in the more distant past, French became the source of vocabulary for new concepts entering Vietnam towards the end of the 19th Century. We can see that all the words listed in tab. 2.11. were transformed to fit Vietnamese orthographic and phonotactic rules. Moreover, it is apparent that most of the Vietnamese syllables carry the least marked tone *ngang*, borrowed syllables with a final plosive carry the tone *sắc2* (D1) and very rarely there are instances of *huyền*. *Ngang*, *huyền* and *sắc* are considered the least marked tones and the fact that they are used so frequently in borrowed lexemes only supports the theory.

2.2.2.6. Coarticulation

Coarticulation of North Vietnamese tones seems to be bi-directional with dominant progressive effects (Brunelle, 2009a). As discrimination of North Vietnamese tones does not rely solely on pitch contour, there is more space for variation because it does not affect tone recognition to such an extent. "Both the height and slope of tones are affected by their tonal context, but the relative position of each tone in the tonal space is overall stable." (Brunelle 2009a: 50). Coarticulation in Vietnamese seems to lead to no phonological changes although Brunelle argued in his recent study (2016) that *sắc* and *huyền* before *ngang* can often be confused. He sees the cause of this confusion in the fact that the peak of *sắc* is normally delayed onto the initial portion of the following tone. This peak delay, however, lacks acoustic and perceptual salience when preceding *ngang*. As a result, *sắc* before *ngang* can be at times perceived as *huyền*. Brunelle concluded that although this phenomenon could lead to a possible emergence of tonal sandhi in the future, it does not have a phonological effect in the current state of Vietnamese.

2.2.2.7. Intonation

Brunelle et al. (2012) admitted the existence of sentential intonation, however, it seems to be more prominent in faster or emotionally affected speech whereas in slow and careful speech

such as reading or TV and radio broadcasting, sentential intonation is clearly overridden by lexical intonation.

> Declarative sentences are found to have a slight overall f0 declination. Interrogatives are described as having a high overall range, or a high range and a rise starting much before the sentence final question marker. (...) Imperatives are also described as having a high overall f0, possibly with an additional final rise and longer duration. (Brunelle 2012:6)

A possible reason for lower salience of sentential intonation in Vietnamese relative to non-tonal languages is its relative redundancy. The Vietnamese language employs final particles in marking communicative functions: particles *hả* or *à* mark yes/no questions (*Người đó là sinh viên của anh **à**?* Is that person your student?); *đi* or *nhé* mark the imperative (*Đóng cửa **đi**!* Close the door!); *nhỉ* marks expected agreement (*Quyển sách này hay **nhỉ**?* This book is interesting, isn't it?); *chứ* or *mà* mark contradiction (*Đây là chị tôi **mà**!* This IS my older sister!) etc. We can see that many functions (including pragmatic ones) expressed through intonation in non-tonal languages are expressed lexically in Vietnamese, which renders sentential intonation in Vietnamese less significant. Nonetheless, Brunelle (et al. 2012) concluded that there are measurable differences in Vietnamese sentential intonation. The core of the issue lies in the fact that despite being measurable, they still seem not to be prominent enough from the perspective of human perception and hence insignificant for real-world communication.

2.2.2.8. Stress

Vietnamese is largely a monosyllabic and syllable-timed (Cunningham 2009) language and majority of syllables represent whole semantic lexemes. The placement of lexical stress in

monosyllabic lexemes is rather straightforward. Nevertheless, the amount of disyllabic lexemes is not negligible, especially due to Sino-Vietnamese vocabulary. Stress placement in disyllabic lexemes is not strictly defined but there seems to be a tendency to place the stress on the second syllable: *sinh* ***viên*** – student; *bưu* ***điện*** – post office; although lexemes with stress on the first syllable also exist: ***tham*** *gia* – take part in; ***tổ*** *chức* – organize (Slavická 2008:66). Lexemes with more than two syllables are rare. There are structures such as *người bán hàng* (a vendor) that translate as one lexeme into most Indo-European languages but it is highly questionable whether they should be treated as one lexeme also in Vietnamese. *Người* is a classifier used for human beings, *bán* is the verb "to sell" and *hàng* means "goods". On the other hand, if we look at a similar lexeme in the English language, "salesman", we can notice a clear parallel between the combining form "man" and the classifier *người.* However, classification of Vietnamese lexical units still has not been universally agreed on and as Vietnamese is generally considered a monosyllabic language, the structures *người bán hang* tend to be treated as three separate lexical units.

Sentential stress is dealt with in Cao Xuân Hạo (2007). Cao claimed that sentential stress is used for distinguishing individual syntagmata as it is often difficult to establish which elements in the sentence belong together due to lack of overt word-class markers or inflection. Cao uses the example of the sentence: ***Lan***//*đi* <u>*mua*</u> ***cá***// *mí lị* ***khế***// <u>*về*</u> <u>*nấu*</u> ***canh***. (**Lan** went to <u>buy</u> **fish** as well as **star fruit** then she <u>returned home</u> to <u>make</u> **soup**.). The individual syntagmata are separated by (//), syllables bearing primary stress are marked bold and syllables with secondary stress are underlined.

Thompson (1965: 106-107) distinguished three types of stress: Heavy stress – singles out the syllable or syllables of each utterance which carry the heaviest burden of conveying information. Weak stress – accompanies syllables which bear the lowest information

conveying load. They often refer to things which have been brought up earlier or which are expectable in the general context. Other syllables are accompanied by medium stress.

Chaudhary (1983) identified intensity as an acoustic correlate of stress in Vietnamese. Nguyễn & Ingram (2006) arrived at a conclusion that duration is closely related to stress as well. We can see in Cao (2007) above that stress is also conditioned by position in the sentence. Sentence or phrase-final syllables tend to the most prominent. Sentential stress is usually not carried by certain grammatical units (pronouns, prepositions, classifiers) but it is by others (vocative elements, temporal elements, intensifiers). In short sentences and phrases, all constituents can be stressed e.g. ***người cao*** (tall people//people are tall); ***chó chạy*** (dogs run). Some disyllabic units can have both syllables stressed such as ***vợ chồng*** (married couple). *Vợ* means "wife" and *chồng* means "husband" when the two words stand alone, which is the reason why both constituents are stressed, whereas in the case of *sinh* ***viên*** (student) the two syllables cannot stand independently, hence the single stress on the second syllable.

Vietnamese, as opposed to Chinese, does not possess atonal syllables. Unstressed syllables in Vietnamese tend to be shorter and the tone contour might not be canonical but the tone is still present. Chinese atonal syllables lost the tone due to their low prominence, which is a tendency that does not occur in Vietnamese.

2.2.3. Saigonese dialect

Up until 1955, when South Vietnam won independence from France, there had never been any reason for perceiving the dialect spoken in the area of the current HCM City as anything other than a regional dialect. The city of Saigon in modern proportions only came to

existence in 1956 by merging the town Sài Gòn surrounding the Gia Định citadel with the Chợ Lớn market and it immediately became the capital city of the Republic of Vietnam, the capitalist and anti-communist state used by the Americans as a tool to fight Communism in South-East Asia. Although the Republic of Vietnam was defeated in 1975 and reunited with Northern Vietnam in 1976, the two decades of its existence were linguistically significant. American military presence in the Republic of Vietnam was strong, especially between 1967-1969, when it hosted more than a million American citizens. Despite the language policy focused mainly on English education of the local population, there was also need for Americans to learn Vietnamese be it for the purposes of the military, secret service, news agencies, charities or religion. In order to teach Vietnamese as a second language, textbooks had to be designed and standard Southern Vietnamese had to be defined. After the reunification in 1976, however, Saigon was renamed to Ho Chi Minh City, public offices as well as media were dominated mostly by Northerners, and North Vietnamese was imposed as the national standard with very little space for regional dialects in the public spheres. This tendency began to change after the economic reforms of Đổi Mới in 1986 and nowadays, it is very common to hear Southern Vietnamese in media, movies and public speeches.

Another reason for the relative inability to define the Saigonese standard dialect can be spotted in its variability. It is as hard to define a truly Saigonese accent as it is to find someone actually speaking it. The considerable variability of the Saigonese accent might be caused by a much larger scale of migration in the area. Whereas Hanoi has been the centre of the Vietnamese (or Việt) culture for more than a millennium, Saigon fully emerged only in 1956 as mentioned above. Moreover, there were two huge migration waves spilling over the area of modern HCM City. The first occurred after the Geneva Accords of 1954 were signed and over a million people, more than 60% of whom were Catholic (Hansen 2009, Picard 2016), feared repercussions in the North and therefore fled to the South. Moreover,

Ngô Đinh Diệm, the prime minister and later the president of the Republic of Vietnam, was a devout Catholic and he actively invited followers of Catholicism to the South. Although Hansen (2009) rejected the theory that Diệm invited Catholics from the North into the suburbs of Saigon in order to construct a "ring of steel", i.e., to surround Saigon by his affiliates to stabilize the region politically, it remains a fact that majority of the Northern refugees at that time resettled in the area of current HCM City. The second immigration wave came after the fall of Saigon in 1975 when members of the North Vietnamese Army as well as followers of the Communist ideology permanently settled in the area to consolidate it after the war. The two migration waves together with the general animosity towards anything reminiscent of the Republic of Vietnam seem to constitute persuasive causes of the rather wide variation of phonetic features in the Saigonese dialect.

From the academic perspective, Saigonese dialect is understudied in comparison with the dialect of Hanoi. Between the mid-70s and late 90s, Saigonese dialect fell entirely out of the scope of scholarly interest. Even nowadays, the most comprehensive description of the Saigonese dialect remains in Thompson (1965) and a few PhD theses from right after the war like Vũ (1982).

2.2.3.1. Phonemes

The Saigonese dialect differentiates between the Hanoian homophones x|s; r|d|gi and ch|tr. Unlike the Hanoian dialect which does not distinguish between the pronunciation of *x* and *s* (both are pronounced [s]), the Saigonese dialect still retains a difference in pronunciation since *x* is also pronounced [s] but the grapheme *s* is pronounced [ʃ]. *Ch* is pronounced [tʃ] similarly to Hanoi but *tr* is "a retroflex stop formed by touching the underside of the tip of the tongue against the alveolar ridge; it is usually slightly affricated – that is, released with a

very short spirant." (Thompson 1965: 89) Graphemes *d* and *gi* are pronounced [j] and there is no [z] sound in the Saigonese dialect whatsoever. The grapheme *r* has the same place of articulation as *tr* but more than one pronunciation may be found even within a single speaker. It may occur as a retroflex fricative [ʐ̻], an alveolar approximant [ɹ], a flap [ɾ], a trill [r], or a fricative tap/trill [ɾ̝, r̝]. The grapheme *v* is also often pronounced as [j], which is a typical and frequently occurring feature of the Saigonese dialects but it is considered somewhat undesirable or not prestigious by the speakers themselves. Among the syllable-final consonants, there is homophony in the Saigonese dialect where there is none in the Hanoian dialect. Graphemes *n* and *nh* pronounced [n] and [ɲ] in Hanoi are both pronounced [n] in Saigon. Similarly, *t* and *c* pronounced [t] and [k] in Hanoi are both pronounced [k] in Saigonese syllabic codas.

Vowels in the Saigonese dialect may shift slightly in quality. However, there are no instances or splits or mergers or any phenomena that could influence tonality. As opposed to the Hanoian dialect that often merges the diphthongs /i͡ɣ̞/ and /ɯ͡ɣ̞/ into /i͡ɣ̞/, Saigonese speakers retain the distinction and their pronunciation of /ɯ͡ɣ̞/. On the other hand, as we can see in tab. (x) below, the dipthongs /i͡ɣ̞/ and /ɯ͡ɣ̞/ tend to be reduced to /i/ and /ɯ/.

Grapheme(s)	Standard Vietnamese	Southern dialect
i	/i/	/i/
y	/i/	/i/
ê	/e/	/e/,/ɤ/
e	/ɛ/	/ɛ/
ư	/ɯ/	/ɯ/
u	/u/	/u/
ô	/o/	/o/
o	/ɔ/	/ɔ/
ơ	/ɤ/	/ɤ/
a	/a/ or /ɑ/	/a/ or /ɑ/
â	/ɤ̆/	/ɤ̆/
ă, a	/ă/	/ă/
iê	/ie/	/i/
yê	/ie/	/i/
ia	/ie/	/ie/
ya	/ie/	/ie/
uô	/uo/	/u/
ua	/uo/	/uo/
ươ	/ɯɤ/	/ɯ/
ưa	/ɯɤ/	/ɯɤ/

Table 2.12. *Vietnamese vowel realizations in the standard dialect compared to the realizations in the Saigonese dialect with differences marked in grey (Phạm & McLeod 2016).*

2.2.3.2. Syllable, stress and intonation

The syllable structure in the Saigonese dialect does not differ from Hanoi. However, "there seems to be generally a much more pronounced difference in intensity or loudness between heavy and medium stresses; and weak stress is accompanied by very short syllables. This gives the typical conversational language a much more syncopated rhythmic impression than Hanoi speech." (Thompson 1965: 93)

2.2.3.3. Tones

Thompson (1965), Hoàng Thị Châu (1989) as well as Brunelle (2009b) and Kirby (2009) agree that the Saigonese dialect employs only five tones because *hỏi* and *ngã* collapse into

one tone that acoustically resembles Hanoian *hỏi* more than *ngã*. Standard orthography in HCMC remains unaltered and so this phenomenon is not reflected in writing with the exception of spelling errors. Language users with lower education tend to spell syllables containing *ngã* according to the orthographic standard with *hỏi* or just generally confuse the two tones in lexemes like **hẽm* instead of *hẻm* (alley) or *xả* (lemon grass) instead of *xã* (commune).

Thompson (1965: 92) described the fused *hỏi-ngã* tone as one with "a long rising contour beginning in low mid-range and rising sometimes as high as *sắc* tone". *Nặng* was mentioned as "quite low (although not so low as *huyền* most of the time) and level with syllables ending in [p, t, k]; with other syllables it dips slightly, then rises." (Thompson 1965: 92)

Brunelle (2009b) argued that South-Vietnamese tones (based on his data-collection methodology they have the same characteristics as what this thesis calls Saigonese tones) are distinguished either solely on pitch contour or at least the role of voice quality does not play such a significant role in tonal discrimination. Fig. 2.13. shows a tone system comparison of the Hanoian (Northern) and Saigonese (Southern) dialects. The charts indicate that most notable differences between the dialects apart from the fusion of *hỏi* and *ngã* will probably be measured in *sắc* that does not have to compete with *ngã* in the Saigonese dialect and then in *nặng* because of the lack of distinction based on voice quality cues in the Saigonese dialect. Therefore, *nặng* has to differ from *huyền* by difference in pitch instead of glottalization. According to Fig. 2.13. its contour goes lower than the contour of *huyền*.

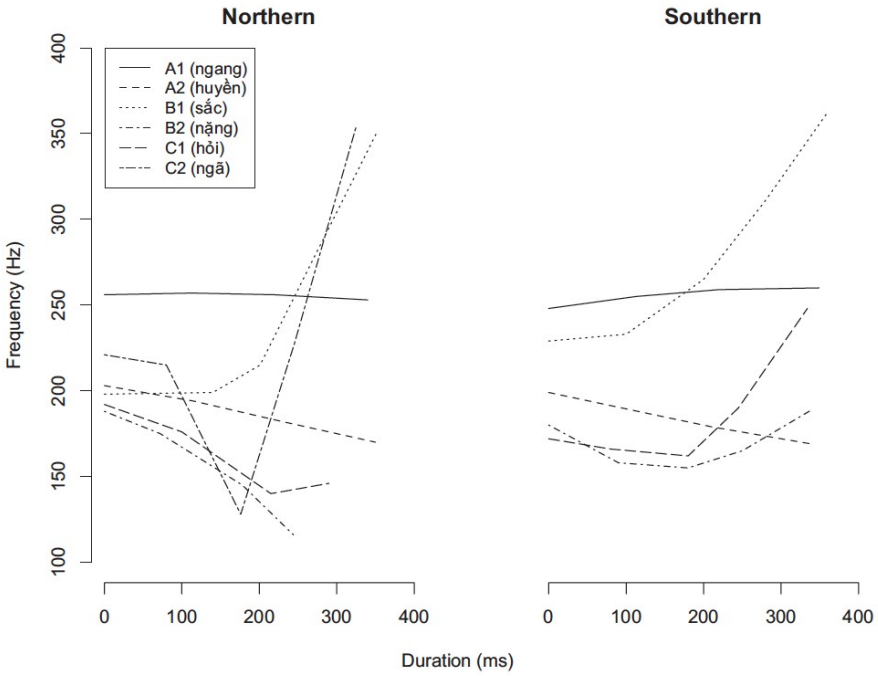


Figure 2.13. *Comparison of NVN and SVN tone systems (Kirby 2010).*

In order to relate Saigonese tones to the Hanoian tones more easily, they can be put into a similar table as Tab. 2.10. describing Hanoian tones in Chao tone letters. Tab. 2.13. lists Chao tone letters as well as other acoustic qualities for the Saigonese dialect based on Kirby (2010), Hoàng (1989) and Hoàng (1986).

Label	Name	Characteristics	Contour	Diacritics
A1	*ngang*	level, higher	44	A
A2	*huyền*	low, breathy	21	À
B1	*sắc*	rising, tense	35	Á
B2	*nặng*	falling with final rise	212	ạ
C1	*hỏi*	fall-rise	214	ả
C2	*ngã*	fall-rise	214	ã

Table 2.13. *Description of Saigonese tones based on Kirby (2010), Hoàng (1989) and Hoàng (1986).*

2.2.3.4. Coarticulation

It was mentioned in the previous section that contrary to Hanoian Vietnamese that uses voice quality as an important cue of tonal discrimination, Saigonese seems to rely fully on pitch contour which has to remain as stable as possible to prevent confusion. Therefore, coarticulation effects in Saigonese seem to be smaller than in the Hanoian dialect, which is also what Brunelle (2009b) inferred.

2.2.4. Tonal development in Vietnamese

The first endeavour to shed light on tonal development in Vietnamese came from Maspero (1912). Although the main object of his scientific interest was Chinese, he managed to come up with revealing discoveries in Vietnamese mainly by means of studying Sino-Vietnamese vocabulary. He grouped the six Vietnamese tones into two sets *ngang-sắc-hỏi* and *huyền-nặng-ngã* claiming that the first set evolved from ancient voiceless onset consonants and the second set evolved from voiced onset consonants. More details on the topic of interaction between voicing and F_0 can be seen in Fig. 2.2. Maspero's ideas were later developed by Haudricourt (1954), who claimed that Vietnamese was a non-tonal language before the 6th Century A.D. when it established three tones roughly corresponding to current *ngang*, *huyền* and *sắc*. *Ngang* appeared in open syllables, *huyền* in syllables that lost final fricatives [s] and [h] and *sắc* in syllables that lost final [x] and glottal stop. In the 12th Century, Vietnamese supposedly lost voicing of initial consonants, which gave rise to the tones *nặng*, *hỏi* and *ngã*. In modern Vietnamese, voicing contrast was re-established in all plosives but in /p/ x /b/ although /p/ exists in certain borrowings from French. Tonal development in Vietnamese demonstrated on syllables *ba* and *pa* can be observed in Tab. 2.14. below.

