Perception of lexical tones by homeland and heritage speakers of Cantonese

by

Wai Man Lam

B. A., The Chinese University of Hong Kong, 2004

M. Phil., The Chinese University of Hong Kong, 2007

A THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES
(Linguistics)

The University of British Columbia
(Vancouver)

November 2018

The following individuals certify that they have read, and recommend to the Faculty of Graduate and Postdoctoral Studies for acceptance, the dissertation entitled:

Perception of lexical tones by homeland and heritage speakers of Cantonese

submitted by Wai Man Lam in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Linguistics.

Examining Committee:

Kathleen Currie Hall, Department of Linguistics
Co-supervisor

Douglas Pulleyblank, Department of Linguistics
Co-supervisor

Molly Babel, Department of Linguistics
Supervisory Committee Member

Valter Ciocca, School of Audiology and Speech Sciences
University Examiner

Márton Sóskuthy, Department of Linguistics
University Examiner

Abstract

This dissertation compares the lexical tone perception abilities of two populations with different bilingual configurations: Cantonese-dominant adults who grew up in Hong Kong (referred to as *homeland speakers*), and English-dominant adults who grew up in a Cantonese-speaking household in Canada (*heritage speakers*). From infancy both were exposed to Cantonese as a first language in terms of chronological order; however, after the onset of schooling, each became dominant in the majority language of their respective society. Given this background, this study investigates whether heritage speakers' perception of lexical tones of a non-dominant first language (Cantonese) exhibits cross-language effects from a dominant second language (English) that does not have a contrastive dimension of tone.

A series of perception experiments was conducted using the word identification paradigm. Eight types of audio stimuli were presented to homeland and heritage speakers (N=34 per group), each of which represented a specific configuration of four variables: whether the acoustic signal contained *segmental* and *tonal* information, whether the target word was isolated or embedded in a carrier sentence with *semantic context*, and whether the meaning of the target word was *congruous* with the carrier sentence. In each trial, participants saw pictures of the target word and minimally contrastive tonal competitors, and were instructed to choose the picture that represented what they heard.

Major findings of this study were: (1) among the eight stimulus types, the accuracy gap between the two groups was the biggest when the

stimuli were low-pass-filtered monosyllables with no segmental information or semantic context, which suggests that homeland speakers have a significantly greater ability to identify tonally contrastive words by solely relying on tonal information. (2) Both groups showed confusion of overlapping subsets of tone pairs, but heritage speakers had a higher error percentage, which indicates a quantitative but not qualitative difference between the two groups. (3) When the target word was semantically incongruous with the carrier sentence, homeland speakers outperformed heritage speakers by attending to acoustic information, while heritage speakers relied on semantic information relatively more often. In other words, the two groups used different listening strategies in tone identification.

Lay Summary

The purpose of this research is to investigate effects of post-childhood linguistic experience on bilingual speakers' ability to perceive speech sounds of a first language that uses pitch variation to distinguish word meaning. Sixty-eight young adults raised by Cantonese-speaking parents performed a listening task. Cantonese words and sentences were played and participants were instructed to select a picture that represented what they heard. Half of the subjects grew up in Hong Kong, and the other half grew up in Canada and generally felt more comfortable with English. Although both groups were exposed to Cantonese from birth, the Hong Kong group was better at paying attention to the pitch of a word, while the Canadian group relied on the overall meaning of a sentence relatively more often. These results inform us how multilingual competence works in the human mind, especially in a world where migration is common.

Preface

This dissertation is an original intellectual product of the author, Wai Man Lam. The name Zoe Wai-Man Lam is also used in other published works by the author.

All projects and associated methods were approved by the Behavioural Research Ethics Board of the University of British Columbia [certificate #H16-00297]. All data collection took place in the Speech in Context Laboratory at the University of British Columbia, Vancouver.

A preliminary version of Chapter 1, Chapter 4, and Chapter 5 was presented at the 29th North American Conference on Chinese Linguistics held at Rutgers University, New Jersey, in June 2017.

Table of Contents

List of Tables

List of Figures

The LSHK Cantonese Romanization Scheme

Consonants

IPA	Jyutping
p	b
$\mathrm{p^h}$	p
m	m
f	f
t	d
$\mathrm{t^h}$	t
n	n
l	l
ts	z
$\mathrm{ts^h}$	c
s	s
j	j
k	g
$\mathrm{k^h}$	k
ŋ	ng
h	h
$\mathrm{k^w}$	gw
$\mathrm{k^{wh}}$	kw
w	w

Vowels and glides

IPA	Jyutping
iː	i
e	i
yː	yu
uː	u
o	u
ɛː	e
ɔː	o
œː	oe
ɵ	eo
ɐ	a
aː	aa
iw	iu
ɵɥ	eoi
uɥ	ui
ej	ei
ɔj	oi
ɐj	ai
ɐw	au
aːj	aai
aːw	aau

Tones

IPA		Jyutping
55 (5)	˥	1
25	˨˥	2
33 (3)	˧	3
21	˨˩	4
23	˨˧	5
22 (2)	˨	6

List of Abbreviations

BLP	Bilingual Language Profile
f0	Fundamental frequency
F2	Second formant
GLMM	Generalized logistic mixed model
Hz	Hertz
IPA	International Phonetic Alphabet
L1	First language (based on the order of acquisition)
L2	Second language (based on the order of acquisition)
LOC	Linguistics Outside the Classroom
LSHK	The Linguistic Society of Hong Kong
ms	Milliseconds
UBC	The University of British Columbia
VOT	Voice onset time

Glossary

bilingual	a person who uses two languages in everyday life, but does not necessarily have an equal mastery of listening, speaking, reading, and writing skills for both languages
Bilingual Language Profile	a questionnaire developed by Birdsong, Gertken & Amengual (2012) to elicit language background information and assess language dominance on a gradient scale using four criteria: language history, language use, language proficiency, and language attitudes
dominant language	a person's default or preferred language for speaking and thinking
heritage language	any language that has a cultural connection to an individual's family or community, but is not the primary language used in government, education, and public communication
heritage speaker	a bilingual who was exposed to a heritage language early in life, but later became dominant in the majority language of the society

homeland speaker	a person whose default or preferred language for speaking and thinking is the language that s/he was exposed to early in life; during the period of exposure, this language not only has a cultural connection to his/her family or community, but is also the language of primary use in government, education, and public communication
lexical tone	the use of pitch to distinguish meaning on the word level
majority language	the primary language used in government, education, and public communication in a society
merger	a sound change whereby two or more contrastive phonological categories are replaced by a single category
tonal perception	the process of extracting relevant auditory cues from a continuous speech signal and mapping pitch attributes to discrete phonological categories

Note: These are basic definitions. The complexities of these terms are discussed in Chapter 2 and Chapter 3.

Acknowledgments

This dissertation was completed thanks to the contribution of numerous mentors, colleagues, and friends. I owe my deepest gratitude to my co-supervisors Kathleen Currie Hall and Douglas Pulleyblank. I could not thank you enough for guiding me through the early stage of formulating the methodology of this study. Your input had made *all* the difference. I thoroughly enjoyed every meeting of ours, not only because of the intellectual stimulation (also known as "grilling"), but also because of your sense of humour, which never failed to give me a warm boost of morale. Kathleen's speedy *and* meticulous comments on my drafts were immensely helpful. The knowledge that I have gained was far beyond statistical analysis—in retrospect every small step paved the way to becoming a better scientist with logical and critical thinking. Doug's insightful feedback reminded me to step back and look at the big picture whenever I was struggling to get the details right. I am extremely grateful for the encouragement that I received when I felt I was not smart enough. From daily interactions with Kathleen and Doug I have learned that academia is not just about being smart—what's more important are hard work, dedication, and humility. Thank you both so much for these life lessons.

I would like to express my greatest appreciation to Molly Babel of my dissertation committee, who piqued my interest in speech perception and sociolinguistic variation. Your seminars on perceptual adaptation, attention and salience, and heritage linguistics played a major role in inspiring me to pursue this dissertation topic. Your passion and enthusiasm shone through your comments on my writing. Thank you for your comprehensive advice and thoughtful suggestions.

My funding sources certainly deserve acknowledgment. This study was supported by awards and scholarships from the University of British Columbia and a research grant awarded to Douglas Pulleyblank by the Social Sciences and Humanities Research Council of Canada [#435-2016-0369].

Throughout my graduate career I had benefited from the wisdom of many professors. I am indebted to Martina Wiltschko, who taught me how to be an effective writer. I am appreciative of Hotze Rullmann's skills of communicating abstract ideas in the clearest way possible. My heartfelt thanks go to Ping Jiang, Virginia Yip, Robert Bauer, and Cathy Wong, who introduced me to the fascinating world of linguistics and encouraged me to pursue a PhD. I would like to acknowledge invaluable advice from Marjorie Chan, who generously spent time to meet me during conferences. I am tremendously fortunate to have met Henry Yu, whose research on Chinese Canadian history gave me an interdisciplinary perspective on my research.

I would not have been able to finish my PhD without peer support. It was an unforgettably humbling experience to study and work with Blake Allen, Joash Johannes, Adriana Osa-Gómez, Oksana Tkachman, Sihwei Chen, Roger Lo, Michael Fry, Andrei Anghelescu, and Natalie Weber at UBC Linguistics. I am also grateful for the moral support from Sweden Xiao, Michelle Chen, Kamila Kolpashnikova, Tomoharu Hirota, Irene Setiawan, Hyunju Kwon, Nathaniel Lim, Foong Yen Chong, Denise Chan, Siu Pong Cheng, and Hoi Wing Chan. Thank you all for giving me a push whenever I needed it.

Special thanks to staff, friends, and students at UBC St John's College, UBC Cantonese Language Program, UBC Asian Canadian and Asian Migration Studies, UBC Graduate Pathways to Success Program, as well as members of the Vancouver Chinatown community. You gave me momentum to make academic research accessible to the general public.

[To my family] 最後我衷心感激家人無限的支持和包容，給我空間和自由追尋自己的夢想。這些年來陪伴你們的時間少了，但你們對我的關愛不僅沒有減退，反而與日俱增。你們永遠是我每天奮鬥的動力。

For my parents
謹以此文獻給先父及家母

Chapter 1

Introduction

Migration of linguistic communities has led to various language contact phenomena throughout history: cross-continental trade gave rise to pidgins and creoles (Holm, 1989); early maritime explorers brought loanwords into their language out of the need to communicate (Cutler, 2000), and colonization and empire created varieties of English around the globe (Schneider, 2007). This dissertation is generally about *immigration*—the relevant keyword for the current century—and its byproduct, a population called *heritage speakers*. The general questions asked in this study are: what happens when children of Cantonese-speaking immigrants grow up in a predominantly English-speaking society? As they enter adulthood, does English affect the way they perceive speech sounds in Cantonese? In particular, will their extensive exposure to a language that does not use pitch to distinguish word meaning (English) affect their perception of a language that uses pitch to distinguish word meaning (Cantonese)?

1.1 Cantonese in Canada

A linguistically diverse country like Canada is an ideal place to study language contact induced by immigration. While English and French are official languages, 207 other languages were reported as a mother tongue in the 2016 Census (Statistics Canada, 2017e). These 207 non-official languages can be divided into two groups: 67 of them are Aboriginal languages, which are traditionally spoken by the First Nations, the Métis, and the Inuit peoples. The other 140 non-official languages are non-Aboriginal. According to the 2016 Census, 7.3 million people, which is 21.1% of the Canadian population, speak a non-official and non-Aboriginal language at home (Statistics Canada, 2017a).

Among these 140 non-official and non-Aboriginal languages, Cantonese is the second most-spoken home language in Canada with a population of 594,705 (see Table 1.1). In the Metro Vancouver area, Cantonese is the mother tongue of a population of 193,030 (see Table 1.2). Many of these Cantonese speakers came from Hong Kong in the 1980s and 1990s. As Table 1.3 shows, Hong Kong was the top origin of immigrants in 1991 and 1996 (Statistics Canada, 2009). These were the years around major political incidents, such as the Tiananmen Massacre in 1989, and the handover of Hong Kong's sovereignty from the United Kingdom to the People's Republic of China in 1997 (Li, 2005; Wong, 1999). In the period of 1980–2006, the total number of people who emigrated from Hong Kong to Canada was 215,430 (Statistics Canada, 2012).

1.2 Heritage speakers of Cantonese

The focus of this dissertation is individuals who grew up in Canada being raised by Cantonese-speaking parents who had emigrated from Hong Kong. For the purpose of the current study, these individuals are hereafter referred to as *heritage speakers*, and Cantonese is considered their *heritage language*. These terms have long been used in discussions on the social politics of multilingualism, language education and pedagogy, especially in North America (Cummins & Danesi, 1990; Kagan & Dillon, 2001; Peyton, Ranard & McGinnis, 2001, among others). In recent years, they have also been used in linguistics (Benmamoun, Montrul & Polinsky, 2010, 2013b; Nagy, 2015, among others). Definitions vary from source to source, but in the most general sense, heritage languages are "languages other than the de facto dominant language in a given social context" (Kelleher, 2010, p.1).

Table 1.1: Top five non-official languages spoken at home in Canada in the 2016 Census (Statistics Canada, 2017c)

	Language	Population
1.	Mandarin	641,100
2.	Cantonese	594,705
3.	Punjabi	568,375
4.	Spanish	553,495
5.	Tagalog	525,375

Table 1.2: Top five non-official languages spoken at home in Metro Vancouver in the 2016 Census (Statistics Canada, 2017b)

	Language	Population
1.	Cantonese	193,030
2.	Mandarin	180,170
3.	Punjabi	163,400
4.	Tagalog	78,830
5.	Korean	47,920

Table 1.3: Top 10 countries of birth of recent immigrants, 1981–2006 (Statistics Canada, 2009)

	1981	1991	1996	2001	2006
1.	U.K.	Hong Kong	Hong Kong	P.R. China	P.R. China
2.	Vietnam	Poland	P.R. China	India	India
3.	U.S.A.	P.R. China	India	Philippines	Philippines
4.	India	India	Philippines	Pakistan	Pakistan
5.	Philippines	Philippines	Sri Lanka	Hong Kong	U.S.A.
6.	Jamaica	U.K.	Poland	Iran	S. Korea
7.	Hong Kong	Vietnam	Taiwan	Taiwan	Romania
8.	Portugal	U.S.A.	Vietnam	U.S.A.	Iran
9.	Taiwan	Lebanon	U.S.A.	S. Korea	U.K.
10.	P.R. China	Portugal	U.K.	Sri Lanka	Colombia

For a detailed discussion on the definitions of heritage languages and their speakers from a linguistic perspective, see Chapter 2.

The term *heritage language* can also be defined by what it is not. First, a heritage language is not "foreign" to its speakers due to personal or cultural connections. Second, an ethnolinguistic community may be a "minority" according to official definitions but not in a sociological sense if its economic power is taken into consideration. For example, Chinese is officially a visible minority according to the Government of Canada (Statistics Canada, 2017d). However, in the past two decades, Chinese immigrants to Canada have brought along wealth and labour, and so China is regarded as a "titanic economic power" that has impacted the economy of Metro Vancouver (Barnes & Hutton, 2016, p.11). In addition, languages of visible minorities may be the most-spoken mother tongues in certain regions, and so they are not "minority" in a numerical sense either. For example, in Richmond, British Columbia, 44.8% of residents speak a variety of Chinese[1] as their mother tongue, and 33.1% speak English as their mother tongue (City of Richmond, 2017, p.4). Taken together, "minority language" is not an appropriate label from both socioeconomic and numerical perspectives.

Lastly, specifically for the context of Cantonese in Metro Vancouver, the term "immigrant language" could be controversial. According to the Government of Canada, immigrant languages in Canada are defined as those "whose presence is initially due to immigration *after* English and French colonization" [italics by author] (Statistics Canada, 2017b, para. 2). However, languages from southern China, including varieties of Cantonese, had been brought to the Pacific Northwest initially by miners and railroad builders since the first half of the 19th century (Yu, 2011) *before* British Columbia joined Confederation on 20 July, 1871 (Ormsby, 1958). It is thus questionable whether Cantonese meets this definition if it had been spoken before the region became part of Canada. Moreover, the relation between colonial settlers and Aboriginal peoples is a complex issue. Unlike other parts of Canada, 95% of British Columbia is on the *unceded territory*

[1]In the report by the City of Richmond, varieties of Chinese such as Cantonese and Mandarin are categorized into one group, namely "Chinese".

of the First Nations peoples (City of Vancouver, 2014), which means the land has never been surrendered or given away to the colonizers. It is therefore debatable who is an "immigrant" on Aboriginal lands. Since this dissertation focuses on the linguistic but not the political aspect of the issue, the term *heritage language* is used throughout the subsequent discussion to avoid tangential controversies.

Anecdotal reports point out that some heritage speakers cannot speak Cantonese fluently even though their parents do. Usually they can express themselves better in the majority language of the society, which is English in the context of Metro Vancouver. Below are three Vancouver-based media stories, in which the interviewees' Cantonese proficiency varies from semi-fluent to mere passive knowledge:

> I usually won't be shy [to speak Cantonese] because if the person I'm talking to only speaks Cantonese, I think the person would appreciate... even if I have an accent, or can't fully express myself. But if we are talking about my younger brother... I think he might not even try. (Cheong & Lee, 2015)

> Doris Chow understands the words but can't speak Cantonese, the first language she learned in life, fluently. It's a language that is slowly disappearing from Vancouver but the city's Chinatown is one place where Cantonese still rules... (Li, 2016, para. 1)

> I am Russell... I am Cantonese myself, but I can't really speak the language... My parents having both immigrated from Hong Kong... Like a lot of other Chinese Canadians, I went to Chinese school from a young age, but since I only spoke English with my parents, I failed to pick up the language that well. My family's loss of a language over just one generation intrigued me. (Chiong et al., 2017)

The phenomena associated with heritage speakers are not only intriguing on a personal level to Russell from the third story above, but also to linguists from an academic perspective. How are heritage speakers different from Cantonese speakers who grew up in Hong Kong? Since Cantonese tones are notoriously difficult for second language learners (Boyle, 1997), do heritage speakers face similar difficulties despite having early exposure to Cantonese at a young age? These general questions will be narrowed down in the next section.

1.3 Goals of this dissertation

This dissertation investigates how Cantonese lexical tones are perceived by two populations of Cantonese-English bilinguals with varying configurations of language dominance, as shown in Figure 1.1. The first population, referred to as *homeland speakers*, consists of Cantonese speakers who grew up in Hong Kong, where English is taught in school as a second language. They are commonly considered the prototypical native speakers. The second population, known as *heritage speakers*, refers to individuals who grew up in a Cantonese-speaking household in Canada, a country on the other side of the Pacific Ocean. These speakers were exposed to Cantonese in a family setting from early childhood. Up to that point their development of Cantonese was similar to that of homeland speakers. However, after the onset of schooling, English started to become their most frequently used language in day-to-day situations. As a result, heritage speakers could be more comfortable with using English than Cantonese. Particularly of note for this study is that Cantonese is a tone language, while English is not. This raises an interesting question: as Cantonese is gradually becoming the weaker language of heritage speakers, does their ability to perceive Cantonese tonal contrasts also become weaker as they enter adulthood?

Given such differences in exposure and domains of language use between homeland and heritage speakers, the current study aims at answering the following questions:

(1) Research questions of the current study

a. Do homeland and heritage speakers behave differently in terms of their ability to identify tonally contrastive words?

b. Do homeland and heritage speakers exhibit similar confusion patterns with respect to lexical tone perception?

c. Do homeland and heritage speakers make use of the same type of information when identifying a word from a tonally contrastive set? In particular, are acoustic and semantic information equally useful?

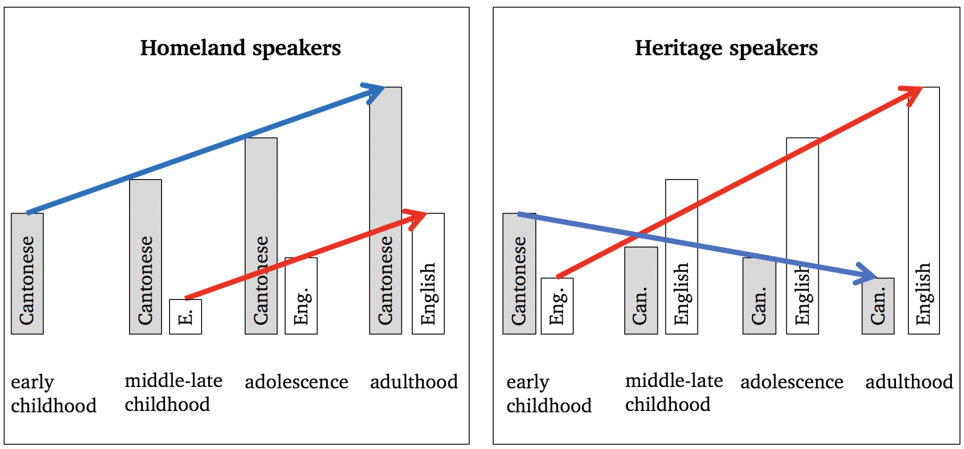


Figure 1.1: Language development of homeland and heritage speakers (adapted from Montrul, 2012)

A personal anecdote that inspired me to ask these questions could be a good illustration of the linguistic phenomenon in question. As a homeland speaker who lives in Canada as an international student, I often interact with heritage speakers born and raised in Canada. At a dinner gathering in Vancouver Chinatown, I was ordering food in Cantonese for a group of heritage speakers. One of the dishes was called *gon1 bin1 sei3 gwai3 dau2* "stir-fried green beans", in which *sei3 gwai3* means "four seasons", as in (2a)[2]. Both *sei3* and *gwai3* are produced with Tone 3, the mid level tone. When I said the name of this dish, one of my friends could not help laughing.

[2]Cantonese words and sentences in this dissertation are romanized following the LSHK Cantonese Romanization Scheme unless otherwise specified. For details, see the explanation in Section 1.4.6.

Seeing my puzzled face, he said in English, "Did you just say *goddamn beans*? That's so funny." After processing his reaction for five seconds, I realized that he probably heard something like (2b). *Sei2 gwai2* "goddamn" is produced with Tone 2, the high rising tone. These two phrases involve two tonally contrastive minimal pairs: the first pair is *sei3* "four" and *sei2* "die", and the second pair is *gwai3* "season" and *gwai2* "ghost".

(2) a. *gon1 bin1* ***sei3 gwai3*** *dau2*
dry stir-fry **four season** bean

'stir-fried **green** beans'

b. *gon1 bin1* ***sei2 gwai2*** *dau2*
dry stir-fry **die ghost** bean

'stir-fried **goddamn** beans'

It never occurred to me, someone from Hong Kong, that these two phrases could make a good pun, because the mid level tone and the high rising tone are perceptually quite distinct to me. My friend, however, must have perceived them as similar sounds to find it funny. In other words, there must be some differences between the perception of tonally contrastive minimal pairs by me (a homeland speaker) and by my friend (a heritage speaker). Using this example, I exemplify my research questions in (3):

(3) An instantiation of the research questions in (1)

a. Are both populations equally able to identify *sei2* "die" and *sei3* "four"?

b. Is it possible that certain minimal pairs (e.g. *sei3-sei2*) are more confusing for one population than for the other?

c. Consider a semantically congruous sentence such as *jat1 nin4 jau5* <u>*sei3*</u> *gwai3* "there are <u>four</u> seasons in a year". If both populations are able to identify the underlined target word as *sei3* "four" but not *sei2* "die", are they using the same information to reach this conclusion? Do they rely on the acoustic cues (i.e. what they heard) or the semantic information (i.e. what makes sense)?

1.4 The Cantonese language

This section introduces aspects of the Cantonese language that are necessary for understanding this dissertation. It addresses three questions. First, where is Cantonese spoken? The second question is asked twice with a different emphasis: is Cantonese *a* language (or multiple languages)? Is Cantonese a *language* (or dialect)? Lastly, what are the linguistic features of this language?

1.4.1 Origin and spread

Cantonese, also known as *Yue* (粵), belongs to the Sinitic branch of the Sino-Tibetan language family. Originating from southern China, it is named after *Canton* (or *Guangzhou* in Mandarin), the capital city of the Guangdong province (Chao, 1947). According to Crissman (2012), Cantonese is spoken by 59.58 million people in the province of Guangdong, the Guangxi Zhuang Autonomous Region, the Hong Kong Special Administrative Region, and the Macau Special Administrative Region of the People's Republic of China. Since migrants from Guangdong dominated trans-Pacific labour migrations from the 19th to early 20th century (Yu, 2011), Cantonese communities can also be found in Malaysia, Singapore, Vietnam, Thailand, Brunei, Indonesia, the Philippines, the Netherlands, Australia, Canada and the United States (cf. Chau, 2011; Clyne & Kipp, 1997; Hashimoto, 1972; Lewis, Simons & Fennig, 2009; Tan, 2005). The total linguistic population worldwide is estimated to be 73.76 million (Lewis et al., 2009).

1.4.2 Is Cantonese a language?

The term *Cantonese* is ambiguous since it can be used in either a broad or narrow sense in the literature. As an umbrella term it refers to the Yue dialect group as a whole. According to Wurm, Li & Baumann (1987), Yue can be classified into seven subgroups: Guangfu (廣府), Siyi (四邑), Gaoyang (高陽), Goulou (勾漏), Wuhua (吳化), Qinlian (欽廉) and Yongxun (邕潯). Their geographical distribution is shown on the map in Figure 1.2 alongside other non-Yue Sinitic languages, namely Min, Hakka, Ping and Mandarin. For detailed discussions on the phonological differences and grammatical diversity across these subgroups, see Hashimoto (1972) and Kwok, Chin & Tsou (2016) respectively.

Among the seven subgroups, Siyi and Guangfu are most relevant to the discussion of overseas Cantonese communities. Historically, according to Yu (2011), 80% of the Chinese migrants to Canada between 1910 and 1923 were from the counties of Taishan (台山), Kaiping (開平), Xinhui (新會), and Enping (恩平), all of which fall into Region 2 in Figure 1.2, where Siyi (四邑, literally "four counties") Cantonese is spoken. Significant Siyi Cantonese communities can also be found in continental United States due to a similar migration history (Chao, 1947). However, in Honolulu, Hawaii, most Chinese migrants were from Zhongshan (中山) (Yu, 2011), which falls into the area of the Guangfu subgroup. In the 1980–1990s, migrants from Hong Kong to Canada (the focus of this dissertation), the United States and Australia mostly spoke Guangfu Cantonese, which is marked as Region 1 in Figure 1.2. Therefore in sociohistorical discussions of overseas Cantonese communities, the general term "Cantonese" does not always refer to the same language variety.

Linguistically, the Siyi and Guangfu varieties are not entirely mutually intelligible, although they share cognates, syntactic and phonological features. According to a listening comprehension test by Szeto (2000), spoken Taishanese (which belongs to the Siyi subgroup) is 31.3% intelligible to speakers of Hong Kong Cantonese (which belongs to the Guangfu subgroup).

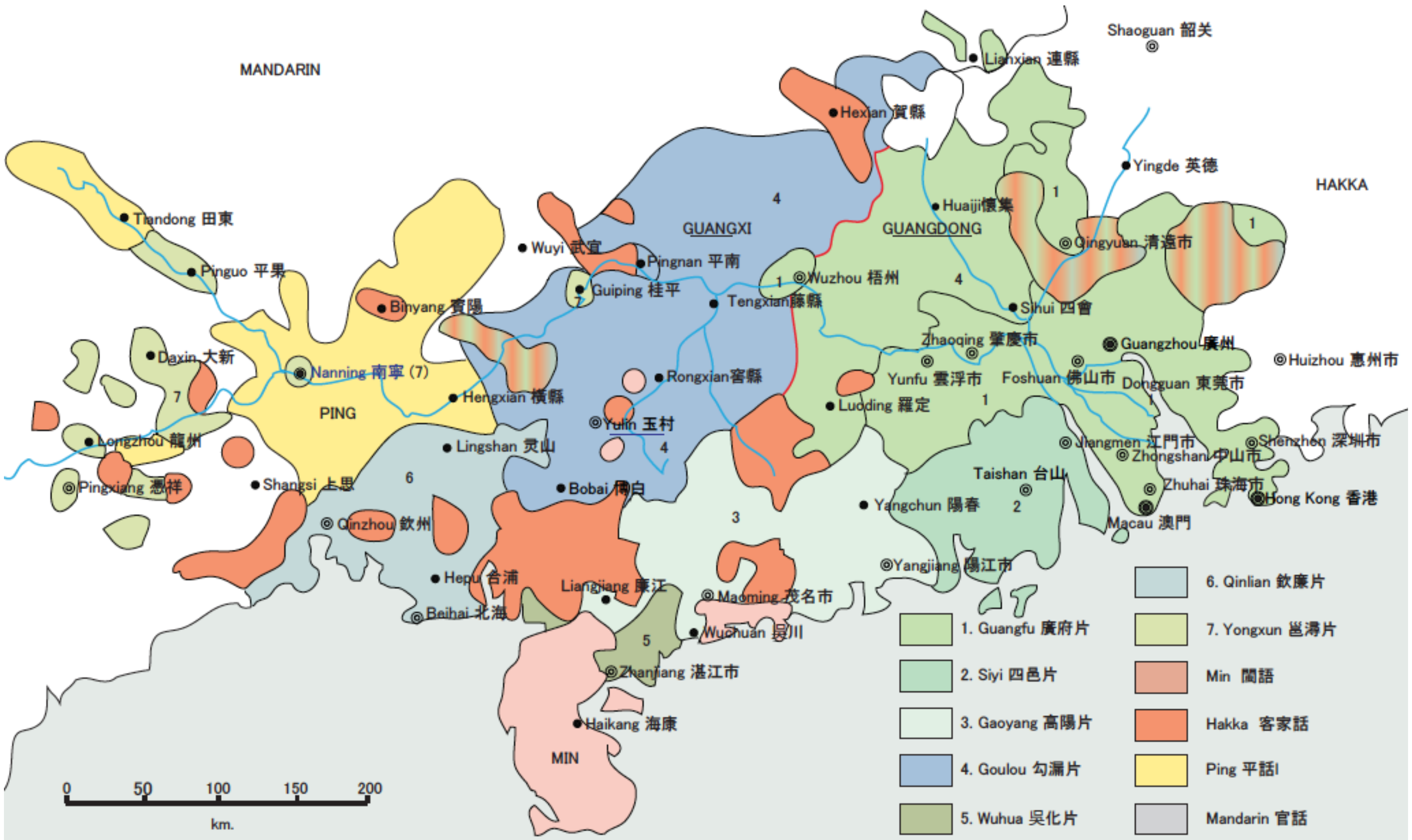


Figure 1.2: Geographical distribution of Yue subgroups based on Wurm et al. (1987)
Image from Iacoponi (2012), used under Creative Commons Attribution 3.0 Unported License

When used in its narrow sense, *Cantonese* refers to the specific variety spoken in Guangzhou, Hong Kong and Macau[3] within the Guangfu subgroup. It is considered a "genuine regional standard" (Ramsey, 1987, p.99) that enjoys a more prestigious social status compared with other varieties of Yue. It is not only the lingua franca for doing business in Southern China (Bauer & Benedict, 1997), but also the language used in television programs produced by Television Broadcasts Limited, the major exporter of Cantonese popular culture from Hong Kong to diaspora communities all over the world (To & Lau, 1995). This variety is thus called *Standard Cantonese* in many linguistic studies (Fung, 2000; Lee, 1993; Yu, 2007, among others), even though there has never been a language standardization agency for Cantonese, nor is it the official language of any independent nation.

Although Guangzhou is the origin of Standard Cantonese, its status as the centre of Cantonese language and culture has shifted away to Hong Kong since the 1950s (Bauer, 2016). The first reason is the adoption of Mandarin, officially known as *Putonghua* (普通話, literally "common speech"), as the only national language of the People's Republic of China. Second, the influx of non-Cantonese-speaking workers from other provinces in China weakened the dominance of Cantonese in Guangzhou. Before 1997 Hong Kong was a British colony and so Cantonese was able to maintain its status as the city's de facto official language, even though it has never been a de jure one (Bauer & Benedict, 1997). Considering the pervasive use of Cantonese in various domains (e.g., government, education, mass media, cultural industry), Bolton (2011) even calls Hong Kong the "Cantonese-speaking capital of the world" (p.9).

Spoken Cantonese, in both its broad and narrow sense, is mutually unintelligible to Mandarin (Bauer, 2016; Bauer & Benedict, 1997; Cheng

[3]The varieties spoken in these three regions are very similar. However, due to its colonial history, Hong Kong Cantonese is characterized by as many as 700 loanwords from English (Wong, Bauer & Lam, 2009), which played a role in expanding the language's syllabary (Bauer, 1985), setting Hong Kong Cantonese apart from Guangzhou or Macau Cantonese. For a detailed explanation of the phonetic and phonological differences between these varieties, see Gui (2005).

& Tang, 2016a; Matthews & Yip, 2013; Olson, 1998). Both Cantonese and Mandarin belong to the Sinitic branch of the Sino-Tibetan family, but they have different sound systems, vocabulary, and morphosyntax (Cheng & Tang, 2016a; Tang & Cheng, 2014). Despite the linguistic differences, Cantonese is often considered a Chinese dialect by the general public (including the speakers themselves) due to similarities in the orthography, as well as sociocultural and political reasons. For a detailed discussion on the language versus dialect debate, see Cheng & Tang (2014). Some authors point out that the English term *dialect* is not an accurate translation for the Chinese term *fangyan* (方言, literally "regional speech"), and thus have proposed other terms such as *regionalect* (DeFrancis, 1986) and *topolect* (Mair, 1991). For the purpose of the current linguistic study, Cantonese is referred to as a language throughout this dissertation.

Since this study focuses on Cantonese speakers from Hong Kong and and heritage speakers whose parents are from Hong Kong, the relevant language variety is Hong Kong Cantonese, hereafter simply referred to as *Cantonese* unless otherwise specified. The following subsections will provide an introduction to the phonology of this language variety.

1.4.3 Segmental inventories

Since tone is the focus of this study, the segmental phonology of Cantonese will be introduced only briefly. There are 19 phonemic consonants in Cantonese (cf. Bauer & Benedict, 1997; Cheng & Tang, 2016a; Zee, 1991), as listed in Table 1.4[4]. For stops and affricates, aspirated and unaspirated consonants with the same place of articulation are contrastive. In other words, /p/ and /pʰ/ are two different phonemes, and so are /ts/ and /tsʰ/. All 19 consonants can occur syllable-initially, but only /p, t, k, m, n, ŋ/ can be syllable-final. When occupying the syllable-final position, /p, t, k/ are realized as unreleased [p̚, t̚, k̚] or as a glottal stop [ʔ]. Another characteristic of Cantonese is syllabic nasals. Only /m/ and /ŋ/ can stand alone as syllabic consonants [m̩] and [ŋ̩]. Lastly, semi-vowels /j,w/ can be

[4] On the consonant chart of Zee (1991), /l/ is described as denti-alveolar, while /ts, tsʰ/ are between alveolar and postalveolar. These are collapsed into "alveolar" in Table 1.4.

offglides. When /j/ is an offglide that follows a rounded non-front nuclear vowel /ɵ/ or /u/, it is phonetically realized as a labial-palatal approximant [ɥ].

Intertalker variation has been observed for the phonetic realization of these phonemes. First, the sibilants /s, ts, ts^h/ have free variants [ʃ, tʃ, $tʃ^h$] when followed by a rounded vowel. Some speakers pronounce /k^w/ as [k], /k^{wh}/ as [k^h], syllable-initial /n/ as [l], syllable-initial /ŋ/ as [ʔ], and syllable-final /ŋ/ as [n] (Bauer & Benedict, 1997; Cheung, 2007; Zee, 1999). Such variation is documented by linguists as ongoing mergers (Law, Fung & Bauer, 2001), but they are often stigmatized as "lazy" or "sloppy" pronunciation in the speaker community (Bauer, 2016).

Table 1.4: The phonemic consonant inventory of Cantonese

Manner of articulation	Place of articulation					
	Labial	Alveolar	Palatal	Velar	Labial-velar	glottal
Stop						
- Unaspirated	p	t		k	k^w	
- Aspirated	p^h	t^h		k^h	k^{wh}	
Nasal	m	n		ŋ		
Fricative	f	s				h
Affricate						
- Unaspirated		ts				
- Aspirated		ts^h				
Approximant		l	j		w	

The classification of phonemic and allophonic vowels is not as straightforward as the consonants, but most works agree that there are 11 vowel phonemes (Bauer & Benedict, 1997; Cheng & Tang, 2016a), which are shown in Figure 1.3[5]. Although vowel length is not contrastive, /iː, yː, uː, ɛː, ɔː, œː, aː/ are conventionally marked with an optional long diacritic

[5]Some works transcribe the vowels differently. For example, Hashimoto (1972) uses /ø/ instead of /ɵ/. Zee (1991) uses /ɪ/ instead of /e/, and /ʊ/ instead of /o/. Barrie (2003) uses /ʌ/ instead of /ɐ/.

to reflect their phonetic property (Zee, 1991). For a detailed discussion on different ways to analyze the vowel system, see Barrie (2003) and Bauer & Benedict (1997).

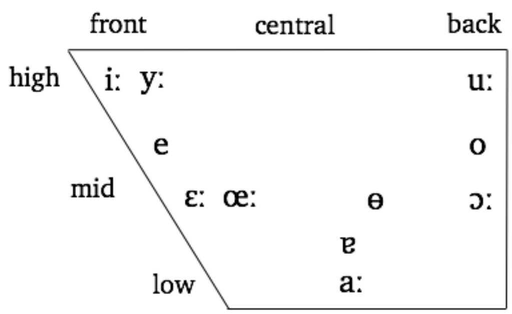


Figure 1.3: The phonemic vowel inventory of Cantonese

1.4.4 Syllable structure

A canonical Cantonese syllable is maximally CV(C), and syllabic nasals N can also be found. Consonant clusters are rare and they only occur in ideophones, onomatopoeia, and loanwords (Bauer, 1985). Gliding vowels [iw, ɵɥ, uɥ, ej, ɔj, ɐj, ɐw, aːj, aːw] are described as diphthongs in language textbooks (Matthews & Yip, 2013); however, since [j, w] are never followed by a consonant, these glides can be phonologically analyzed as coda consonants (Bauer & Benedict, 1997). Examples of each syllable type are listed in Table 1.5.

Table 1.5: Examples of Cantonese syllables

	Syllable type	Example	Gloss
CV	open syllable	[saː^{55}]	“sand”
CVC$_{[j,w]}$	closed syllable with an offglide	[saːj^{55}]	“waste”
CVC$_{[m,n,ŋ]}$	closed syllable with a nasal coda	[saːm^{55}]	“three”
CVC$_{[p,t,k]}$	closed syllable with an obstruent coda	[saːt^{3}]	“kill”
N	syllable with a nasal nucleus	[m̩^{21}]	“not”

The duration of syllables with an obstruent coda is found to be significantly shorter than those with a sonorant coda (Kao, 1971). This interacts with the tone system: tones of syllables ending with [p, t, k] have "checked" or "entering" tones, as in [saːt^3] in Table 1.5. This will be elaborated in the next subsection.

1.4.5 Tones[6]

Cantonese is a tone language, which means that changing the pitch of a word also changes the meaning of the word (Yip, 2002). For example, as shown in Table 1.6, when the syllable [sɛː] is produced with a high level pitch, it means "some", but when it is produced with a low level pitch, it means "shoot". The way it works is analogous to changing vowels in English: when the low vowel [æ] in "bad" [bæd] is changed to the low-mid vowel [ɛ], as in "bed" [bɛd], the meaning of the word changes completely from "the opposite of good" to "an object where we sleep on". In other words, just like consonants and vowels, tone is *contrastive*.

Table 1.6: The phonemic tone inventory of Cantonese; tone numerals are based on Bauer & Benedict (1997)

Tone	Description	Tone numerals	Example	Gloss
1	high level	55	[sɛː^{55}]	"some"
2	high rising	25	[sɛː^{25}]	"write"
3	mid level	33	[sɛː^{33}]	"diarrhea"
4	low falling	21	[sɛː^{21}]	"snake"
5	low rising	23	[sɛː^{23}]	"society"
6	low level	22	[sɛː^{22}]	"shoot"

There are six phonemic lexical tones in Cantonese (Bauer & Benedict, 1997), which are presented in Table 1.6. Each of their pitch contours is shown in Figure 1.4. The first tone (T1) is the high level[7] tone which marks

[6]This subsection is a basic introduction of Cantonese tones. A detailed literature review on Cantonese tonal perception will be presented in Chapter 3.

[7]Historically, T1 is a high falling tone [51] (Chao, 1947). Some Guangzhou speakers and older Hong Kong speakers still produce the high falling tone. Since the current study focuses on college-age individuals, the high falling tone is treated as a free variant of T1.

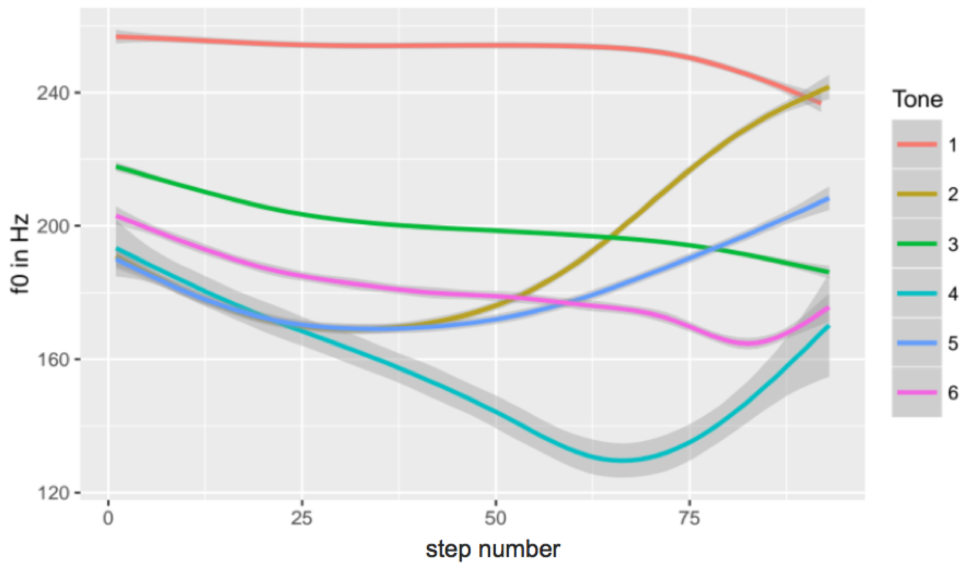


Figure 1.4: Pitch contours of the six phonemic tones produced by a female homeland speaker who participated in the current study. The x-axis represents 100 equally spaced steps in the vocalic portion of a syllable. Average f0 values were extracted from 162 word tokens. Shaded areas around each contour indicate values within a confidence interval of 95%.

the upper boundary of the tonal space. According to the tone numeral system originally developed by (Chao, 1947), T1 is represented as [55], meaning that it starts with a high pitch and also ends with a high pitch. In Figure 1.4, the ending pitch is slightly lower than the starting pitch arguably due to declination (Li, Lee & Qian, 2002; Wong, 2006), whereby a drop in subglottal pressure over the course of an utterance leads to a drop in fundamental frequency (Ladd, 1984; Lieberman, 1966). However, since the effect is phonetic, the ending pitch is still represented as [5].

The second tone (T2) is the high rising tone. Although Chao (1947) and earlier works represent it as [35], Bauer & Benedict (1997) provide phonetic evidence to show that the starting point of the high rising tone is as low as the starting point of the low rising tone [23]. They suggest that this contour tone may have undergone a sound change from 1947 to the 21st century.

This contour tone ends with a pitch that is as high as the ending pitch of T1. Therefore, the tone numerals for the high rising tone are [25].

The third tone (T3) is the mid level tone. Its starting pitch is higher than that of the low falling tone [21], low rising tone [23], and the low level tone [22], and so the tone numerals for T3 are [33]. Similar to T1, T3's ending pitch is slightly lower than its starting pitch due to phonetic declination effects. Its ending pitch is still represented as [3].

The fourth tone (T4) is the low falling tone, which marks the lower boundary of the tonal space. Its starting pitch is similar to that of the low rising [23] tone and low level tone [22]. Therefore, the tone numerals for T4 are [21]. Previous studies found that T4 is sometimes produced with creaky voice (Yu & Lam, 2014). They also point out that adding creaky voice to T4 enhances its accuracy rate in a tone identification task. In particular, it helps to distinguish T4 from T6, the low level tone.

The fifth tone (T5) is the low rising tone. It is represented as [13] in some works (e.g. Cheung, 2007), but is more commonly represented as [23] (e.g. Bauer & Benedict, 1997). As shown in Figure 1.4, the first 40% of its pitch contour is the same as T2 high rising. Both T2 and T5 have a rising contour, and they only differ by the magnitude of pitch change towards the end of the tone. T5 has a lower ending pitch than that of T2, and so its tone numerals are [23]. Ciocca & Lui (2003) found that the distinction between T2 and T5 is the hardest for children acquiring Cantonese. By age 10, most children are able to achieve an accuracy rate of 90% for most tone pairs, but for T2-T5, the accuracy rate is just 70%. It is also well documented that even adult native speakers are merging the two rising tones (Bauer, Cheung & Cheung, 2003; Fung & Wong, 2011; Mok, Zuo & Wong, 2013), which suggests an ongoing sound change in its initial stage. For a detailed discussion, see Section 3.3.2.

The sixth tone (T6) is the low level tone. Its starting pitch is a bit higher but still similar to that of T4 and T5, so the tone numerals for T6 are [22]. Similar to the other level tones, declination effects can be seen towards the end of the tone.

Apart from the six phonemic tones, Cantonese has three allotones, also known as *checked* or *entering tones* (Chao, 1947). They are T7 high checked [5], T8 mid checked [3], and T9 low checked [2]. They only occur in syllables with an obstruent coda [p, t, k]. Since they have a shorter duration (Zee, 1991), they are only represented by one-digit tone numerals, as in Table 1.7. These checked tones are allophonic to their respective level tone: the high checked tone (T7) is an allotone of the high level tone (T1), mid checked (T8) is an allotone of mid level (T3), and lastly, low checked (T9) is an allotone of low level (T6). These three allotones are not the focus of the present study, and none of the target words used in the experiment had checked tones. For a detailed discussion of the history and phonology of checked tones, see Bauer & Benedict (1997).

Table 1.7: Allotones in Cantonese; tone numerals are based on Bauer & Benedict (1997)

Tone	Description	Tone numeral	Example	Gloss
7	high checked	5	[sek^{5}]	"colour"
8	mid checked	3	[sɛk^{3}]	"kiss"
9	low checked	2	[sɛk^{2}]	"stone"

In general, there is no tone sandhi in Cantonese, except that certain morphological processes may change a word's tone into the high rising tone. An example of such processes is attenuative reduplication, which is illustrated in Table 1.8. None of the sentences used as stimuli in the current study involved these morphological processes. For details of morphologically induced tone change, see Bauer & Benedict (1997) and Yu (2007).

Table 1.8: Derived high rising tone in attenuative reduplication

Regular adjective	Gloss	Reduplicated form	Gloss
[k^{w}ɐj^{33}]	"costly"	[k^{w}ɐj^{33} k^{w}ɐj$^{\mathbf{25}}$ tej^{25}]	"a little bit costly"
[haːm^{21}]	"salty"	[haːm^{21} haːm$^{\mathbf{25}}$ tej^{25}]	"a little bit salty"
[lok^{22}]	"green"	[lok^{22} lok$^{\mathbf{25}}$ tej^{25}]	"a little bit green"

1.4.6 Romanization

Although numerous ways to romanize Cantonese have been proposed by missionaries and authors of language textbooks, the Jyutping (粵拼) system developed by the Linguistic Society of Hong Kong, officially known as the LSHK Cantonese Romanization Scheme (Tang et al., 2002), is the de facto standard in academia (Cheng & Tang, 2016b). For this reason Cantonese words and sentences of the current study are romanized following the Jyutping convention. Table 1.9 is a list of Jyutping symbols and their corresponding International Phonetic Alphabet (IPA) symbols.

Table 1.9: The LSHK Cantonese Romanization Scheme (Jyutping)

Consonants		**Vowels and glides**		**Tones**		
IPA	Jyutping	IPA	Jyutping	IPA		Jyutping
p	b	iː	i	55 (5)	˥	1
p^h	p	e	i	25	˧˥	2
m	m	yː	yu	33 (3)	˧	3
f	f	uː	u	21	˨˩	4
t	d	o	u	23	˨˧	5
t^h	t	ɛː	e	22 (2)	˨	6
n	n	ɔː	o			
l	l	œː	oe			
ts	z	ɵ	eo			
ts^h	c	ɐ	a			
s	s	aː	aa			
j	j	iw	iu			
k	g	ɵɥ	eoi			
k^h	k	uɥ	ui			
ŋ	ng	ej	ei			
h	h	ɔj	oi			
k^w	gw	ɐj	ai			
k^{wh}	kw	ɐw	au			
w	w	aːj	aai			
		aːw	aau			

Readers who are familiar with IPA but unfamiliar with Jyutping may pay attention to the following points. First, unaspirated stops and affricates [p, t, k, k^w, ts] are romanized as *b, d, g, gw, z*. Note that there are no voiced obstruents in Cantonese and the romanization does not reflect the actual voicing properties of these sounds. Second, syllable-initial [j] is also *j* in the romanization, but [j] as an offglide is romanized as *i*. Similarly, [w] is romanized as *u* when it is an offglide. Third, [ɐ] and [aː] are represented by *a* and *aa* respectively in Jyutping. Fourth, [iː] and [e] are both romanized as *i*, and [uː] and [o] are both romanized as *u*. This is because the two vowels in each pair are in complementary distribution: [e] and [o] only occur before velars, and [iː] and [uː] occur elsewhere. For a detailed analysis of these alternations, see Bauer & Benedict (1997). Since [e, o] are accounted for, the Jyutping symbol *e* and *o* actually represent [ɛː] and [ɔː] respectively. Last but not least, tones are represented by numbers. The romanization scheme does not distinguish checked tones from regular tones, since they can be predicted by the segmental environment. Therefore both 55 (high level) and 5 (high checked) are represented as *1*.

1.4.7 Writing system

Although the current study focuses on the spoken language, two concerns regarding the writing system need to be explained as they will be relevant in Chapter 4 with regard to research methodology.

The dominant writing script in Hong Kong is Traditional Chinese[8]. The non-alphabetic orthography is notoriously complex, as characters are learned as logograms (Baron & Strawson, 1976). In particular, 90% of Chinese characters are ideophonetic compounds, while 10% of them are pictographs or ideographs (Zhu, 1987). According to a corpus study of Chinese textbooks in Hong Kong, a primary school student will have been introduced to 2,570 to 3,844 characters by the end of the sixth school year (Chung & Leung, 2008). It is important to note that heritage speakers of Cantonese (or any variety of Chinese) do not acquire reading and writing skills by being exposed solely to the spoken language at home. Xiao (2006) compares the performance of non-heritage and heritage students (who

[8]In recent years, Simplified Chinese has become more visible due to the influx of tourists from mainland China (Choi, Liu, Pang & Chow, 2008). Readers may also note that Guangzhou Cantonese speakers write in Simplified Chinese.

were exposed to some variety of Chinese at home) in an intensive Chinese course in a New England university, and found that the heritage students did significantly better in speaking and listening than reading and writing. This point is crucial to some of the methodological decisions made for the present study, which are elaborated in Section 4.2.2 and Section 4.4.1.2. In particular, since not all heritage speakers have reading proficiency, pictures instead of Chinese characters were used.

The issue of writing is complicated by the fact that there is a distinction between two kinds of "written Cantonese". The first is simply writing Cantonese the way it is spoken, which often contains lexical items that are unfamiliar to Mandarin speakers. An example is provided in (4), which shows how "to be available" is expressed in spoken Cantonese. This way to write is common in informal contexts, such as texting with friends and family, or discussion on social media and internet forums. Snow (2004) comments that written Cantonese is somewhat stigmatized due to its connection with lower-class life. The second kind of "written Cantonese" is the way that educated people typically write, and refers to written Standard Chinese, which can then be read aloud with Cantonese pronunciation. In formal contexts such as official notices and academic writing, Standard Chinese grammar must be followed, and lexical items of Standard Chinese must be used, as in (5). When the text needs to be read aloud (e.g. in a Chinese class of a primary school that adopts Cantonese as the language of instruction), the characters would be pronounced with Cantonese pronunciation. For comparison, the Mandarin pronunciation of the same characters are shown in (6). The distinction between written Cantonese and Standard Chinese will be important in Section 4.2 regarding which words should be used for an experiment involving heritage speakers. For example, even though (5) can be considered a "Cantonese" word in formal texts, a heritage speaker may have never heard of it in the family setting, unless s/he had been to a Chinese school for formal language instruction. To ensure that heritage speakers actually know the words being used in the experiment, words from Standard Chinese like (5) were avoided.

(4) Written Cantonese being read in Cantonese: 得閒
dak1 *haan4*
obtain free-time

'to be available'

(5) Written Standard Chinese being read in Cantonese: 有空
jau5 *hung1*
have empty

'to be available'

(6) Written Standard Chinese being read in Mandarin: 有空
yǒu *kōng*
have empty

'to be available'

1.4.8 Summary

To sum up, varieties of Cantonese are spoken in different overseas communities in the world. In the current study, the baseline language for comparing homeland and heritage speakers is Hong Kong Cantonese. Both populations were exposed to this variety in their family. In this language there are six phonemic lexical tones and three allophonic checked tones. The Chinese writing system is complex and literacy usually requires formal education, which means heritage speakers are less likely to be fully literate in Chinese.

1.5 The structure of this dissertation

This introductory chapter is followed by two literature review chapters. Chapter 2 discusses the definition of heritage speakers and their role in the bilingualism literature. Chapter 3 is a review of previous studies on

Cantonese tonal perception, providing the research context of the current study. Chapter 4 describes the experimental paradigm, materials used, and participants of the study. Chapter 5 explains the statistical tests being used for data analysis and presents results of the experiment. Lastly, Chapter 6 discusses implications of major findings and concludes the dissertation.

Chapter 2

Who Are Heritage Speakers?

This chapter is a literature review focusing on heritage speakers (vis-à-vis homeland speakers). Previous studies have approached this topic from different perspectives, such as language policy, applied linguistics, and theoretical linguistics. Consequently, the definition of key terms varies among authors. Section 2.1 compares their motivations and defines these key terms in the context of this dissertation. Section 2.2 outlines what is known on the linguistic behaviour of heritage speakers, and provides the research background that leads to the current study on the perception of lexical tones. Tone-related studies will be reviewed in Chapter 3.

2.1 Defining key terms

The heritage speaker population is known for its heterogeneity and variance, which makes defining it a highly challenging and difficult task (Benmamoun et al., 2013a; Montrul, 2013; Polinsky & Kagan, 2007; Valdés, 2001; Van Deusen-Scholl, 2003; Wiley, 2001; Zyzik, 2016). For the purpose of jumpstarting the discussion, I propose the following working definition:

(7) Definition of a heritage speaker (Version 1)
A heritage speaker is a bilingual who was exposed to a heritage language early in life, but later became dominant in the majority language of the society.

The working definition in (7) contains keywords that can be interpreted in various ways, such as “heritage language”, “bilingual”, and “dominant”. To ensure that readers of different backgrounds can come to the same understanding of (7), each of the subsections that follow will focus on one keyword, discuss possible ways to interpret it, and explain why a particular interpretation is adopted in the context of this dissertation.

2.1.1 Heritage languages

The term *heritage language* is conventionally used in the literature of language policy and education, often in countries with a history of colonization and immigration, such as Canada (Cummins, 1992, 2005; Cummins & Danesi, 1990; Duff, 2008; Duff & Li, 2009), the United States (Fishman, 2014; García, 2005; Hornberger & Wang, 2008; Peyton et al., 2001), and Australia (Brinton, Kagan & Bauckus, 2008; Elder, 2005, 2009; Hornberger, 2005). Definitions given by governmental institutions are usually specific to the sociopolitical context of the relevant country. This subsection introduces the definition of heritage languages according to the federal and provincial governments of Canada—the country of residence of heritage Cantonese speakers under investigation in the present study. The discussion is then followed by a comparison with the sociolinguistic definition.

When defined sociopolitically, heritage languages in Canada refer to languages other than the two official languages, namely English and French. According to Cummins (2005), the term *heritage languages* emerged in 1977, when the Ontario Heritage Language Program was introduced in schools to offer two and a half hours of heritage language instruction per week if 25 students expressed interest in their family language, such as Italian, Portuguese, and Ukrainian (Ontario Ministry of Education, 1991). In 1991, the Senate and House of Commons of Canada enacted the Canadian Heritage Languages Institute Act, which defined a heritage language as “a language, other than one of the official languages of Canada, that contributes to the linguistic heritage of Canada” (Government of Canada, 1991). A more recent definition by the Ministry of Education in Manitoba states that heritage languages are “all languages other than English, French, or Aboriginal, taught in the public school system during the regular school day, either as a regular subject, or as a language of instruction, or as a language of instruction in an enhanced heritage language program” (Government of Manitoba, 2018).

The sociolinguistic definition of heritage languages differs from those provided by governmental institutions, in that it is not country-specific and can be applied to any given social context to focus on the de facto (non-)dominance of languages in a community. In the literature of applied linguistics, a heritage language can refer to any language that has a cultural connection to an individual’s family or community, but is not the primary language used in government, education, and public communication (Fishman, 2001; Kelleher, 2010). As such, this definition is useful for identifying heritage languages in countries without an official language, such as the United States (Peyton et al., 2001). Since English is the de facto dominant language in the United States, heritage languages are languages other than English (Fishman, 2001). Besides, the sociolinguistic definition is useful for describing the language ecology of countries where the official language is not the de facto primary language of the society. For example, Wales has two official languages, namely Welsh and English. According to the 2011 census, only 11% (310,600) of respondents above the age of three reported that they could speak Welsh fluently (Welsh Government and Welsh Language Commissioner, 2015). Although Welsh has official status in Wales, its de facto status fits the sociolinguistic definition of a heritage language, as English is the de facto majority language of Wales. This special case shows that the sociopolitical and sociolinguistic ways of classification do not always yield the same result.

In the context of this dissertation, both the sociopolitical and sociolinguistic definitions can accurately describe the status of Cantonese in Canada. Since the general goal of the current study is to contribute to the field of heritage linguistics, the sociolinguistic definition is more useful, as it is meaningful in non-Canadian contexts as well. It allows comparison between heritage Cantonese speakers in Canada and other heritage speaker communities in any part of the world with a similar language ecology. Adopting the sociolinguistic definition, (8) is a revised version of (7), where new content is underlined:

(8) Definition of a heritage speaker (Version 2)
A heritage speaker is a bilingual; early in life s/he was exposed to a language that has a cultural connection to his/her family or community, but later became dominant in a different language that is of primary use in government, education, and public communication.

2.1.2 Bilingualism

In its most general sense, a *bilingual* is a person who uses two languages in everyday life (Grosjean, 1982). In the bilingualism literature, there is diverse opinion on the specific meaning of "use". The first kind of debate pertains to the level of language proficiency. If a person is very fluent in one language but only semi-fluent in the other, is s/he bilingual? The second type of debate has to do with the degree of importance of four language skills, namely speaking, listening, reading, and writing. If a person has all four language skills for one language but is illiterate in the other, is s/he bilingual? The rest of this subsection is going to elaborate on the narrow and broad definitions of bilingualism, each of which would lead to a different response to these debates. The subsection will be concluded with an explanation of why the broad definition is adopted for the current study.

When defined narrowly, bilingualism is an equally excellent mastery of two languages. This view is commonly found in discussions of second language acquisition. Bloomfield (1933) states that bilingualism is an outcome of "perfect foreign-language learning" (pp.55–56), when a foreign-language learner has reached a level of proficiency that is indistinguishable from native speakers, and at the same time has maintained his/her native language. A bilingual is therefore an individual who has "native-like control of two languages" (pp.55–56). In language testing, the term *bilingual* is sometimes used as a descriptor of the top level of proficiency. For example, in the scale of the United States Foreign Service Institute, Level Five (the highest level) is called "native or bilingual proficiency", which means the individual has "complete fluency in the language such that his speech on all levels is fully accepted by educated native speakers in all its features" (Fulcher, 2014, p.227). This interpretation of bilingual is close to laypeople's understanding of the term. LinkedIn, a business networking website with 500 million users from 200 countries (Darrow, 2017), adopts a similar scale with "native or bilingual proficiency" as the highest level (Ali, 2015). To sum up, the narrow or lay definition is largely motivated by the need to describe an ideal outcome of language learning.

For researchers who are interested in describing the linguistic competence of actual individuals, the idea of a perfectly balanced bilingual who has mastered all language skills for both languages may be unrealistic or even "mythical" (Valdés, 2001, p.40). While perfectly balanced bilinguals do exist, they are rare because it is unlikely for an individual to have the exact same amount of exposure to two languages, and it is also unlikely that the two languages are spoken equally frequently in the same domains of language use (Dornic, 1978; Grosjean, 1998; Myers-Scotton, 2005). Although children of parents who each speak a different native language may be more likely to receive relatively balanced linguistic input, the longitudinal study of Yip & Matthews (2007) shows that the ratio of linguistic input is often affected by factors other than the parents. For example, the child may spend more time with either paternal or maternal grandparents and relatives, because often one side does not live in the same country as the child. Among the six children being studied, only one was truly a balanced bilingual. It shows that if the narrow definition is adopted, a vast majority of actual individuals would fall outside the definition.

By contrast, the broader definition views bilingualism as a continuum: a bilingual is maximally fluent in two languages, or minimally competent in at least one of the four language skills (listening, speaking, reading, and writing) for one of the two languages (Gertken, Amengual & Birdsong, 2014; Macnamara, 1967; Valdés, 2001). Between these two ends there is a wide range of bilinguals with different abilities. In previous studies of heritage speakers (Amengual, 2017; Casillas, 2015, among others), the majority of the subjects fall between the two ends of the continuum. Since the current study is about heritage speakers as well, it was expected that participants would also fall between the two ends of the continuum. Therefore, in the context of this dissertation, the broad definition of a bilingual can better describe the actual language abilities of the subjects. Applying the broad definition, (9) is a revised definition of a heritage speaker, where new content is underlined:

(9) Definition of a heritage speaker (Version 3)
A heritage speaker is <u>a person who uses two languages in everyday life, but does not necessarily have an equal mastery of listening, speaking, reading, and writing skills for both languages</u>; early in life s/he was exposed to a language that has a cultural connection to his/her family or community, but later became dominant in a different language that is of primary use in government, education, and public communication.

2.1.3 Language dominance

Derived from the state of having two or more languages in the mind, *dominance* is a psychological construct with a relativistic nature (Gertken et al., 2014; Grosjean, 1998). A dominant language is the "default language for speaking and thinking" (Harris et al., 2006, p. 264), and is sometimes called the preferred language (Dodson, 1981). According to Gertken et al. (2014), dominance is a function of four components: age of acquisition, frequency of use, language proficiency, and language attitudes. Young age of acquisition, high frequency of use, high language proficiency, and a strong cultural identification to the language are positively correlated with dominance. For a detailed explanation of how these components can be assessed, see Section 4.4.3.2 on the discussion of the Bilingual Language Profile (Birdsong, Gertken & Amengual, 2012).

Proficiency and dominance are distinct concepts even though they are correlated (Schmeißer, Hager, Gil, Jansen, Geveler, Eichler, Patuto & Müller, 2016). First, dominance requires at least a bilingual context but proficiency does not. A monolingual person's language proficiency can be assessed and discussed, but the relativistic concept of dominance would be irrelevant and inapplicable if there is only one language in question (Gertken et al., 2014). Second, dominance is affected by the psychosocial factor of language attitudes. A bilingual person who is equally proficient in two languages can have a stronger cultural identification to one language than the other (Marian & Kaushanskaya, 2004), which affects which language s/he wants to use in a given situation. Third, it is possible for a bilingual's dominant language to be the less proficient one. For example, immigrants who have been immersed in an L2 environment for many years may become L2-dominant, even if this L2 remains the less proficient language compared with L1 (Harris et al., 2006). Taken together, although proficiency is a component of dominance, proficiency alone is not sufficient to predict dominance.

