

Chapter 1

Introduction

Bilingual and monolingual linguistic experiences differ drastically. This sentiment is captured by the often-repeated observation that a “bilingual is NOT the sum of two complete or incomplete monolinguals; rather, [they have] a unique and specific linguistic configuration...a different but complete linguistic entity” (Grosjean, 1989, p. 6). One of the defining characteristics of a bilingual’s linguistic configuration is a shared phonetic space where both languages are produced and perceived (Flege & Bohn, 2021). Broadly, this dissertation is concerned with the implications of such a space and what aspects of sound are shared across languages in speech production. More specifically, this dissertation leverages the study of phonetic variation to ask whether and how individuals share a common structure across languages.

In pursuit of this goal, this dissertation first introduces a brand new speech corpus, developed as a key component of this dissertation and made freely available to the research community. The SpiCE corpus of Speech in Cantonese and English is described in detail in Chapter 2 (based on Johnson et al., 2020a), and can be accessed via Scholars Portal Dataverse (Johnson, 2021b). Broadly, SpiCE consists of conversational speech from 34 early Cantonese-English bilinguals residing in the Vancouver, BC, Canada area. The development of this corpus was a major undertaking, and its contribution is likely to be one of the most enduring and impactful parts of this dissertation. The studies described in this dissertation represent some

of the first examples of research with SpiCE but are unlikely to be the last.

The studies presented in this dissertation approach the larger question of structure in phonetic variation at different levels in the phonetic space—voice in Chapter 3 and sound categories in Chapter 4. While the focus of this dissertation remains squarely in production, both studies draw motivation from the perceptual domain, as crosslinguistic similarity for these different levels of phonetic variation have been proposed to account for how listeners can identify multilingual talkers in more than one language. While there is no perception component to this dissertation, the production studies provide some groundwork for developing better hypotheses about bilingual speech perception—some of these will be described in the general discussion in Chapter 5. This introduction briefly sets up that shared motivation, which ties two methodologically different studies together. Given the unique angle of each study, most of the literature is reviewed in the relevant chapters.

The introduction proceeds as follows. Section 1.1 gives context to the study of bilingualism in phonetics in broad strokes—that is, who is considered to be bilingual and what are some of the key characteristics that define them. The goal is to set up later chapters rather than provide a comprehensive discussion. Section 1.2 reviews some of the literature on how multilingual phonetic variation is perceived, emphasizing talker identification and how this dissertation provides a parallel for some of the claims made regarding perception. Section 1.3 motivates the need to attend to speaking style and argues that spontaneous speech corpora greatly facilitate the study of multilingual phonetic variation. Lastly, Section 1.4 provides the specific goal or research question for Chapters 2, 3, and 4.

1.1 Bilingualism

The population of interest in this dissertation comprises early bilinguals who are comfortable speaking and comprehending their two main languages—Cantonese and English. Chapter 2 provides a detailed description of the early bilinguals comprising the sample population.

In the most general sense, a bilingual is someone with knowledge of two or

more languages (Grosjean, 1989). This incredibly broad definition includes a diverse range of types of bilinguals. This range is perhaps best described on a continuum from first language (L1) to second language (L2) dominance, the bookends of which are monolinguals and replacive bilinguals, with learners, attriters, and balanced bilinguals in the middle (Gertken et al., 2014). Using a continuum in this way reflects the heterogeneous nature of bilingualism, even if it only captures a particular facet of bilingual competence. Dominance—and other aspects like patterns of use—are affected by factors such as age of acquisition, immersion environment, frequency, and social and communicative context (Gertken et al., 2014; Marian & Hayakawa, 2021).

While a spectrum may better reflect the reality of bilingualism, much of the literature focuses on more narrowly defined groups at discrete points of the spectrum, such as language learners or early (balanced) bilinguals. Typically, early bilinguals have learned both languages from their first years of life. A common cutoff is age five, or the age at which children begin regularly attending primary school (Amengual, 2017), as this marks a qualitative change in the kind of linguistic input bilinguals receive. Regardless of when bilinguals acquire a language, they do not necessarily use their languages to the same extent across different domains. To provide a relevant example, a Cantonese-English bilingual in Vancouver, BC, Canada, might use English at school and Cantonese at home. Bilingual language experience varies in still other ways, including code-switching (Fricke et al., 2016a), immersion environment (Sancier & Fowler, 1997), and formal instruction (Fricke et al., 2019). Each of these factors has a demonstrated effect on speech production. Such variety leads to markedly different linguistic experiences across groups of bilinguals, and as a result, markedly different patterns in speech production.

