博士論文

The perception and production of English voicing contrast by Mandarin–Japanese bilinguals
（中国語・日本語バイリンガルによる英語の破裂音知覚と生成）

青木　理香

Table of Contents

List of Tables

List of Figures

List of Appendixes

List of Symbols and Abbreviations

Symbol	Definition
AM	Amplitude-modulated component
ANOVA	Analysis of variance
AOA	Age of arrival
BL	Bilingual
BLIP	Bi-Level Input Processing model
EB	Early bilingual
L1	First language
L2	Second language
L3	Third language
LB	Late bilingual
LOR	Length of residence
PAM	Perceptual Assimilation Model
SLM	Speech Learning Model
VOT	Voice Onset Time.

Acknowledgments

I would like to express my endless and deepest gratitude to Professor Izabelle Grenon, my supervisor and much more. I cannot describe how encouraging and patient she has been to me in all these years since she started to taking me on as a supervisee in 2013. Thank you for getting me back on the right track and for continuing to encourage me throughout. Izabelle's guidance is always fair, positive, clear and inspiring --- she always cheers me up when my research is stuck and I get depressed. She has shown me what sort of a teacher I want to be to my students. I feel very honored to be her first Ph.D. student!

My sincere thanks also go to Yuki Hirose and Shinichi Tanaka, my other thesis committee members. Hirose-sensei and Tanaka-sensei have been always supportive since I entered the Master's program at University of Tokyo. With Hirose-sensei, I also appreciated her help in recruiting participants of my experiments and her valuable advice on my statistical analysis. Tanaka-sensei always patiently listened to my research and gave me constructive and gentle comments.

I also would like to thank Chris Sheppard and Tom Gally, who joined to the defense committee. They kindly agreed to be committee members of my PhD. Dissertation and arranged a meeting at the last minute. Without Chris, I could not have conducted appropriate statistical analysis. Thank you very much for taking the time to answer my questions and demystify statistics. Comments from Gally-sensei especially reminded me how important it is to think and discuss with an objective eye.

For technical matters, I first thank Fumiaki Nishihara, Masaki Sone, and Prof.

Douglas Roland for the statistical help and advice on study design. Nishihara-san also helped me setting up my experiments and recruiting bilingual participants: without Nishihara-san, I could not have possibly started the data collection. Sone-san and Roland-sensei taught me exactly how to make a dataset, analyze, and interpret the data sparing their busy time.

My thank-you also goes to many people who may not be directly involved in the process of the thesis, but led me to where I am now. Professor Hiroko Saito, who sparked and cultivated my interest in linguistics and phonetics in my first year undergraduate, always cares me even after I graduated Tokyo University of Foreign Studies. Professor Paul Rossiter, who was a supervisor of mine until he retired in 2013, and introduced me to Izabelle. What you taught me have still served me well in writing an academic paper, and this dissertation.

In the year I entered the Master's program at University of Tokyo, I met people who have been important to me both as academic colleagues and friends. To Haruna Hirohata, Noriko Nishino, Yuko Nishio, and Yoko Suzuki, thank you for your friendship which made the tough times bearable and the good times even better. It is always pleasant and inspiring to talk with you. I am also thankful to Satoru Uchida, who has been my great "sempai" for such a long time.

Finally, I would like to say a big thank you to my family for their endless support and encouragement throughout my life. I'm deeply grateful to have them. Dad, Mom, and Yuriko are always there to provide unconditional love and support. Kenichi got me through the tough moments throughout the years. You are the best chef on this planet. Arigato!

I still cannot believe that I am submitting a Ph.D. dissertation and getting closer to finishing an eight-year journey. I am very excited that I have come such a long way!!

Abstract

The current research focuses on the investigation of whether mastering two languages facilitates or impedes the perception and production of a newly acquired language (L3) and whether the first language (L1) or second language (L2) affects the L3 performance. The present study conducted experiments to answer these questions and evaluate L1, L2 and L3 production, and bilingual's perception of L3. The participants were Mandarin-Japanese bilinguals who were L3 learners of English. They immigrated to Japan before or around puberty, and their varying age of arrival (AOA) in Japan was from 0 to 15 years old. They were divided into two groups based on their AOA to investigate AOA's effect in their L3 perception and production. The bilinguals who arrived in Japan between ages 0 and 6 years were called early bilinguals, while the ones whose AOA was between 10 and 15 years were late bilinguals. The results of the experiments were compared to the predictions made by the Bi-Level Input Processing (BLIP) model (Grenon, 2010).