Applying the aforementioned definition of dominance, (10) is a revised version of (9), where new content is underlined:

(10) Definition of a heritage speaker (Version 4)

A heritage speaker is a person who uses two languages in everyday life, but does not necessarily have an equal mastery of listening, speaking, reading, and writing skills for both languages; early in life s/he was exposed to a language that has a cultural connection to his/her family or community, but later a different language that is of primary use in government, education, and public communication has become his/her <u>default or preferred language for speaking and thinking</u>.

2.1.4 Heritage and homeland speakers

The previous subsections have discussed key terms that are crucial to defining who a heritage speaker is. The two definitions in (11) and (12) convey the same message, but the latter is explicit about the meaning of "heritage languages", "bilingualism", and "dominance":

(11) Definition of a prototypical heritage speaker (condensed version, same as (7))
A heritage speaker is a bilingual who was exposed to a heritage language early in life, but later became dominant in the majority language of the society.

(12) Definition of a prototypical heritage speaker (elaborated version, same as (10))
A heritage speaker is a person who uses two languages in everyday life, but does not necessarily have an equal mastery of listening, speaking, reading, and writing skills for both languages; early in life s/he was exposed to a language that has a cultural connection to his/her family or community, but later a different language that is of primary use in government, education, and public communication has become his/her default or preferred language for speaking and thinking.

It should be noted that (11) and (12) are definitions of a *prototypical* heritage speaker. Since the heritage speaker population is highly heterogeneous, there will always be heritage speakers who fall outside a given definition. Zyzik (2016) points out that although no definition is perfect, it is useful to characterize a prototype in heritage language research, so that a population of similar attributes can be identified and studied. In the prototype model of categorization, the boundary between categories could be fuzzy (Rosch, 1973). The *prototype* is central in the category, and group members away from the centre fall on a gradient scale of typicality—the farther away it is from the centre, the less typical it is. Therefore, (11) and (12) by no means suggest that perfectly balanced Cantonese-English bilinguals born and raised in Canada are not heritage speakers. Instead, they suggest that such individuals are less typical of a heritage speaker compared with English-dominant ones.

Now that a prototypical heritage speaker is defined, it is necessary to define a prototypical homeland speaker, whose linguistic behaviour will be compared with that of heritage speakers in the current study:

(13) Definition of a prototypical homeland speaker
A homeland speaker is a person whose default or preferred language for speaking and thinking is the language that s/he was exposed to early in life; during the period of exposure, this language not only has a cultural connection to his/her family or community, but is also the language of primary use in government, education, and public communication.

There are several reasons to use the term *homeland speakers* as opposed to other choices. First, the term *native speaker* is not used even though native speakers typically have the attributes in (13). If *heritage speakers* and *native speakers* are viewed as two distinct populations, it would suggest that heritage speakers are not native speakers. Such a view requires a clear definition of a native speaker, which is out of the scope of this dissertation. Due to reasons stated in Rothman & Treffers-Daller (2014), I assume that it is possible for heritage speakers to be native speakers of their heritage language. The use of the term *native speaker* vis-à-vis *heritage speaker* is therefore avoided. Second, the term *non-heritage speaker* is not used due to its ambiguity. L2 learners who have no familial or cultural connection to the target language can also be called non-heritage speakers. The term *homeland speakers* can exclude L2 learners and avoid ambiguity. Lastly, language-specific terms like *Canadian Cantonese speakers* and *Hong Kong Cantonese speakers* are not used, because they are parallel to names of dialectal varieties, such as *Guangzhou Cantonese speakers* or *Macau Cantonese speakers*. While geographical difference does play a role in shaping language ecology, bilingual configuration is the crucial difference between the two populations in the current study. To sum up, the terms *heritage* and *homeland* are relatively unambiguous and able to highlight the crux of the phenomenon in question without problematic assumptions.

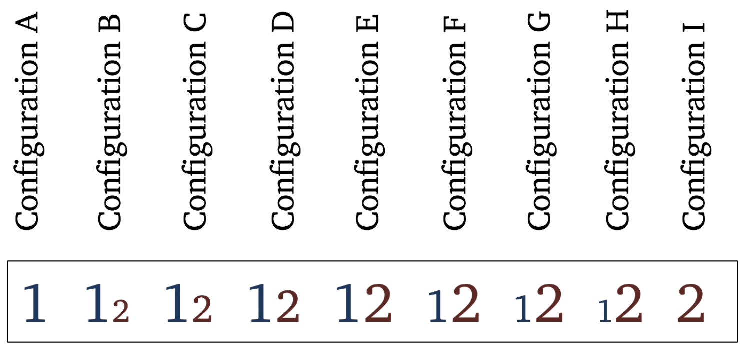
The next section is going to locate homeland and heritage speakers on a bigger picture of the bilingual continuum, and review previous studies on the linguistic behaviour of individuals with different bilingual configurations.

2.2 Configurations on the bilingual continuum

Now that keywords have been defined, this section summarizes what is known about different types of speakers on the bilingual continuum, and sets the stage for the research questions of the current study.

Although most bilinguals can be discretely classified as dominant in one language, Grosjean (2001) points out that dominance can be gradient and not necessarily dichotomous. If dominance is only dichotomous, it would not be able to describe differences in terms of the *degree* of dominance. A dichotomous view of dominance would also obscure longitudinal changes of the two languages' relative strengths over a bilingual individual's lifetime. Therefore, a gradient view of dominance can better capture intra-group and intra-speaker variation.

The bilingual continuum in Figure 2.1 is a visualization of different degrees of dominance. Nine configurations are labelled with English letters for easy identification in subsequent discussion. For each configuration, the numbers "1" and "2" indicate the first language (L1) and second language (L2) respectively, based on the order of acquisition. Configurations A and I represent monolinguals, and so they only have "1" or "2" respectively. These two configurations mark the two ends of the continuum. For each bilingual configuration between A and I, the relative strength of the two languages are indicated by their font size. In Configuration E, which is the middle of the continuum, "1" and "2" have the same font size, which represents balanced bilingualism. L1-dominant configurations are between (but do not include) A and E, with the strength of L2 increasing from left to right. In the present



1 = First language based on the order of acquisition
2 = Second language based on the order of acquisition
Font size indicates language dominance

Figure 2.1: The bilingual continuum (adapted from Valdés, 2001, p.41)

study, homeland speakers of Cantonese were expected to fall within this range of L1-dominant configurations. Lastly, L2-dominant configurations are between (but do not include) E and I, with the strength of L1 decreasing from left to right. Heritage speakers of this study were anticipated to fall within this range of L2-dominant configurations.

Using the bilingual continuum as an anchor, the following subsections will discuss a range of speakers documented in the existing literature, each of which is representative of a certain configuration on the continuum.

2.2.1 Configuration A: Monolinguals

Configuration A in Figure 2.1 represents a monolingual who has not acquired or learned a second language. In linguistic inquiry, an adult with this configuration (and with no hearing impairment or speech disorder) often represents an ideal native speaker who has achieved complete acquisition of a target language. Such speakers' speech is a valuable source of data for understanding the grammar of their language. In experimental studies (e.g., Flege, 1987; Flege & Eefting, 1988), monolingual speakers often form the control group, whose linguistic behaviour is compared with that of another population, such as L2 learners of the target language or monolingual speakers of another language.

In an age of globalization, monolinguals are becoming a minority. According to Crystal (2012), two-thirds of children in the world grow up in a bilingual environment (p.17), and non-native speakers of English have outnumbered native English speakers (p.69). In Hong Kong, for example, the government adopts a language education policy of biliteracy (Chinese and English) and trilingualism (Cantonese, English, and Mandarin) (Bolton, 2011). Therefore, in previous linguistic studies involving adult Cantonese speakers recruited from the university community (e.g. Ma, Ciocca & Whitehill, 2011), participants were not monolinguals, as they had learned English as a second language. For the current study of adult speakers, it was expected that no participants would be monolingual Cantonese speakers with Configuration A.

Bilingual configurations are to the right of Configuration A on the continuum. As Grosjean (1989) points out, bilinguals are not simply two monolinguals living in separate compartments within one person. There is no way for a speaker to "switch off" one language completely when using the other. Psycholinguistic studies show that cross-language activation in processing leads to cross-language competition in comprehension (Dijkstra, 2005) and production (La Heij, 2005). Such cross-language effects were found to be bidirectional: L1 may influence L2, and L2 may also influence L1 (Flege, 1987). The subsequent discussion will focus on previous studies on how L2 influences L1, since it is most relevant to hypotheses of the current study.

2.2.2 Configurations B to D: L1-dominant bilinguals

Configurations B, C, and D all represent L1-dominance, but they differ from each other by the strength of L2. In Figure 2.1, L2 in Configuration D is comparatively stronger than in C; similarly, L2 in C is comparatively stronger than in B, albeit being non-dominant. A psycholinguistic study that pertains to this range of configurations is Linck, Kroll & Sunderman (2009), which found that immersion in L2 can attenuate lexical access to a dominant L1. Their participants were native English speakers learning Spanish as L2, and were matched for self-rated Spanish and English proficiency. They were categorized into two different groups based on immersion experience: the *classroom group* was attending an intermediate-level Spanish course in an American university and had no immersion experience, while the *immersion group* had been studying in Spain for three months. Due to immersion the latter group had a higher frequency of using Spanish (at the time the study was being conducted) compared with the classroom group, which correlates to a greater strength of Spanish. Therefore, the classroom group and the immersion group can be mapped onto the bilingual continuum as Configuration B and Configuration C respectively[1]. In a verbal-fluency task, both groups were presented with one category name at a time (such as "animals"), and were asked to produce as many category exemplars (such as "dog" and "cat") as possible within 30 seconds. The task was done in both Spanish and English in two separate blocks. In their results, the immersion group produced more Spanish exemplars than the classroom group, which is not surprising. A perhaps more interesting finding is that the immersion group produced significantly fewer English exemplars than the classroom group, despite the fact that both groups consisted of native English speakers. The interaction between language and group was significant. The authors conclude that L2 immersion has inhibitory effects on L1 access. The significance of this study is that even immersion in L2 for just three months can leave an impact on L1. Heritage speakers, the focus of this dissertation, have had "L2 immersion" most of their lives. This raises the question of whether a long period of L2 immersion will further influence other processes such as phoneme production.

Flege (1987) found that phonetic spaces of adult speakers can be restructured if they are highly experienced in their L2. Such restructuring

[1]The classroom group and the immersion group could be anywhere between Configuration A and E, as long as the classroom group is to the left of the immersion group. For simplicity I map them onto the continuum as Configuration B and Configuration C respectively.

may affect phoneme production of L1, when a phoneme is used in both languages but realized with phonetic differences. For example, /t/ is used in both English and French, but the English /t/ is produced with a longer voice onset time (VOT) at 77 milliseconds (ms) on average, while the French /t/ has a shorter VOT at 33 ms on average. Groups of L1 English speakers who started learning French as L2 in late adolescence or early adulthood were asked to produce English words (e.g. *two*) and French words (e.g. *tous*), and their VOT values were compared with those of monolingual English and monolingual French speakers. Comparable to Configuration C, the first group of subjects consisted of American university students who had studied in France for less than a year but had already returned to the US for at least three months when the study was conducted. Their average VOT values for English /t/ (72 ms) did not differ from that of monolingual English speakers (77 ms). Another group of subjects was comparable to Configuration D, and they were Americans married to French spouses and had lived in France for an average of 11.7 years when the study was conducted. Their average VOT values for English /t/ (49 ms) were considerably shorter than that of English monolinguals (77 ms). In other words, their English /t/ had become more French-like. To match this group of L1 English speakers who were highly experienced in French, the author recruited native speakers of French who had lived in Chicago for an average of 12.2 years at the time of the study to do the same production task. They produced the French /t/ with a significantly longer VOT (50 ms) than monolingual French speakers (33 ms). In other words, their French /t/ had become more English-like. The implication of this study is that a dominant L1 can be vulnerable to influences from an L2 learned after late adolescence. For the case of heritage speakers, will a non-dominant L1 be even more vulnerable to influences from a dominant L2 acquired in early childhood? Can L2 affect not only production, but also perception?

In their perception study, Samuel & Larraza (2015) point out that extensive L2 experience may entail exposure to L2-accented speech of L1, which makes speakers adapt by accepting “wrong” pronunciations of L1 phonemes as allophonic variations. In Basque, the voiceless predorsoalveolar affricate /ts̻/ and the voiceless alveopalatal affricate /tʃ/ are contrastive, but in Spanish only /tʃ/ is used. L1 Spanish speakers may find it challenging to produce Basque words with /ts̻/, and would produce a more /tʃ/-like consonant instead. In their experiment, highly proficient Basque-Spanish early bilinguals (comparable to Configuration D) were trained to match unusual objects with new Basque words that contained either /ts̻/ or /tʃ/. Since they were new words, L1 Basque speakers should have never heard of them being pronounced with a Spanish accent. After the training session, L1 Basque and L1 Spanish bilinguals performed a picture-name matching task. A picture of an object was presented on the screen and at the same time a spoken word was played over the headphone. Participants had to press a button to indicate whether the picture matched with the word. They were told explicitly that there would be minor deviations in the pronunciation of some words, which should be considered a mismatch. Surprisingly, even L1 Basque speakers failed to reject non-words half of the time, although their accuracy was still significantly higher than that of L1 Spanish speakers. Another part of the experiment was an AXB discrimination task, in which the L1 Basque participants did very well. Their high discrimination accuracy eliminated the possibility that they were unable to detect the acoustic differences between the two affricates. The authors concluded that L1 Basque speakers' acceptance of "mispronounced" words was not a sign of poor perception, but it was a dual-mapping process, in which two phonetic variants are mapped onto a single lexical representation. Such process can facilitate efficient lexical access in an environment where L2-accented speech is often heard. The significance of this study is that L2 effects on L1 perception may not always suggest perceptual "impairment"; rather, it could be a sign of perceptual flexibility and adaptation as a strategy for efficient communication with other speakers.

2.2.3 Configuration E: Perfectly balanced bilinguals

Configuration E represents perfectly balanced bilinguals. As mentioned previously, even children of parents who each speak a different native language tend to be dominant in one of the two languages (Yip & Matthews, 2007). Individuals raised in bilingual societies are usually dominant in one of the two languages as well. For example, in a study by Sebastián-Gallés, Echeverría & Bosch (2005) on Catalan-Spanish simultaneous bilinguals in Barcelona, all 40 participants had one Spanish-speaking parent and one Catalan-speaking parent. Although it may seem to be the perfect environment for producing balanced bilinguals, the participants were found to be either Catalan-dominant or Spanish-dominant, as they had different patterns of language use at home and for socializing. Therefore, even though bilinguals can be equally proficient in both languages, the perfectly balanced bilinguals in terms of dominance are extremely rare. Amengual Watson (2013) even comments that "the perfectly balanced bilingual probably does not exist" (p.7).

2.2.4 Configurations F to H: L2-dominant bilinguals

Heritage speakers, the subject of investigation in this dissertation, are typically L2-dominant, which means they fall to the right of Configuration E in Figure 2.1, where L2 is stronger than than L1. They are highly heterogeneous in terms of the strength of their heritage language (L1), ranging from fluent speakers (Configuration F), to semi-fluent speakers (Configuration G), to receptive listeners who can understand but barely speak or cannot speak the language at all (Configuration H).

This range of configurations is particularly useful for describing the change of language use patterns across different generations of immigrants. According to Valdés (2001), every new generation moves closer to the right end of the bilingual continuum compared with their previous generation. Having migrated from their home country to a host country, the first generation is typically either monolingual or L1-dominant, hence comparable to Configurations A to D. Their children, the second generation, were born and raised in the host country. They tend to be dominant in the majority language of the society, but they continue to have at least some proficiency of the heritage language in order to communicate with the first generation. Therefore, their bilingual configuration often falls to the right of Configuration E. The third generation are children of heritage speakers. Since their parents are L2-dominant, at home they may communicate in the majority language of the society more often than in the heritage language. If they spend time with family members of the first generation, they may be exposed to the heritage language more often than those who do not. Therefore, the bilingual configuration of the third generation may drift further to the right on the continuum.

Recognizing these cross-generational differences, the Heritage Language Documentation Corpus constructed by Nagy (2009) contains language data from all three generations mentioned above. Nagy (2015) is a study based on this corpus, which shows that first-generation Ukrainian speakers in the Greater Toronto area have significantly shorter VOTs (an average of 26 ms) than those of the newer generations when producing Ukrainian words with /p t k/. On the other hand, the average VOTs of the second (38 ms) and third generations (43 ms) did not differ from each other significantly. Such linguistic variation across generations has important implications for this dissertation. Although the second and third generations can both be considered heritage speakers, the second generation received language input mostly from homeland speakers, while the third generation received language input mostly from heritage speakers. To control potential effects due to differences in the input, only second-generation speakers were included in the present study (see Section 4.4.3.2). This ensured that both homeland and heritage speakers were exposed to the same baseline variety of Cantonese spoken by homeland speakers in the parental generation. This way any observed difference between the two groups in the experiment, if any, would be a reflection of their different perception of lexical tones, but not a reflection of linguistic variation in the input from parents of different generations.

Apart from cross-generational comparisons, previous studies on heritage speakers have also looked into L2 effects across different aspects of grammar. Benmamoun et al. (2013b), Montrul (2013), and Polinsky & Kagan (2007) argue that morphology, syntax and semantics are more vulnerable domains compared with phonetics and phonology. For morphology, English-dominant receptive listeners of Inuttitut in Canada accepted ungrammatical sentences that omit case morphemes, incurring an error rate of 40% in a grammaticality judgment task (Sherkina-Lieber, Pérez-Leroux & Johns, 2011). Heritage speakers of classifier languages such as Mandarin (Ming & Tao, 2008) and Cantonese (Wei & Lee, 2001) sometimes paired a noun with the wrong classifier or omitted classifiers completely. For syntax, adult heritage speakers of Russian in the United States found it challenging to process object relative clauses in a picture-matching task (Polinsky, 2008). For semantics, heritage Spanish speakers in the United States were found to be insensitive to nuances of the subjunctive mood (Montrul, 2009). Phonetics and phonology, however, are considered the less vulnerable areas of linguistic knowledge. Researchers agree that heritage speakers have more native-like pronunciation than L2 learners (Benmamoun et al., 2013b; Montrul, 2013), even though homeland speakers might sometimes find it "off" or "funny" (Montrul, 2013, p.378).

The phonetic and phonological knowledge of L2-dominant bilinguals raises interesting questions. First, the previous subsection on Configurations B to D has shown that even a dominant L1 can be affected by a non-dominant L2. For L2-dominant bilinguals, to what extent are L1 phonetics and phonology (in)vulnerable to L2 effects? Second, if homeland and heritage speakers share similar phonetic and phonological knowledge, would it suggest that early exposure guarantees maintenance of this knowledge for the rest of one's life? Third, if there are L2 effects on heritage speakers' L1 phonetics or phonology (as in those observed in L1-dominant bilinguals), are they symmetric between production and perception? Last but not least, what would be the mechanism of such effects, if any? The rest of this subsection will review previous studies on speech production and perception that have shed light on these questions, with the goal of motivating the research questions of the current study.

Chang, Yao, Haynes & Rhodes (2011) compare the production of American English /u/ and Mandarin /u/ by three groups of English-Mandarin bilinguals residing in the United States: (1) L1-dominant Mandarin speakers born and educated in mainland China or Taiwan up to at least seventh grade, who learned English as their L2; (2) L1-dominant English speakers born and educated in the United States who started to learn Mandarin as L2 after the age of 18, and (3) heritage speakers of Mandarin comparable to Configurations F or G, who grew up in the United States and spoke Mandarin in a family setting. All participants were instructed to read aloud English and Mandarin words that contained /u/. The second formant (F2) of their /u/'s was measured. An /u/ with a higher F2 is more English-like, and an /u/ with a lower F2 is more Mandarin-like. After obtaining the F2 values, the authors measured the acoustic distance between each subject's English and Mandarin /u/'s. Two possible situations may lead to a small acoustic distance: a speaker produces a Mandarin-like /u/ for both English and Mandarin words, or a speaker produces an English-like /u/ for both English and Mandarin words. In their results, the two L1-dominant groups (those from mainland China or Taiwan, and those from the United States learning Mandarin in school) produced a relatively small acoustic distance between their English and Mandarin /u/'s, as their F2 values were within the proximity of their respective L1 vowel space. In contrast, the heritage group had the biggest acoustic separation for the two languages. Their English /u/ had a higher F2 (hence more English-like) than L1-dominant speakers', and their Mandarin /u/ had a lower F2 (hence more Mandarin-like) than L1-dominant speakers'. This means among the three groups, heritage speakers' vowels showed the closest approximation to the phonetic norms of both languages. The findings echo with those of Saadah (2011) on the production of Arabic and English vowels by heritage speakers of Arabic, whose vowel spaces have shown that they had two separate vowel categories for English and Arabic. A possible interpretation of these results is that early exposure to both languages allowed heritage speakers to be sensitive to the fine-grained phonetic details in the input, dissimilate cross-language vowel categories, and create more polarized phonetic spaces for two languages. On the other hand, L1-dominant speakers who learned the L2 later in life may have assimilated L2 categories into existing L1 categories, producing L2 vowels that were closer to L1 phonetic norms.

In another production study, Antoniou, Best, Tyler & Kroos (2010, 2011) looked into the production of English and Greek consonants by heritage speakers of Greek in Australia, and found that cross-language effects were different between unilingual and code-switching contexts. In particular, L1 effects were observed in code-switching contexts, even though the participants were L2-dominant. All subjects were exposed to Greek from birth, and quickly became English-dominant after the onset of schooling between age three and four. According to self-reports, they kept using Greek on a daily basis, and so were comparable to Configuration F or G. Voiceless stops /p t k/ are phonetically realized with a short-lag VOT in Greek but a long-lag VOT in English. When asked to produce English-only or Greek-only sentences (such as "say pa again" or "*λέει πα άλλο*"), heritage speakers produced consonants with VOT values that were indistinguishable from those of monolingual speakers of each of the two languages. However, when asked to produce the target words in carrier sentences of the other language (such as "say *πα* again" or "*λέει pa άλλο*"), cross-language effects were unidirectional, in that L1 affected the production of L2 targets, but L2 did not affect L1 targets. When switching from Greek to English ("*λέει pa άλλο*"), heritage speakers produced English targets with Greek-like VOT values. However, when switching from English to Greek ("say *πα* again"), their Greek targets had Greek-like VOT values without any sign of effects from English. The significance of this study is that cross-language interference is not "across the board" but can vary in more complex linguistic processes such as code-switching, in which L1 effects can be observed despite the subjects' dominance in L2. Certain methodological decisions in this dissertation were made in the light of these findings. Since code-switching is not relevant to the current study, measures were taken to make sure that participants would operate in a unilingual mode. As Chapter 4 will explain, spoken Cantonese was used in all task instructions, and a Cantonese story-listening task was inserted before the actual experiment.

These ensured that all participants, who were residing in Canada when this study was conducted, were attuned to a unilingual Cantonese listening environment.

Configuration H has the weakest L1 among the L2-dominant configurations in Figure 2.1, which is comparable to receptive bilinguals who have at least some degree of listening competence but no production ability of their L1. For this reason receptive bilinguals are usually recruited for perception instead of production studies (e.g., Celata & Cancila, 2010; Tees & Werker, 1984). Since their L1 is so much weaker compared with Configurations F and G, will their ability to discriminate L1 phonemes differ from homeland speakers'? Or could it be that early exposure to two languages would enable them to separate two phonological systems successfully in perception, just like what happened to heritage speakers' production in Chang et al. (2011)? Two perception studies offer different answers to this question.

Tees & Werker (1984) point out that linguistic perceptual abilities can be maintained even after a long period of disuse if an individual had early experience in hearing the relevant contrasts. In their study the critical contrast was the dental /t̪/ versus the retroflex /ʈ/ in Hindi. Two groups of students from a Hindi language course in a Canadian university performed a category-change discrimination task, in which they were asked to press a button when hearing a change in a stream of sounds such as [ʈa ʈa ʈa t̪a t̪a...]. The first group had no experience of Hindi prior to the course and were typical L2 learners. The second group had early experience in Hindi but had almost none or very limited ability to speak or understand it when the course started (hence comparable to Configuration H). They either had lived in India in the first year or two of their lives and stopped using any Indian language with this place contrast in the family after moving to North America, or had a Hindi-speaking relative living with their family in North America in the first year or two of their lives. Both groups were tested twice: the first test took place in the first or second week of the Hindi language course, and the second test was held one year after the course started. Results of the first test showed significant difference between the two groups. The "early experience group" had an accuracy rate of 90%, comparable to seven-month-old infants in Werker, Gilbert, Humphrey & Tees (1981), while the "no early experience group" had an accuracy rate of lower than 10%. As for the second test, the early experience group made an improvement and went over 90%, close to the native adult speakers of Hindi. The "no early experience group" also showed an increase, but their accuracy was still less than 20% a year after the course started. It is noteworthy that the two groups' average grades for the course did not differ significantly in the end, which suggests that the "early experience group" did not have developmental privileges for other aspects of grammar such as vocabulary or syntax. These findings were similar to those of Oh, Jun, Knightly & Au (2003) on the perception of Korean consonants by individuals who heard Korean regularly during childhood but had minimal exposure to the language after childhood. The authors of both studies posit that early linguistic experience, however limited it might be, is beneficial to the long-term maintenance of perceptual ability.

Contrary to the findings of Tees & Werker (1984), Celata & Cancila (2010) reported phonological-perceptual attrition in heritage speakers of Lucchese (spoken in Lucchesia, northern Tuscany, Italy). The phonemic contrasts under investigation were singleton versus geminate consonants in Lucchese, such as *casa* /ˈkasa/ "house" versus *cassa* /ˈkasːa/ "box". Carrier sentences with target words containing singleton or geminate consonants were read aloud, and participants were asked to indicate what they believed was uttered on a piece of paper printed with choices written in Italian orthography. Participants in the heritage group (or what the authors called "second-generation immigrants") were born and raised in the United States who reported that they could not speak the language although they could understand their parents at least to some extent when they spoke Lucchese to each other. In their results, heritage speakers (or rather, *listeners*) had an error rate of 46.19%, which is significantly higher than the error rate of 8.75% found for Lucchese speakers residing in Italy. In the same perception study, non-words with singleton or geminate consonants (such as /asa/ and /asːa/) were manipulated to create stimuli with varying consonant lengths on a phonetic continuum with six steps (Step 1 singleton, Step 6 geminate). Lucchese speakers residing in Italy showed categorical perception, showing a steep increase of "geminate" responses at Step 3 or Step 4. However, for the heritage group, the increase of "geminate" responses was gradual along the continuum, showing little sign of categorical perception. The authors concluded that the heritage group showed strong impairment in the perception of the consonant length feature in both experiments. According to the authors, heritage listeners' insensitivity to the singleton and geminate contrasts was owing to their reliance on the phonological system of American English. Since consonant length is not a useful cue for making lexical contrasts in American English, heritage listeners of Lucchese may have adjusted their processing strategies and only attended to contrasts that are relevant to American English, the dominant language being used extensively in daily life. In sum, early exposure to L1 does not necessarily guarantee native-like perception for the rest of one's life. Continuous exposure and use of L2 may result in a change of listening strategies (see also Bruggeman, 2016; Rafat, Mohaghegh & Stevenson, 2017).

2.2.5 Configuration I: Replacive bilinguals

Marking the end of the bilingual continuum, Configuration I represents individuals who have no conscious recollection of the language that they were first exposed to in life, which is common for international adoptees. Since these individuals have neither receptive nor productive competence in their pre-adoption L1, they are functionally monolingual. For this reason in Figure 2.1, the number "1" is absent from Configuration I and only "2" is shown. De Geer (1992) and Gauthier & Genesee (2011) call the language acquired after adoption a "second first language"—"second" in terms of chronological order but it is the "first" of which they have conscious memory of exposure. Although individuals with Configuration I are functionally monolingual, they are not as monolingual as those with Configuration A, in the sense that they were exposed to two languages in total during their life. Therefore, Configuration I is still of interest to researchers of bilingualism.

For example, Yip (2013) uses the term “replacive bilingualism” (p.120) to refer to cases where an adopted child’s native language was replaced entirely. These cases raise an important question of whether early childhood language memory can remain accessible in adulthood after a long period of zero exposure.

Oh, Au & Jun (2010) show that international adoptees have an advantage over novice L2 learners when (re)learning phonemes of the pre-adoption language. The critical contrast in the study was the three-way distinction of stops in Korean: lenis /t/, tense /t*/, and aspirated /t^h/. Their subjects were two groups of young adults attending a first-semester Korean language course in an American university. The first group was English-speaking L2 learners who had no Korean exposure in childhood. The second group was monolingual English-speaking individuals adopted from Korea to the United States as young children between three months to one year old. They differed from the Hindi listeners in Tees & Werker (1984) in that they had no conscious recollection of the pre-adoption language. Therefore, these functionally monolingual adoptees were comparable to Configuration I. All participants were asked to do an ABX discrimination task, where the first two words (A and B) were produced by the same talker, and the last word (X) was produced by a different talker. Participants responded by pressing a button to indicate whether X was the same word as A or B. According to their results, Korean adoptees were significantly better than novice L2 learners at distinguishing lenis and aspirated stops but not the tense stop. The authors see the adoptees’ phoneme distinction ability as a sign of retention and re-activation of long-ago childhood language memory, which was not lost completely but was only inactive. Their findings provide converging evidence to the claim in other studies that early linguistic exposure benefits phoneme perception despite the lack of post-childhood exposure.

2.2.6 Summary

To sum up, previous research suggests that early exposure *allows* but does not *guarantee* maintenance of productive or perceptual abilities for L1 across all contexts of language use. On the one hand, it is possible for heritage speakers to have identical performance to L1-dominant native speakers in production (Antoniou et al., 2010; Saadah, 2011) and perception (Tees & Werker, 1984), even for individuals who had stopped receiving L1 input for an extended period of time (Oh et al., 2010, 2003). On the other hand, heritage speakers may have different linguistic behaviour from that of homeland speakers due to language attrition (Celata & Cancila, 2010) or having more polarized phonetic spaces for two phonologies (Chang et al., 2011). Lastly, cross-language effects can come from a dominant language (Celata & Cancila, 2010), a non-dominant language (Antoniou et al., 2011; Flege, 1987; Linck et al., 2009; Samuel & Larraza, 2015), or the state of having two languages in the mind (Chang et al., 2011). There is not a single generalization that can account for all aforementioned cases.

How can *tone*, the contrastive dimension under investigation in this dissertation, contribute to the literature of bilingualism, and in particular, the growing field of heritage linguistics? The reviewed studies so far have been dealing with segmental phonemes, namely consonants and vowels. Even though the L1 and L2 in each study have different consonant and vowel inventories, on a general level both language systems share the similarity of using segmental contrasts to encode lexical contrasts. All observed cross-language influences on vowels and consonants pertain to the same contrastive dimension. For a language pair like Cantonese and English, however, the two systems do not share the same dimensions for lexical contrasts. Cantonese has both segmental and tonal dimensions, while English has the segmental but not the tonal dimension. On the one hand, it can be argued that a non-existent lexical tone system of English cannot possibly contain anything to affect the tonal dimension of Cantonese in a bilingual's mind. On the other hand, it can be argued that both languages have suprasegmental phonologies, as English does make use of pitch variations for stress and intonation, which carry meaning on the phrase or sentence level. From this perspective it may be possible for cross-language effects to happen between the suprasegmental phonologies of two languages. Before exploring which of the two claims is empirically supported, the next chapter will explain the tone system of Cantonese and what is known about homeland speakers' tonal perception, the basis for comparison with heritage speakers' perceptual abilities for lexical tones.

Chapter 3

What Is Tonal Perception?

The focus of this literature review chapter is tone, or more precisely, the *perception* of tone. Section 3.1 clarifies the relationship among three concepts, namely *fundamental frequency*, *pitch*, and *tone*. It is then followed by a summary of previous studies on Cantonese tonal perception by three groups of individuals[1]: children acquiring Cantonese (Section 3.2), adult homeland speakers (Section 3.3), and non-Cantonese speakers (Section 3.4). Results of these studies constitute the basis for comparison with results of the current study in Chapter 5. Section 3.5 discusses studies related to tone and heritage speakers. Lastly, Section 3.6 summarizes the two literature review chapters, which leads to hypotheses to be tested in this study.

3.1 The acoustic and perceptual aspects of lexical tone

Before the discussion of previous studies, the meanings of three related but not equivalent terms need to be clarified—*fundamental frequency*, *pitch*, and *tone*. The first one, *fundamental frequency* (f0), is an acoustic term referring to the number of cycles per second in a sound wave measured in hertz (Hz) (Titze, 1994). It pertains to the physical property of the acoustic signal

[1]All participants in the studies mentioned in this chapter had no reported hearing impairment, speech disorder, or abnormalities in cognitive development.

itself, and is not dependent on the perspective of the receiver of the signal. For example, an f0 of 22,000 Hz is measurable, regardless of whether it is audible to the human ear.

The second term, *pitch*, is perceptual in nature. It is dependent on the perspective of a hearer, whose auditory system processes the acoustic signal and determines what is heard. Due to the special structure of the human's basilar membrane, f0 values do not translate linearly into pitch (von Békésy, 1960). For example, to a human hearer, the pitch difference between 200 Hz and 300 Hz is much bigger than that of 12,000 Hz and 12,100 Hz, even though the f0 difference in both cases is the same at 100 Hz. Therefore, pitch is only relevant when a signal is processed by a hearer.

While pitch can be an attribute of speech or non-speech signals (such as music or a fire alarm), *tone* is a different term that is linguistic in nature. It refers to the use of pitch variations (among other things, such as duration and phonation type) to mark lexical contrasts, morphological or syntactic categories (Yip, 2002). This makes tone a phonological category like consonants and vowels. For tone languages with contrasts involving multiple dimensions (e.g., pitch, duration, phonation type), f0 is often the most important acoustic correlate of tone, while relative pitch height (and the change thereof) is often the most important perceptual correlate of tone (Gandour, 1978; Yip, 2002).

Adding everything up, tonal perception is the process of extracting relevant auditory cues from a continuous speech signal and mapping pitch attributes to discrete phonological categories. A speaker of a tone language has the ability to locate boundaries between separate tonal categories along a continuous dimension (Gandour, 1978; Gandour & Krishnan, 2015). The basic question asked in this dissertation is whether homeland and heritage speakers of Cantonese share the same phonological knowledge of tone.

There is a huge body of literature on Cantonese tonal perception, which will be summarized in the upcoming sections. Readers who have not read the introduction to Cantonese tones in Section 1.4.5 are recommended to do so before diving into the rest of this chapter. The phonemic tone inventory of Cantonese is repeated below as Table 3.1 for readers' easy reference.

Table 3.1: The phonemic tone inventory of Cantonese

Tone	Description	Tone numerals	Example	Gloss
1	high level	55	*se1*	“some”
2	high rising	25	*se2*	“write”
3	mid level	33	*se3*	“diarrhea”
4	low falling	21	*se4*	“snake”
5	low rising	23	*se5*	“society”
6	low level	22	*se6*	“shoot”

3.2 Perception of Cantonese tones by Cantonese-learning infants and children

Although subjects of the current study were adults, previous works on Cantonese-learning infants and children serve as important references as to the kind of phonological knowledge shared by homeland and heritage speakers before their language development went on diverging paths towards different bilingual configurations, as in Figure 1.1 from Chapter 1. These references can be pointers[2] to whether an observed difference between homeland and heritage speakers is due to language attrition (i.e. it was acquired before but eroded or lost later) or incomplete acquisition (i.e. it has not been acquired in the first place). If heritage speakers lack a certain type of phonological knowledge that Cantonese-learning infants or children possess, it may be a sign of language attrition in heritage speakers. However, if heritage speakers, infants and children all lack a certain kind of phonological knowledge that adult homeland speakers possess, it may be a sign of incomplete acquisition by heritage speakers.

Converging evidence from previous research suggests that Cantonese-learning infants are capable of detecting tonal contrasts as early as four months of age. Yeung, Chen & Werker (2013) is a Vancouver-based study

[2]I call them “pointers” only as I recognize that different methodological paradigms were adopted in these studies, and adult heritage speakers were infants at a different point in time from the infants being tested.

on the tone discrimination abilities of Cantonese- and English-learning infants at four months and nine months of age. In particular they focused on the Cantonese tonal contrast of T2 [25] and T3 [33]. According to the parents' reports, all Cantonese-learning infants were exposed to Cantonese exclusively at least 90% of the time, and did not spend a lot of time with English speakers. In other words, these infants had the potential to become heritage speakers of Cantonese after the onset of schooling, since English is the majority language for education and public communication in Vancouver. Results of the study show that all four-month-old infants (both Cantonese- and English-learning) were sensitive to the T2-T3 contrast. However, the performance of nine-month-old infants differed: nine-month-old Cantonese-learning infants were able to maintain their tone discrimination ability for T2 and T3, but English-learning infants failed to do so. The implication for adult heritage speakers is that if they fail to discriminate T2 [25] and T3 [33], it would more likely be a sign of attrition than incomplete acquisition.

Several Hong Kong-based studies confirm the general tone discrimination ability of Cantonese-learning infants and children, though their performance for different tone pairs varied. Lei (2007) adopted the Conditioned Head Turn procedure (Eilers, Wilson & Moore, 1977) to investigate the perception of three level tones (T1 [55], T3 [33], and T6 [22]) by Cantonese-learning infants at six and eight months of age. Results suggest a possible relationship between acoustic distance and ease of discrimination. T1 [55] and T6 [22] are most different in terms of pitch height, and were best discriminated by the subjects. On the other hand, T3 [33] and T6 [22] have more similar pitch heights, and were comparatively more difficult for the infants.

In another Hong Kong-based study on children of two to three years of age, Lee, Chiu & van Hasselt (2002) found that f0 onsets play a more important role in children's tonal perception than f0 offsets do. Their stimuli only included three tones with the most salient perceptual cues: T1 [55] has the highest pitch height among the six tones and marks the upper boundary of the Cantonese tonal space; T2 [25] has the biggest magnitude of pitch change; lastly, T4 [21] marks the lower boundary of the tonal space. The experimenter read aloud words and non-words (presented as names of dolls with different facial expressions and costumes) live, and subjects were asked to point at a picture (for words) or a doll (for non-words) between two choices that represented a tonally contrastive minimal pair. Their overall accuracy was at 90.6% for words and 72.7% for non-words, but their performance for T2-T4 was significantly worse (87% for words and 66% for non-words) than that of T1-T2 (93% for words and 77% for non-words) and T1-T4 (92% for words and 75% for non-words). The authors point out that T2 [25] and T4 [21] have similar f0 onsets, which makes them more difficult to tell apart than the other two pairs with more distinct f0 onsets. This study concludes that three-year-old children's tonal perception abilities were only partial.

As for Cantonese-learning children of four to six years of age, two Hong Kong-based studies confirmed improvement of tonal perception abilities during this period, but their accuracy is still significantly lower than adults'. In Ciocca & Lui (2003), all contrastive pairs in the Cantonese tone inventory were tested with children of four, six, and ten years old as well as adults. They listened to recorded stimuli and were instructed to choose one of the two pictures for a minimal pair on a screen. Significant improvement of accuracy was observed between four and six years of age, as well as between six and ten years of age; after ten years of age no significant improvement was observed. Among all tone pairs, T3-T6 and T2-T5 were the last to be acquired in perception (see also Ciocca & Ip, 2008). For the level tones T3 [33] and T6 [22], four-year-olds' accuracy was at chance (50%), six-year-olds' was close to 80%, showing significant improvement, and ten-year-olds' was near 90%, close to adults' 95%. For the rising tones T2 [25] and T5 [23], no significant improvement was observed between age four and age six—both groups achieved an accuracy rate close to 65%. Children at age ten were able to reach 70%, the lowest within this age group. Adults' accuracy did not reach ceiling and was only at 80%. In a more recent study by Wong & Leung (2018) who adopted a slightly different picture identification task (with four pictures instead of two: one target word, one tonally contrastive competitor, two segmentally constrastive competitors), six-year-olds had fairly high accuracy (over 80% for all tones), but in general it was still significantly lower than adults' (99–100%). In sum, in both Ciocca & Lui (2003) and Wong & Leung (2018), six-year-olds' performance had not reached adult accuracy.

These findings raise interesting questions for the case of heritage speakers: if tonal perception is still not fully acquired at six years of age, what will happen if a child starts to receive significantly more exposure to English around this time? Will s/he continue to acquire Cantonese tonal contrasts, or will the acquisition be interrupted? If s/he continues to acquire Cantonese tonal contrasts, will it be done through the lens of English phonology?

3.3 Perception of Cantonese tones by adult homeland speakers

Cantonese tonal perception by adult homeland speakers has been studied extensively from a variety of perspectives, including contextual effects (Fox & Qi, 1990; Francis, Ciocca, Wong, Leung & Chu, 2006; Gu & Lee, 2007; Li, Lee & Qian, 2002; Wong, 2007; Zhang, Peng, Wang & Wang, 2015; Zheng, Peng, Tsang & Wang, 2006), speech rate effects (Wong, 2011), auditory attention and memory (Law et al., 2013), normalization for intra- and inter-talker variation (Chang, Yao & Huang, 2017; Wong & Diehl, 2003; Zhang, Peng & Wang, 2011), temporal resolution (Yu, 2017), tone-intonation interaction (Ma, Ciocca & Whitehill, 2006, 2011; Vance, 1976), cross-modal perception (Burnham, Ciocca, Lauw, Lau & Stokes, 2000; Burnham, Lau, Tam & Schoknecht, 2001), and comparison with machine recognition (Lee, Lau, Wong & Ching, 2002; Peng & Wang, 2005; Yu, 2017). This section will only review two areas that are most relevant to the present study, namely the acoustic and perceptual correlates of tone, and sound change in progress. Results for homeland speakers in the present study were expected to be generally similar to those in prior research.

3.3.1 Acoustic and perceptual correlates of tone identity

In the literature on Cantonese tonal perception, it is generally agreed that f0 is the main acoustic correlate of tone identity (Fok-Chan, 1974; Gandour, 1981; Khouw & Ciocca, 2007; Lee et al., 2015; Vance, 1977). In particular, Tong, Lee, Lee & Burnham (2015) found that the three most important acoustic cues for accurate tonal perception among adult speakers are average f0, f0 onset, and f0 major slope. As for non-f0 cues, duration and intensity are not particularly useful for tonal perception (Fok-Chan, 1974; Khouw & Ciocca, 2007; Tong et al., 2015), even though T2 [25] has the longest duration and T1 [55] has the highest intensity among all six tones (Tong et al., 2015). Lastly, creaky phonation is useful but not required for accurate identification of T4 [21] (Yu & Lam, 2014). Each of the exploited perceptual cues is explained as follows.

The six Cantonese tones can be perceptually divided into "level" and "contour" according to whether there is a change of relative pitch height within the syllable (Fok-Chan, 1974). Within the "no change of relative pitch" group, tones can be divided into "high level" T1 [55], "mid level" T3 [33], and "low level" T6 [22] by pitch height (Gandour, 1981). It should be noted that these tones are perceptually level, but acoustically they show some f0 declination towards the end of the syllable (Li et al., 2002; Wong, 2006). Results of the tone identification task in Francis, Ciocca & Ng (2003) provide evidence that the perception of level tones is categorical. Synthesized stimuli were created with a tonal continuum from Step 1 "low" [22] to Step 10 "high" [55]. In their results, the category boundary between low level and mid level was clear around Step 4, while the boundary between mid level and high level was also clear between Step 7 and Step 8. According to a study on the relationship between linguistic tones and musical tones (Yiu, 2013), the perceptual distance between T1 [55] and T3 [33] is two semitones, while the distance between T3 [33] and T6 [22] is only one semitone. T1 [55] is separated from other tones in the perceptual tonal space, and often receives the highest accuracy score in perception studies (e.g. Lee et al., 2015). On the other hand, T3 [33] and

T6 [22] are closer together in the tonal space, and so the confusion rate of T3-T6 is often higher than that of T1-T3 (e.g. Mok & Wong, 2010).

Contour tones fall into the “with change of relative pitch” group, which can be further divided into “rising” (T2 [25], T5 [23]) and “falling” (T4 [21]) by the direction of pitch change (Fok-Chan, 1974; Gandour, 1981; Khouw & Ciocca, 2007). Francis et al. (2003) confirmed that the perception of T2, T4, and T5 is categorical. Synthesized stimuli with a tonal continuum from Step 1 “low falling” [21] to Step 10 “high rising” [25] were created. In their results, the category boundary between low falling and low rising was clear, showing a crossover between Step 3 and 4. The boundary between low rising and high rising was also clear, showing a crossover at Step 7.

Although perceptually T4 [21] and T6 [22] belong to the “contour” and “level” group respectively, acoustically the falling slope of T4 [21] is similar to that of T6 [22] in declination (Wong, 2006). Despite their acoustic similarity, T4 [21] often receives a high score in word-identification tasks (e.g. Khouw & Ciocca, 2007), and the tone pair T4-T6 has low confusion rates in AX discrimination tasks (e.g. Mok & Wong, 2010). A possible explanation is that T4 [21] is sometimes but not consistently realized with creaky voice quality. In the perception study by Yu & Lam (2014), creaky voice was added to some T6 [22] utterances in the stimuli. Participants responded with T4 significantly more often when the stimuli had creaky voice, even though the stimuli’s f0 had T6 [22] properties. The authors concluded that creaky voice facilitates identification of T4 [21] in addition to its low f0.

The last perceptual dimension, namely “magnitude of pitch change”, is primarily relevant for distinguishing between T2 [25] and T5 [23] (Fok-Chan, 1974; Gandour, 1981; Khouw & Ciocca, 2007; Vance, 1977). Both of these tones are “contour” and “rising”, but T2 [25] has a larger magnitude of pitch change due to its higher peak in the offset compared with T5 [23]. Since both tones have similar pitch onsets, the latter half of the syllable contains most of the cues to separate these two tones (Khouw & Ciocca, 2007; Lam, Hall & Pulleyblank, 2016). T2-T5 is often the most confusable pair in perception studies (e.g. the control group in Mok & Wong, 2010), which led researchers to believe that the two tonal categories are in the process of being merged. The next subsection will elaborate on this sound change in progress.

3.3.2 Tone mergers

Cantonese speakers' confusion between T2 [25] and T5 [23] in production and perception has been observed since the 2000s (Kej, Smyth, So, Lau & Capell, 2002). Over the years the phenomenon went from being called "tone production errors" (Kej et al., 2002, p.35), to "some kind of change is going on" (Bauer, Cheung & Cheung, 2003, p.222), to "initial stages of tone merging in progress" (Mok et al., 2013, p.364). Individuals who have merged the tones show a lot of inter- and intra-speaker variations, which are explained as follows.

The first type of such variation is how the tones are merged: T5 merging into T2, T2 merging into T5, or having one general rising tone that is different from T2 and T5. In the production study by Kej et al. (2002), six out of 15 participants were found to have difficulties with T2 and T5, but their tonal production patterns differed. Three of them produced T5 with T2-like offsets, two of them produced T2 with T5-like offsets, and one of them produced only one rising tone midway between T2 and T5. Fung & Wong (2011) looked into the acoustics of this "midway" rising tone produced by six subjects in their early 20s, and found that the onset of this rising tone is higher than T2 and T5, while its offset is similar to T2's[3]. As a result, this rising tone has a slope similar to T5's, even though it has a T2-like offset. Fung, Wong & Law (2011) recorded words produced by two age groups (average 22.3 and 53.17 years), and concluded that older speakers tend to merge T5 into T2, while younger speakers tend to have the "midway" rising tone. The two age groups demonstrated merger by transfer and merger by approximation respectively.

Apart from inter-speaker variation, intra-speaker variation was also observed. Bauer et al. (2003) investigated tone production by two speakers

[3]The authors did not say this, but as far as I understand, this "midway" tone would be [45] in Chao's tone numerals.

(30 and 35 years of age) in two experimental conditions. In the first condition, participants were instructed to read aloud Cantonese words one by one, each of which had one of the six phonemic Cantonese tones. In the second condition, all words to be read aloud were T2-T5 minimal pairs presented in two columns, and so participants were more aware of what was being tested. The first subject produced relatively different offsets for T2 and T5 in the first condition, but produced only T5-like tones in the second condition. The second subject produced both T2 and T5 with T2-like offsets in both conditions. In sum, their production patterns varied depending on the experimental condition. Seeing that merging may not happen across all words with the same tone, Mok et al. (2013) explore the relationship between word frequency and merger, but did not find any correlation between the two.

Speakers who merge in production do not necessarily merge in perception. In the AX discrimination task of Mok & Wong (2010) and Mok et al. (2013), participants who did not produce the T2-T5 contrast achieved a high discrimination accuracy rate of 90%, although their reaction time was longer than those who did not merge the tones in production. The authors commented that it could be a sign that the merger was still in progress, or the nature of the AX discrimination task caused participants to attend to acoustic details that they might not normally notice in naturalistic spontaneous speech.

If the T2-T5 contrast is difficult for homeland speakers, will it be difficult for heritage speakers as well? On the one hand, if heritage speakers also find these two tones perceptually similar, it is possible that the failure to perceive the contrast will lead to T2-T5 merger. On the other hand, in social dialectology, the language variety spoken in a geographically separated community is relatively more resistant to language change than that of the mainstream linguistic population (Wolfram & Schilling-Estes, 2003). If the parents of heritage speakers in the present study had migrated away from Hong Kong before T2-T5 merger became a trend, they may have passed on a merger-free variety of Cantonese to their children. It remains a question whether the merger phenomenon is unique to the homeland speaker population.

3.4 Perception of Cantonese tones by non-Cantonese speakers

Tonal perception by naïve listeners with no Cantonese proficiency can tease apart which tonal contrasts are perceptually distinct on a universal basis, and which contrasts are better perceived by Cantonese speakers only. Previous studies have looked into the perception of Cantonese tones by native speakers of tone and non-tone languages, including English (Francis, Ciocca, Ma & Fenn, 2008; Qin & Mok, 2011, 2013), French (Qin & Mok, 2011, 2013), Tagalog (Chung, 2009), Thai (Burnham, Lau, Tam & Schoknecht, 2001; Chung, 2009), and Mandarin (Chang et al., 2017; Francis et al., 2008; Lee et al., 1996; Qin & Mok, 2011, 2013). Because English is the dominant language of heritage Cantonese speakers in Canada, this section will only review perception studies about English speakers with no prior Cantonese knowledge. If a dominant L2 is able to impact L1 phonology, heritage speakers' confusion patterns in the current study are expected to share similarities with those of Cantonese-naïve English speakers.

Two studies that implemented different experimental paradigms came to the same conclusion that English speakers find it challenging to distinguish low tones that only differ in the direction of pitch change. In Francis et al. (2008), English speakers with no prior knowledge of any tone language were instructed to perform a forced-choice Cantonese word-identification task before and after a training phase (ten hours in total over the course of 16 to 30 days). In each trial, six tonally contrastive lexical items were shown on the screen along with their pitch contour pictures, romanization, and English translation. Before the training phase, the participants achieved an accuracy rate of at least 80% for T1 [55], T2 [25], and T3 [33]. They had a lower accuracy for the three low tones with different contours, namely T4 [21] (61%), T5 [23] (45%), and T6 [22] (18%). The subjects heard T4 [21] as T6 [22] 32% of the time, and T6 [22] was heard as T4 [21] 21% of the time. T5 [23] was heard as T3 [33] 30% of the time, but confusion in the other direction was rare (only 5%). After the training phase, the subjects' overall accuracy increased by 15%. Their post-training accuracy rates for T4 [21], T5 [23], and T6 [22] were 74%, 77%, and 51% respectively. Although they showed improvement, T6 [22] was still the hardest after the training phase. Results of another study by Qin & Mok (2011) were in accordance with those of Francis et al. (2008). In their AX discrimination task in which participants were asked whether two sounds were the same or different, English speakers erroneously responded with "the same" 60% of the time when T5 [23] and T6 [22] were presented. Both Francis et al. (2008) and Qin & Mok (2011) performed multidimensional scaling analyses for their own similarity rating task. Although their interpretations of the dimensions are different, results of both studies agree that T5 [23] and T6 [22] are very close in the perceptual space of English speakers.

One difference between the two studies' results, however, was the English speakers' confusion patterns of T2 [25], T3 [33], and T5 [23]. In the AX discrimination task of Qin & Mok (2011), the error rate for the pair T2-T5 was 75%—the highest among all pairs. The error rate for the pair T3-T5 was only 12.5%. In other words, the low rising tone T5 [23] was perceived as more similar to the high rising tone T2 [25] than to the mid level tone T3 [33]. In their multidimensional scaling analysis, one dimension was labelled as "starting pitch height". Since T2 [25] and T5 [23] have similar starting pitch heights, they are close along this dimension. On the other hand, T5 [23] and T3 [33] have different starting pitch heights, and so are farther apart along this dimension. By contrast, in Francis et al. (2008), T5 [23] was rarely heard as T2 [25] (5% in pre-training, 2% in post-training); instead, T5 [23] was confused with T3 [33] more often (30% in pre-training, 7% in post-training). Therefore, in this study's multidimensional scaling analysis, one dimension was labelled as "pitch height" in an overall sense, not referring to the starting or ending pitch in particular. Since T5 [23] and T3 [33] have similar overall pitch heights, they are close along this dimension. T2 [25] and T5 [23] have a bigger difference of overall pitch height, and are father apart along this dimension. In sum, English speakers' perception of T2-T5 in Qin & Mok (2011) was more similar to that of homeland Cantonese speakers.

Authors of both studies relate their findings to suprasegmental patterns in English prosody. Qin & Mok (2011) explain that f0 is one of the acoustic correlates of English lexical stress (Lieberman, 1960). In addition, in English intonation there are high or low boundary tones at the edge of phrases or sentences (Liberman, 1975; Pierrehumbert, 1980). Therefore, English speakers have sensitivity to pitch height differences, which explains why they did well in discriminating the three level tones in Cantonese. Since English boundary tones are either high or low, Cantonese low falling, low rising, and low level tones were perceived as the same category, namely "low". Therefore, English speakers did less well in perceiving the difference between level and contour tones within the lower pitch range. Lastly, Francis et al. (2008) point out that English questions have a rising intonation comparable to Cantonese T2 [25]. English speakers may have assimilated a Cantonese lexical category into an English intonational category, and so were able to do well. As for T5 [23], although it is also a rising tone, its magnitude of pitch change may not be sufficient to be assimilated into the rising category in English intonation. This may explain why T5 [23] was rarely confused with T2 [25].

In sum, both Francis et al. (2008) and Qin & Mok (2011) agree that Cantonese-naïve English speakers are more sensitive to the average pitch height than the direction or magnitude of pitch change. If heritage speakers in the present study also have difficulty discriminating T4 [21], T5 [23], and T6 [22], it may be evidence of cross-language effects from the non-lexical suprasegmental phonology of a dominant L2 to the lexical-tonal suprasegmental phonology of a non-dominant L1.

3.5 Tone and heritage speakers of Cantonese

Two studies on adult Cantonese speakers in Canada reported different findings from each other with respect to tonal production and perception. The earlier one by So (2000) was based in Metro Vancouver, British

Columbia, and involved three groups of participants categorized by their age of arrival in Canada: before seven, between 10 and 15, and after 16. The later study by Soo & Monahan (2017) was based in Toronto, Ontario. They recruited participants from Canada and Hong Kong, who were categorized into two groups by their dominant language. A comparison of their methodologies and outcomes is presented as follows.

With regard to production, So (2000) found that participants who moved to Canada before age seven (and so were most similar to heritage speakers as defined in this dissertation in terms of linguistic background) had a significantly smaller tonal space than those who arrived after 16. In her study, target words were elicited in both isolated and embedded forms. f0 values were first transformed into musical semitones, and were further transformed into Chao's tone numerals. Participants who arrived in Canada before age seven produced the high level tone (T1) with a pitch that was comparable to [44]. Since the upper boundary of the tonal space was lower than the usual [55], their whole tonal space was compressed. The offset of the high rising tone (T2, normally [25]) became lower, resulting in a rising tone like [2 3.5]. Similarly, the offset of the low rising tone (T5, normally [23]) also became lower, resulting in a rising tone like [2 2.5]. As a result, the difference of the offsets between T2 and T5 became smaller compared with participants who arrived after age 16, the group that was most similar to homeland speakers. The author also measured percentage change in f0 along different sections of the syllable for contour tones, and confirmed that the two groups produced rising tones with significantly different slopes. These findings coincide with Chang & Yao (2016) on heritage speakers of Mandarin, in that heritage speakers diverged from native norms in terms of tonal production.

Soo & Monahan (2017), however, reported that homeland[4] and heritage speakers produced tones with similar slopes. In their study, participants were instructed to read aloud words in isolation. f0 values were extracted at eight equally spaced points along the duration of the syllable and

[4]In the original study, this group was referred to as "native speakers".

normalized. Results showed that homeland and heritage speakers produced overlapping pitch contours for each of the six lexical tones, and no deviation of tonal production was found in the heritage group.

Results of the perception portion of the two studies also differed. So (2000) adopted the word identification paradigm, where participants heard a monosyllabic word and were asked to choose one of the six pictures that represented what was heard. The average accuracy of those who arrived before age seven (59.38%) was significantly lower than those who arrived after age 16 (87.50%). The former group's confusion patterns showed that the tonal contrasts of T2-T5, T3-T6, and T4-T6 were difficult to perceive. In addition, incorrect T1 responses were spread over to target tones T2, T3, T4, T5, and T6. The author concluded that those who arrived in Canada before age seven experienced confusion with all tones.

Using the AX discrimination paradigm, Soo & Monahan (2017) reported that their two groups' performances did not differ significantly. In each trial participants listened to a pair of tonally contrastive words, and were asked to choose between "same" or "different". D-prime scores of the two groups were compared, and there was no effect of group. The authors admitted that the results were somewhat surprising, and pointed out that the nature of the AX discrimination task taps into the phonetic-perceptual level of processing, but not the abstract-phonological level. While the task was able to show that heritage speakers were aware of the phonetic differences between two stimuli, it was not able to show whether the listeners used this phonetic information to draw lexical contrasts. This comment was similar to that of Mok et al. (2013), who found that homeland speakers merging tones in production could still do well in the AX discrimination task. To sum up, the different methodologies being used, different linguistic dynamics between the Metro Vancouver and Greater Toronto areas, along with a 17-year gap between So (2000) and Soo & Monahan (2017), may have led to different results and conclusions.

3.6 Hypotheses to be tested

Building on previous research discussed in Chapter 2 and Chapter 3, this dissertation posits three hypotheses in (14) with regard to homeland and heritage speakers' perception of Cantonese lexical tones. First, studies like Celata & Cancila (2010) point out that heritage speakers are less sensitive to L1 sound contrasts that are not phonemic in L2, their dominant language. Since Cantonese is a tone language and English is not, I hypothesize that heritage speakers of Cantonese, who are English-dominant, make less use of tonal information for word identification.

Second, basing on the findings of So (2000) that individuals who had migrated from Hong Kong to Canada before seven years of age experienced confusion with all tones, I hypothesize that heritage speakers born and raised in Canada will have a similar pattern, in that they show more confusion than homeland speakers when asked to identify a word from a tonally contrastive set. Although less confusion is anticipated from homeland speakers, they are expected to show confusion between T2 [25] and T5 [23] due to the ongoing sound change mentioned in Section 3.3.2. I am agnostic as to whether heritage speakers follow this particular trend of tone merger.

The last hypothesis stems from the first: if heritage speakers make less use of tonal information, how can they understand a tone language? In everyday utterances, words come in sentences that provide semantic context, which serves as non-acoustic top-down cues for the listener. I hypothesize that semantic information is especially useful for heritage speakers. Even if a word is perceptually confusable with its tonally contrastive set, semantic context can help to resolve any potential ambiguity.

(14) Hypotheses of the current study

a. Compared with homeland speakers, heritage speakers make less use of tonal information for word identification.

b. Homeland and heritage speakers exhibit different confusion patterns with respect to lexical tone perception.

c. Compared with homeland speakers, heritage speakers rely on semantic information to a greater degree.

Apart from filling an empirical gap, the current study also has the objective of building a methodology that can better tackle the following issues. First, how can we be sure that differences between homeland and heritage speakers, if any, are mainly due to their tonal perception abilities, but not due to lower overall Cantonese proficiency, or lower comfort level with performing a Cantonese task in a laboratory setting? Second, in daily language use, words are embedded in sentences and are rarely uttered in isolated forms. How do heritage speakers make use of acoustic tonal information when non-acoustic cues are available as well? The next chapter will explain how variable manipulation in the present study can tease tone apart from general factors, and at the same time put tone into competition with semantic information.

Chapter 4

Methodology

This chapter is a walkthrough of the experimental design and procedures of two pilot studies and the main study. Section 4.1 presents an overview of the word identification paradigm and independent variables being manipulated. Section 4.2 and Section 4.3 provide a detailed explanation of how target words and carrier phrases were selected through Pilot Study 1 and 2 respectively. Section 4.4 describes the materials, procedures, and participants of the main study.

4.1 An overview of the experimental design

Before the discussion on research methodology, I recapitulate the purpose of this study. Research questions listed below are repeated from (1) in Chapter 1.

(1) Research questions of the current study (repeated)

a. Do homeland and heritage speakers behave differently in terms of their ability to identify tonally contrastive words?

b. Do homeland and heritage speakers exhibit similar confusion patterns with respect to lexical tone perception?

c. Do homeland and heritage speakers make use of the same type of information when identifying a word from a tonally contrastive set? In particular, are acoustic and semantic information equally useful?

The following subsections will discuss how the experimental design of the current study addressed these research questions.

4.1.1 The word identification paradigm

To investigate homeland and heritage speakers' tonal perception on the lexical phonological level, a forced-choice word identification paradigm was adopted. Every trial in the experiment followed the same procedures: pictures representing the target word and its tonally contrastive competitors were presented on a computer screen. Each of these pictures corresponded to a button on a response device. At the same time an audio stimulus containing the target word was presented. In all cases, this stimulus ended with the target word, but the exact nature of the stimulus varied during the course of the experiment, as will be described in Section 4.1.3. Participants were asked to identify the word they heard and respond by pushing the corresponding button on the device. Their accuracy rate—the dependent variable of this study—was measured and compared. The use of the same paradigm throughout the study allowed direct comparison of results.

Although only one paradigm was adopted, different types of stimuli were used to address different questions. A summary of these stimuli is given in Table 4.1. The "Type" column shows how each stimulus type is referred to in subsequent discussion. Types 1, 2, and 3 were monosyllablic words for answering the question in (1a) : do homeland and heritage speakers behave differently in terms of their ability to identify tonally contrastive words? Types 4, 5A, 5B, 6A, and 6B were sentences with a semantic context. Their purpose was to answer (1c) : do the two groups find acoustic and semantic information equally useful when discriminating tonally contrastive words? All stimulus types can answer the question in (1b) : do the two groups exhibit similar confusion patterns with respect to lexical tone perception?