Across different kinds of phonetics research in bilingualism, there is a common trend of comparing bilinguals to “closely matched” monolingual populations. However, given the sheer heterogeneity within and across bilingual populations, an appropriate monolingual comparison group may not always exist. Hong Kong Cantonese offers a clear example of this. Given the incredibly multilingual nature of Hong Kong and the strong presence

of both English and Mandarin in daily life, few if any Cantonese speakers in Hong Kong would consider themselves monolingual. Further, Grosjean (1989) and many others have argued that such comparisons are often inappropriate, as using monolingual benchmarks implicitly and unjustly sets monolingual language varieties as the standard to achieve.¹ As a result, drawing comparisons between monolinguals and bilinguals may not always be fruitful or necessary, depending on the circumstances. There are cases where such comparisons are useful, but they should be justified and done with tact.

This recentring of bilingualism is reflected by a shift in the literature towards examining bilinguals on a within-population (e.g., Chan et al., 2020) or within-talker basis (e.g., Simonet & Amengual, 2019), or by comparing separate bilingual populations with different characteristics (e.g., Brown & Harper, 2009). This range of study designs will be apparent in the literature reviews of the following chapters. While there remains a broad range of comparisons in the literature—including with monolingual populations—in all cases, there is a strong push to consider bilinguals as the complete speakers they are. The analyses in this dissertation are based on within-talker comparisons in the domain of voice (Chapter 3) and sound categories (Chapter 4), and thus do not rely on a monolingual benchmark.

1.2 Processing bilingual talkers

Communicating in more than one language doesn't just involve the language produced by bilingual talkers; it also impacts how listeners perceive those talkers. As noted earlier in this chapter, one of the major consequences of bilingualism is a shared phonetic space (Flege & Bohn, 2021), in which bilinguals presumably (i.e., are hypothesized to) use similar voice quality to produce similar sound categories when speaking each of their languages.² While the idea of a shared phonetic space will be unpacked in greater de-

¹For extensive discussion on this topic, readers are referred to the literature on *translanguaging* (e.g., Wei, 2018).

²The speech production literature discussing similarity in voice quality and sound categories will be reviewed in greater detail within Chapters 3 and 4, respectively.

tail in Chapter 4, the important point here is that crosslinguistic similarity exists in the signal. Note, however, that the shape and extent of that similarity remain an active area of research, including, but not limited to this dissertation.

From the point of view of perception, crosslinguistic similarity of various kinds has been used to account for how bilingual listeners and talkers perceive and process one another’s speech (Chang, 2015). The clearest example of this comes from multilingual talker identification. Orena et al. (2019) report on a bilingual talker identification study within French-English bilingual and English monolingual listeners.³ The authors found that bilingual listeners were better at generalizing talker identification across languages than English monolinguals were, whether they were initially exposed to the talker in French and tested in English, or vice versa. However, both groups performed above chance, demonstrating that bilingual experience represents an advantage and not a prerequisite. A secondary analysis demonstrates that there is an additional advantage for bilinguals with more experience mixing languages.

The results of Orena et al. (2019) support a view where bilinguals use both talker-indexical and linguistic information to track a talker across languages, while monolinguals are restricted to using only talker-indexical information, as they lack the knowledge needed to extract linguistic information from French. Orena et al. offer several potential explanations for the advantage that bilingual listeners have: “that there are systematic changes in indexical information...[or] systematic consistencies in linguistic information across bilingual speech” (2019, p. EL308). As bilingual listeners are highly sensitive to subtle differences in acoustic input (Ju & Luce, 2004), the presence of systematicity in both talker-indexical and linguistic information—however subtle—would be accessible to bilingual listeners. Orena et al. also suggest that the results could be explained because the bilinguals “were familiar with both languages...while the monolinguals were only familiar with

³The paper does not specify where participants resided at the time of the study. Presumably, the bilinguals were recruited in Montreal, QC, Canada, though the monolinguals could have been recruited there or in Storrs, CT, USA, given the author affiliations.

one of the languages” (2019, p. EL309), though this account would be difficult to separate from the previous talker-indexical and linguistic accounts.