In the production experiments of English, the voice-onset-time (VOT) values produced by bilinguals were measured and compared with those of English, Mandarin, and Japanese monolinguals. The production experiments were also conducted on L1 (Mandarin) and L2 (Japanese) to understand the difference between monolinguals and bilinguals. The results suggest that, regardless of their AOA in Japan, bilinguals produced L3 VOT similar to that of English monolinguals without being significantly influenced by L1 and L2. These results are compatible with the prediction inferred from the BLIP model, which I supposed would predict the following: provided that the bilinguals have forged a new neural map and phonemic category for the Japanese [+voice] feature, they should have all the neural maps or

categories to produce the plosive contrasts in all three languages. Thus, knowing two languages helped them to produce a wide range of VOT and therefore enabled them to produce L3 VOT accurately too. The patterns of the L1 and L2 production of bilinguals were also identical to those of Mandarin and Japanese monolinguals. There were some exceptions in the results, however, based on the place of articulation.

In the perception experiment of English, the bilinguals listened to a VOT continuum of the words *back* and *pack* ranging from lead VOT to long-lag VOT, and had to classify each word as corresponding to the word *back* or *pack*. The results show that there was an L1 influence on L3 perception pattern of bilinguals despite the fact that L2 perception pattern of Japanese monolinguals was much more similar to the L3 perception pattern than the L1 perception pattern of Mandarin Chinese monolinguals. Here, it is indicated that in L3 perception, bilinguals had difficulty in associating the underlying features with VOT neural maps at the phonological level, because of the difference in the number of underlying features among L1, L2, and L3. Thus, knowing two languages did not help bilinguals to perceive L3 stop contrast native-likely.

Hence, it can be concluded that bilinguals can produce L3 VOTs with a native-like norm without L1 or L2 transfer; on the other hand, their L3 perception pattern was distinct from a native-like norm, and was influenced by L1 or L2. Differences between perception and production results may, however, partly be due to differences in the features that the tasks measure. While the production tasks evaluate the mean VOT values, the perception task evaluates the categorical boundary along the VOT dimension. In addition, the present study also showed that the predictions on L3 production inferred from the BLIP model, which was originally a model of speech processing in perception with implications for the study of language acquisition, were compatible with the results of the production experiments. In other words, the present study suggests that the BLIP model has the potential to be extended to predict speech processing not only in perception but also in production. Implications of these results for a better understanding of L3 production and perception have been discussed.

1. Introduction

1.1 Background and context

There is a common belief according to which people who have learned more than one language may be able to learn new languages more easily than those who have learned only one language, presumably because the learning experience of additional language(s) may give them language learning strategies. Also, having acquired multiple different language systems and being able to switch from one language to another might make their newly learned sound system more flexible, and therefore they could adapt to the new system more easily. For example, bilingual students with knowledge of two orthographies showed significantly higher proficiency in their additional language orthography/phonological decoding compared to students who have knowledge of only one orthographic type, suggesting that bilinguals tend to have higher orthographical and phonological ability. (Abu-Rabia & Sanitsky, 2010).

However, there still remain important unsolved issues, possibly because research on second language (L2) acquisition and bilingualism has only emerged over the past few decades, and research on multilingualism is even scarcer. For example, does the knowledge of multiple languages facilitate both production and perception of the newly acquired language? Do the first language (L1) or L2 of bilingual/multilingual speakers affect each other? Are they different from those of monolingual speakers? In an attempt to investigate these issues, the current research conducted experiments evaluating the performance of bilinguals on a third language (L3) sound contrast, to see if having two languages facilitates the perception and production of an L3 contrast. More specifically, I evaluated the perception and production of stop voicing contrasts in terms of changes of Voice-Onset-Time (VOT) in

English (e.g. /b, d, g/ vs. /p, t, k/) by bilinguals of Mandarin Chinese (L1) and Japanese (L2) who have learned English as an L3.

In the English production experiment, VOT of the English (L3) word-initial plosives by bilinguals was compared with that of Mandarin monolinguals, Japanese monolinguals, and English monolinguals in order to see whether the L3 production pattern of bilinguals was similar to that of L1/L2/L3 monolinguals and to determine from which of the three languages bilinguals' production was affected. In addition, the production experiments in L1 (Mandarin Chinese) and L2 (Japanese) were conducted in the aim of observing whether bilinguals' L1 and L2 production was identical to or distinct from monolinguals'.

The L3 perception of bilinguals was also compared with that of L1/L2/L3 monolinguals so that I could determine whether their L1/L2/L3 had an impact on their L3 perception. In the perception experiment, bilinguals, Mandarin monolinguals, Japanese monolinguals and English monolinguals distinguished a VOT continuum into a voiced (unaspirated) plosive /b/ or a voiceless (aspirated) plosive /p/. I compared the categorizing patterns of each group.