Table 4.1: A summary of stimulus types and procedures of the main study

						Day 1	**No. of trials**
					Pre-tasks:	Story listening	1
						Picture learning	31
						Practice trials	8
Block	**Type**	**Segment**	**Tone**	**Context**	**Congruity**	**Example of stimuli**	**No. of trials**
First	1	✓	✓	✗	not appl.	*fan3* “sleep”	60
First	2	✓	✗	✗	not appl.	*fan_* “sleep”	60
First	3	✗	✓	✗	not appl.	*___3* “sleep”	60
Second	4	✓	✗	✓	✓	*sap_ ji_ dim_ zung_ hou_ soeng_ cong_ fan_* “At twelve (you) should go to bed and sleep”	60
					Post-task:	Language background questionnaire	

						Day 2	**No. of trials**
					Pre-tasks:	Picture learning	31
						Practice trials	8
Block	**Type**	**Segment**	**Tone**	**Context**	**Congruity**	**Example of stimuli**	**No. of trials**
Third	5A	✓	✓	✓	✓	*sap6 ji6 dim2 zung1 hou2 soeng5 cong4 fan3* “At twelve (you) should go to bed and sleep”	60
Third	5B	✓	✓	✓	✗	*sap6 ji6 dim2 zung1 hou2 soeng5 cong4 fan2* “At twelve (you) should go to bed and powder”	60*3
Third	6A	✗	✓	✓	✓	*sap6 ji6 dim2 zung1 hou2 soeng5 cong4 ___3* “At twelve (you) should go to bed and sleep”	60
Third	6B	✗	✓	✓	✗	*sap6 ji6 dim2 zung1 hou2 soeng5 cong4 ___2* “At twelve (you) should go to bed and powder”	60*3

4.1.2 Variables being controlled

Although Types 1, 2, and 3 were all monosyllables and Types 4, 5A, 5B, 6A, and 6B were all sentences, each of them represented a different way of controlling four independent variables in Figure 4.1: the availability of (i) segmental information, (ii) tonal information, (iii) semantic context, and (iv) semantic congruity of the target word with the carrier phrase. In Table 4.1, the tick ✓ and cross ✗ signs in the columns "Segment", "Tone", "Context", and "Congruity" indicate how each stimulus type manipulated these variables. Note that "congruity" differs from the other three variables in two ways. First, its applicability depends on another variable, namely "context". If a stimulus was a monosyllabic word, it would have no semantic context, and in this case "congruity" would be "not applicable", as in Types 1–3 in Table 4.1. Therefore, "congruity" is immediately under "context" in Figure 4.1, rather than immediately under "semantic information". Second, "segmental information", "tonal information", and "semantic context" were either present or absent; for congruity, when applicable, its configuration was either congruous or incongruous, but not present or absent.

4.1.3 Stimulus types

This subsection explains the purpose of the eight stimulus types. (For details of procedures, such as the number of trials per stimulus type, see Section 4.4.2.) The expected result will also be discussed based on the hypothesis that heritage speakers rely less on tonal information but more on semantic information.

4.1.3.1 Monosyllabic words

The first three types of stimuli were all monosyllabic words. (For details of how these words were chosen, see Section 4.2.) They were similar in terms of the lack of semantic context. Therefore, a cross sign ✗ was put in the "Context" column for Types 1, 2 and 3. Since the target words were not in a carrier sentence, semantic congruity was not applicable.

The difference among Types 1, 2, and 3 lay in the kind of acoustic

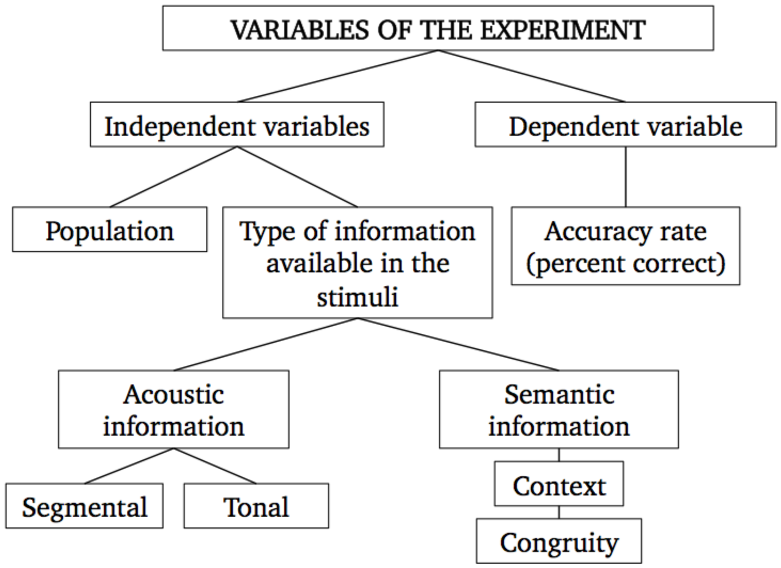


Figure 4.1: A summary of dependent and independent variables of this study

information that was available. Type 1 stimuli were unmanipulated, so both segmental and tonal information were available. For example, participants would hear *fan3*, and had to decide whether the word was *fan2* "powder", *fan3* "sleep", or some other tonally contrastive competitor shown on the screen. Since semantic information was unavailable but all acoustic information was available, this stimulus type was expected to be fairly challenging but not extremely challenging for heritage speakers. As for homeland speakers who were hypothesized to rely less on semantic information, Type 1 was not expected to be challenging at all.

Type 1 alone, however, would not be adequate to answer the first research question. Even if the heritage group indeed had a lower accuracy than the homeland group, it could be due to general language proficiency differences, such as weaker listening comprehension skills or lower comfort level to complete a task in their non-dominant language. In this case heritage speakers would perform worse than homeland speakers in any kind of Cantonese listening tasks. Therefore, this stimulus type alone would not be sufficient to tease tonal perception abilities apart from overall language abilities.

Types 2 and 3 were therefore designed to contrast the two populations' tone discrimination abilities specifically. Containing only segmental but not tonal information, Type 2 was a control task that was expected to be equally challenging for both populations. (For details of how tonal information was removed, see Section 4.4.1.1.) For example, participants would hear *fan*_, and had to decide whether the word was *fan2* "powder", *fan3* "sleep", or some other tonally contrastive competitor shown on the screen. Since the target words and their tonally contrastive competitors were segmentally the same (all were *fan*), it was anticipated that participants could only make a guess what the stimulus was. The expected accuracy for both groups was equal to chance, which would show that heritage speakers did not simply do worse than homeland speakers in *any* kind of Cantonese listening task.

Unlike Type 2, Type 3 only contained tonal information but not segmental information. (For details of how segmental information was removed, see Section 4.4.1.1.) Pitch was the only available acoustic information, and no consonants or vowels could be identified. For example, participants would hear ___*3* (a pitch that corresponded to T3 [33]), and had to decide whether the word was *fan3* "sleep", *fan2* "powder", or some other competitor. Since the target word and its tonally contrastive competitors were segmentally the same, segmental information was expected to be redundant for these stimuli. A person who is capable of using tonal information should be able to identify the target word even when segmental information is unavailable. As Type 1 was unmanipulated speech but Type 3 sounded unnatural (comparable to hummed speech), it was expected that both populations would show a higher accuracy for Type 1 than Type 3. Given the hypothesis that tonal information was less useful for heritage speakers, Type 3 was anticipated to be particularly difficult for the heritage group, so the biggest accuracy gap between homeland and heritage speakers was expected.

Types 1, 2, and 3 were randomized in the same block during the experiment. They were not put into separate blocks, because Type 2 (and possibly Type 3) in its own block could be a potentially frustrating task. To motivate participants to pay attention, monosyllabic stimuli of different levels of difficulty were mixed in the same block.

4.1.3.2 Sentences

Designed to answer the third research question regarding the use of acoustic versus semantic information, the rest of the stimulus types (4, 5A, 5B, 6A, 6B) offered semantic information in addition to segmental or tonal information. These stimuli had target words embedded in a carrier sentence. (For details of how to decide which sentences to use, see Section 4.3.) Since a context was provided, all of them got a ✓ sign for the "Context" column in Table 4.1.

Presented on Day 1 after Types 1–3, Type 4 was a control task containing sentences with no tonal information. Participants were asked to identify the last word of the sentence, such as *sap_ ji_ dim_ zung_ hou_ soeng_ cong_ fan_* "At twelve (you) should go to bed and sleep", where the underscore indicates an absence of pitch information. The monotonous stimuli resembled alaryngeal speech produced by individuals who use an electrolarynx to speak after surgical removal of the larynx. Law, Ma & Yiu (2009) report that although alaryngeal Cantonese speakers had difficulties producing varying pitches, the sentences that they produced were fairly intelligible. In their sentence intelligibility test, Cantonese speakers of age 25–33 with no speech impairment and no prior experience with alaryngeal speech were asked to transcribe sentences uttered by electrolarynx users. According to their results, electrolaryngeal speech received an average intelligibility score of 77.3%. For this reason, in the current study, monotonous Type 4 stimuli were expected to be reasonably comprehensible as well, and semantic information was considered available in these stimuli. Therefore in Table 4.1, Type 4 got a ✓ sign for the "Context" column. Since the (non-)use of tonal information was hypothesized to be the key difference between homeland and heritage speakers, it was anticipated that removing such information would render the two groups equal. The predicted result for this task was that heritage and homeland speakers would achieve a similar level of accuracy.

The rest of the stimulus types were presented on Day 2 of the experiment. Types 5A and 5B contained the same acoustic information (hence both were called Type 5), and they only differed by congruity: 5A was congruous but 5B was incongruous. Acoustically they were both unmanipulated, hence keeping all segmental and tonal information. However, semantically, Types 5A and 5B were not the same. For Type 5A, the target word was congruous with its carrier phrase in terms of meaning. For example, in *sap6 ji6 dim2 zung1 hou2 soeng5 cong4 fan3* "At twelve (you) should go to bed and sleep", the target word *fan3* "sleep" was relevant to the rest of the sentence, as the function of a bed is to provide a sleeping surface. Note that 5A was the only stimulus type that received a ✓ mark for all four variables. Since it offered the most information, it should be the easiest task for both populations.

Type 5A alone, however, would not be sufficient to answer the third research question. If someone obtained a high accuracy for this stimulus type, there could be several possible explanations. First, s/he attended to the tonal information and made a decision solely based on what was available in the acoustic signal. Second, s/he did not pay attention to the tonal information at all, but simply chose a word that would make sense in a given sentence. Third, s/he made use of both the acoustic and semantic information in the stimuli. This would fail to answer the question of whether heritage speakers tend to rely more on one type of information than another. The next stimulus types were designed to solve this problem.

Type 5B stimuli were sentences that did not make sense, as the target word was semantically incongruous with the carrier phrase. An example was *sap6 ji6 dim2 zung1 hou2 soeng5 cong4 fan2* "At twelve (you) should go to bed and powder". In terms of meaning there is no relationship between going to bed at twelve and the action of powdering. Note that in the beginning of the experiment, participants were explicitly told that some sentences might not make sense, and all they had to do was to identify the last word that they heard. (For details of instructions given to participants, see Section 4.4.1.4.) In other words, they were instructed to attend to acoustic information. If someone was able to pick the correct answer (*fan2* “powder”), it would show that s/he was able to use tonal information and actively choose not to rely on the semantic context. However, if someone consistently chose the one that made sense (*fan3* “sleep”), it would suggest that s/he over-relied on semantic information. Given the hypothesis that heritage speakers rely more on semantic than acoustic information, it was expected that this task would be significantly more challenging for heritage than homeland speakers.

The last two stimulus types, namely 6A and 6B, contained less acoustic information compared to 5A and 5B: all segmental information of the target word[1] was removed, but its tonal information was kept. Type 6A was the congruous version, such as *sap6 ji6 dim2 zung1 hou2 soeng5 cong4 __3* “At twelve (you) should go to bed and sleep”. Type 6B, on the other hand, was the incongruous version, such as *sap6 ji6 dim2 zung1 hou2 soeng5 cong4 __2* “At twelve (you) should go to bed and powder”. Lacking segmental information, these two stimulus types were expected to be more challenging than 5A and 5B for heritage speakers.

Types 5A, 5B, 6A, and 6B were randomized and presented as one block on Day 2. If they were presented in the order of 5A, 5B, 6A, and 6B as four separate blocks, it would potentially be a frustrating experience for the participants, as they would have to listen to many incongruous sentences in a row (5B and 6B), and they would also have to listen to many sentences in which the last word had no segments in a row (6A and 6B). In this case, even if Type 5A had the highest accuracy and 6B had a low accuracy as expected, it could be merely due to participants’ loss of attention over the course of the experiment session. To maintain morale of the participants and to avoid potential noise in the results, these stimulus types were mixed in the same block.

[1]Segmental information of words in the carrier phrase was not removed (if removed, it would be like *__6 __6 __2 __1 __2 __5 __4 __3* “At twelve (you) should go to bed and sleep”), because the carrier phrase would be incomprehensible, hence unable to create a semantic context to get the predicted effect.

4.1.4 Summary

To sum up, only one paradigm, namely the word identification paradigm, was adopted for easy comparison of results. Eight types of stimuli were designed to control four variables: the availability of segmental information, tonal information, semantic context, and semantic congruity. The stimulus types in Table 4.2 are the same as those of Table 4.1, but they are arranged by the anticipated accuracy gap between homeland and heritage speakers. The predictions were based on the hypothesis that tonal information was least useful while semantic context was most useful for heritage speakers. It was expected that both populations would have similar performance for stimuli with no tone (Types 2 and 4). Type 5A which contained everything should reflect the baseline Cantonese listening abilities of the two groups. Types 1, 5B, 6A, and 6B were predicted to be more difficult as they lacked certain information. Lastly, Type 3 only contained tonal information, which was hypothesized to be least useful for heritage speakers. As a result, their accuracy was predicted to be far lower than that of homeland speakers.

Table 4.2: Summary of stimulus types, arranged by the anticipated accuracy gap between homeland and heritage speakers (from smallest to largest)

Type	Description	Segment	Tone	Context	Congruity	Predicted result		
2	Words with no tone	✓	✗	✗	Not appl.	Homeland	=	Heritage
4	Sentences with no tone	✓	✗	✓	✓	Homeland	=	Heritage
5A	Normal sentences (congruous)	✓	✓	✓	✓	Homeland	>	Heritage
5B	Normal sentences (incongruous)	✓	✓	✓	✗	Homeland	>>	Heritage
6A	The last word of the sentence has no segments (congruous)	✗	✓	✓	✓	Homeland	>>	Heritage
6B	The last word of the sentence has no segments (incongruous)	✗	✓	✓	✗	Homeland	>>	Heritage
1	Normal words	✓	✓	✗	Not appl.	Homeland	>>	Heritage
3	Words with no segments	✗	✓	✗	Not appl.	Homeland	>>>	Heritage

4.2 Pilot Study 1: Familiarity with target words

This section explains the purpose, procedures, and results of Pilot Study 1, a word familiarity rating task that determined target words to be used for the main study.

4.2.1 Background and purpose

Due to heritage speakers' weaker Chinese literacy skills on average, choosing target words for a Cantonese word-identification task can be challenging. In the present study, the following criteria were used for choosing target words: availability of minimally contrastive sets, usage in daily life, imageability, and lastly, familiarity, which is the focus of Pilot Study 1. Each of these criteria is explained as follows.

Since this dissertation aims at investigating homeland and heritage speakers' tone discrimination abilities, the target words to be used should be tonally contrastive minimal sets. In other words, members of a word set should have the same segmental phonemes, but differ in lexical tone only. In previous works on homeland Cantonese speakers (e.g., Burnham, Ciocca, Lauw, Lau & Stokes, 2000; Ciocca & Lui, 2003; Fok-Chan, 1974), minimal sextuplets were used as target words, which are summarized in Table 4.3. Although all phonotactically well-formed syllables without an obstruent coda can be uttered with any of the six lexical tones, not all syllable-tone matches are meaningful. Take the syllable *wu* as an example: *wu1* "dirt", *wu2* "pot", *wu3* "nuisance", *wu4* "lake", and *wu6* "mutual" are all real words, but *wu5* does not mean anything in Cantonese. In other words, *wu5* is an accidental gap. In fact, such gaps are everywhere in the lexicon. Only a limited set of syllables (*fan*, *fu*, *jan*, *jau*, *ji*, *jyun*, *lau*, *se*, *seoi*, *si*, and *wai*) can match with every single lexical tone and result in a meaningful word.

In addition to the limited number of full contrastive sets, a word's usage in daily life (and the lack thereof) should be taken into consideration, particularly for a study about heritage speakers. For instance, the syllable *jyun* has a full contrastive set, as in Table 4.3. However, the word *jyun1*

"Mandarin duck" is commonly used in romantic poetry, but rarely heard in daily conversations. Since heritage speakers acquire the language in a family setting and receive education in a dominantly English environment, it is unlikely for them to be familiar with this literary term. Furthermore, as discussed in Section 1.4.7 previously, certain words are more frequently used in Standard Chinese, which must be learned in formal language classes. Examples from Table 4.3 include *fan5* "diligence", *jan6* "pregnancy", *seoi5* "mental state", *jau3* "young", *se1* "some", *wai5* "great", *jau1* "rest", and *ji3* "idea", which have more colloquial counterparts in spoken Cantonese. Using these words in the study may lead to undesirable consequences: if heritage speakers demonstrate a low accuracy in a word-identification task, it would be difficult to conclude whether it is due to their inability to distinguish lexical tones, or their lack of familiarity with these lexical items.

Meanings that are hard to express through pictures cannot be used either. Written Chinese characters are largely opaque to their pronunciation and must be learned through formal education. As data from the linguistic background questionnaire in Section 4.4.3.3 will show, most heritage speakers in the subject pool were not literate in Chinese. To eliminate the possibility that a low accuracy rate was due to participants' lack of reading skills, pictures instead of written characters were presented as choices in the experiment. Among the words in Table 4.3, *jau5* "have", *lau3* "instigate", *seoi3* "tax", *wai1* "might", and *wai6* "position" have relatively abstract meanings that are difficult to be represented by drawings. Therefore, even though they are used in daily contexts, their respective tonal sextuplets were not used.

So far within the limited set of syllables that can generate meaningful tonal sextuplets, *fan*, *jan*, *jau*, *ji*, *jyun*, *lau*, *se*, *seoi*, and *wai* have failed to meet the criteria above perfectly; only *si* and *fu* were left as usable candidates. If all 720 trials in the current study were *si* and *fu* only, the experiment might become a tedious task, and participants might lose their attention quickly, affecting their accuracy. In addition, using a variety of syllables and a robust set of lexical items can enhance the generalizability of results. It was thus necessary to explore options other than tonal sextuplets.

Table 4.3: Minimal sextuplets used in previous studies

Author(s)	Syllable	T1	T2	T3	T4	T5	T6
Fok-Chan (1974)	*fu*	man	bitter	richness	help	woman	father
	jyun	Mandarin duck	gentle	grumble	finish	soft	wish
Yiu & Fok (1995)	*ji*	cure	chair	idea	child	ear	two
Burnham et al. (2000)	*fu*	husband	tiger	rich	hold	woman	father
So (2000)	*fu*	husband	tiger	pants	symbol	woman	tofu
	si	lion	history	attempt	time	city	trained person
Ciocca & Lui (2003)	*ji*	clothing	chair	Italian	child	ear	two
Khouw & Ciocca (2007)	*si*	poetry	history	try	time	city	surname
Francis et al. (2008)	*se*	some	write	spill	snake	society	shoot
	si	thought	history	try	time	market	event
	jau	rest	grapefruit	young	from	have	and
	fan	separate	noodle	command	burn	diligence	portion
	fu	husband	bitter	rich	appropriate	woman	negative
	ji	cure	chair	meaning	son	ear	two
Kung et al. (2014)	*jan*	joy	endure	print	human	pull/draw	pregnancy
	si	teacher	history	try	time	market	matter
	seoi	bad	water	tax	hang down	mental state	sleep
	fu	husband	bitter	rich	symbol	woman	negative
	wai	might	destroy	comfort	encircle	great	position
Yu & Lam (2014)	*lau*	angry	twist	instigate	stay	willow	leak
Lam et al. (2016)	*fu*	exhale	bitter	wealth	match	woman	load
	se	some	write	diarrhea	snake	society	shoot
	si	poetry	history	time	time	market	right

A solution to the aforementioned problem was to use counterbalanced tonal quadruplets instead of sextuplets, as in Table 4.4. The first column of the table shows the list of all possible four-tone combinations. Since there are six lexical tones in the inventory, the number of possible four-tone combinations is 15. All permutations of the same four tones (e.g., [1 2 3 4], [2 1 3 4], [3 1 2 4], and [4 1 2 3]) were considered the same set, namely [1 2 3 4]. (For details of how the four tones were ordered in the main study, see Section 4.4.1.3.) The second column of the table shows the five syllables to be used: *fan*, *fu*, *ji*, *se*, and *si*. Each syllable was matched with three tone sets. For example, the syllable *ji* was matched with [1 2 4 5], [1 4 5 6], and [2 4 5 6], thus avoiding the abstract word *ji3* "idea". Similarly, the syllable *se* was matched with [2 3 4 5], [2 3 5 6], and [3 4 5 6], hence avoiding *se1* "some", which is hard to be represented by a picture. Lastly, *fan* was matched with [1 2 3 6], [1 3 4 6], and [2 3 4 6], avoiding *fan5* "diligence", which is more commonly used in Standard Chinese. To sum up, three syllables (*ji, se, fan*) occurred with only five of the six tones, while the other two syllables (*fu, si*) occurred with each of the six tones. The number of unique words was therefore (3*5 + 2*6), yielding 27 .

If a syllable-tone match has many homophones, the meaning with the highest imageability (along with familiarity, which will be discussed later) would be chosen. As shown in Table 4.3, *si1* has many homophones. When spoken, it could mean "lion", "poetry", "thought", or "teacher". Since pictures were to be presented to elicit response, "poetry" and "thought" were less desirable candidates; instead, "lion" was chosen to be used in the experiment. A noteworthy case is *si2*, which can mean "history" or "poop" unless disambiguated by writing. In all studies listed in Table 4.3, only the character for "history" was used, possibly because words related to bodily functions were not preferred in contexts like an academic experiment. For the purpose of the current study, however, *si2* "poop" matched more criteria to be a usable target word. First, it can easily be represented by a picture (see Appendix A for the actual picture used in the main study). Second, it is a common word in infant-directed speech. According to Frank, Braginsky, Yurovsky & Marchman (2017) and Tardif, Fletcher, Liang & Kaciroti (2009),

Table 4.4: Tonal quadruplets used in the current study

Tone set	Syllable	T1	T2	T3	T4	T5	T6
1 2 3 4	*si*	lion	poop	try	key		
1 2 3 5	*fu*	exhale	tiger	pants		woman	
1 2 3 6	*fan*	share	powder	sleep			portion
1 2 4 5	*ji*	cure	chair		child	ear	
1 2 4 6	*si*	lion	poop		key		nurse/trained person
1 2 5 6	*si*	lion	poop			market	nurse/trained person
1 3 4 5	*fu*	exhale		pants	help	woman	
1 3 4 6	*fan*	share		sleep	tomb		portion
1 3 5 6	*fu*	exhale		pants		woman	negative
1 4 5 6	*ji*	cure			child	ear	two
2 3 4 5	*se*		write	diarrhea	snake	society	
2 3 4 6	*fan*		powder	sleep	tomb		portion
2 3 5 6	*se*		write	diarrhea		society	shoot
2 4 5 6	*ji*		chair		child	ear	two
3 4 5 6	*se*			diarrhea	snake	society	shoot

Total number of unique words: 27

most Cantonese-learning infants from Hong Kong had acquired the word *si2* "poop" by 30 months of age. Therefore, adult heritage speakers who grew up listening to Cantonese in a family setting were expected to be familiar with this lexical item.

Lastly, similar pictures within the same tone set were avoided. Consider the example of *fu* from the first row of Table 4.3. *Fu1* "man" and *fu6* "father" are both male human beings. It was possible to draw a baby next to the "father" to distinguish it from "man", but still "father" logically entails "man". To solve this problem, only words that do not incur semantic confusion with another word in the tone set were chosen. For instance, in Table 4.4, *fu1* "exhale", *fu2* "tiger", *fu3*, "pants", *fu4* "help by holding someone's arm", *fu5* "woman", and *fu6* "negative" are all semantically distinct. The same principle was applied to pictures for all tone sets. For more details about pictures, see Section 4.4.1.2 and Appendix A.

The last criterion for choosing target words was familiarity, which is not the same as word frequency. Familiarity is a subjective measure based on familiarity ratings to reflect a speaker's experiential encounter with a lexical item (Connine, Mullennix, Shernoff & Yelen, 1990). Word frequency, on the other hand, is an objective measure based on the number of occurrences of a word, usually in a corpus. A word can have a low frequency but receive a high familiarity rating. Using the data from *The Teacher's Word Book of 30,000 Words* (Thorndike & Lorge, 1963), Gernsbacher (1984) points out that the English word *ultra* had a low frequency (237 occurrences per one million words), but it was rated as highly familiar, comparable to *super*, a word that was both highly familiar and highly frequent (8,031 occurrences per one million words). Although *ultra* (237 occurrences) was similar to *twixt* (287 occurrences) in terms of word frequency, *ultra* received a high familiarity rating, while *twixt* had a low familiarity rating. A summary of the comparison is provided in Table 4.5. These examples show that there is no direct mapping between word frequency and familiarity.

Table 4.5: Comparison of word frequency per one million words and familiarity ratings of three English words (Gernsbacher, 1984)

Word	Frequency	Familiarity rating
super	high (8,031)	high
ultra	low (237)	high
twixt	low (287)	low

As for Cantonese, word frequency data are available from the Hong Kong Cantonese Corpus (Luke & Wong, 2015), a database comprising 180,000 word tokens and representing spontaneous speech produced by 100 speakers from Hong Kong. A word search for the proposed target words in Table 4.4 was performed, and their frequency in the corpus is summarized in Table 4.6. A large range of frequency can be observed: *ji6* "two" was very frequent (199 occurrences), but *ji2* "chair" was not even found in the corpus (0 occurrences). If word frequency was the sole criterion for choosing target words, most of the words in Table 4.6 could not be used.

Table 4.6: Word frequency of target words out of a total of 180,000 word tokens in the Hong Kong Cantonese Corpus (Luke & Wong, 2015)

Written form	Word	Meaning	Frequency
分	*fan1*	share	124
粉	*fan2*	powder	7
瞓	*fan3*	sleep	30
墳	*fan4*	tomb	0
份	*fan6*	portion	12
呼	*fu1*	exhale	8
虎	*fu2*	tiger	1
褲	*fu3*	pants	6
扶	*fu4*	help by holding another person's arm	1
婦	*fu5*	woman	9
負	*fu6*	negative	10
醫	*ji1*	cure	60
椅	*ji2*	chair	0
兒	*ji4*	child	2
耳	*ji5*	ear	6
二	*ji6*	two	199
寫	*se2*	write	68
瀉	*se3*	diarrhea	1
蛇	*se4*	snake	6
社	*se5*	society	1
射	*se6*	shoot	5
獅	*si1*	lion	3
屎	*si2*	poop	1
試	*si3*	try	77
匙	*si4*	key	3
市	*si5*	market	1
士	*si6*	nurse/trained person	43

Since word frequency does not equal familiarity, a separate measure for familiarity was used. To confirm that Cantonese speakers were indeed familiar with the words listed in Table 4.4, Pilot Study 1 was conducted in the form of a word familiarity rating task. Although it was assumed that all homeland speakers would be very familiar with all proposed target words, it was not the case for heritage speakers, who had varied exposure to the language. It was thus necessary to make sure that heritage speakers were reasonably familiar with these words. If they did not know a particular word at all, they would not know its tone either. In this case a low accuracy rate for that word in the main study would not help conclude anything about their tone discrimination ability. For this reason, the heritage group was the population of interest of this pilot study, even though data from both homeland and heritage speakers were collected and analyzed.

4.2.2 Procedures

Pilot Study 1 was conducted in the form of an online questionnaire hosted on UBC FluidSurveys (FluidSurveys, 2017). Since the main target population of this pilot study was heritage speakers, the online survey was circulated on social media platforms mainly among UBC student communities and interest clubs, rather than Hong Kong-based online communities. The first page of the survey was a consent form, stating that participation was on a voluntary basis, no personally identifiable information would be elicited, and all collected data would be used for academic purpose only. Anyone who had at least one Cantonese-speaking parent was invited to participate. A high fluency in Cantonese or the ability to read Chinese was not required. Since there were international students from Hong Kong studying at UBC, it was anticipated that the survey might reach homeland speakers as well. To separate their data from heritage speakers', a language background questionnaire[2] was included between the consent form and the main part of the pilot study. It contained questions on language proficiency, countries lived in, and parents' native languages.

The main part of the questionnaire was entitled "How familiar are you with this SPOKEN word?" The word "spoken" was emphasized, so it was clear that participants were not supposed to give ratings for a written character. A list of 27 target words (see Table 4.4) was presented in random

[2]The language background questionnaire used in Pilot Study 1 was a condensed version of the one used for the main study. For details of the latter, see Section 4.4.3.2.

order. For each word, its romanized form and translation were provided, while its written form in Traditional Chinese was displayed in parentheses for reference. Since heritage speakers were expected to have limited reading skills for Chinese, the audio file of the word was available as well, as in Figure 4.2. Participants could click and listen to the word spoken by the author. They were asked to rate its familiarity on a four-point scale: “very familiar”, “quite familiar”, “not so familiar”, and “not familiar at all”.

Figure 4.2: A screenshot of Pilot Study 1

4.2.3 Participants

A total of 648 individuals attempted the questionnaire. Responses from 286 individuals were excluded from analysis, either because they did not complete the whole survey, or they did not have any Cantonese-speaking parents. Among the 362 individuals who answered all questions on the survey and reported to have Cantonese-speaking parents, 237 of them grew up in Canada and so were categorized as heritage speakers. As expected, although the online survey was mainly circulated among communities of a Canadian university, it also reached international students originally from Hong Kong, or friends and relatives of heritage speakers from Hong Kong. Among the 362 individuals whose data were analyzed, 125 of them grew up in Hong Kong, and so were categorized as homeland speakers.

4.2.4 Results

Results of the Pilot Study 1 are presented as histograms in Figure 4.3 and as boxplots in Figure 4.4. The former compares the overall distribution of ratings by the two groups, while the latter contrasts their familiarity with individual lexical items. Each figure is discussed as follows.

Figure 4.3a and Figure 4.3b show that both homeland and heritage speakers rated most lexical items as "very familiar". Although the number of data points was different between homeland (125 participants * 27 words = 3,429 data points) and heritage speakers (237 participants * 27 words = 6,399 data points), the purpose of the histograms was to compare the overall distribution of their ratings, rather than the number of responses. In both figures, the x-axis represents familiarity ratings on a four-point scale. As expected, almost all lexical items were rated "4" (very familiar) by homeland speakers in Figure 4.3a. There were only two responses of "1" (not familiar at all), two responses for "2" (not so familiar), and 15 responses for "3" (quite familiar), which can barely be seen in Figure 4.3a. As for heritage speakers in Figure 4.3b, they had relatively more ratings below "4" compared with homeland speakers: there were 429 responses for "1" (not familiar at all), 440 responses for "2" (not so familiar), and 482 responses for "3" (quite familiar). However, there were still 5,026 responses for "4" (very familiar), which outnumbered the other responses predominantly. It can be concluded that both groups shared a skewed distribution towards "very familiar", although the homeland group's distribution was even more skewed than that of the heritage group.

The boxplots in Figure 4.4 contrast the two groups' ratings for individual words. In both figures, the y-axis represents ratings on the same four-point scale. Homeland speakers in Figure 4.4a rated almost all words as "very familiar" (M = 3.99, SD = 0.01). As for heritage speakers in Figure 4.4b, the mean rating of all words was between "quite familiar" and "very familiar" (M = 3.58, SD = 0.32). Words that received the lowest ratings were *fu6* "negative" (M = 3.00, SD = 0.03), *fan4* "tomb" (M = 2.94, SD = 0.03), and *se3* "diarrhea" (M = 2.89, SD = 0.03). Even though their ratings were relatively low, they were still very close to 3, which translates to "quite familiar".

To sum up, Pilot Study 1 confirmed that heritage speakers were reasonably familiar with the 27 proposed target words listed in Table 4.4. Therefore, they were appropriate to be used as target words in the main study[3].

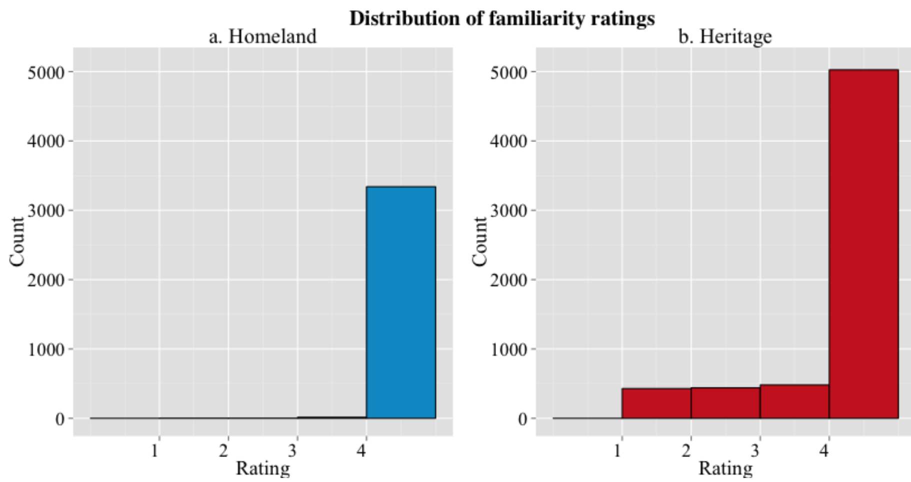


Figure 4.3: Distribution of homeland and heritage speakers' familiarity ratings; 1="not familiar at all", 4="very familiar"

[3]In the main study, a picture learning task was inserted before the experimental block as an extra measure to ensure that heritage speakers were aware of the meaning of all target words, especially those that received relatively low ratings in Pilot Study 1. For details of the picture learning task, see Section 4.4.2.2

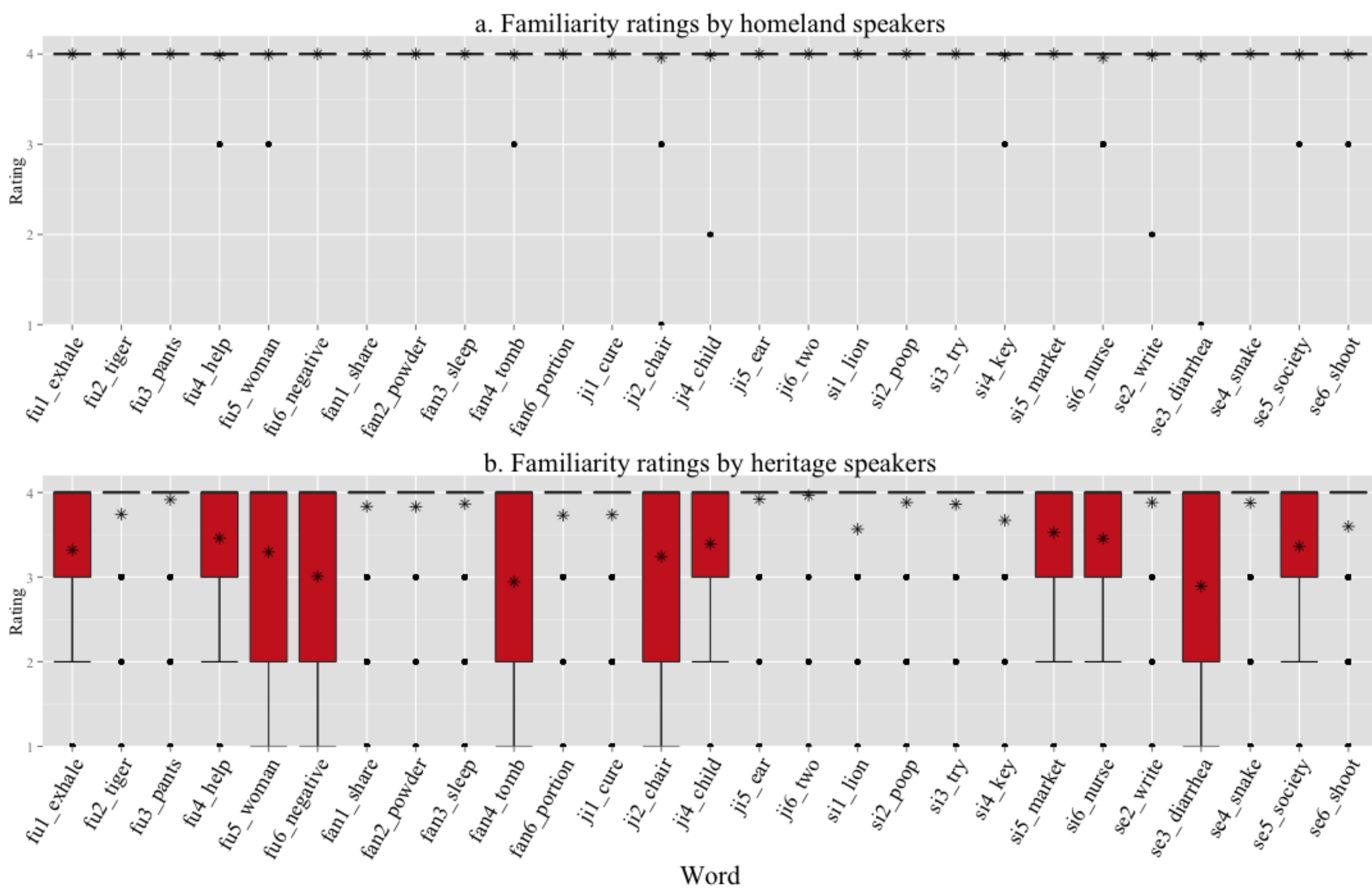


Figure 4.4: Comparison of homeland and heritage speakers' ratings for individual words; 1="not familiar at all", 4="very familiar"

4.3 Pilot Study 2: Semantic congruity of sentences

This section explains how carrier phrases were chosen for the experiment, which involved Pilot Study 2, a congruity judgment task.

4.3.1 Background and purpose

As mentioned in Section 4.1.3.2, Types 5 and 6 stimuli were sentences rather than monosyllabic words. Each of these sentences consisted of an initial introductory carrier phrase, with the target word following as the last word. Three factors were considered in the process of choosing carrier phrases.

First, the length of the sentences was controlled. If some sentences were shorter than others, different sentence-final pitch declination patterns might be observed (Li et al., 2002). To avoid any potential effects of pitch declination on tonal perception, the number of syllables of all carrier phrases was set to seven, and so the number of syllables of a complete sentence (carrier plus target word) was always eight. See Appendix A for the full list of 147 sentences.

Second, to avoid potential confusion, none of the carrier phrases contained any of the 27 words listed in Table 4.4. For example, a sentence like *bei2 ji1 sang1 tai2 haa5 nei5 zek3 ji5* "Let the doctor see your ear" would not be usable, because the carrier phrase contained the syllable *ji1* "cure" (as in *ji1 sang1* "doctor, one who cures"), which might be confused with the target word *ji5* "ear". It was possible that some participants might choose the picture for "cure" in this case, even though they were instructed to pay attention to the last word.

Last but not least, since one goal of this study is to compare how homeland and heritage speakers make use of acoustic and semantic cues, it was important that carrier phrases created a semantic context that was more relevant to a particular word than its tonal competitors. Table 4.7 shows four examples involving the tone set [1 2 3 4]. The carrier phrase *gwo3 nin4 tong4 jan4 gaai1 jau5 mou5...* "There is ... dance in Chinatown during the Lunar New Year" contained the words "Chinatown", "Lunar New Year", and "dance", which constructed a semantic context that was likely to be associated with "lion", since the lion dance is a traditional art performed during the Lunar New Year. The word *si1* "lion" was therefore semantically congruous with this carrier phrase. Its tonal competitors, however, yield semantically incongruous sentences, namely "There is poop dance in Chinatown during the Lunar New Year", "There is try dance in Chinatown during the Lunar New Year", and "There is key dance during the Lunar New Year". To confirm that other Cantonese speakers had the same congruity judgments as the author, Pilot Study 2 was conducted.

Table 4.7: Examples of carrier phrases and (in)congruous target words

Carrier phrase	Congruous	Incongruous		
gwo3 nin4 tong4 jan4 gaai1 jau5 mou5... There is ... dance in Chinatown during the Lunar New Year	*si1* lion	*si2* poop	*si3* try	*si4* key
fong3 gau2 gei3 zyu6 zap1 faan1 di1... Clean up the ... after walking your dog	*si2* poop	*si1* lion	*si3* try	*si4* key
gam1 ci3 m4 dak1 haa6 ci3 zoi3... If you fail this time, ... again next time	*si3* try	*si1* lion	*si2* poop	*si4* key
ceot1 mun4 hau2 gei3 dak1 daai3 so2... Remember to take your ... when leaving home	*si4* key	*si1* lion	*si2* poop	*si3* try

4.3.2 Procedures

Similar to Pilot Study 1, Pilot Study 2 was conducted in the form of an online questionnaire hosted on UBC FluidSurveys (FluidSurveys, 2017). Since this pilot study did not target at a particular population, it was circulated on social media platforms of university student groups in UBC as well as universities in Hong Kong. Individuals who had Cantonese-speaking parents were invited to complete the study. The first page of the questionnaire was a consent form, followed by a short survey of language background.

The main part of the questionnaire had a header that read in Traditional Chinese "Do you think the sentences below make sense?", followed by a list of sentences, as in Figure 4.5. Congruous (N = 27) and incongruous (N = 127) sentences were randomized. Two options were offered below each sentence, namely "it makes sense" (literally "the meaning is complete" in Cantonese) and "it does not make sense" (literally "the meaning is not through" in Cantonese).

Unlike Pilot Study 1 which provided bilingual instructions, Pilot Study 2 was in written Chinese only. The purpose was to disambiguate tonally similar words. If the sentences were presented in an audio format only, some participants might answer "it makes sense" for an intentionally incongruous sentence, simply because they thought a congruous sentence was uttered in a funny way or the speaker had an accent. Since Chinese literacy was required to complete the survey, it was anticipated that most participants of Pilot Study 2 were homeland speakers.

Figure 4.5: A screenshot of Pilot Study 2

4.3.3 Participants

A total of 541 individuals attempted the online survey, but only 371 of them had Cantonese-speaking parents and completed all questions. The data of these 371 individuals were included in the following analysis.

4.3.4 Results

Results of Pilot Study 2 are summarized in Figure 4.6. The x-axis shows two groups of sentences: those intended to be congruous (e.g. “There is lion dance in Chinatown during the Lunar New Year”) and those intended to be incongruous (e.g. “There is key dance in Chinatown during the Lunar New Year”). The y-axis represents mean ratings given by participants: 1 means “it makes sense”, while 0 translates to “it does not make sense”.

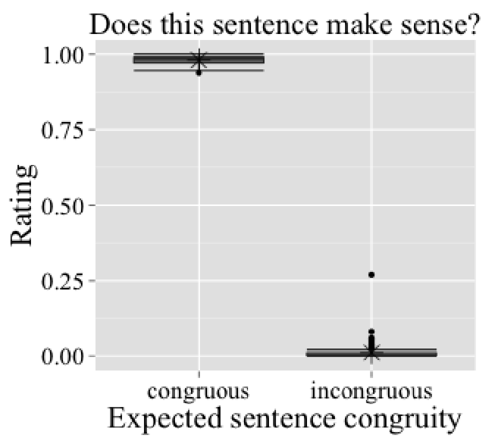


Figure 4.6: Results of Pilot Study 2

In general, participants’ judgments matched with the expected sentence congruity. All sentences that were intended to be congruous received ratings close to 1 (M = 0.98, SD = 0.01). Most sentences that were intended to be incongruous received ratings close to 0 (M = 0.01, SD = 0.02). The outlier with a mean rating of 0.27 is the incongruous sentence *nei5 bong1 ngo5 sik6 maai5 ngo5 go2 fan1* “You can help me eat my share (verb)”. Its congruous counterpart has the target word *fan6* “portion” instead, which reads “You can help me eat my portion”. A possible explanation for a relatively high rating for this incongruous sentence is the orthographic similarity of 分 *fan1* “share (verb)” and 份 *fan6* “portion”. It was assumed that orthographic confusion would not affect the main study, in which only audio stimuli were to be presented, and options were to be given in the form of pictures (see

Section 4.4.1.2 for details).

To conclude, Pilot Study 2 confirmed that the proposed carrier phrases were able to create the desired semantic effect, and were thus appropriate to be used as stimuli for the main study.

4.4 Main study

This section describes the materials, procedures, and participants of the main study.

4.4.1 Materials

Materials being discussed in this subsection are stimuli, pictures presented as options, tones and their corresponding buttons on the response device, and instructions for participants.

4.4.1.1 Stimuli

All stimuli for the experimental block were recorded by a 25-year-old female[4] native Cantonese speaker born and raised in Hong Kong, whose parents were also native Cantonese speakers born and raised in Hong Kong. After completing her undergraduate degree in Hong Kong, she moved to Vancouver when she was 22 years old. This means that she had lived in Canada for three years when doing the recording. She spoke Cantonese on a daily basis to communicate with her family and friends. No known hearing loss or speech disorder was reported. She was not linguistically trained and was naïve about the purpose of the study. She was compensated C$10 per hour for her time spent on the task.

The recording session took place in a sound-attenuated booth at the UBC Interdisciplinary Speech Research Lab. All stimuli were recorded at a sampling rate of 44,100 Hz with a USBPre 2 High-Resolution Audio

[4]The author had recruited both male and female talkers, but it happened that all male talkers either had extremely creaky voice especially when producing T4 [21] words, or they seemed to have merged the two rising tones (T2 [25] and T5 [23]) in production according to the author's judgment. The female talker was selected in the end because of her voice quality and her pitch contours, not because of her gender.

Interface manufactured by Sound Devices. The talker was presented with a list of 27 words and 147 sentences written in Traditional Chinese characters, and was asked to say each item three times at a natural speed. If necessary (e.g., if an item did not sound natural, if she sounded like she was laughing when producing the semantically incongruous sentences, or if she wanted to clear her throat), she could repeat an item for as many times as she wanted.

Special attention was paid to words with T4 [21] and semantically incongruous sentences. As pointed out by Yu & Lam (2014), creaky voice facilitates the identification of T4. To control the effect of phonation on accuracy rate, the talker was instructed to produce speech sounds with normal phonation as much as possible. Although the talker was not linguistically trained and did not know the meaning of normal phonation, the author demonstrated the production of several words with normal phonation and with creaky voice, and asked the talker to try her best to do the former. In some of her T4 words, creaky voice could still be found, but they were not creaky for the entirety of the syllable. As for sentences (especially the semantically incongruous ones), the talker was requested to re-record them if they were produced hesitantly, or if a pause was added before the last word. After listening to all recorded tokens, the author selected the best ones to be used as stimuli based on the naturalness of the production, and how well a pitch contour represented the respective lexical tone.

Note that the aforementioned female talker only recorded stimuli for the experimental blocks. All materials that were not part of the experimental block (task instructions, the story listening task, the picture learning task) were recorded by the author. For more details, see Section 4.4.2.

After the selection of recorded tokens to be used as stimuli, the sound files were edited for the purpose of the experiment. As mentioned in Section 4.1, two of the variables being controlled in this study were tonal and segmental information. In Table 4.1, target words in Types 3 and 6 had their segmental information filtered to test how well heritage speakers could identify a tone by solely relying on tonal information. In Types 2 and 4, tonal information was removed while segmental information was kept.

These stimuli were created by using functions in Praat (Boersma, 2002), which will be explained as follows.

For stimuli of Type 3 (monosyllabic words with no segments) and Types 6A and 6B (sentences in which the last word had no segments), a low-pass filter was applied to remove segmental information through the Praat function “Filter (pass Hann band)”. The cutoff frequency was 350 Hz, while the smoothing frequency was 20 Hz. This configuration was used for several reasons. First, the talker’s T1 [55] was approximately 250 Hz, so in order to retain all tonal information, the cutoff point must be higher than 250 Hz. Second, a cut-off of 350 Hz successfully filtered away the relevant acoustic signal for identifying vowels and consonants for this speaker. For example, compare an unmanipulated *fu2* “tiger” in Figure 4.7 with its low-pass-filtered version in Figure 4.8. In Figure 4.8, all energy above 350 Hz was removed, and so there was no clear F1 and F2, which means the vowel would be unidentifiable. The high frequency noise for identifying fricatives was also removed. Lastly, although other cut-off frequencies were tested, the output files sounded irritating to the human ear. The chosen configuration created stimuli that best resembles muffled speech, which is common in phone conversations when reception is poor. This helped minimize any listening discomfort that might affect participants’ accuracy rate in the experiment.

As for Type 2 (monosyllabic words with no tone) and Type 4 (sentences with no tone), tonal information was “removed” in a different sense. Since all vocalic elements must have f0 values, f0 cannot be removed without affecting segmental information. Rather, the f0 of these stimuli was reset to a uniform frequency at 200 Hz by the Praat functions “remove pitch point” and “add pitch point”. For instance, compare the unmanipulated *fu2* “tiger” in Figure 4.7 with its “no tone” version in Figure 4.9. The rising pitch contour in Figure 4.7 was completely flattened in Figure 4.9. As a result, certain perceptual correlates of tone, namely the direction and magnitude of pitch change, were made unavailable. The frequency at 200 Hz was chosen, because the talker’s T3 [33] was produced at a similar frequency, as in *fu3* “pants” in Figure 4.10. Although a naturally produced T3 [33] showed f0 declination towards the end of the syllable, its average pitch was still very similar to that of a Type 2 stimulus after pitch reset, as in Figure 4.9. This provided the ground for making predictions: when the only available perceptual cue was pitch height, participants who did pay attention to this cue should always respond that they heard T3 [33] when presented with Type 2 stimuli.

To ensure that no stimulus was louder than others (which might startle participants), the amplitude of all eight types of stimuli was normalized by the command `Scale... 0.99996948` in a Praat script adapted from Crosswhite (2009), which served to scale each file in a given directory in amplitude, so that all output files had the same peak amplitude. Syllable duration, however, was not normalized. Although duration is found to be an important perceptual correlate of Mandarin tones (Blicher, Diehl & Cohen, 1990), it is not the case for Cantonese. There has been no evidence that duration affects the perception of Cantonese tones (Fok-Chan, 1974; Tong, McBride & Burnham, 2014; Vance, 1976). To keep the stimuli as natural as possible, no action was taken to control syllable duration.

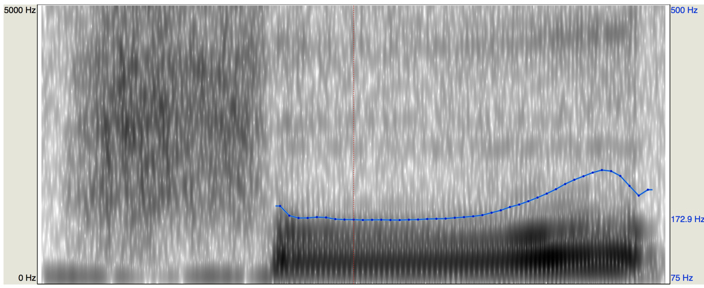


Figure 4.7: Spectrogram of the syllable *fu2* (high rising tone, unmanipulated)

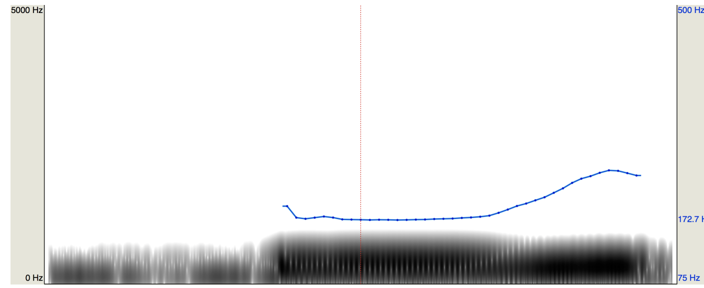


Figure 4.8: Spectrogram of the syllable *fu2* (high rising tone, low-pass filter applied)

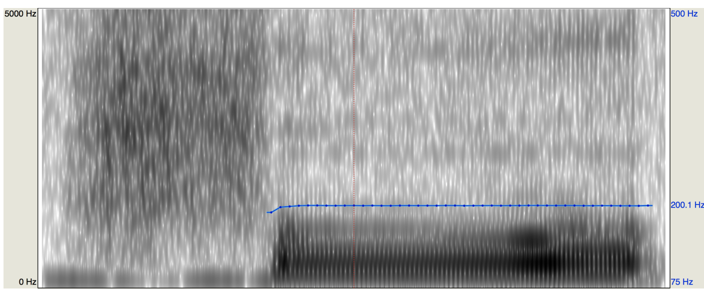


Figure 4.9: Spectrogram of the syllable *fu2* (high rising tone, pitch being reset at 200 Hz)

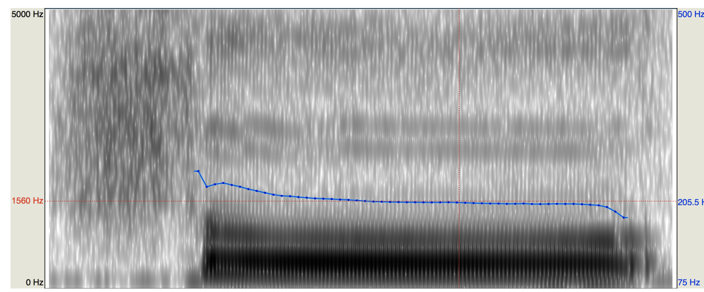
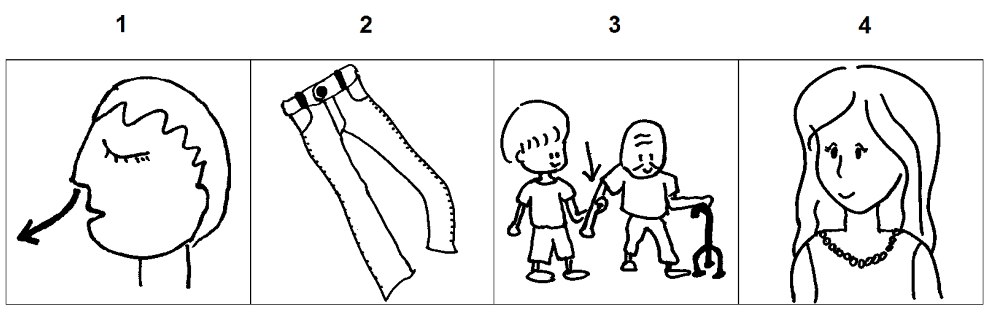


Figure 4.10: Spectrogram of the syllable *fu3* (mid level tone, unmanipulated)

4.4.1.2 Pictures

In their summary of results from the National Heritage Language Survey conducted in the United States, Carreira & Kagan (2011) report that more than 80% of the respondents who were heritage speakers of Cantonese or Mandarin rated their Chinese reading and writing skills “low” to “intermediate”. It was anticipated that heritage Cantonese speakers from Canada in the current study would have weak Chinese literacy skills. To ensure that all participants—especially those who had not received formal education in Chinese—could understand the options given in the forced-choice task, pictures instead of Chinese characters were presented as choices, as in Figure 4.11. All pictures were hand-drawn by the author only, so that no particular picture was visually more salient than another due to different artistic styles. Hand-drawn pictures were scanned as black and white bitmap files on a computer, so that no picture was more eye-catching than another in terms of colour. For the full collection of pictures, see Appendix A .



Note: Numbers refer to buttons on the button box (not tones).

Figure 4.11: A sample picture set: *fu1* “exhale”, *fu3* “pants”, *fu4* “help by holding another person’s arm”, and *fu5* “woman”

Special attention was paid to potentially confusing pictures. For example, *fu1* “exhale”, *fu4* “help by holding another person’s arm” , and *fu5* “woman” were in the same tone set [1 3 4 5]. In Figure 4.11, the exhaling person for the *fu1* picture and the helper/helpee in the *fu4* picture did not have stereotypical female features, so that they would not be confused with *fu5* “woman”. Another pair of examples was *fan1* “share” and *fan6* “portion”, which could be visually ambiguous, as the action of sharing would produce multiple portions of the object being shared. Earlier versions of these pictures were presented to a Cantonese instructor at UBC who did not participate in the experiment. Based on the instructor’s feedback, amendments were made to the pictures to make sure that the intended meaning would be conveyed.

In addition to the measures above, all pictures and their intended meaning were presented in a picture learning task that preceded the experimental block, so that participants were guided to the intended interpretation of a picture for the purpose of the experiment. For details of the picture learning task, see Section 4.4.2.2.

4.4.1.3 Tones and buttons

Measures were taken to counterbalance the correspondence between tones and response buttons on the response device for the experiment. Consider the numerically ordered tone sets [1 2 3 4], [1 3 4 5], and [1 4 5 6]. T4 corresponds to the fourth button in [1 2 3 4], the third in [1 3 4 5], and the second in [1 4 5 6]. However, T1 always corresponds to the first button across these three tone sets. If a participant responded by pressing Button 1 most of the time, it could be due to two possible reasons: s/he either had a T1 bias, or s/he simply preferred pushing the first button regardless of what was heard. To avoid this situation, the position of a tone within a set must vary.

Table 4.8 illustrates the solution to this potential problem through the example of [1 4 5 6]. The second column shows that tones within this set were ordered in four different ways: [1 4 5 6], [4 5 6 1], [5 6 1 4], and [6 1 4 5]. Tone-button correspondence was completely different in every row, so that every tone could correspond to any button. Although this was not an exhaustive list of all possible orders (e.g. [6 5 4 1] was a possible order but it was not represented in Table 4.8), it was sufficient to make each tone occur with each button exactly once. This way a participant who had a bias towards a particular tone could be distinguished from a participant who preferred a particular button regardless of what was heard: the former would press the four buttons more or less equally frequently, while the latter would press only Button 1 most of the time.

Table 4.8: An example of how tone-button correspondence was counterbalanced for a tone set

Tone set	Order	Button 1	Button 2	Button 3	Button 4
[1 4 5 6]	[1 4 5 6]	T1	T4	T5	T6
	[4 5 6 1]	T4	T5	T6	T1
	[5 6 1 4]	T5	T6	T1	T4
	[6 1 4 5]	T6	T1	T4	T5

4.4.1.4 Instructions

All instructions were recorded by the author in spoken Cantonese. This medium had an advantage over written Cantonese: it ensured that all participants, regardless of their Chinese reading proficiency, would be able to follow. Although written English could be understood by both groups as well, research has shown that silent reading activates the phonology of the language being written (McCutchen & Perfetti, 1982; Newman & Connolly, 2004; Perfetti, Bell & Delaney, 1988). This might lead to effects of interlanguage interference in online processing, which is out of the scope of this study. To avoid the activation of English phonology during the experiment, English—whether spoken or written—was avoided. The content of task instructions will be explained in the next subsection. For a full transcript of instructions in written Chinese and Cantonese romanization, see Appendix A.

4.4.2 Procedures

Each participant was invited to attend two experiment sessions separated by at least 24 hours, both of which took place at the UBC Speech in Context Lab. All tasks were programmed on E-Prime Version 2.0

(Schneider, Eschman & Zuccolotto, 2002). A step-by-step walkthrough of the procedures in Table 4.1 is provided as follows. Readers are recommended to bookmark Table 4.1 for easy reference.

On Day 1, all participants were presented with a printed copy of a consent form stating the purpose, risks and benefits of the experiment, as well as contact information of the experimenter. They were asked to sign on the form if they agreed to participate. However, in all collected data, participants were only identified by a subject number to ensure anonymity. After signing the consent form, participants were brought to a sound-attenuated computer booth equipped with a response button box and an AKG K240 Studio headphone. After they were seated at the booth, they were instructed to put on the headphone and press any button on the box to start.

4.4.2.1 Story listening task

The experiment session began with a story listening task. Its purpose was to tune participants into a unilingual Cantonese-speaking environment. Since all participants were recruited from the university student community in Vancouver, both homeland and heritage speakers were attending classes conducted in English except for foreign language courses. By listening to a Cantonese story before the experimental block, both groups had a chance to warm up and get ready to process Cantonese speech. This task also helped minimize potential effects of code-switching on bilingual language processing (Antoniou et al., 2010, 2011).

A 40-second story called *The Sun and the Wind* was played to the participants at the beginning of Day 1. The story did not contain any words that were used as target words in the experimental block. (For a full script of the story, see Appendix A.) The story was recorded by the author, a different voice from that of the experimental block. To make sure that participants would listen carefully, they were told that they would have to answer a question after listening to the story. A picture of a headphone was shown in the centre of the screen while the story was being played, as in Figure 4.12. Participants had a chance to let the experimenter know if the headphones were not working, or if the volume was not comfortable. When the story finished, a multiple-choice question was posed, asking "who won in the end?" in spoken Cantonese. Three pictures (Sun, Wind, Man) were displayed on the screen, each corresponding to a button on the button box. Participants were instructed to respond by pushing a button on the device. Their response to this question, however, was not used to assess their comprehension of spoken Cantonese. Since this was the very first question requiring the use of the button box, participants were still getting used to the sensitivity of the buttons. Therefore, responses to this question played no role in determining the inclusion or exclusion of a participant's data.

Figure 4.12: Picture shown during the story listening task

4.4.2.2 Picture learning task

Performed after the story listening task and before the practice block, the picture learning task served to clarify the intended meaning of the pictures, each of which could be compatible with multiple words. For example, when participants saw the picture of *se6* "shoot", it was possible for them to interpret it as *coeng1* "gun". Another concern was that some words were more commonly used in their disyllabic form in spoken Cantonese. Generally, the word *fu2* in isolation can mean either "bitter" or "tiger". In spoken contexts, "bitter" is always monosyllabic, but "tiger" is more commonly referred to as *lou5 fu5* (in which *lou5* literally means "old", but *lou5 fu2* is a set phrase which makes no reference to a tiger's age). The picture learning procedure helped participants associate the monosyllabic word *fu2* with "tiger" instead of its homophone "bitter".

The procedures of the picture learning tasks were as follows. A picture was shown in the centre of the screen, as in Figure 4.13. At the same time, participants would hear the sentence "X, this picture refers to X, as in XY", where X was the target word, and Y was another word that collocates with X often in Cantonese. Y could precede or follow X, depending on the word. Participants were only allowed to listen to each sentence once. They could press any button on the box to continue to the next picture. The total number of pictures presented was 31, since there were 27 target words for the main task (as in Table 4.4) and 4 for the practice block (see the next subsection).

Audio: ***fu2**, li1 fuk1 tou4 hai6* ***lou5 fu2*** *go3* ***fu2***
"Tiger$_{\text{monosyllabic}}$, this picture is tiger$_{\text{monosyllabic}}$, as in tiger$_{\text{disyllabic}}$"

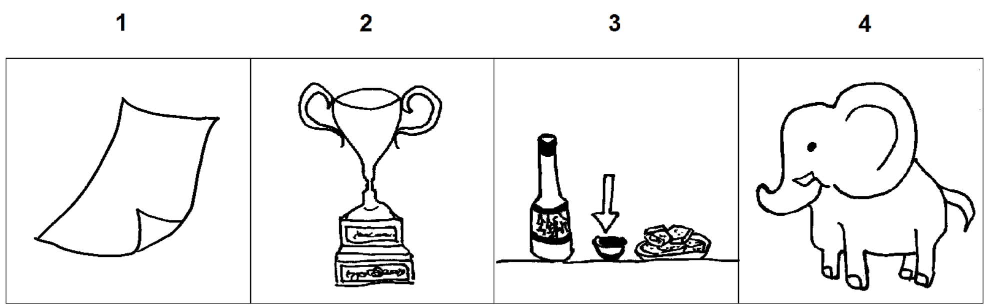
Figure 4.13: An example of the picture learning task

4.4.2.3 Practice trials for monosyllabic words

The purpose of the practice block was to familiarize participants with the format of the task and the use of the button box. In this block, all words and pictures were different from those of the experimental block. A word set with the syllable *zoeng* was used: *zoeng1* "piece (of paper)", *zoeng2* "prize", *zoeng3* "sauce", and *zoeng6* "elephant".

English translation of the spoken instructions is provided in (15). When the spoken instructions were being played, the same headphone picture in Figure 4.12 was shown on the screen. After listening to the instructions, the participant could press any button on the box to start.