Original	6th Century	12th Century	Modern
Pa	pa	pa	ba
Ba	ba	pà	bà
pas/pah	pà	pả	bả
bas/bah	bà	pã	bã
pax/paʔ	pá	pá	bá
bax/baʔ	bá	pạ	bạ

Table 2.14. *Reconstruction of tonal development in Vietnamese according to Hadricourt (1954).*

Hadricourt did not address the process by means of which *hỏi* and *ngã* evolved from *huyền* but his theory is nonetheless valued as the most accurate explanation of Vietnamese tones development up to now. Hoàng Thị Châu (1989) claimed that the disappearing consonant in fact left traces on the tones in form of phonation types. Final glottal stops supposedly left glottalization in *sắc* and *nặng* and the disappearance of final fricatives left traces of breathiness and creakiness in *ngã* and *hỏi*. Phạm (2003) argued that although it seemed likely that the loss of glottal stop left traces of glottalization in *nặng*, there was no glottalization in *sắc*. Similarly, fricatives leaving behind breathiness as in *hỏi* seemed quite logical but why it turned into creakiness in *ngã* was disclosed neither by Hadricourt nor by Hoàng Thị Châu. Ferlus (2004) pointed to the fact that some modern Việt-Mường languages such as Khmu contain tones as well as final glottal stops, and mentioned Diffloth's (1989) hypothesis that there was no glottal stop in Proto-Vietic but there was a contrast between modal and creaky voice where present day *sắc* and *nặng* reflect the creaky voice whereas *ngang* and *huyền* modal voice. Ferlus researched this hypothesis further and altered it by claiming that there in fact was a glottal stop in Proto-Vietic but it was lost only in sesquisyllables that later on also disappeared from Old Vietnamese under the influence of Old Chinese but remained in other Việt-Mường languages that were not affected by Old Chinese as directly as Old Vietnamese was.

2.2.5. Tonal interference into other languages

Similarly to the speakers of non-tonal languages attempting to learn Vietnamese, Vietnamese native speakers struggle when acquiring phonetic systems of non-tonal languages. They cannot get rid of lexical intonation in a foreign language and tend to pronounce all open and nasal-closed syllables with the *ngang* tone and all closed syllables with *sắc*. Moreover, they simplify consonant clusters, drop final consonants (or pronounce them unreleased) and shorten vowels in closed syllables. Therefore, the sentence *I would like to speak good English* [aɪ wəd laɪk tə spiːk gʊd ˈɪŋglɪʃ] would probably be pronounced like *ai út lai tơ píc gút ing lít* [ai(44) ut(35) lai(44) tɤ(44) pik(35) gut(35) iŋ(44) lit(35)]. Tonality in Vietnamese can be used in the learning process of sentential intonation in non-tonal languages. For instance, if Vietnamese learners of English end their yes-no questions with a *sắc* syllable and statements with a *huyền* syllable, it renders the utterances very easy to distinguish.

3. Methodology

In order to carry out a reliable and accurate comparison of the Hanoian and Saigonese tonal systems, it was crucial to collect data that would not be compromised in terms of 1) speaker selection, 2) acoustically in the recording stage or 3) during the processing stage. Therefore,

it was essential to devise dependable methodology to address all three issues and ensure data reliability. The methodology is based largely on the findings of Thomas (2002), McGuire (2010) and Niebuhr & Michaud (2015) who addressed the topics of phonetic field research – particularly speaker selection, recording conditions and linguistic use of perception tests – in substantial depth.

3.1. Speaker selection

With respect to Niebuhr & Michaud (2015), attention had to be paid to careful selection of suitable speakers based on clearly defined criteria concerning age, gender and family background so that each group of speakers would be as homogenous as possible. At first, all subjects were given questionnaires where they were asked to fill in the necessary information regarding the criteria in question in order to obtain a sound set of metadata. Niebuhr & Michaud (2015) also stated that the sample size should be at least 10 speakers in order to assess individual differences within the group regarding gender or dialect. The aim was to record 12 speakers of each dialect, ideally equally divided in terms of gender. The speakers were supposed to be born into and grow up with the dialect without any long-term periods of living away from it. Their parents were ideally also supposed to spend their life within the dialect region or in its close vicinity. The age group was set to late teens to early thirties because younger speakers could still be in process of language development, and with older speakers there is the danger of their language being obsolete. Furthermore, anonymity of all subjects was preserved and they were assured that all the collected information would be used for research purposes only and no third parties would be allowed to access it.

For the purposes of the perception test, only one male and one female speaker was recorded for every dialect totalling 4 speakers. We applied the same ethical framework as with the 24 speakers recorded in the tasks mentioned above.

3.1.1. Syllables, text and semi-spontaneous speech

Twelve inhabitants of Hanoi and the same amount in the Ho Chi Minh City were recorded for the purposes of this analysis. As we only managed to record 5 male subjects in Hanoi, we decided to set the ratio to 5 males to 7 females for both dialects. All subjects were born between 1985 and 1997. Subjects recorded in Hanoi grew up and spent most of their lives in Hanoi and subjects recorded in Ho Chi Minh City were born and spent most of their lives there with the exception of two who spent a considerable part of their life in Đồng Nai a province just behind the eastern rim of HCMC. All subjects were asked to provide information about their family background and it was an important aspect of their selection that their parents would be if not from Hanoi and Saigon respectively, then at least from the North or the South. This condition was especially significant for the speakers of the Saigonese dialect for the reasons described in 2.2.3. Recording subjects who grew up in a family environment linguistically different from the regional dialect might negatively influence the gathered data. Thomas (2002) mentioned the fact that reading fluency varies among speakers, which might potentially compromise the results. On the other hand, excluding speakers who are not fluent readers could distort linguistic reality especially in combination with another factor – choosing university students as subjects for recording experimental data based on their easy accessibility. Furthermore, Thomas claimed that university students are not ideal subjects especially for experiments distinguishing between dialects as they are frequently exposed to outside influences and they tend to lose touch with their vernacular. In Vietnamese, however, variation even within one dialect can be

extremely broad. University students therefore guarantee a reasonable consistency and homogeneity of the data. In section 2.2.2.1., we described a speech impediment frequently occurring among poorly educated inhabitants of the Hanoi region, who are unable to distinguish between initial /l/ and /n/. Section 2.2.3. mentioned the generally higher variability of the Saigonese accent. Working with university students should filter out these two issues. Tabs 3.1. and 3.2. introduce the recorded speakers from both Hanoi and HCMC in terms of gender, age and regional origin.

Code	Gender	DoB	Origin	Education	Comments
HN1	M	1992	Hanoi	Uni	
HN2	F	1990	Hanoi	Uni	
HN3	F	1987	Hanoi	Uni	
HN4	M	1993	Hanoi	Uni	
HN5	F	1993	Hanoi	Uni	
HN6	F	1990	Hanoi	Uni	
HN7	M	1988	Hanoi	Uni	
HN8	M	1996	Hanoi	Uni	2 years in Germany
HN9	F	1994	Hanoi	Uni	
HN10	F	1995	Hanoi	Uni	
HN11	F	1993	Hanoi	Uni	
HN12	M	1994	Hanoi	Uni	

Table 3.1. *Overview of the recorded Hanoian speakers. (DoB stands for "date of birth")*

3.1.2. Perception test speaker selection

Two speakers (male and female) were recorded for each dialect. They were all in their thirties and the criteria of choice were identical to the section above. As a secondary selection cue we were looking for subjects who were identified by their surroundings as speakers of stereotypically Hanoian or Saigonese accent.

Code	Gender	DoB	Origin	Education	Comments
SG1	F	1985	HCM	Uni	mom from C. High.

SG2	F	1995	Đồng Nai	Secondary	
SG3	M	1992	HCM	Secondary	
SG4	F	1990	HCM	Uni	
SG5	F	1988	HCM	Uni	
SG6	M	1993	HCM	Uni	
SG7	M	1992	HCM	College	
SG8	F	1995	Đồng Nai	Secondary	family from C. VN
SG9	F	1988	HCM	Uni	6 years in Finland
SG10	F	1900	HCM	Uni	
SG11	M	1992	HCM	Secondary	
SG12	M	1987	HCM	Uni	

Table 3.2. *Overview of the recorded Saigonese speakers. Two of the respondents were not directly from HCMC but from a province 30 km north-east of HCMC. Parents of one of the speakers came from Central Vietnam and mom of another was born in the Central Highlands. (DoB stands for "date of birth")*

3.2. Material

Each of the analyses required a separate bulk of recorded material. The first task comparing syllables in context with syllables in isolation used a recording of an artificially created short story containing 8 syllables carrying all the Vietnamese tonal variants in high and low left contexts. The right context was not considered because Brunelle's (2012) study showed that coarticulation in Vietnamese, if any, is mostly progressive. The same syllables were also recorded in isolation by the same speakers prior to the recording of the text. All the syllables have the same CV(C) structure, i.e. six open syllables with an initial nasal consonant and two checked syllables with an initial nasal consonant and final voiceless plosive. It was not feasible to utilize an identical syllable in all tonal realizations as very few Vietnamese syllables carry the full inventory of tones and even if so, they occur in dramatically different contexts hence their incorporation into a coherent story would be problematic. Moreover, as it can be seen in 2.1.3., vowel quality has negligible effect on tonal contour and therefore it

constitutes a variable that should not cause any data distortion. Selection of proper initial consonants seemed to be more vital than vowel quality because the transition from the consonant to the vocalic core affects the F_0 (see 2.1.2.) to an extent that could potentially compromise the analysis if the consonants were not chosen carefully. Relatively high sonority of nasal consonants ensures smoother transition between the segments than, for instance, plosives or fricatives (including their voiced variants). All the selected syllables were positioned so that they would manifest heavy stress as described in 2.2.2.7. and 2.2.3.2. The experimental text can be found in the Appendix 1, the set of syllables chosen for task 1 can be seen in Tab. 3.3.

Tone	**Syll.**	**Tone**	**Syll.**
A1	*ngô*	C1	*mổ*
A2	*mù*	C2	*Ngã*
B1	*má*	D1	*Ngáp*
B2	*nhẹ*	D2	*ngọt*

Table 3.3. *Syllables chosen for the comparison of tones in isolation to tones in context.*

The second task utilizes the reading material recorded in task one and compares it with semi-spontaneous speech. In order to obtain comparable data, we attempted to elicit as many syllables identical to those in the reading material as possible. To do so, the recorded subjects were asked to retell the story in their own words after a five-minute pause when they were finished recording the reading task. This proved to be a rather problematic endeavour as many of the subjects struggled to produce coherent material of reasonable length. In the end, we fortunately managed to obtain at least 30 seconds of recording from all subjects.

The third task employs the syllable [ma] in all tonal realizations uttered by two speakers (1 male and 1 female) from each dialect. Syllable [ma] was selected because it is one of the structurally least complex Vietnamese syllables which, moreover, carries lexical meaning in all tonal realizations, whereas many tonal realizations of other syllables are semantically empty despite being phonotactically plausible.

3.3. Material recording

According to Niebuhr and Michaud (2015), professional recording equipment should be used set to the sampling frequency 44.1 kHz (sampling frequency of CDs) and bit-depth of 16-bit or higher. It can be argued that the sampling frequency 44.1 kHz is meant for music and recording speech can be carried out successfully at a lower rate without any undesirable effect on the data.

All recordings used in this dissertation were captured using the device MEDELI DR2. It has 4 microphones allowing stereo recording, sampling frequency up to 48 000 Hz, MP3 (up to 640kB)/WAV (16/32bit) format and a button to adjust sound input. The format WAV 16bit and sampling frequency 48 000 Hz (that was subsequently resampled to 32 000 Hz) was set as default setting for all the recordings. It might be argued that such high quality recording was unnecessary as the nature of human speech does not allow for using the full potential of the setting but at the current state of technology, there is no longer need for data economization and so it was decided to take this “better safe than sorry” approach to data quality. McGuire (2010) mentioned an important point – normalization of signal amplitude. In order to secure comparability of the recorded data as well as the stimuli used in the perception test, all sounds were normalized to -3 dB.

Niebuhr & Michaud (2015) also discussed the importance of suitable recording space; if a laboratory is unavailable, the recording space should be carefully chosen and optimized by means of reducing open space, background noise, and eliminating reverberant surfaces. Due to the lack of access to a recording lab, the recording for the purposes of this thesis took place in standard residential rooms mostly belonging to the subjects themselves. The rooms were selected depending on whether they were equipped with furniture at least to some extent in order to prevent echoes and other acoustic disruptions that could be caused by empty space. The recording was carried out with the subject sitting in front of a table with a towel or bed sheet spread on it, pillows placed around the recording area and another towel hung in the background also to prevent sound waves from bouncing off hard flat surfaces and hence creating undesirable acoustic effects. The recorder was installed onto a small tripod and tipped slightly so the microphones would face the speaker and then placed on the table roughly 30 centimetres from the pillows and 50 centimetres from the hanging towel. Space in front of the recorder was created for the text sheet so it could rest on the table freely to avoid rustling caused by manipulation with the sheet throughout the recording.

3.3.1. Syllables, reading and semi-spontaneous speech

The speakers were seated approximately 50 centimetres in front of the recorder and they were asked to read a part of the text as a test of sound intensity and sound input was adjusted accordingly. Unfortunately, even after the adjustment the intensity differed to a great extent and this issue had to be subsequently dealt with in the stage of material processing. Firstly, the subjects were asked to read the 8 selected syllables in isolation. Due to the fact that the readings of syllables towards the end tended to be hastier and less clear, the respondents were asked to read the sets of syllables twice in reverse order so we could subsequently handpick the most prominent syllables with most canonical contour. Afterwards they were

given approximately 5 minutes to familiarize themselves with the story and then they were instructed to read it as clearly as possible at their natural pace. If they stumbled or made a reading error, they were free to correct themselves but not forced to do so. When the text recording finished, the subjects were given a break of around 5 minutes to relax or drink water after which they faced their final task – to recall the story from the text and tell it again in their own words. As mentioned above, it constituted a problem to produce a coherent bulk of semi-spontaneous speech data for many of the subjects. However, we managed to gather between 30 seconds to 2 minutes of semi-spontaneous speech from every subject.

There were a few technical difficulties that must be taken in consideration for further recording: Hanoi as well as HCMC are generally very noisy due to the exorbitant number of motorcycles and growing amount of car traffic. Engines create low-pitch noise that the microphones do not pick up but honking can potentially disrupt the recorded speech signal irreversibly. The tropical climate of Vietnam calls for the use of fans or air conditioners in order to keep the room temperature at a bearable level. However, these devices produce humming noise affecting the recording. On the other hand, not using them affects the well-being of the recorded subjects but there were no objections in this respect as everybody was willing to undergo mild physical discomfort in the name of science.

3.3.2. Perception test

Recording of the material used in the perception test was less demanding in terms of the amount of data as well as the number of speakers. Only 2 speakers of each dialect (1 male and 1 female) were used. The syllable MA was recorded in full tonal register in three stress levels in order to find out to what extent the tonal contour canonicity affects the ability of correct tone identification. In order to achieve constant and neutral tonal environment, we

opted for the sentence *tôi xem con dao đen* (I watch a black knife). The three final elements *con dao đen* represent the three stress levels described in 2.2.2.7. and 2.2.3.2. because *con* is a classifier with weak semantic role and it is placed before a noun *dao* (stress level 2) and a phrase-final adjective *đen* (stress level 3). The subjects were asked to read the sentence once and then the 18 variations placing all six tones into the three stress level slots as it can be seen in tab. 3.4.

Tôi	**xem**	**con**	**dao**	**Đen**
		ma	***ma***	***Ma***
		mà	***mà***	***Mà***
		mã	***mã***	***Mã***
		mả	***mả***	***mả***
		má	***má***	***Má***
		mạ	***mạ***	***mạ***

Table 3.4. *Framework for recording the material used to carry out the perception test.*

3.4. Material processing

The recorded data was transformed from stereo to mono preserving the channel with better sound quality, resampled to 32 000 Hz sampling frequency and normalized to -3 dB using SONY Sound Forge and Adobe Audition. The software used for the analyses can only analyze one sound channel at a time, therefore, recording in stereo does not benefit it in any way, the only advantage is that if one of the channels suffers damage, there is still hope that the other will be intact. As it was mentioned above, we experienced difficulties with signal intensity so we had to normalize it, i.e., boost the weak signals and decrease the strong ones. The strong signal was in some places so intensive that it exceeded the caption capacity of the recorder and so the heights were lost beyond remedy. Fortunately, these places were relatively scarce and F0 extraction is in such cases unaffected.

3.4.1. Syllables

The .wav files with syllables read in isolation were opened in PRAAT (Boersma & Weenink 2017) and annotated to text grids (TGs) containing two interval tiers. The bottom tier called *syllable* was used to determine syllabic boundaries and the top tier labelled *core* determined the area of the syllabic core within which the F_0 would be measured. After manual segmentation of the *core* tier, a point tier labelled *tone* was created on the top. Using a PRAAT script, 5 equidistant points were placed into the area of every syllabic core with the last point being at the very end of the syllabic core and the initial point was placed 15ms after the beginning of the syllabic core to avoid F_0 contour fluctuations in the transition between the initial consonant and the vocalic core. Finally, we created pitch tiers (PTs) for all files. The PTs also had to be manually checked in order to correct some of the mistakes miscalculated by the software, especially in areas with heavy glottalization where the F_0 contour can be unclear and lead the software to assuming that the F_0 is unrealistically high. By grouping TGs and PTs we were able to measure F_0 value of the 5 points in the tier *core*.

3.4.2. Processing of reading and semi-spontaneous speech

3.4.2.1. Breath group segmentation in PRAAT

Similarly to the syllable files, we opened the text and semi-spontaneous speech files in PRAAT (Boersma & Weenink 2017), annotated them to single-tier TGs and subdivided them into smaller units labelled *breath groups* (BGs). Ideally, these units should capture speech signal between the speaker's inhalations. However, given that the speakers were not trained orators the division mostly had to be carried out based on the syntactic structure in order to preserve semantic coherence of the text. In the segmentation process of the data from Hanoi, we tried to preserve "natural" BGs but it led to variation in the number of BGs

between the speakers (21 to 36) which later turned out to render the material more difficult to search in. For this reason, it was decided to segment the Saigonese data into 21 "ideal" BGs in order to end up with a clearly organized data set. Having divided the data into BGs, we could begin to employ the force alignment software and segment the data in a more detailed manner.

The procedure of processing the semi-spontaneous speech data was identical to the text data with the exception that the files were smaller and the content varied. Therefore, it was not helpful to use the force alignment software and all BGs were further segmented manually.

3.4.2.2. Forced alignment

A specialized software Prague Labeller (Pollák, Volín & Skarnitzl, 2007) compares spectrograms and oscillograms to written text in order to establish boundaries between individual speech sounds. As the software was designed to work with Czech speech data, the text had to be transcribed from Vietnamese to Czech orthography (the transcription can be found in the Appendix 1). When all the text had been transcribed, the aligner could be applied and as a result we gained two extra tiers: *word*, and *phone*.

Although the aligner was not trained on Vietnamese data, it handled the segmentation with surprising accuracy. However, time still had to be invested in the manual segmentation mainly due to dysfluencies and hesitations that the aligner was not able to deal with.