The results were compared with predictions inferred from the Bi-Level Input Processing (BLIP) model proposed by Grenon (2010). The BLIP model accounts for "the categorical processing of speech sound contrasts in perception" (Grenon, 2010, p.74) of L2 learners. This study adopted the BLIP model because it separates between phonetic processing and phonological encoding. It is crucial to the comparison of VOT perception across languages as the VOT cue is used differently in the three languages I investigated. As for L1/L2/L3 perception, it can be inferred that it would be more difficult for bilinguals to categorize English VOT than Japanese monolinguals because they have both

Mandarin-attuned map and Japanese-attuned map: while the English contrast requiring two maps, namely [+voice] and [-voice], bilinguals have three maps, namely [+ spread glottis] map in Mandarin, [-voice] map in Mandarin and Japanese, and [+voice] map in Japanese. On the other hand, Japanese monolinguals would have fewer problems in categorizing English VOT, for they have the same number of neural maps associated with the same underlying features [+voice] and [-voice]. Even though the VOT values in Japanese are different than in English for the same underlying features, the BLIP model predicts that since there are only 2 maps in the VOT space in Japanese and 2 in English, which are also associated with the same underlying features, the maps should be sufficiently flexible to adapt to the perception of the L2 voicing contrast, at least with enough exposure.

The BLIP model does not make any prediction about production. However, in the current study, I extend the assumptions of the BLIP model to study production. As mentioned above, bilinguals would have to resolve the conflict between the three neural maps associated with different features in L1 Mandarin and L2 Japanese in order to produce the L3 VOT contrast. It is unclear if the presence of the three neural maps would facilitate or impede on L3 production. For a comparison and a better understanding of the results, Perceptual Assimilation Model (Best, 1995) and Speech Learning Model (Flege, 1995) are also introduced.

1.2 Dissertation outline

The current research on L3 production and perception by bilingual speakers is meant to provide a better understanding of L3 phonetic acquisition, in which L1, L2 and L3 are intricately intertwined with each other. For this purpose, I compared the bilinguals' L3 perception as well as their L1, L2, and L3 production of stops voicing contrasts with monolinguals of each of the three languages of interest: Mandarin Chinese, Japanese and English (North American) monolingual speakers.

Chapter 2 describes possible influential factors on L2/L3 phonetic acquisition and the L2 phonetic acquisition model that will be used in the current study. First, section 2.1 introduces previous studies on L2 phonetic acquisition regarding the possible effect of L1 transfer (2.1.1) and L2 transfer (2.1.2). In 2.1.3, previous studies on the possible effect of L2 experience are reviewed, followed by the effect of age of arrival in the country where the L2 is spoken (2.1.4) and length of residence in the L2-speaking country (2.1.5), and the influence of phonetic similarities between L1 and L2 (2.1.6). In section 2.2, previous studies on L3 phonetic acquisition are reviewed focusing on two possible types of transfer, language transfer (2.2.1) and language distance (2.2.2), followed by a discussion of the limitations in the previous studies (2.2.3). The following section 2.3 introduces a theoretical model that has been developed based on neural studies and models of sound processing: Speech Learning Model (SLM) in 2.3.1, Perceptual Assimilation Model (PAM) in 2.3.2, and the Bi-Level Input Processing (BLIP) model in 2.3.3. Section 2.4 introduces the acoustic features related to the stop contrasts in English, Mandarin Chinese and Japanese, focusing on the difference between the values of VOT in the three languages, which is the focus of this investigation (2.4.1). Then, the definitions of bilingual, monolingual, L1, L2 and L3 are given in 2.4.2. After that, I will present the hypotheses of the current study based on the BLIP model, for perception in 2.4.3, and for production in 2.4.4.

In Chapter 3, the methodology and the result of three production experiments are presented: the Mandarin production experiment (3.1), the Japanese production experiment

(3.2), and the English production experiment (3.3). In these production experiments, I observed and compared the VOT values by Mandarin-Japanese bilinguals and monolinguals of the target language to evaluate, first, if being proficient in two languages has an impact on their production of L1, L2 and L3 VOT contrasts. Section 3.4 summarizes and discusses the results of the three production experiments.

Chapter 4 reports the English perception experiment, where a VOT continuum was categorized into a voiced (unaspirated) plosive /b/ or a voiceless (aspirated) plosive /p/. Here, the participants were the Mandarin-Japanese bilinguals and the monolinguals speakers of English, Mandarin and Japanese (4.1). I first introduce the stimuli (4.2) and the procedure used in the experiment (4.3). Section 4.4 provides a summary and discussion of the results of the perception experiment.