The following procedures applied to not only practice trials, but also trials in all experimental blocks. In each trial, participants would hear one stimulus on the headphone, and at the same time a picture set like Figure 4.14 was shown on the screen. They were asked to press a button on the box that corresponded to the picture that represented the word being heard. No feedback was provided, which means participants were not told whether their answer was correct. After the participant responded by pressing a button, the screen would be blank for 500 ms. After that another stimulus would be played and at the same time another picture set would be shown.



Note: Numbers refer to buttons on the button box (not tones).

Figure 4.14: Pictures used in practice trials: *zoeng1* “piece (of paper)”, *zoeng2* “prize”, *zoeng3* “sauce”, and *zoeng6* “elephant”

(15) English translation of instructions for Day 1 practice trials, originally in spoken Cantonese:
You are going to listen to some words. Which picture represents the word that you heard? You can only choose one picture and respond by using the button box. Note that sometimes the word may be unclear, and it is intentional. You just need to try your best to answer. When you are ready, press any button on the box to start.

There were eight trials in the practice block, which formed a representative sample of the upcoming experimental block with normal words, words with no tone, and words with no segment, as in Table 4.9. The sample of different types of stimuli helped participants understand that some stimuli were intentionally unclear or unnatural, so that they would not be surprised when they heard these stimulus types in the experimental block. Note that participants were not told that these eight trials were for training purpose, and so they were expected to respond as they normally would in the experimental block. Since the practice trials only served to familiarize participants with the task, their performance in this block was not analyzed.

Table 4.9: Examples of practice trials on Day 1

Type	Target	Competitors		
Normal	*zoeng2* prize	*zoeng1* piece	*zoeng3* sauce	*zoeng6* elephant
No tone	*zoeng_* piece	*zoeng_* prize	*zoeng_* sauce	*zoeng_* elephant
No segment	*____3* sauce	*____1* piece	*____2* prize	*____6* elephant

4.4.2.4 The first experimental block: Types 1–3 randomized

The first experimental block followed the practice block immediately. It consisted of randomized stimuli of Type 1 (normal words), Type 2 (words with no tone), and Type 3 (words with no segments). For each stimulus type, there were 60 trials (15 tone sets * 4 words per set). Therefore, there were a total of 180 trials in the first experimental block (60 trials * 3 stimulus types).

4.4.2.5 The second experimental block: Type 4

Instructions for the second experimental block in (16) were played immediately after the first experimental block. In this block, there were only stimuli of Type 4 (sentences with no tone). Sentences representing different tone sets were randomized. The total number of trials for this block was 60 (15 tone sets * 4 words per set).

(16) English translation of instructions for the second experimental block, originally in spoken Cantonese:
You are going to listen to some sentences. What is the last word of the sentence? Note that the sentences may sound unnatural, which is intentional. Just try your best to answer. If you have questions, please let the experimenter know. When you are ready, press any button on the box to start.

Although Type 4 stimuli were sentences, they were not presented on Day 2 with other stimulus types that were sentences (Types 5A, 5B, 6A, and 6B). The first experimental block on Day 1 (Types 1–3) contained 180 trials, but on Day 2, the third block (Types 5A, 5B, 6A, and 6B) contained 480 trials. If Type 4 was presented on Day 2, the experiment session might take more than 60 minutes. To avoid listening fatigue on Day 2, which might potentially add noise to the results, Type 4 was presented on Day 1 instead.

4.4.2.6 Language background questionnaire

After the completion of the second experimental block, participants were instructed to complete a language background questionnaire hosted on UBC FluidSurveys (FluidSurveys, 2017). It contained questions about their age, education level, language history, language use, language proficiency, language attitudes, and the native languages of their parents. For a detailed explanation of this questionnaire and a full list of questions, see Section 4.4.3.2 and Appendix B respectively.

Participants were asked to complete the questionnaire on Day 1 instead of Day 2 for two reasons. First, since participation was voluntary, some participants might not return to complete the session on Day 2. In this case, their language background information would still be available if they completed the questionnaire on Day 1. Second, Day 1 tasks took approximately 40 minutes to complete, while Day 2 tasks took approximately 60 minutes to complete. To avoid making the session too long on Day 2, the questionnaire was arranged to be completed on Day 1.

After filling out the questionnaire, participants could sign up for a time slot for the second experiment session. Day 1 and Day 2 were separated by a minimum of 24 hours. Due to scheduling factors (e.g. within a school term students always had free time on the same day of the week), most participants attended the second experiment session one week after their first.

4.4.2.7 Picture learning task (repeated)

The session on Day 2 began with the same picture learning task described in Section 4.4.2.2. It served as a reminder of the intended meaning of the pictures.

4.4.2.8 Practice trials for sentences

Like Day 1, a practice block with eight trials preceded the experimental block on Day 2. The target words were also *zoeng1* “piece (of paper)”, *zoeng2* “prize”, *zoeng3* “sauce”, or *zoeng6* “elephant”, and the same picture set in Figure 4.14 was used. However, unlike Day 1 practice trials which were monosyllabic words, Day 2 practice trials were normal sentences, congruous and incongruous sentences, as well as sentences with a segmentless target word at the end, as in Table 4.10. They formed a representative sample of the upcoming experimental block. This helped participants understand that some stimuli were intended to be semantically anomalous, and some words were intentionally unclear.

English translation of the spoken instructions is provided in (17). It is important to note that participants were told explicitly that some sentences might not make sense, and their task was to identify the last word that they heard. If they were unsure about the instructions, they had the opportunity to clarify with the experimenter during the practice trials, so that their performance in the experimental block would not be affected. However, they were not told that these eight trials were for training purpose, and they were expected to respond as they normally would.

(17) English translation of instructions for Day 2 practice trials, originally in spoken Cantonese:
You are going to listen to some sentences. What is the last word of the sentence? Some words may be unclear, and you just need to try your best. Note that some sentences may not make sense, which is intentional. All you need to do is identify the last word that you heard. For example, if you hear "there are too many student names to remember", then you should respond with "remember" (*gei3*). However, if you hear "there are too many student names to airplane", then you should respond with "airplane" (*gei1*) but not "remember" (*gei3*). If you have questions, please let the experimenter know. When you are ready, press any button on the box to start.

4.4.2.9 The third experimental block: Types 5A–6B randomized

Immediately following the practice block, the third experimental block consisted of randomized stimuli of Types 5A (normal, congruous sentences), 5B (normal, incongruous sentences), 6A (congruous sentences in which the last word had no segments), and 6B (incongruous sentences in which the last word had no segments). Since this was a long block with a total of 480 trials, there was a break every 120 trials. Participants could press any button on the box to resume whenever they were ready to continue.

The number of trials per stimulus type in this block is explained as follows through examples from the tone set [2 3 4 6]. All stimuli in Table 4.11a, Table 4.11b, Table 4.11c, and Table 4.11d shared the same carrier phrase *sap6 ji6 dim2 zung1 hou2 soeng5 cong4...* "at twelve (you) should go to bed and..." and all of their targets and competitors represented

Table 4.10: Examples of practice trials on Day 2

Type	Carrier phrase	Target	Competitors		
Normal (congruous)	*dung6 mat6 jyun4 zeoi3 daai6 ge3 hai6...*	*zoeng6*	*zoeng1*	*zoeng2*	*zoeng3*
	The biggest (animal) in the zoo is the ...	elephant	piece	prize	sauce
Normal (incongruous)	*dung6 mat6 jyun4 zeoi3 daai6 ge3 hai6...*	*zoeng3*	*zoeng1*	*zoeng2*	*zoeng6*
	The biggest (animal) in the zoo is the ...	sauce	piece	prize	elephant
No segment (congruous)	*dung6 mat6 jyun4 zeoi3 daai6 ge3 hai6...*	___*6*	___*1*	___*2*	___*3*
	The biggest (animal) in the zoo is the ...	elephant	piece	prize	sauce
No segment (incongruous)	*dung6 mat6 jyun4 zeoi3 daai6 ge3 hai6...*	___*3*	___*1*	___*2*	___*6*
	The biggest (animal) in the zoo is the ...	sauce	piece	prize	elephant

the tone set [2 3 4 6]. However, they differed by whether the target word was semantically congruous with the carrier phrase, and whether the last word contained segmental information.

The total number of trials for Type 5A was 60. In Table 4.11a, the word *fan3* "sleep" was semantically congruous with this particular carrier phrase. However, the word *fan2* was congruous with another carrier phrase, namely *bong1 bi4 bi1 caa4 di1 song2 san1...* "put some baby ... on the baby". In other words, for each tone set, each member of the quadruplet was semantically congruous with one unique carrier phrase. Since there were 15 tone sets in total (see Table 4.4), the total number of trials for Type 5A was 60 (1 member of a quadruplet * 4 unique carrier phrases * 15 tone sets).

The same applied to Type 6A stimuli, which were also congruous sentences and therefore had a total of 60 trials. However, for Type 6A, the last word of the sentence had no segments, as in Table 4.11c.

The total number of trials for Type 5B was 180. In Table 4.11b, there were three rows, unlike Table 4.11a which had one only. This was because only one member in a tonally contrastive quadruplet was meant to be semantically congruous with a given carrier phrase, while the other three members were not. Since Type 5B was incongruous sentences, its total number of trials was 180 (3 members of a quadruplet * 4 unique carrier phrases * 15 tone sets).

The same applied to Type 6B stimuli, which were also incongruous sentences and therefore had a total of 180 trials. However, for Type 6B, the last word of the sentence had no segments, as in Table 4.11d.

After the last trial of the third experimental block, the English phrase "The end" was shown on the screen. Participants were given either cash or course credit for their participation (see Section 4.4.3.1 on subject recruitment). This was the end of the second (also the last) experiment session.

Table 4.11: A sample of the third experimental block representing the tone set [2 3 4 6]

a. Type 5A: Normal (congruous)

Carrier phrase	Target	Competitors		
sap6 ji6 dim2 zung1 hou2 soeng5 cong4...	*fan3*	*fan2*	*fan4*	*fan6*
At twelve (you) should go to bed and ...	sleep	powder	tomb	portion

b. Type 5B: Normal (incongruous)

Carrier phrase	Target	Competitors		
sap6 ji6 dim2 zung1 hou2 soeng5 cong4...	*fan2*	*fan3*	*fan4*	*fan6*
At twelve (you) should go to bed and ...	powder	sleep	tomb	portion
sap6 ji6 dim2 zung1 hou2 soeng5 cong4...	*fan4*	*fan2*	*fan3*	*fan6*
At twelve (you) should go to bed and ...	tomb	powder	sleep	portion
sap6 ji6 dim2 zung1 hou2 soeng5 cong4...	*fan6*	*fan2*	*fan3*	*fan4*
At twelve (you) should go to bed and ...	portion	powder	sleep	tomb

c. Type 6A: The last word has no segments (congruous)

Carrier phrase	Target	Competitors		
sap6 ji6 dim2 zung1 hou2 soeng5 cong4...	___*3*	___*2*	___*4*	___*6*
At twelve (you) should go to bed and ...	sleep	powder	tomb	portion

d. Type 6B: The last word has no segments (incongruous)

Carrier phrase	Target	Competitors		
sap6 ji6 dim2 zung1 hou2 soeng5 cong4...	___*2*	___*3*	___*4*	___*6*
At twelve (you) should go to bed and ...	powder	sleep	tomb	portion
sap6 ji6 dim2 zung1 hou2 soeng5 cong4...	___*4*	___*2*	___*3*	___*6*
At twelve (you) should go to bed and ...	tomb	powder	sleep	portion
sap6 ji6 dim2 zung1 hou2 soeng5 cong4...	___*6*	___*2*	___*3*	___*4*
At twelve (you) should go to bed and ...	portion	powder	sleep	tomb

4.4.3 Participants

This subsection describes how participants for the main study were recruited and screened, and reports on the demographic information of included participants.

4.4.3.1 Recruitment

All participants were recruited from the university community in Vancouver. Electronic recruitment flyers were circulated on social media like Facebook. They were also sent via email to three student clubs at UBC, namely the UBC Mahjong Club, the UBC Hong Kong Student Association, and Hong Kong at Heart. The flyers were written in both Chinese and English, so that both homeland and heritage speakers could read them. The flyer stated that all individuals who had at least one Cantonese-speaking parent from Hong Kong were welcome to sign up; a high fluency in Cantonese or the ability to read Chinese was not required. Interested individuals could contact the author via email to sign up for a time slot. Participants recruited through social media and student clubs were compensated C$10 for Day 1 and C$15 for Day 2. A higher payment for Day 2 was an incentive for participants to return and complete the second part of the experiment.

Another way to recruit participants was *Linguistics Outside the Classroom* (LOC), an initiative to encourage undergraduate students to get involved in research conducted by members of the Department of Linguistics (UBC Department of Linguistics, 2014). Course instructors had the option to include LOC credits in their course syllabi. Students could satisfy LOC requirements either by writing a summary of a research seminar or colloquium that they attended, or by participating in experiments. Students could view available time slots and sign up on the UBC Linguistics Sign-up System (Sona Systems Ltd., 2017). Each user was identified by a unique participant code instead of his or her name. As a general rule of LOC, no one should be refused to participate in any experiment. As a result, Cantonese speakers from China, Macau, Malaysia, and even non-Cantonese speakers were allowed to sign up, even though their data were not usable. All participants recruited through LOC were awarded course credit, regardless of the usability of their data.

4.4.3.2 Screening

A total of 100 individuals participated in the study. However, only 68 participants' data were included for data analysis in Chapter 5. The data of 32 participants were excluded for various reasons. Decisions of inclusion were made based on responses to a language background questionnaire, which was also used to categorize participants into homeland and heritage speakers. The screening process is summarized in Figure 4.15, and the purpose of specific questions on the questionnaire is explained as follows.

As shown in Figure 4.15, the first screening criterion was whether a participant had completed all tasks in the experiment. Three participants only attended the experiment session on Day 1 but did not return on Day 2. To balance the number of responses across all types of stimuli, their data were excluded.

The rest of the screening criteria in Figure 4.15 were based on participants' responses on a language background questionnaire. At the end of the experiment session on Day 1, all participants were asked to complete an online questionnaire. It comprised 30 questions about demographic information and language background, which took 10 to 15 minutes to complete. All responses were anonymous and no personally identifiable information was elicited. Out of the 30 questions, 19 were from the Bilingual Language Profile (Birdsong et al., 2012), and 11 of them were designed specifically for the current study. Both types of questions will be explained below. For the complete list of questions, see Appendix B.

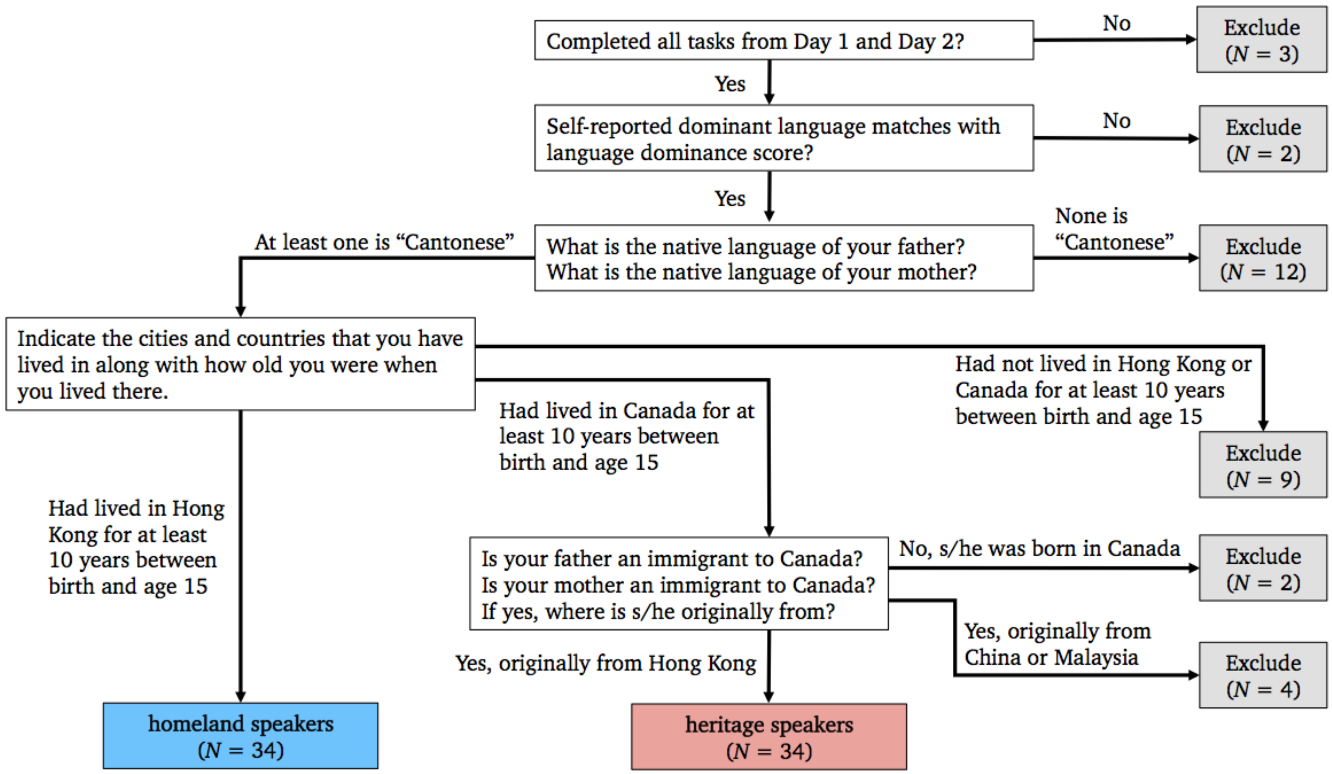


Figure 4.15: Procedures to screen and categorize participants

The Bilingual Language Profile (BLP) developed by Birdsong et al. (2012) was used to elicit language background information. It started with general demographic information, such as age and education level. The main part of the questionnaire was designed to assess language dominance on a gradient scale using four criteria: language history, language use, language proficiency, and language attitudes. Participants were asked the same set of questions for each of their two languages. For example, if there is question like "How well do you speak English?", then there is also a question like "How well do you speak Cantonese?" Responses to the questions add up to a global language score for each language. Subtracting the global language score for Cantonese from that of English yields the language dominance score, which ranges from –218 (extremely Cantonese-dominant) to 218 (extremely English-dominant)[5]. A dominance score at zero indicates balanced bilingualism. Table 4.12 illustrates the calculation using the data of Subject #345 as an example. Her global language score for Cantonese was 113.78 out of 218, while her global language score for English was 192.96 out of 218. Thus, her language dominance score was 192.96 minus 113.78, which is 79.18. Since 79.18 is above zero, Subject #345 was English-dominant.

Table 4.12: Calculation of Subject #345's language dominance score

Module score (out of 54.5)		Cantonese	English
I.	Language history	30.42	45.85
II.	Language use	15.26	38.15
III.	Language proficiency	38.59	54.48
IV.	Language attitudes	29.51	54.48
Global language score (out of 218)		113.78	192.96

Language dominance score = 192.96 –113.78 = $\underline{\underline{79.18}}$

[5]These positive and negative numbers on the scale are arbitrary and by no means imply value judgment of whether it is better or worse to be dominant in a particular language.

The BLP was supplemented with an additional set of questions for the purpose of this study. First, "What is your dominant language?" was added to understand participants' own judgment of their language dominance. All except two participants reported a dominant language that matched their BLP language dominance score. The two special cases were Subjects #331 and #339, who had dominance scores of 26.52 and 7.00 respectively. Although their scores were above zero, which suggests dominance in English, their self-reported dominant language was Cantonese. According to the discussion on measures of language dominance by Gollan, Weissberger, Runnqvist, Montoya & Cera (2012), if there is discrepancy between the result of a measurement method and self-assessment by an individual, s/he may not be "wrong"; rather, it indicates that s/he may have focused on factors not included in the measurement method. Since every individual takes different factors into account when assessing their own language dominance, no single measure can be a perfectly "complete" assessment. Due to this discrepancy, the dominant language of Subjects #331 and #339 was considered undetermined, and so their data were excluded. In addition, they were the only subjects who reported that they attended an international school when they were living in Hong Kong. According to Lai, Li & Gong (2016), international schools in Hong Kong adopt the curricula of the United States, Australia, or Canada. In the majority of these schools, Chinese language classes are offered as a subject, but they are taught in Mandarin rather than Cantonese. As for the demographics of students, international schools mainly cater for children of expatriates, and on average 76% of students are non-local. In other words, English is not only the main language of instruction, but also the language for social interaction at these schools. This distinguished Subjects #331 and #339 from individuals who went to a local school in Hong Kong, where Cantonese is the majority language in various domains of language use.

Another question added for the purpose of this study was "Indicate the cities and countries that you have lived in along with how old you were when you lived there". Although the BLP already has the question "How many years have you spent in a country/region where Cantonese is spoken?", it does not specify a particular Cantonese-speaking region. An additional open-ended question was therefore added to separate Hong Kong Cantonese speakers from speakers of other varieties of Cantonese. Nine participants reported that they grew up in the Guangdong province of China or Malaysia. Their data were excluded from analysis, because tonal features of these varieties may affect their perception of Hong Kong Cantonese. For example, Guangzhou Cantonese has undergone a merger of the non-high level tones (T3 [33] and T6 [22]) (Ou, 2012), which differs from the merger of rising tones (T2 [25] and T5 [23]) in Hong Kong Cantonese. As for Malaysian Cantonese, there are only five phonemic tones in the system (Hsiar, 2007), while Hong Kong Cantonese has six.

A few additional questions were about the background of participants' parents. "What is the native language of your father/mother?" served to exclude individuals who did not have at least one Cantonese-speaking parent. Twelve participants were excluded since neither of their parents spoke Cantonese; in fact, these 12 participants did not speak or barely spoke Cantonese according to their self-ratings for Cantonese proficiency. Out of 66 individuals whose data were included for analysis, 65 of them had two Cantonese-speaking parents. The participant who had only one Cantonese-speaking parent was Subject #314, who has a Cantonese-speaking mother and a Spanish-speaking father from Peru. Since she met the criterion of having at least one Cantonese-speaking parent, her data were included for analysis. No unusual patterns were observed in her experiment responses.

Two more questions were added to exclude heritage speakers whose parents were also heritage speakers. As mentioned in Chapter 2, second- and third-generation heritage speakers of Ukrainian in Toronto produced Ukrainian obstruent consonants with different VOTs. In other words, the linguistic input that the third generation received may be different from the input that the second generation had received from the homeland generation. For the purpose of the current study, it is important that both homeland and heritage speakers were exposed to the same baseline variety of the language for fair comparison. This way any observed difference between the two groups in the experiment, if any, would be a reflection of their different perception of lexical tones, but not a reflection of linguistic variation in the input from parents of different generations. For this reason, the following questions were included in the survey: “Is your father/mother an immigrant to Canada? Where is s/he originally from?” Two participants reported that their parents were born and raised in Canada, and so they were excluded from data analysis. Lastly, four Canadian-born participants reported that their parents were originally from Guangzhou. For the reason stated previously regarding tonal features of regional varieties of Cantonese, the data of these four participants were excluded.

Participants who had not been excluded due to aforementioned reasons would be categorized as either homeland or heritage speakers based on their response to the question “Indicate the cities and countries that you have lived in along with how old you were when you lived there”. Those who had lived in Hong Kong for at least 10 years between birth and age 15 were categorized as homeland speakers, while those who had lived in Canada for at least ten years between birth and age 15 were categorized as heritage speakers.

To conclude, only participants who met *all* the following criteria were included: (i) they completed all required tasks; (ii) their self-reported dominant language matched their language dominance score; (iii) they had at least one parent from Hong Kong who spoke Cantonese as a native language, and (iv) they had lived in Hong Kong or Canada for at least 10 years between birth to age 15. Based on these criteria, 32 participants were excluded. Among the 68 participants who were included, 34 of them had spent at least ten years in Hong Kong from birth to age 15, and were therefore considered homeland speakers; 34 of them had spent at least ten years in Canada from birth to age 15, and were therefore considered heritage speakers[6]. The demographic information of these 68 individuals will be presented in the next subsection.

[6]The balanced numbers were a result of continuous data collection until there were 34 participants in each group.

4.4.3.3 Demographics of included participants

The age, language dominance scores, and language proficiency of included participants are reported as follows.

As mentioned in Chapter 3, Cantonese tone merger in Hong Kong is more common among speakers in their 20s than those in their 50s (Fung et al., 2011), so it is important that the two groups in the current study fell within the same age range. Since the 68 included participants were recruited from the university student community, they were all young adults (M = 20.78 years of age, SD = 3.10). As summarized in Table 4.13, the average age of the homeland speaker group was 20.85 years (SD = 2.97), and the average age of the heritage speaker group was 20.71 years (SD = 3.27).

Table 4.13: Age of included participants (in years)

	Minimum	Maximum	*M*	*SD*
homeland	18	33	20.85	2.97
heritage	18	31	20.71	3.27

So far no known studies have suggested any relationship between gender and Cantonese tonal perception, and so no attempt was made to balance the number of male and female participants. In the current study there were more female than male participants, and this was true for both populations. In the homeland group there were nine male and 25 female participants; as for the heritage group, there were 12 male and 22 female participants.

As mentioned Section 4.4.3.2, the language background questionnaire not only elicited demographic details like age and gender, but also collected the required information to generate language dominance scores, following the method introduced by Gertken et al. (2014). In Figure 4.16, the two ends of the x-axis are marked by the minimal and maximal endpoints of the BPL scale, which are –218 (extremely Cantonese-dominant) and 218 (extremely English-dominant) respectively. The dotted line in the middle marks zero, which represents balanced bilingualism.

In the current study, all homeland speakers had negative scores (hence Cantonese-dominant), and all heritage speakers had positive scores (hence English-dominant). Indicated by an asterisk on the boxplot, the mean score of the homeland group was –63.66 (*SD* = 30.09), while the mean score of the heritage group was 89.80 (*SD* = 36.12). The homeland group was closer to zero, which can be attributed to the fact that all homeland speakers were residing in Canada when they participated in the study. They had a relatively high level of English proficiency in order to study in a post-secondary institution. English was also used for day-to-day interactions on and off campus.

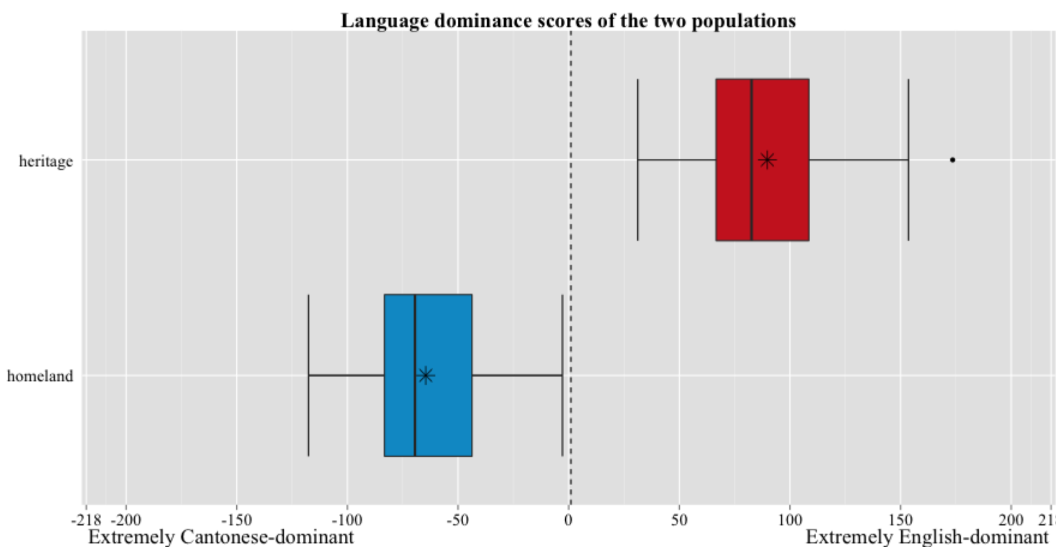


Figure 4.16: Language dominance scores of the two populations

Recall that Chapter 2 has discussed nine configurations on the bilingual continuum. Although the nine configurations were introduced for the purpose of a conceptual discussion and are by no means translatable to dominance scores linearly, the two end points and the mid point of Figure 2.1 and Figure 4.16 are comparable to each other. To obtain a dominance score of –218, the global language score for L1 must be 218 (the highest possible score) and the global language score for L2 must be 0 (the lowest possible score). The reverse is true for a dominance score of

218: the global language score for L1 must be 0 (the lowest possible score) and the global language score for L2 must be 218 (the highest possible score). Therefore, Configuration A and a dominance score of –218 can each indicate L1 monolingualism. Configuration I and a dominance score of 218 can each indicate L2 monolingualism. Configuration E and a dominance score of 0 can each indicate balanced bilingualism. By comparing the bilingual continuum in Figure 2.1 and the language dominance scale in Figure 4.16, it can be concluded that both groups' dominance scores landed on their expected range respectively. In Section 2.2 it was mentioned that homeland and heritage speakers were expected to have L1-dominant configurations (between A and E) and L2-dominant configurations (between E and I) respectively. Results in Figure 4.16 match with these expectations: homeland speakers' scores were between –218 and 0, and heritage speakers' scores were between 0 and 218.

As discussed in Section 2.1.3, dominance and proficiency are related but different concepts (Schmeißer et al., 2016), so data about participants' language proficiency are reported separately as follows. All participants were instructed to rate their listening, speaking, reading and writing skills of both languages on a scale of 0–6: 0 means "not well at all" and 6 means "very well". Table 4.14a and Table 4.14b summarize their mean self-ratings for Cantonese and English respectively, and present results of *t* tests for for two independent samples with Bonferroni correction, so as to compare the two groups' ratings for each skill and each language. Figure 4.17 visualizes the same data in the form of boxplots, in which mean values are marked with an asterisk, and median values are marked with centre lines in the box. Observations from these data and their implications for the main study are discussed as follows.

Table 4.14: *t*-test comparison of homeland and heritage speakers' self-rated language proficiency on a scale of 0–6: 0="not well at all" and 6="very well"

a. Cantonese language skills

	Homeland		Heritage		*t*	*df*	*p*
	M	*SD*	*M*	*SD*			
Listening	5.65	0.64	4.47	1.13	5.26	52.36	$<$.001
Speaking	5.79	0.48	3.53	1.50	8.38	39.63	$<$.001
Reading	5.56	0.70	1.82	1.47	12.68	45.83	$<$.001
Writing	5.32	0.94	1.41	1.35	13.84	59.04	$<$.001

b. English language skills

	Homeland		Heritage		*t*	*df*	*p*
	M	*SD*	*M*	*SD*			
Listening	4.79	0.73	5.94	0.24	–8.71	40.00	$<$.001
Speaking	4.56	0.93	5.94	0.24	–8.42	37.36	$<$.001
Reading	4.64	0.77	5.91	0.29	–8.93	41.96	$<$.001
Writing	4.23	1.10	5.82	0.46	–7.76	44.09	$<$.001

Note: Since there were eight pairwise comparisons, a Bonferroni correction was made. The alpha level after correction was 0.00625.

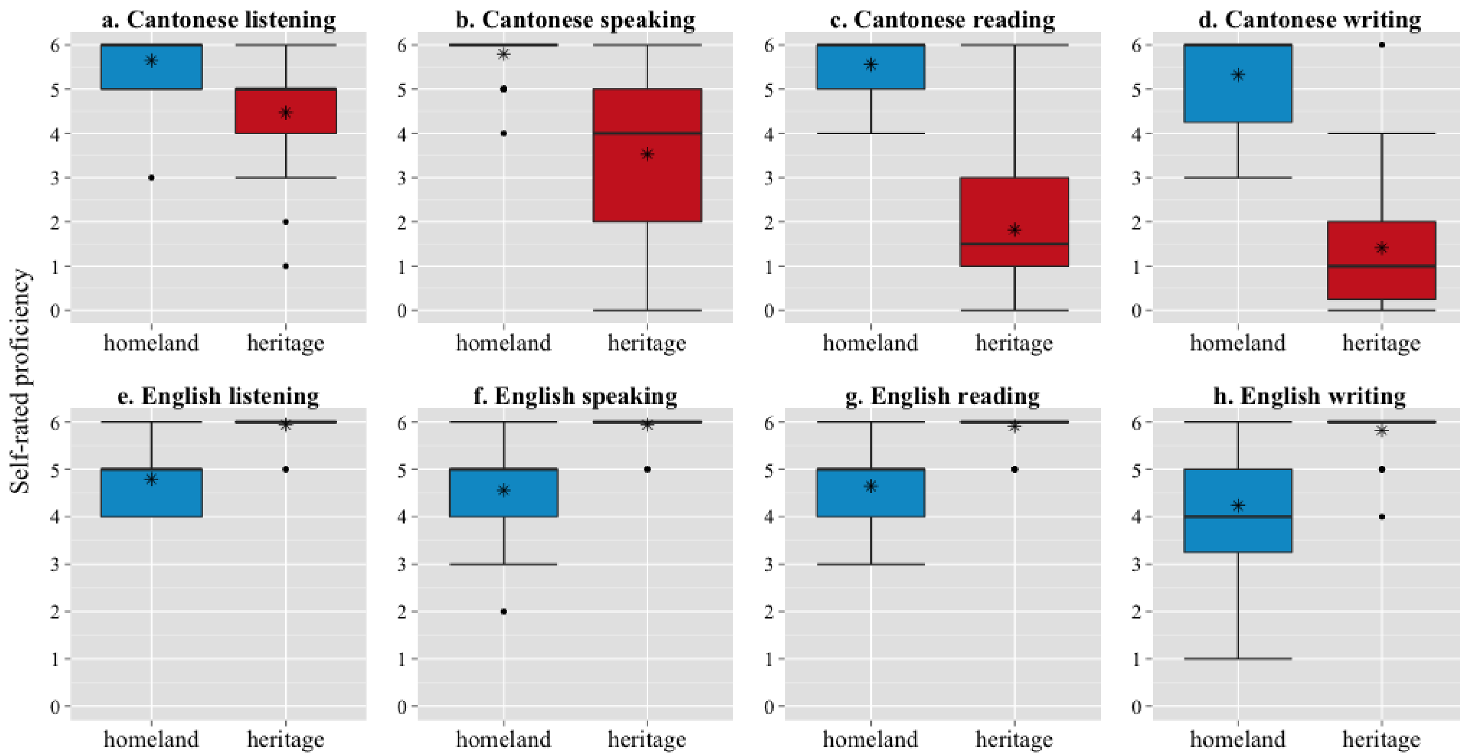


Figure 4.17: Self-rated language proficiency of homeland and heritage speakers on a scale of 0–6; 6="very well" and 0="not well at all"

As an overview, no heritage speakers rated 0 for all four language skills for Cantonese. In Figure 4.17a ("Cantonese listening"), the lowest rating given was 1. In other words, all heritage speakers (and homeland speakers too) in the current study met the definition of a bilingual given in Chapter 2: they were minimally competent in at least one of the four language skills for each of the two languages, even though they did not have an equal mastery of listening, speaking, reading, and writing skills for both languages.

Several observations can be made regarding heritage speakers' self-rated Cantonese proficiency. First, their self-rated listening ability was the highest (M = 4.47, SD = 1.13) among the four language skills. Second, their self-rated speaking ability was generally fair (M = 3.53), but its standard deviation was the highest (1.50) among the four skills, indicating a high degree of variability. Such variability is also demonstrated by the height of the red box in Figure 4.17b, which has whiskers spanning from 0 to 6. Third, the heritage group's literacy skills were much lower compared with their listening and speaking skills. Their average self-ratings for reading (M = 1.82, SD = 1.47) and writing (M = 1.41, SD = 1.35) were both below 2, and hence these two red boxes are pulled farther away from the blue ones that represent the homeland group. Therefore, the four red boxes in Figure 4.17 appear to be descending from left to right.

These observations about heritage speakers' Cantonese language skills are consistent with those from previous studies. In both Canada (the current study) and the United States (Carreira & Kagan, 2011), heritage Cantonese speakers gave higher ratings for their aural skills (listening and speaking) than their literacy skills (reading and writing). Their low ratings for reading also met earlier expectations, which led to the decision to use pictures rather than written Chinese characters as options in the forced-choice experiment.

The high variability in terms of Cantonese proficiency (especially for listening skills) within the heritage population suggests that a high variance of accuracy could also be expected in results of the word-identification study. It was a possibility to have two clearly split groups within the heritage population, namely highly proficient speakers with extremely high accuracy, and not so proficient speakers with extremely low accuracy. In this case, the mean alone could lead to misleading interpretations. For this reason, other measures were taken to make fair generalizations about the population. To address within-group variability, generalized logistic mixed models with both fixed and random effects were used for data analysis. For details, see Section 5.1.

As for heritage speakers' self-rated English proficiency, it was high and close to ceiling across all four language skills. The mean ratings for listening, speaking, reading, and writing were 5.94, 5.94, 5.91, and 5.82 respectively. The standard deviations were all below 0.50, which means there was very little variability. In the bottom row of Figure 4.17, the "red boxes" are rendered as lines because the fences and whiskers overlap.

Homeland speakers also gave high ratings for their dominant language, which is, in this case, Cantonese. All four skills had a mean rating above 5 in Table 4.14a: 5.65 for listening (SD = 0.64), 5.79 for speaking (SD = 0.48), 5.56 for reading (SD = 0.70), and 5.32 for writing (SD = 0.94).

Similar to heritage speakers, homeland speakers gave lower ratings for their non-dominant language than their dominant language, but the difference was less drastic than heritage speakers. Their mean ratings for English language skills were all below 5 but above 4: 4.79 for listening (SD = 0.73), 4.56 for speaking (SD = 0.93), 4.64 for reading (SD = 0.77), and 4.23 for writing (SD = 1.10). In the bottom row of Figure 4.17, the asterisks inside the blue boxes are somewhat descending from left to right, but it is not as dramatic as the red boxes in the top row.

On the whole, the average self-ratings of homeland and heritage speakers for each language and each skill were significantly different, as shown in Table 4.14. Since the word-identification tasks in the main study only involved listening comprehension in Cantonese, their ratings for this skill were most noteworthy. Homeland speakers' self-rated Cantonese listening proficiency (M = 5.65, SD = 0.64) was significantly higher than heritage speakers' (M = 4.47, SD = 1.13), $t(52.36) = 5.26$, $p < .001$. Due to this difference of listening abilities in Cantonese, an accuracy gap would also be expected for results in the main study. If the accuracy gaps across stimulus types were static, it would suggest that the observed differences between two groups were merely a reflection of their different Cantonese proficiency. However, if the accuracy gap was bigger for a particular stimulus type (e.g. Type 3, words with no segments) but not others, it would suggest that the observed differences were due to factors beyond language proficiency (e.g. their tone discrimination abilities).

To sum up, both homeland and heritage speakers gave higher ratings for their respective dominant language. However, for heritage speakers, the difference between dominant and non-dominant language was more extreme. Ratings for their dominant language (English) were extremely high and variability was extremely small, but ratings for their non-dominant language (Cantonese) varied a lot from skill to skill; even for a particular skill, there was a lot of variability within the same population. In general, these ratings met earlier expectations.

Chapter 5

Results

This chapter is divided into four sections. Section 5.1 presents an overview of results by analyzing outputs of generalized logistic mixed models. The other three sections discuss the three research questions in (1) respectively in detail: Section 5.2 responds to the first research question by discussing the effects of variable manipulation on accuracy; Section 5.3 answers the second research question by reporting patterns of tonal confusion; finally, Section 5.4 addresses the third research question by comparing the two populations' use of acoustic and semantic cues.

(1) Research questions of the current study (repeated)

a. Do homeland and heritage speakers behave differently in terms of their ability to identify tonally contrastive words?

b. Do homeland and heritage speakers exhibit similar confusion patterns with respect to lexical tone perception?

c. Do homeland and heritage speakers make use of the same type of information when identifying a word from a tonally contrastive set? In particular, are acoustic and semantic information equally useful?

5.1 Overview: Hypothesis testing with generalized logistic mixed models

This section is an overview exploring whether the manipulated independent variables, such as the presence or absence of segmental or tonal information, were effective predictors of the dependent variable of this study, namely accuracy. Results of the word-identification experiment were analyzed in R (R Core Team, 2013) using the `lme4` package (Bates, Mächler, Bolker & Walker, 2015). Generalized logistic mixed models (GLMMs) were fit to the experiment data to see which combination of model parameters could best describe the data. In all models being fit, `accuracy` was treated as a discrete variable, where 1 means correct and 0 means incorrect. Therefore, logistic regression was implemented by the `glmer` function with the argument `family="binomial"`.

GLMMs are "mixed" because predictors are a combination of fixed and random effects. The fixed effects to be examined are the experimentally manipulated independent variables, namely `population` and the type of information available in the stimuli: `is_there_segment`, `is_there_tone`, and `context`. Among them, `population`, `is_there_segment`, and `is_there_tone` were binary variables: 0 refers to the reference category ("homeland" or "there is no X"), and 1 refers to the non-reference category ("heritage" or "there is X"). As for `context`, there were three levels: "yes congruous", "yes incongruous", and "no". Among the three levels, "yes congruous" was chosen to be the reference category to facilitate the comparison between no context and congruous contexts, and the comparison between congruous and incongruous contexts.

Interactions of the variables above were also fixed effects. These interactions may change the predictive power of a model drastically, because they can show whether the effect of Predictor A is significantly different for different values of Predictor B. For example, coefficients for the interaction of `is_there_tone` and `population` (expressed with an asterisk, as in `is_there_tone*population`) can show whether the effect of the presence or absence of tone for homeland speakers was significantly different from the effect of the presence or absence of tone for heritage speakers, which is exactly what was being tested in this dissertation.

Four-way (is_there_segment*is_there_tone*context*population) and three-way interactions (is_there_tone*context*population) were considered, but these models failed to converge. Multiple two-way interactions in one model were also considered, but the models failed to converge. When only one two-way interaction was included at a time, the models converged successfully. They were referred to as Models I, II, and III, which had is_there_tone*population, is_there_segment*population, and context*population as a fixed effect respectively, as in Table 5.1.

Table 5.1: Fixed and random effects of three generalized logistic mixed models predicting accuracy

	Effects	
	Fixed	Random
Model I	is_there_segment + is_there_tone + context + population + is_there_tone*population	random slopes for subject and syllable for is_there_segment, is_there_tone, and context
Model II	is_there_segment + is_there_tone + context + population + is_there_segment*population	random slopes for subject and syllable for is_there_segment, is_there_tone, and context
Model III	is_there_segment + is_there_tone + context + population + context*population	by-subject and by-syllable random intercepts

To improve the models' ability to assess the effects of fixed-effect variables, random effects were included to account for idiosyncratic variations that were unpredictable. For example, the following situations could have happened (though they were not actually reported): Participant A was startled when hearing the word *si2* "poop" during the experiment and would press the wrong button whenever this word came up. Participant B found the picture for *se3* "diarrhea" amusing and would always be distracted by it, hence not paying attention to the stimulus. Participant C knew the word *ji1* "to cure" particularly well compared with other Tone 1 words because his father was a doctor (*ji1 sang1*) and he grew up hearing this word every day. To account for such within-population variability, `subject` and `syllable` were examined as random effects in the models.

As emphasized by Barr, Levy, Scheepers & Tily (2013), in order to minimize Type I error rates in confirmatory hypothesis testing, random effect structures should be as maximally specified as possible. To follow their recommendation, a random effect structure with all of the following was considered for every model: by-`subject` and by-`syllable` random intercepts, as well as random slopes for `subject` and `syllable` for `is_there_segment`, `is_there_tone`, and `context`.

A model with the interaction `context*population` and both random slopes and random intercepts was considered, but it failed to converge. A modified model including random slopes but excluding random intercepts was considered, but it also failed to converge. Therefore, the random effect structure of Model III in Table 5.1 only had random intercepts but not random slopes.

Models with random intercepts, random slopes, and fixed effects of `is_there_tone*population` or `is_there_segment*population` converged successfully. To test whether random intercepts improved the predictive power of a model, pairs of models only differed by the inclusion or exclusion of random intercepts were compared through a likelihood ratio test using the `anova()` function. Results of the test show that by-`subject` and by-`syllable` random intercepts did not improve the models, $\chi^2(1) = 0, p = 1$. For this reason, in Models I and II, random intercepts were dropped but random slopes were kept, as shown in Table 5.1.

Each of Models I, II, and III included the interaction of `population` with a different variable. Their outputs will be reported and analyzed in the rest of this section. Note that logistic regression coefficients are log odds[1], so they are by no means the predicted value of accuracy. However, these coefficients can show the general trend of whether accuracy was boosted or lowered.

Coefficient estimates of fixed effects in Model I are summarized in Table 5.2. The intercept refers to reference categories, namely "homeland", "there is no X", or "yes congruous". Effects of `is_there_segment` (β = 0.630, *SE* = 0.128, z = 4.925, $p < .001$) and `is_there_tone` (β = 2.962, *SE* = 0.314, z = 9.375, $p < .001$) were significant. This means the presence of segments or tone helped homeland speakers to boost their performance significantly. As for `context`, "yes incongruous" had a negative coefficient, which indicates that going from "yes congruous" to "yes incongruous" led to a significant decrease of accuracy for homeland speakers (β = –2.408, *SE* = 0.209, z = –11.543, $p < .001$). In a similar fashion, going from "yes congruous" to "no" also led to a significant decrease of accuracy for homeland speakers (β = –2.796, *SE* = 0.208, z = –13.429, $p < .001$). These results were not surprising, as turning a congruous context into an incongruous one or taking away the semantic context altogether was expected to enhance the level of difficulty of the task in general.

Since heritage speakers were of particular interest, the variable `population` and its interaction with `is_there_tone` are the focus of the discussion for Model I. In Table 5.2, the effect of `population` was not significant (β = –0.094, *SE* = 0.095, z = –0.997, p = .319). It should be clarified that this coefficient was compared against the intercept, which means when there was a congruous context but there were no segments and no tone, going from "homeland" to "heritage" would not make a significant effect on accuracy. Although no actual stimuli had no segments and no tone at the same time, it would have been an impossible task for both homeland and heritage speakers. Therefore, it was not surprising that the effect of

[1] Conversion from log odds to probabilities can be done using the formula p = odds/(1 + odds), where the odds are calculated by exponentiating the sums of the relevant coefficients in the model. For details, see Chapter 8 of Sweet & Grace-Martin (1999).

Table 5.2: Summary of fixed effects of Model I, a generalized logistic mixed model that included the interaction of `is_there_tone` and `population`, predicting `accuracy`

	β	*SE*	z	p	
(Intercept)	1.452	0.245	5.914	$<$ .001	***
is_there_segment(yes)	0.630	0.128	4.925	$<$ .001	***
is_there_tone(yes)	2.962	0.314	9.375	$<$ .001	***
context(yes_incong)	–2.408	0.209	–11.543	$<$ .001	***
context(no)	–2.796	0.208	–13.429	$<$ .001	***
population(heritage)	–0.094	0.095	–0.997	.319	
is_there_tone(yes) *population(heritage)	–1.404	0.212	–6.623	$<$ .001	***

Note: * $p < .05$; ** $p < .01$; *** $p < .001$

Number of observations:	48960.0
Akaike's Information Criterion:	37924.1
Bayesian Information Criterion:	38248.9
Log-likelihood:	–18925.0

`population` was not significant.

The most important observation in Table 5.2 is that effect of the interaction `is_there_tone*population` was significant (β = –1.404, *SE* = 0.212, z = –6.623, $p < .001$). In other words, the presence or absence of tonal information had significantly different effects for different populations. For homeland speakers, going from "there is no tone" to "there is tone" means an increase of log odds from 1.452 (intercept) to 4.414 (1.452 + 2.962). Going from *both* "homeland" to "heritage" *and* "there is no tone" to "there is tone" leads to an increase of log odds from 1.358 (1.452 – 0.094) to 2.916 (1.452 – 0.094 + 2.962 – 1.404). In other words, the presence of tonal information gave a boost for both homeland and heritage speakers, but the boost for homeland speakers (2.962) was significantly bigger than for heritage speakers (1.558). This supports the hypothesis that heritage speakers make less use of tonal information when identifying a word from a tonally contrastive set.

Coefficient estimates of fixed effects in Model II are reported as follows. The effects of interest in this model are is_there_segment and its interaction with population. As summarized in Table 5.3, the effect of is_there_segment was significant (β = 0.513, *SE* = 0.153, z = 3.349, $p < .001$). This means going from "there are no segments" to "there are segments" enhanced homeland speakers' accuracy. However, the effect of is_there_segment*population was not significant (β = 0.211, *SE* = 0.169, z = 1.251, p = .211). In other words, the presence or absence of segmental information affected both homeland and heritage speakers, but its impact on homeland speakers was not significantly different from its impact on heritage speakers.

Table 5.3: Summary of fixed effects of Model II, a generalized logistic mixed model that included the interaction of is_there_segment and population, predicting accuracy

	β	*SE*	z	p	
(Intercept)	1.733	0.252	6.874	$< .001$	***
is_there_segment(yes)	0.513	0.153	3.349	$< .001$	***
is_there_tone(yes)	2.254	0.312	7.238	$< .001$	***
context(yes_incong)	–2.417	0.210	–11.533	$< .001$	***
context(no)	–2.808	0.209	–13.435	$< .001$	***
population(heritage)	–0.620	0.180	–3.445	$< .001$	***
is_there_segment(yes) *population(heritage)	0.211	0.169	1.251	.211	

Note: * $p < .05$; ** $p < .01$; *** $p < .001$

Number of observations:	48960.0
Akaike's Information Criterion:	37956.7
Bayesian Information Criterion:	38281.6
Log-likelihood:	–18941.4

Coefficient estimates of Model III, which included context and its interactions, are reported as follows. As mentioned previously, the random effect structure of Model III was different from that of Models I and II. Due to convergence problems, Model III did not include random slopes for subject and syllable for is_there_segment, is_there_tone, and context. However, by-subject and by-syllable intercepts were included.

Fixed effects of Model III are summarized in Table 5.4. The effects of `context(yes_incong)` (β = –1.589, *SE* = 0.066, z = –24.24, $p < .001$) and `context(no)` (β = –2.240, *SE* = 0.063, z = –35.52, $p < .001$) were significant. These coefficients were negative numbers, which means going from "yes congruous" to "yes incongruous" lowered the accuracy of homeland speakers. Going from "yes congruous" to "no" also lowered the accuracy of homeland speakers. This was not surprising, as taking away the semantic context or making a sentence incongruous would enhance the level of difficulty of the task in general.

Table 5.4: Summary of fixed effects of Model III, a generalized logistic mixed model that included the interaction of `context` and `population`, predicting `accuracy`

	β	*SE*	z	p	
`(Intercept)`	1.565	0.173	9.06	$< .001$	***
`is_there_segment(yes)`	0.540	0.027	19.78	$< .001$	***
`is_there_tone(yes)`	1.871	0.040	46.80	$< .001$	***
`context(yes_incong)`	–1.589	0.066	–24.24	$< .001$	***
`context(no)`	–2.240	0.063	–35.52	$< .001$	***
`population(heritage)`	–1.069	0.183	–5.85	$< .001$	***
`context(yes_incong) *population(heritage)`	–0.446	0.075	–5.92	$< .001$	***
`context(no) *population(heritage)`	0.113	0.079	1.43	.152	

Note: * $p < .05$; ** $p < .01$; *** $p < .001$

Number of observations:	48960.0
Akaike's Information Criterion:	42779.4
Bayesian Information Criterion:	42867.2
Log-likelihood:	–21379.7

Since heritage speakers were of particular interest, the interaction of `context` and `population` is the focus of the discussion. Since `context` had three levels, there were two lines of interactions in Table 5.4. First, the effect of the interaction of `context(yes_incong)` with `population(heritage)` was significant (β = –0.446, *SE* = 0.075, z = –5.92, $p < .001$). This means going from "yes congruous" to "yes incongruous" made a negative impact on the accuracy of both homeland and heritage speakers. However, this impact was bigger on heritage speakers compared with homeland speakers. Specifically, for homeland speakers, the impact was –1.589 in log odds. For heritage speakers, the impact was –2.035 (–1.589–0.446) in log odds. The fact that the heritage group's performance suffered more supports the hypothesis that heritage speakers found it more difficult to actively ignore semantic context and use tonal information when the target words was semantically incongruous with the carrier phrase.

However, the effect of `context(no)*population(heritage)` was not significant ($\beta = 0.113$, $SE = 0.079$, $z = 1.43$, $p = 0.152$). This suggests that going from "yes congruous" to "no" made an impact on the accuracy of both homeland and heritage speakers, but the impact was not significantly different between the two groups.

Overall, results of logistic regression with fixed and random effects confirmed that the presence of tonal information gave a significantly bigger boost for homeland speakers' performance than heritage speakers'. When the target word was semantically incongruous with the carrier phrase, the performance of the heritage group suffered significantly more than the homeland group. Since heritage speakers are known to be a heterogeneous group, the inclusion of random effects made sure that inter-subject variability was factored into the analysis, and so the scope of inference can be extended to the entire population.

Following this overview section, Section 5.2 will compare the two groups' average accuracy for each stimulus type in the word-identification experiment.

5.2 Response to Research Question 1: Accuracy

The rest of this chapter addresses the three research questions of this dissertation in more detail. This section responds to the first research question, which is re-stated below with its null and alternative hypotheses:

> **Research Question 1:**
>
> Do homeland and heritage speakers behave differently in terms of their ability to identify tonally contrastive words?
>
> H_0: There is no difference between homeland and heritage speakers in terms of their ability to identify tonally contrastive words.
>
> H_1: Compared with homeland speakers, heritage speakers are less able to identify tonally contrastive words.

Previously in Section 3.6, it was hypothesized that heritage speakers make less use of tonal information for word identification than homeland speakers. In general, this hypothesis predicts that H_0 will be rejected in favour of H_1. However, on the level of specific stimulus types, the predictions for H_0 and H_1 may differ. Recall from Section 4.1.3 that each of the eight stimulus types represented a specific way to manipulate four variables: segmental information, tonal, semantic context, and semantic congruity. A recap of the tested configurations is provided in Table 5.5. Consider Type 2 (words with no tone) and Type 4 (sentences with no tone) as examples. Since the removal of tonal information was hypothesized to render the two groups equal, no difference between homeland and heritage speakers was expected, hence predicting that H_1 would be rejected in favour of H_0. As for the rest of the stimulus types, an accuracy gap between the two groups was expected in favour of H_1. However, the crucial observation to be made here is not only the presence or absence of an accuracy gap, but also the size of the gap. Therefore, the subsequent discussion will address two questions for each stimulus type: first, was any observed difference between homeland and heritage speakers statistically significant? Second, if there was a significant difference, what was its magnitude?

To answer the first question regarding statistical significance, *t* tests for

Table 5.5: Recap of stimulus types and predicted results

Type	Segment	Tone	Context	Congruity	Predicted result		
1	✓	✓	✗	not appl.	homeland	>>	heritage
2	✓	✗	✗	not appl.	homeland	=	heritage
3	✗	✓	✗	not appl.	homeland	>>>	heritage
4	✓	✗	✓	✓	homeland	=	heritage
5A	✓	✓	✓	✓	homeland	>	heritage
5B	✓	✓	✓	✗	homeland	>>	heritage
6A	✗	✓	✓	✓	homeland	>>	heritage
6B	✗	✓	✓	✗	homeland	>>	heritage

two independent samples were conducted to compare the average accuracy of homeland and heritage speakers for each stimulus type. To decide whether t tests assuming equal or unequal variances should be used, F tests for comparing variances of two samples were performed. Results of the F tests indicate that variances of homeland and heritage speakers were significantly different for Types 1, 3, 5A, 5B, 6A, and 6B, and so t tests assuming unequal variances were used. For Types 2 and 4, variances were not significantly different, and so t tests assuming equal variances were used. An overview of the results is presented in Table 5.6. Since eight pairwise t tests were performed on a single data set, a Bonferroni correction was made to reduce the chances of obtaining false-positive results. The alpha level was adjusted to a more conservative value. This was done by dividing the normal alpha level (0.05) by the number of pairwise comparisons (in this case, 8), which yields 0.00625. This means a p value above 0.00625 indicates that an observed difference is statistically non-significant, while a p value below 0.00625 indicates that the difference is statistically significant.

Table 5.6: Accuracy rates arranged by effect size in form of Cohen's *d* (smallest to largest)

Type	Population	*M* (in %)	*SD*	*t*	*df*	*p*	Cohen's *d*	Predicted result		
2	homeland heritage	30.05 28.97	45.86 45.37	1.13	63.64	.26	0.27	homeland	=	heritage
4	homeland heritage	89.22 82.25	31.03 38.21	1.71	60.14	.09	0.41	homeland	=	heritage
5A	homeland heritage	95.83 86.37	19.99 34.32	3.99	57.60	<.001	0.97	homeland	>	heritage
5B	homeland heritage	87.14 66.31	33.37 47.27	5.07	42.36	<.001	1.22	homeland	>>	heritage
6A	homeland heritage	94.17 80.54	23.44 39.60	5.26	58.49	<.001	1.28	homeland	>>	heritage
6B	homeland heritage	85.34 56.72	35.44 49.55	5.67	48.32	<.001	1.37	homeland	>>	heritage
1	homeland heritage	91.37 69.22	28.08 46.17	7.89	42.83	<.001	1.91	homeland	>>	heritage
3	homeland heritage	80.54 49.41	39.60 50.01	8.28	56.96	<.001	2.00	homeland	>>>	heritage

Note: The alpha level after Bonferroni correction was 0.00625.

To answer the second question regarding the magnitude of difference, effect sizes in Cohen's *d* (Cohen, 1988) were measured. This was calculated by dividing the difference between two means by the pooled standard deviation of the data. The output value indicates the effect size, which can be interpreted as shown in Table 5.7. The eight stimulus types in Table 5.6 were arranged by effect size from the smallest to the largest. These results will be further discussed in the rest of this section.

Table 5.7: Interpretation of Cohen's *d* (Cohen, 1988; Sawilowsky, 2009)

Cohen's *d*	Effect size
2.00	huge
1.20	very large
0.80	large
0.50	medium
0.20	small
0.01	very small

In subsequent discussion, data will be visualized using boxplots, as in the example in Figure 5.1 based on real data from this study. The central line and the asterisk in the box indicate the median and the mean respectively. The lower and upper boundaries of the box show the lower and upper quartiles respectively. Whiskers are drawn to include any data points within a certain distance of the box: the interquartile range is multiplied by 1.5, and this number is added to the upper quartile and subtracted from the lower quartile. The lower and upper whiskers show the lowest and highest datum within this limit respectively. Points outside this limit are outliers.

Each of the four subsections that follow will illustrate the effect of one independent variable. Subsets of stimulus types in Table 5.5 will be singled out for discussion, and effect sizes will be checked against predictions made in Chapter 4. Section 5.2.5 will provide an interim summary and revisit the first research question of this dissertation.

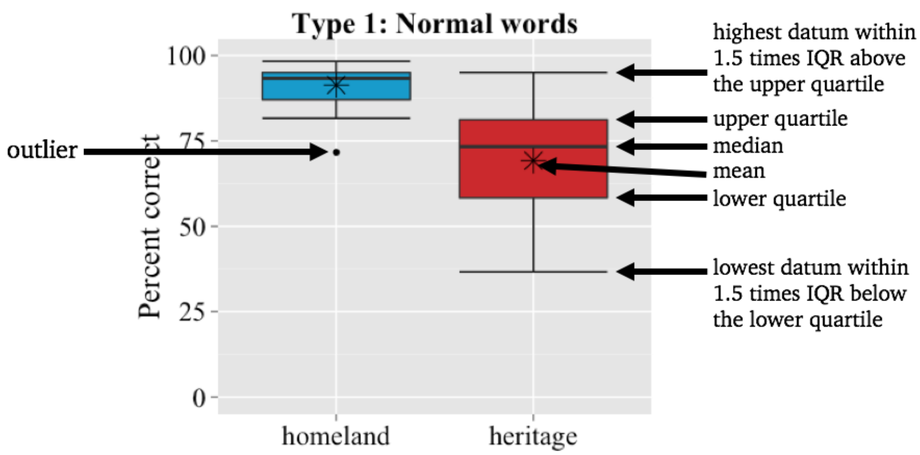


Figure 5.1: A sample boxplot

5.2.1 With vs. without context

The first effect being examined is the presence or absence of semantic context. Stimulus types that only differ by this variable will be discussed side by side: normal words will be compared with normal congruous sentences, words with no tone will be compared with sentences with no tone, and words with no segments will be compared with congruous sentences with no segments.

The first pair for comparison is Type 1 (normal words) and Type 5A (normal, congruous sentences), and their variable configuration is summarized in Table 5.8. This pair is presented first for discussion, because these two stimulus types can be a window to understanding the two groups' baseline listening ability in Cantonese. For Type 5A, it has a ✓ mark for all four variables, which means it contained all of the following listening cues: segmental information, tonal information, semantic context, and semantic congruity. Therefore, this stimulus type should be the least challenging task for both homeland and heritage speakers. Recall that in Section 4.4.3.3, homeland (M = 5.65 out of 6, SD = 0.64) and heritage speakers (M = 4.47 out of 6, SD= 1.13) gave significantly different self-ratings for their Cantonese listening abilities. It was expected that their results for Type 5A would reflect this baseline difference.

As for Type 1, although these stimuli also contained segmental and tonal information, the target words were not embedded in a carrier phrase, and so listeners did not have a frame of reference of the talker's pitch. Generally speaking, monosyllabic words are more difficult than sentences. Both groups were expected to have a lower accuracy compared with Type 5A. However, what is of particular interest here is not the difference of accuracy *within* one population across stimulus types; rather, it is the comparison of accuracy gaps *between* the two groups across stimulus types. If heritage speakers indeed relied on semantic context more often than they did for tonal information in word identification, a larger accuracy gap between homeland and heritage speakers would be expected for Type 1 compared with Type 5A. However, if homeland and heritage speakers did not differ in terms of how much they relied on semantic information, then the accuracy gaps for Type 1 and Type 5A were expected to be the same, which would just be a reflection of their difference in Cantonese proficiency in general.

Results for Type 5A in Figure 5.2b met expectations. Accuracy of homeland speakers was near ceiling (M = 95.83%, SD = 19.99). Four outliers fell below 90%, but they were all above 70%. The short whiskers suggest very little variability among homeland speakers. Heritage speakers also did well in this task (M = 86.37%, SD = 34.32), but their accuracy was significantly lower than that of homeland speakers, $t(57.61) = 3.99$, $p < .001$. The relatively high standard deviation and longer whiskers in the plot suggests large variability within the heritage group.

The effect size for Type 5A was large ($d = 0.97$), which has important implications for other stimulus types: if heritage speakers' lower overall Cantonese proficiency was the sole reason for any difference between the two populations, then a similar effect size (0.97) should be expected across all eight stimulus types. However, if a larger effect size was observed in other stimulus types, such magnitude of difference must be due to reasons other than heritage speakers' lower overall Cantonese proficiency.