3.4.2.3. Manual segmentation

As it was touched upon in the previous section, the forced aligner saved many hours of tedious and repetitive manual segmentation from scratch but it was still necessary to carry

out careful manual correction so the data could yield accurate results. Speech sound boundaries were adjusted according to the guidelines suggested by Machač & Skarnitzl (2009) which proved to be useful despite not being designed for Vietnamese.

In the end, the tier *phone* was not used and it only mattered to determine the boundaries within the tier *word* and *core*. After this step the procedure was the same for the text as well as the semi-spontaneous speech. We created a tier labelled *core* that was supposed the vocalic core, i.e., the domain of our measurements. With tier core ready, we created the interval tier *tone* and using the same script as in 3.4.1., we inserted 5 equidistant points into each vocalic core. When this was accomplished, we created another interval tier called *stress* and manually decided on the sentential stress distribution according to 2.2.2.7. and 2.2.3.2. with three levels (level 1 being the weakest and level 3 the heaviest). As the last step, we created PTs to all files and because the software to measure and calculate F_0 is not infallible either, all PTs had to be manually checked as well.

3.4.3. Perception test preparation

In order to design and perform the perception test, it was necessary to extract the syllables from the sentences described in Tab. 3.4. This was done by means of another PRAAT script the purpose of which was to cut the sound file according to the segments marked in the TG. The syllable files were then used as material for tonal discrimination a test that was composed in ALVIN (Hillenbrand 2015), a software designed for implementation of phonetic experiments.

The objective of the test was to listen to the individual syllables in both dialects with three stress levels and decide which of the tones they carried. We added the *replay* button so the

test subjects could listen to the segments repeatedly but the maximum number of replays was limited to 3. Whenever the subjects felt they recognized the tone, they pressed a button labelled with tone in question and the programme automatically played the next syllable. The interface of the perception test can be seen in Fig. 3.1. below.

3.5. Data extraction

This section aims to introduce the means of extracting data for the purposes of task 1 (comparison of syllables in context) and task 2 (comparison of reading and semi-spontaneous speech). As mentioned in section 3.4., the speech sounds were annotated to Text Grids (TGs) and Pitch Tiers (PTs) in PRAAT (Boersma & Weenink 2017) and the data for the analysis were extracted from these two types of files.

3.5.1. Individual syllables without context + syllables in context

F0 of the selected syllables was measured in five points placed equidistantly into the tier *core* with the last point at the very end of the syllabic core and the first one 15ms after the transition from the syllabic onset to the vocalic core. The measurements were taken in Hz but then a PRAAT script was employed that converted them to semitones and subsequently normalized by the speaker's mean F_0 to render the data directly comparable. The mean F_0 of every speaker used as an anchor for normalization was calculated by means of another PRAAT script prior to the actual data extraction.

HN male	121.6	*SG male*	153.4
HN female	219.8	*SG female*	222.9

Table 3.5. *Mean F_0 for male and female speakers of the Hanoian and Saigonese dialect measured in Hz. Whereas the mean F_0 of female speakers differs merely by 3.1 Hz, Saigonese men seem to speak significantly higher than Hanoian men.*

Measurements of F_0 were pasted to and MS Excel table and categorized according to dialect, gender, style (isolated syllables, syllables extracted from reading, syllables extracted from semi-spontaneous speech) and left context (0, High, Low). The reason for not analyzing right context can be found in Brunelle (2009a), who concluded that anticipatory tonal coarticulation in Vietnamese is not prominent, whereas progressive tonal coarticulation manifests itself more clearly.

3.5.2. Reading and semi-spontaneous speech

The means of measuring F_0 were identical to the section 3.5.1. above. The research objective shifted from a group of handpicked syllables to the reading material and semi-spontaneous speech in general. The presupposition was that changing focus from individual syllables to connected speech should yield results that would correspond more accurately with linguistic reality. Employing the same method as in 3.5.1., we managed to gather all the data and transfer them into an MS Excel table. As tonal coarticulation is not phonologized in Vietnamese, the variable "context" was no longer traced but we replaced it by the variable "stress" (on the scale 1 to 3 with 3 being the most prominent to reflect the impact of sentential stress on tone contours. The rest of the variables stayed the same – dialect, gender and style (reading, semi-spontaneous speech).

One point that cannot be omitted regarding sections 3.5.1. and 3.5.2. is that the method of analysis was devised to capture the continuous contour of F_0 only. Therefore, it is unable to yield any kind of quantitative assessment of voice quality that might also serve as a cue for tonal distinction. Due to interpolation, areas where F_0 was distorted by glottalization or laryngalization were smoothened out and the charts in sections 4.1. and 4.2. do not reflect them. Moreover, F_0 was measured in 5 equidistant points disregarding syllabic duration. In

order to account for these methodological obstacles, manually selected examples of voice quality features will be mentioned in sections 4.1.3. and 4.2.3. and parts of sections 4.1. and 4.2. are devoted to syllabic duration.

3.6. Perception test administration

When administering perception test, it is necessary to make a decision whether to use earphones. The advantage of earphones is the deliverance of stimuli in uncompromised quality but, on the other hand, the respondent might feel detached from the surroundings, which could potentially affect the results. Our respondents had to use earphones mainly because there is a significantly large amount of background noise present in Vietnamese cities due to lively traffic and particularly motorcycles.

The perception test design drew inspiration from McGuire (2010) according to whom it is necessary to establish the amount of stimuli in one set and the amount of sets, the length of pauses and the sound signalizing them. A test should not be longer than 10 to 15 minutes in order for the subjects to maintain focus. It is also important for the subsequent data processing to note down reaction time and whether the subjects interacted with the person administrating the test. There should be at least three test stimuli so the subjects can adjust to the task. The test stimuli should be discarded and not included into the data. Our perception test also marked the first three trials as test trials and did not include them in the results. The average completion time of the test consisting of 144+3 trials was 11 minutes 34 seconds.

The perception test devised for the purposes of this thesis employs a discrimination test design that McGuire (2010) called “labelling”, i.e. only a single stimulus is presented in each trial and the subjects must apply a label to the stimulus (one of the six Vietnamese

tones in our case). The test was administered to 9 Hanoian speakers and 9 speakers of the Saigonese dialect. The test subjects were informed that they would listen to a syllable “ma” in all tonal variations uttered by Hanoian and Saigonese male and female speakers. They were supposed to decide which tone they heard and press an accordingly labelled button (see fig. 3.1.). Pressing the tone button would then elicit the next syllable. There was also a possibility to replay each stimulus up to 4 times. However, the average number of replays per stimulus was merely 0.72. The set of syllables was played in a random order and every syllable was included twice.

Figure 3.1. *A screenshot of the perception test programmed in Alvin 3.12*

4. Analysis

With the data collection and extraction described, we will now present the research hypotheses and, subsequently, the results of the measurements conducted in the three analytical tasks.

4.1. Individual syllables + syllables in high, low and 0 context

The first task set two goals. Firstly, to model average contours of all tonal variations in both researched accents. The aim is to determine to what extent the tonal contours differ between the dialects and whether they follow the patterns described by the previous studies mentioned in 2.2.2.6. and 2.2.3.4. Secondly, this task investigates how much progressive coarticulation described by Brunelle (2009a) affects tonal contours and whether they differ significantly from their counterparts uttered in isolation.

4.1.1. Hypotheses

$H1_0$ – Tonal contours of both dialects are identical.

$H1_A$ – Each dialect exhibits notably different tonal contours. Moreover, according to Brunnelle (2009b), Hanoian tones combine pitch and voice quality cues whereas in Saigonese tones, pitch contour is the chief decisive factor for tonal discrimination.

$H2_0$ – Contrary to Brunelle's (2009a) claim admitting the existence of progressive coarticulation effects, there are no difference in tonal contours between the syllables in isolation and in context.

$H2_A$ – There is a statistically significant difference in the initial segment of tonal contour based on the preceding context. If the left context is low and the onset of the following syllable is high, there is a rise in the first segment of the tonal contour although it does not

correspond to the shape of the canonical contour. Similarly, if the left context is high and the onset of the syllable is low, a fall occurs in the first segment of its tonal contour.

4.1.2. Results

If we disregard the category of dialect and represent all the gathered data in a chart, we get Fig. 4.1. The representation could be called "super urban Vietnamese" as it captures a blend of the two largest cities in Vietnam with possibly one fifth of the country's population. However, the chart is, in fact, very much removed from linguistic reality as it describes the language of non-existent speakers. On the other hand, Fig. 4.1. helps us familiarize with Vietnamese generic tone patterns and more importantly it serves as a valuable control group to assure ourselves that our data are in agreement with the findings of previous researches on the topic of Vietnamese tones listed in Chapter 2. Nevertheless, in order to begin with the description of real-world situation, the data must be broken down according to the dialects. As it was already mentioned in section 3.4., the tone contour charts do not consider duration of the researched syllables hence all the lines are drawn equivalent in length although the tones B2, D1 are D2 substantially shorter than the others (see Fig. 4.11.).

In Fig. 4.2., we can observe the tone inventory of Hanoian Vietnamese. Out of the two analysed dialects, Hanoian tones have been described in greater depth by a larger number of researchers. For this reason, we deem it desirable to deal with it first and use it as a means of uncovering, for instance, any mistakes in our methodology. Had the contours been significantly different from those in the papers and studies listed in Chapter 2, it could signal either a ground-breaking discovery (which is rarely the case) or a flaw in our research method. Majority of the contours answer to the expectations based on the studies by Đoàn (1977), Pham (2003) and Brunelle (2003, 2009a, 2009b).

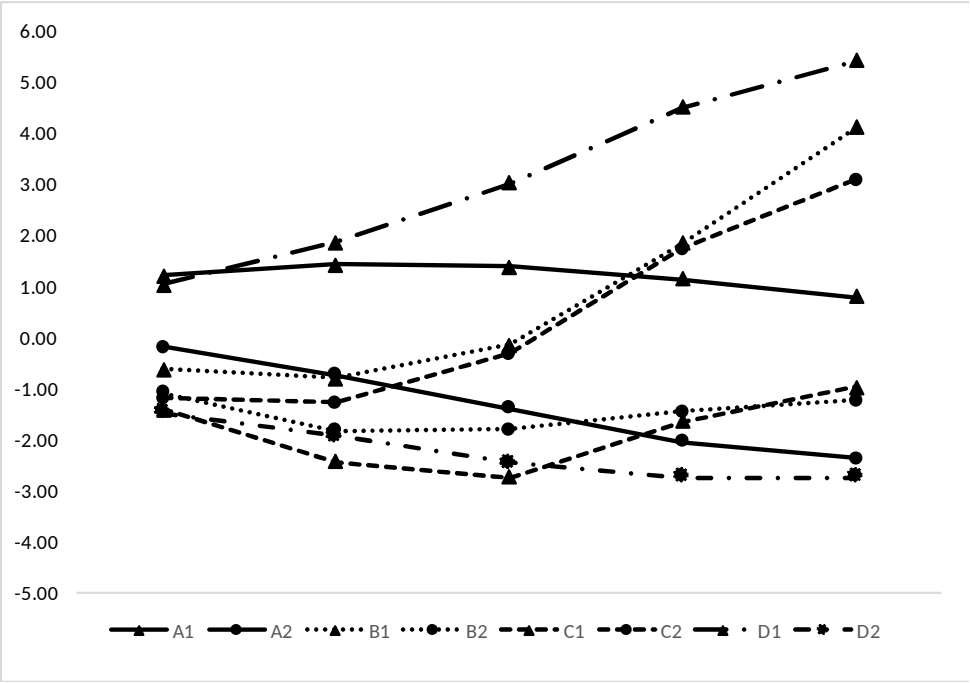


Figure 4.1. *Average tone inventory of Vietnamese tones after combining the data from both dialects. This representation of abstract underlying shapes does not reflect linguistic reality. (axis y – semitones)*

Substantial differences between the contours of B1 x D1 and B2 x D2 speak in favour of Pham's 8-tone classification. Tones A2, B2, C1 and C2 begin almost exactly on the average F_0 level of every speaker (i.e., the value of 0 in the normalized data). Tone A1 conforms to the "high-level" classification, A2 is lower and falling, D1 is high and rising steadily and D2 low and falling slightly. Although the contours of A2 and D2 appear extremely similar, their confusion is impossible firstly because D2 only appears in the "checked syllables" (see section 2.2.2.5.), and secondly because it tends to be significantly shorter (see Fig. 4.11.). Tone B1 begins more than two semitones below the average level of F_0 and the contour is level in the first half followed by an abrupt and steep rise towards the end. B2 is usually classified as a low tone but it appears that it is actually higher than A2. As opposed to A2, B2 is shorter (see Fig. 4.11.) and heavily glottalized (see fig (x). Standard Hanoian C1 is

normally classified as a lower fall-rise but our data clearly shows that the rise is not present. It is, in fact, a tone ending the lowest of all the Hanoian tones. Fig. 4.11. also shows that it is shorter than A2. Finally, the tone C2 begins on the average F_0 level and finishes more than 3 semitones above the average. However, what the chart is unable to show, for the reasons mentioned in 3.5. is the heavy glottalization in the mid section of the contour as can be seen in Fig. 4.13.

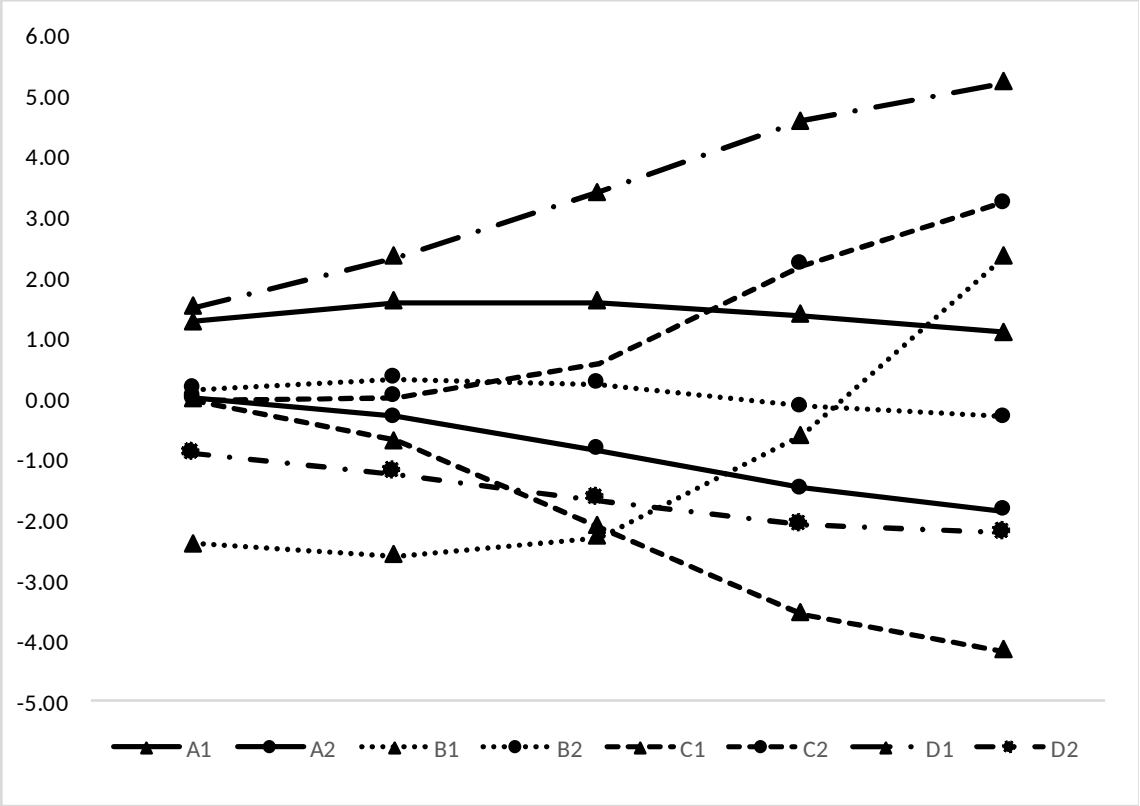


Figure 4.2. *Average tone inventory of Hanoian Vietnamese. (axis y – semitones)*

Fig. 4.3. illustrates the tonal inventory of Saigonese Vietnamese. It is quite apparent that the 8-tone classification advocated by Pham (2003) cannot be applied to this dialect. Tones B1 x D1, B2 x D2 and to an extent also C1 x C2 appear homophonous. We can notice that the tones are generally spaced farther apart from each other than in the Hanoian dialect. This observation conforms to Brunelle's (2009a) claims that the Saigonese dialect depends on

pitch contour as a cue for tone discrimination more than the dialect used in Hanoi. Tones A1 and A2 exhibit a contour very similar to A1 and A2 in Hanoi, they are just farther apart. Saigonese B1 and D1 start a bit lower but also resemble the Hanoian D1. Analogically, Saigonese B2 and D2 bear resemblance to the Hanoian D2. Although the general public as well as the researchers of Vietnamese tones unanimously agree that the Saigonese tonal inventory contains merely 5 tones, our data show that there might be an audible difference between the realizations of C1 and C2. On the other hand, the contours are shaped very similarly and there is no perceptional necessity to classify Saigonese C1 and C2 as two separate tones. Compared to the dialect of Hanoi, contours of Saigonese C1 and C2 seem to resemble the Hanoian C2 whereas the Hanoian C1 does not have a counterpart.

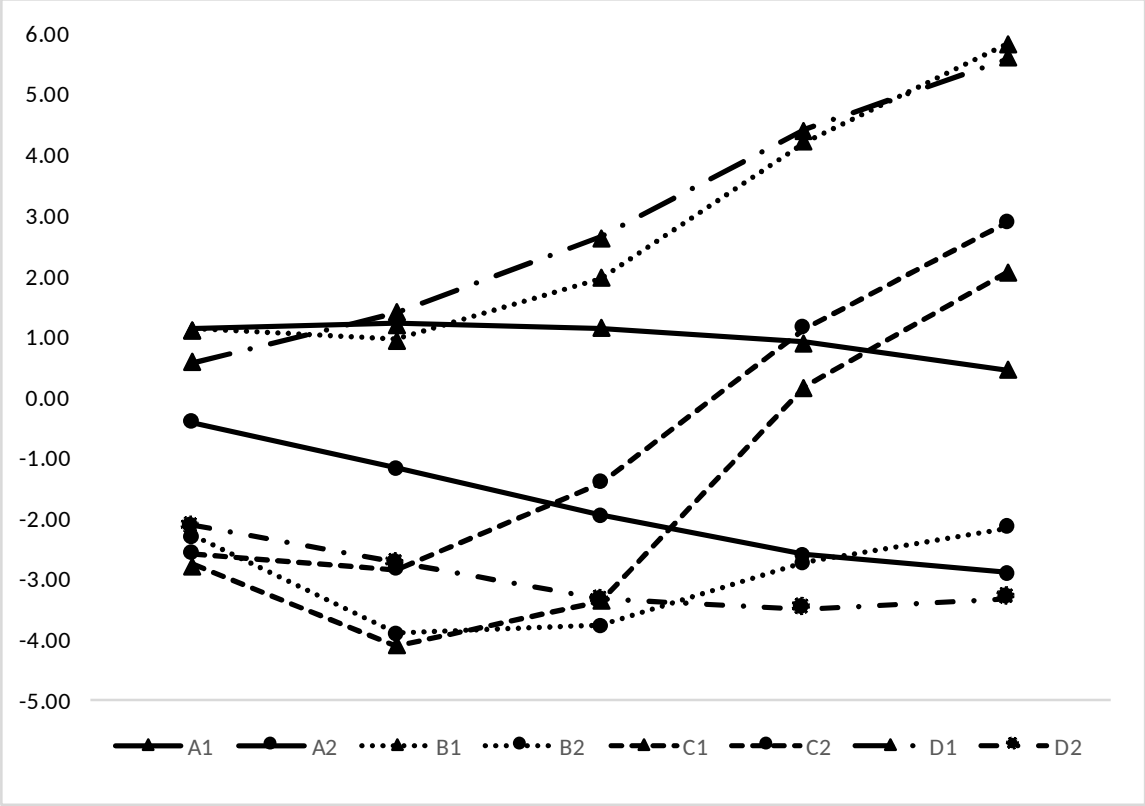


Figure 4.3. *Average inventory of Saigonese Vietnamese. (axis y – semitones)*

In the following set of paired figures (Fig. 4.4. to 4.7) a more detailed observation of the individual tones is offered. Fig. 4.4. compares tones A1 and A2 of both dialects. Both contours are very similar across the dialects, just the Saigonese tones seem to drop more towards the end compared to their Hanoian counterparts. The difference is, however, under or around 1 semitone, which is not necessarily significant.