In Chapter 5, the results of the production and perception experiments are linked together and a general discussion follows. Finally, Chapter 6 concludes the dissertation with implications and a suggestion for future research on L3 phonetic acquisition.

2. Literature review

In this chapter, previous studies on phonetic acquisition are briefly reviewed with a focus on VOT acquisition. Before discussing L3 phonetic acquisition, the studies on L2 phonetic acquisition, including both monolingual and bilingual speakers, are overviewed in 2.1, for the research on L2 phonetic acquisition has a longer and deeper history than L3 phonetic acquisition. In 2.2, the studies on L3 phonetic acquisition are reviewed, followed by the discussion on the limitation of these studies. Next, L2 phonetic acquisition models are introduced in 2.3. Finally, based on the predictions of the Bi-Level Input Processing (henceforth BLIP) model, hypotheses for the current study are discussed in 2.4.

2.1 Learning an L2 phonetic system

There are numerous studies on L2 phonetic acquisition providing evidence that L2 phonetic acquisition pattern is different from L1 acquisition pattern. Monolingual speakers are equipped with the sound system and a series of sound rules tuned specifically for their L1. That is, how speakers distinguish and categorize a sound input, or produce a sound output is processed by the L1 sound system. This makes it difficult for them to acquire another language that they encounter after they are fully attuned to the L1 system (that is, possibly past puberty). Researchers have reported various influential factors on the acquisition of L2 phonetic system, for example, transfer from L1 (L1 transfer), transfer from L2 (L2 transfer), the amount/length of L2 experience (L2 experience), and the phonetic similarities between L1 and L2. In the next following subsections, I will review the previous studies focusing on these factors.

Of course, there are more factors that affect the acquisition of L2 phonetic system.

The amount of L2 use is also an influential factor, that is, the frequency of learners' use of L2 in everyday life. For example, the frequency of English use after immigration from Italy to Canada affected the rate of goodness in the English vowel production by Italian-English bilinguals as rated by English monolingual speakers (Piske et al., 2002). As another factor, the effects of word familiarity were investigated in Flege et al. (1996), observing identification of English word-initial /r-l/ and /w-d/ minimal pairs by experienced Japanese learners of English who had lived in the U.S. for 21 years on average and inexperienced Japanese learners of English whose average length of residence in the U.S. was 2 years. The result was that experienced Japanese learners scored better for high-familiarity words compared to low-familiarity words, which reveals an effect of word familiarity on L2 performance at least for high proficiency L2 speakers. As for another factor, the effect of nonce word, Piske et al. (2011) observed the production of six English vowels by early/late Italian-English bilinguals with high/low L1 use in non-words and conversational speech. Generally, "early learners" indicate the learners who begin to learn L2 early in childhood, while "late learners" indicate those who begin to learn L2 in late adolescence or early adulthood. Although the criterion for classifying early/late learners vary among the studies, generally learners categorized as early learners start to learn the L2 around six years old. In the case of this study, the early learners began learning the L2 at the age of seven, while the late learners began at the age of twenty. The authors found that errors observed in non-words by early bilinguals with relatively high Italian use were not apparent in conversational speech. Therefore, they concluded that the production elicited in experiments using nonce words does not always reveal the accuracy in conversational speech. Hence, the production materials used in this study will be the real words in the three languages.

Among a great number of influential factors, L1 transfer, L2 transfer, L2 proficiency, age factors, and the phonetic similarities between L1 and L2 are reported to have a substantial impact on the acquisition of the L2 phonetic system. In the current study, the degree of L1 transfer on L2 production and L2 transfer on L1 production will be examined and the participants will be divided into two groups according to their age of arrival in L2 country. L2 proficiency and phonetic similarities, however, will not be evaluated in the present study. In the following sections, I will briefly present the results of previous studies on L1 transfer (2.1.1), L2 transfer (2.1.2), L2 proficiency (2.1.3), Age of acquisition (AOA) (2.1.4), Length of residence (LOR), phonetic similarities of L1 and L2 (2.1.6), and the reason why I include or exclude these factors from the research.