Results for Type 1 in Figure 5.2a also met expectations. The difference between homeland (M = 91.37%, SD = 28.08) and heritage speakers

Table 5.8: Stimulus types with all acoustic information

Type	Segment	Tone	Context	Congruity	Predicted result		
1	✓	✓	✗	not appl.	homeland	>>	heritage
5A	✓	✓	✓	✓	homeland	>	heritage

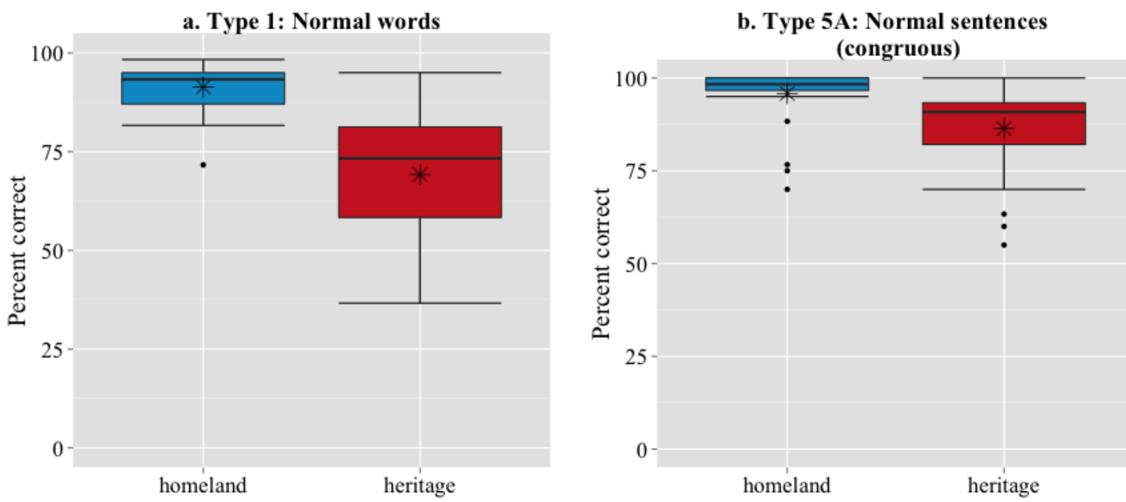


Figure 5.2: Comparison of stimulus types with all acoustic information

(M = 69.22%, SD = 46.17) was significant, $t(42.83) = 7.89$, $p < .001$. Whiskers for the heritage group span between 36.67% and 95.00%, which suggests an even higher degree of variability compared to Type 5A. The effect size for Type 1 stimuli was close to huge ($d = 1.91$).

The actual results above matched the predicted results: even when all acoustic information was available, the magnitude of difference between the two groups for normal words was huge. When semantic context was available in Type 5A, however, the gap between the two groups became smaller. This implies that semantic context had helped both groups to achieve a higher accuracy, but it helped heritage speakers even more proportionally.

The second pair for comparison is Type 2 (words with no tone) and Type 4 (sentences with no tone), and their variable configuration is summarized in Table 5.9. According to the hypotheses of this dissertation (see Section 3.6), the difference between homeland and heritage speakers lay in the usefulness of tonal information in word identification. Therefore, removing tonal information from the stimuli would render the two groups equal (or close to equal, considering their different language proficiency in general). In particular, Type 2 (words with no tone) was expected to be an impossible task for both groups, as the target word and competitors had identical segments and only differed by tone. It was predicted that accuracy rates of both groups would be at chance (25%). As for Type 4, no significant difference was expected, as it was hypothesized that both populations would be equally able to make use of semantic information for word identification.

Table 5.9: Stimulus types with no tone

Type	Segment	Tone	Context	Congruity	Predicted result		
2	✓	✗	✗	not appl.	homeland	=	heritage
4	✓	✗	✓	✓	homeland	=	heritage

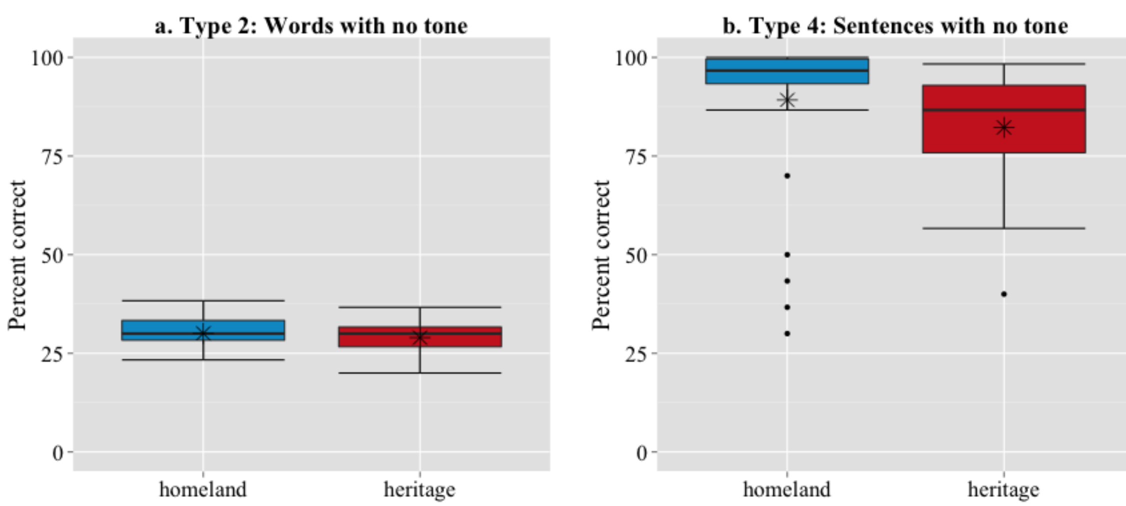


Figure 5.3: Comparison of stimulus types with no tone

Results for Type 2 in Figure 5.3a confirmed that both groups' performance was close to chance (25%). The accuracy of homeland (M = 30.05%, SD = 45.86) and heritage speakers (M = 28.97%, SD = 45.37) was very similar, $t(63.64) = 1.13$, $p = .26$, $d = 0.27$. Both groups had similar standard deviations, which suggests that no particular population had more variability within the group. These met the expectation that Type 2 was an impossible task for both groups.

For Type 4 in Figure 5.3b, the difference between homeland (M = 89.22%, SD = 31.03) and heritage speakers (M = 82.25%, SD = 38.21) was also not significant, $t(60.14) = 1.71$, $p = .09$, $d = 0.41$. It is noteworthy that five outliers in the homeland group scored below 70%. These individuals mostly chose words with the mid level tone instead of words that would make sense in the given sentence. These cases will be discussed in detail in Section 5.3. To sum up, results of Type 2 and Type 4 show that when tonal information was unavailable, the two groups performed similarly—either they both performed poorly (as in Type 2), or they both performed well (as in Type 4).

The third pair for comparison is Type 3 (words with no segments) and Type 6A (the last word of the congruous sentence has no segments), and their variable configuration is summarized in Table 5.10. Although Type 6A stimuli were congruous sentences, they were predicted to be relatively challenging for heritage speakers due to their lack of segmental information plus the fact that all congruous and incongruous sentences were mixed in one experimental block. As for Type 3, its lack of semantic context was predicted to make the task even harder for heritage speakers. Therefore, the magnitude of difference between the two groups was expected to be larger for Type 3 than Type 6A.

Results for Type 3 in Figure 5.4a show that although this task was extremely challenging for heritage speakers, their performance was above chance (M = 49.41%, SD = 50.01). As expected, homeland speakers did significantly better (M = 80.54%, SD = 39.60) than heritage speakers, $t(56.96) = 8.28$, $p < .001$. The effect size of this comparison was huge ($d = 2.00$), the largest of all stimulus types. For Type 6A in Figure 5.4b, the difference between homeland (M = 94.17%, SD = 23.44) and heritage speakers (M = 80.54%, SD = 39.60) was significant, $t(58.49) = 5.26$, $p < .001$. The effect size was not as huge as that of Type 3, but it was still very large ($d = 1.28$). Since Types 3 and 6A only differed by the presence or absence of semantic context, it can be concluded that the presence of semantic context significantly decreased the gap between homeland and heritage speakers.

Table 5.10: Stimulus types with no segments

Type	Segment	Tone	Context	Congruity	Predicted result		
3	✗	✓	✗	not appl.	homeland	>>>	heritage
6A	✗	✓	✓	✓	homeland	>>	heritage

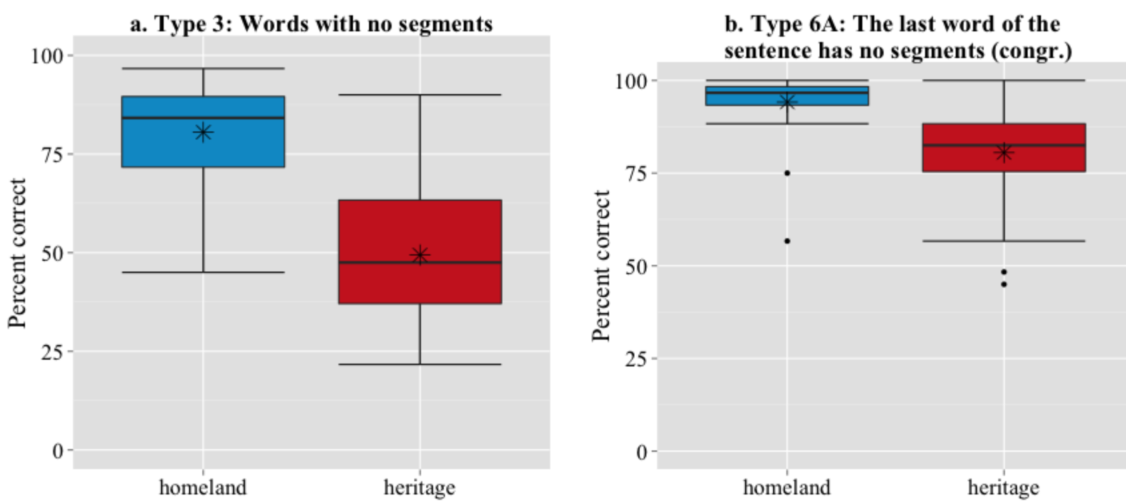


Figure 5.4: Comparison of stimulus types with no segments

5.2.2 With vs. without congruity

This subsection examines the effect of congruity by comparing stimulus types that only differed by this variable. The first pair for comparison is Type 5A (normal, congruous sentences) and Type 5B (normal, incongruous sentences). Their variable configuration is summarized in Table 5.11. As mentioned previously, the two groups' performance in Type 5A was expected to reflect their baseline Cantonese listening proficiency. Therefore, an accuracy gap was expected for Type 5A stimuli. As for Type 5B, target words were semantically incongruous with the carrier phrase. It was hypothesized that heritage speakers would find it challenging to actively ignore the semantic context, and identify a word by relying on tonal information only. Therefore, the gap between homeland and heritage speakers was predicted to be larger for Type 5B compared with 5A.

Table 5.11: Stimulus types with context and all acoustic information

Type	Segment	Tone	Context	Congruity	Predicted result		
5A	✓	✓	✓	✓	homeland	>	heritage
5B	✓	✓	✓	✗	homeland	>>	heritage

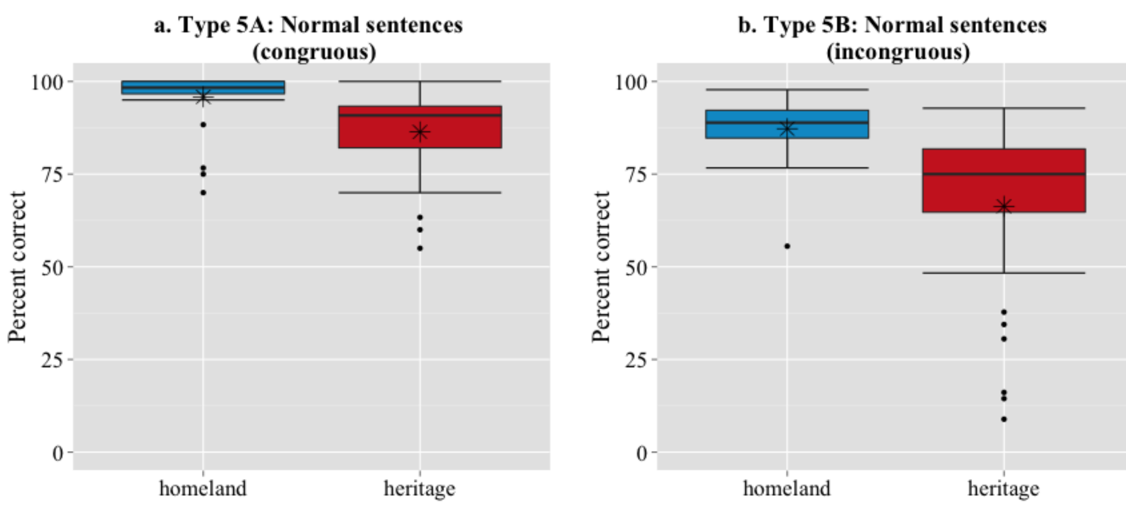


Figure 5.5: Comparison of stimulus types with context and all acoustic information

Results presented in Figure 5.5 show that this prediction was correct. For Type 5A, the difference between homeland (M = 95.83%, SD = 19.99) and heritage speakers (M = 86.37%, SD = 34.32) was significant,

$t(57.60) = 3.99$, $p < .001$; the effect size was large ($d = 0.97$). For Type 5B, the difference between homeland (M = 87.14%, SD = 33.37) and heritage speakers (M = 66.31%, SD = 47.27) was also significant, $t(42.36)=5.07$, $p < .001$; the effect size was very large ($d = 1.22$). Both populations showed more intra-group variability for Type 5B compared with 5A. For Type 5B in Figure 5.5b, the most accurate listener from the homeland group achieved 97.78% accuracy, and the bulk of this group achieved at least 73.47% accuracy. There was one outlier who achieved only 55.56% accuracy, suggesting that this homeland speaker also found Type 5B challenging. As for heritage speakers, their accuracy range was even larger. On the one hand, the most accurate listener from this group achieved 92.78% accuracy, which was comparable to the performance of homeland speakers. On the other hand, there were six outliers in the heritage group. In particular, three of the six outliers fell below chance (25%), and the lowest outlier was at 8.89%. An accuracy below chance suggests that these errors were not made randomly. Section 5.4 will explain that heritage speakers relied on semantic context relatively more often than homeland speakers did, and so they made errors due to choosing a word that was semantically congruous with the carrier phrase. This pulled the two groups apart, and led to a larger accuracy gap in Type 5B than in Type 5A.

The next pair for comparison is Type 6A (the last word of the congruous sentence has no segments) and Type 6B (the last word of the incongruous sentence has no segments). Their variable configuration is summarized in Table 5.12. Note that 5A-5B and 6A-6B were both congruous-incongruous pairs. In general, "B" stimuli were expected to be harder for heritage speakers than their "A" counterparts due to the difference in congruity. However, for 6A and 6B, there was an additional complication of the lack of segmental information, which means heritage speakers would have to rely on tonal information for both 6A and 6B. Given the hypothesis that heritage speakers were less good at using tonal information, Types 6A and 6B were both predicted to be challenging for them.

Results show that Types 6A and 6B had similar effect sizes. For Type 6A in Figure 5.6a, the difference between homeland (M = 94.17%, *SD* = 23.44) and heritage speakers (*M* = 80.54%, *SD* = 39.60) was significant, $t(58.49) = 5.26$, $p < .001$. For Type 6B in Figure 5.6b, the difference of the average accuracy between homeland (*M* = 85.34%, *SD* = 35.44) and heritage speakers (*M* = 56.72%, *SD* = 49.55) was also significant, $t(48.32) = 5.67$, $p < .001$. Effect sizes of Type 6A ($d = 1.28$) and 6B ($d = 1.37$) were similar and both translate to "very large".

Table 5.12: Stimulus types with context and tonal information only

Type	Segment	Tone	Context	Congruity	Predicted result		
6A	✗	✓	✓	✓	homeland	>>	heritage
6B	✗	✓	✓	✗	homeland	>>	heritage

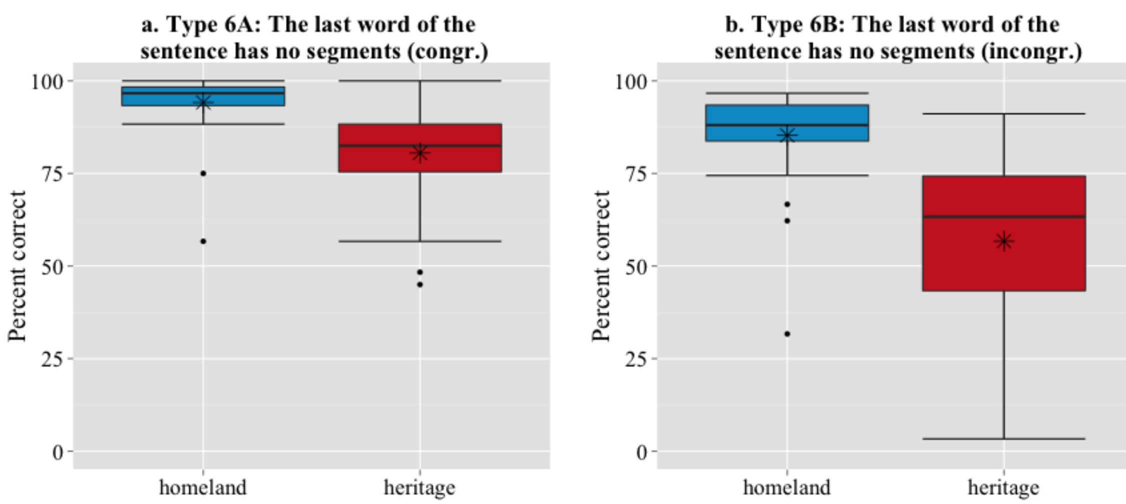


Figure 5.6: Comparison of stimulus types with no segments

For each population, a wider accuracy range was observed for Type 6B compared with 6A. In Figure 5.6b, the most accurate listener from the homeland group achieved 96.67% accuracy, while the lowest outlier in the same group was at 31.67%. The heritage group had even more extreme performance, ranging from 3.33% to 91.11%. Six heritage speakers' accuracy fell below chance (25%), which suggests that their errors were not random. Section 5.4 will further investigate the patterns in these errors and show that it was comparatively harder for heritage speakers to ignore the semantic context when the target word was semantically incongruous with its carrier phrase.

At this point the boxplots of all eight stimulus types have been shown. In the two subsections that follow, different pairs from the same set of boxplots will be juxtaposed to illustrate the effect of variables that have not been discussed. Relevant statistics such as mean accuracy and standard deviations will be repeated for clarity.

5.2.3 With vs. without segmental information

This subsection discusses the effect of the presence or absence of segmental information by comparing stimulus types that only differed by this variable. The first pair for comparison is Type 1 (normal words) and Type 3 (words with no segments). Their variable configuration is summarized in Table 5.13. Type 1 stimuli were tonally contrastive monosyllabic words which did not provide any semantic context. Since semantic context was hypothesized to be an important cue for heritage speakers, a big accuracy gap between the two populations was expected for Type 1. In addition to the lack of context, Type 3 also lacked segmental information. Although segmental information was redundant in this task as target words and competitors were segmentally the same, segmental information does facilitate lexical activation in general. When segmental information was also taken away, the only listening cue left was tonal information, which was hypothesized to be the least helpful for the heritage group. Therefore, the gap between the two groups was expected to be even bigger for Type 3 compared with Type 1.

Results in Figure 5.7 generally met expectations. For Type 1, the homeland group (M = 91.37%, SD = 28.08) did significantly better than the heritage group (M = 69.22%, SD = 46.17), $t(42.83) = 7.89$, $p < .001$. For Type 3, the homeland group (M = 80.54%, SD = 39.60) also did significantly better than the heritage group (M = 49.41%, SD = 50.01),

$t(56.96) = 8.28$, $p < .001$. Effect sizes of both stimulus types were on the "huge" end of the scale. Although the effect size of Type 1 ($d = 1.91$) was smaller than that of Type 3 ($d = 2.00$), this difference was smaller than the distance between two thresholds on the effect size scale.

Table 5.13: Stimulus types with no context

Type	Segment	Tone	Context	Congruity	Predicted result		
1	✓	✓	✗	not appl.	homeland	>>	heritage
3	✗	✓	✗	not appl.	homeland	>>>	heritage

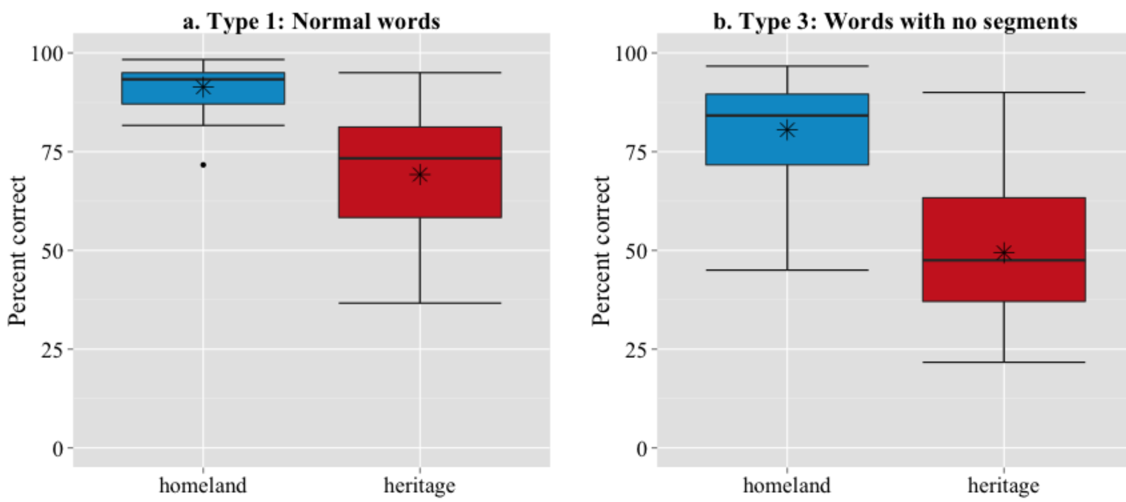


Figure 5.7: Comparison of stimulus types with no context

For Type 3, a wide range of accuracy was observed for both populations. However, the two groups' mean accuracy had skewed distributions in opposite directions in Figure 5.7b. For the homeland group, the lower whisker was longer than the upper whisker, and the median was closer to the upper quartile than the lower quartile. This suggests that while some homeland speakers did find this task challenging, they were the minority. The opposite can be observed in the heritage group: the upper whisker was longer than the lower whisker, and the median was closer to the lower quartile than the upper quartile. This suggests that while it was possible for heritage speakers to achieve an accuracy comparable to homeland speakers', they were the minority. The skewed distributions pulled the two groups apart, which led to a wider gap between them in Type 3 compared with Type 1.

The second pair for comparison is Type 5A (normal congruous sentences) and 6A (the last word of the congruous sentence has no segments). Their variable configuration is summarized in Table 5.14. Similar to the case of Type 1 vs. Type 3, since the only difference between Type 5A and Type 6A was that Type 6A lacked segmental information, it was predicted that this would enhance the level of difficulty for heritage speakers.

Table 5.14: Stimulus types with context and congruity

Type	Segment	Tone	Context	Congruity	Predicted result		
5A	✓	✓	✓	✓	homeland	$>$	heritage
6A	✗	✓	✓	✓	homeland	$>>$	heritage

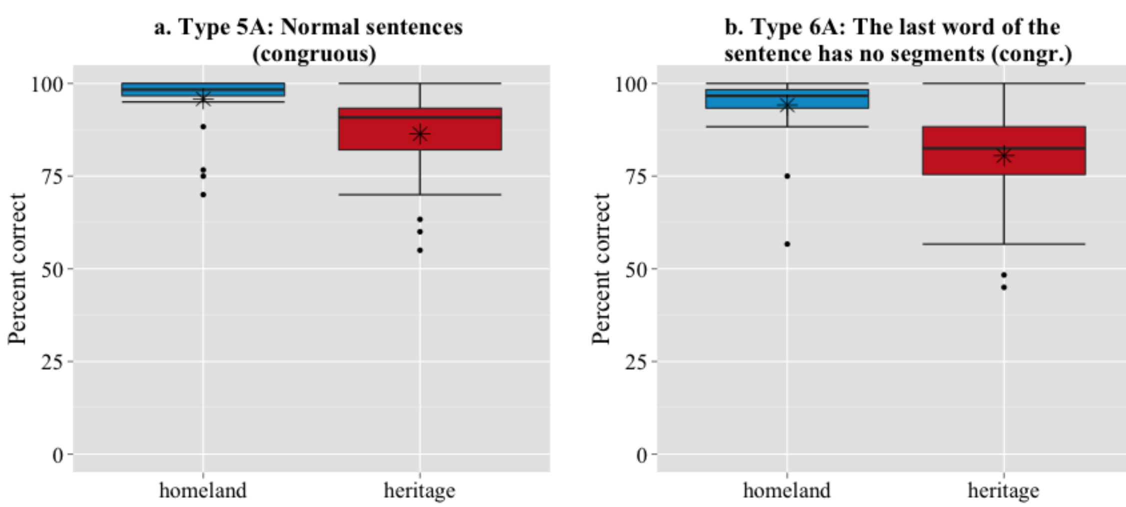


Figure 5.8: Comparison of stimulus types with context and congruity

Results in Figure 5.8 show that the prediction was correct. For Type 5A, the difference between homeland (M = 95.83%, SD = 19.99) and heritage speakers (M = 86.37%, SD = 34.32) was significant, $t(57.60) = 3.99$, $p < .001$. For Type 6A, the difference between homeland (M = 94.17%, SD = 23.44) and heritage speakers (M = 80.54%, SD = 39.60) was also significant, $t(58.49) = 5.26$, $p < .001$. The effect size of Type 5A was large ($d = 0.97$), while that of Type 6A was very large ($d = 1.28$). This difference in effect size confirmed that the lack of segmental information widened the gap between the two groups in Type 6A compared with Type 5A.

The last pair for comparison in this subsection is Type 5B (normal incongruous sentences) and 6B (the last word of the incongruous sentence has no segments). Their variable configuration is summarized in Table 5.15. Similar to 5A-6A, 5B-6B were also "with segment - no segment" pairs of sentences. In general, the "6" stimuli without segmental information were expected to be more difficult than their "5" counterparts with segmental information. However, Types 5B and 6B had an additional complication of having target words that were semantically incongruous with the carrier phrase. Given the hypothesis that heritage speakers would find it hard to actively ignore the semantic context and identify the word solely based on acoustic information, all kinds of incongruous sentences would be more challenging for heritage speakers than homeland speakers. Therefore, it was predicted that 5B and 6B would show a similar magnitude of difference between the two groups.

Results in Figure 5.9 are explained as follows. For Type 5B, the difference between homeland (M = 87.14%, SD = 33.37) and heritage speakers (M = 66.31%, SD = 47.27) was significant, $t(42.36)=5.07$, $p < .001$. For Type 6B, the difference of the average accuracy between homeland (M = 85.34%, SD = 35.44) and heritage speakers (M = 57.72%, SD = 49.55) was also significant, $t(48.32) = 5.67$, $p < .001$. The effect sizes for Type 5B and Type 6B were 1.22 and 1.37 respectively, both of which translate to "very large". Although the two effect sizes were not the same, their difference was smaller than the distance between two thresholds on the effect size scale. This was similar to the difference of effect sizes of Type

1 (normal words) and Type 3 (words with no segments).

Table 5.15: Stimulus types with context but no congruity

Type	Segment	Tone	Context	Congruity	Predicted result		
5B	✓	✓	✓	✗	homeland	>>	heritage
6B	✗	✓	✓	✗	homeland	>>	heritage

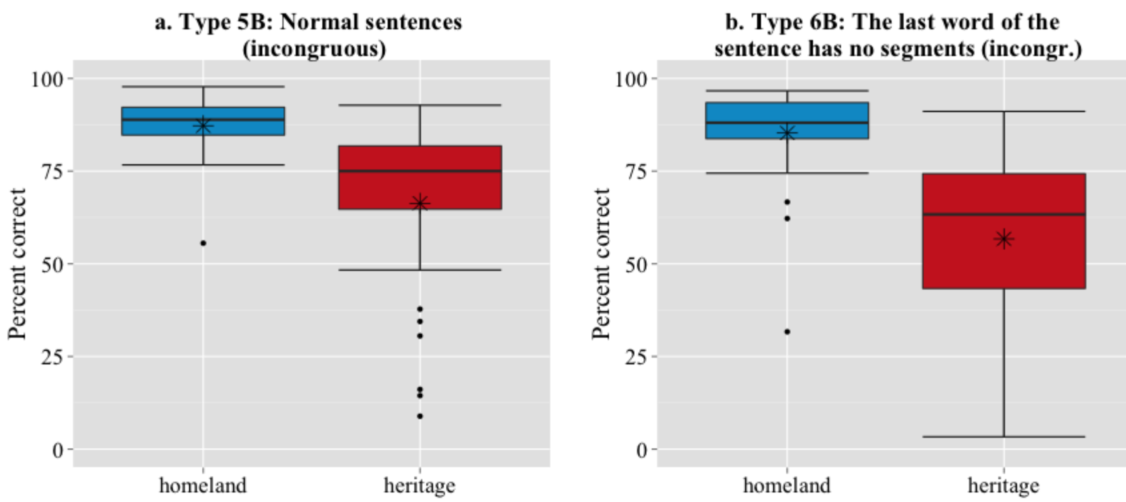


Figure 5.9: Comparison of stimulus types with context but no congruity

Taken together, the absence of segmental information made more impact on the accuracy gap for tasks that were relatively easy (Type 5A vs. 6A). For stimulus types that were already complicated by the absence of semantic context (Type 1 vs. Type 3) or incongruity (Type 5B vs. Type 6B), the absence of segmental information made less impact on the accuracy gap between the homeland and heritage groups.

5.2.4 With vs. without tonal information

Last but not least, this subsection discusses the impact of the presence or absence of tonal information—the variable of most interest in this study. In general, the lack of tonal information should make a task more difficult. Perhaps a more interesting question is how much difference the presence of context alone would make for each group in the absence of tonal information.

The first pair for comparison is Type 1 (normal words) and Type 2 (words with no tone). Their variable configuration is summarized in Table 5.16. As mentioned previously, Type 1 stimuli were tonally contrastive monosyllabic words. As there was no semantic context and only acoustic information was available, a large accuracy gap was expected between homeland and heritage speakers. For Type 2, the removal of tonal information from tonally contrastive monosyllabic words was expected to render the two groups closer to equal, in that both groups' accuracy would be near chance (25%). Therefore, the accuracy gap for Type 1 was expected to be larger than that of Type 2.

Results for Type 1 in Figure 5.10a show that compared with homeland speakers (M = 91.37%, SD = 28.08), heritage speakers (M = 69.22%, SD = 46.17) were indeed less able to identify a word from its tonally contrastive competitors in the absence of semantic context, $t(42.83) = 7.89$, $p < .001$. However, it was not the case that heritage speakers were completely unable to distinguish between tonally contrastive words in isolation. If they were unable to use tonal information at all, their performance in Type 1 would have been closer to chance (25%). As for Type 2, results in Figure 5.10b confirmed that it was an impossible task, and both groups' performance was close to chance. The mean accuracy of homeland (M = 30.05%, SD = 45.86) and heritage speakers (M = 28.97%, SD = 45.37) was very similar, $t(63.64) = 1.13$, $p = .26$. As predicted, the accuracy gaps of these two stimulus types were very different: the effect size was close to huge for Type 1 (d = 1.91) but only small for Type 2 (d = 0.27). This suggests that the absence of tonal information closed the accuracy gap between homeland and heritage speakers.

Table 5.16: Stimulus types with segmental information but no context

Type	Segment	Tone	Context	Congruity	Predicted result		
1	✓	✓	✗	not appl.	homeland	>>	heritage
2	✓	✗	✗	not appl.	homeland	=	heritage

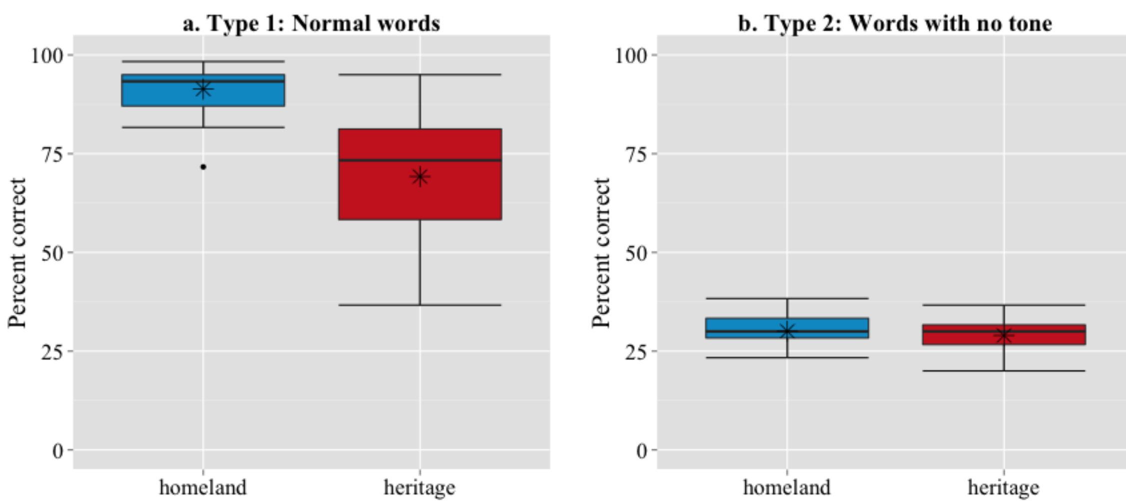


Figure 5.10: Comparison of stimulus types with segmental information but no context

The last pair for comparison is Type 5A (normal congruous sentences) and Type 4 (sentences with no tone). Their variable configuration is summarized in Table 5.17. As mentioned previously, Type 5A was expected to reflect the difference of homeland and heritage speakers' baseline listening proficiency in Cantonese. As for Type 4 stimuli, they were "toneless" sentences in the sense that the f0 of the entire sentence (both the carrier phrase and the target word at the end) was reset to 200 Hz, which was similar to the talker's mid level tone. The stimuli block only contained congruous sentences, and so correct answers were words that made sense in that carrier phrase. Given the hypothesis that the difference between homeland and heritage speakers was their ability to make use of tonal information, it was predicted that the removal of tonal information would render the two groups equal with regard to their performance in Type 4.

Table 5.17: Stimulus types with segments, context, and congruity

Type	Segment	Tone	Context	Congruity	Predicted result		
5A	✓	✓	✓	✓	homeland	$>$	heritage
4	✓	✗	✓	✓	homeland	$=$	heritage

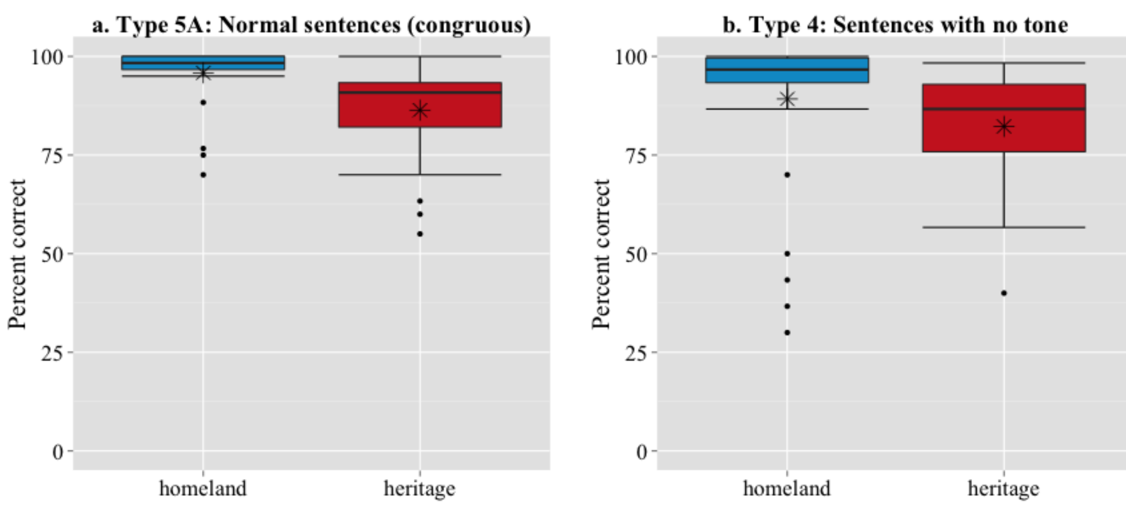


Figure 5.11: Comparison of stimulus types with segments, context, and congruity

For results of Type 5A in Figure 5.11a, the difference between homeland (M = 95.83%, SD = 19.99) and heritage speakers (M = 86.37%, SD = 34.32) was significant, $t(57.60)=3.99$, $p < .001$. However, for Type 4 in Figure 5.11b, the difference between homeland (M = 89.22%, SD = 31.03) and heritage speakers (M = 82.25%, SD = 38.21) was not significant, $t(60.14) = 1.71$, $p = .09$. The effect size was large for Type 5A ($d = 0.97$) but it was small for Type 4 ($d = 0.41$). Although the effect size for Type 4 was not as small as expected, the difference between Types 5A and 4 in terms of effect size met previous expectations.

Taken together, results in this subsection confirmed that the absence of tonal information minimized the accuracy gap between homeland and heritage speakers. As expected, results of Type 2 and Type 4 provide evidence that homeland speakers did not simply do better than heritage speakers across the board due to the difference of baseline proficiency. The crucial factor to determine the magnitude of difference between the two groups was the configuration of variables.

5.2.5 Interim summary

To sum up, the absence of tonal information brought homeland and heritage speakers closer to each other in terms of accuracy, but the absence of segmental information, semantic context or congruity widened the gap between the two populations. In general, results presented in this section support the hypothesis that heritage speakers are less able to distinguish tonally contrastive words compared with heritage speakers. More importantly, this difference between the two populations was not static across stimulus types; rather, the magnitude of difference can be predicted by the type of information available, which is summarized in Table 5.18. When tone was the only available type of information, the accuracy gap was the largest, as in Type 3. The presence of segmental information played a role to make the accuracy gap smaller (e.g. Type 5A vs. Type 6A), but the presence of semantic context and congruity made an even bigger impact on minimizing the accuracy gap (Type 1 vs. Type 5A).

Table 5.18: Summary of stimulus types and variables, arranged by effect size in form of Cohen's d (smallest to largest)

Type	Description of stimulus type	Seg.	T.	Ct.	Cg.	Cohen's d
2	Words with no tone	✓	✗	✗	not appl.	0.27
4	Sentences with no tone	✓	✗	✓	✓	0.41
5A	Normal sentences (congr.)	✓	✓	✓	✓	0.97
5B	Normal sentences (incongr.)	✓	✓	✓	✗	1.22
6A	Last word of the sentence has no segments (congruous)	✗	✓	✓	✓	1.28
6B	Last word of the sentence has no segments (incongruous)	✗	✓	✓	✗	1.37
1	Normal words	✓	✓	✗	not appl.	1.91
3	Words with no segments	✗	✓	✗	not appl.	2.00

Seg. = Segment, T. = Tone, Ct. = Context, Cg. = Congruity

Cohen's d thresholds: 0.20=small, 0.50=medium, 0.80=large, 1.20=very large, 2.00=huge

5.3 Response to Research Question 2: Confusion patterns

This section addresses the second research question of this dissertation: do homeland and heritage speakers exhibit similar confusion patterns with respect to lexical tone perception? The null and alternative hypotheses are listed as follows:

> **Research Question 2:**
> Do homeland and heritage speakers exhibit similar confusion patterns with respect to lexical tone perception?
> H_0: There is no difference between homeland and heritage speakers in terms of their confusion patterns.
> H_1: Homeland and heritage speakers have different confusion patterns.

Previously in Section 3.6, it was hypothesized that heritage speakers would experience more confusion than homeland speakers when asked to identify a word from a tonally contrastive set. Therefore, it was anticipated that H_0 would be rejected in favour of H_1.

In order to measure differences in terms of confusion patterns, data will be presented as confusion matrices in this section. Section 5.3.1 is an introduction to confusion matrices as a type of data visualization. It is then followed by Section 5.3.2, which explains the Mantel test for comparing global similarity of matrices (Mantel, 1967). Data analysis for the current study will be presented in Section 5.3.3, followed by an interim summary in Section 5.3.4.

5.3.1 How to read a confusion matrix

This subsection is an introduction to confusion matrices. Readers who are familiar with this type of data visualization may skip and proceed to Section 5.3.2 on the statistical test used for comparing confusion matrices.

A confusion matrix is a two-dimensional table with two axes, as in Figure 5.12. The horizontal axis represents the target category, which in this study means the intended tone. The vertical axis represents which category a stimulus was perceived as, which in this case means participants' responses. Since there are six tones, the matrix has 36 (6*6) cells, each of which has a "nickname" in the form of [x,y]. For example, [5,2] is the cell that represents "the stimulus was T5; participants answered T2", and [6,6] means "the stimulus was T6; participants answered T6". Numbers printed on the cells indicate percentage of the target category instances that were classified with the given row label. If the percentage is high, the cell is coloured in a proportionally darker shade. Percentages of each column always add up to 100.

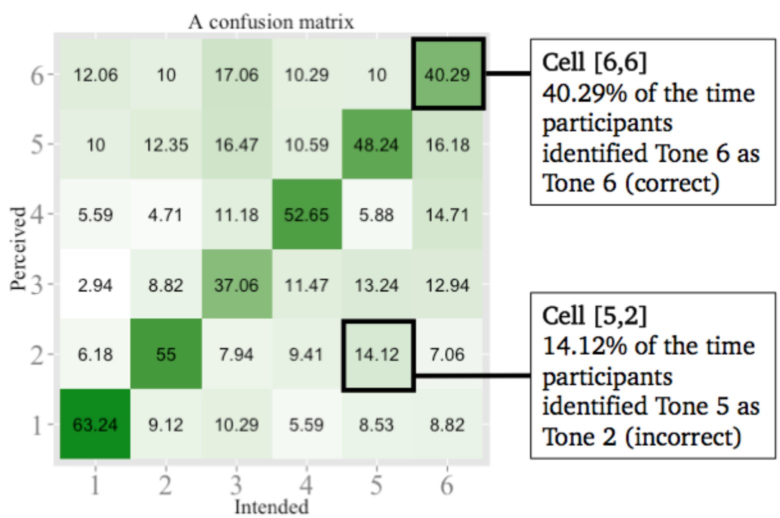


Figure 5.12: How to interpret a confusion matrix

Four hypothetical situations and their respective patterns are presented as follows. First, if participants have an accuracy rate of 100%, their confusion pattern will look like Figure 5.13a. The cells [1,1], [2,2], [3,3], [4,4], [5,5], and [6,6] are instances when participants identify the correct tone. These cells form a diagonal from the lower left corner to the upper right corner[2]. Therefore, a dark diagonal is a sign of high accuracy.

[2]In the literature, confusion matrices often have diagonals from the upper left to the lower right corner. Those in this section, however, have diagonals from the lower left to the upper right corner, so that tone numbers are arranged in a more intuitive order: numbers on the x-axis go from left to right, while those on the y-axis go from bottom to top.

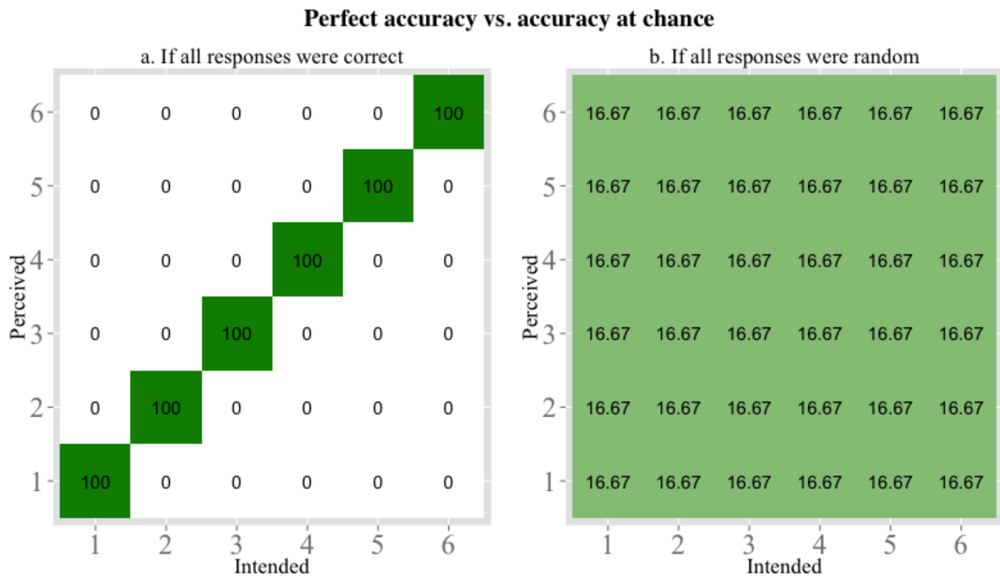


Figure 5.13: Confusion matrices showing perfect accuracy and accuracy at chance respectively

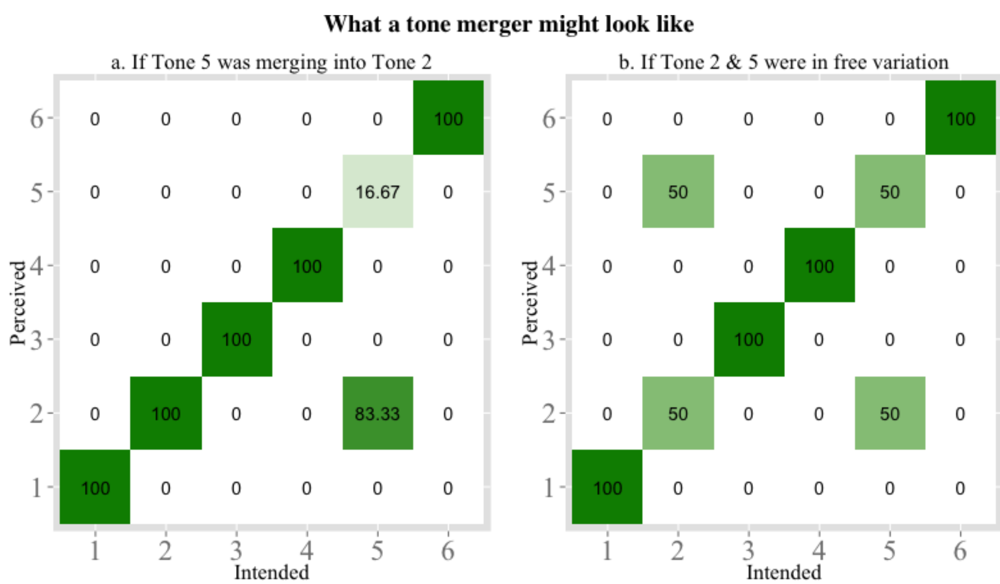


Figure 5.14: Confusion matrices showing two possibilities of T2-T5 merger

The second hypothetical situation is at the other end of the extreme, namely random responses. Figure 5.13b has no clear diagonal at all, and all cells are equally dark. It means that each intended tone has been perceived as every tone equally frequently. Therefore, an evenly shaded matrix indicates random responses.

The third and fourth situations pertain to the tone merger phenomenon discussed in Section 3.3.2. Figure 5.14 shows two possible ways of merging T2 (high rising) and T5 (low rising). The first way is merging one category into another existing category, which involves confusion in only one direction. In Figure 5.14a, the cell [5,2] has a much darker shade than [5,5], which means T5 is perceived as T2 most of the time (83.33%). However, T2 is still consistently perceived as T2 (see [2,2]), and is never identified as T5 (see [2,5]). In other words, T5 is merging into T2 in perception.

The second kind of merger is merging two categories into one new category, which involves confusion in two directions. In Figure 5.14b, the cells [2,2], [2,5], [5,2], and [5,5] have the same shade. T2 is perceived as T2 or T5 at equal chance, and similarly T5 is perceived as T2 or T5 at equal chance. In this case T2 and T5 are merging into a new category (a general rising tone, for example) in perception. To sum up, both types of merger reduce the number of contrastive rising tones from two to one: in the first case the rising tone that remains in the inventory would be the high rising tone, while in the second case it would be a general rising tone.

Figure 5.14a and Figure 5.14b look more similar to each other than Figure 5.13a and Figure 5.13b are. When more pairs with varying degree of similarity are compared, however, it will be a question of how similarity can be measured and ranked. The next subsection is going to explain how similarity can be quantified statistically.

5.3.2 The Mantel test for comparing global similarity of matrices

This subsection explains the Mantel test (Mantel, 1967) to be used for assessing the strength of correlation between two distance matrices.

Readers who are familiar with this statistical method may skip and proceed to Section 5.3.3 for results.

Consider a basic question: what does *similarity* mean for confusion matrices? Figure 5.15 and Figure 5.16 present two pairs of toy matrices, which show the confusion patterns of six hypothetical categories α, β, γ, δ, ϵ, and ζ. Each of the two pairs can be argued as more similar than the other due to different reasons. On the one hand, one can argue that Toy Matrices A and C in Figure 5.16 are more similar than A and B are in Figure 5.15, because A and C have the same number of shaded non-diagonal cells (which is two), and the percentages shown on these non-diagonal cells are the same as well (which is 16.67). On the other hand, one can argue that Toy Matrices A and B are more similar than A and C are. Even though their number of shaded non-diagonal cells is different (A has two but B has three), these cells overlap in terms of distribution. The cells [β,ϵ] and [ϵ,β] are shaded in both matrices. It is noteworthy that these cells are "mirrored" versions of each other, a sign of mutual confusion. However in Toy Matrix C, the cells [α,ζ] and [δ,α] are shaded, while their counterparts in A are not. The sets of categories being confused do not overlap, and mutual confusion is only found in Toy Matrix A but not C. In this respect A and C are quite different.

The Mantel test for comparing global similarity of matrices shares the same rationale as the latter argument above—it is crucial to recognize that cells in a confusion matrix are not independent from each other, in that any given cell and its "mirrored twin" inform us about confusion of the same set of categories. The output of the test is known as the Mantel *r* statistic, which indicates the strength level of correlation coefficients. A guide to interpreting this value is provided in Table 5.19. A Mantel *r* statistic of 0 and $\pm$ 1 indicate no correlation and perfect correlation respectively. Numbers between 0 and $\pm$ 1 are intermediate levels from weakly correlated to very strongly correlated. The Mantel test was conducted for the Toy Matrices in Figure 5.15 and Figure 5.16. Results suggest that Toy Matrices A and B are very strongly correlated (r = .92, p = .02), while A and C are only modestly correlated (r = .23, p = .16). Therefore, "similarity" of confusion matrices as measured by the Mantel test does not simply pertain to the total number

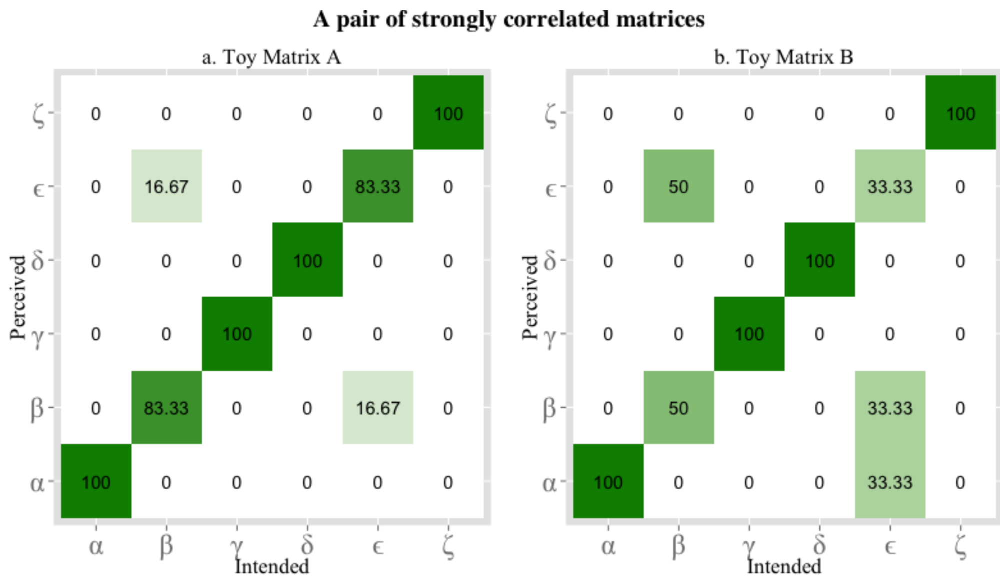


Figure 5.15: Toy matrices demonstrating a strong correlation

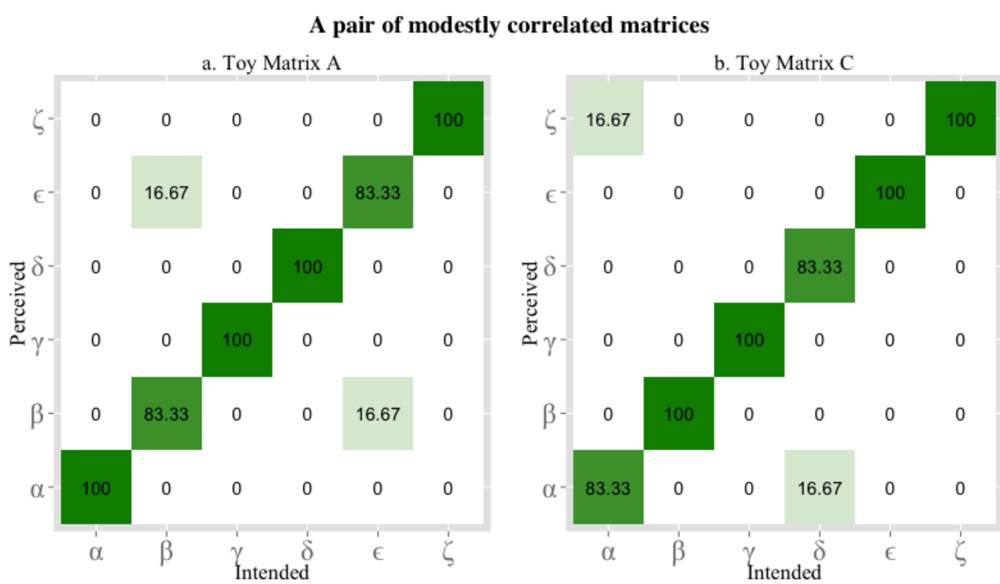


Figure 5.16: Toy matrices demonstrating a modest correlation

of shaded non-diagonal cells or the total percentages of those cells; what is more crucial is the relationship between these shaded cells and the sets of categories being confused.

Table 5.19: Interpretation of the Mantel *r* statistic (Mantel, 1967)

Mantel *r* statistic			Correlation
	$\pm$ 1		perfect
$\pm$0.8	to	$\pm$0.9	very strong
$\pm$0.5	to	$\pm$0.8	strong
$\pm$0.3	to	$\pm$0.5	moderate
$\pm$0.1	to	$\pm$0.3	modest
	below $\pm$0.1		weak
	0		none

Information presented so far should be adequate for readers to interpret results presented in Section 5.3.3. Readers who are interested in how to calculate the Mantel *r* statistic and significance level may keep reading this subsection, which dives into details of the steps. Table 5.20 is an overview of all procedures, which were implemented in R (R Core Team, 2013) with the `vegan` package (Oksanen et al., 2018). Among the seven steps, Step 6 (compute correlation coefficient) and Step 7 (calculate significance level) are directly relevant to data analysis. As Step 6 requires distance matrices as input, Steps 2–5 are intermediate steps with the purpose of transforming raw counts into the right input for data analysis; the output of each step becomes the input being fed to the next step.

To illustrate the steps, a walk-through of the transformation of Toy Matrix D from raw counts (Table 5.21) to distance (Table 5.25) is presented as follows. First, raw counts were obtained and presented in Table 5.21. As usual the x- and y-axes represent the intended and perceived categories respectively. In this table some cells had a zero. These zeros would be a potential problem for Step 4 (convert proportion to similarity). To avoid potential problems caused by zeros, the next step served the purpose of replacing zeros with a non-zero value.

At Step 2, a smoothing technique (Witten & Bell, 1991) was employed

Table 5.20: Procedures to implement the Mantel test, adapted from Tang (2015)

Step	Description	Method
1	Obtain raw counts of responses	For each intended category, count how many times it was perceived as each category
2	"Smooth" the matrix	Witten & Bell (1991)
3	Convert counts to proportion	For each cell, the value is divided by the total number of times that intended category was presented
4	Convert proportion to similarity	Shepard (1972)
5	Convert similarity to distance	Shepard (1972)
6	Compute correlation coefficient	Kendall's tau
7	Calculate significance level	10,000 permutations

Table 5.21: Toy Matrix D at Step 1 (raw counts)

ζ	0	0	0	0	0	100
ϵ	0	50	0	0	33	0
δ	0	0	0	100	0	0
γ	0	0	100	0	0	0
β	0	50	0	0	33	0
α	100	0	0	0	34	0
	α	β	γ	δ	ϵ	ζ

to decrease the occurrences of non-zero counts (a process known as "discounting"), and redistribute them to cells that had zeros (a process known as "backoff"). Compare the matrix before (Table 5.21) and after smoothing (Table 5.22). Cells that used to have a value of 100 are now 98.522, while cells that used to have a value of 0 are now 0.328. Here is how this was done: first, for non-zero cells, the discounting formula in (18) was applied. What this formula does is, for every non-zero cell, take the original value and multiply it by a fraction. This fraction's numerator is the total number of responses in the matrix, and its denominator is the sum of the total number of responses and the number of non-zero cells in a matrix. In other words, its numerator will always be smaller than its denominator. As a result, the product after multiplication by this fraction will always be smaller than the original value, and this is why the process is called "discounting". For example, in Table 5.21, the number of non-zero cells was 9, the total number of responses was 600, and the cell [α,α] had the value of 100. Therefore, the new value for this cell after discounting would be 100 * (600/(600+9)), which yields 98.522, as in Table 5.22.

Table 5.22: Toy Matrix D after Step 2 (smoothing)

ζ	0.328	0.328	0.328	0.328	0.328	98.522
ϵ	0.328	49.261	0.328	0.328	32.512	0.328
δ	0.328	0.328	0.328	98.522	0.328	0.328
γ	0.328	0.328	98.522	0.328	0.328	0.328
β	0.328	49.261	0.328	0.328	32.512	0.328
α	98.522	0.328	0.328	0.328	33.498	0.328
	α	β	γ	δ	ϵ	ζ

(18) Formula for discounting non-zero responses (Witten & Bell, 1991)

$$V_{\text{!zero}} = R \times \frac{N_{\text{T}}}{N_{\text{T}} + N_{\text{!zero}}}$$

- $V_{\text{!zero}}$ is the new value for a non-zero cell after smoothing
- R is the original raw count in a non-zero cell before smoothing
- N_{T} is the total number of responses
- $N_{\text{!zero}}$ is the number of non-zero cells

Second, for zero cells, the backoff formula in (19) was applied. What this formula does is, take the ratio of non-zero to zero cells in the matrix, and then multiply this ratio by the same fraction described above (the numerator is the total number of responses in the matrix, and the denominator is the sum of the total number of responses and the number of non-zero cells in a matrix). For example, the total number of zero cells was 27 in Table 5.21. The new value for the cell [α,β] after backoff would be (9/27) * (600/(600+9)), which yields 0.328. Therefore, cells that used to be 0 in Table 5.21 have become 0.328 in Table 5.22. The matrix has been "smoothed" and is ready to be fed to the next step.

(19) Backoff formula for zero responses (Witten & Bell, 1991)

$$V_{\text{zero}} = \frac{N_{\text{!zero}}}{N_{\text{zero}}} \times \frac{N_{\text{T}}}{N_{\text{T}} + N_{\text{!zero}}}$$

- V_{zero} is the new value for a zero cell after smoothing
- $N_{\text{!zero}}$ is the number of non-zero cells
- N_{zero} is the number of zero cells
- N_{T} is the total number of responses

Step 3, namely converting smoothed values to proportions, is quite straightforward. For each cell, the smoothed value was divided by the number of times the respective intended category was presented (i.e. add up the values of that column). Take the cell [α,α] as an example: its value was 98.522 after Step 2 in Table 5.22, and the sum of the whole α column was 100. Therefore, the new value after Step 3 would be 98.522 divided by 100, hence the result was 0.985 in Table 5.23. Similarly, the cell [α,β] used to be 0.328; after Step 3 it has become 0.003. Now that the values have become proportion, they are ready to be converted to similarity.

Table 5.23: Toy Matrix D after Step 3 (proportion)

ζ	0.003	0.003	0.003	0.003	0.003	0.985
ϵ	0.003	0.493	0.003	0.003	0.327	0.003
δ	0.003	0.003	0.003	0.985	0.003	0.003
γ	0.003	0.003	0.985	0.003	0.003	0.003
β	0.003	0.493	0.003	0.003	0.327	0.003
α	0.985	0.003	0.003	0.003	0.337	0.003
	α	β	γ	δ	ϵ	ζ

The purpose of Step 4 is to obtain a value to quantify the perceived similarity of two categories. The metric proposed by Shepard (1972) in (20) was adopted, which gives an output value from 0 (not similar at all) to 1 (extremely similar). Conceptually the formula can be understood as a ratio of how often two categories are confused with each other, to how often these two categories are correctly perceived. If two categories are always perceived correctly with no confusion at all, the ratio would be 0:1, or 0/1 when represented as a fraction, which gives the output value 0. This means these categories are perceived as “not similar at all”. However, if two categories are confused with each other half of the time, and perceived correctly half of the time, the ratio would be 0.5:0.5, or 0.5/0.5, which yields 1. This result means the two categories are perceived as “extremely similar”. Note that if two categories are never perceived correctly as the intended category, the denominator of the fraction in (20) will be zero, which means the result will always be “undefined”, and the rest of the steps can never be taken. This explains why the zeros in Table 5.21 must be avoided by the smoothing technique at Step 2.

To illustrate how this formula works, consider the categories β and ϵ in Table 5.23. The numerator of the fraction would be the sum of [β,ϵ] (0.493) and [ϵ,β] (0.327), which yields 0.82. The denominator of the fraction would be the sum of [β,β] (0.493) and [ϵ,ϵ] (0.327), which is also 0.82. The fact that the numerator and the denominator are identical suggests that the categories β and ϵ were confused with each other as often as they were perceived correctly. The output of the formula is therefore 0.82/0.82, resulting in 1, as in Table 5.24. In other words, β and ϵ were perceived as "extremely similar".

(20) Formula for converting proportion to similarity (Shepard, 1972)

$$S_{xy} = \frac{p_{(x,y)} + p_{(y,x)}}{p_{(x,x)} + p_{(y,y)}}$$

- S_{xy} is the similarity value between two categories x and y.
- $p_{(x,y)}$ is the proportion of times that x was incorrectly perceived as y.
- $p_{(y,x)}$ is the proportion of times that y was incorrectly perceived as x.
- $p_{(x,x)}$ is the proportion of times that x was correctly perceived as x.
- $p_{(y,y)}$ is the proportion of times that y was correctly perceived as y.

Table 5.24: Toy Matrix D after Step 4 (similarity)

ζ	0.003	0.004	0.003	0.003	0.005	1
ϵ	0.259	1	0.005	0.005	1	0.005
δ	0.003	0.004	0.003	1	0.005	0.003
γ	0.003	0.004	1	0.003	0.005	0.003
β	0.004	1	0.004	0.004	1	0.004
α	1	0.004	0.003	0.003	0.259	0.003
	α	β	γ	δ	ϵ	ζ

This formula in (20) has two important consequences. First, the similarity between β and ϵ will always be the same as the similarity between ϵ and β. At Step 3 (proportion), the cell [β,ϵ] and its "mirrored twin" [ϵ,β] had different values, as in Table 5.23. However, after Step 4, the values of both cells have become 1, as in Table 5.24. Another consequence is that after conversion to similarity, values in the diagonal cells (such as

[α,α] and [β,β]) will always be 1, regardless of how different they were from each other before conversion. This is because the numerator and the denominator for the formula will always be identical for diagonal cells. For example, to calculate the similarity between β and β, the numerator will be ($p_{(\beta,\beta)}+p_{(\beta,\beta)}$) and the denominator will also be ($p_{(\beta,\beta)}+p_{(\beta,\beta)}$); the only possible result is 1.

The purpose of Step 5 is to convert similarity to distance. Similarity and distance are like two sides of a coin: if two categories are extremely similar, the perceptual distance between them must be very small or even non-existent. Therefore, a similarity value of 1 (extremely similar) would be equivalent to a perceptual distance of 0 (no perceptual distance). To implement the conversion, Shepard's law (Shepard, 1972) was used, as in (21). The idea of this formula is turn the similarity value between two categories into its negative logarithm. As a result, a similarity value of 1 becomes a distance value of 0, as in Table 5.25. A relatively small similarity value (e.g. 0.003 in Table 5.24) becomes a relatively large distance value (e.g. 5.809 in Table 5.25). The consequences mentioned in Step 4 still hold after Step 5: all diagonal cells will always have a distance value of 0, and any given cell and its "mirrored twin" will always have identical values. This fact about "mirrored twins" is especially important for the next step, because it reflects the assumption that cells in a confusion matrix are not independent from each other. A distance matrix like Table 5.25 is now an appropriate input for correlation analysis via the Mantel test.