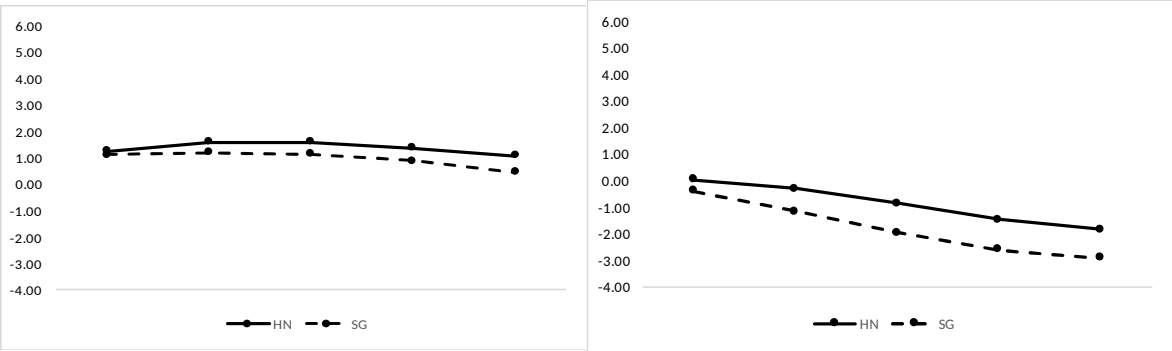


Figure 4.4. *Representations of tones A1 (left) and A2 (right) from both dialects. (axis y – semitones; HN = Hanoian, SG = Saigonese)*

Contours of tones B1 and B2 uttered in Hanoi and HCMC are similar in contour shape but different in height. The rising tone B1 in HCMC begins and ends more than 3 semitones higher than in Hanoi. On the other hand, B2 in HCMC begins and ends approximately 2 semitones lower than in Hanoi.

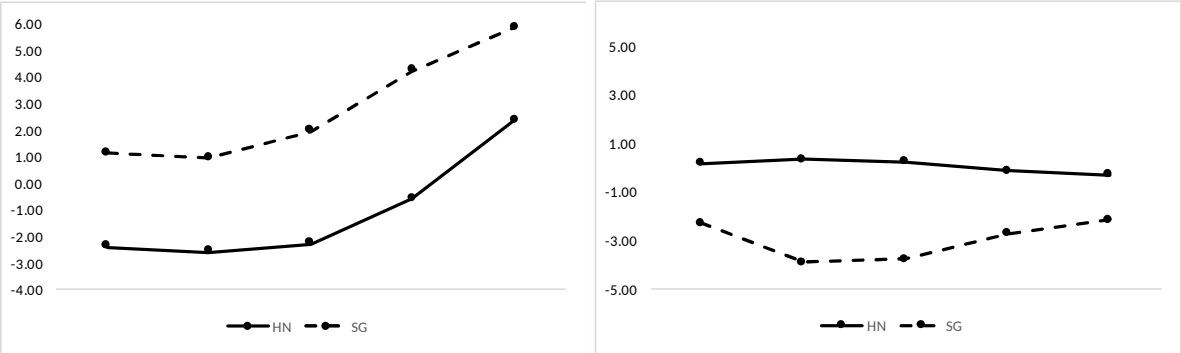


Figure 4.5. *Representations of tones B1 (left) and B2 (right) from both dialects. (axis y – semitones; HN = Hanoian, SG = Saigonese)*

Although the tone C2 in HCMC begins more than 2 semitones lower than in Hanoi, the target height is very similar. This cannot be said about the contour of C1 because C1 in HCMC is clearly a rising tone whereas in Hanoi, it has the contour of a steeply falling low tone.

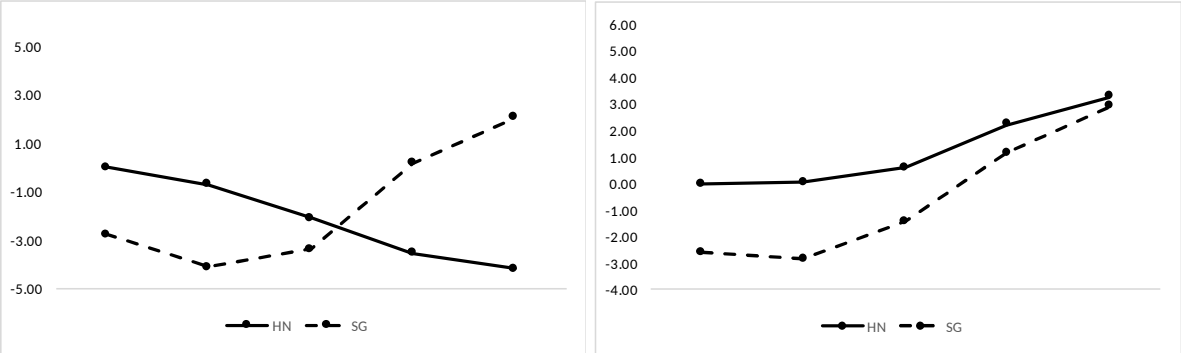


Figure 4.6. *Representations of tones C1 (left) and C2 (right) from both dialects. (axis y – semitones; HN = Hanoian, SG = Saigonese)*

Tones D1 and D2 as pictured in Fig. 4.7. are similar in contour shape, height and even duration (see Fig. 4.11.)

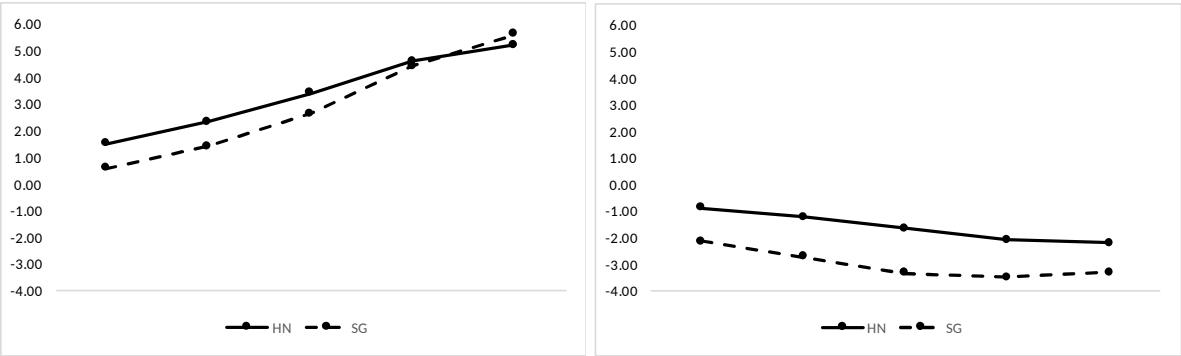


Figure 4.7. *Representations of tones D1 (left) and D2 (right) from both dialects. (axis y – semitones; HN = Hanoian, SG = Saigonese)*

Figures 4.8. and 4.9. present the results of experimenting with progressive tonal coarticulation based on Brunelle (2009a). Fig. 4.8. shows Hanoian tones A1 and B1 in three variants – high, zero and low preceding context; Fig. 4.9. is an analogical representation of Saigonese tones A2 and B2. Other tones in both dialects conform to the same trend,

therefore it is not necessary to list them there We can notice that the zero context syllable has the most neutral contour whereas the first section of all contours in H/L context is always affected because it must compensate for the overlap of the preceding tone. It is also interesting that the contours of H and zero context syllables become very similar after the second measurement whereas the L context contour remains lower for the entire duration of the tone.

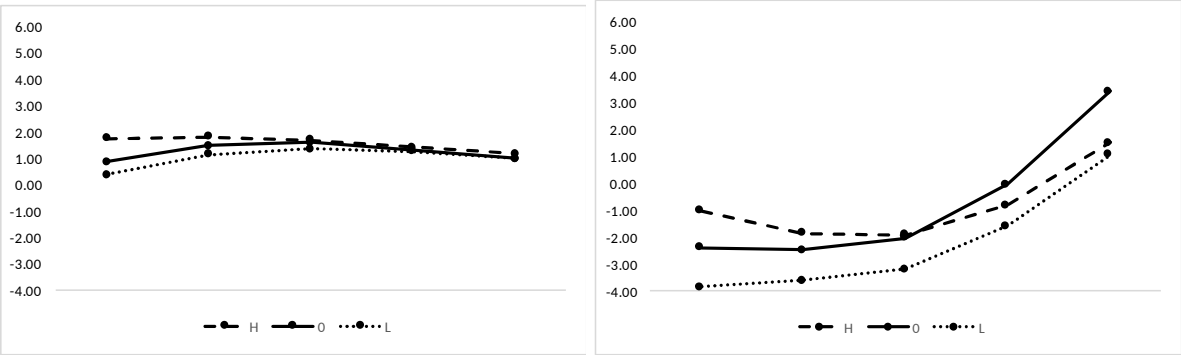


Figure 4.8. *Hanoian tones A1 and B1 uttered in high, zero and low left context. (axis y – semitones)*

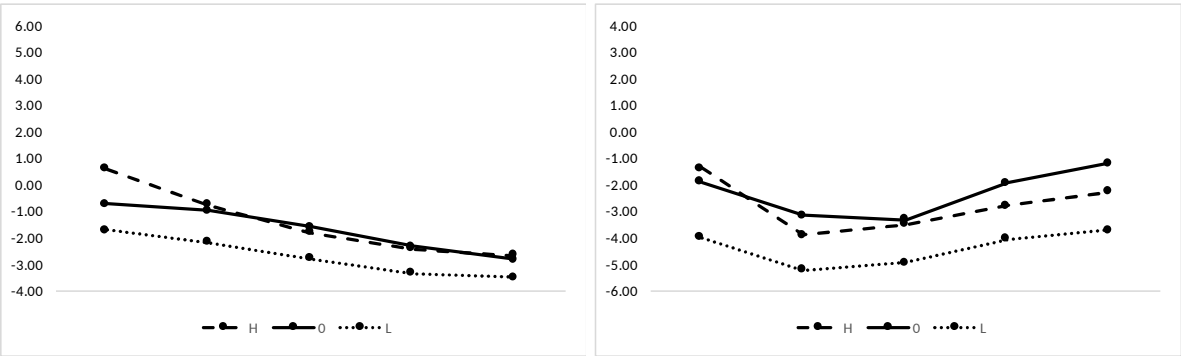


Figure 4.9. *Saigonese tones A2 and B2 in high, zero and low preceding context. (axis y – semitones)*

Although the dialect is the most significant variable influencing the contour of Vietnamese tones, there are other variables that need to be taken in consideration. Age might be one of them but because the group of recorded subjects for the purposes of this thesis was very homogenous in this respect, we were mostly concerned with gender. There are languages

such as Russian where gender is a highly influential prosodic factor. However, as it can be seen in Fig. 4.10., gender does not bear any significance for tone contours in Vietnamese. Although the contours in the Saigonese dialect are farther apart, they still firmly cluster together. The other tones follow the same pattern and so there is no need to represent them here further.

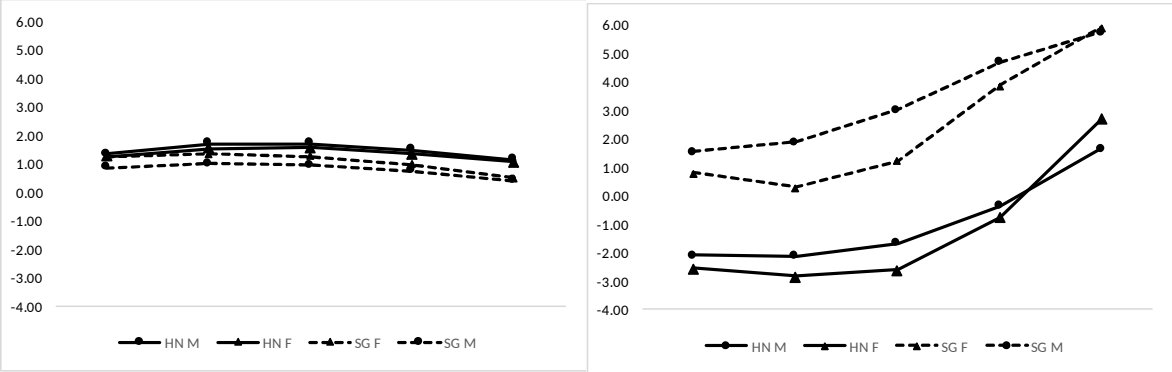


Figure 4.10. *Representations of tones A1 (left) and B1 (right) from both dialects with respect to gender. (M – male, F – female; HN = Hanoian, SG = Saigonese; axis y – semitones)*

As it was mentioned above, the tone contours are all represented by lines of the same length because the measurements were carried out in equidistant points inside the respective syllables regardless of their duration. However, the duration is also likely to be one of the cues for tone discrimination in Vietnamese and hence it is necessary to dedicate it some space. Fig. 4.11. indicates average durations of tones from both dialects uttered in isolation. Saigonese syllables are generally longer. In tones A1, A2 and B2, the difference is only 21-30 ms. In C2, the difference is approximately 60 ms similarly to D1 and D2. C1 and B2 in the Hanoian dialect are by about 130 and 170 ms shorter than their Saigonese counterparts. Tones D1 and D2 are generally shorter because the voiceless plosives in syllabic coda do not bear any F_0. As opposed to Saigonese B2 where there is milder glottalization present throughout the tone, the Hanoian B2 ends in glottalization so heavy that the F_0 disappears similarly to the tones D1 and D2. The tonal contour of Hanoian B2 is therefore shorter in

comparison with the Saigonese B2. The most surprising results concern the Hanoian tone C1. Fig. 4.2. and 4.6. showed the loss of the final rise in Hanoian C1. Furthermore, it appears to be significantly shorter than the other tones that are not ended by voiceless plosives or heavy glottalization. On average, Hanoian C1 is by 70 ms or 25 % shorter than Hanoian A2, a low tone with a very similar contour.

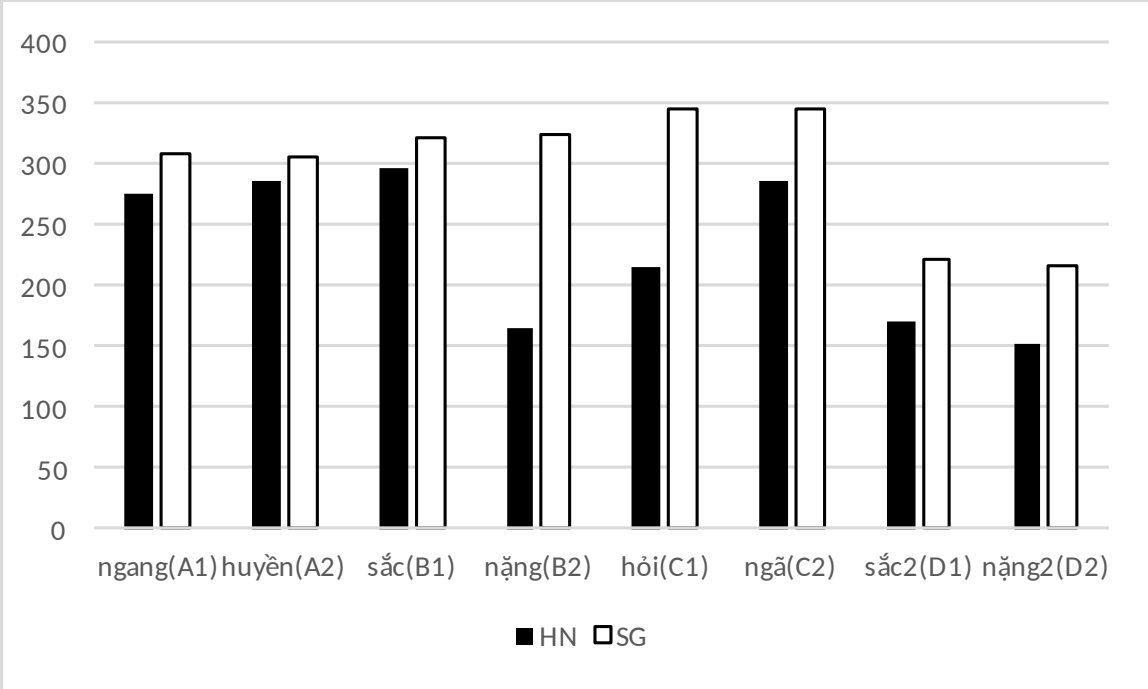


Figure 4.11. *Average duration of tones uttered in isolation in both dialects measured in milliseconds.*

4.1.3. Discussion

Pham (2003) proposed an 8-tone classification for Hanoian Vietnamese due to the fact that contours of B1 x D1 and B2 x D2 are different. This classification is rather useless for the purposes of communication because B1 x D1 and B2 x D2 are merely allotones in complementary distribution but from the perspective of detailed speech description, the findings of this experiment speak in favour of her propositions as can be seen in Fig. 4.2. On the other hand, Fig. 4.3. shows that Pham's classification is not applicable to the Saigonese

dialect. Apart from the fact that the contours of tones C1 and C2 are very similar and the tones are generally considered homophonous, the contours of tones B1 x D1 and B2 x D2 are also very close to each other. Therefore, it could be claimed that Hanoian Vietnamese truly has an inventory of 8 tones but Saigonese Vietnamese contains merely 5 tones.

In Hanoian Vietnamese, the tone B1 begins more than two semitones below the average value of F_0 and the contour is flat in the first half followed by an abrupt and steep rise towards the end, whereas D1 begins high and its contour is straight and equally steep. In Saigonese, both B1 and D1 resemble the contour of the Hanoian D1.

Hanoian B2 is almost 2 semitones higher than D2. In fact, it is even higher than A2. However, confusion of A2 and B2 is prevented by shorter duration (see Fig. 4.11.) and glottalization (see Fig. 4.12.) of B2.