2.1.1 L1 transfer

L1 transfer is usually defined as "the incorporation of features of the L1 into the knowledge system of the L2 which the learner is trying to build" (Ellis, 1994, p. 28). For many decades, L2 learners' difficulties due to L1 transfer have been reported by numerous researchers, just to mention those on L2 VOT acquisition in production (Flege & Port, 1981; Flege & Hillenbrand, 1984; Flege & Hillenbrand, 1987b; Flege, 1987; Flege & Wang, 1989; González-Bueno, 1997), in perception (Flege & Schmidt, 1995; Curtin et al., 1998), and in both (Flege & Eefting, 1987). For example, Flege & Eefting (1987) conducted production and perception experiments with native Spanish adult learners of English and children who started to learn English at the age of 5-6, and age-matched English/Spanish monolinguals. Here, the mean VOT values in Spanish are generally shorter than those in English, regardless of the participant groups. In the production experiment initial plosives were produced, while in the perception experiment an identification judgment test was done using a VOT continuum ranging from /da/ to /ta/. These tests found that the English VOT values produced by Spanish speakers were shorter than English monolinguals' VOT. Also, the results of the adult Spanish speakers in the identification judgment test showed that the mean category boundary for English /da/ occurred at significantly shorter VOT values than that of English monolinguals. Therefore, the authors concluded that Spanish listeners who have learned English from around 5 or 6 years old seemed to develop a new L2 English aspirated category somewhat, but not fully native-like. That is, there was still an influence from L1.

In addition, L1 influence the way of using phonetic cues in perception. When French and English native speakers distinguished the English word-final stops /pɛk-pɛg/ and /pɪk-pɪg/, release burst cue had a significant impact on French listeners but not on English listeners (Flege & Hillenbrand, 1987a). This result can be attributed to the fact that word-final stops are released more consistently in French than in English. Thus, it suggests L1 phonetic rules influenced L2 perception.

L1 transfer is also observed in the speech of bilinguals learning two languages sequentially (i.e. a speaker has exposure to L1 at birth and then begins to have exposure to L2 later in childhood or adulthood). For example, in an imitation task of a /d-t/ VOT continuum, native monolingual English adults/children, native monolingual Spanish adults/children, and early/late Spanish-English bilinguals who had begun learning English before the age of six years, imitated short-lag stops (which appear both in Spanish and English), long-lag stops (which appear only in English), and lead stops (which appear only in Spanish) (Flege & Eefting, 1988). The results showed that the bilinguals could produce lead/short-lag/long-lag stops, while English monolinguals produced the continuum differentiating short-lag stops and long-lag stops, and Spanish monolinguals produced the stimuli differentiating lead VOT stops and short-lag stops. However, the VOT values of the long-lag stops by the bilinguals were shorter than native speakers of English. Thus, the authors concluded that the bilinguals established a new L2 /t^h/ category, but failed to reach the native range.

It does not mean, however, that L1 interference or transfer makes it impossible for L2 learners to establish a new L2 sound category. For example, English monolinguals and Spanish adult learners of English rated the goodness of the bilabial stops as realizations of the English /p/ listening to the /b-p/ continuum with VOT values ranged from 10 ms to 320 ms (Flege & Schmidt, 1995). The authors found that both English monolinguals and Spanish learners rated highest for VOT of 50 ms in spite of the fact that Spanish /p/ is short-lag in contrast to English long-lag /p/. Thus, these results suggest that L2 adult learners can establish a new L2 English category without being influenced by L1 Spanish category. Similarly, as for bilinguals, Spanish-English bilinguals identified Spanish /t, d/ and English /t, d/ in consonant-vowel syllables produced by nine monolinguals speakers of Spanish and English (Bohn & Flege, 1993). The results showed that in terms of the frequency of /t/ responses, bilinguals identified the stimuli of both Spanish and English in much the same way as monolingual speakers of Spanish and English. Furthermore, a longitudinal study of the acquisition of English word-initial VOT by a three-year-old Dutch child showed that over a period of seven months he successfully acquired English VOT contrast and even gradually changed his L1 prevoiced and short-lag stop contrast toward L2 short-lag and long-lag stop system (Simon, 2010). In this case, not only L2 learners overcame L1 transfer but also L2 had an inverse influence on L1. The next section introduces previous studies reporting L2 transfer to L1.

2.1.2 L2 transfer

L2 transfer can be defined as the influence of a speaker's L2 knowledge on L1 performance. In most studies on L2 transfer on L1, the participants resided in the L2 countries and belonged to the L2 speaking community, which means that the amount of L2 input was substantial (Flege, 1987; Major, 1990; Major, 1992; Thornburgh & Ryalls, 1998). For example, in Flege (1987), English native speakers living in Paris for 12 years with a French-speaking family as well as French native speakers living in Chicago for 12 years with an English-speaking family produced initial /t/ in both languages. Flege found that the mean English VOT of English-French bilinguals was shorter than that of English monolinguals (56 ms vs. 77 ms), and the mean French VOT of French-English bilinguals was longer than French monolinguals (51 ms vs. 33 ms). This result indicates that L1 VOT of the bilinguals became closer to L2 presumably due to an influence from L2. A comparable result was also found in the study on the English VOT of /p t k/ by English native speakers living in Brazil for 12-35 years (Major, 1992). His finding suggests that their L1 English VOT values decreased and drifted closer to the Portuguese values in casual speech but not in formal speech, suggesting that their L1 was significantly affected by their L2 in casual speech.