(21) Formula for converting similarity to distance (Shepard, 1972)

$$D_{xy} = -(\log S_{xy})$$

- D_{xy} is the distance value between two categories x and y.
- S_{xy} is the similarity value between two categories x and y.

Table 5.25: Toy Matrix D after Step 5 (distance)

ζ	5.809	5.521	5.809	5.809	5.298	0
ϵ	1.351	0	5.298	5.298	0	5.298
δ	5.809	5.521	5.809	0	5.298	5.809
γ	5.809	5.521	0	5.809	5.298	5.809
β	5.521	0	5.521	5.521	0	5.521
α	0	5.521	5.809	5.809	1.351	5.809
	α	β	γ	δ	ϵ	ζ

The last two steps are procedures of the Mantel test itself. It compares two distance matrices and computes their correlation coefficient (known as the Mantel *r* statistic) and significance level (the *p* value). To compute the correlation coefficient (Step 6), Kendall's tau[3] was used. What the formula in (22) does is, take the difference between the number of concordant and discordant pairs, and divide this difference by the total number of pair combinations. This gives an output value of 0 (no correlation) to ± 1 (perfect correlation), as listed in Table 5.19.

(22) Formula for computing the correlation coefficient of two distance matrices

$$r_\tau = \frac{(\textit{number of concordant pairs}) - (\textit{number of discordant pairs})}{n(n-1)/2}$$

- r_τ is the correlation coefficient (Kendall's tau).
- n is the number of observations.

To obtain a significance level (Step 7), 10,000 permutations were performed. A permutation means shuffling the rows and columns of the matrices, and then recomputing the correlation coefficient of the shuffled matrices. The purpose is to see if the recomputed correlation coefficient

[3]To avoid potential errors made by using Spearman's rho, Kendall's tau was used instead of Spearman's rho. For more details, see the discussion in Tang (2015).

would be larger than the original coefficient. If this process is repeated 10,000 times[4], and most of the time the recomputed coefficient of shuffled matrices has a greater value than that of the unshuffled matrices, it would suggest that any observed correlation could be due to chance. In this case a high *p* value will be obtained, which means the correlation is not statistically significant. However, if we can only find a few instances in which the recomputed coefficient is larger than the original one, it would suggest that the observed correlation in unshuffled matrices is less likely to have happened by chance. In this case a low *p* value will be obtained, which means the the correlation is statistically significant.

In the next subsection, confusion matrices of the actual experiment will be discussed. Although distance matrices are the input for the Mantel test, they are not suitable for visualizing the directionality of confusion, because cells that are mirrored twins will always have the same value. Therefore, in the next subsection, values in the cells will indicate the percentage of responses, which can better display (a)symmetrical patterns of potential tone mergers. Although the transformation of each matrix from raw counts to distance will not be shown, readers should bear in mind that all matrices in the next subsection have gone through the seven steps in order to compute the Mantel *r* statistic and significance level. Therefore matrices that seem similar to the human eye (e.g. those in Figure 5.16) may be weakly correlated according to the Mantel test.

5.3.3 Comparison of confusion patterns

Although boxplots in Section 5.2 and confusion matrices in the current subsection were based on responses of the same group of participants from the same tasks, the focus of this subsection is different. The purpose of Section 5.2 was to compare homeland and heritage speakers' average percentage of *correct* responses. However, this subsection is mainly about

[4]In theory the number of permutations should be as large as possible so as to lower the uncertainty and to enhance the reliability of the result. As discussed in Tang (2015), 10,000 permutations is a good number as it is larger than the minimum number of permutations (1,000) and at the same time the computation time required is reasonable.

their *incorrect* responses: were the errors random? Were there patterns in the errors? If there were patterns, were they different between the two groups? To answer these questions (which are iterations of Research Question 2), incorrect responses represented by non-diagonal cells will be discussed in more detail.

As an overview, Table 5.26 summarizes results of the Mantel test for comparing global similarity of homeland and heritage speakers' confusion matrices. The eight stimulus types were arranged by the Mantel *r* statistic from the largest to the smallest. Overall, the *r* values were between 0.30 and 0.81. In other words, all homeland-heritage pairs fell between "modestly correlated" and "strongly correlated" (see Table 5.19). This suggests that the two groups' confusion matrices were quite similar in general. However, their degree of similarity varied across stimulus types. When the stimuli were incongruous sentences (Types 5B and 6B), the two groups had the most strongly correlated confusion matrices with *r* values above 0.60. Monosyllabic words (Type 1, 2, 3) came next with *r* values around 0.50, showing moderate to strong correlations. Lastly, stimuli that were congruous sentences (Types 4, 5A, 6A) had *r* values close to 0.30, showing modest to moderate correlations. Taken together, stimulus types representing different variable configurations fell into different ranges on the scale of correlation strength.

While the Mantel *r* statistic and *p* values were calculated based on distance scores, the cells in all matrices presented below will show proportions in percentages. Recall from Section 5.3.2 that the conversion from proportions to distances erases any asymmetry in the direction of confusion. Presenting proportions instead of distances allows the comparison of homeland and heritage speakers' directionality of confusion, which is crucial for answering questions raised in Chapter 3 regarding trends of tone merger.

The rest of this subsection will compare confusion matrices of homeland and heritage speakers for each stimulus type. The eight homeland-heritage pairs will be discussed in the same order shown in Table 5.26. In other words, the most correlated pair will be discussed first, while the least

Table 5.26: Summary of Mantel test results comparing global similarity of homeland and heritage speakers' confusion matrices; rows were arranged by *r* values (largest to smallest)

Type	Description of stimulus type	Seg.	T.	Ct.	Cg.	*r*	*p*
6B	Last word of the sentence has no segments (incongruous)	✗	✓	✓	✗	.81	.001
5B	Normal sentences (incongr.)	✓	✓	✓	✗	.64	.001
3	Words with no segments	✗	✓	✗	n.a.	.58	.007
1	Normal words	✓	✓	✗	n.a.	.50	.013
2	Words with no tone	✓	✗	✗	n.a.	.49	.015
5A	Normal sentences (congr.)	✓	✓	✓	✓	.33	.078
6A	Last word of the sentence has no segments (congruous)	✗	✓	✓	✓	.33	.046
4	Sentences with no tone	✓	✗	✓	✓	.30	.057

Seg. = Segment, T. = Tone, Ct. = Context, Cg. = Congruity

correlated pair will be discussed last. For each pair of matrices, the left one represents responses of homeland speakers, while the right one represents those of heritage speakers, as in Figure 5.17a and Figure 5.17b respectively. When perusing the matrices, readers may find it useful to bookmark Table 1.6 and Figure 1.4 on page 16 for a quick review on tone numbers and their respective tone numerals or contours. Following this subsection, an interim summary in Section 5.3.4 will discuss possible interpretations of the observed patterns.

The two most similar pairs of matrices were Type 6B (the last word of the incongruous sentence has no segments) and 5B (normal, incongruous sentences), which are represented by Figure 5.17 and Figure 5.18 respectively. The Mantel *r* statistic was 0.81 for Type 6B (p = .001) and 0.64 for Type 5B (p = .001), which is a sign of strong correlation between the homeland and heritage matrices. In both Type 6B (Figure 5.17) and 5B (Figure 5.18), non-diagonal cells were not evenly shaded, which suggests that errors were not random. For both homeland and heritage speakers, the same set of *paired* cells, namely [6,3]&[3,6] and [5,2]&[2,5], stood out as relatively darker than other non-diagonal cells in the same matrix. In other

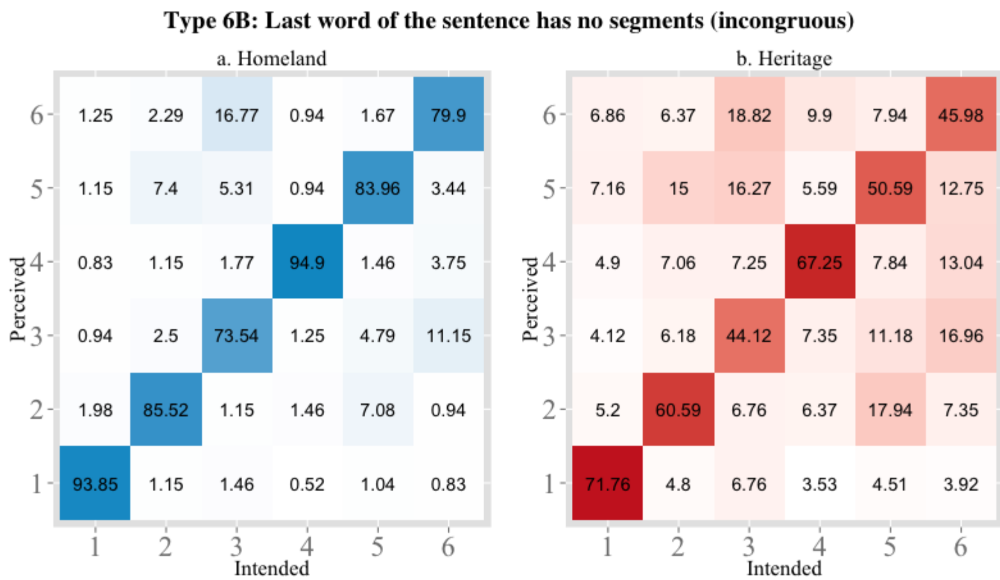


Figure 5.17: Confusion patterns of homeland and heritage speakers for Type 6B stimuli

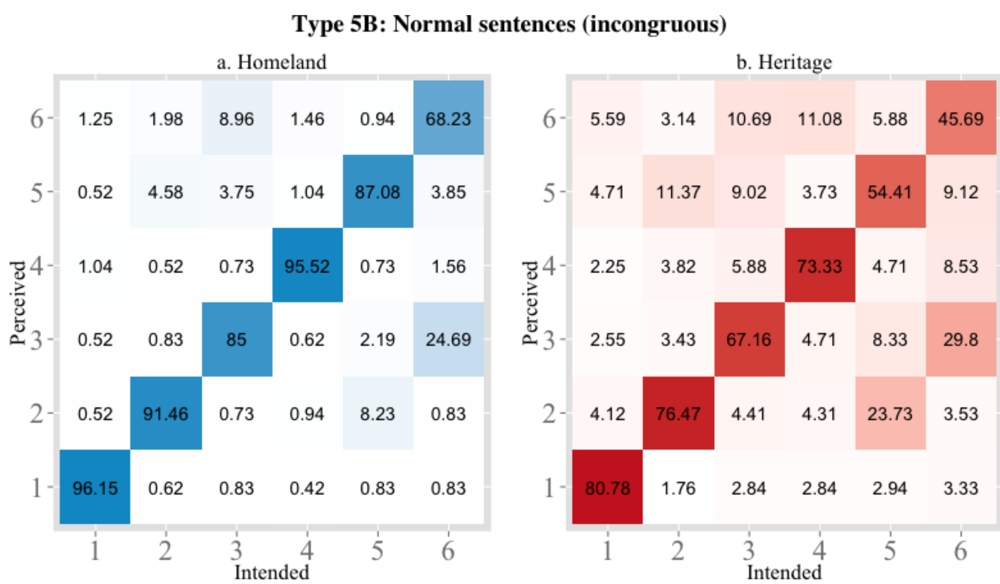


Figure 5.18: Confusion patterns of homeland and heritage speakers for Type 5B stimuli

words, when the stimuli were incongruous sentences, both populations found T2-T5 (high rising and low rising) and T3-T6 (mid level and low level) relatively more confusable than other tone pairs, possibly due to their similar pitch contours. For each of the cells that stood out (e.g. [5,2] in Figure 5.18a and Figure 5.18b), the one in heritage is always darker than its counterpart in homeland, which shows that confusion between these tones happened more often for the heritage group. It can be concluded that homeland and heritage speakers shared similar *types* of errors, but they differed mostly by the *quantity* of such errors.

A difference between Type 5B and 6B was that patterns of the latter were more symmetrical than those of the former. In Type 5B (Figure 5.18), the cell [6,3] was a lot darker than [3,6] (homeland: 24.69% vs 8.96%; heritage: 29.8% vs 10.69%), and [5,2] was much darker than [2,5] (homeland: 8.23% vs 4.58%; heritage: 23.73% vs 11.37%). The asymmetrical patterns suggest that confusion took place in one direction (e.g. T5 heard as T2) more often than the other (e.g. T2 heard as T5). However, in Type 6B (Figure 5.17), [3,6] and [6,3] had a similar shade, and this was true for both populations (homeland: 16.77% vs 11.15%; heritage: 18.82% vs 16.96%). The same can be observed for [2,5] and [5,2] (homeland: 7.40% vs 7.08%; heritage 15% vs 17.94%). The symmetry indicates that confusion took place in both directions: T2 was heard as T5, and T5 was also heard as T2. A possible reason is that as segmental information was taken away from Type 6B stimuli, the task became more difficult than Type 5B, which led to more mutual confusion for the tone pairs that were already confusable even when segmental information was available.

A qualitative difference between homeland and heritage speakers was the confusion between T4 (low falling) and T6 (low level). In Figure 5.17b (heritage), the cell [6,4] was quite dark (13.04%) compared with other non-diagonal cells, but its counterpart in Figure 5.17a (homeland) was not (3.75%). Similarly in Figure 5.18b (heritage), the cell [4,6] was shaded (11.08%), but the same cell in Figure 5.18a (homeland) was not (1.46%). This suggests that T4-T6 confusion was unique to heritage speakers.

The next most correlated pair was Type 3 (word with no segments). Its Mantel *r* statistic was 0.58 ($p = .007$), which indicates a strong correlation. On the whole, even though the diagonal in Figure 5.19b (heritage) was visible, its non-diagonal cells were shaded relatively evenly compared with Figure 5.19a (homeland). These features make Figure 5.19b (heritage) look like a blend of the two sample matrices for perfect accuracy and random answers in Figure 5.13. This lack of obvious patterns in non-diagonal cells suggests that when the only available cue was tonal information, heritage speakers—who were hypothesized to be struggling with using this cue—had little to rely on to make a judgment; therefore, the answer had to be picked more or less randomly. As for homeland speakers, they exhibited T3-T6 confusion again, but this time the direction of confusion was different. In Figure 5.19a, the cell [3,6] stood out as a darker non-diagonal cell (26.76%) compared with [6,3] (6.18%). In other words, T3 was heard as T6 more often. This was different from Type 5B, in which the other direction of confusion was observed more often.

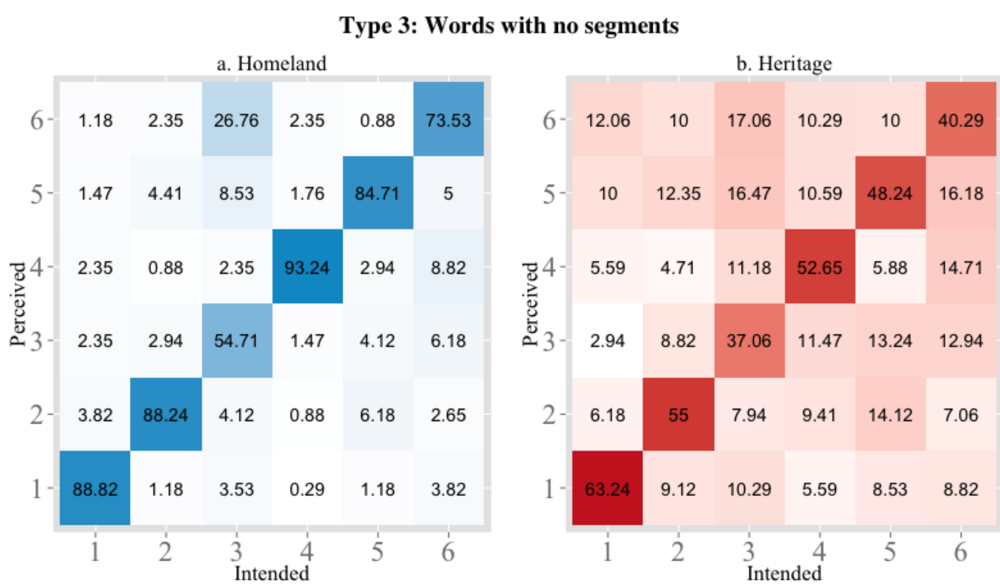


Figure 5.19: Confusion patterns of homeland and heritage speakers for Type 3 stimuli

Type 1 (normal words) matrices had a Mantel r statistic of 0.50 (p = .013), which indicates that the correlation was between moderate and strong. Some recurring patterns from previous stimulus types can be found. The first recurring pattern was T3-T6 (mid level and low level) confusion. In Figure 5.20a (homeland), [6,3] and [3,6] stood out as two of the darkest non-diagonal cells (11.47% vs 10.00%). However, in Figure 5.20b (heritage), [6,3] was darker than [3,6] (21.76% vs 11.18%). This suggests that confusion took place in both directions among homeland speakers, which was similar to what happened for Type 6B but different from Types 5B and 3. However, among heritage speakers, T3-T6 confusion took place in one direction more often than the other, which was similar to what happened in Type 5B but different from Types 6B and 3. Thus far, the direction of T3-T6 confusion has been inconsistent.

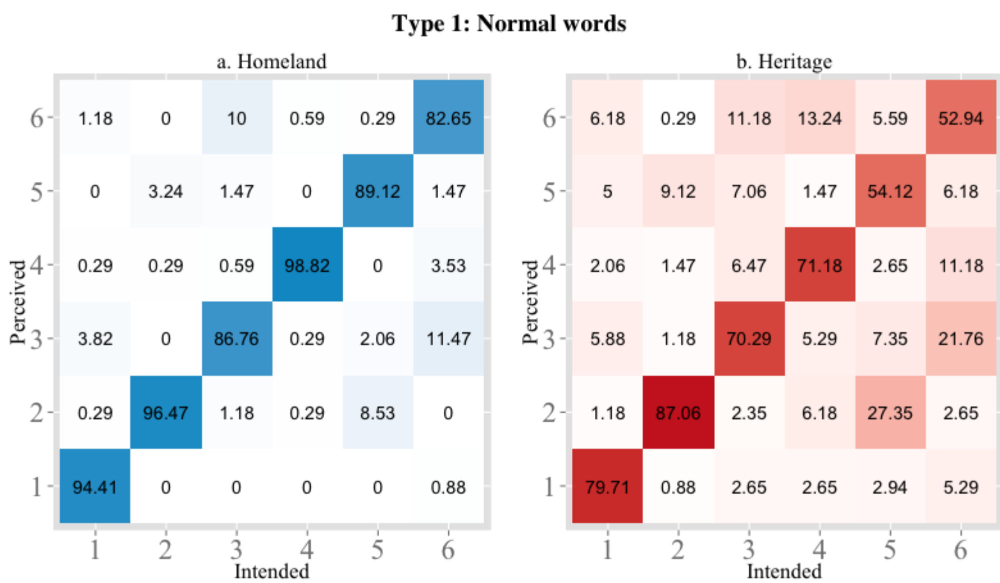


Figure 5.20: Confusion patterns of homeland and heritage speakers for Type 1 stimuli

The second recurring pattern was T2-T5 (high rising and low rising) confusion. In both Figure 5.20a and Figure 5.20b, the cells [5,2] and [2,5] show that confusion took place mostly in one direction (T5 heard as T2), and this was true for both homeland and heritage speakers. Thus far, directions of T2-T5 confusion have been more consistent across stimulus types than those of T3-T6 confusion.

In the discussion of Types 6B and 5B, it was mentioned that T4-T6 (low falling and low level) confusion was unique to heritage speakers. This was true for patterns observed in Type 1 too. The cells [6,4] and [4,6] were shaded in the heritage matrix in Figure 5.20b (11.18% and 13.24%) but not so much in the homeland matrix in Figure 5.20a (3.53% and 0.59%). This suggests that heritage speakers had T4-T6 confusion across a variety of stimulus types including both words and sentences.

Type 2 (words with no tone) matrices were moderately correlated (r = .49, p = .015). Recall from Section 4.4.1.1 that the f0 of Type 2 was reset to a uniform pitch (200 Hz) close to the talker's T3 (mid level tone). These stimuli had "no tone" in the sense that the pitch height and pitch contour of the intended tone, which are important perceptual correlates of Cantonese tones, were made unavailable. Since the semantic context was not provided, the task was predicted to be impossible for homeland and heritage speakers, and so no clear diagonals were expected. As seen in Figure 5.21, there were indeed no clear diagonals for both groups of participants, which met previous expectations. Although these two matrices may look extremely similar to the human eye, none of the shaded cells were paired like [6,3] and [3,6] in Figure 5.17, which explains why Type 2's Mantel statistic was lower than Type 6B's.

One obvious pattern in Type 2 (Figure 5.21) was that both groups of speakers identified a high proportion of the stimuli as T3 (mid level), as indicated by the darker shading in the whole T3 row. This indicates a general T3 bias, presumably due to the stimuli's uniform pitch at 200 Hz, which was close to the f0 of this talker's T3. The cell [3,3] showed the highest accuracy as the pitch of the stimuli after manipulation was similar to that of the intended tone. However, it was not true that the participants had a T3 bias for *all* Type 2 stimuli. In both the homeland and heritage matrices, the top row for T6 was somewhat shaded as well, which means the stimuli were sometimes identified as T6 (low level). This was not surprising, given

Figure 5.21: Confusion patterns of homeland and heritage speakers for Type 2 stimuli

that both groups exhibited T3-T6 confusion in other stimulus types as well. The last observation was that the cell [4,4] in both matrices was shaded, which suggests that in some instances both groups of participants were able to identify T4 (low falling) despite not having access to tonal information. This was possibly due to the availability of non-tonal cues such as occasional creaky voice in the stimuli. Overall, confusion often went in one direction (i.e. some tone was heard as T3), and so the two groups' patterns were similar in this respect.

The last three stimulus types to be discussed were all congruous sentences. Type 5A (normal, congruous sentences) had a Mantel *r* statistic of 0.33 (p = .078), which means they were moderately correlated[5]. Since this type of stimuli contained both acoustic and semantic information, it was predicted to be the least challenging task for both homeland and heritage

[5]Since the focus of this section is the evaluation of the range of the strength of the correlation across stimulus types, rather than an evaluation of whether any given type shows a significant degree of correlation, the interpretation of the last three stimulus types is included despite their relatively higher p-values.

speakers. In Figure 5.22, both matrices had a clear diagonal, which indicates a high accuracy rate. In Figure 5.22a (homeland), all percentages printed on non-diagonal cells were so small that the cells were close to white, which indicates an almost perfect accuracy. As a result, any errors made by heritage speakers would pull the two populations apart. The most common error among heritage speakers was represented by the cell [5,2] in Figure 5.22b, which stood out as the darkest non-diagonal cell, while its counterpart in Figure 5.22a (homeland) was not as dark. This shaded shell indicates that even when both acoustic and semantic cues were available, 13.82% of the time heritage speakers identified T5 (low rising) as T2 (high rising). This direction of T2-T5 confusion was consistent with patterns observed from the stimulus types above. In other words, heritage speakers exhibited T2-T5 confusion in this specific direction across both difficult (e.g. Type 6B) and easy stimulus types (e.g. Type 5A). As for homeland speakers, although they showed T2-T5 and T3-T6 confusion in Types 6B, 5B, 3, and 1, they were able to perceive the intended tonal categories almost perfectly when acoustic and semantic cues were both available.

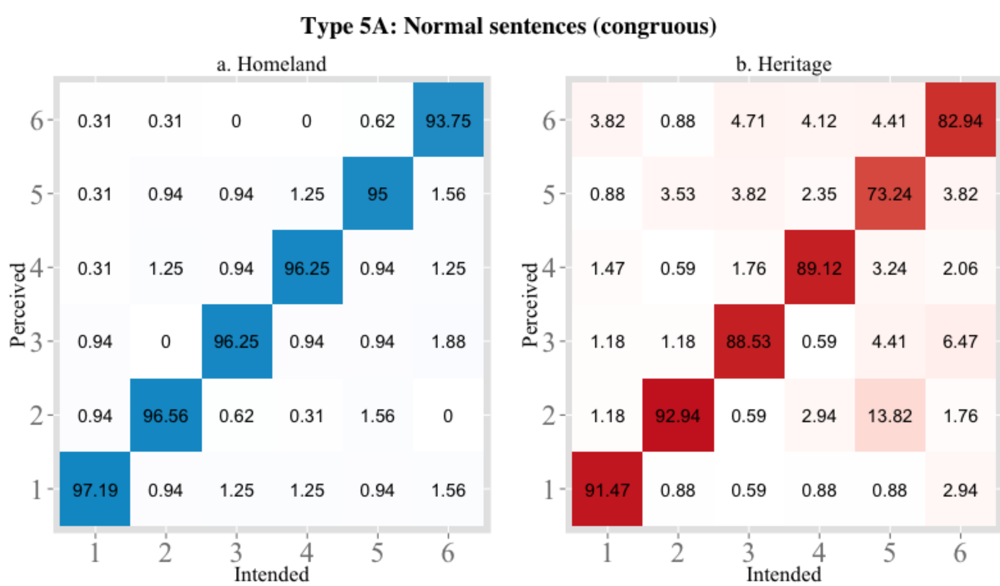


Figure 5.22: Confusion patterns of homeland and heritage speakers for Type 5A stimuli

Similar to those for Type 5A, confusion matrices for Type 6A (the last word of the congruous sentence has no segments) had a Mantel *r* statistic of 0.33 (p = .046), which indicates moderate correlation. The homeland matrices in Figure 5.23a and Figure 5.22a were very similar to each other, in that most non-diagonal cells were almost white (except [3,6] in Figure 5.23a, which read 8.12%). Since the homeland group barely made any errors, any errors made by heritage speakers would increase the difference between the two groups. This was indeed what happened to the heritage matrix in Figure 5.23b: many cells such as [3,5], [5,2], [6,3], [6,4], and [6,5] were shaded (albeit relatively lightly compared with Figure 5.19), but their counterparts in "homeland" were close to white. Among these shaded cells, [5,2] (10.88%) and [3,6] (10.88%) were relatively dark. They provided additional evidence for heritage speakers' T3-T6 and T2-T5 confusions.

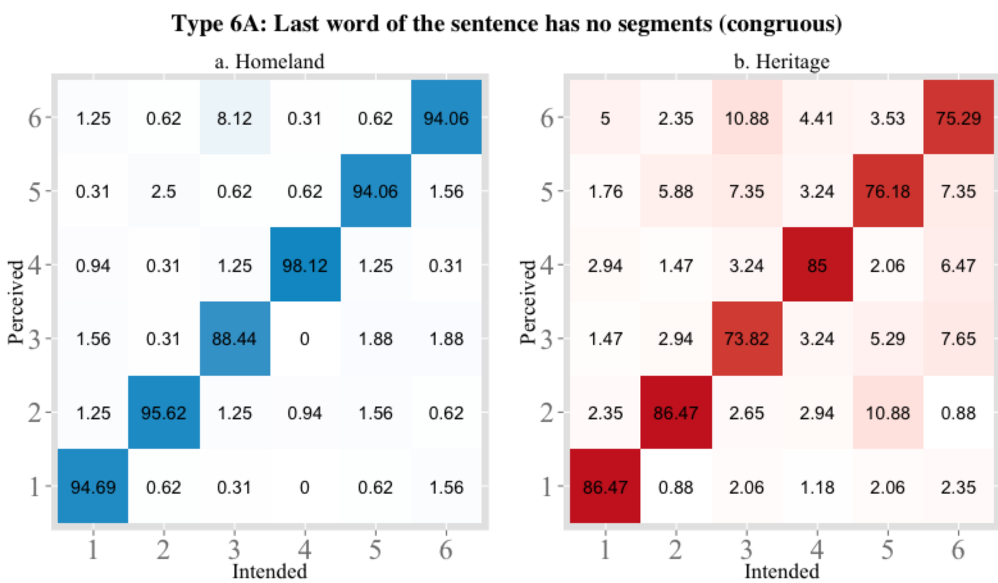


Figure 5.23: Confusion patterns of homeland and heritage speakers for Type 6A stimuli

The least correlated pair was Type 4 (sentences with no tone) in Figure 5.24 according to the Mantel test (r = .30, p = .057). Although this was the least correlated pair among the eight stimulus types, it was still moderately correlated and thus quite similar to Types 5A and 6A in terms of strength of the correlation. Type 4 stimuli contained no tonal information, but the target word was embedded in a congruous sentence that offered semantic context, so it was predicted that the two populations would not differ significantly in terms of accuracy. In Figure 5.24 both matrices had a dark diagonal, a sign of high accuracy rates (homeland 89.22%, heritage 82.25%). However, their error patterns were quite different: in Figure 5.24a (homeland) there was a very weak T3 bias as indicated by a lightly shaded T3 row. However, in Figure 5.24b (heritage), T1 was sometimes misidentified as other tones as indicated by a lightly shaded T1 column.

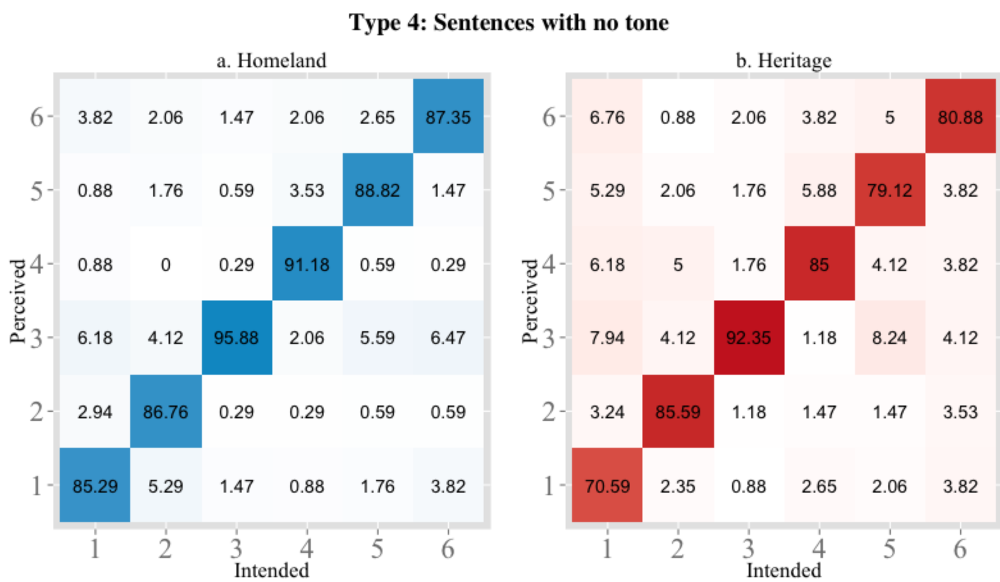


Figure 5.24: Confusion patterns of homeland and heritage speakers for Type 4 stimuli

The weak T3 bias in the homeland group could be attributed to a few subjects who might have interpreted the instructions differently from the majority of participants. These subjects were also the outliers who scored below 70% in Figure 5.3. Recall that Type 4 was a separate block on its own, and the question asked of the participants was "What is the last word of the sentence?" It was up to the participants to decide whether they should make use of semantic cues, or ignore them and use acoustic cues only. Although the majority of homeland speakers used semantic cues for tone identification, a few participants used acoustic cues more often. For example in Figure 5.25, Subject #320 often responded with T3 (mid level). This was why this figure looked more similar to Type 2 (Figure 5.21a) even though it belonged to a different type of stimuli. Interestingly (but not surprisingly), no one from the heritage group did the same in a consistent manner. A possible explanation is that if heritage speakers attended to semantic cues more often, it might not occur to them that there was another way to interpret the instructions.

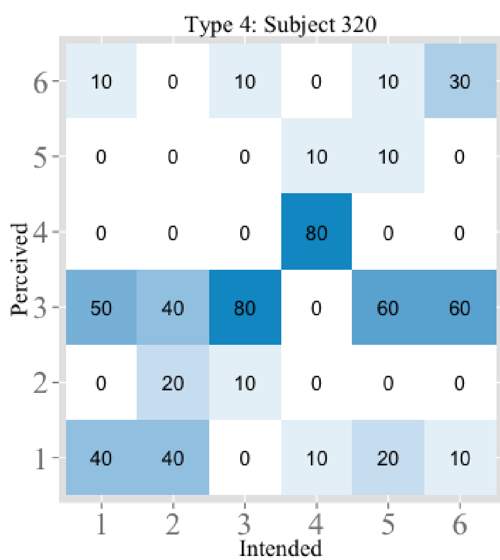


Figure 5.25: Confusion patterns of Subject #320 for Type 4 stimuli

5.3.4 Interim summary

In this interim summary, the second research question of this dissertation is revisited: Do homeland and heritage speakers exhibit similar confusion patterns with respect to lexical tone perception? Overall, results of the

Mantel test suggest that homeland and heritage speakers had fairly similar confusion matrices. Although the strength of correlation varied depending on the stimulus type, all r values fell within the range between “moderately correlated” and “strongly correlated”.

Homeland and heritage speakers showed confusion for overlapping subsets of lexical tones. First, they both showed T2-T5 (high rising and low rising) confusion. As discussed in Section 3.3.2, Cantonese in Hong Kong has been undergoing a T2-T5 merger in both production and perception. Results of the current study confirmed that heritage Cantonese in Canada is following a similar sound change trend in perception. In particular, T5 was heard as T2 more often than T2 being heard as T5. This direction of confusion suggests that T5 is merging into T2.

The second pattern shared by both groups was T3-T6 (mid level and low level) confusion. However, unlike T2-T5, the direction of confusion for T3-T6 was inconsistent across stimulus types. For example, for both groups, T6 was heard as T3 more often than the other direction for Type 5B (normal, incongruous sentences), but T3 was heard as T6 more often than the other direction for Type 3 (words with no segments).

Although the types of confusion were similar, the frequency of confusion differed between the two groups. For every confused tone pair, heritage speakers had a higher error percentage than homeland speakers, as shown in the matrices for Types 1, 5A, 5B, 6A, and 6B. This suggests that both groups perceived the same subset of tonal categories as similar to each other, but these categories were perceived as even more similar for heritage speakers.

Another difference between the two groups was that T4-T6 (low falling and low level) confusion was unique to heritage speakers. Recall from Chapter 3 that Cantonese-naïve English speakers in Francis et al. (2008) also exhibited T4-T6 confusion, possibly due to influence from English intonational categories. Since English intonation has utterance-final “high” or “low” boundary tones (Liberman, 1975; Pierrehumbert, 1980), all low tones in Cantonese (T4 low falling, T5 low rising, T6 low level) may be perceptually assimilated into the “low” category. In terms of the discrimination of T4 and T6, heritage Cantonese speakers in the current study were similar to Cantonese-naïve English speakers. However, they did not share other tone discrimination patterns. In the current study, heritage Cantonese speakers did not show a lot of T5-T6 confusion, which was common among Cantonese-naïve English speakers in Francis et al. (2008) and Qin & Mok (2011).

To conclude, the hypothesis that homeland and heritage speakers exhibit different confusion patterns with respect to lexical tone perception stands only from the quantitative perspective, but does not stand from the qualitative perspective.

5.4 Response to Research Question 3: Use of acoustic and semantic cues

This section addresses the third question regarding the type of information used in word identification. The null and alternative hypotheses are listed as follows:

> **Research Question 3:**
>
> Do homeland and heritage speakers make use of the same type of information when identifying a word from a tonally contrastive set? In particular, are acoustic and semantic information equally useful?
>
> H_0: There is no difference between homeland and heritage speakers in terms of what information they use in tone identification.
>
> H_1: Homeland and heritage speakers use different information in tone identification.

Previously in Section 3.6, it was hypothesized that heritage speakers would rely on semantic information more often than homeland speakers. Therefore, it was anticipated that H_0 would be rejected in favour of H_1. To test this hypothesis, the stimuli must contain both acoustic information and semantic contexts, so that both types of cues were made available for participants to choose from. Therefore, monosyllabic stimuli (Types 1, 2, and 3) that did not have semantic contexts are not useful for testing this hypothesis.

Congruous sentences (Types 5A and 6A) are not useful for testing the hypothesis either, as acoustic and semantic information were not in conflict. Consider this situation: the stimulus was *sap6 ji6 dim2 zung1 hou2 soeng5 cong4 fan3* "at twelve o'clock you'd better go to bed and sleep", and participants were asked whether the last word was *fan1* "share", *fan2* "powder", *fan3* "sleep", or *fan4* "tomb". Participant A paid attention to the tone and picked a word with Tone 3, which is *fan3* "sleep", the correct answer. Participant B, on the other hand, considered the semantic content and picked a word that would make sense in this sentence, which was *fan3* “sleep”, the correct answer. Even though the two participants used different cues, they ended up with the same answer. In other words, it would not be possible to conclude which particular cue was used, or whether both cues were used simultaneously.

Incongruous sentences (Types 5B and 6B) are most useful for testing the hypothesis, as acoustic and semantic information were in conflict and participants would have to use only one of them. As mentioned in Section 4.4.2, participants were told that some sentences might not make sense, and all they had to do was to identify the last word that they heard. In other words, they were instructed to ignore semantic cues and focus on acoustic cues. The explicit instructions ensured that participants’ different responses would be a reflection of different word-identification strategies, but not a reflection of different ways to interpret the instructions. Consider this incongruous sentence: the stimulus was *sap6 ji6 dim2 zung1 hou2 soeng5 cong4 fan2* “at twelve o’clock you’d better go to bed and powder”. Participant A, B, and C responded with *fan2* “powder”, *fan3* “sleep”, and *fan4* “tomb” respectively. Only Participant A was correct. From their answers it can be concluded that Participant A made the decision by using acoustic cues (especially tonal information since the choices were tonally contrastive), Participant B relied on semantic information and picked a word that would make sense (despite being instructed to focus on acoustic cues), and Participant C used neither type of cue. Therefore, data from Types 5B and 6B were able to tease apart how various cues were used in tone identification. The rest of this section will focus on these two stimulus types.

All responses to Type 5B (normal, incongruous sentences) and Type 6B (the last word of the incongruous sentence has no segments) stimuli were grouped into three categories representing the kind of cues being used: “acoustic”, “semantic”, and “neither”. Counts and percentages are presented in Table 5.27, and the same data are visualized in Figure 5.26. Pearson’s chi-squared test of goodness-of-fit was performed to determine whether the distributions of “acoustic”, “semantic”, and “neither” were similar between the two populations for each of the two stimulus types separately. Results show that homeland and heritage speakers had significantly different

Table 5.27: Counts and percentages of different cues used by homeland and heritage speakers for Type 5B (normal, incongruous sentences) and Type 6B stimuli (the last word of the incongruous sentence has no segments)

	Type 5B				Type 6B			
	homeland		heritage		homeland		heritage	
acoustic	5333	87.14%	4058	66.31%	5224	85.34%	3471	56.72%
semantic	403	6.58%	1175	19.20%	402	6.57%	1596	26.08%
neither	384	6.27%	887	14.49%	494	8.07%	1053	17.21%
total	6120	100%	6120	100%	6120	100%	6120	100%

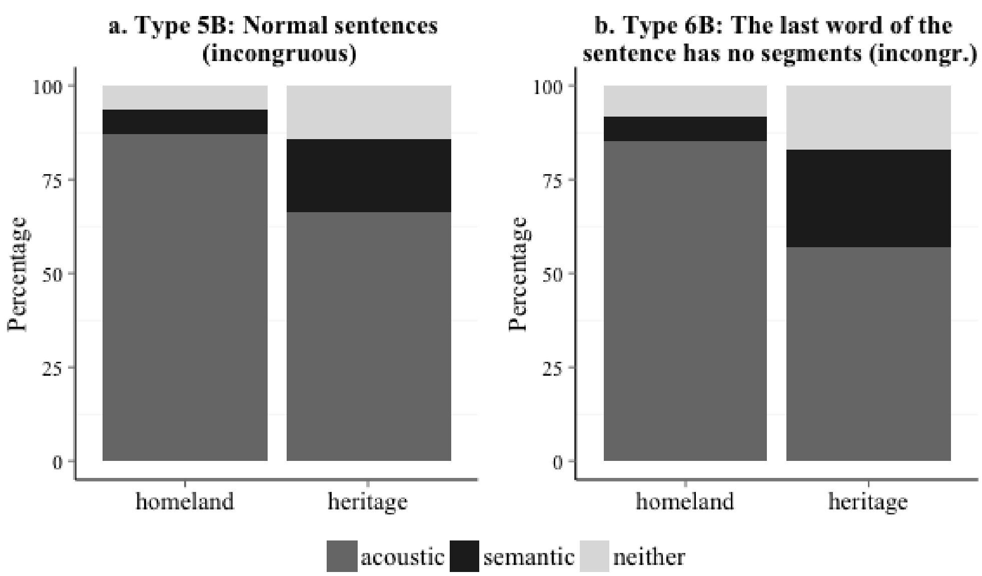


Figure 5.26: Comparison of cues used by homeland and heritage speakers for Type 5B and Type 6B stimuli

distributions of cues used for Type 5B, $\chi^2(2) = 749.85$, $p < .001$, as well as for Type 6B, $\chi^2(2) = 1268.94$, $p < .001$. Two-sample binomial tests confirmed that the two groups' ratios for each type of cue and for each stimulus type were significantly different: $p < .001$ for "acoustic", $p < .001$ for "semantic", and $p < .001$ for "neither".

In Figure 5.26 homeland and heritage speakers seem to have similar error patterns. In the "homeland" bars for both Type 5B and Type 6B, the ratio of "semantic" (in black) to "neither" (in light grey) is around 50:50. In the "heritage" bar for Type 5B, the ratio of "semantic" to "neither" also looks close to 50:50. This raises the question of whether the two groups in fact had similar error patterns: when they did make mistakes, 50% of the time the error was due to the use of semantic cues, and 50% of the time they made random errors, which means there was an equal chance for both. Would it be the case that these results were evidence for similarities rather than differences between the two groups?

To answer this question, the two populations' ratios of "semantic" to "neither" (i.e. black to light grey) were compared. Table 5.27 and Table 5.28 were based on the same data, but the "acoustic" row was removed from the latter, and percentages were redistributed. For Type 5B, homeland speakers used semantic cues in 403 out of 787 erroneous responses, while heritage speakers used semantic cues in 1175 out of 2062 errors. A two-sample binomial test confirmed that 403/787 is significantly smaller than 1175/2062, $p = .001$. The test was performed the other way round and confirmed that 1175/2062 is significantly greater than 403/787, $p < .001$. As for Type 6B, the two populations' ratios were significantly different as well, $p < .001$ (both ways).

To conclude, although heritage speakers were able to use acoustic cues to achieve an accuracy rate above chance when incongruous sentences were presented, they used semantic cues relatively more often than homeland speakers did.

Table 5.28: Counts and percentages of two types of incorrect responses for Type 5B (normal, incongruous sentences) and Type 6B stimuli (the last word of the incongruous sentence has no segments)

	Type 5B				Type 6B			
	homeland		heritage		homeland		heritage	
semantic	403	51.21%	1175	56.98%	402	44.87%	1596	60.25%
neither	384	48.79%	887	43.02%	494	55.13%	1053	39.75%
total	787	100%	2062	100%	896	100 %	2649	100%

Chapter 6

Discussion and conclusion

This final chapter summarizes major findings, discusses their implications, and concludes the dissertation.

6.1 Summary of research findings

Major findings of this dissertation are summarized below as answers to the three research questions introduced in Chapter 1.

Research Question 1:

Do homeland and heritage speakers behave differently in terms of their ability to identify tonally contrastive words?

Answer to Research Question 1:

On average, homeland speakers outperformed heritage speakers in the word-identification task, but the magnitude of their difference was not static across stimulus types. The accuracy gap between the two populations depended on the variables being manipulated in the stimuli. First, when tonal information was the *only* type of information available (as in Type 3, monosyllabic words with no segments), the accuracy gap between the two groups was the largest (80.54% for homeland and 49.41% for heritage; chance level was 25%). This suggests that homeland speakers had a significantly greater ability to distinguish tonally contrastive words by solely relying on tonal information.

Second, when semantic context was available along with tonal information (as in Types 5A, 5B, 6A, and 6B, which were all sentences), the accuracy gap was smaller, although the difference between the two groups was still significant. This suggests that semantic context was helpful for heritage speakers to distinguish tonally contrastive words, but this additional information was not enough to close the accuracy gap.

Lastly, when tonal information was not available at all, the two populations' accuracy rates were on par: they performed equally poorly (as in Type 2, monosyllabic words with no tone) or almost equally well (as in Type 4, sentences with no tone). This was crucial to show that the accuracy gap being observed in other stimulus types was not merely a reflection of heritage speakers' lower Cantonese proficiency, but was due to the two groups' different abilities to make use of tonal information in word identification.

Research Question 2:

Do homeland and heritage speakers exhibit similar confusion patterns with respect to lexical tone perception?

Answer to Research Question 2:

In general, confusion patterns of homeland and heritage speakers were similar. Although the levels of similarity varied across stimulus types, they all fell within the range between "moderately correlated" and "strongly correlated".

Error patterns of both groups exhibited trends of tone merger for the two rising tones, namely T2 [25] (high rising) and T5 [23] (low rising). For both homeland and heritage speakers, T5 was mistaken for T2 more often than T2 being mistaken as T5, which can be a sign that T5 is merging into the T2 category in perception. Although both groups exhibited confusion for T2 and T5, heritage speakers had a higher error percentage for this tone pair than homeland speakers, which suggests that heritage speakers are ahead of homeland speakers in this ongoing sound change.

Apart from confusion between two rising tones, confusion between two level tones, namely T3 [33] (mid level) and T6 [22] (low level), was also observed in both groups. Unlike the asymmetrical pattern of T2-T5 confusion, no particular direction of T3-T6 confusion was predominantly more frequent than the other across the board. Although both groups showed confusion for this tone pair, heritage speakers had a higher error percentage than homeland speakers.

Lastly, confusion between T4 [21] (low falling) and T6 [22] (low level) was found among heritage speakers but it was very rare for homeland speakers. This is a qualitative difference between the two groups.

To sum up, there were more similarities than differences between the two groups' confusion patterns. Their differences were more quantitative than qualitative, in that both groups had confusion for overlapping subsets of categories, but the heritage group showed a higher degree of confusion compared with the homeland group.

Research Question 3:

Do homeland and heritage speakers make use of the same type of information when identifying a word from a tonally contrastive set? In particular, are acoustic and semantic information equally useful?

Answer to Research Question 3:

Overall, homeland speakers were better at using acoustic information compared with heritage speakers, while heritage speakers had a relatively higher tendency to rely on semantic information compared with homeland speakers. This was most clearly shown when the target word was semantically incongruous with the carrier phrase (e.g. *sap6 ji6 dim2 zung1 hou2 soeng5 cong4 fan2* "at twelve o'clock you'd better go to bed and powder"). Homeland speakers were significantly better at actively rejecting the word that would make sense in that sentence (e.g. *fan3* "sleep"), and selecting the actual word presented (e.g. *fan2* "powder") by attending to tonal information. The fact that heritage speakers achieved accuracy rates well above chance (66.31% for Type 5B and 56.72% for Type 6B; chance level was 25%) suggests that they were not completely unable to attend to acoustic information. Compared with homeland speakers, they made relatively more errors due to relying on semantic information (e.g. chose *fan3* "sleep" as the answer). This is evidence that the two groups have different listening strategies during word identification even when they were told explicitly to focus on what they heard instead of what made sense.

6.2 Discussion and implications

This section discusses how findings of the present study can address unanswered questions in the existing literature identified in Chapter 2 and Chapter 3.

6.2.1 Sound change trends in heritage Cantonese

In Chapter 3 it was pointed out that as participants from Canada were at an average age of 20.78 years at the time of the experiment, their parents very likely had migrated from Hong Kong to Canada before the 2000s—the time when T2-T5 merger started to be documented in the literature (Bauer et al., 2003; Kej et al., 2002). Their migration to a geographically remote country across the Pacific Ocean conveniently allowed a natural experiment of sound change: will a merger-free variety of Cantonese be passed on to their children who grew up in Canada?

Results in Section 5.3 suggest that homeland and heritage speakers have similar tone merger trends in perception; moreover, heritage speakers' higher error percentage is a sign that they are ahead of homeland speakers in these trends. There are at least four possible reasons for this. First, although T2-T5 confusion started to catch the attention of researchers in the 2000s, it is possible that the trend had actually started for a period of time before it was documented in the literature. If the parents had merged the tones in production before migrating to Canada, it would follow that their children acquired the same variety of Cantonese with signs of sound change.

Another possible reason is that although Hong Kong and Canada are geographically separated, socio-culturally there are strong ties between the two regions. Heritage speakers can easily access Cantonese popular culture such as film and music from Hong Kong through the internet, and they may also visit their family and relatives in Hong Kong regularly. As a result, even if their parents do not merge the tones, they may have acquired the sound change from other speakers residing in Hong Kong. The two reasons mentioned so far, however, are not adequate to explain all observations. If heritage speakers merely acquired the sound change from homeland speakers, the two groups would be expected to have no quantitative difference in accuracy for the tone pairs in question. Why did heritage speakers have a higher level of confusion?

The third possible explanation is that tone merger was induced by language-internal factors but not idiosyncratic factors unique to speakers in Hong Kong. As such, sound change would take place regardless of external factors like the speakers' geographic location. These language-internal factors may include inherent characteristics of the tones in question, such as acoustic similarity, or information-theoretic properties of the tones, such as functional load. Tsui (2012) proposes that it takes both factors to motivate tone merger. According to his analysis, the pair T2-T5 has the smallest acoustic distance *and* carries the second lowest functional load among all tone pairs. The combination of small acoustic distance and low functional load makes T2-T5 more susceptible to tone merger. If this is the case, two geographically separated speaker communities could ultimately end up on the same path towards merger, and it is possible for either community to go faster than the other. However, this still does not explain why heritage speakers had moved further other than a result of chance.

To account for their faster sound change trends, the reason must be something that lies within the heritage speaker population, which led to their lower sensitivity to the relevant tonal contrasts for word identification. Could this population-internal factor be the incomplete acquisition of tonal perception, or language attrition resulting from cross-language effects from a dominant language, namely English? This question will be addressed in the next subsection through a comparison among heritage speakers of the current study, Cantonese-learning children discussed in Section 3.2, and Cantonese-naïve English speakers discussed in Section 3.4.

6.2.2 Tonal perception and heritage bilingualism

Several important questions raised in the literature review chapters should be revisited. Previous works such as Benmamoun et al. (2013b), Montrul (2013), and Polinsky & Kagan (2007) comment that phonetics and phonology are the most stable domains in heritage speakers' grammar compared with morphology, syntax, and semantics. While this claim is supported by studies like Tees & Werker (1984), there are also cases like Celata & Cancila (2010) where heritage speakers have lost sensitivity to certain sound contrasts of their non-dominant L1 that are not phonemic in their dominant L2. What is common among these studies is that the sound contrasts in question are mostly consonants or vowels. What happens if one's non-dominant L1 has the contrastive dimension of lexical tone, but the dominant L2 lacks such a dimension altogether? Can an intonational suprasegmental phonology (as in English) impact a tonal suprasegmental phonology (as in Cantonese)?

The statement that heritage speakers have *lost* sensitivity to certain sound contrasts presupposes that they once possessed such sensitivity. What if these contrasts were not acquired in the first place, and so they cannot really be *lost*? In Ciocca & Lui (2003) and Wong & Leung (2018), the six-year-olds residing in Hong Kong had not reached adult accuracy with respect to tone identification. For children residing in Canada, if the onset of schooling is around five years of age, the switch of dominant language may happen before they fully acquire all tonal contrasts. Extensive L2 input and reduced L1 exposure together is a common cause of incomplete acquisition of L1 in immigrant communities (Levine, 2015; Montrul, 2008). If the tonal grammar of adult heritage speakers is a fossilized form of what they had acquired so far in early childhood, tone pairs with the lowest accuracy for heritage speakers in the current study should match those in previous studies on young children. As discussed in Section 5.3, heritage speakers showed most confusion for T2-T5 and T3-T6. In Ciocca & Lui (2003) these were the last two tonal contrasts acquired in perception by Cantonese-learning children in Hong Kong. From the matching results it can be concluded that incomplete acquisition is a *possible* reason for heritage speakers' lower sensitivity to T2-T5 and T3-T6, which may also explain why they appear to have moved further than homeland speakers in sound change trends. Further research on the longitudinal language development of Cantonese-learning children in Canada is necessary to provide empirical evidence to support this hypothesis.

To tap into potential cross-language effects from English, the linguistic behaviour of heritage speakers in the present study should be compared with those of Cantonese-naïve English speakers in Francis et al. (2008) and Qin & Mok (2013). In their studies, English speakers found tones with a similar average pitch height (T4-T5, T4-T6, T5-T6) harder to distinguish than tones with the same direction of pitch change (T2-T5 and T3-T6). In both studies' multidimensional scaling analyses, T5 [23] and T6 [22] are very close in the perceptual space of English speakers. Authors of both studies discuss potential influences from English intonation, in which "high" or "low" boundary tones can be found at the edge of a phrase or sentence. T4 [21], T5 [23], and T6 [22] in Cantonese are therefore perceptually assimilated to the "low" category in English prosody. This is quite different from error patterns in the current study. Although heritage speakers did exhibit T4-T6 confusion sometimes, they hardly made errors for T4-T5 or T5-T6. In other words, there was little evidence that heritage speakers' tonal perception was affected by intonational categories in English.

Although incomplete acquisition of tonal contrasts may better account for heritage speakers' specific confusion patterns, it still cannot explain their overall lower ability to distinguish tonally contrastive words by solely relying on tonal information in the acoustic signal, as shown in the big accuracy gap for Type 3 stimuli (monosyllabic words with no segments) compared with other stimulus types. If incomplete acquisition was the *sole* difference between homeland and heritage speakers, T2-T5 and T3-T6 would have been the most confused tone pairs across stimulus types. In other words, the cells [2,5], [5,2], [3,6], and [6,3] would have been the darkest non-diagonal cells in heritage speakers' confusion matrices across the board. However, in the actual results for Type 3, the confusion matrix of heritage speakers had relatively evenly shaded non-diagonal cells (see Figure 5.19), which suggests a global decline of accuracy regardless of the tone of the target word. At least two speculations can be made with regard to heritage speakers' overall lower ability to rely on tonal information alone in the word-identification experiment.

The first speculation pertains to different quantities of linguistic input received by homeland and heritage speakers during their acquisition of Cantonese. In the current study, to ensure that both tested populations were exposed to the same baseline variety of Cantonese, the screening process excluded heritage speakers whose parents were also heritage speakers, and only included individuals who were born to parents originally from Hong Kong. This was an attempt to control the *quality* but not the *quantity* of the input. Homeland speakers grew up in an environment where Cantonese is used in a variety of domains, including education and public communication. This environment allowed homeland speakers to be exposed to Cantonese spoken by a large number of talkers with both familiar and unfamiliar voices. However, heritage speakers grew up in an English-dominant environment, where the use of Cantonese is mostly restricted to the family setting. The Cantonese input that they had received is mainly from a limited number of family members with familiar voices. This difference may affect the two groups' abilities of perceptual normalization for Cantonese tones produced by an unfamiliar voice in the experiment. In a study about talker familiarity effects on speech intelligibility, Nygaard & Pisoni (1998) found that sensitivity to talker-specific indexical information in a voice facilitates the extraction of meaningful linguistic units from a speech signal. Although there was only one talker in the current study and her voice was unfamiliar to all participants, homeland speakers had more prior experience of extracting tonal categories from speech produced with unfamiliar voices. In other words, their exposure to Cantonese spoken by a large number of talkers may have allowed them to perceptually adapt to tones produced with a novel voice more effectively. As a result, homeland speakers may be less affected by talker familiarity effects. Heritage speakers, on the other hand, are used to processing Cantonese tones produced by a fewer number of talkers. Therefore, their performance in the word-identification task may be more impacted by talker familiarity effects, especially when tone was the only available type of information in the speech signal. Further research on heritage speakers' perceptual flexibility is necessary to confirm this speculation.

The second speculation is cross-language effects from English by virtue of its lack of tone as a contrastive dimension. Although specific confusion patterns of heritage speakers showed little evidence of effects from English intonational phonology, English as a non-tonal language may still leave a global impact on heritage speakers' ability to make use of tonal information in Cantonese word identification. After the onset of schooling, heritage speakers had received regular and extensive English input, and Cantonese input was relatively reduced. To optimize the perceptual system for a frequently used non-tonal dominant language, reorganization of cognitive resources and adjustment of listening strategies may occur. In Bruggeman (2016), Dutch emigrants in Australia ignored stress cues that are useful for processing Dutch but not useful for processing English. In Celata & Cancila (2010), English-dominant heritage speakers of Lucchese became insensitive to the singleton-geminate consonant distinction in Lucchese, because this cue is not useful for distinguishing English words. It is possible that heritage speakers of Cantonese in the current study have experienced a similar switch of perception strategy from one that favours the processing of Cantonese lexical tones (as in Vancouver-based Cantonese-learning infants in Yeung et al., 2013), to one that can improve the efficiency of processing English. Further research on tonal perception by school-age Cantonese-learning children or teenagers in Canada will help to confirm whether there is a switch of strategy and, if so, when it happens.

If a re-allocation of cognitive resources in favour of a dominant L2 has occurred, this will have interesting implications for the permanence vs. contingency debate in the language acquisition literature. It has been established that perceptual narrowing happens early in life during a critical period, when infants and children show a decline of sensitivity to non-native phonetic contrasts (Maurer & Werker, 2014; Tees & Werker, 1984; Werker

& Hensch, 2015), and exhibit signs of a neural commitment to language-specific auditory patterns (Kuhl et al., 2006; Zhang et al., 2005). As a result, the way individuals listen to language in general is tailored to their L1 (Cutler, 2012). There is, however, not yet a consensus on the reversibility of this neural commitment to L1. On the one hand, the permanence hypothesis posits that resources dedicated to a specific language cannot be re-allocated, and therefore speech perception strategies established during infancy or early childhood will be permanent throughout life (Benmamoun et al., 2013b). This hypothesis is supported by studies on phoneme discrimination by international adoptees, such as Oh et al. (2010) and Pierce, Klein, Chen, Delcenserie & Genesee (2014). On the other hand, the contingency hypothesis posits that the persistence of linguistic knowledge depends on continuous input; neural commitments can be altered in favour of another language when the exposure of L1 decreases or stops (Benmamoun et al., 2013b). This hypothesis is supported by the aforementioned studies on immigrant communities (Bruggeman, 2016; Celata & Cancila, 2010).

Empirical facts from the current study suggest that the two hypotheses are not necessarily mutually exclusive. Indeed, heritage Cantonese speakers have maintained a lexical tone system in their perception despite being English-dominant. In this regard, it may be true that the phonological knowledge of tone as a dimension for lexical contrast is persistent even after reduced exposure of Cantonese. However, this study has also shown that heritage speakers were less good at using tonal cues compared with homeland speakers. It could be that the process of mapping listening cues to phonological categories in the lexical tone system was altered for efficient processing of a non-tonal dominant language. The change may not result in complete abandonment of Cantonese-specific listening strategies, but it may favour English and put Cantonese to a lower priority. Further studies on different aspects of heritage speakers' perceptual grammar may provide a more nuanced view on the permanence and contingency hypotheses.

6.2.3 Language pedagogy for heritage learners of Cantonese

This study also has implications for Cantonese language teaching, especially in the context of heritage language maintenance in North America, where a growing number of universities offer Cantonese courses, such as the University of British Columbia (Pai, 2016), New York University (NYU College of Arts and Science, 2018), the Ohio State University (OSU East Asian Studies Center, 2014), and Stanford University (Stanford Language Center, 2018). Existing works on Cantonese pedagogy mostly focus on teaching Cantonese as a foreign language (Lee, 2004; Pai, 2016). As the present study has shown, heritage speakers do not have the same linguistic behaviour as homeland Cantonese speakers or Cantonese-naïve English speakers, and as a result they are expected to have different learning needs in the language classroom. Below are a few take-home messages for educators and curriculum developers.

The case of heritage speakers challenges the traditional practice of using a monolingual native speaker norm as the standard for evaluating linguistic competence of any language user. In an increasingly bilingual or multilingual world, it is imperative to recognize that bilingual language users, as Grosjean (1989) puts it, are never two monolinguals living in one person. Therefore, it is unreasonable to expect or require that bilinguals have the exact same linguistic behaviour as monolingual native speakers. In Samuel & Larraza (2015), L1 Basque speakers who are fluent in Spanish failed to reject non-words with deviated pronunciations half of the time in a picture-name matching task in Basque, even though they did very well in an AXB discrimination task. The authors see it as perceptual adaptation to their linguistic environment, where Basque speakers often interact with people who speak Spanish-accented Basque. In this dissertation, heritage Cantonese speakers in Canada were good at using semantic context as a listening cue but less good at using acoustic information. A comment by Lynch (2003) can be a good explanation: the nature of L1 acquisition by heritage speakers is “dialogic, discursive and absolutely contextual from the beginning” (p.11). Hence, it is no surprise that they are accustomed to using top-down processing strategies to extract the message of a Cantonese utterance in a discursive context. In both Samuel & Larraza (2015) and the present study, low accuracy in one listening task does not always imply overall perceptual "impairment". Instead, bilinguals with a special language background may have *different* but not necessarily *poorer* language processing strategies compared with prototypical native speakers.

As for curriculum design, a heritage learner-oriented approach should be meaning-focussed and communicative. Cummins (1979) made a distinction between two pedagogical styles: *cognitive/academic language proficiency*, and *interpersonal communication skills*. The former is a more formal, grammar-based approach that emphasizes drills on conscious knowledge, such as pronunciation and rules. The latter emphasizes what learners can do using their language in a social setting in the real world. In this dissertation, heritage speakers of Cantonese definitely have tone as a contrastive dimension in their perceptual grammar, which means tone is already part of their unconscious linguistic knowledge. As such, the cognitive/academic approach may not be suitable for heritage speakers, especially when there are second language learners in the same classroom. As Krashen (2000) describes, non-heritage language learners often outperform heritage learners when linguistic competence is assessed as declarative knowledge. This could be psychologically devastating and traumatizing for heritage speakers, who may not understand that they do possess procedural knowledge of the language. Therefore, the interpersonal approach that focuses on the communicative content is more suitable for heritage learners.

This communicative style of pedagogy can be supplemented with listening practice targeting at commonly confused tone pairs. Although heritage bilinguals should not be expected or required to behave exactly like prototypical native speakers, it is undeniable that tonal confusion may obstruct everyday communication. To strengthen heritage learners' tone discrimination ability, listening materials with multiple unfamiliar voices may allow learners to get used to extracting auditory cues from novel voices and facilitate intertalker normalization in tonal processing. In general, it is important to offer ample opportunities of re-exposure to the heritage language, which may serve as a triggering experience for linguistic knowledge that has not been accessed for an extended period of time.

6.3 Conclusions

To conclude, it is certain that heritage speakers' perceptual grammar of Cantonese has tone as a contrastive dimension. In the word-identification task, heritage speakers' accuracy was significantly above chance across stimulus types (except for Type 2, words with no tone, which was intended to be an impossible task). This provides evidence that they were able to maintain the contrastive dimension of tone after becoming English-dominant.

Heritage speakers' perceptual grammar of Cantonese is, however, not identical to that of homeland speakers. On average, homeland speakers achieved a higher accuracy than heritage speakers, but the difference between the two groups varied among stimulus types, depending on what type of information was available. Although tonal information alone was not very useful for heritage speakers, semantic context did help to decrease the accuracy gap. Two speculations were made with regard to heritage speakers' lower ability to use tonal cues, namely the challenge of perceiving pitch attributes from an unfamiliar voice due to limited Cantonese input, and the change of listening strategy in favour of English, a non-tonal dominant language.