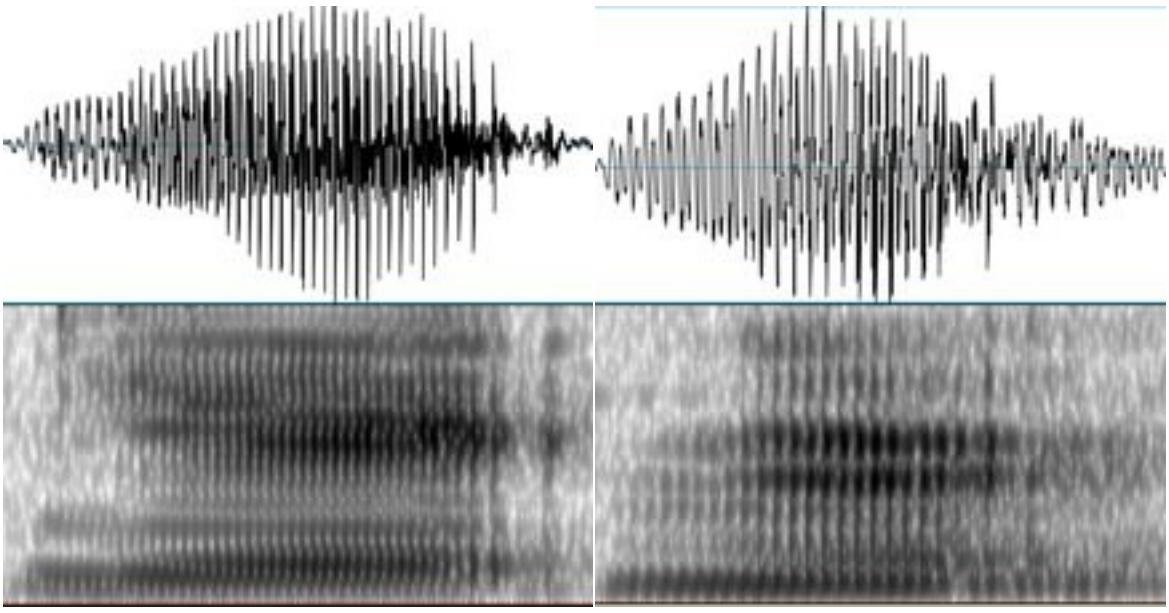
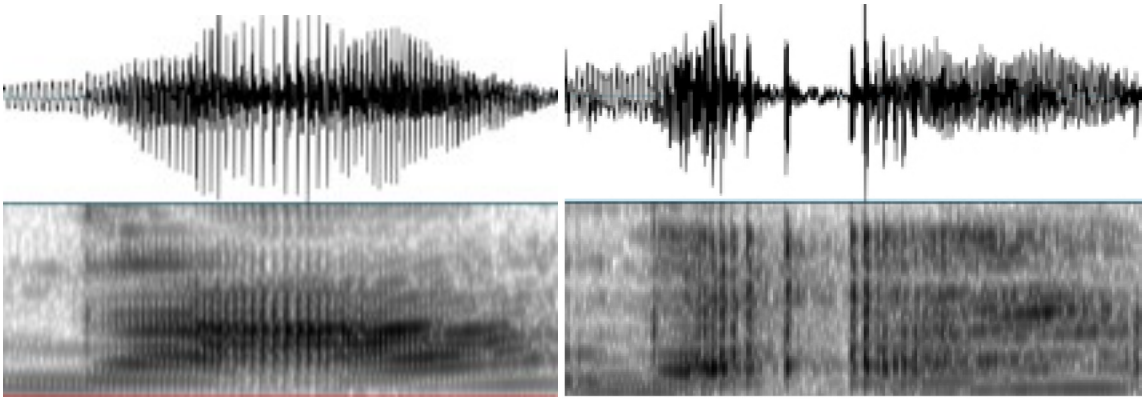


Fig. 4.12. *Realizations of the tone B2 (syllable nhẹ) by speakers HN10 (left) and HN12 (right).*

Standard Hanoian C1 is normally classified as a lower fall-rise but our data clearly show that the rise is not present. Moreover, it ends the lowest of all the Hanoian tones. Although Fig. 4.11. suggests that C1 is shorter than A2, Fig. 4.22. indicates that the difference in duration is ameliorated in reading or speech. This phenomenon could possibly lead into merging of tones A2 and C1 in Hanoian Vietnamese in the future.

The tone C2 begins on the average F_0 level and finishes approximately 3 semitones above the average. The charts do not capture heavy glottalization in the mid-section of the contour as can be seen in Fig. 4.13. The central glottalization very much seems to be an important cue for tonal identification.



F**igure 4.13.** *The tone C2 (syllable ngã) uttered by speakers HN3 (left) and SG7 (right).*

4.2. Tones in Reading and Semi-Spontaneous Speech

The second task attempts to infer tonal tendencies from data closely resembling natural speech, i.e. it does not take in consideration merely preselected syllables or syllables recorded in isolation but it looks at the data as a whole. The aim is to determine whether language data examined in laboratory conditions in order to limit variables and variation can yield results that might be accepted as an accurate representation of linguistic reality.

Therefore, the data analysis in this task employ quantitative methods that should be able to illustrate how the contours change based on changes in variables.

4.2.1. Hypotheses

$H3_0$ – There is no difference in tonal contours between the canonical contours, reading and semi-spontaneous speech.

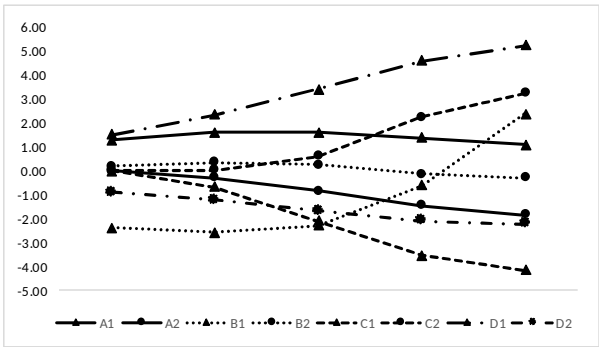
$H3_A$ – The degree of spontaneity in speech production is a factor influencing tonal contour in Vietnamese.

$H4_0$ – Sentential stress does not influence contours of Vietnamese tones.

$H4_A$ – Tonal contours in connected speech differ from the canonical contours in terms of prominence and duration.

4.2.2. Results

The first question to answer for this task is whether style affects contours of Vietnamese tones. Figures 4.14. and 4.15. suggest that style does indeed affect contours of Hanoian Vietnamese rather significanly. Especially the representation of Hanoian tones from semi-spontaneous speech indicates that the tones are flattened and they converge towards the average F_0 of the speakers. Many of the tones differ from each other by a mere 1 semitone or less. It is also interesting how the final rise of B1 decreases in reading and semi-spontaneous speech to such an extent that the tone is nearly level and it is even lower than the tone B2. The same phenomenon can be observed in the data with stress level distinction (see Fig. 4.17.).



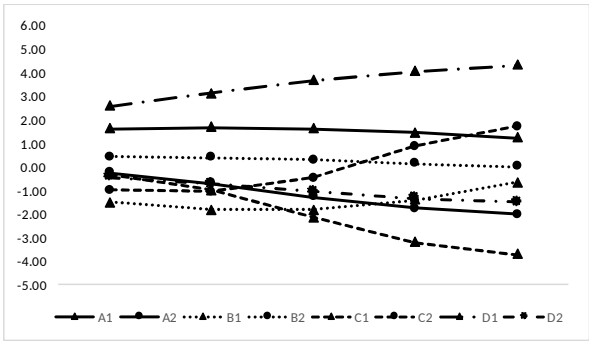
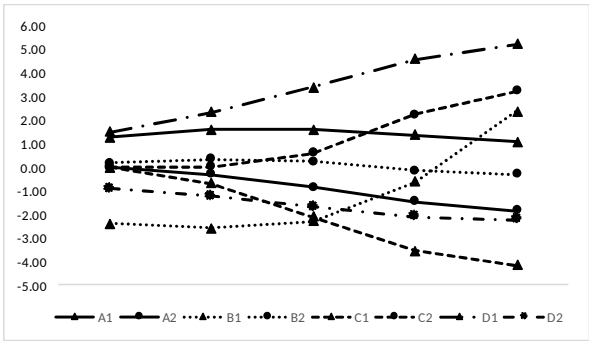


Figure 4.14. *Comparison of Hanoian tone inventory representations based on "canonical realizations" as described in section 4.1. (left) and data collected from reading (right).*



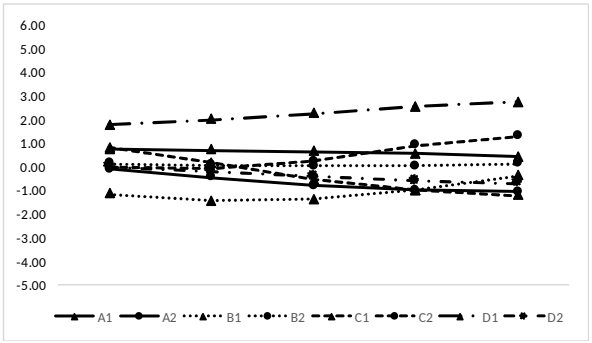


Figure 4.15. *Comparison of Hanoian tone inventory representations based on "canonical realizations" as described in section 4.1. (left) and data collected from semi-spontaneous speech (right).*

It was already noted in section 4.1. that the Saigonese tones tend to be positioned father apart from each other and this tendency prevails even with a change of style. Although figures 4.16. and 4.17. show that the effects are similar as in the Hanoian tones above, the span of Saigonese tones is clearly broader than the span of the tones spoken in Hanoi.

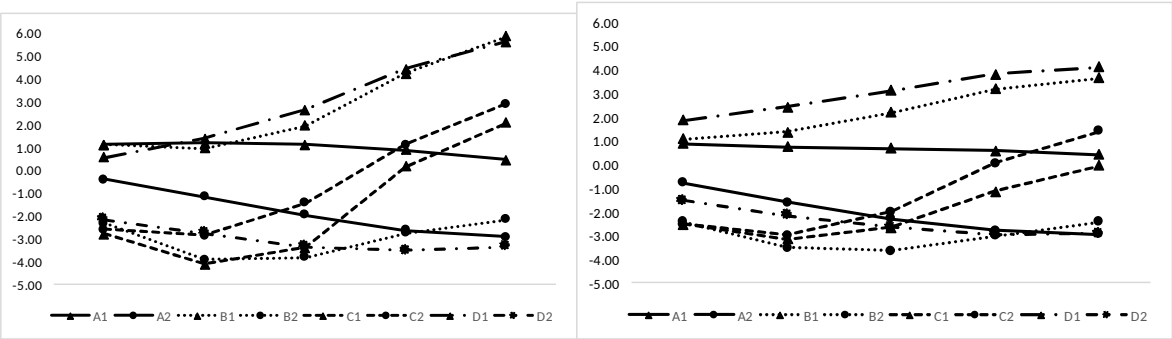


Figure 4.16. *Comparison of Saigonese tone inventory representations based on "canonical realizations" as described in section 4.1. (left) and data collected from reading (right).*

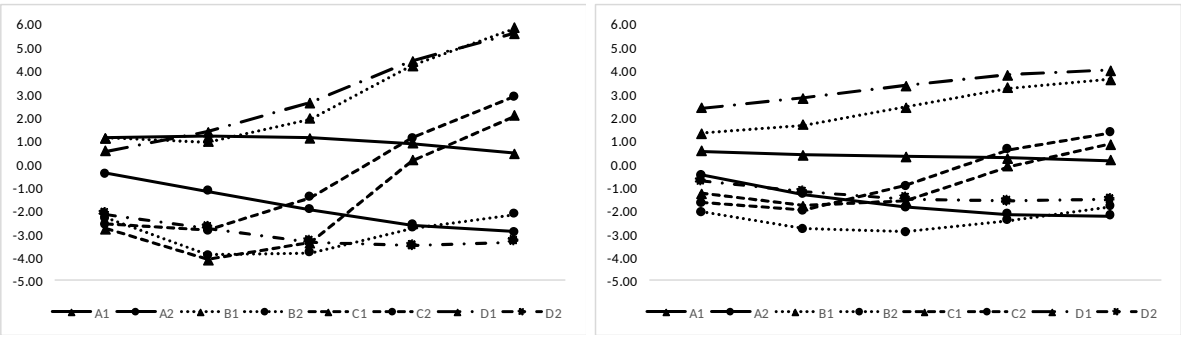


Figure 4.17. *Comparison of Saigonese tone inventory representations based on "canonical realizations" as described in section 4.1. (left) and data collected from semi-spontaneous speech (right).*

Fig. 4.18. portrays the tones A2 and B1 in both dialects. It is quite clearly visible that there is a tendency towards levelling the tonal contours and converging to the average F_0 in both dialects.

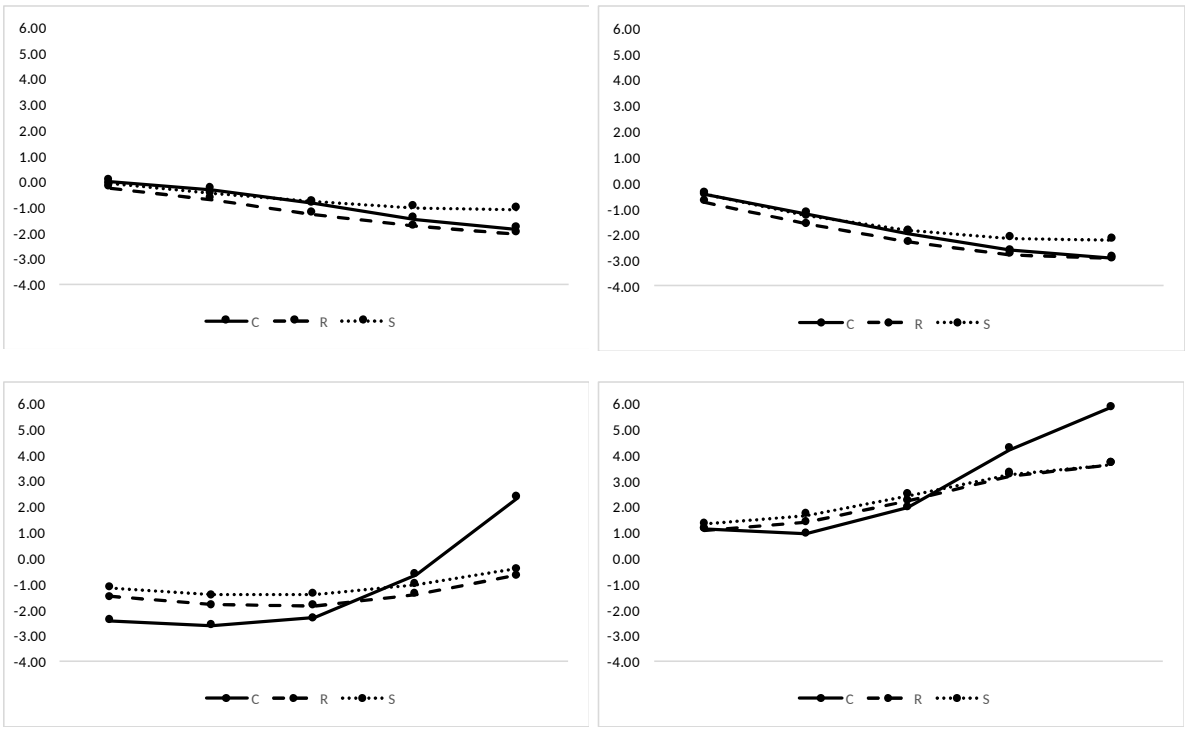


Figure 4.18. *Representations of tones A2 (top) and B1 (bottom) from Hanoi (left) and HCMC (right) in the canonical realization (C), reading (R) and speaking (S).*

The purpose of Figures 4.19. and 4.20. is similar to Figures 4.14. – 4.17. However, instead of the effects of style on tone contour, they strive to depict the effects of different levels of sentential stress. The tendency appears to be quite analogical to the effects of style – weakening stress causes the tones to flatten their contour and converge towards the average F_0. Similarly to style, the Saigonese tones manage to maintain greater distance even on the weakest stress level.

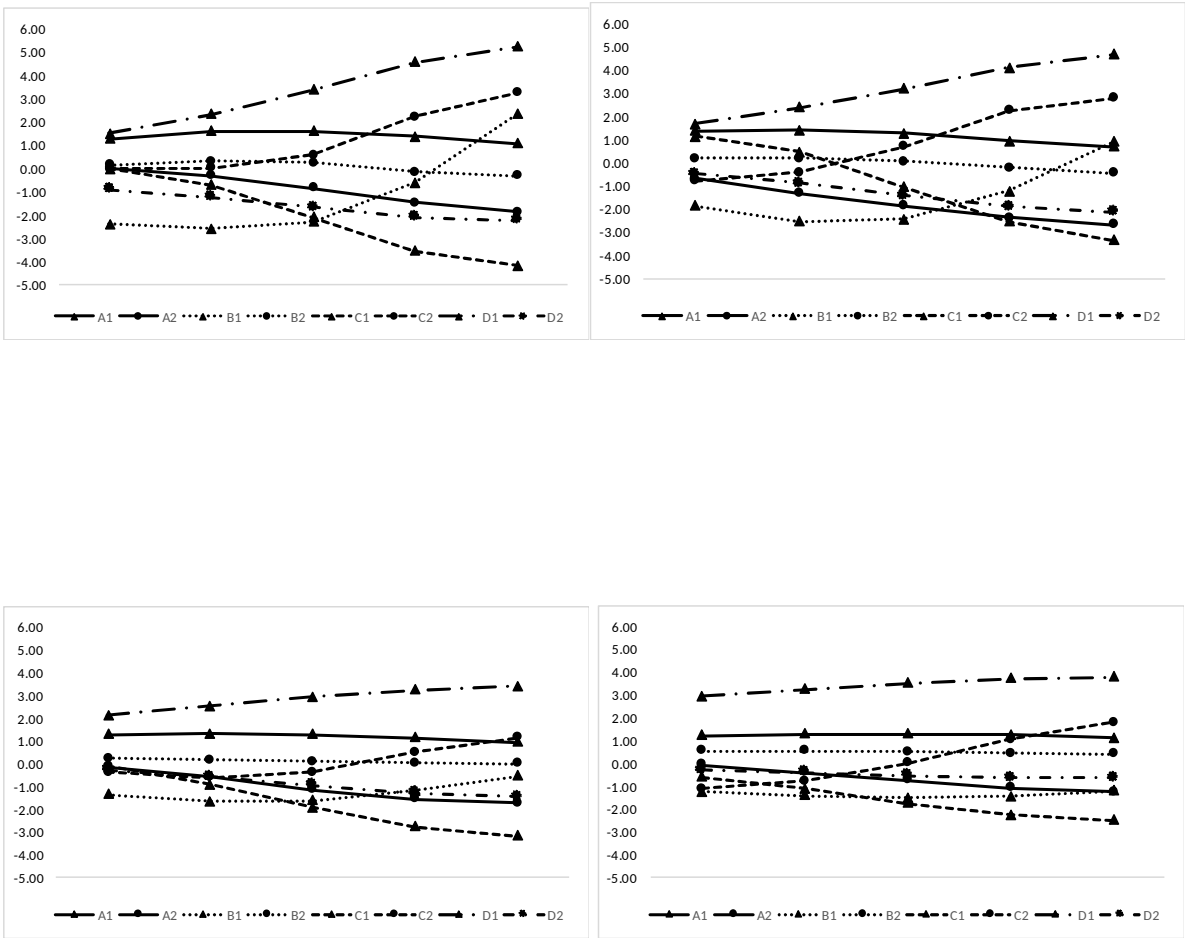


Figure 4.19. *Comparison of Hanoian tone inventory representations based on "canonical realizations" as described in section 4.1. (top left) and on data from syllables bearing stress level 3 (top right), 2 (bottom left) and 1 (bottom right)*