A small number of previous studies reported that L2 has an impact on L1 not only for speakers living in an L2 speaking community but also for those who remain in the L1 speaking community (Caramazza et al., 1973; Sancier and Fowler, 1997; Lord, 2008). When English native speakers with high proficiency in Spanish living in the States produced English and Spanish /p t k/, the VOT values of both English and Spanish were intermediate to those by monolinguals of the two languages, meaning that their L1 English was influenced by their L2 Spanish (Lord, 2008). These studies indicate that the L1 English phonetic system can be affected by the L2 system as a result of L2 acquisition even when a speaker remains in the L1 speaking communities. This is called regressive transfer or L2 transfer.

Since the Mandarin-Japanese bilinguals of the present study lived in the L2 country (Japan) and belonged to the L2-speaking community, it is possible for them to be more susceptible to L2 transfer than L1 transfer. In this case, Mandarin VOT production by bilinguals would be Japanese-accented, that is, shorter Mandarin VOT than Mandarin monolinguals. The degree of L1/L2 accentedness has been reported to be determined by the proficiency of one's L2. In the next section, previous studies on L2 proficiency are briefly reviewed.

2.1.3 L2 proficiency

In a number of previous studies, the degree of L2 proficiency (the overall degree of L2 knowledge) is found to be a predictor of L2 accuracy, where learners with lower L2 proficiency are more susceptible to L1 transfer. For example, the aforementioned study by Flege & Hillenbrand (1984) showed that L2 proficiency was influential for the production of the L2 sound which had a counterpart in the L1 (that is, /u/). In another study by Flege (1987), there were three groups of native English speakers with lower/intermediate/highest proficiency of French produced French and English stop /t/ and vowels /u, i, y/. The authors found that the more proficient a learner was, the more similarly they produced L2 sounds to the L2 monolingual speakers: in fact, the VOT in /t/ of the English learners of French with highest proficiency was shorter than that of the learners with lowest proficiency.

The problem of the L2 proficiency is that it is not always possible to assess the L2 proficiency of a speaker accurately and objectively. Therefore, the present study will not adopt L2 proficiency as an independent variable, but rather employ age of L2 acquisition, which is introduced in the next subsection.

2.1.4 Age of acquisition (AOA)

Age of acquisition is often taken as the Age of Arrival (AOA), which refers to the learners' age of arrival in the country where the L2 is the main spoken language. In many studies, especially those on bilingual learners, AOA has been reported to have an impact on L2 speech. Generally speaking, it is supposed that the younger a learner starts to live in the country where the L2 is a predominant language, the more accurate is their production (Munro et al., 1996; Piske et al., 2002) or both production and perception (Tsukuda et al., 2005).

Flege and his colleagues have examined the effect of AOA with a series of experiments of native Italian speakers who immigrated to Canada (Flege et al., 1995a, 1995b, 1999, 2002, 2003; Flege & MacKay, 2004, 2011; MacKay et al., 2001; Munro et al., 1996). Their general conclusion is that a learner's age of first extensive exposure to L2 is inversely related to their L2 accentedness, that is, the younger a learner is exposed to the L2-dominant environment, the less his/her L2 is accented.

Specifically regarding the relationship between AOA and VOT, English VOT production by Spanish-English early/late sequential bilinguals was found to be significantly affected by AOA (Flege, 1991). That is, the VOT values of /p, t, k/ were shorter for the Spanish monolinguals and longer for both the English monolinguals and the early bilinguals in any positions in an utterance, while VOT of the late bilinguals were intermediate between the short-lag values of Spanish monolinguals and the long-lag values of English monolinguals. Thus, the author concluded that early learners could establish L2 categories that were different from their L1, while late learners could not.

Thus, in the present study, it is predicted that Mandarin-Japanese bilinguals who immigrated to Japan at a young age will perform more accurately than those with older AOA, hence, we will separate the early bilinguals from the late bilinguals throughout the results sections. In the next subsection, length of residence in L2 country, which has been discussed in parallel with AOA in the previous studies, are briefly reviewed.