The argument Orena et al. (2019) make is that if something in the speech signal is talker-indexical, then any listener (regardless of language background) would be able to access it. If something is linguistic, then the more familiarity a listener has with that language, the better they will be able to leverage it in multilingual talker identification. This argument, however, hinges on crosslinguistic similarity being present in the bilingual speech production signal. The central goal of this dissertation could be construed as a parallel of this argument in production. While Orena et al. (2019) point to some prior work supporting indexical and linguistic accounts of bilingual talker identification (Theodore & Flanagan, 2019; Sundara et al., 2006, e.g.), convincing evidence remains scarce. This dissertation directly addresses these accounts from the perspective of crosslinguistic similarity and structured phonetic variation.

Using the corpus described in Chapter 2, Chapter 3 asks if there is shared structure in talker-indexical aspects of voice variation. A positive result here would help validate the argument that all listeners have access to talker-indexical information by virtue of demonstrating that such similarity, in fact, exists. Chapter 4 asks if there is shared structure at a linguistic level in the long-lag stop series of each language. If such structure exists, then it could be argued to facilitate a bilingual advantage in bilingual talker identification. While using distinct methods and addressing different aspects of phonetics, each of these chapters represents an aspect of the speech signal that may facilitate crosslinguistic talker identification. To preview the results of this dissertation, there is strong evidence for shared structure in voice variation. In contrast, the evidence for shared sound categories is promising but not as clear. Together, the studies offer support for the speculations made in Orena et al. (2019).

1.3 Variability in conversational speech

Though some of the motivation for this dissertation stems from perception, the primary goal is to document and investigate the structure of phonetic variation.⁴ While variability is inherent to the speech signal (cf. the lack of invariance problem; Liberman et al., 1967), spontaneous speech encompasses a greater degree of variability than other speaking styles (e.g., reduction phenomena; Johnson, 2004). Spontaneous speech—in the form of conversations—is thus the focus of study in this dissertation.

Additionally, as the motivation for the studies in Chapters 3 and 4 stems from all listeners’ ability—and bilinguals’ enhanced ability—to identify talkers in more than one language, using conversational speech also supports the external validity of this dissertation. Conversational speech better reflects the range of forms that people use and perceive in their daily lives. Additionally, given the potential range of variability, it is also necessary to study a large enough sample to cover the range of variation for the particular communicative situation. For similar reasons, Tanner et al. (2020) argue that large-scale corpus studies are uniquely valuable for understanding phonetic variation. In this vein, this dissertation introduces a new and sufficiently large corpus and subsequently leverages it in the study of conversational speech (SpiCE; Johnson, 2021b).

1.4 Thesis goals and research questions

While the main content chapters in this dissertation are united by common motivation, each has a unique focus or research question. These are as follows:

Chapter 2 expands on the motivation behind studying spontaneous speech and introduces the SpiCE corpus of spontaneous bilingual Speech in Cantonese and English (Johnson, 2021b). The development and dissemination of this corpus comprise a major portion of this dissertation.

⁴While the brevity of this section may not signal such a focus, the relevant phonetic variation literature is reviewed in detail in the following chapters.

Chapter 3 focuses on the structure of voice variation. In broad terms, it asks: do bilinguals have the same voice in each of their languages? More specifically, do bilinguals exhibit similar spectral properties and lower-dimensional structure in their voice across Cantonese and English? The answer is an emphatic *yes*. Chapter 3 also addresses a methodological question regarding the sample size necessary for the methods used and offers context for prior work in the literature.

Chapter 4 focuses on the structure of sound categories. In broad terms, it asks: Do bilinguals produce long-lag stops in the same way in each of their languages? More specifically, it describes the structure and sources of variation in how bilinguals produce voice-onset time in conversational speech. The results of Chapter 4 give some evidence for structure but require a nuanced interpretation.

Chapter 5 provides a brief recap of the preceding chapters, ties together the results, revisits some themes introduced in this chapter, and outlines limitations, implications, and future directions.

Together, the SpiCE corpus and these studies provide insight into the nature of crosslinguistic similarity and phonetic structure in conversational, bilingual speech, improving our empirical and theoretical understanding of these topics.