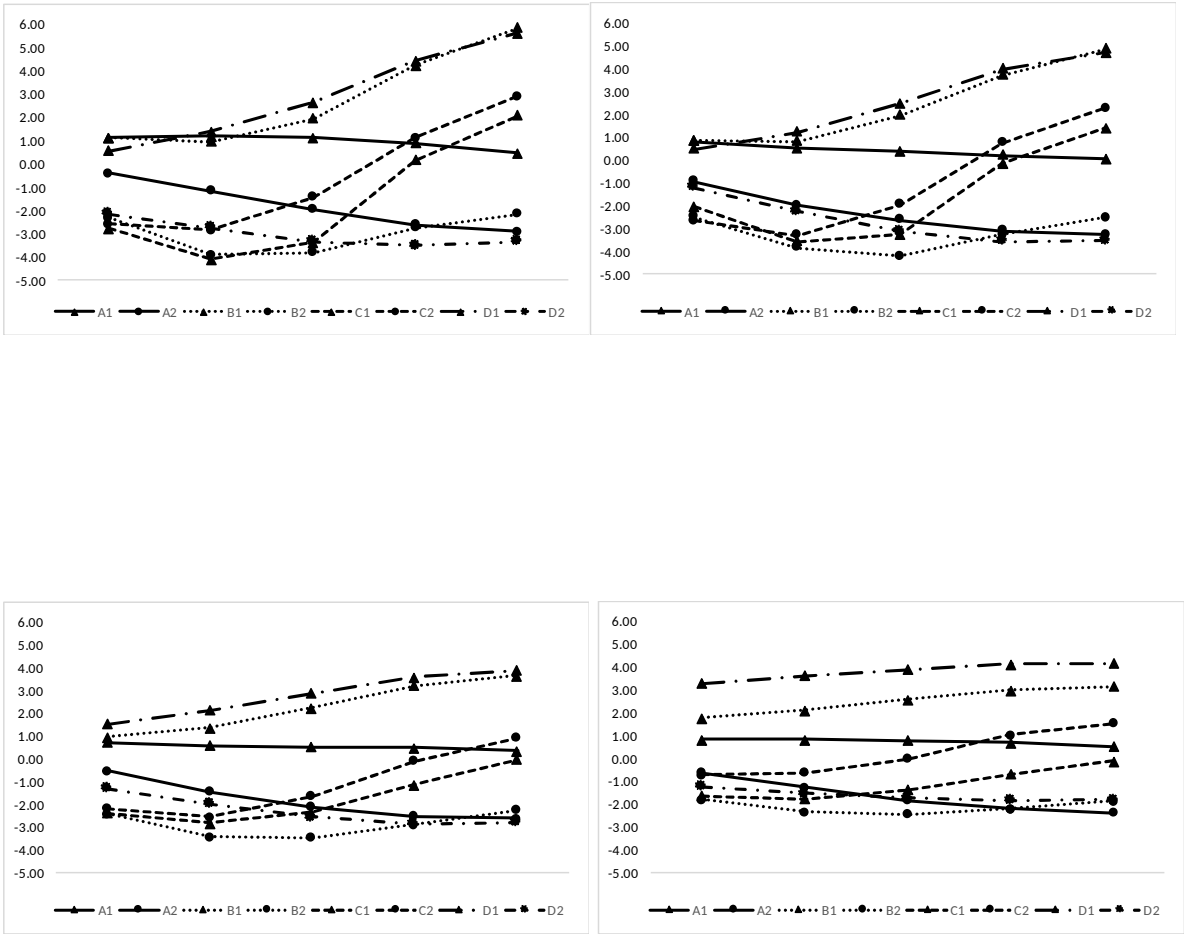
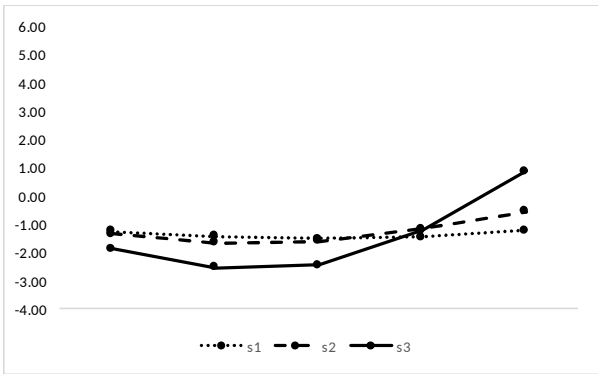


Figure 4.20. *Comparison of Saigonese tone inventory representations based on "canonical realizations" as described in section 4.1. (top left) and on data from syllables bearing stress level 3 (top right), 2 (bottom left) and 1 (bottom right).*

Fig. 4.21. is quite analogical to Fig. 4.18. above. It depicts the tone B2 uttered in both dialects on three levels of sentential stress. Weaker stress flattens the contour and makes it convergent towards the average F_0.



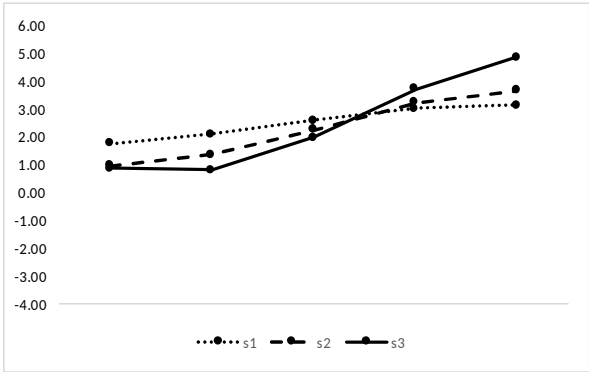


Figure 4.21. *Representations of the tone B1 from Hanoi (left) and HCMC (right) on stress levels 1, 2 and 3.*

So far, this section managed to prove that both style and sentential stress have the ability to affect tonal contour. Figures 4.22. and 4.23. compare the most canonical realizations of tonal contours from both dialects with contours of tones occurring in semi-spontaneous speech bearing the weakest stress. The realizations do not differ from Figures 4.15. and 4.17. to a great extent. Any discrepancies, might be in fact caused by the relatively small corpus of data in the category “semi-spontaneous speech, stress level 1”.

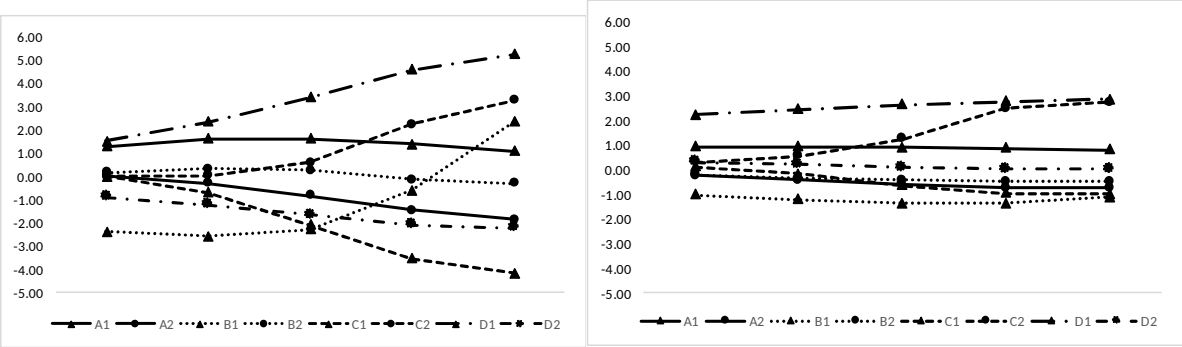


Figure 4.22. *Comparison of Hanoian tone inventory representations based on "canonical realizations" as described in section 4.1. (left) and on data from syllables from semi-spontaneous speech bearing stress level 1.*

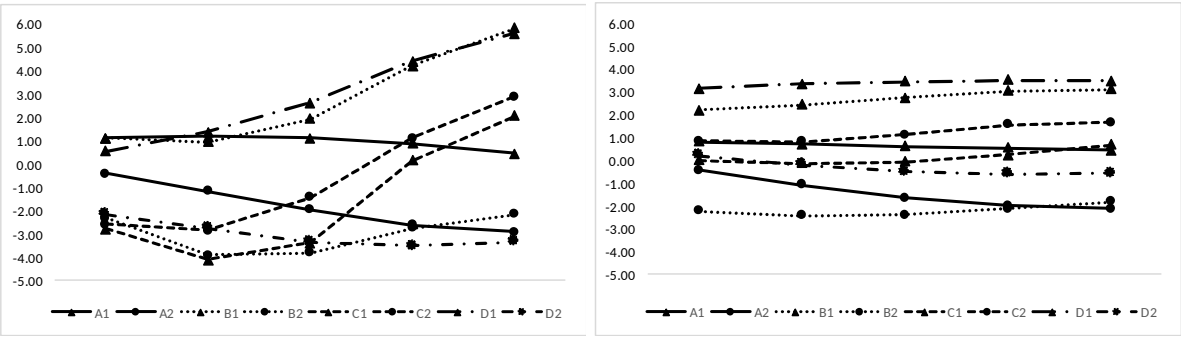


Figure 4.23. *Comparison of Saigonese tone inventory representations based on "canonical realizations" as described in section 4.1. (left) and on data from syllables from semi-spontaneous speech bearing stress level 1.*

Average duration of tones was already discussed in section 4.1. but only for the syllables in isolation and in selected context. Fig. 4.24. adds average durations of syllables in both dialects in reading (r) and semi-spontaneous speech (s). It is noteworthy that the charts do not seem to show any significant difference in duration between (r) and (s) in either dialect. However, both styles indicate duration substantially shorter than in the preselected syllables used in section 4.1. Duration in the Saigonese dialect seems proportional but a surprising feature emerged from the Hanoian data – in the canonical realization, the tone C1 is notably shorter than A2 but the duration equalizes in more natural speech and the tones become indistinguishable by duration.

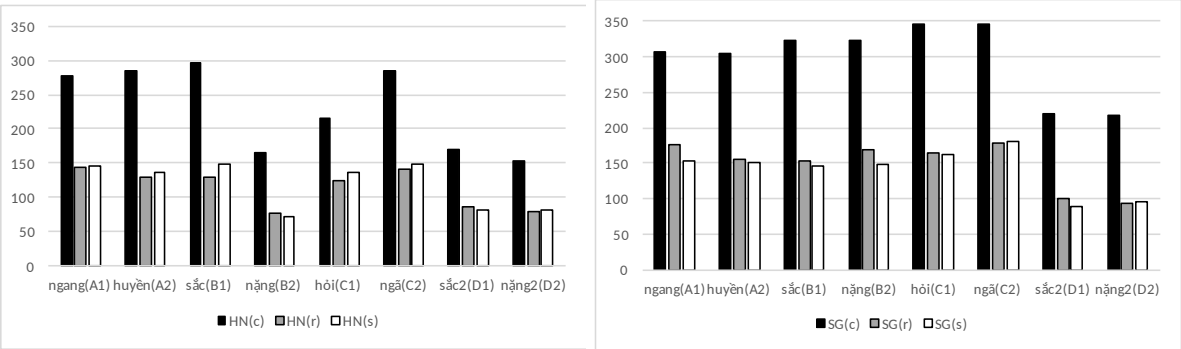


Figure 4.24. *Average duration of tone in the canonical realizations (c), reading (r) and speaking (s) in Hanoian Vietnamese (left) and Saigonese Vietnamese (right) measured in milliseconds.*

4.2.3. Discussion

The representations of Hanoian tones from reading, semi-spontaneous speech as well as the weak stress level show a clear tendency towards contour levelling and converging towards the average F_0 of the speakers. The most apparent example is the tone B1 that nearly loses the final rising segment in semi-spontaneous speech as well as in syllables with weak stress as demonstrated by Fig. 4.25. The graph also shows that the tone B1 is lower than both A1 and B2 by almost 2 semitones although it usually rises more than one semitone over A1 and 2-3 semitones over B2. Therefore, F_0 cannot serve as the chief cue for their distinction in connected speech. Other factors like duration, voice quality and most importantly context play a decisive role in tone discrimination.

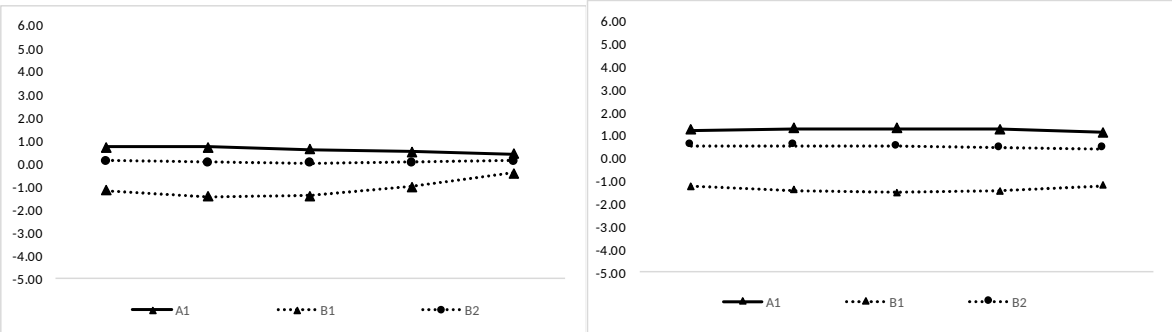


Figure 4.25. *Comparison of tones A1, B1 and B2 in semi-spontaneous speech (left) and stress level 1 (right) in the Hanoian dialect.*

Broader span of Saigonese tones was discussed already in section 4.1. but the tones are father apart even in realizations gathered from connected speech, which supports the observation postulated by Brunelle (2009a) that discrimination of Saigonese tones is based on F_0 differences to a greater extent than in the case of Hanoian tones.

Average duration of tones was also already discussed in section 4.1. but only for the syllables in isolation and in selected context. Duration in the Saigonese dialect seems proportional. In Hanoi, however, the tone C1, which has lost its final rise and probably uses duration as one of the cues to distinguish itself from A2 in canonical realizations, is not shorter than A2 when it comes to connected speech. If this tendency persists, it is quite possible that the Hanoian tones A2 and C1 might merge into one in the future.

4.3. Perception Test

The perception test is a complementary task in this thesis not necessarily comparable to the previous two by the amount of data gathered and work invested. Nevertheless, after the first two tasks focusing on tone production, it is desirable to look into the means of cross-dialectal tonal perception. The data presented in this section show how accurate the test

subjects from Hanoi and HCMC were in differentiating between the tones in both dialects, also with respect to the stress level from weak to medium and strong.

4.3.1. Hypotheses

$H5_0$ – Test subjects from Hanoi and HCMC are able to discriminate tones in the test equally accurately.

$H5_A$ – There is a statistically significant difference in tonal discrimination between the speakers of two dialects.

$H6_0$ – Stress level of experimental syllables does not influence the ability of test subjects to discriminate Vietnamese tones.

$H6_A$ – Stress level is one of the effective factors in discriminating Vietnamese tones.

4.3.2. Results

This section attempts to describe the data collected from the perception test. The description begins by listing the success rate of tone discrimination within the gathered data in general as well as with respect to the individual respondents. The next task is the comparison of success rates across dialects of stimuli as well as respondents. After that we focus on the overall success rates of the individual tones that are subsequently broken down based on the respondents' dialects. Then the attention is drawn to the influence of stress levels on the recognition of tones. The last task of this section is to determine confusion patterns in the discrimination of tones in order to establish groups of tones prone to mutual confusion.

The overall success rate of all respondents combined was 53%., which means that the respondents misjudged almost every second trial. If we look at the Hanoian and Saigonese respondents separately, the people from Hanoi scored the average 53.9% whereas the people

from HCMC 52%. Although the figure 53% is well above the chance level considering the fact that the respondents had to discriminate between six tones, it seems rather low since the research concerns the field of communication where the message must be transmitted extremely accurately in order for communication to carry on smoothly.

Fig 4.26. breaks down the success rate according to the individual respondents. We can notice that the respondents from Hanoi (marked in black) are more compact spanning from 43.1% to 62.5% whereas the respondents from the HCM City (marked in white) exhibit greater dispersion with the extremes at 38.2% and 65.3%. In other words, the respondent with the poorest judgement identified correctly a bit over one in three trials and the most successful respondent identified correctly roughly two out of every three trials. The chart indicates that the Hanoian respondents tend to be more consistent whereas respondents from the HCMC show a large degree of variation. However, it must be noted that a sample of 9 individuals from each dialect might not be a sound representation of the whole population.

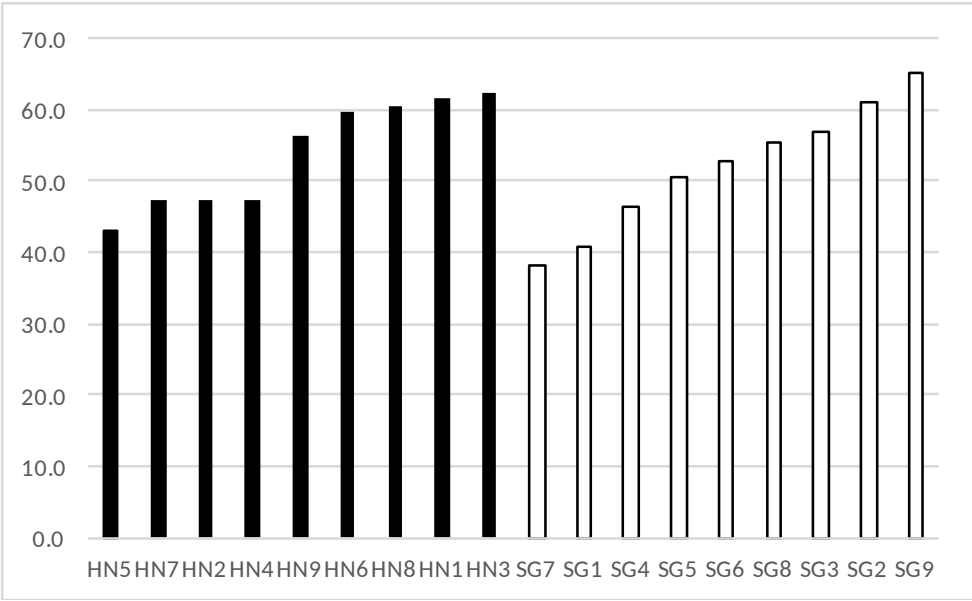


Figure 4.26. *Overall success rate of Vietnamese tones discrimination displaying the respondents from both dialects individually. Respondents from Hanoi are marked black while respondents from the HCMC are marked white.*

In Fig. 4.27. below, we can see how the success rate changed when the respondents listened to stimuli in the same dialect as their own and in the other dialect. When the respondents listened to their own dialect, their success rate was 54.4% for the Hanoian speakers, and 57.1% for the Saigonese speakers, i.e., above the overall average of both groups. When the respondents were played the other dialect, the success rate dropped to 53.5% for the respondents from Hanoi, and 46.9% for the respondents from Saigon. As opposed to Hanoi, where the difference is merely 1%, the success rate of the Saigonese respondents dropped by more than 10% when they were discriminating among tones of the Hanoian dialect.