2.1.5 Length of residence (LOR)

Length of residence (LOR) "specifies the number of years spent in a community where the L2 is the predominant language" (Piske et al., 2001, p.197). As for the studies on LOR, it is mainly concluded that learners with longer LOR perform more similarly to native speakers of L2 compared to those with shorter LOR in production (Flege et al., 1995), in perception (Meador et al., 2000; Flege & Liu, 2001), and in both production and perception (Flege et al., 1997). For example, the four groups of L2 English learners having different L1, German, Spanish, Mandarin and Korean, with longer and shorter LOR took the production and perception tests in Flege et al. (1997). In the production test, they produced English test words containing /i, ɪ, ɛ, æ/, while in the perception test they identified the vowels in synthetic *beat*-*bit* /i-ɪ/ and *bat*-*bet* /æ-ɛ/ continua with 11 spectral steps and three duration steps. The results show that not only the effect of L1 transfer but also the effect of LOR were found to be significant in both production and perception. More specifically, in production the learners with longer LOR were assessed as more accurate than those with shorter LOR.

Similarly, in perception the learners with longer LOR used spectral information in /i-ɪ/ and /ɛ-æ/ contrasts in the same manner as the native speakers of English, while those with shorter LOR used the temporal cue more frequently.

Comparing AOA and LOR, researchers argue that LOR was a less important predictor of L2 performance accuracy than AOA (Oyama, 1976; Flege and Fletcher, 1992). In fact, when the English production by Italian-English bilinguals in Canada with various LOR was rated by the native speakers of English, additional years of experience in the L2-speaking environment were not likely to have a further ameliorative effect on the ratings of L2 pronunciation (Piske et al., 2001). What is more, there are some researchers claiming that the important factors on L2 accentedness are the amount of L2 input and the degree of L1 system strength, rather than age-related factors such as AOA and LOR (Flege and MacKay, 2011). They emphasized that AOA interacts with multiple factors bringing age-related effects, and therefore what has been considered as the influence from AOA may be affected by several other factors. For example, in a production study of English /b/, the late Italian-English bilinguals, whose mean AOA was 20 years old, showed a stronger L1 influence than the early Italian-English bilinguals, whose mean AOA was 8 years old (MacKay et al., 2001). Particularly, the late bilinguals produced a larger percentage of prevoiced (i.e. Italian-like) English /b/ than early bilinguals and English monolinguals, and production in the two languages was correlated in that they produced less prevoicing in their L1 Italian (i.e. in a less Italian-like manner) if they produced less prevoicing in their L2 English (i.e. in a more English-like manner). Also, in their perception test, the late bilinguals misidentified short-lag English /b, d, g/ as /p, t, k/ more often than the early bilinguals and English monolinguals. The authors attributed these AOA influences to the quantity and quality of the L2 input they had received: that is, late bilinguals were more likely to receive less amount of L2 input, and much of their L2 input was L1-accented. Many L2 acquisition and bilingual studies, however, still use AOA for categorizing subjects, or set AOA as an independent variable for L2 performance for the sake of its manageability and objectivity. In the current experiments, the bilingual participants are going to be divided into two groups according to their AOA, too, since I am interested in the possible effect of age of learning an L2 on bilinguals' L3 acquisition. LOR is nevertheless identified for each participant of the current study, in case it would have an effect on the performance of bilinguals.

2.1.6 Phonetic similarities of L1 and L2

A series of foreign accent studies in L2 phonetic acquisition found that L2 performance is influenced by perceived phonetic similarities between L1 and L2 sounds (Flege, 1987; Flege & Hillenbrand, 1987b; Flege et al., 2003; Tsukuda et al., 2005). Learners' difficulty in distinguishing L1 and L2 sounds is determined by whether perceived L2 category is close or distant from L1 category, or whether their L1 has a counterpart sound of L2 sound or not. For example, when English learners of French produced two L2 sounds, /u/ (which has a counterpart in L1) and /y/ (which was supposed to be new because their L1 does not have /y/ in its inventory), L2 learners produced the new L2 sound more accurately than the L2 sound which has a counterpart (Flege, 1987; Flege & Hillenbrand, 1987). This result suggests that the equivalence classification hinders learners from acquiring target-like L2 sounds which are similar to L1 sounds, while it does not influence the acquisition of L2 sounds which are distant from any L1 sounds.

In contrasting L2 sounds, L2 learners tend to behave in a characteristic way:

experienced learners produce significantly larger L2 sound contrast or show more exaggerated movement than inexperienced learners (Flege et al., 2003; Tsukuda et al., 2005). In Flege et al. (2003), the effect of L1 on English /e^{ɪ}/ production by Italian-English bilinguals with early/late AOA and high/low L1 use was investigated. The authors found that the tongue movement of early bilinguals was greater than that of late bilinguals and English monolinguals, which was attributed to the dissimilation of a phonetic category they formed for L2 English /e^{ɪ}/ from L1 Italian /e/.