When it comes to specific confusion patterns, heritage speakers were more similar to homeland speakers than to Cantonese-naïve English speakers. Both homeland and heritage speakers found tones with the same direction of pitch change confusing, while Cantonese-naïve English speakers tend to find tones with similar overall pitch heights confusing. Heritage speakers' higher level of confusion for tone pairs with the same direction of pitch change indicates that they are ahead of homeland speakers in sound change trends. Such difference may be owing to incomplete acquisition of tonal perception by heritage speakers, though further research is required to confirm this.

All in all, findings of the current study agree with previous research, in that heritage speakers have unique linguistic behaviour that is different from homeland speakers. In the context of heritage language maintenance through language education, the uniqueness of this population implies that a heritage learner-specific curriculum will be more effective than a typical curriculum of teaching Cantonese as a foreign language.

References

Ali, D. (2015, January 14). Language proficiency on LinkedIn. Retrieved from http://www.linkedin.com/pulse/language-proficiency-linkedin-duaa-ali → pages 29

Amengual, M. (2017). Type of early bilingualism and its effect on the acoustic realization of allophonic variants: Early sequential and simultaneous bilinguals. *International Journal of Bilingualism*, 1–17. doi:10.1177/1367006917741364 → pages 30

Amengual Watson, M. (2013). *An experimental approach to phonetic transfer in the production and perception of early Spanish-Catalan bilinguals*. PhD thesis, The University of Texas at Austin. → pages 40

Antoniou, M., Best, C. T., Tyler, M. D., & Kroos, C. (2010). Language context elicits native-like stop voicing in early bilinguals' productions in both L1 and L2. *Journal of Phonetics*, *38*(4), 640–653. doi:10.1016/j.wocn.2010.09.005 → pages 44, 49, 103

Antoniou, M., Best, C. T., Tyler, M. D., & Kroos, C. (2011). Inter-language interference in VOT production by L2-dominant bilinguals: Asymmetries in phonetic code-switching. *Journal of Phonetics*, *39*(4), 558–570. doi:10.1016/j.wocn.2011.03.001 → pages 44, 49, 103

Barnes, T. & Hutton, T. (2016). Dynamics of economic change in Metro Vancouver: Networked economies and globalizing urban regions. Retrieved from http://www.mvprosperity.org/Documents/DynamicsofEconomicChangeinMetroVancouver.pdf → pages 4

Baron, J. & Strawson, C. (1976). Use of orthographic and word-specific knowledge in reading words aloud. *Journal of Experimental Psychology: Human Perception and Performance*, *2*(3), 386. doi:10.1037/0096-1523.2.3.386 → pages 21

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. doi:10.1016/j.jml.2012.11.001 → pages 132

Barrie, M. (2003). Contrast in Cantonese vowels. *Toronto Working Papers in Linguistics*, *20*. → pages 14, 15

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. doi:10.18637/jss.v067.i01 → pages 130

Bauer, R. S. (1985). The expanding syllabary of Hong Kong Cantonese. *Cahiers de Linguistique Asie Orientale*, *14*(1), 99–111. doi:10.1163/19606028_014_01-05 → pages 12, 15

Bauer, R. S. (2016). The Hong Kong Cantonese language: Current features and future prospects. *Global Chinese*, *2*(2), 115–161. doi:10.1515/glochi-2016-0007 → pages 12, 14

Bauer, R. S. & Benedict, P. K. (1997). *Modern Cantonese phonology*. Berlin, Germany: Walter de Gruyter. → pages xi, 12, 13, 14, 15, 16, 17, 18, 19, 21

Bauer, R. S., Cheung, K.-H., & Cheung, P.-M. (2003). Variation and merger of the rising tones in Hong Kong Cantonese. *Language Variation and Change*, *15*(02), 211–225. doi:10.1017/S0954394503152039 → pages 18, 59, 198

Benmamoun, E., Montrul, S., & Polinsky, M. (2010). Prolegomena to heritage linguistics [white paper]. University of Illinois at Urbana-Champaign and Harvard University. Retrieved from www.nhlrc.ucla.edu/pdf/HL-whitepaper.pdf → pages 2

Benmamoun, E., Montrul, S., & Polinsky, M. (2013a). Defining an "ideal" heritage speaker: Theoretical and methodological challenges reply to peer commentaries. *Theoretical Linguistics*, *39*(3-4), 259–294. doi:10.1515/tl-2013-0018 → pages 25

Benmamoun, E., Montrul, S., & Polinsky, M. (2013b). Heritage languages and their speakers: Opportunities and challenges for linguistics. *Theoretical Linguistics*, *39*(3-4), 129–181. doi:10.1515/tl-2013-0009 → pages 2, 42, 200, 204

Birdsong, D., Gertken, L. M., & Amengual, M. (2012). Bilingual Language Profile: An easy-to-use instrument to assess bilingualism. *COERLL, University of Texas at Austin*. → pages xxi, 30, 115, 117, 276

Blicher, D. L., Diehl, R. L., & Cohen, L. B. (1990). Effects of syllable duration on the perception of the Mandarin Tone 2/Tone 3 distinction: Evidence of auditory enhancement. *Journal of Phonetics*. → pages 98

Bloomfield, L. (1933). *Language*. London: George Allen & Unwin Ltd. → pages 28

Boersma, P. (2002). Praat, a system for doing phonetics by computer. *Glot international*, *5*(9/10), 341–345. Retrieved from http://hdl.handle.net/11245/1.200596 → pages 97

Bolton, K. (2011). Language policy and planning in Hong Kong: Colonial and post-colonial perspectives. *Applied Linguistics Review*, *2*, 51–74. doi:10.1515/9783110239331.51 → pages 12, 36

Boyle, J. (1997). Success and failure in learning Cantonese. *Language Learning Journal*, *16*(1), 82–86. doi:10.1080/09571739785200341 → pages 6

Brinton, D., Kagan, O., & Bauckus, S. (2008). *Heritage language education: A new field emerging*. New York, NY: Routledge. → pages 26

Bruggeman, L. (2016). *Nativeness, dominance, and the flexibility of listening to spoken language*. PhD thesis, Western Sydney University. → pages 47, 203, 204

Burnham, D., Ciocca, V., Lauw, C., Lau, S., & Stokes, S. (2000). Perception of visual information for Cantonese tones. *Proceedings of the 8th Australian International Conference on Speech Science and Technology*, 86–91. Retrived from http://www.assta.org/sst/SST-00/cache/SST-00-Chapter4-p2.pdf → pages 56, 79, 81

Burnham, D., Lau, S., Tam, H., & Schoknecht, C. (2001). Visual discrimination of Cantonese tone by tonal but non-Cantonese speakers, and by non-tonal language speakers. *AVSP 2001 International Conference on Auditory-Visual Speech Processing*, 155–160. Retrieved from https://www.isca-speech.org/archive_open/archive_papers/avsp01/av01_155.pdf → pages 56, 61

Carreira, M. & Kagan, O. (2011). The results of the National Heritage Language Survey: Implications for teaching, curriculum design, and professional development. *Foreign Language Annals*, *44*(1), 40–64. doi:10.1111/j.1944-9720.2010.01118.x → pages 100, 126

Casillas, J. (2015). Production and perception of the/i/-/I/vowel contrast: The case of L2-dominant early learners of English. *Phonetica*, *72*(2-3), 182–205. doi:10.1159/000431101 → pages 30

Celata, C. & Cancila, J. (2010). Phonological attrition and the perception of geminate consonants in the Lucchese community of San Francisco (CA). *International Journal of Bilingualism*, *14*(2), 185–209. doi:10.1177/1367006910363058 → pages 45, 46, 49, 66, 200, 203, 204

Chang, C. B. & Yao, Y. (2016). Toward an understanding of heritage prosody: Acoustic and perceptual properties of tone produced by heritage, native, and second language speakers of Mandarin. *Heritage Language Journal*, *13*(2), 134–160. → pages 64

Chang, C. B., Yao, Y., Haynes, E. F., & Rhodes, R. (2011). Production of phonetic and phonological contrast by heritage speakers of Mandarin. *The Journal of the Acoustical Society of America*, *129*(6), 3964–3980. doi:10.1121/1.3569736 → pages 42, 45, 49

Chang, Y.-h. S., Yao, Y., & Huang, B. H. (2017). Effects of linguistic experience on the perception of high-variability non-native tones. *The Journal of the Acoustical Society of America*, *141*(2), EL120–EL126. doi:10.1121/1.4976037 → pages 56, 61

Chao, Y. R. (1947). *Cantonese primer*. Cambridge, MA: The Harvard-Yenching Institute [by] Harvard University Press. → pages 9, 10, 16, 17, 19

Chau, W. (2011). The influence of Dutch on Cantonese in the Netherlands: Rural versus urban areas. Master's thesis, Leiden University. → pages 9

Cheng, S.-P. & Tang, S.-W. (2014). Languagehood of Cantonese: A renewed front in an old debate. *Open Journal of Modern Linguistics*, *4*(03), 389–398. doi:10.4236/ojml.2014.43032 → pages 13

Cheng, S.-P. & Tang, S.-W. (2016a). Cantonese. In S.-W. Chan (Ed.), *The Routledge Encyclopedia of the Chinese Language* (pp. 18–34). Milton Park, Abingdon, Oxon: Routledge. → pages 12, 13, 14

Cheng, S.-P. & Tang, S.-W. (2016b). Cantonese romanization. In S.-W. Chan (Ed.), *The Routledge Encyclopedia of the Chinese Language* (pp. 35–50). Milton Park, Abingdon, Oxon: Routledge. → pages 20

Cheong, E. & Lee, L. (2015). Cantonese: Passing. Short movie screened at the Vancouver Asian Film Festival 2015. → pages 5

Cheung, S. H.-n. (2007). 香港粵語語法的研究 *Cantonese as spoken in Hong Kong (Revised Edition)*. The Chinese University of Hong Kong. → pages 14, 18

Chiong, R., Lau, W. T., Ng, W. M., Sun, Y., Wong, T. H., & Xie, H. (2017). Have you eaten yet? Investigating language and identity [online video]. Retrieved August 5, 2017, from https://www.youtube.com/watch?v=2C6IgDNZpDo → pages 5

Choi, T.-M., Liu, S.-C., Pang, K.-M., & Chow, P.-S. (2008). Shopping behaviors of individual tourists from the Chinese Mainland to Hong Kong. *Tourism Management*, *29*(4), 811–820. doi:10.1016/j.tourman.2007.07.009 → pages 21

Chung, F. H.-K. & Leung, M.-T. (2008). Data analysis of Chinese characters in primary school corpora of Hong Kong and mainland China: Preliminary theoretical interpretations. *Clinical Linguistics & Phonetics*, *22*(4-5), 379–389. doi:10.1080/02699200701776757 → pages 21

Chung, L. M. V. (2009). Perception and production of Cantonese tones by Thai and Filipino students in Hong Kong secondary school. Master's thesis, The Chinese University of Hong Kong. → pages 61

Ciocca, V. & Ip, V. W.-K. (2008). Development of tone perception and tone production in Cantonese-learning children aged 2 to 5 years. *Proceedings of the 9th Annual Conference of the International Speech Communication Association*. Retrieved from https://www.isca-speech.org/archive/archive_papers/interspeech_2008/i08_0623.pdf → pages 55

Ciocca, V. & Lui, J. (2003). The development of the perception of Cantonese lexical tones. *Journal of Multilingual Communication Disorders*, *1*(2), 141–147. doi:10.1080/1476967031000090971 → pages 18, 55, 56, 79, 81, 200

City of Richmond. (2017). Languages hot facts. Retrieved from https://www.richmond.ca/__shared/assets/Languages6251.pdf → pages 4

City of Vancouver. (2014). First peoples: A guide for newcomers. Retrieved from https://vancouver.ca/files/cov/First-Peoples-A-Guide-for-Newcomers.pdf → pages 5

Clyne, M. & Kipp, S. (1997). Linguistic diversity in Australia. *People and Place*, *5*(3), 6. → pages 9

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. New York, NY: Lawrence Erlbaum Associates. → pages xiii, 141

Connine, C. M., Mullennix, J., Shernoff, E., & Yelen, J. (1990). Word familiarity and frequency in visual and auditory word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*(6), 1084–1096. doi:10.1037/0278-7393.16.6.1084 → pages 84

Crissman, L. W. (2012). Digital language atlas of China. Harvard Dataverse. doi:1902.1/18939 → pages 9

Crosswhite, K. (2009). Praat script for adjusting intensity. Retrieved August 31, 2016, from http://phonetics.linguistics.ucla.edu/facilities/acoustic/adjust_intensity_whole_file.txt → pages 98

Crystal, D. (2012). *English as a global language*. Cambridge, England: Cambridge University Press. → pages 36

Cummins, J. (1979). Cognitive/academic language proficiency, linguistic interdependence, the optimum age question and some other matters. *Working Papers on Bilingualism*, *19*, 121–129. → pages 206

Cummins, J. (1992). Heritage language teaching in Canadian schools. *Journal of Curriculum Studies*, *24*(3), 281–286. doi:10.1080/0022027920240306 → pages 26

Cummins, J. (2005). A proposal for action: Strategies for recognizing heritage language competence as a learning resource within the mainstream classroom. *Modern Language Journal*, *89*(4), 585–592. → pages 26

Cummins, J. & Danesi, M. (1990). *Heritage languages: The development and denial of Canada's linguistic resources*. Toronto, ON: James Lorimer & Company. → pages 2, 26

Cutler, A. (2012). *Native listening: Language experience and the recognition of spoken words*. Cambridge, MA: MIT Press. → pages 204

Cutler, C. L. (2000). *O brave new words!: Native American loanwords in current English*. Norman, OK: University of Oklahoma Press. → pages 1

Darrow, B. (2017). LinkedIn claims half a billion users. *Fortune*. Retrieved April 1, 2018, from http://fortune.com/2017/04/24/linkedin-users → pages 29

De Geer, B. (1992). Internationally adopted children in communication: A developmental study. *Working Papers in Linguistics*, *39*, 1–200. Retrieved from http://journals.lub.lu.se/index.php/LWPL/issue/view/2526/388. → pages 47

DeFrancis, J. (1986). *The Chinese language: Fact and fantasy*. Honolulu, HI: University of Hawai'i Press. → pages 13

Dijkstra, T. (2005). Bilingual visual word recognition and lexical access. In J. F. Kroll & A. M. B. De Groot (Eds.), *Handbook of bilingualism: Psycholinguistic approaches* (pp. 179–201). New York, NY: Oxford University Press. → pages 36

Dodson, C. J. (1981). A reappraisal of bilingual development and education: Some theoretical and practical considerations. In H. B. Beardsmore (Ed.), *Elements of Bilingual Theory* (pp. 14–27). Brussels: Vrije Universiteit Brussel. → pages 30

Dornic, S. (1978). The bilingual's performance: Language dominance, stress, and individual differences. In D. Gerver & H. Sinaiko (Eds.), *Language Interpretation and Communication* (pp. 259–271). New York, NY: Plenum Press. → pages 29

Duff, P. (2008). Heritage language education in Canada. In D. Brinton, O. Kagan, & S. Bauckus (Eds.), *Heritage language education: A new field emerging* (pp. 70–91). New York, NY: Routledge. → pages 26

Duff, P. A. & Li, D. (2009). Indigenous, minority, and heritage language education in canada: Policies, contexts, and issues. *Canadian Modern Language Review*, *66*(1), 1–8. doi::10.3138/cmlr.66.1.001 → pages 26

Eilers, R. E., Wilson, W. R., & Moore, J. M. (1977). Developmental changes in speech discrimination in infants. *Journal of Speech, Language, and*

Hearing Research, *20*(4), 766–780. doi:10.1044/jshr.2004.766 → pages 54

Elder, C. (2005). Evaluating the effectiveness of heritage language education: What role for testing? *International Journal of Bilingual Education and Bilingualism*, *8*(2-3), 196–212. doi:10.1080/13670050508668607 → pages 26

Elder, C. (2009). Reconciling accountability and development needs in heritage language education: A communication challenge for the evaluation consultant. *Language Teaching Research*, *13*(1), 15–33. doi:10.1177/1362168808095521 → pages 26

Fishman, J. A. (2001). Three hundred-plus years of heritage language education in the United States. In J. K. Peyton, D. A. Ranard, & S. McGinnis (Eds.), *Heritage languages in America: Preserving a national resource* (pp. 81–97). McHenry, Illinois: Center for Applied Linguistics and Delta Systems Co. Inc. → pages 27

Fishman, J. A. (2014). Three hundred-plus years of heritage language education in the United States. In *Handbook of heritage, community, and native American languages in the United States* (pp. 50–58). New York, NY: Routledge. → pages 26

Flege, J. E. (1987). The production of "new" and "similar" phones in a foreign language: Evidence for the effect of equivalence classification. *Journal of Phonetics*, *15*(1), 47–65. → pages 35, 36, 37, 49

Flege, J. E. & Eefting, W. (1988). Imitation of a VOT continuum by native speakers of English and Spanish: Evidence for phonetic category formation. *The Journal of the Acoustical Society of America*, *83*(2), 729–740. doi:10.1121/1.396115 → pages 35

FluidSurveys (2017). UBC FluidSurveys. Retrieved February 20, 2017, from http://survey.ubc.ca → pages 86, 92, 108

Fok-Chan, Y.-Y. (1974). *A perceptual study of tones in Cantonese*. University of Hong Kong, Centre of Asian studies. → pages 57, 58, 79, 81, 98

Fox, R. A. & Qi, Y.-Y. (1990). Context effects in the perception of lexical tone. *Journal of Chinese Linguistics*, *18*(2), 261–284. → pages 56

Francis, A. L., Ciocca, V., Ma, L., & Fenn, K. (2008). Perceptual learning of Cantonese lexical tones by tone and non-tone language speakers. *Journal of Phonetics*, *36*(2), 268–294. doi:10.1016/j.wocn.2007.06.005 → pages 61, 62, 63, 81, 188, 189, 201

Francis, A. L., Ciocca, V., & Ng, B. K. C. (2003). On the (non)categorical perception of lexical tones. *Perception & Psychophysics*, *65*(7), 1029–1044. doi:10.3758/BF03194832 → pages 57, 58

Francis, A. L., Ciocca, V., Wong, N. K. Y., Leung, W. H. Y., & Chu, P. C. Y. (2006). Extrinsic context affects perceptual normalization of lexical tone. *The Journal of the Acoustical Society of America*, *119*(3), 1712–1726. doi:10.1121/1.2149768 → pages 56

Frank, M. C., Braginsky, M., Yurovsky, D., & Marchman, V. A. (2017). Wordbank: An open repository for developmental vocabulary data. *Journal of Child Language*, *44*(3), 677–694. doi:10.1017/S0305000916000209 → pages 82

Fulcher, G. (2014). *Testing second language speaking*. Milton Park, Abingdon, Oxon: Routledge. → pages 29

Fung, R. S. & Wong, C. S. (2011). Acoustic analysis of the new rising tone in Hong Kong Cantonese. *Proceedings of the 17th International Congress of Phonetic Sciences*, 716–718. Retrieved from https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2011/OnlineProceedings/RegularSession/Fung/Fung.pdf → pages 18, 59

Fung, R. S., Wong, C. S., & Law, S. (2011). The mechanism of rising tone merger in Hong Kong Cantonese: An acoustic approach. *Phonetics & Phonology In Iberia*. Retrieved from http://hdl.handle.net/10722/136295 → pages 59, 121

Fung, R. S.-Y. (2000). *Final particles in Standard Cantonese: Semantic extension and pragmatic inference*. PhD thesis, The Ohio State University. → pages 12

Gandour, J. (1981). Perceptual dimensions of tone: Evidence from Cantonese. *Journal of Chinese Linguistics*, 20–36. Retrieved from https://www.jstor.org/stable/23753516 → pages 57, 58

Gandour, J. T. (1978). The perception of tone. In V. A. Fromkin (Ed.), *Tone: A linguistic review* (pp. 41–76). New York, NY: Academic Press. → pages 52

Gandour, J. T. & Krishnan, A. (2015). Processing tone languages. In G. Hickok & S. Small (Eds.), *Neurobiology of language* (pp. 1095–1107). London, England: Academic Press. doi:10.1016/C2011-0-07351-9 → pages 52

García, O. (2005). Positioning heritage languages in the United States. *The Modern Language Journal*, *89*(4), 601–605. Retrieved from https://www.jstor.org/stable/3588631 → pages 26

Gauthier, K. & Genesee, F. (2011). Language development in internationally adopted children: A special case of early second language learning. *Child Development*, *82*(3), 887–901. doi:10.1111/j.1467-8624.2011.01578.x → pages 47

Gernsbacher, M. A. (1984). Resolving 20 years of inconsistent interactions between lexical familiarity and orthography, concreteness, and polysemy. *Journal of Experimental Psychology: General*, *113*(2), 256–281. doi:10.1037/0096-3445.113.2.256 → pages xii, 84

Gertken, L. M., Amengual, M., & Birdsong, D. (2014). Assessing language dominance with the Bilingual Language Profile. In P. Leclercq, A. Edmonds, & H. Hilton (Eds.), *Measuring L2 proficiency: Perspectives from SLA* (pp. 208–225). Multilingual Matters. → pages 29, 30, 31, 121

Gollan, T. H., Weissberger, G. H., Runnqvist, E., Montoya, R. I., & Cera, C. M. (2012). Self-ratings of spoken language dominance: A Multilingual Naming Test (MINT) and preliminary norms for young and aging Spanish–English bilinguals. *Bilingualism: Language and Cognition*, *15*(3), 594–615. doi:10.1017/S1366728911000332 → pages 118

Government of Canada. (1991). Canadian Heritage Languages Institute Act. Retrieved from http://laws-lois.justice.gc.ca/eng/acts/C-17.6/20050401/P1TT3xt3.html → pages 27

Government of Manitoba. (2018). Policy for heritage language instruction. Retrieved March 20, 2018, from http://www.edu.gov.mb.ca/k12/docs/policy/heritage/index.html → pages 27

Grosjean, F. (1982). *Life with two languages: An introduction to bilingualism*. Cambridge, MA: Harvard University Press. → pages 28

Grosjean, F. (1989). Neurolinguists, beware! The bilingual is not two monolinguals in one person. *Brain and Language*, *36*(1), 3–15. doi:10.1016/0093-934X(89)90048-5 → pages 36, 205

Grosjean, F. (1998). Studying bilinguals: Methodological and conceptual issues. *Bilingualism: Language and Cognition*, *1*(2), 131–149. doi:10.1017/S136672899800025X → pages 29, 30

Grosjean, F. (2001). The bilingual's language modes. In J. L. Nicol (Ed.), *One mind, two languages: Bilingual Language Processing* (pp. 1–22). Oxford, England: Blackwell. → pages 34

Gu, W. & Lee, T. (2007). Effects of tonal context and focus on Cantonese f0. *Proceedings of the 16th International Congress of Phonetic Sciences*, 1033–1036. Retrieved from http://www.icphs2007.de/conference/Papers/1689/1689.pdf → pages 56

Gui, M. C. (2005). *The phonology of Guangzhou Cantonese*. Munich, Germany: Lincom Europa. → pages 12

Harris, C. L., Gleason, J. B., & Aycicegi, A. (2006). When is a first language more emotional? Psychophysiological evidence from bilingual speakers. In A. Pavlenko (Ed.), *Bilingual education and bilingualism* (pp. 257–283). Multilingual Matters. → pages 30, 31

Hashimoto, O.-k. Y. (1972). *Studies in Yue dialects 1: Phonology of Cantonese*. New York, NY: Cambridge University Press. → pages 9, 10, 14

Holm, J. A. (1989). *Pidgins and creoles: Volume 2, Reference survey*. Cambridge, England: Cambridge University Press. → pages 1

Hornberger, N. H. (2005). Heritage/community language education: US and Australian perspectives. *International Journal of Bilingual Education and Bilingualism*, *8*(2-3), 101–108. doi:10.1080/13670050508668599 → pages 26

Hornberger, N. H. & Wang, S. C. (2008). Who are our heritage language learners? Identity and biliteracy in heritage language education in the United States. In D. Brinton, O. Kagan, & S. Bauckus (Eds.), *Heritage language education: A new field emerging* (pp. 3–35). New York, NY: Routledge. → pages 26

Hsiar, O. Y. (2007). Phonological elision in Malaysian Cantonese casual speech. Master's thesis, National University of Singapore, Singapore. → pages 119

Iacoponi, L. (2012). Synchronic and diachronic variation of Cantonese tone change in Optimality Theory. Master's thesis, Università degli Studi di Pisa, Italy. → pages 11

Kagan, O. & Dillon, K. (2001). A new perspective on teaching Russian: Focus on the heritage learner. *The Slavic and East European Journal*, *45*(3), 507–518. doi:10.2307/3086367 → pages 2

Kao, D. L. (1971). *Structure of the syllable in Cantonese*. The Hague: Mouton & Co. → pages 16

Kej, J., Smyth, V., So, L. K., Lau, C., & Capell, K. (2002). Assessing the accuracy of production of Cantonese lexical tones: a comparison between perceptual judgement and an instrumental measure. *Asia Pacific Journal of Speech, Language and Hearing*, *7*(1), 25–38. doi:10.1179/136132802805576535 → pages 59, 198

Kelleher, A. (2010). What is a heritage language? *Heritage Briefs*, 1–3. Retrieved from http://www.cal.org/heritage/pdfs/briefs/What-is-a-Heritage-Language.pdf → pages 2, 27

Khouw, E. & Ciocca, V. (2007). Perceptual correlates of Cantonese tones. *Journal of Phonetics*, *35*(1), 104–117. doi:10.1016/j.wocn.2005.10.003 → pages 57, 58, 81

Krashen, S. (2000). Bilingual education, the acquisition of English, and the retention and loss of Spanish. In A. Roca (Ed.), *Research on Spanish in the US: Linguistic Issues and Challenges* (pp. 432–444). Somerville, MA: Cascadilla Press. → pages 206

Kuhl, P. K., Stevens, E., Hayashi, A., Deguchi, T., Kiritani, S., & Iverson, P. (2006). Infants show a facilitation effect for native language phonetic perception between 6 and 12 months. *Developmental Science*, *9*(2), F13–F21. doi:10.1111/j.1467-7687.2006.00468.x → pages 204

Kung, C., Chwilla, D. J., & Schriefers, H. (2014). The interaction of lexical tone, intonation and semantic context in on-line spoken word recognition: An ERP study on Cantonese Chinese. *Neuropsychologia*, *53*, 293–309. doi:10.1016/j.neuropsychologia.2013.11.020 → pages 81

Kwok, B.-C., Chin, A. C., & Tsou, B. K. (2016). Grammatical diversity across the Yue dialects. *Journal of Chinese Linguistics*, *44*(1), 109–152. doi:10.1353/jcl.2016.0002 → pages 10

La Heij, W. (2005). Selection processes in monolingual and bilingual lexical access. In J. F. Kroll & A. M. B. De Groot (Eds.), *Handbook of bilingualism: Psycholinguistic approaches* (pp. 289–307). New York, NY: Oxford University Press. → pages 36

Ladd, D. R. (1984). Declination: A review and some hypotheses. *Phonology*, *1*, 53–74. doi:10.1017/S0952675700000294 → pages 17

Lai, C., Li, Z., & Gong, Y. (2016). Teacher agency and professional learning in cross-cultural teaching contexts: Accounts of Chinese teachers from international schools in Hong Kong. *Teaching and Teacher Education*, *54*, 12–21. doi:10.1016/j.tate.2015.11.007 → pages 118

Lam, Z. W.-M., Hall, K. C., & Pulleyblank, D. (2016). Temporal location of perceptual cues for Cantonese tone identification. In *The 3rd Workshop on Innovations in Cantonese Linguistics (WICL-3)*, Columbus, OH. The Ohio State University. → pages 58, 81

Law, I. K.-Y., Ma, E. P.-M., & Yiu, E. M.-L. (2009). Speech intelligibility, acceptability, and communication-related quality of life in Chinese alaryngeal speakers. *Archives of Otolaryngology–Head & Neck Surgery*, *135*(7), 704–711. doi:10.1001/archoto.2009.71 → pages 74

Law, S.-P., Fung, R. S.-Y., & Bauer, R. S. (2001). Perception and production of Cantonese consonant endings. *Asia Pacific Journal of Speech, Language and Hearing*, *6*(3), 179–195. doi:10.1179/136132801805576590 → pages 14

Law, S.-P., Fung, R. S.-Y., & Kung, C. (2013). An ERP study of good production vis-à-vis poor perception of tones in Cantonese: Implications for top-down speech processing. *PLoS One*, *8*(1), e54396. doi:10.1371/journal.pone.0054396 → pages 56

Lee, G. M. (1993). *Comparative, diachronic and experimental perspectives on the interaction between tone and vowel in Standard Cantonese*. PhD thesis, The Ohio State University. → pages 12

Lee, K. Y., Chan, K. T., Lam, J. H., Van Hasselt, C., & Tong, M. C. (2015). Lexical tone perception in native speakers of Cantonese. *International*

Journal of Speech-Language Pathology, *17*(1), 53–62. doi:10.3109/17549507.2014.898096 → pages 57

Lee, S. L. (2004). *History and current trends of teaching Cantonese as a foreign Language: Investigating approaches to teaching and learning Cantonese*. EdD thesis, University of Leicester. → pages 205

Lee, T., Lau, W., Wong, Y. W., & Ching, P. (2002). Using tone information in Cantonese continuous speech recognition. *ACM Transactions on Asian Language Information Processing*, *1*(1), 83–102. doi:10.1145/595576.595581 → pages 56

Lee, Y.-S., Vakoch, D. A., & Wurm, L. H. (1996). Tone perception in Cantonese and Mandarin: A cross-linguistic comparison. *Journal of Psycholinguistic Research*, *25*(5), 527–542. doi:10.1007/BF01758181 → pages 61

Lee, Y. S. K., Chiu, S. N., & van Hasselt, C. A. (2002). Tone perception ability of Cantonese-speaking children. *Language and Speech*, *45*(4), 387–406. doi:10.1177/00238309020450040401 → pages 54

Lei, M. K.-Y. (2007). Discrimination of level tones in Cantonese-learning infants. *Proceedings of the 16th International Congress of Phonetic Sciences*, 1313–1316. Retrieved from http://www.icphs2007.de/conference/Papers/1620/1620.pdf → pages 54

Levine, G. S. (2015). *Incomplete L1 acquisition in the immigrant situation: Yiddish in the United States*. Max Niemeyer Verlag Tübingen. → pages 200

Lewis, M. P., Simons, G. F., & Fennig, C. D. (2009). *Ethnologue: Languages of the world*, volume 16. SIL international. → pages 9

Li, P. S. (2005). The rise and fall of Chinese immigration to Canada: Newcomers from Hong Kong Special Administrative Region of China and Mainland china, 1980–2002. *International Migration*, *43*(3), 9–34. doi:10.1111/j.1468-2435.2005.00324.x → pages 2

Li, W. (2016, September 20). New spin on Chinese school focuses on Chinatown's Cantonese conversations. *Metronews Vancouver*. Retrieved August 5, 2017, from http://www.metronews.ca/news/vancouver/2016/09/20/chinese-school-in-chinatown-focuses-on-survival-cantonese.html → pages 5

Li, Y., Lee, T., & Qian, Y. (2002). Acoustical f0 analysis of continuous Cantonese speech. *Proceedings of the International Symposium on Chinese Spoken Language Processing*. Retrieved from https://www.isca-speech.org/archive_open/archive_papers/iscslp2002/clp2_072.pdf → pages 17, 56, 57, 91

Liberman, M. Y. (1975). *The intonational system of English.* PhD thesis, Massachusetts Institute of Technology. → pages 63, 188

Lieberman, P. (1960). Some acoustic correlates of word stress in American English. *The Journal of the Acoustical Society of America*, *32*(4), 451–454. doi:10.1121/1.1908095 → pages 63

Lieberman, P. (1966). *Intonation, perception, and language*. PhD thesis, Massachusetts Institute of Technology. → pages 17

Linck, J. A., Kroll, J. F., & Sunderman, G. (2009). Losing access to the native language while immersed in a second language: Evidence for the role of inhibition in second-language learning. *Psychological Science*, *20*(12), 1507–1515. Retrieved from https://www.jstor.org/stable/40575218 → pages 36, 49

Luke, K.-K. & Wong, M. L. (2015). The Hong Kong Cantonese Corpus: Design and uses. *Journal of Chinese Linguistics*, *25*(2015), 309–330. → pages xii, 84, 85

Lynch, A. (2003). The relationship between second and heritage language acquisition: Notes on research and theory building. Retrieved July 1, 2018, from http://international.ucla.edu/institute/article/3615 → pages 205

Ma, J. K.-Y., Ciocca, V., & Whitehill, T. L. (2006). Effect of intonation on Cantonese lexical tones. *The Journal of the Acoustical Society of America*, *120*(6), 3978–3987. doi10.1121/1.2363927 → pages 56

Ma, J. K.-Y., Ciocca, V., & Whitehill, T. L. (2011). The perception of intonation questions and statements in Cantonese. *The Journal of the Acoustical Society of America*, *129*(2), 1012–1023. doi:10.1121/1.3531840 → pages 36, 56

Macnamara, J. (1967). The bilingual's linguistic performance—A psychological overview. *Journal of Social Issues*, *23*(2), 58–77. doi:10.1111/j.1540-4560.1967.tb00576.x → pages 30

Mair, V. H. (1991). What is a Chinese “dialect/topolect”?: Reflections on some key Sino-English linguistic terms. *Sino-Platonic Papers*, *29*, 2–30. Retrieved from http://sino-platonic.org/complete/spp029_chinese_dialect.pdf → pages 13

Mantel, N. (1967). The detection of disease clustering and a generalized regression approach. *Cancer Research*, *27*(2 Part 1), 209–220. → pages xiii, 161, 164, 167

Marian, V. & Kaushanskaya, M. (2004). Self-construal and emotion in bicultural bilinguals. *Journal of Memory and Language*, *51*(2), 190–201. doi:10.1016/j.jml.2004.04.003 → pages 31

Matthews, S. & Yip, V. (2013). *Cantonese: A comprehensive grammar*. Milton Park, Abingdon, Oxon: Routledge. → pages 13, 15

Maurer, D. & Werker, J. F. (2014). Perceptual narrowing during infancy: A comparison of language and faces. *Developmental Psychobiology*, *56*(2), 154–178. doi:10.1002/dev.21177 → pages 203

McCutchen, D. & Perfetti, C. A. (1982). The visual tongue-twister effect: Phonological activation in silent reading. *Journal of Verbal Learning and Verbal Behavior*, *21*(6), 672–687. → pages 102

Ming, T. & Tao, H. (2008). Developing a Chinese heritage language corpus: Issues and a preliminary report. In A. W. He & Y. Xiao (Eds.), *Chinese as a heritage language: Fostering rooted world citizenry* (pp. 167–188). Honolulu, HI: University of Hawai’i, National Foreign Language Resource Center Honolulu. → pages 42

Mok, P. P.-K. & Wong, P. W.-Y. (2010). Perception of the merging tones in Hong Kong Cantonese: Preliminary data on monosyllables. *Proceedings of the 5th International Conference on Speech Prosody*. Retrieved from https://www.isca-speech.org/archive/sp2010/papers/sp10_916.pdf → pages 58, 60

Mok, P. P.-K., Zuo, D., & Wong, P. W.-Y. (2013). Production and perception of a sound change in progress: Tone merging in Hong Kong Cantonese. *Language Variation and Change*, *25*(3), 341–370. doi:10.1017/S0954394513000161 → pages 18, 59, 60, 65

Montrul, S. (2009). Knowledge of tense-aspect and mood in Spanish heritage speakers. *International Journal of Bilingualism*, *13*(2), 239–269. doi:10.1177/1367006909339816 → pages 42

Montrul, S. (2013). Bilingualism and the heritage language speaker. In T. K. Bhatia & W. C. Ritchie (Eds.), *The handbook of bilingualism* (pp. 168–189). Malden, MA: Blackwell. → pages 25, 42, 200

Montrul, S. A. (2008). *Incomplete Acquisition in Bilingualism: Re-examining the age factor*. Amsterdam: John Benjamins. → pages 200

Montrul, S. A. (2012). Is the heritage language like a second language? *Eurosla Yearbook*, *12*(1), 1–29. doi:10.1075/eurosla.12.03mon → pages xv, 7

Myers-Scotton, C. (2005). *Multiple voices: An introduction to bilingualism*. Malden, MA: Blackwell. → pages 29

Nagy, N. (2009). Heritage language variation and change. Retrieved August, 31, 2017, from http://projects.chass.utoronto.ca/ngn/HLVC → pages 41

Nagy, N. (2015). A sociolinguistic view of null subjects and VOT in Toronto heritage languages. *Lingua*, *164*, 309–327. doi:10.1016/j.lingua.2014.04.012 → pages 2, 41

Newman, R. L. & Connolly, J. F. (2004). Determining the role of phonology in silent reading using event-related brain potentials. *Cognitive Brain Research*, *21*(1), 94–105. doi:10.1016/j.cogbrainres.2004.05.006 → pages 102

Nygaard, L. C. & Pisoni, D. B. (1998). Talker-specific learning in speech perception. *Perception & Psychophysics*, *60*(3), 355–376. doi:10.3758/BF03206860 → pages 202

NYU College of Arts and Science. (2018). Language courses. Retrieved July 1, 2018, from https://as.nyu.edu/sca/current-students/language-courses.html → pages 205

Oh, J. S., Au, T. K.-f., & Jun, S.-A. (2010). Early childhood language memory in the speech perception of international adoptees. *Journal of Child Language*, *37*(05), 1123–1132. doi:10.1017/S0305000909990286 → pages 48, 49, 204

Oh, J. S., Jun, S.-A., Knightly, L. M., & Au, T. K.-f. (2003). Holding on to childhood language memory. *Cognition*, *86*(3), B53–B64. doi:10.1016/S0010-0277(02)00175-0 → pages 46, 49

Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlinn, D., Minchin, P. R., O'Hara, R. B., Simpson, G. L., Solymos, P., Henry, M., Stevens, H., Szoecs, E., & Wagner, H. (2018). vegan: Community ecology package. Retrieved from https//CRAN.R-project.org/package=vegan → pages 167

Olson, J. S. (1998). *An ethnohistorical dictionary of China*. Westport, CT: Greenwood Publishing Group. → pages 13

Ontario Ministry of Education. (1991). *Heritage languages: Kindergarten to Grade 8*. Toronto, ON: Queen's Printer. → pages 26

Ormsby, M. A. (1958). *British Columbia: A history*. Toronto, ON: Macmillan. → pages 4

OSU East Asian Studies Center. (2014). Focus on Cantonese: OSU's Cantonese program grows into language, area studies courses. Retrieved July 1, 2018, from https://cpb-us-w2.wpmucdn.com/u.osu.edu/dist/7/1615/files/2014/05/EASC-expOSUre-Sp2014-pp4-5-2hpy01c.pdf → pages 205

Ou, J. (2012). Tone merger in Guangzhou Cantonese. Master's thesis, The Hong Kong Polytechnic University, Hong Kong. → pages 119

Pai, R. (2016). Cantonese as a foreign language (CFL) curriculum design based on learner needs in North America. In *The 3rd Workshop on Innovations in Cantonese Linguistics (WICL-3)*, Columbus, OH. The Ohio State University. → pages 205

Peng, G. & Wang, W. S.-Y. (2005). Tone recognition of continuous Cantonese speech based on support vector machines. *Speech Communication*, *45*(1), 49–62. doi:10.1016/j.specom.2004.09.004 → pages 56

Perfetti, C. A., Bell, L. C., & Delaney, S. M. (1988). Automatic (prelexical) phonetic activation in silent word reading: Evidence from backward masking. *Journal of Memory and Language*, *27*(1), 59–70. doi:10.1016/0749-596X(88)90048-4 → pages 102

Peyton, J. K., Ranard, D. A., & McGinnis, S. (2001). Charting a new course: Heritage language education in the United States. In J. K. Peyton, D. A. Ranard, & S. McGinnis (Eds.), *Heritage languages in America: Preserving a national resource* (pp. 3–26). McHenry, Illinois: Center for Applied Linguistics and Delta Systems Co. Inc. → pages 2, 26, 27

Pierce, L. J., Klein, D., Chen, J.-K., Delcenserie, A., & Genesee, F. (2014). Mapping the unconscious maintenance of a lost first language. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(48), 17314–17319. doi:10.1073/pnas.1409411111 → pages 204

Pierrehumbert, J. B. (1980). *The phonology and phonetics of English intonation*. PhD thesis, Massachusetts Institute of Technology. → pages 63, 188

Polinsky, M. (2008). Relative clauses in heritage Russian: Fossilization or divergent grammar. *Formal Approaches to Slavic Linguistics #17: The Yale Meeting 2008*, 333–358. → pages 42

Polinsky, M. & Kagan, O. (2007). Heritage languages: In the 'wild' and in the classroom. *Language and Linguistics Compass*, *1*(5), 368–395. doi:10.1111/j.1749-818X.2007.00022.x → pages 25, 42, 200

Qin, Z. & Mok, P. P.-K. (2011). Perception of Cantonese tones by Mandarin, English and French speakers. *Proceedings of the 17th International Congress of Phonetic Sciences*, 1654–1657. Retrieved from http://www.phonetics.ucla.edu/voiceproject/Publications/Shue-etal_2011_ICPhS.pdf → pages 61, 62, 63, 189

Qin, Z. & Mok, P. P.-K. (2013). Discrimination of Cantonese tones by speakers of tone and non-tone languages. *Kansas Working Papers in Linguistics*, *34*. doi:10.17161/KWPL.1808.12864 → pages 61, 201

R Core Team (2013). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. → pages 130, 167

Rafat, Y., Mohaghegh, M., & Stevenson, R. (2017). Geminate attrition across three generations of Farsi-English bilinguals living in Canada: An acoustic study. *Ilha do Desterro*, *70*(3), 151–168. doi:10.5007/2175-8026.2017v70n3p151 → pages 47

Ramsey, S. R. (1987). *The languages of China*. Princeton, NJ: Princeton University Press. → pages 12

Rosch, E. H. (1973). On the internal structure of perceptual and semantic categories. In T. E. Moore (Ed.), *Cognitive development and acquisition of language* (pp. 111–144). New York, NY: Academic Press. doi:10.1016/B978-0-12-505850-6.50010-4 → pages 32

Rothman, J. & Treffers-Daller, J. (2014). A prolegomenon to the construct of the native speaker: Heritage speaker bilinguals are natives too! *Applied Linguistics*, *35*(1), 93–98. doi:10.1093/applin/amt049 → pages 33

Saadah, E. (2011). *The production of Arabic vowels by English L2 learners and heritage speakers of Arabic*. PhD thesis, University of Illinois at Urbana-Champaign. → pages 43, 49

Samuel, A. G. & Larraza, S. (2015). Does listening to non-native speech impair speech perception? *Journal of Memory and Language*, *81*, 51–71. doi:10.1016/j.jml.2015.01.003 → pages 38, 49, 205, 206

Sawilowsky, S. S. (2009). New effect size rules of thumb. *Journal of Modern Applied Statistical Methods*, *8*(2), 597–599. Retrieved from https://digitalcommons.wayne.edu/coe_tbf/4/ → pages xiii, 141

Schmeißer, A., Hager, M., Gil, L. A., Jansen, V., Geveler, J., Eichler, N., Patuto, M., & Müller, N. (2016). Related but different: The two concepts of language dominance and language proficiency. In J. Treffers-Daller & C. Silva-Corvalán (Eds.), *Language dominance in bilinguals: Issues of operationalization and measurement* (pp. 36–65). Cambridge, England: Cambridge University Press. → pages 31, 123

Schneider, E. W. (2007). *Postcolonial English: Varieties around the world*. New York, NY: Cambridge University Press. → pages 1

Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-Prime: User's guide*. Psychology Software Incorporated. → pages 103

Sebastián-Gallés, N., Echeverría, S., & Bosch, L. (2005). The influence of initial exposure on lexical representation: Comparing early and simultaneous bilinguals. *Journal of Memory and Language*, *52*(2), 240–255. doi:10.1016/j.jml.2004.11.001 → pages 40

Shepard, R. N. (1972). Psychological representation of speech sounds. In E. David & P. Denes (Eds.), *Human Communication: A Unified View* (pp. 67–113). New York, NY: McGraw-Hill. → pages 168, 171, 172, 173

Sherkina-Lieber, M., Pérez-Leroux, A. T., & Johns, A. (2011). Grammar without speech production: The case of Labrador Inuttitut heritage receptive bilinguals. *Bilingualism: Language and Cognition*, *14*(3), 301–317. doi:10.1017/S1366728910000210 → pages 42

Snow, D. (2004). *Cantonese as written language: The growth of a written Chinese vernacular*. Hong Kong: Hong Kong University Press. → pages 22

So, K. L. C. (2000). Tonal production and perception patterns of Canadian raised Cantonese speakers. Master's thesis, Simon Fraser University, Burnaby, Canada. → pages 63, 64, 65, 66, 81

Sona Systems Ltd. (2017). UBC Linguistics Sign-up System. Retrieved February 20, 2017, from https://ubclinguistics.sona-systems.com → pages 114

Soo, R. & Monahan, P. J. (2017). Language exposure modulates the role of tone in perception and long-term memory: Evidence from Cantonese native and heritage speakers. *Proceedings of the 43rd Annual Meeting of the Berkeley Linguistics Society*, *2*, 47–54. Retrieved from http://linguistics.berkeley.edu/bls/previous_proceedings/bls43_2.pdf → pages 64, 65

Stanford Language Center. (2018). Cantonese language program. Retrieved July 1, 2018, from https://cantonese.stanford.edu/courses → pages 205

Statistics Canada. (2009). Top 10 countries of birth of recent immigrants, 1981 to 2006. Retrieved February 13, 2017, from http://www12.statcan.ca/census-recensement/2006/as-sa/97-557/table/t1-eng.cfm → pages xi, 2, 3

Statistics Canada. (2012). Ethnic diversity of immigration. Retrieved February 13, 2017, from http://www.statcan.gc.ca/pub/11-402-x/2012000/pdf/ethnic-ethnique-eng.pdf → pages 2

Statistics Canada. (2017a). 2016 census: Immigrant languages in Canada. Retrieved August 5, 2017, from http://www.statcan.gc.ca/pub/11-627-m/11-627-m2017025-eng.htm → pages 2

Statistics Canada. (2017b). Census in brief: Linguistic diversity and multilingualism in Canadian homes. Retrieved September 1, 2017, from http://www12.statcan.gc.ca/census-recensement/2016/as-sa/98-200-x/2016010/98-200-x2016010-eng.cfm → pages xi, 3, 4

Statistics Canada. (2017c). An increasingly diverse linguistic profile: Corrected data from the 2016 Census. Retrieved July 1, 2018, from https://www150.statcan.gc.ca/n1/daily-quotidien/170817/dq170817a-eng.htm → pages xi, 3

Statistics Canada. (2017d). Previous standard - Visible minority. Retrieved September 1, 2017, from https://www.statcan.gc.ca/eng/concepts/definitions/previous/preminority → pages 4

Statistics Canada. (2017e). Proportion of mother tongue responses for various regions in Canada, 2016 Census. Retrieved August 5, 2018, from http://www12.statcan.gc.ca/census-recensement/2016/dp-pd/dv-vd/lang/index-eng.cfm → pages 1

Sweet, S. A. & Grace-Martin, K. (1999). *Data analysis with SPSS*. Boston, MA: Allyn & Bacon. → pages 133

Szeto, C. (2000). Testing intelligibility among Sinitic dialects. *Proceedings of ALS2K, the 2000 conference of the Australian Linguistic Society*. Retrieved from http://www.als.asn.au/proceedings/als2000/szeto.pdf → pages 10

Tan, C.-B. (2005). Chinese in Malaysia. In M. Ember, C. R. Ember, & I. Skoggard (Eds.), *Encyclopedia of Diasporas: Immigrant and Refugee Cultures Around the World* (pp. 697–706). New York, NY: Springer. → pages 9

Tang, K. (2015). *Naturalistic speech misperception*. PhD thesis, University College London. → pages xiii, 168, 174, 175

Tang, S.-W. & Cheng, S.-P. (2014). Aspects of Cantonese grammar. In C.-T. J. Huang, Y.-H. A. Li, & A. Simpson (Eds.), *The Handbook of Chinese Linguistics* (pp. 599–628). West Sussex, England: Blackwell. doi:10.1002/9781118584552.ch23 → pages 13

Tang, S.-W., Kwok, F., Lee, T. H.-T., Lun, C., Luke, K. K., Tung, P., & Cheung, K. H. (2002). Guide to LSHK Cantonese romanization of Chinese characters. Retrieved from https://www.lshk.org/jyutping → pages 20

Tardif, T., Fletcher, P., Liang, W., & Kaciroti, N. (2009). Early vocabulary development in Mandarin (Putonghua) and Cantonese. *Journal of Child*

Language, *36*(5), 1115–1144. doi:10.1017/S0305000908009185 → pages 82

Tees, R. C. & Werker, J. F. (1984). Perceptual flexibility: Maintenance or recovery of the ability to discriminate non-native speech sounds. *Canadian Journal of Psychology*, *38*(4), 579–590. doi:10.1037/h0080868 → pages 45, 46, 48, 49, 200, 203

Thorndike, E. L. & Lorge, I. (1963). *The Teacher's Word Book of 30000 Words*. New York, NY: Teachers College, Columbia University. → pages 84

Titze, I. R. (1994). *Principles of Voice Production*. Englewood Cliffs, NJ: Prentice Hall. → pages 51

To, Y.-m. & Lau, T.-y. (1995). Global export of Hong Kong television: Television Broadcasts Limited. *Asian Journal of Communication*, *5*(2), 108–121. doi:10.1080/01292989509364726 → pages 12

Tong, X., Lee, S. M. K., Lee, M. M. L., & Burnham, D. (2015). A tale of two features: Perception of Cantonese lexical tone and English lexical stress in Cantonese-English bilinguals. *PloS one*, *10*(11), e0142896. doi:10.1371/journal.pone.0142896 → pages 57

Tong, X., McBride, C., & Burnham, D. (2014). Cues for lexical tone perception in children: Acoustic correlates and phonetic context effects. *Journal of Speech, Language, and Hearing Research*, *57*(5), 1589–1605. doi:10.1044/2014_JSLHR-S-13-0145 → pages 98

Tsui, T.-H. (2012). Tonal variation in Hong Kong Cantonese acoustic distance and functional load. *Proceedings from the Annual Meeting of the Chicago Linguistic Society*, *48*(1), 579–588. → pages 199

UBC Department of Linguistics. (2014). Linguistics Outside the Classroom (LOC). Retrieved February 20, 2017, from http://linguistics.ubc.ca/undergrad/current-students/linguistics-outside-the-classroom-loc → pages 114

Valdés, G. (2001). Heritage language students: Profiles and possibilities. In Peyton, Joy Kreeft and Ranard, Donald A. and McGinnis, Scott (Ed.), *Heritage languages in America: Preserving a national resource* (pp. 37–80). McHenry, Illinois: Center for Applied Linguistics and Delta Systems Co. Inc. → pages xv, 25, 29, 30, 35, 40

Van Deusen-Scholl, N. (2003). Toward a definition of heritage language: Sociopolitical and pedagogical considerations. *Journal of Language, Identity, and Education*, *2*(3), 211–230. doi:10.1207/S15327701JLIE0203_4 → pages 25

Vance, T. J. (1976). An experimental investigation of tone and intonation in Cantonese. *Phonetica*, *33*(5), 368–392. doi:10.1159/000259793 → pages 56, 98

Vance, T. J. (1977). Tonal distinctions in Cantonese. *Phonetica*, *34*(2), 93–107. doi:10.1159/000259872 → pages 57, 58

von Békésy, G. (1960). *Experiments in hearing*. New York, NY: McGraw-Hill. → pages 52

Wei, L. & Lee, S. (2001). The use of Cantonese classifiers and quantifiers by young British-born Chinese in Tyneside. *International Journal of Bilingual Education and Bilingualism*, *14*(6), 359–382. → pages 42

Welsh Government and Welsh Language Commissioner. (2015). National survey for Wales, 2013-14: Welsh language use survey. Retrieved from http://www.comisiynyddygymraeg.cymru/English/Publications%20List/20150129%20DG%20S%20Welsh%20Language%20Use%20Survey%202013-14%20-%20Main%20report.pdf → pages 27

Werker, J. F., Gilbert, J. H., Humphrey, K., & Tees, R. C. (1981). Developmental aspects of cross-language speech perception. *Child Development*, 349–355. Retrieved from https://www.jstor.org/stable/1129249 → pages 46

Werker, J. F. & Hensch, T. K. (2015). Critical periods in speech perception: New directions. *Annual Review of Psychology*, *66*, 173–196. doi:0.1146/annurev-psych-010814-015104 → pages 203

Wiley, T. G. (2001). On defining heritage languages and their speakers. In Peyton, Joy Kreeft and Ranard, Donald A. and McGinnis, Scott (Ed.), *Heritage languages in America: Preserving a national resource* (pp. 29–36). McHenry, Illinois: Center for Applied Linguistics and Delta Systems Co. Inc. → pages 25

Witten, I. H. & Bell, T. C. (1991). The zero-frequency problem: Estimating the probabilities of novel events in adaptive text compression. *IEEE Transactions on Information Theory*, *37*(4), 1085–1094. doi:10.1109/18.87000 → pages 167, 168, 169, 170

Wolfram, W. & Schilling-Estes, N. (2003). Language change in “conservative” dialects: The case of past tense be in Southern enclave communities. *American Speech*, *78*(2), 209–228. → pages 60

Wong, C. S. P., Bauer, R. S., & Lam, Z. W. M. (2009). The integration of English loanwords in Hong Kong Cantonese. *Journal of the Southeast Asian Linguistics Society*, *1*, 251–266. Retrieved from http://hdl.handle.net/10397/5824 → pages 12

Wong, P. & Leung, C. T.-T. (2018). Suprasegmental features are not acquired early: Perception and production of monosyllabic Cantonese lexical tones in 4- to 6-year-old preschool children. *Journal of Speech, Language, and Hearing Research*, *61*(5), 1070–1085. Retrieved from https://jslhr.pubs.asha.org/article.aspx?articleid=2680411 → pages 55, 56, 200

Wong, P. C. & Diehl, R. L. (2003). Perceptual normalization for inter-and intratalker variation in Cantonese level tones. *Journal of Speech, Language, and Hearing Research*, *46*(2), 413–421. doi:10.1044/1092-4388(2003/034) → pages 56

Wong, S.-L. (1999). Deciding to stay, deciding to move, deciding not to decide. In G. G. Hamilton (Ed.), *Cosmopolitan capitalists: Hong Kong & the Chinese diaspora at the end of the 20th century* (pp. 135–151). Seattle, WA: University of Washington Press. → pages 2

Wong, Y. W. (2006). Contextual tonal variations and pitch targets in Cantonese. *Proceedings of the 3rd International Conference on Speech Prosody*, 317–320. Retrieved from https://www.isca-speech.org/archive/sp2006/papers/sp06_199.pdf → pages 17, 57, 58

Wong, Y. W. (2007). Production and perception of tones in Cantonese continuous speech. Master’s thesis, The Chinese University of Hong Kong. → pages 56

Wong, Y. W. (2011). *Sound changes in Hong Kong Cantonese: A multi-perspective study*. PhD thesis, The Chinese University of Hong Kong. → pages 56

Wurm, S. A., Li, R., & Baumann, T. (1987). *Language Atlas of China*. Australian Academy of the Humanities; Longman Group (Far East). → pages xv, 9, 11

Xiao, Y. (2006). Heritage learners in the Chinese language classroom: Home background. *Heritage Language Journal*, *4*(1), 47–56. → pages 21

Yeung, H. H., Chen, K. H., & Werker, J. F. (2013). When does native language input affect phonetic perception? The precocious case of lexical tone. *Journal of Memory and Language*, *68*(2), 123–139. doi:10.1016/j.jml.2012.09.004 → pages 53, 203

Yip, M. (2002). *Tone*. Cambridge, England: Cambridge University Press. → pages 16, 52

Yip, V. (2013). Simultaneous language acquisition. In F. Grosjean & P. Li (Eds.), *The psycholinguistics of bilingualism* (pp. 119–144). Oxford, England: Blackwell. → pages 48

Yip, V. & Matthews, S. (2007). *The bilingual child: Early development and language contact*. New York, NY: Cambridge University Press. → pages 29, 40

Yiu, E. M.-L. & Fok, A. Y.-Y. (1995). Lexical tone disruption in Cantonese aphasic speakers. *Clinical Linguistics & Phonetics*, *9*(1), 79–92. doi:10.3109/02699209508985326 → pages 81

Yiu, S. S. (2013). Cantonese tones and musical intervals. *Proceedings of the International Conference on Phonetics of the Languages in China, ICPLC-2013*, 155–158. Retrieved from https://hub.hku.hk/bitstream/10722/205625/1/Content.pdf → pages 57

Yu, A. C. L. (2007). Understanding near mergers: The case of morphological tone in Cantonese. *Phonology*, *24*(01), 187–214. doi:10.1017/S0952675707001157 → pages 12, 19

Yu, H. (2011). The intermittent rhythms of the Cantonese Pacific. In D. R. Gabaccía & D. Hoerder (Eds.), *Connecting Seas and Connected Ocean Rims* (pp. 393–414). Leiden, The Netherlands: Brill. → pages 4, 9, 10

Yu, K. M. (2017). The role of time in phonetic spaces: Temporal resolution in Cantonese tone perception. *Journal of Phonetics*, *65*, 126–144. doi:10.1016/j.wocn.2017.06.004 → pages 56

Yu, K. M. & Lam, H. W. (2014). The role of creaky voice in Cantonese tonal perception. *The Journal of the Acoustical Society of America*, *136*(3), 1320–1333. doi:10.1121/1.4887462 → pages 18, 57, 58, 81, 96

Zee, E. (1991). Chinese (Hong Kong Cantonese). *Journal of the International Phonetic Association*, *21*(1), 46–48. doi:10.1017/S0025100300006058 → pages 13, 14, 15, 19

Zee, E. (1999). Change and variation in the syllable-initial and syllable-final consonants in Hong Kong Cantonese. *Journal of Chinese Linguistics*, 120–167. → pages 14

Zhang, C., Peng, G., & Wang, W. S. (2011). Inter-talker variation as a source of confusion in Cantonese tone perception. *Proceedings of the 17th International Congress of the Phonetic Sciences*, 2276–2279. Retrieved from https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2011/OnlineProceedings/RegularSession/Zhang,%20Caicai/Zhang,%20Caicai.pdf → pages 56

Zhang, C., Peng, G., Wang, X., & Wang, W. S. (2015). Cumulative effects of phonetic context on speech perception. *Proceedings of the 18th International Congress of Phonetic Sciences*. Retrieved from https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2015/Papers/ICPHS0085.pdf → pages 56

Zhang, Y., Kuhl, P. K., Imada, T., Kotani, M., & Tohkura, Y. (2005). Effects of language experience: Neural commitment to language-specific auditory patterns. *NeuroImage*, *26*(3), 703–720. doi::10.1016/j.neuroimage.2005.02.040 → pages 204

Zheng, H., Peng, G., Tsang, P. W., & Wang, W. S. (2006). Perception of Cantonese level tones influenced by context position. *Proceedings of the 3rd International Conference on Speech Prosody*. Retrieved from https://www.isca-speech.org/archive/sp2006/papers/sp06_178.pdf → pages 56

Zhu, Y. (1987). *Analysis of cuing functions of the phonetic in modern China*. Unpublished manuscript, East China Normal University. (In Chinese). → pages 21

Zyzik, E. (2016). Toward a prototype model of the heritage language learner: Understanding strengths and needs. In M. Fairclough & S. M. Beaudrie (Eds.), *Innovative strategies for heritage language teaching: A practical guide for the classroom* (pp. 19–38). Washington, DC: Georgetown University Press. → pages 25, 32

Appendix A

Materials Used in the Experiment

This appendix contains materials used in the word-identification experiment described in Chapter 4. Target words listed in Section A.1 were selected based on results of Pilot Study 1 described in Section 4.2. Sentences listed in Section A.2 were selected based on results of Pilot Study 2 described in Section 4.3. Pictures presented in Section A.3 were discussed in Section 4.4.1.2. Instructions shown in Section A.4 were discussed in Section 4.4.1.4. Lastly, the story listening task in Section A.5 was described in Section 4.4.2.1.

A.1 Words

Table A.1 is a list of 27 target words.

Table A.1: Words used in the main study

Written form	Word	Meaning
分	*fan1*	share
粉	*fan2*	powder
瞓	*fan3*	sleep
墳	*fan4*	tomb
份	*fan6*	portion
呼	*fu1*	exhale
虎	*fu2*	tiger
褲	*fu3*	pants
扶	*fu4*	help by holding another person's arm
婦	*fu5*	woman
負	*fu6*	negative
醫	*ji1*	cure
椅	*ji2*	chair
兒	*ji4*	child
耳	*ji5*	ear
二	*ji6*	two
寫	*se2*	write
瀉	*se3*	diarrhea
蛇	*se4*	snake
社	*se5*	society
射	*se6*	shoot
獅	*si1*	lion
屎	*si2*	poop
試	*si3*	try
匙	*si4*	key
市	*si5*	market
士	*si6*	nurse/trained person

Table A.2 summarizes tonally contrastive quadruplets used in the main study, and shows why there the number of unique target words was 27.

Table A.2: Tonal quadruplets used in the current study (identical to Table 4.4)

Tone set	Syllable	T1	T2	T3	T4	T5	T6
1 2 3 4	*si*	lion	poop	try	key		
1 2 3 5	*fu*	exhale	tiger	pants		woman	
1 2 3 6	*fan*	share	powder	sleep			portion
1 2 4 5	*ji*	cure	chair		child	ear	
1 2 4 6	*si*	lion	poop		key		nurse/trained person
1 2 5 6	*si*	lion	poop			market	nurse/trained person
1 3 4 5	*fu*	exhale		pants	help	woman	
1 3 4 6	*fan*	share		sleep	tomb		portion
1 3 5 6	*fu*	exhale		pants		woman	negative
1 4 5 6	*ji*	cure			child	ear	two
2 3 4 5	*se*		write	diarrhea	snake	society	
2 3 4 6	*fan*		powder	sleep	tomb		portion
2 3 5 6	*se*		write	diarrhea		society	shoot
2 4 5 6	*ji*		chair		child	ear	two
3 4 5 6	*se*			diarrhea	snake	society	shoot

Total number of unique words: 27

Note that *fan5* “diligence”, *ji3* “idea”, and *se1* “some” were not used deliberately. For a detailed explanation of why they were avoided, see Section 4.2.

A.2 Sentences

The sentences below are grouped into sets by carrier phrase. In each set, all sentences share the same carrier phrase but differ in the lexical tone of the last word. Only one sentence in each set is semantically congruous with the carrier phrase. Incongruous sentences are marked with a pound sign #. Note that the *fu* and *si* series have six sentences (a, b, c, d, e, f) in each set, but the *fan*, *ji*, and *se* series only have five sentences (a, b, c, d, e) in each set. This is because *fan5* "diligence", *ji3* "idea", and *se1* "some" were not used for reasons stated in Section 4.2. In the main study, all sentences were randomized. For details, see Section 4.4.2.

(1) a.

	叫	好	多	碟	餸	大家	分
	giu3	*hou2*	*do1*	*dip6*	*sung3*	*daai6gaa1*	***fan1***
	order	very	many	plate	food	everyone	**share**

'(Let's) order many dishes for everyone to **share**.'

b.

#	叫	好	多	碟	餸	大家	粉
	giu3	*hou2*	*do1*	*dip6*	*sung3*	*daai6gaa1*	***fan2***
	order	very	many	plate	food	everyone	**powder**

'(Let's) order many dishes for everyone to **powder**.'

c.