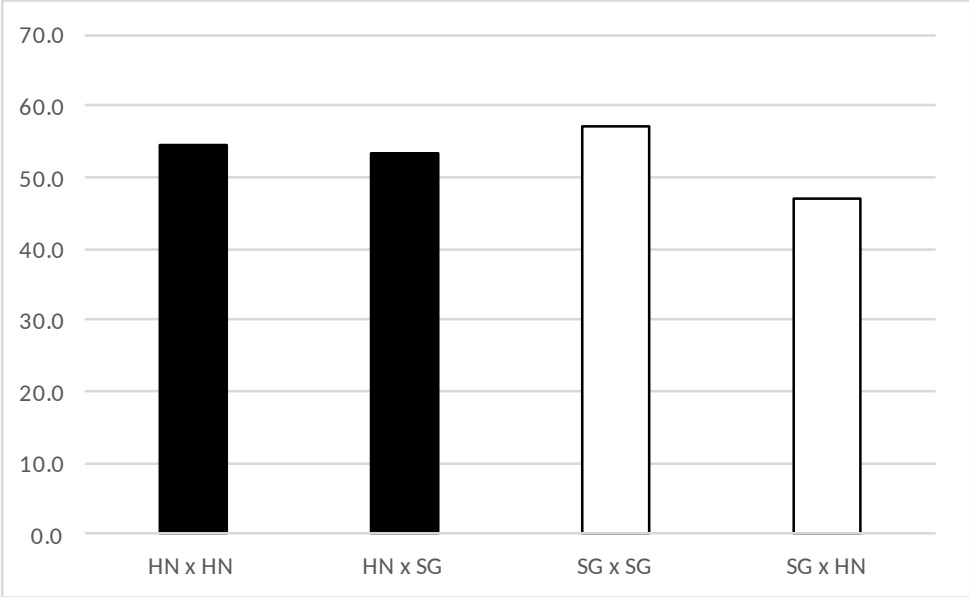


Figure 4.27. *Success rate of respondents within and across dialects. Hanoian speakers listening to Hanoian and Saigonese stimuli, Saigonese speakers listening to Saigonese and Hanoian stimuli.*

Fig. 4.28. illustrates the overall success rates across the individual tones. The low glottalized tone B2 manifests the highest success rate of 71.5% followed by the centrally glottalized fall-rise C1 with 66.2%, rising B1 with 65% and high level A1 with 63.4% success rate. The low falling tone A2 is identified with significantly lower success of 36.1% and the fall-rise C2 manifests the success rate of merely 15.5%, which could be classified as borderline chance level.

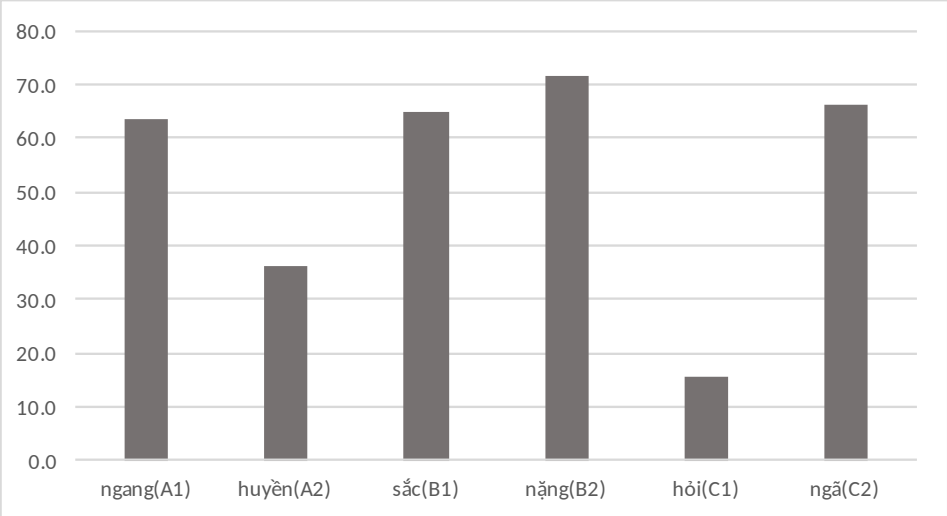


Figure 4.28. *Overall success rate of the 6 individual tones labelled by the Vietnamese syllables and also by the alphanumeric norm adopted by the international academia.*

Fig. 4.29. breaks down the overall success rate results according to the dialects of the respondents. The respondents from HCMC were more accurate at discriminating only the tone B2 (0.5% difference) and the tone A2 where the difference was more significant, 4.6%. Tones A1, B1, C1 and C2 were discriminated more accurately by the respondents from Hanoi. The success rate difference in A1, B1 and C2 is not very large but the tone C1 showed a difference of 7.8%, which seems more significant especially concerning the generally low SR of C1 discrimination.

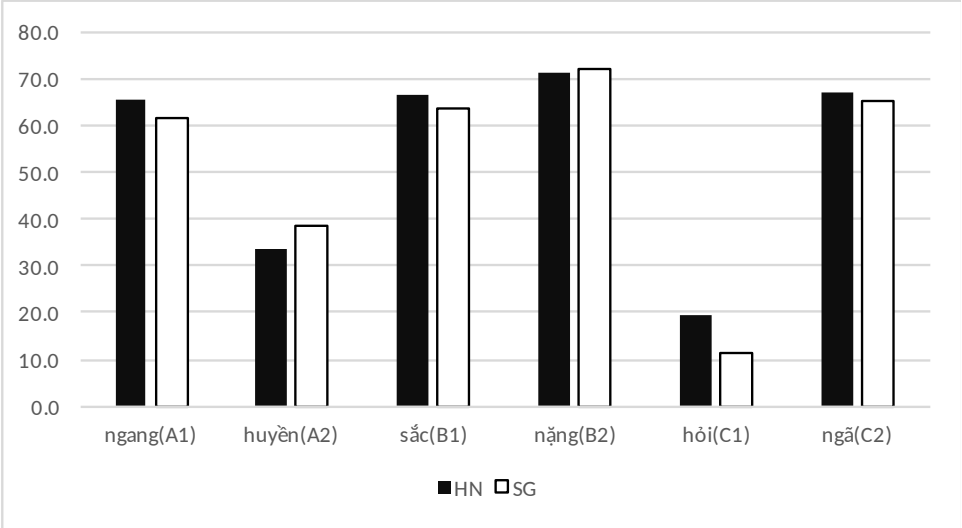


Figure 4.29. *Overall success rates of the individual tones with respect to the respondents' dialect.*

Stress level seems to play a significant role in tone discrimination as the success rate increases substantially in heavily stressed syllables as can be observed in Fig. 4.30. The overall success rate of both dialects combined in syllables carrying weak stress was merely 42.5% rising to 72.3% in the syllables on the highest stress level increasing the success rate by 29.8%.

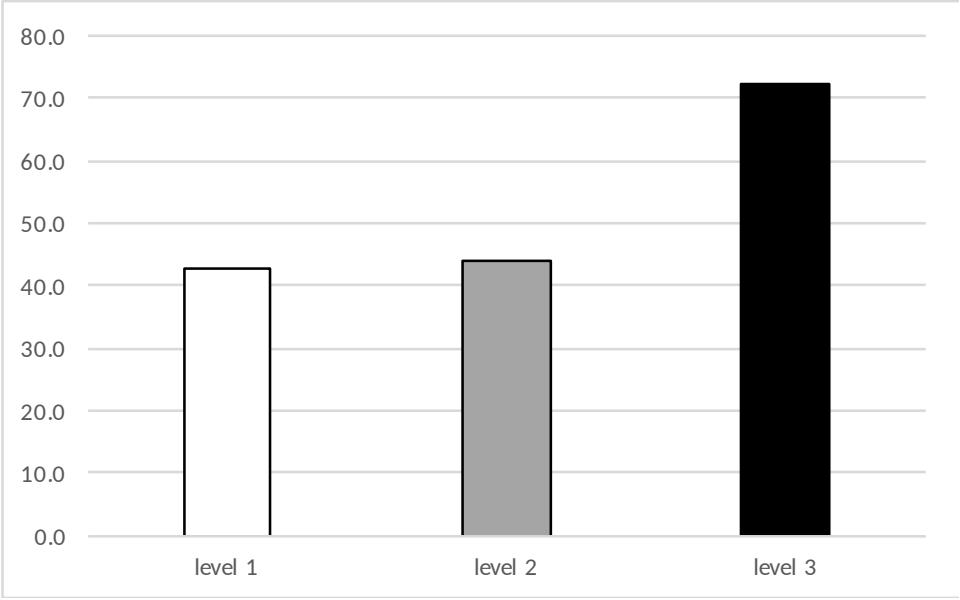


Figure 4.30. *Success rate in respect to the stress level (level 3 being the most prominent).*

Fig. 4.31. breaks down the success rate based on the individual tones within the three stress levels. It seems very likely that the increase in success rate is directly proportional to stress level. The heaviest stress level (marked red) manifested the highest success rate in all tones. In five out of six tones, the success rate of the highest stress level nearly reached or even surpassed 80%. Even in case of the tone C1 that has produced generally very low success rate, the figure nearly doubled from 13.2% to 23.6%, which is still very low compared to the other tones. The situation of tones A2 and B2 are also noteworthy. The success rate of A2 at the stress level 1 (marked blue) was the lowest of all the tones with 11.1% but it increased more than seven times to 79.8% at the stress level 3 (marked red). The tone B2, on the other hand, had a very high success rate even at the first stress level and its increase was very small by merely 2.8% at the third level.

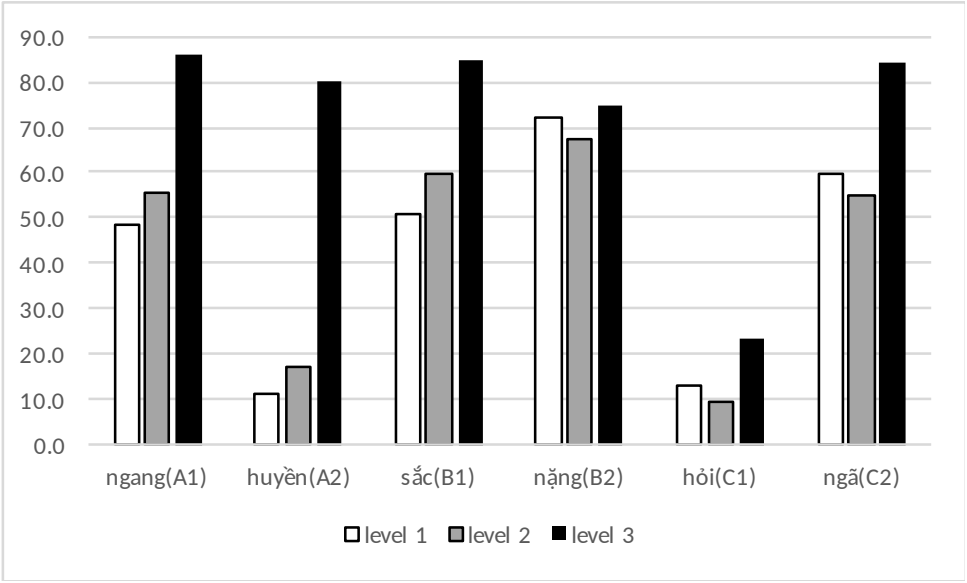


Figure 4.31. *Success rate of the individual tones in respect to the stress level (blue being the weakest level and red the most prominent).*

The confusion matrix in Fig. 4.32. captures confusion patterns among the tones. The matrix juxtaposes expected (E) and observed (O) tones. In an ideal scenario, if the respondents guessed everything correctly, there should be the number 100.0 in the grey box and 0.0 in all the others. If we look at the line *ngang* in Fig. 4.32., we can see that the tone *ngang* was identified as *ngang* in 63.4% of cases, *huyền* in 6% *ngã* in 0.5%, *hỏi* in 0.0%, *sắc* in 6% and *nặng* in 24.1%. It further shows that the tone *nặng* (B2) was the most frequent source of misjudgement. It was chosen as an answer more often than any other tone that was played to the listeners. All the tones with the exception of *ngã* (C2) expressed the confusion rate with *nặng* (B2) from almost 20% in case of *hỏi* (C1) to 45.4% in case of *huyền* (A2). The tone *hỏi* (C1) was confused with the tone *ngã* (C2) in 39.4% trials. The tone *sắc* (B1), a high rising tone, was confused with *nặng* (B2), a low glottalized tone, in 23.1% trials, which is definitely noteworthy as the two tones are supposed to have very contrasting contours.

E \ O	**Ngang**	**huyền**	**ngã**	**hỏi**	**sắc**	**nặng**
ngang	63.4	6.0	0.5	0.0	6.0	24.1
huyền	9.7	36.1	3.9	2.8	2.1	45.4
ngã	0.9	0.5	66.9	6.7	18.3	6.7
hỏi	3.9	12.3	39.4	15.5	9.7	19.2
sắc	3.0	0.5	6.7	1.6	65.0	23.1
nặng	1.6	3.5	11.6	4.4	7.4	71.5

Figure 4.32. *Confusion matrix of percentages of judgements for all data. E – expected, O – observed.*

E \ O	**ngang**	**Huyền**	**ngã**	**hỏi**	**sắc**	**nặng**
ngang	65.3	5.1	0.5	0.0	8.8	20.4
huyền	9.3	33.8	2.3	3.2	2.8	48.6
ngã	0.0	0.0	68.1	2.8	25.9	3.2
hỏi	6.0	6.9	33.8	19.4	16.2	17.6
sắc	3.7	0.9	4.2	0.9	66.7	23.6
nặng	0.9	3.2	9.3	3.7	11.6	71.3

Figure 4.33. *Confusion matrix of percentages of judgements for Hanoian respondents. E – expected, O – observed.*

E \ O	**Ngang**	**huyền**	**ngã**	**hỏi**	**sắc**	**nặng**
ngang	61.6	6.9	0.5	0.0	3.2	27.8
huyền	10.2	38.4	5.6	2.3	1.4	42.1
ngã	1.9	0.9	65.7	10.6	10.6	10.2
hỏi	1.9	17.6	44.9	11.6	3.2	20.8
sắc	2.3	0.0	9.3	2.3	63.4	22.7
nặng	2.3	3.7	13.9	5.1	3.2	71.8

Figure 4.34. *Confusion matrix of percentages of judgements for Saigonese respondents. E – expected, O – observed.*

4.3.3. Discussion

Most of the subjects apparently struggled while performing the test. On the one hand, they were afraid to make a mistake and usually used all replay possibilities. They were also rather disgruntled by the fact that they could not see the results immediately after each step, which

would go rather against the nature of the experiment as revealing correct answers throughout the test would artificially boost their performance. These issues might have been caused by the Vietnamese education system that is very result-oriented and making mistakes is considered highly undesirable. On the other hand, it was quite difficult for the subjects to perceive the tone as a separate category that merely manifests itself on a syllable. It seems more likely that the speakers think about the tonal variations as a set of separate lexemes more than a syllable with the potential of tonal combinations. Non-native speakers of Vietnamese clearly nest the syllables together, hence storing together lexemes with no semantic relation. However, such a strategy seems to be unnatural for native speakers. If we were to come up with an example from the English language, we could perceive lexemes like *bed* x *bet* or *fit* x *fish* as variants of one form. Although there might be some EFL potential for vocabulary memorization, the native speakers would likely deem such a strategy rather odd and they would retreat to semantic categorization or categorization of frequency in usage.

The overall success rate is very low with merely 53% although it rises to 72.2% for the most prominent stress level. Nevertheless, the figures suggest that an average Vietnamese speaker generally understands only half of all lexemes he/she hears out of context or not even three quarters if the syllables are uttered carefully. This observation indicates that context must be a decisive factor in discriminating Vietnamese tones. Moreover, the findings of section 4.2. also support this claim.

In terms of the dialects and individual speakers, Hanoian speakers proved to be more consistent as their success rates varied to a lesser extent than in the case of Saigonese speakers. Hanoian speakers were not as successful in discrimination of tones within their own dialect but their success rate in discriminating tones in the opposite dialect was higher

by almost 7% in comparison with the Saigonese speakers. Better success rate in the opposite dialect is likely to be caused by the fact that Hanoian tone inventory is more complex and it is therefore easier to discriminate within a simpler inventory than vice versa.

The by far lowest success rate was observed on the tone C1 (hỏi) with only 15.5% on average. This phenomenon was probably cause by three factors. Firstly, the tone C1 is considered one of the two most marked tones with most complex contour in Vietnamese (although the results of section 4.1. yielded only a simple falling contour). Secondly, tones C1 and C2 are perceptionally homophonous in the Saigonese dialect. Therefore, when the Saigonese respondents were choosing the label C1 and C2 for syllables uttered by Saigonese speakers, they might have been doing it by chance. The claim is supported by the average confusion rate of C1 with C2 amounting to 39.4%. Finally, the tones C1 and C2 in the Saigonese dialect have a contour similar to the Hanoian C2. Therefore, Hanoian respondents might have been tempted to classify all C1 and C2 tones uttered by Saigonese speakers as C2 and the Saigonese respondents might have been unfamiliar with Hanoian C1, which might have led to a generally high error rate.

On the other hand, the tone B2 (nặng) had the overall success rate of 71.5%, which, however, seems a little misleading due to the confusion rate of other tones with B2. In general, nearly a half of A2 tones was misclassified as B2, almost one fourth of A1s and B1s was also classified as B2 and even one fifth of C1s was erroneously taken for the tone B2. In sections 4.1. and 4.2., it was already mentioned that the tone B2 has generally the shortest duration. This feature could have misled the respondents into thinking that various tones uttered on lower stress levels was in fact the tone B2.

5. Conclusion

It can be concluded that the two main ambitions of this dissertation have been achieved. Firstly, we managed to devise a method of F_0 representation suitable for description of the Hanoian as well as the Saigonese dialect. Secondly, we created a corpus of almost 4000 tokens from 12 speakers for each dialect and added a voluminous database containing a large bulk of metadata categorized on: dialect, speaker, gender, tone, stress level, style, context, duration and F_0 measurements. By doing so, we hopefully rendered the analysed data more comparable to the real-life linguistic situation.

Results gathered from the preselected syllables and syllables in isolation largely conformed to the findings of the previous studies discussed in Section 2.2. We confirmed Pham's hypothesis of 8-tone classification of Hanoian tone inventory because the contours of B1 x D1 and B2 x D2 indeed turned out differently (even if from the systemic point of view, the idea of allotonicity should not be ruled out).

The same analysis of the Saigonese tone inventory showed a different setting. The tonal pairs in the Saigonese dialect have almost identical contour that resembles the one of Hanoian D1 x D2. We also confirmed that the contours of Saigonese tones C1 and C2 are shaped very similarly although C1 seems to be slightly lower. However, the shape seems to be superior to height as the tone C1 and C2 in HCMC are perceptually identical. Saigonese tones are also generally positioned farther from each other than the tones in Hanoi, which supports the claim by Brunelle (2009a) that discrimination of Saigonese tones depends on F_0 differences more than in the case of Hanoian tones. The findings concerning Hanoian B2 and C1 have also been previously discussed but only tentatively. Although B2 is taken for a low and/or falling tone, its average F_0 in this study turned out to be level and only slightly

lower than high level A1. In the case of C1, it has been claimed that its contour has lost its final rise in colloquial speech but our research suggest that the rise has been lost even in careful speech. It is also noteworthy that its duration in careful annunciation is significantly shorter than of other tones with the exception of B2, D1 and D2.

The analysis of the effects of the speech style and sentential stress on tone contours indicates that both categories influence the tones in an analogical way in both dialects. With weakening stress and more spontaneous speech, the tones tend to be flatter in contour, shorter and converging to the speaker's average F_0. Figure 4.25. even suggests that the rising tone B1 in Hanoi, when occurring on syllables with weak stress or in semi-spontaneous speech, loses its properties and becomes a low and level tone below the contours of A1 and B2. In terms of duration, all tones in both dialects proportionally decrease their duration with weakening of stress and/or increasing the speech rate with the exception of Hanoian C1 that becomes equally long as A2. In this setting C1 and A2 in Hanoi are very similar in term of contour as well as duration and they might be easily confused, which partly manifested itself in Section 4.3.

The perception test was not performed utilizing as much data as tasks 4.1. and 4.2. but it nevertheless also led to certain notable observations. Firstly, respondents were having difficulties in the test administration process. It seems that abstracting tones from syllables is something that the native speakers of Vietnamese are not very comfortable with, which might have been a partial cause of their rather poor performance. What the test results reliably indicate is that the respondents were more successful discriminating tones in their own dialects. Furthermore, the Hanoian respondents scored better with stimuli from Saigonese speakers that Saigonese speakers with Hanoian stimuli. This fact can be explained by higher complexity of the tone inventory in Hanoi. Saigonese speakers using a simpler

tonal inventory have greater difficulties with orientation within a more complex system. The test also indicated that Hanoian C1 was quite often misjudged as A2, which could be used as a supportive argument for possible merging of A2 and C1 in the Hanoian dialect in the future.

Finally, it is necessary to say that this dissertation has merely skimmed off the top of Vietnamese tonality as there are many more Vietnamese dialects that have not been addressed. Moreover, even the issues that have been addressed could be investigated in greater detail but that would be far beyond the scope of this dissertation. There are, however, certain issues that deserve to be addressed in the near future. Firstly, the method of F_0 representation should be adjusted so that it also reflected duration of the individual tones. Secondly, subsequent research should be devoted to quantitative analysis of voice quality cues for tonal discrimination. Finally, the perception test should be replicated with more data and under more controlled conditions to ensure that the surprisingly low success rates reflect reality and they are not merely a manifestation of a flaw in the research method.

6. Bibliography

Abramson, A. (1975) The Tones of Central Thai: Some Perceptual Experiments. In Harris and Chamberlain (eds.), *Studies in Thai Linguistics*. Bangkok: Central Institute of English Language, 1-16.

Boersma, P. & Weenink, D. (2017). *Praat: Doing Phonetics by Computer*.

Brunelle, M. (2003). Tone Coarticulation in Northern Vietnamese. *Proceedings of the 15th International Congress of Phonetic Sciences*. 2673-2676.

Brunelle, M. (2009a). Northern and Southern Vietnamese Tone Coarticulation: A Comparative Case Study. *Journal of the Southeast Asian Linguistics Society*.

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

Brunelle, M. et al. (2012). *Intonation in Northern Vietnamese.* University of Ottawa Press.

Brunelle, M. & Phương Hạ, Kiều & Grice, M. (2016). Inconspicuous coarticulation: A complex path to sound change in the tone system of Hanoi Vietnamese. *Journal of Phonetics*. 59.

Cao Xuân Hạo. (2010). *Tiếng Việt, mấy vấn đề ngữ âm, ngữ pháp, ngữ nghĩa* (Vietnamese, Selected Problems from Phonetics, Grammar and Semantics). Hà Nội. NXB Giáo dục.

Chao, Y. R. (1930). A system of tone letters. *Le maître phonétique 45*, 24-27.

Chao, Y.R. (1968). *A grammar of spoken Chinese.* Berkeley.

Chaudhary, C. C. (1983). Word stress in Vietnamese: A preliminary investigation. *Indian Linguistics 44*, 1-10.

Chen, G-T. (1974). The pitch range of English and Chinese speakers, Journal of Chinese Linguistics 2, 159–171.

Clumeck, H. (1980). The Acquisition of Tone. In Yeni-komshian, Kavanagh and Ferguson (eds.). *Child Phonology, Vol. I, Production.* New York: Academic Press. 257-275.

Cruttenden, A. (1997). Intonation. Cambridge: Cambridge University Press.

Crystal, D. (1969). *Prosodic Systems and Intonation in English*. Cambridge: CUP.

Cunningham, U. (2009). Phonetic Correlates of Unintelligibility in Vietnamese-accented English. *Proceedings, Fonetik 2009*. University of Stockholm.

Cutler, A. and Chen, H. C. (1997). Lexical Tone in Cantonese: Spoken-word Processing. *Perception and Psychophysics*. Dallas.

Čermák, F. (2004). *Jazyk a jazykověda*. Praha: Karolinum.

Dediu, D. & Ladd, R.D. (2007). Linguistic tone is related to the population frequency of the adaptive haplogroups of two brain size genes, *ASPM* and Microcephalin. *Proceedings of the National Academy of Sciences Jun 2007*, 104 (26)

Deutsch, D., Dooley, K., Henthorn, T., and Head, B. (2009). Absolute pitch among students in an American music conservatory: association with tone language fluency, *Journal of Acoustical Society of America 125*. 2398–2403.

Đoàn Thiện Thuật. (1977). *Ngữ âm tiếng Việt* (Vietnamese Phonetics). Hà Nội: NXB Đại học và trung học chuyên nghiệp.

Diffloth, G. (1989). Proto-Austroasiatic Creaky Voice. *Mon-Khmer Studies 15*:.139-54.

Everett, C. et al. (2015). Climate, Vocal Folds, and Tonal Languages: Connecting the Physiological and Geographic Dots. *Proceedings of the National Academy of Sciences,* 112. 1322-27.

Ewan, W. G. (1976). *Laryngeal Behavior in Speech*. Ph.D. Dissertation. University of California, Berkeley.

Ferlus, M. (2004). The origin of tones in Viet-Muong, in *Papers from the Eleventh Annual Meeting of the Southeast Asian Linguistics Society*, S. Burusphat (ed), Arizona, pp. 297-313.

Fu, Q. J. and Zeng, F. G. (2000). Identification of Temporal Envelope Cues in Chinese Tone Recognition. *Asia Pacific Journal of Speech, Language and Hearing 5*. 45-57.

Gordina, M. V. & Bystrov, I. S. (1984). *Фонетический строй вьетнамского языка* (Phonetic Structure of the Vietnamese Language). Moskva: Nauka.

Greenberg, S. and Zee, E. (1979). On the Perception of Contour Tones. *UCLA Working Papers in Phonetics 45*. 150-165.

Gussenhoven, C. (2004). *The Phonology of Tone and Intonation*. Cambridge: Cambridge University.

Harris, M.S. and N. Umeda (1987). Difference Limens for Fundamental Frequency Contours in Sentences. *Journal of the Acoustical Society of America 81*. 1139-1145.

Harrison, P. A. (2000). Acquiring the Phonology of Lexical Tone in Infancy. *Lingua 110*. 581-616.

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

Hayes, B. (1995). *Metrical Stress Theory: Principles and Case Studies*. The University of Chicago Press.

Healy, D. (2004). *Teach Yourself Vietnamese*. London: McGraw-Hill.

Hermes, D. J. 2006. Stylization of Pitch Contours. In W. de Gruyter (ed.) *Methods in Empirical Prosody Research*. ISBN 978-311018856-1.

Hillenbrand, J. M. and Gayvert, R. T. (2015). Phonetics exercises using the Alvin experiment-control software. *Journal of Speech, Language, and Hearing Research,* 1-14.

Hoàng Cao Cường. (1986). Suy Nghĩ Thêm về Thanh Điệu Tiếng Việt (More Thoughts about Vietnamese Tones). *Ngôn ngữ.* 3:19–38.

Hoàng Thị Châu. (1989). *Tiếng Việt trên các miền đất nước – Phương ngữ* (Vietnamese Across the Region – Dialectology). Hà Nội, NXB Khoa học xã hội.

Hombert, J. M., Ohala, J. J. and Ewan, W. G. (1979). Phonetic Explanations for the Development of Tones. *Language* 55: 37-58.

Hữu Quỳnh & Vương Lộc. (1980). *Khái quát về lịch sử tiếng Việt và ngữ âm tiếng Việt hiện đại* (Brief History of the Vietnamese Language and Modern Vietnamese Phonetics). Hà Nội: NXB Giáo Dục.

Johns-Lewis, C. (1986). Prosodic differentiation of discourse modes, in *Intonation in Discourse*, pp. 199–219.

Kirby, J. P. (2010). Dialect Experience in Vietnamese Tone Perception. *Journal of the Acoustical Society of America* 127(4), 3749-3757.

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

Klatt, D. (1973). Discrimination of Fundamental Frequency Contours in Synthetic Speech Duplications for Models of Pitch Perception. *Journal of the Acoustical Society of America 53*. 8-16.

Ladd, D. R. (1997). *Intonational phonology*. Cambridge: Cambridge University Press.

Law, S. P. (1990). *The Syntax and Phonology of Cantonese Sentence-final Particles*. PhD Thesis. Boston University.

Li, C. and Thompson, S. (1977). The Acquisition of Tone in Mandarin-speaking Children. *Journal of Child Language 4*. 185-199.

Maddieson, I. (1997). Phonetic universals. In *The Handbook of Phonetic Sciences*, ed. W. Hardcastle & J. Laver. Blackwell Publishers, Oxford. 619-639.

Maddieson, I. (2013). Tone. In: Dryer, Matthew S. & Haspelmath, Martin (eds.) *The World Atlas of Language Structures Online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. (Available online at http://wals.info/chapter/13, Accessed on 2018-05-21.)

Maeda, S. (1975). Electromyographic Study of Intonational Attributes. *Progress Report*. Research Laboratory of Electronic, MIT. 261-269.

Machač, P., Skarnitzl R. (2009). *Principles of Phonetic Segmentation.* Praha: Epocha.

Maspero, Henri. (1912). *Phonétique historique de la langue annamite: les initiales*. Bulletin de l'Ecole Française d'Extrême-Orient 12(1): 1-127.

Matisoff, J. A. (1973). Tonogenesis in Southeast Asia. Consonant Types and Tone, In: L. M. Hyman (ed.) *Southern California Occasional Papers in Linguistics*. Los Angeles: Linguistics Program University of Southern California.

McGuire, Grant. (2010). *A brief primer on experimental designs for speech perception research.* Laboratory Report. 77.

Michaud, A. (2004). Final Consonants and Glottalization: New Perspectives from Hanoi Vietnamese. *Phonetica 61*. 119–146.

Miller, G. A. (1956). The Magical Number Seven, Plus or Minus Two: Some Limits on our Capacity for Processing Information. *Psychological Review 63(2)*. 81-97.

Ngô Thanh Nhân. (1984). *The Syllabeme and Pattern of Word Formation in Vietnamese*. Ph.D. dissertation. New York University.

Nguyễn, V. L. and Edmondson, J. (1997). *Tones and voice quality in Modern Northern Vietnamese: Instrumental case studies*. Mon-Khmer Studies 28: 1-18.

Nguyễn, T. & Ingram, J. (2006). Stress, tone and word prosody in Vietnamese compounds. In: Paul Warren and Catherine I. Watson, Proceedings of the 11th Australasian International Conference on Speech Science & Technology. *Eleventh Australasian International Conference on Speech Science and Technology 2006*. Auckland, NZ (193-198). 6-8

Niebuhr, O., & Michaud, A. (2015). *Speech Data Acquisition - The Underestimated Challenge*. Kieler Arbeiten in Linguistik und Phonetik (KALIPHO), 3, 1-42.

Nooteboom, S. (1999). The Prosody of Speech: Melody and Rhythm. *The Handbook of Phonetic Science*. Blackwell Reference Online.

Ohala, J. J. (1972). The physiology of tone. In: L. M. Hyman (ed.), *Consonant types and tone. South California Occasional Papers in Linguistics* (Univ. of So. Calif.) 1.1-14.

Ohala, J.J. (1978). Production of tone. In Fromkin, V.A. (ed.), Tone: a linguistic survey 5-39.

O'Grady, W., Dobrovolsky, M., & Aronoff, M. (1997). *Contemporary linguistics: an introduction.* NY: St. Martin's Press.

Palková, Z. (1994). *Fonetika a fonologie češtiny*. Praha: Karolinum.

Peiros, Ilia. (1998). *Comparative Linguistics in Southeast Asia.* Pacific Linguistics Series C, No. 142. Canberra: Australian National University.

Peng, S. H. (1997). Production and Perception of Taiwanese Tones in Different Tonal and Prosodic Context. *Journal of Phonetics 25*. 371-400.

Peng, S. H. (2000). Lexical versus 'Phonological' Representations of Mandarin Sandhi Tones. *Acquisition and the Lexicon: Papers in Laboratory Phonology V*. Cambridge: CUP. 152-167.

Peterson, G. E. & Barney, H. L. (1952): Control methods used in a study of the vowels. *Journal of the Acoustical Society of America* 24: 175–184.

Phạm, A. H. (2003). *Vietnamese Tone: A New Analysis*. New York: Routlege.

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

Pierrehumbert, J. (1980) *The Phonology and Phonetics of English Intonation*. PhD thesis, MIT. Distributed 1988, Indiana University Linguistics Club.

Pike, K. (1948). *Tone languages*. Ann Arbor: The University of Michigan Press.

Pollák, P., Volín, J. & Skarnitzl, R. (2007). HMM-based phonetic segmentation in Praat environment. In: *Proceedings of XIIth "Speech and Computer – SPECOM 2007"*, 537–541.

Pollack, I. (1952). "The Information of Elementary Auditory Displays". *Journal of the Acoustical Society of America 24.6*. 745-749.

Sidwell, Paul. (2009). *Classifying the Austroasiatic languages: History and the State of the Art*. LINCOM studies in Asian linguistics, 76. Munich: Lincom Europa.

Skalička, V. (2004). *Souborné dílo, I. díl.* (Collected Works I). Praha: Karolinum.

Skalička, V. (2004). *Souborné dílo, II. díl.* (Collected Works II). Praha: Karolinum.

Slavická, B. (2008). *Praktická fonetika vietnamštiny*. Praha: Karolinum.

Smith, N. V. (1968). Tone in Ewe. *MIT Research Laboratory of Electronic Quarterly Progress Report 88*. 290-304.

Sundberg, Johan. (1973). Data on maximum speed of pitch changes. *Quarterly Progress and Status Report 14*. Stockholm: Speech Transmission Laboratory. 39–47.

Takefuta, Y., Jancosek, E. G., and Brunt, M. (1972). A statistical analysis of melody curves in the intonation of American English, in *Proceedings of the 7th International Congress of Phonetic Sciences*, Montreal 1971, 1035–1039.

Thomas, Erik R. (2002). Sociophonetic Applications of Speech Perception Experiments. *American Speech, Volume 77, Number 2*. Duke University Press. pp. 115-147.

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

Trần Trí Dõi. 2011. *Giáo trình lịch sử tiếng Việt* (Textbook to the History of the Vietnamese Language). Hà Nội: NXB giáo dục Việt Nam.

Vũ, T. P. (1982). Phonetic Properties of Vietnamese Tones across dialects. In: D. Bradley (ed.), Papers in Southeast Asian Linguistics. 55-75. Sydney, Australian National University.

Xu, Y. (1994). Production and Perception of Coarticulated Tones. *Journal of Acoustical Society of America 95.4*. 2240-2253.

Xu, Y. (1999a). F_0 Peak Delay: When, Where, and Why It Occurs. In Ohala, J. (ed.), *International Congress of Phonetic Sciences 1999*. San Francisco. 1881-4.

Xu, Y. (1999b). Effects of Tone and Focus on the Formation and Alignment of F0 Contours. *Journal of Phonetics 27*. 55-105.

Yip, M. (2002). *Tone*. Cambridge: CUP.

Zhang, J. (2000). Phonetic duration effects on contour tone distribution. In M. Hirotani, A. Coetzee, N. Hall, and J-Y. Kim (eds.), Proceedings of the 30th annual meeting of the North East Linguistic Society (NELS 30). GLSA Publications, Amherst, MA.

Appendix 1

Reading:

Bác Hùng ngã xe đạp, khiến da bác ấy bị trầy nhẹ. Không để ý gì tới đường đi, bác vừa lái xe, vừa ăn kem. Túi ngô và mấy củ khoai ngọt rơi lăn lóc trên đường.
Đúng lúc đó, một thầy bói mù đi qua. Vì thiếu ngủ, ông ngáp. Chân trượt vào bắp ngô, ông ngã ra đường như đang vồ ếch. Mặt đập xuống đất làm đỏ ửng má. Ông cứ ngồi ở đấy mà ăn vạ:
„Mẹ cái bắp ngô, đúng là của nợ!“
Trên vỉa hè, bà bán khoai và ngô vừa ngáp vừa phe phẩy cái quạt trông có vẻ mệt mỏi. Nghe thấy tiếng chửi, bà bực mình quay ra:
„Mẹ cha thằng mù, bà đã cho ăn rồi mà còn chửi hàng bà! Bà mổ cả gia đình nhà mày!“
Ông thầy bói cãi lại:
„Chim chíc mà ghẹo bồ nông. Đến khi nó mổ, lạy ông xin chừa!“
Bà như bị sỉ nhục, tay cầm dép, ném vào ông kia. Dép trúng má, làm chảy máu và sây sát thêm. Một con chó nhân vụ kiện, lẳng lặng chạy ra đánh hơi. Thấy khoai rơi trên đường có vị ngọt như mật ong, nó liếm nhẹ, ngoạm lấy vào mõm và chạy mất.

Transliteration for Prague Labeller:

bak hum nga se dap, chien za bak ej bi čaj ně. Chom de í gí toj duong di, bak vủa lái sé, vủa an kém. Tủj ngo vá maj kủ choáj ngot zoj lan lok čén duong. Dum luk dó, mot thaj bói mủ di kvá. Ví thiu ngủ om ngap. Čan čuot vao bap ngo. om nga za duong nhu dang vo eť. Mat dap xuong dat lam do ung má. Om kủ ngoj o daj ma an va: Me káj bap ngo, dum la kua no. Čen via he, ba bán choaj va ngo vua ngap vua fe faj kái kvat čom kó ve met moj. Ngé thaj tyeng čủj, ba buk miň kvaj zá: Me čá thang mu, ba da čó an roj ma kon čủj hang ba. Ba mo ka zá diň ňá maj. Om thaj boi

kai laj. Čím čík ma geo bo nong. Den chi no mo, laj om sin čua. Ba nhu bi si ňuk, taj kam zep, nem vao om kia. Zep čum ma, lam čaj mau va saj sat them. Mot kon čo ňan vu kien, lang lang čaj ra danh hoi. Thay khoaj zoi čen duong kó vi ngot nhu mat om, no liem ňe, ngoam laj vao mom va čaj mat.