#	叫	好	多	碟	餸	大家	瞓
	giu3	*hou2*	*do1*	*dip6*	*sung3*	*daai6gaa1*	***fan3***
	order	very	many	plate	food	everyone	**sleep**

'(Let's) order many dishes for everyone to **sleep**.'

d.

#	叫	好	多	碟	餸	大家	墳
	giu3	*hou2*	*do1*	*dip6*	*sung3*	*daai6gaa1*	***fan4***
	order	very	many	plate	food	everyone	**tomb**

'(Let's) order many dishes for everyone to **tomb**.'

e.

#	叫	好	多	碟	餸	大家	份
	giu3	*hou2*	*do1*	*dip6*	*sung3*	*daai6gaa1*	***fan6***
	order	very	many	plate	food	everyone	**portion**

'(Let's) order many dishes for everyone to **portion**.'

(2) a.

# 幫	啤啤	搽	啲	爽	身	分
bong1	*bi4 bi1*	*caa4*	*di1*	*song2*	*san1*	***fan1***
help	baby	apply	DET	dry	body	**share**

‘(Please) put some baby dry body **share** on the baby.’

b.

幫	啤啤	搽	啲	爽	身	粉
bong1	*bi4 bi1*	*caa4*	*di1*	*song2*	*san1*	***fan2***
help	baby	apply	DET	dry	body	**powder**

‘(Please) put some baby **powder** on the baby.’

c.

# 幫	啤啤	搽	啲	爽	身	瞓
bong1	*bi4 bi1*	*caa4*	*di1*	*song2*	*san1*	***fan3***
help	baby	apply	DET	dry	body	**sleep**

‘(Please) put some dry body **sleep** on the baby.’

d.

# 幫	啤啤	搽	啲	爽	身	墳
bong1	*bi4 bi1*	*caa4*	*di1*	*song2*	*san1*	***fan4***
help	baby	apply	DET	dry	body	**tomb**

‘(Please) put some dry body **tomb** on the baby.’

e.

# 幫	啤啤	搽	啲	爽	身	份
bong1	*bi4 bi1*	*caa4*	*di1*	*song2*	*san1*	***fan6***
help	baby	apply	DET	dry	body	**portion**

‘(Please) put some dry body **portion** on the baby.’

(3) a.

# 十	二	點	鐘	好	上	床	分
sap6	*ji6*	*dim2*	*zung1*	*hou2*	*soeng5*	*cong4*	***fan1***
ten	two	point	clock	better	up	bed	**share**

‘At twelve o’clock (you’d) better go to bed and **share**.’

b.

# 十	二	點	鐘	好	上	床	粉
sap6	*ji6*	*dim2*	*zung1*	*hou2*	*soeng5*	*cong4*	***fan2***
ten	two	point	clock	better	up	bed	**powder**

‘At twelve o’clock (you’d) better go to bed and **powder**.’

c. 十 二 點 鐘 好 上 床 瞓
sap6 ji6 dim2 zung1 hou2 soeng5 cong4 ***fan3***
ten two point clock better up bed **sleep**
'At twelve o'clock (you'd) better go to bed and **sleep**.'

d. # 十 二 點 鐘 好 上 床 墳
sap6 ji6 dim2 zung1 hou2 soeng5 cong4 ***fan4***
ten two point clock better up bed **tomb**
'At twelve o'clock (you'd) better go to bed and **tomb**.'

e. # 十 二 點 鐘 好 上 床 份
sap6 ji6 dim2 zung1 hou2 soeng5 cong4 ***fan6***
ten two point clock better up bed **portion**
'At twelve o'clock (you'd) better go to bed and **portion**.'

(4) a. # 有 啲 賊 仔 專 掘 山 分
jau5 di1 caak6 zai2 zyun1 gwat6 saan1 ***fan1***
have CL thief DIM specialize dig hill **share**
'There are some thieves who specialize in digging up **share** on the hills.'

b. # 有 啲 賊 仔 專 掘 山 粉
jau5 di1 caak6 zai2 zyun1 gwat6 saan1 ***fan2***
have CL thief DIM specialize dig hill **powder**
'There are some thieves who specialize in digging up **powder** on the hills.'

c. # 有 啲 賊 仔 專 掘 山 瞓
jau5 di1 caak6 zai2 zyun1 gwat6 saan1 ***fan3***
have CL thief DIM specialize dig hill **sleep**
'There are some thieves who specialize in digging up **sleep** on the hills.'

d. 有 啲 賊 仔 專 掘 山 墳
jau5 di1 caak6 zai2 zyun1 gwat6 saan1 ***fan4***
have CL thief DIM specialize dig hill **tomb**
'There are some thieves who specialize in digging up **tombs** on the hills.'

e. # 有 啲 賊 仔 專 掘 山 份
jau5 di1 caak6 zai2 zyun1 gwat6 saan1 ***fan6***
have CL thief DIM specialize dig hill **portion**
'There are some thieves who specialize in digging up **portion** on the hills.'

(5) a. # 你 幫 我 食 埋 我 嗰 分
nei5 bong1 ngo5 sik6 maai4 ngo5 go2 ***fan1***
2.SG help 1.SG eat ASP 1.SG DEM **share**
'You help me eat my **share**.'

b. # 你 幫 我 食 埋 我 嗰 粉
nei5 bong1 ngo5 sik6 maai4 ngo5 go2 ***fan2***
2.SG help 1.SG eat ASP 1.SG DEM **powder**
'You help me eat my **powder**.'

c. # 你 幫 我 食 埋 我 嗰 瞓
nei5 bong1 ngo5 sik6 maai4 ngo5 go2 ***fan3***
2.SG help 1.SG eat ASP 1.SG DEM **sleep**
'You help me eat my **sleep**.'

d. # 你 幫 我 食 埋 我 嗰 墳
nei5 bong1 ngo5 sik6 maai4 ngo5 go2 ***fan4***
2.SG help 1.SG eat ASP 1.SG DEM **tomb**
'You help me eat my **tomb**.'

e. 你 幫 我 食 埋 我 嗰 份

nei5 bong1 ngo5 sik6 maai4 ngo5 go2 ***fan6***

2.SG help 1.SG eat ASP 1.SG DEM **portion**

'You help me eat my **portion**.'

(6) a. 游 水 換 氣 要 慢 慢 呼

jau4 seoi2 wun6 hei3 jiu3 maan6 maan2 ***fu1***

swim water exchange air should slow slow **exhale**

'When (you) swim (you) should **exhale** slowly.'

b. # 游 水 換 氣 要 慢 慢 虎

jau4 seoi2 wun6 hei3 jiu3 maan6 maan2 ***fu2***

swim water exchange air should slow slow **tiger**

'When (you) swim (you) should **tiger** slowly.'

c. # 游 水 換 氣 要 慢 慢 褲

jau4 seoi2 wun6 hei3 jiu3 maan6 maan2 ***fu3***

swim water exchange air should slow slow **pant**

'When (you) swim (you) should **pant (as in trousers)** slowly.'

d. # 游 水 換 氣 要 慢 慢 扶

jau4 seoi2 wun6 hei3 jiu3 maan6 maan2 ***fu4***

swim water exchange air should slow slow **lift**

'When (you) swim (you) should **lift** (as in helping someone balance by holding his/her arm) slowly.'

e. # 游 水 換 氣 要 慢 慢 婦

jau4 seoi2 wun6 hei3 jiu3 maan6 maan2 ***fu5***

swim water exchange air should slow slow **woman**

'When (you) swim (you) should **woman** slowly.'

f. # 游 水 換 氣 要 慢 慢 負

jau4 seoi2 wun6 hei3 jiu3 maan6 maan2 ***fu6***

swim water exchange air should slow slow **negative**

'When (you) swim (you) should **negative** slowly.'

(7) a. # 動物園 有 兩 隻 老 呼
dung6mat6jyun4 jau5 loeng5 zek3 lou5 ***fu1***
zoo have two CL old **exhale**
'There are two old **exhale** in the zoo.'

b. 動物園 有 兩 隻 老 虎
dung6mat6jyun4 jau5 loeng5 zek3 lou5 ***fu2***
zoo have two CL old **tiger**
'There are two **tigers** in the zoo.'

c. # 動物園 有 兩 隻 老 褲
dung6mat6jyun4 jau5 loeng5 zek3 lou5 ***fu3***
zoo have two CL old **pant**
'There are two old **pants** in the zoo.'

d. # 動物園 有 兩 隻 老 扶
dung6mat6jyun4 jau5 loeng5 zek3 lou5 ***fu4***
zoo have two CL old **lift**
'There are two old **lift** (as in helping someone balance by holding his/her arm) in the zoo.'

e. # 動物園 有 兩 隻 老 婦
dung6mat6jyun4 jau5 loeng5 zek3 lou5 ***fu5***
zoo have two CL old **woman**
'There are two old **women** in the zoo.'

f. # 動物園 有 兩 隻 老 負
dung6mat6jyun4 jau5 loeng5 zek3 lou5 ***fu6***
zoo have two CL old **negative**
'There are two old **negatives** in the zoo.'

(8) a. # 佢 唔 鍾意 太 闊 嘅 呼
keoi5 m4 zung1ji3 taai3 fut3 ge3 ***fu1***
3.SG NEG like too wide REL **exhale**
'S/he does not like **exhale** that is too wide.'

b. # 佢 唔 鍾意 太 闊 嘅 虎
keoi5 m4 zung1ji3 taai3 fut3 ge3 ***fu2***
3.SG NEG like too wide REL **tiger**

'S/he does not like **tigers** that are too wide.'

c. 佢 唔 鍾意 太 闊 嘅 褲
keoi5 m4 zung1ji3 taai3 fut3 ge3 ***fu3***
3.SG NEG like too wide REL **pant**

'S/he does not like **pants** that are too loose.'

d. # 佢 唔 鍾意 太 闊 嘅 扶
keoi5 m4 zung1ji3 taai3 fut3 ge3 ***fu4***
3.SG NEG like too wide REL **lift**

'S/he does not like **lift** (as in helping someone balance by holding his/her arm) that is too wide.'

e. # 佢 唔 鍾意 太 闊 嘅 婦
keoi5 m4 zung1ji3 taai3 fut3 ge3 ***fu5***
3.SG NEG like too wide REL **woman**

'S/he does not like **women** that are too wide.'

f. # 佢 唔 鍾意 太 闊 嘅 負
keoi5 m4 zung1ji3 taai3 fut3 ge3 ***fu6***
3.SG NEG like too wide REL **negative**

'S/he does not like **negative** that is too wide.'

(9) a. # 婆婆 行 唔 到 要 人 呼
po4po2 haang4 m4 dou2 jiu3 jan4 ***fu1***
grandma walk NEG able need person **exhale**

'Grandma cannot walk and needs someone to **exhale**.'

b. # 婆婆 行 唔 到 要 人 虎
po4po2 haang4 m4 dou2 jiu3 jan4 ***fu2***
grandma walk NEG able need person **lift**

'Grandma cannot walk and needs someone **tiger**.'

c. # 婆婆 行 唔 到 要 人 褲
po4po2 haang4 m4 dou2 jiu3 jan4 ***fu3***
grandma walk NEG able need person **pant**
'Grandma cannot walk and needs someone's **pants**.'

d. 婆婆 行 唔 到 要 人 扶
po4po2 haang4 m4 dou2 jiu3 jan4 ***fu4***
grandma walk NEG able need person **lift**
'Grandma cannot walk and needs someone's **lift** (as in helping someone balance by holding his/her arm).'

e. # 婆婆 行 唔 到 要 人 婦
po4po2 haang4 m4 dou2 jiu3 jan4 ***fu5***
grandma walk NEG able need person **woman**
'Grandma cannot walk and needs someone's **woman**.'

f. # 婆婆 行 唔 到 要 人 負
po4po2 haang4 m4 dou2 jiu3 jan4 ***fu6***
grandma walk NEG able need person **negative**
'Grandma cannot walk and needs someone's **negative**.'

(10) a. # 要 做 個 有 錢 嘅 貴 呼
jiu3 zou6 go3 jau5 cin2 ge3 gwai3 ***fu1***
need be CL have money GE3 elegant **exhale**
'(I) need to be a rich, elegant **exhale**.'

b. # 要 做 個 有 錢 嘅 貴 虎
jiu3 zou6 go3 jau5 cin2 ge3 gwai3 ***fu2***
need be CL have money GE3 elegant **tiger**
'(I) need to be a rich, elegant **tiger**.'

c. # 要 做 個 有 錢 嘅 貴 褲
jiu3 zou6 go3 jau5 cin2 ge3 gwai3 ***fu3***
need be CL have money GE3 elegant **pant**
'(I) need to be a rich, elegant **pant**.'

d. # 要 做 個 有 錢 嘅 貴 扶
jiu3 zou6 go3 jau5 cin2 ge3 gwai3 ***fu4***
need be CL have money GE3 elegant **lift**
'(I) need to be a rich, elegant **lift** (as in helping someone balance by holding his/her arm).'

e. 要 做 個 有 錢 嘅 貴 婦
jiu3 zou6 go3 jau5 cin2 ge3 gwai3 ***fu5***
need be CL have money GE3 elegant **woman**
'(I) need to be a rich, elegant **woman**.'

f. # 要 做 個 有 錢 嘅 貴 負
jiu3 zou6 go3 jau5 cin2 ge3 gwai3 ***fu6***
need be CL have money GE3 elegant **negative**
'(I) need to be a rich, elegant **negative**.'

(11) a. # 磁 場 都 有 分 正 同 呼
ci4 coeng4 dou1 jau5 fan1 zing3 tung4 ***fu1***
magnetic field also have divide positive and **exhale**
'There are positive and **exhale** magnetic fields.'

b. # 磁 場 都 有 分 正 同 虎
ci4 coeng4 dou1 jau5 fan1 zing3 tung4 ***fu2***
magnetic field also have divide positive and **tiger**
'There are positive and **tiger** magnetic fields.'

c. # 磁 場 都 有 分 正 同 褲
ci4 coeng4 dou1 jau5 fan1 zing3 tung4 ***fu3***
magnetic field also have divide positive and **pant**
'There are positive and **pant** magnetic fields.'

d. # 磁 場 都 有 分 正 同 扶
ci4 coeng4 dou1 jau5 fan1 zing3 tung4 ***fu4***
magnetic field also have divide positive and **lift**

‘There are positive and **lift** (as in helping someone balance by holding his/her arm) magnetic fields.’

e. # 磁 場 都 有 分 正 同 婦
ci4 coeng4 dou1 jau5 fan1 zing3 tung4 ***fu5***
magnetic field also have divide positive and **woman**

‘There are positive and **woman** magnetic fields.’

f. 磁 場 都 有 分 正 同 負
ci4 coeng4 dou1 jau5 fan1 zing3 tung4 ***fu6***
magnetic field also have divide positive and **negative**

‘There are positive and **negative** magnetic fields.’

(12) a. 呢 種 怪 病 係 冇 得 醫
lei1 zung2 gwaai3 beng6 hai6 mou5 dak1 ***ji1***
DEM type strange disease be NEG able **cure**

‘There is no **cure** for this type of strange disease.’

b. # 呢 種 怪 病 係 冇 得 椅
lei1 zung2 gwaai3 beng6 hai6 mou5 dak1 ***ji2***
DEM type strange disease be NEG able **chair**

‘There is no **chair** for this type of strange disease.’

c. # 呢 種 怪 病 係 冇 得 兒
lei1 zung2 gwaai3 beng6 hai6 mou5 dak1 ***ji4***
DEM type strange disease be NEG able **child**

‘There is no **child** for this type of strange disease.’

d. # 呢 種 怪 病 係 冇 得 耳
lei1 zung2 gwaai3 beng6 hai6 mou5 dak1 ***ji5***
DEM type strange disease be NEG able **ear**

‘There is no **ear** for this type of strange disease.’

e. # 呢 種 怪 病 係 冇 得 二
lei1 zung2 gwaai3 beng6 hai6 mou5 dak1 ***ji6***
DEM type strange disease be NEG able **two**
'There is no **two** for this type of strange disease.'

(13) a. # 媽咪 話 要 坐 嗰 張 醫
maa1mi4 waa6 jiu3 co5 go2 zoeng1 ***ji1***
Mommy say want sit DEM CL **cure**
'Mommy said she wanted to sit on that **cure**.'

b. 媽咪 話 要 坐 嗰 張 椅
maa1mi4 waa6 jiu3 co5 go2 zoeng1 ***ji2***
Mommy say want sit DEM CL **chair**
'Mommy said she wanted to sit on that **chair**.'

c. # 媽咪 話 要 坐 嗰 張 兒
maa1mi4 waa6 jiu3 co5 go2 zoeng1 ***ji4***
Mommy say want sit DEM CL **child**
'Mommy said she wanted to sit on that **child**.'

d. # 媽咪 話 要 坐 嗰 張 耳
maa1mi4 waa6 jiu3 co5 go2 zoeng1 ***ji5***
Mommy say want sit DEM CL **ear**
'Mommy said she wanted to sit on that **ear**.'

e. # 媽咪 話 要 坐 嗰 張 二
maa1mi4 waa6 jiu3 co5 go2 zoeng1 ***ji6***
Mommy say want sit DEM CL **two**
'Mommy said she wanted to sit on that **two**.'

(14) a. # 兩 歲 以下 就 算 幼 醫
loeng5 seoi3 ji5haa6 zau6 syun3 jau3 ***ji1***
two year below then consider young **cure**
'(Babies) below two years of age are considered young **cure**.'

b. #

兩	歲	以下	就	算	幼	椅
loeng5	*seoi3*	*ji5haa6*	*zau6*	*syun3*	*jau3*	***ji2***
two	year	below	then	consider	young	**chair**

‘(Babies) below two years of age are considered young **chairs**.’

c.

兩	歲	以下	就	算	幼	兒
loeng5	*seoi3*	*ji5haa6*	*zau6*	*syun3*	*jau3*	***ji4***
two	year	below	then	consider	young	**child**

‘(Babies) below two years of age are considered **infants**.’

d. #

兩	歲	以下	就	算	幼	耳
loeng5	*seoi3*	*ji5haa6*	*zau6*	*syun3*	*jau3*	***ji5***
two	year	below	then	consider	young	**ear**

‘(Babies) below two years of age are considered young **ears**.’

e. #

兩	歲	以下	就	算	幼	二
loeng5	*seoi3*	*ji5haa6*	*zau6*	*syun3*	*jau3*	***ji6***
two	year	below	then	consider	young	**two**

‘(Babies) below two years of age are considered young **two**.’

(15) a. #

小	心	清潔	對	眼	同	醫
siu2	*sam1*	*cing1git3*	*deoi3*	*ngaan5*	*tung4*	***ji1***
small	heart	clean	pair	eye	and	**cure**

‘Clean (your) eyes and **cure** carefully.’

b. #

小	心	清潔	對	眼	同	椅
siu2	*sam1*	*cing1git3*	*deoi3*	*ngaan5*	*tung4*	***ji2***
small	heart	clean	pair	eye	and	**chair**

‘Clean (your) eyes and **chairs** carefully.’

c. #

小	心	清潔	對	眼	同	兒
siu2	*sam1*	*cing1git3*	*deoi3*	*ngaan5*	*tung4*	***ji4***
small	heart	clean	pair	eye	and	**child**

‘Clean (your) eyes and **children** carefully.’

d. 小 心 清潔 對 眼 同 耳
siu2 sam1 cing1git3 deoi3 ngaan5 tung4 ***ji5***
small heart clean pair eye and **ear**
'Clean (your) eyes and **ears** carefully.'

e. # 小 心 清潔 對 眼 同 二
siu2 sam1 cing1git3 deoi3 ngaan5 tung4 ***ji6***
small heart clean pair eye and **two**
'Clean (your) eyes and **two** carefully.'

(16) a. # 一 加 一 結果 等如 醫
jat1 gaa1 jat1 git3gwo2 dang2jyu4 ***ji1***
one add one result equal **cure**
'One plus one equals **cure**.'

b. # 一 加 一 結果 等如 椅
jat1 gaa1 jat1 git3gwo2 dang2jyu4 ***ji2***
one add one result equal **chair**
'One plus one equals **chair**.'

c. # 一 加 一 結果 等如 兒
jat1 gaa1 jat1 git3gwo2 dang2jyu4 ***ji4***
one add one result equal **child**
'One plus one equals **child**.'

d. # 一 加 一 結果 等如 耳
jat1 gaa1 jat1 git3gwo2 dang2jyu4 ***ji5***
one add one result equal **ear**
'One plus one equals **ear**.'

e. 一 加 一 結果 等如 二
jat1 gaa1 jat1 git3gwo2 dang2jyu4 ***ji6***
one add one result equal **two**
'One plus one equals **two**.'

(17) a. 我 有 好 多 字 唔 識 寫
ngo5 jau5 hou2 do1 zi6 m4 sik1 ***se2***
1.SG have very many word NEG know **write**
'There are many words that I don't know how to **write**.'

b. # 我 有 好 多 字 唔 識 瀉
ngo5 jau5 hou2 do1 zi6 m4 sik1 ***se3***
1.SG have very many word NEG know **diarrhea**
'There are many words that I don't know how to **diarrhea**.'

c. # 我 有 好 多 字 唔 識 蛇
ngo5 jau5 hou2 do1 zi6 m4 sik1 ***se4***
1.SG have very many word NEG know **snake**
'There are many words that I don't know how to **snake**.'

d. # 我 有 好 多 字 唔 識 社
ngo5 jau5 hou2 do1 zi6 m4 sik1 ***se5***
1.SG have very many word NEG know **society**
'There are many words that I don't know how to **society**.'

e. # 我 有 好 多 字 唔 識 射
ngo5 jau5 hou2 do1 zi6 m4 sik1 ***se6***
1.SG have very many word NEG know **shoot**
'There are many words that I don't know how to **shoot**.'

(18) a. # 前 日 我 食 錯 嘢 肚 寫
cin4 jat6 ngo5 sik6 co3 je5 tou5 ***se2***
before day 1.SG eat wrong thing belly **write**
'The day before yesterday I ate something bad and got belly **write**.'

b. 前 日 我 食 錯 嘢 肚 瀉
cin4 jat6 ngo5 sik6 co3 je5 tou5 ***se3***
before day 1.SG eat wrong thing belly **diarrhea**
'The day before yesterday I ate something bad and got **diarrhea**.'

c. # 前 日 我 食 錯 嘢 肚 **蛇**
cin4 jat6 ngo5 sik6 co3 je5 tou5 ***se4***
before day 1.SG eat wrong thing belly **snake**

'The day before yesterday I ate something bad and got belly **snake**.'

d. # 前 日 我 食 錯 嘢 肚 **社**
cin4 jat6 ngo5 sik6 co3 je5 tou5 ***se5***
before day 1.SG eat wrong thing belly **society**

'The day before yesterday I ate something bad and got belly **society**.'

e. # 前 日 我 食 錯 嘢 肚 **射**
cin4 jat6 ngo5 sik6 co3 je5 tou5 ***se6***
before day 1.SG eat wrong thing belly **shoot**

'The day before yesterday I ate something bad and got belly **shoot**.'

(19) a. # 行 山 時 小 心 有 毒 **寫**
haang4 saan1 si4 siu2 sam1 jau5 duk6 ***se2***
walk hill time small heart have poison **write**

'Beware of poisonous **write** when (you) go hiking.'

b. # 行 山 時 小 心 有 毒 **瀉**
haang4 saan1 si4 siu2 sam1 jau5 duk6 ***se3***
walk hill time small heart have poison **diarrhea**

'Beware of poisonous **diarrhea** when (you) go hiking.'

c. 行 山 時 小 心 有 毒 **蛇**
haang4 saan1 si4 siu2 sam1 jau5 duk6 ***se4***
walk hill time small heart have poison **snake**

'Beware of poisonous **snakes** when (you) go hiking.'

d. # 行 山 時 小 心 有 毒 社

行	山	時	小	心	有	毒	社
haang4	*saan1*	*si4*	*siu2*	*sam1*	*jau5*	*duk6*	***se5***
walk	hill	time	small	heart	have	poison	**society**

‘Beware of poisonous **society** when (you) go hiking.’

e. #

行	山	時	小	心	有	毒	射
haang4	*saan1*	*si4*	*siu2*	*sam1*	*jau5*	*duk6*	***se6***
walk	hill	time	small	heart	have	poison	**shoot**

‘Beware of poisonous **shoot** when (you) go hiking.’

(20) a. #

我	公公	有	個	粵	劇	寫
ngo5	*gung1gung1*	*jau5*	*go3*	*jyut6*	*kek6*	***se2***
1.SG	grandpa	have	CL	Cantonese	opera	**write**

‘My grandpa has a Cantonese opera **write**.’

b. #

我	公公	有	個	粵	劇	瀉
ngo5	*gung1gung1*	*jau5*	*go3*	*jyut6*	*kek6*	***se3***
1.SG	grandpa	have	CL	Cantonese	opera	**diarrhea**

‘My grandpa has a Cantonese opera **diarrhea**.’

c. #

我	公公	有	個	粵	劇	蛇
ngo5	*gung1gung1*	*jau5*	*go3*	*jyut6*	*kek6*	***se4***
1.SG	grandpa	have	CL	Cantonese	opera	**snake**

‘My grandpa has a Cantonese opera **snake**.’

d.

我	公公	有	個	粵	劇	社
ngo5	*gung1gung1*	*jau5*	*go3*	*jyut6*	*kek6*	***se5***
1.SG	grandpa	have	CL	Cantonese	opera	**society**

‘My grandpa has a Cantonese opera **society**.’

e. #

我	公公	有	個	粵	劇	射
ngo5	*gung1gung1*	*jau5*	*go3*	*jyut6*	*kek6*	***se6***
1.SG	grandpa	have	CL	Cantonese	opera	**shoot**

‘My grandpa has a Cantonese opera **shoot**.’

(21) a. # 開 槍 要 對 準 目標 寫

hoi1 coeng1 jiu3 deoi3 zeon2 muk6biu1 ***se2***

open gun need aim accurate target **write**

'(You) should aim at the target and **write** when using a gun.'

b. # 開 槍 要 對 準 目標 瀉

hoi1 coeng1 jiu3 deoi3 zeon2 muk6biu1 ***se3***

open gun need aim accurate target **diarrhea**

'(You) should aim at the target and **diarrhea** when using a gun.'

c. # 開 槍 要 對 準 目標 蛇

hoi1 coeng1 jiu3 deoi3 zeon2 muk6biu1 ***se4***

open gun need aim accurate target **snake**

'(You) should aim at the target and **snake** when using a gun.'

d. # 開 槍 要 對 準 目標 社

hoi1 coeng1 jiu3 deoi3 zeon2 muk6biu1 ***se5***

open gun need aim accurate target **society**

'(You) should aim at the target and **society** when using a gun.'

e. 開 槍 要 對 準 目標 射

hoi1 coeng1 jiu3 deoi3 zeon2 muk6biu1 ***se6***

open gun need aim accurate target **shoot**

'(You) should aim at the target and **shoot** when using a gun.'

(22) a. 過年 唐人街 有 舞 獅

gwo3nin4 tong4jan4gaai1 jau5 mou5 ***si1***

Lunar New Year Chinatown have dance **lion**

'There is **lion** dance in Chinatown during the Lunar New Year.'

b. # 過年 唐人街 有 舞 屎

gwo3nin4 tong4jan4gaai1 jau5 mou5 ***si2***

Lunar New Year Chinatown have dance **poop**

'There is **poop** dance in Chinatown during the Lunar New Year.'

c. # 過年 唐人街 有 舞 試

gwo3nin4 tong4jan4gaai1 jau5 mou5 ***si3***

Lunar New Year Chinatown have dance **try**

'There is **try** dance in Chinatown during the Lunar New Year.'

d. # 過年 唐人街 有 舞 匙

gwo3nin4 tong4jan4gaai1 jau5 mou5 ***si4***

Lunar New Year Chinatown have dance **key**

'There is **key** dance in Chinatown during the Lunar New Year.'

e. # 過年 唐人街 有 舞 市

gwo3nin4 tong4jan4gaai1 jau5 mou5 ***si5***

Lunar New Year Chinatown have dance **market**

'There is **market** dance in Chinatown during the Lunar New Year.'

f. # 過年 唐人街 有 舞 士

gwo3nin4 tong4jan4gaai1 jau5 mou5 ***si6***

Lunar New Year Chinatown have dance **trained person**

'There is **trained person** dance in Chinatown during the Lunar New Year.'

(23) a. # 放 狗 記住 執 番 啲 獅

fong3 gau2 gei3zyu6 zap1 faan1 di1 ***si1***

release dog remember pick up back DET **lion**

'When you walk your dog, remember to pick up the **lion**.'

b. 放 狗 記住 執 番 啲 屎

fong3 gau2 gei3zyu6 zap1 faan1 di1 ***si2***

release dog remember pick up back DET **poop**

'When you walk your dog, remember to pick up the **poop**.'

c. # 放 狗 記住 執 番 啲 試

fong3 gau2 gei3zyu6 zap1 faan1 di1 ***si3***

release dog remember pick up back DET **try**

'When you walk your dog, remember to pick up the **try**.'

d. # 放 狗 記住 執 番 啲 匙

fong3 gau2 gei3zyu6 zap1 faan1 di1 ***si4***

release dog remember pick up back DET **key**

'When you walk your dog, remember to pick up the **key**.'

e. # 放 狗 記住 執 番 啲 市

fong3 gau2 gei3zyu6 zap1 faan1 di1 ***si5***

release dog remember pick up back DET **market**

'When you walk your dog, remember to pick up the **market**.'

f. # 放 狗 記住 執 番 啲 士

fong3 gau2 gei3zyu6 zap1 faan1 di1 ***si6***

release dog remember pick up back DET **trained person**

'When you walk your dog, remember to pick up the **trained person**.'

(24) a. # 今 次 唔 得 下 次 再 獅

gam1 ci3 m4 dak1 haa6 ci3 zoi3 ***si1***

this time NEG okay next time again **lion**

'If you fail this time, **lion** again next time.'

b. # 今 次 唔 得 下 次 再 屎

gam1 ci3 m4 dak1 haa6 ci3 zoi3 ***si2***

this time NEG okay next time again **poop**

'If you fail this time, **poop** again next time.'

c. 今 次 唔 得 下 次 再 試

gam1 ci3 m4 dak1 haa6 ci3 zoi3 ***si3***

this time NEG okay next time again **try**

'If you fail this time, **try** again next time.'

d. # 今 次 唔 得 下 次 再 匙

gam1 ci3 m4 dak1 haa6 ci3 zoi3 ***si4***

this time NEG okay next time again **key**

'If you fail this time, **key** again next time.'

e. # 今 次 唔 得 下 次 再 市

gam1	*ci3*	*m4*	*dak1*	*haa6*	*ci3*	*zoi3*	***si5***
this	time	NEG	okay	next	time	again	**market**

'If you fail this time, **market** again next time.'

f. # 今 次 唔 得 下 次 再 士

gam1	*ci3*	*m4*	*dak1*	*haa6*	*ci3*	*zoi3*	***si6***
this	time	NEG	okay	next	time	again	**trained person**

'If you fail this time, **trained person** again next time.'

(25) a. # 出 門 口 記得 帶 鎖 獅

ceot1	*mun4*	*hau2*	*gei3dak1*	*daai3*	*so2*	***si1***
exit	door	mouth	remember	bring	lock	**lion**

'Remember to bring lock **lion** when leaving the house.'

b. # 出 門 口 記得 帶 鎖 屎

ceot1	*mun4*	*hau2*	*gei3dak1*	*daai3*	*so2*	***si2***
exit	door	mouth	remember	bring	lock	**poop**

'Remember to bring lock **poop** when leaving the house.'

c. # 出 門 口 記得 帶 鎖 試

ceot1	*mun4*	*hau2*	*gei3dak1*	*daai3*	*so2*	***si3***
exit	door	mouth	remember	bring	lock	**try**

'Remember to bring lock **try** when leaving the house.'

d. 出 門 口 記得 帶 鎖 匙

ceot1	*mun4*	*hau2*	*gei3dak1*	*daai3*	*so2*	***si4***
exit	door	mouth	remember	bring	lock	**key**

'Remember to bring (your) **key** when leaving the house.'

e. # 出 門 口 記得 帶 鎖 市

ceot1	*mun4*	*hau2*	*gei3dak1*	*daai3*	*so2*	***si5***
exit	door	mouth	remember	bring	lock	**market**

'Remember to bring lock **market** when leaving the house.'

f. # 出 門 口 記得 帶 鎖 士

ceot1 mun4 hau2 gei3dak1 daai3 so2 ***si6***

exit door mouth remember bring lock **trained person**

'Remember to bring lock **trained person** when leaving the house.'

(26) a. # 啲 經紀 睇 實 個 股 獅

di1 ging1gei2 tai2 sat6 go3 gu2 ***si1***

DET broker watch closely CL stock **lion**

'The brokers are watching the stock **lion** closely.'

b. # 啲 經紀 睇 實 個 股 屎

di1 ging1gei2 tai2 sat6 go3 gu2 ***si2***

DET broker watch closely CL stock **poop**

'The brokers are watching the stock **poop** closely.'

c. # 啲 經紀 睇 實 個 股 試

di1 ging1gei2 tai2 sat6 go3 gu2 ***si3***

DET broker watch closely CL stock **try**

'The brokers are watching the stock **try** closely.'

d. # 啲 經紀 睇 實 個 股 匙

di1 ging1gei2 tai2 sat6 go3 gu2 ***si4***

DET broker watch closely CL stock **key**

'The brokers are watching the stock **key** closely.'

e. 啲 經紀 睇 實 個 股 市

di1 ging1gei2 tai2 sat6 go3 gu2 ***si5***

DET broker watch closely CL stock **market**

'The brokers are watching the stock **market** closely.'

f. # 啲 經紀 睇 實 個 股 士

di1 ging1gei2 tai2 sat6 go3 gu2 ***si6***

DET broker watch closely CL stock **trained person**

'The brokers are watching the stock **trained person** closely.'

(27) a. # 手術 室 入 面 有 護 獅
sau2seot6 sat1 jap6 min6 jau5 wu6 ***si1***
operation room in side have care **lion**
'There are care **lions** in the operating theatre.'

b. # 手術 室 入 面 有 護 屎
sau2seot6 sat1 jap6 min6 jau5 wu6 ***si2***
operation room in side have care **poop**
'There is care **poop** in the operating theatre.'

c. # 手術 室 入 面 有 護 試
sau2seot6 sat1 jap6 min6 jau5 wu6 ***si3***
operation room in side have care **try**
'There is care **try** in the operating theatre.'

d. # 手術 室 入 面 有 護 匙
sau2seot6 sat1 jap6 min6 jau5 wu6 ***si4***
operation room in side have care **key**
'There are care **keys** in the operating theatre.'

e. # 手術 室 入 面 有 護 市
sau2seot6 sat1 jap6 min6 jau5 wu6 ***si5***
operation room in side have care **market**
'There are care **markets** in the operating theatre.'

f. 手術 室 入 面 有 護 士
sau2seot6 sat1 jap6 min6 jau5 wu6 ***si6***
operation room in side have care **trained person**
'There are **nurses** in the operating theatre.'

A.3 Pictures

The following pictures were used in the main study.

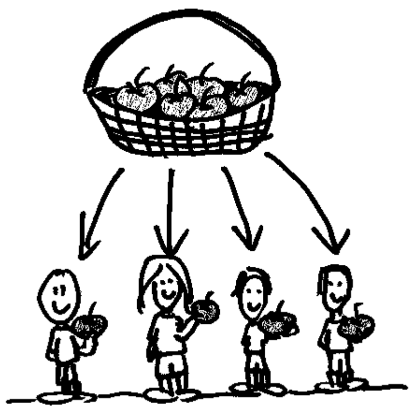


Figure A.1: *fan1* "share"

Figure A.2: *fan2* "powder"

Figure A.3: *fan3* "sleep"

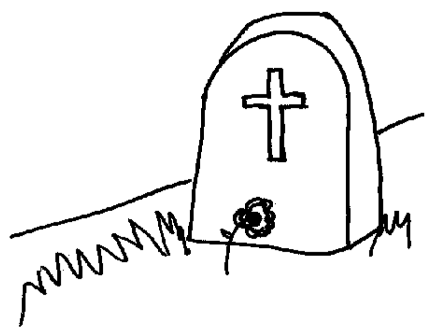


Figure A.4: *fan4* "tomb"

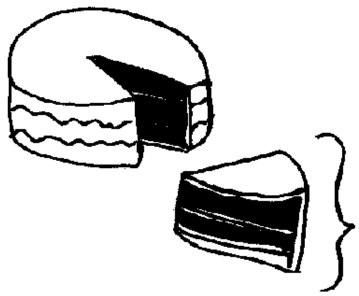


Figure A.5: *fan6* "portion"

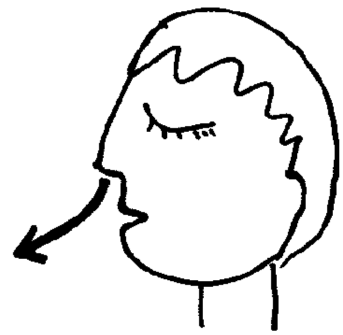


Figure A.6: *fu1* "exhale"

Figure A.7: *fu2* "tiger"

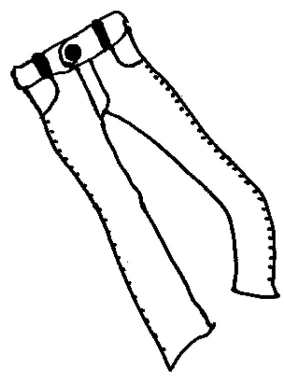


Figure A.8: *fu3* "pants"

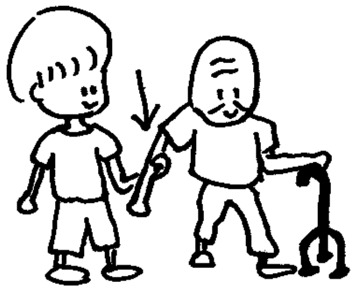


Figure A.9: *fu4* “help by holding another person’s arm”

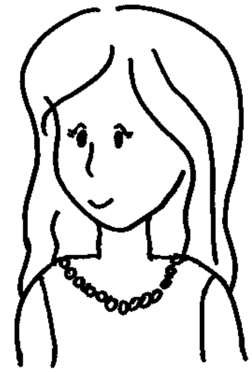


Figure A.10: *fu5* “woman”

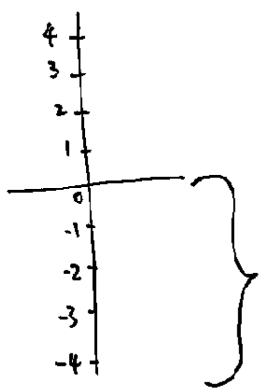


Figure A.11: *fu6* “negative”

Figure A.12: *ji1* “cure”

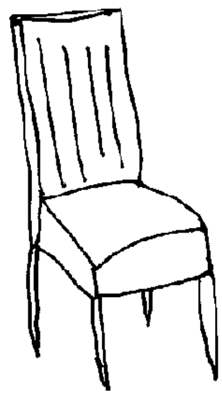


Figure A.13: *ji2* “chair”

Figure A.14: *ji4* “infant/child”

Figure A.15: *ji5* “ear”

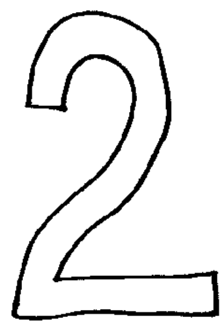


Figure A.16: *ji6* “two”

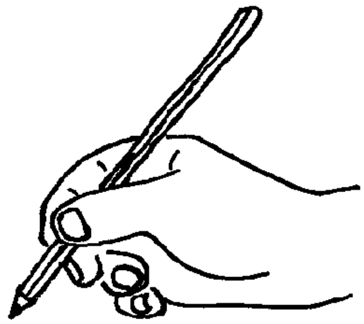


Figure A.17: *se2* “write”

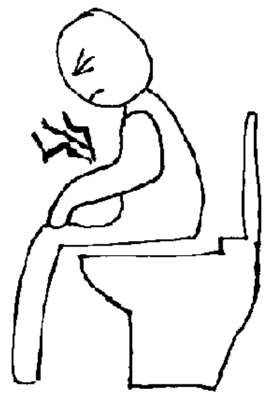


Figure A.18: *se3* “diarrhea”

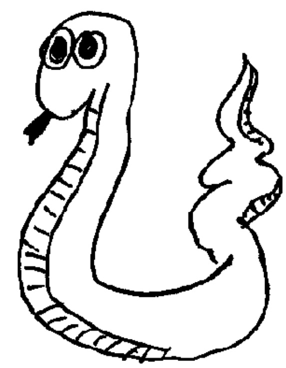


Figure A.19: *se4* “snake”

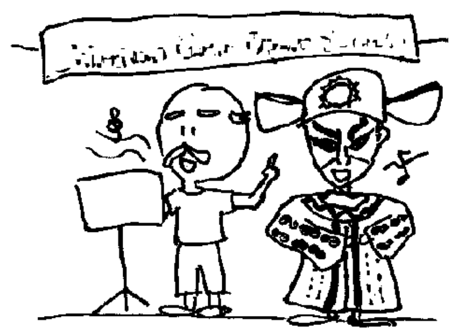


Figure A.20: *se5* “society”

Figure A.21: *se6* "shoot"

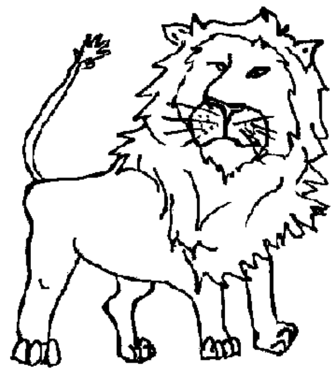


Figure A.22: *si1* "lion"

Figure A.23: *si2* "poop"

Figure A.24: *si3* "try"

Figure A.25: *si4* “key”

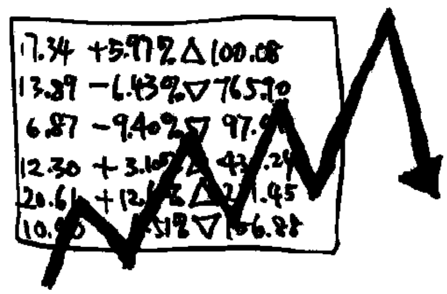


Figure A.26: *si5* “market”

Figure A.27: *si6* “nurse”

The following pictures were used for practice trials.

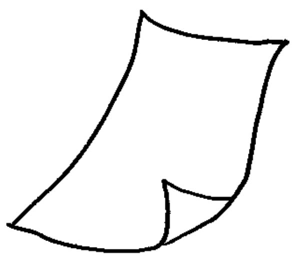


Figure A.28: *zoeng1* “piece (of paper)”

Figure A.29: *zoeng2* “prize”

Figure A.30: *zoeng3* “sauce”

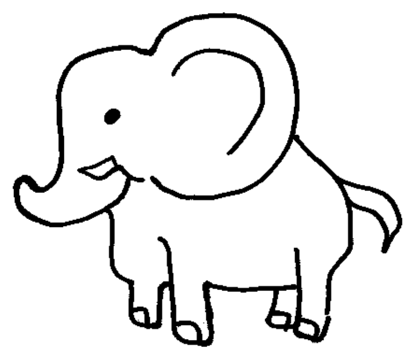


Figure A.31: *zoeng6* “elephant”

The following pictures were used for the story listening task.

Figure A.32: *jan4* “human”

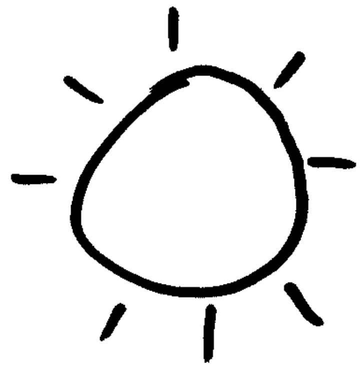


Figure A.33: *taai3 joeng4* “the sun”

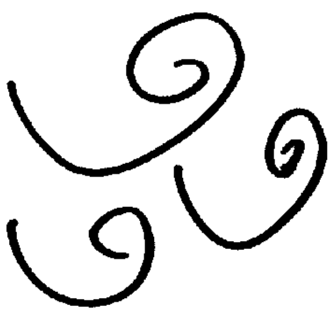


Figure A.34: *bak1 fung1* “the north wind”

A.4 Instructions

Note that all instructions were given in spoken Cantonese in the experiment. The text in written Cantonese, romanization, and English translation below are for readers' reference only.

A.4.1 Written Cantonese

Instructions for the story listening task

而家你會聽一個故仔。如果耳筒有問題，可以話俾研究員知。聽完故仔之後我會問你一個問題。

頭先個故仔裡面邊個贏咗呢？請撳 button box 上面嘅數字回答。

Instructions for the picture learning task

跟住落嚟我會介紹陣間實驗需要用嘅圖畫。請你留心聽，聽完之後撳 button box 上面任何一個掣繼續就可以喇。如果準備好，請你而家就撳 button box 上面任何一個掣。

Instructions for the first experimental block (Types 1, 2, 3)

跟住落嚟你會聽到啲字。邊幅圖畫代表你聽到嘅字呢？你只可以揀一幅圖畫，撳 button box 上面嘅數字回答。請留意，有時你可能會聽唔清楚某啲字，係特登嘅。只要盡你嘅能力回答就得㗎喇。準備好嘅話請你撳 button box 上面任何一個制開始。

Instructions for the second experimental block (Type 4)

跟住落嚟你會聽到啲句子。句子最後嗰個係乜嘢字呢？請留意，句子聽落去可能有少少唔自然，係特登嘅。你只要盡你嘅能力回答就得喇。如果唔明可以問研究員。明白嘅話可以撳 button box 上面任何一個掣開始。

Instructions for the third experimental block (Types 5A, 5B, 6A, 6B)

跟住落嚟你會聽到啲句子。句字最後嗰個係乜嘢字呢？同上個部份一樣，有啲字唔係咁清楚。你盡力回答就得㗎喇。請留意，有啲句字嘅意思係怪怪地，係特登嘅。你只要回答你聽到嘅最後嗰個字係乜就得喇。例如如果你聽到「咁多學生嘅名好難記」，咁你應該揀「記得」個「記」。 但係如果你聽到「咁多學生嘅名好難機」，咁你應該揀「飛機」個「機」，而唔係揀「記得」個「記」。如果唔明，可以問研究員。準備好嘅話請你撳 button box 上面任何一個制開始。

A.4.2 Romanization

Instructions for the story listening task

Ji4 gaa1 nei5 wui5 teng1 jat1 go3 gu2 zai2. Jyu4 gwo2 ji5 tung2 jau5 man6 tai4, ho2 ji5 waa6 bei2 jin4 gau3 jyun4 zi1. Teng1 jyun4 gu2 zai2 zi1 hau6 ngo5 wui5 man6 nei5 jat1 go3 man6 tai4. Tau4 sin1 go3 gu2 zai2 leoi5 min6 bin1 go3 jeng4 zo2 le1? Cing2 gam6 button box soeng6 min6 ge3 sou3 zi6 wui6 daap3.

Instructions for the picture learning task

Gan1 zyu6 lok6 lai4 ngo5 wui5 gaai3 siu6 zan6 gaan1 sat6 jim6 seoi1 jiu3 jung6 ge3 tou4 waa2. Cing2 nei5 lau4 sam1 teng1, teng1 jyun4 zi1 hau6 gam6 button box soeng6 min6 jam6 ho4 jat1 go3 zai3 gai3 zuk6 zau6 ho2 ji5 laa3. Jyu4 gwo2 zeon2 bei6 hou2, cing2 nei5 ji4 gaa1 zau6 gam6 button box soeng6 min6 jam6 ho4 jat1 go3 zai3.

Instructions for the first experimental block (Types 1, 2, 3)

Gan1 zyu6 lok6 lai4 nei5 wui5 teng1 dou2 di1 zi6. Bin1 fuk1 tou4 waa2 doi6 biu2 nei5 teng1 dou2 ge3 zi6 le1? Nei5 zi2 ho2 ji5 gaan2 jat1 fuk1 tou4 waa2, gam6 button box soeng6 min6 ge3 sou3 zi6 wui4 daap3. Cing2 lau4 ji6, jau5 si4 nei5 ho2 nang4 wui5 teng1 m4 cing1 co2 mau5 di1 zi6, hai6 dak6 dang1 ge2. Zi2 jiu3 zeon6 nei5 ge3 nang4 lik6 wui4 daap3 zau6 dak1 gaa3 laa3. Zeon2 bei6 hou2 ge3 waa2 cing2 nei5 gam6 button box soeng6 min6 jam6 ho4 jat1 go3 zai3 hoi1 ci2.

Instructions for the second experimental block (Type 4)

Gan1 zyu6 lok6 lai4 nei5 wui5 teng1 dou2 di1 geoi3 zi2. Geoi3 zi2 zeoi3 hau6 go2 go3 hai6 mat1 je5 zi6 le1? Cing2 lau4 ji6, geoi6 zi2 teng1 lok6 heoi3 ho2 nang4 jau5 siu2 siu2 m4 zi6 jin6, hai6 dak6 dang1 ge2. Nei5 zi2 jiu3 zeon6 nei5 ge3 nang4 lik6 wui4 daap3 zau6 dak1 laa3. Jyu4 gwo2 m4 ming4 ho2 ji5 man6 jin4 gau3 jyun. Ming4 baak3 ge3 waa2 ho2 ji5 gam6 buton box soeng6 min6 jam6 ho4 jat1 go3 zai3 hoi1 ci2.

Instructions for the third experimental block (Types 5A, 5B, 6A, 6B)

Gan1 zyu6 lok6 lai4 nei5 wui5 teng1 dou2 di1 geoi3 zi2. Geoi3 zi2 zeoi3 hau6 go2 go3 hai6 mat1 je5 zi6 le1? Tung4 soeng6 go3 bou6 fan6 jat1 joeng6, jau5 di1 zi6 m4 hai6 gam3 cing1 co2. Nei5 zeon6 lik6 wui4 daap3 zau6 dak1 gaa3 laa3. Cing1 lau4 ji3, jau5 di1 geoi3 zi2 ge3 ji3 si1 hai6 gwaai3 gwaai2 dei2, hai6 dak6 dang1 ge2. Nei5 zi2 jiu3 wui4 daap6 nei5 teng1 dou2 ge3 zeoi3 hau6 go2 go3 zi6 hai6 mat1 zau6 dak1 laa3. Lai6 jyu4 jyu4 gwo2 nei5 teng1 dou2 "gam6 do1 hok6 saang1 ge3 meng2 hou2 naan4 gei3", gam2 nei5 jing1 goi1 gaan2 "gei3 dak1" go3 "gei3". Daan6 hai6 jyu4 gwo2 nei5 teng1 dou2 "gam6 do1 hok6 saang1 ge3 meng2 hou2 naan4 gei1", gam2 nei5 jing1 goi1 gaan2 "fei1 gei1 " go "gei1", ji4 m4 hai6 "gei3 dak1" go "gei3". Jyu4 gwo2 m4 ming4, ho2 ji5 man6 jin4 gau3 jyun4. Zeon2 bei6 hou2 ge3 waa2 cing2 nei5 gam6 button box soeng6 min6 jam6 ho4 jat1 go3 zai3 hoi1 ci2.

A.4.3 English translation

Instructions for the story listening task

Now you are going to listen to a story. If your headphone is not working, please let the experimenter know. After the story I will ask you a question. Who won in the story? Please respond by pressing a number on the button box.

Instructions for the picture learning task

I am going to introduce the pictures to be used in the experiment. Please listen carefully, and press any button on the box to continue. When you are ready, press any button on the box to start.

Instructions for the first experimental block (Types 1, 2, 3)

You are going to listen to some words. Which picture represents the word that you heard? You can only choose one picture and respond by using the button box. Note that sometimes the word may be unclear, and it is intentional. You just need to try your best to answer. When you are ready, press any button on the box to start.

Instructions for the second experimental block (Type 4)

You are going to listen to some sentences. What is the last word of the sentence? Note that the sentences may sound unnatural, which is intentional. Just try your best to answer. If you have questions, please let the experimenter know. When you are ready, press any button on the box to start.

Instructions for the third experimental block (Types 5A, 5B, 6A, 6B)

You are going to listen to some sentences. What is the last word of the sentence? Just like the previous section, some words may be unclear, and you just need to try your best. Note that some sentences may not make sense, which is intentional. All you need to do is identify the last word that you heard. For example, if you hear “there are too many student names to remember”, then you should respond with “remember”. However, if you hear “there are too many student names to airplane”, then you respond with “airplane” but not “remember”. If you have questions, please let the experimenter know. When you are ready, press any button on the box to start.

A.5 Story used for the story listening task

Note that participants of the study only heard the audio file and were not provided with any written texts.

A.5.1 Written Cantonese

有一日，北風同太陽喺到嘈究竟邊個犀利啲。呢個時候，咁啱有個著住一件大褸嘅路人經過。於是，佢哋決定邊個能夠令到嗰個人剝咗佢件褸嘅話，邊個就贏。

北風盡力咁吹，但係越吹得大力，路人反而將件褸搵得越緊。北風唯有放棄。輪到太陽出馬嘅時候，太陽猛力咁曬，曬到路人流晒大汗，而且即刻將件褸除咗落嚟。最後，北風唯有承認太陽比佢犀利。

A.5.2 Romanization

Jau5 jat1 jat6, bak1 fung1 tung4 taai3 joeng4 hai2 dou6 cou4 gau3 ging2 bin1 go3 sai1 lei di1. Ni1 go3 si4 hau6, gam3 aam1 jau5 go3 zeok3 zyu6 jat1 gin6 daai6 lau1 ge3 lou6 jan4 ging1 gwo3. Jyu1 si6, keoi5 dei6 kyut3 ding6 bin1 go3 nang4 gau3 ling6 dou3 go2 go3 jan4 mok1 zo2 keoi5 gin6 lau1 ge3 waa2, bin1 go3 zau6 jeng4.

Bak1 fung1 zeon6 lik6 gam2 ceoi1, daan6 hai6 jyut6 ceoi1 dak1 daai6 lik6, lou6 jan4 faan2 ji4 zoeng1 gin6 lau1 meng1 dak1 jyut6 gan2. Bak1 fung1 wai4 jau5 fong3 hei3. Leon4 dou3 taai3 joeng4 ceot1 maa5 ge3 si4 hau6, taai3 joeng4 maang5 lik6 gam2 saai3, saai3 dou3 lou6 jan4 lau4 saai3 *daai6 hon6, ji4 ce2 zik1 hak1 zoeng1 gin6 lau1 ceoi4 zo2 lok6 lai4. Zeoi3 hau6, bak1 fung1 wai4 jau5 sing4 jing6 taai3 joeng4 bei2 keoi5 sai1 lei.*

A.5.3 English Translation

The Wind and the Sun were disputing which was the stronger. Suddenly they saw a traveller coming down the road. They decided that whichever of the two can cause that traveller to take off his cloak shall be regarded as the stronger.

The Wind began to blow as hard as it could upon the traveller. But the harder he blew the more closely did the traveller wrap his cloak round him, till at last the Wind had to give up in despair. Then the Sun came out and shone in all his glory upon the traveller, who soon found it too hot to walk with his cloak on. At last, the Wind had to admit that the Sun was the stronger.

Appendix B

Language background questionnaire

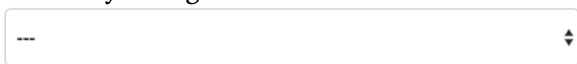
Participants were asked to fill out this questionnaire on a computer at the end of the experiment session on Day 1.

The following questions were from the Bilingual Language Profile (Birdsong et al., 2012): 1, 2, 5, 7, 9, 10, 11, 12, 24, 25, 26, 27, 28.

The following questions were added for the purpose of this study: 3, 4, 6, 8, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 29, 30.

For a detailed discussion on this questionnaire, see Section 4.4.3.2.

1. What is your age?[1]



2. What gender do you identify yourself with?

   ❍ Male

   ❍ Female

   ❍ Other: ________________

3. Do you have any hearing disorder?

   ❍ Yes (Please specify:________________)

   ❍ No

4. What is your dominant language?

   ❍ Cantonese

   ❍ English

   ❍ Mandarin

   ❍ A Chinese dialect (e.g. Hokkien, Hakka); Please specify: ________

   ❍ Other (e.g. French, Japanese); Please specify: ________________

[1]This is a drop-down menu with options from "17 or below", "18", "19" ... up to "60+".

5. How would you rate your proficiency of the following?
   0 = not at all; 6 = very well

	0	1	2	3	4	5	6
Cantonese - listening	❍	❍	❍	❍	❍	❍	❍
Cantonese - speaking	❍	❍	❍	❍	❍	❍	❍
Cantonese - reading	❍	❍	❍	❍	❍	❍	❍
Cantonese - writing	❍	❍	❍	❍	❍	❍	❍
English - listening	❍	❍	❍	❍	❍	❍	❍
English - speaking	❍	❍	❍	❍	❍	❍	❍
English - reading	❍	❍	❍	❍	❍	❍	❍
English - writing	❍	❍	❍	❍	❍	❍	❍

6. *"A heritage speaker of Cantonese is someone who in early childhood grew up with Cantonese (and possibly other languages) in the environment; however, by school age or shortly after, English was the primary language used in day-to-day situations."*

   Based on this definition, would you describe yourself as a heritage speaker of Cantonese?

   ❍ Yes, I consider myself a heritage speaker of Cantonese.

   ❍ No, I do not consider myself a heritage speaker of Cantonese.

7. What is your current place of residence?

   City ______________________________

   State (optional) ______________________________

   Country ______________________________

8. Please indicate the cities and countries that you have lived in along with how old you were when you lived there. List first the place where you were born, and list each town/city you have lived in.
   Example:
   Toronto, Canada — birth to 3
   Hong Kong – 4 to 10
   Vancouver, Canada — 11 to 19

9. At what age did you[2]...?

   Start learning Cantonese

   Start learning English

   ---

10. At what age did you[3]...?

    Start to feel comfortable using Cantonese

    Start to feel comfortable using English

    ---

[2]Options on this drop-down menu are “Never”, “Since birth”, “1”, “2” ... up to “20+”.

[3]Options on this drop-down menu are “Not yet”, “As early as I can remember”, “1”, “2” ... up to “20+”.

11. How many years have you[4]...?

Had classes (history, math, literature etc) in Cantonese (primary school through university, including Saturday School)

Had classes (history, math, literature etc) in English (primary school through university)

Spent in a country/region where Cantonese is spoken

Spent in a country/region where English is spoken

Spent in a family where Cantonese is spoken

Spent in a family where English is spoken

Spent in a work environment where Cantonese is spoken

Spent in a work environment where English is spoken

---

[4]Options on this drop-down menu are “0”, “1”, “2” ... up to “20+”.

12. What is your highest level of formal education?

    If you are a current undergraduate student, choose “Undergraduate degree”.

    ❍ Kindergarten
    ❍ Elementary/ primary school
    ❍ Junior high/ middle school
    ❍ High school diploma or equivalent
    ❍ Undergraduate degree
    ❍ Graduate degree

13. What was/is the main language of instruction in your...?[5]

    Kindergarten
    [--- ▾]

    Elementary/ primary school
    [--- ▾]

    Junior high/ middle school
    [--- ▾]

    High school diploma or equivalent
    [--- ▾]

    Undergraduate degree
    [--- ▾]

    Graduate degree
    [--- ▾]

[5]Options on this drop-down menu are: “Cantonese”, “English”, “Mandarin”, “Other”, and “Not applicable”.

14. Have you ever attended a Chinese school in North America?

   ❍ Yes[6]

   ❍ No[7]

15. Where was the Chinese school that you attended? (City, State, Country)

   ____

16. From what age to what age did you attend the Chinese school?

   Example: *6 to 14*

   ____

17. What was the language of instruction of the Chinese school that you attended?

   ❍ Cantonese

   ❍ Mandarin

[6] Participants who answered “yes” would be directed to a page that contained Question 15 to 17, and then continue with Question 18.

[7] Participants who answered “no” would be directed to Question 18.

18. In an average week, what percentage of time do you use the following languages with family? The percentages should add up to 100[8].

    Cantonese

    Another Chinese dialect (e.g. Hoisan/Toisan/Taishan, Zhongshan, Hakka, Hokkien, Teochew)

    Mandarin

    English

    Other languages

    ---

19. Is your father an immigrant to Canada?

    ❍ Yes. He is originally from:________________

    ❍ No, he was born in Canada.

    ❍ No, he does not live in Canada.

    ❍ Not applicable. I was adopted.

20. Is your mother an immigrant to Canada?

    ❍ Yes. He is originally from:________________

    ❍ No, he was born in Canada.

    ❍ No, he does not live in Canada.

    ❍ Not applicable. I was adopted.

[8]Options on this drop-down menu are "0%", "10%", "20%"... up to "100%"

21. What was/is the native language of your[9]...?

Father

Mother

Paternal grandfather

Paternal grandmother

Maternal grandfather

Maternal grandmother

---

22. If you have chosen "Other" or "Another Chinese dialect" in the previous question, please specify below which language your family member speaks. Leave this question blank if it is not applicable to you.

Example: *My father speaks Toisan. My mother speaks Tagalog.*

[9]Options on this drop-down menu are "Cantonese", "Another Chinese dialect (e.g. Hoisan/Toisan/Taishan, Zhongshan, Hakka, Hokkien, Teochew)", "English", "Mandarin", "Other", and "I don't know".

23. What language(s) is/are used in the following contexts? Please specify the percentage. If not applicable, leave that box blank. Example: *10% Cantonese, 90% English*

    You speaking to your father ________________________________
    Your father speaking to you ________________________________
    You speaking to your mother ________________________________
    Your mother speaking to you ________________________________
    Your parents speaking to each other _________________________
    You speaking to your sibling(s) _____________________________
    Your sibling(s) speaking to you _____________________________
    You speaking to your grandparent(s) _________________________
    Your grandparent(s) speaking to you _________________________

24. In an average week, what percentage of time do you use the following languages with **friends**? The percentages should add up to 100[10].

    Cantonese

    Another Chinese dialect (e.g. Hoisan/Toisan/Taishan, Zhongshan, Hakka, Hokkien, Teochew)

    Mandarin

    English

    Other languages
    ---

[10] Options on the drop-down menu of Question 24-27 are “0%”, “10%”, “20%”... up to “100%”

25. In an average week, what percentage of time do you use the following languages **at school or work**? The percentages should add up to 100.

Cantonese

Another Chinese dialect (e.g. Hoisan/Toisan/Taishan, Zhongshan, Hakka, Hokkien, Teochew)

Mandarin

English

Other languages

---

26. When you talk to yourself, how often do you talk to yourself in...? The percentages should add up to 100.

Cantonese

Another Chinese dialect (e.g. Hoisan/Toisan/Taishan, Zhongshan, Hakka, Hokkien, Teochew)

Mandarin

English

Other languages

---

27. When you count, how often do you count in...? The percentages should add up to 100.

Cantonese

Another Chinese dialect (e.g. Hoisan/Toisan/Taishan, Zhongshan, Hakka, Hokkien, Teochew)

Mandarin

English

Other languages

---

28. To what extent do you agree with the following statements?
0 = not agree at all; 6 = totally agree

	0	1	2	3	4	5	6
I feel like myself when I speak Cantonese.	❍	❍	❍	❍	❍	❍	❍
I feel like myself when I speak English.	❍	❍	❍	❍	❍	❍	❍
I identify with a Cantonese-speaking culture.	❍	❍	❍	❍	❍	❍	❍
I identify with an English-speaking culture.	❍	❍	❍	❍	❍	❍	❍
It is important to me to use (or eventually use) Cantonese like a native speaker.	❍	❍	❍	❍	❍	❍	❍
It is important to me to use (or eventually use) English like a native speaker.	❍	❍	❍	❍	❍	❍	❍
I want others to think I am a native speaker of Cantonese.	❍	❍	❍	❍	❍	❍	❍
I want others to think I am a native speaker of English.	❍	❍	❍	❍	❍	❍	❍

29. What do you think the experiment is about?

____

30. Is there anything else that you want to tell us?

❍ Yes, I think you might want to know that ____________________

❍ No