Thus, in the previous studies, the two languages are compared based on phonetic (perceptual) similarity. However, the three languages investigated in the current studies are compared in terms of phonological features, because the VOT cue in these languages is used differently for equivalent underlying phonological representations. Therefore, the present study will use the BLIP model proposed by Grenon (2010) for the prediction of the experiments, which will be introduced in 2.3. Before that, I will introduce the previous studies on L3 phonetic acquisition in the following section 2.2,

2.2 Learning an L3 phonetic system

The studies mentioned in 2.1 focus on the relationship between two languages, L1 and L2. When there are more than two language systems in one's brain, however, the relationship between each pair of languages becomes more complicated. In fact, it has been considered that "L3/L*n* language learners are distinct from typical adult L2 acquirers since the former possess a larger repertoire of linguistic and metalinguistic knowledge" (Cabrelli Amaro, 2012, p.33-34). Thus, it is difficult to observe whether the knowledge of more than two languages facilitates or interferes with the acquisition of an additional language (i.e. L3), and the number of previous studies on L3 phonetic acquisition is remarkably small compared to those on L2 phonetic acquisition. According to the previous studies on L3 phonetic acquisition, the two main factors that have been reported to be an important predictor of L3 accuracy are language transfer from L1 or L2 and language distance among the three languages in a trilingual speaker. In the following subsections, I will review the previous studies of language transfer in 2.2.1 and language distance in 2.2.2. Finally, 2.2.3 discusses the limitation in the previous studies on L3 phonetic acquisition.

2.2.1 Language transfer

The effect of language transfer can be categorized into four patterns: L1 transfer, L2 transfer, combined transfer, and regressive transfer. These four patterns are briefly discussed in the subsequent subsections.

2.2.1.1 L1 transfer

Previous studies on L3 phonetic acquisition, especially those in 1980s, reported significant L1 influence on L3 production (Llisteri & Poch-Olivé, 1987; Ringbom, 1987; Gut, 2010; Wrembel, 2012a; Wrembel, 2012b). For example, Finnish-Swedish bilinguals learning English as their L3 produced English intonation with Finnish accent regardless of their proficiency of English (Ringbom, 1987). However, when four trilingual speakers with different L1 (Polish, Russian, Hungarian, Spanish), and with high L2 (German or English) proficiency produced three speech styles in L3 (English or German), the L3 speech rhythm and vowel reduction of an L3 speaker whose L1 values were phonetically similar to L3 did not show a significant difference from those of L3 speakers whose L1 values were phonetically distant from L3 (Gut, 2010). The author concluded that L1 has no influence on L3 performance, suggesting that the transfer from L1 to L3 would be limited when a speaker's L2 proficiency is high, because high L2 proficiency may make an L3 learner rely more on L2 than L1, which would be called L2 transfer, as discussed in the next section.

2.2.1.2 L2 transfer

The L2 sound system of a speaker may have an influence on L3 performance (Hammarberg & Hammarberg, 2005; Gut, 2010; Llama et al., 2010; Wrembel, 2010). For example, VOT values of Spanish plosive production by English-French bilinguals and French-English bilinguals were significantly influenced by the L2 for both groups of bilinguals (Llama et al., 2010). This result suggests that L2 transfer is the stronger predictor of L3 performance rather than L1 transfer, or even language distance, which can be defined as the classification of languages according to their generic relatedness or structural characteristics, such as phonological systems, writing systems, and word order (Proctor, August, Snow, & Barr, 2010).

Hammarberg & Hammarberg (2005) argues that the influential language on L3 performance is switched from L2 to L1 along the stages of L3 acquisition. In their longitudinal study of L3 phonetic acquisition, the Swedish speech of an English-German bilingual at two points in time was observed in order to determine whether the L3 performance was more influenced by L1 or L2. They concluded that in early learning, the first eight months of development, the speaker relied on L2 more, while in later learning, as L3 proficiency increased, the L2 effects attenuated allowing L1 to influence L3 (see Wrembel, 2010 for similar results). This result suggests that L2 effect is stronger than L1 effect at the onset of L3 acquisition, while L1 effect becomes stronger than L2 effect as the L3 proficiency increases. In the current study, we may therefore expect that the effect of L2 (Japanese) is stronger than that of L1 (Mandarin) on L3 (English) performance, because all bilinguals in the current study had an English learning experience of more than a few years.

2.2.1.3 Combined transfer

Several previous studies found a combined effect of L1 and L2 in the acquisition of L3 pronunciation caused by a combined transfer, in other words, a transfer with multiple sources, mainly L1 and L2 (Wunder, 2010; Sypiańska, 2013; Wrembel, 2014). For example, when L3 English speakers with L1 Polish and L2 Danish produced word-initial /p, t, k/ in the three languages, their L3 VOT patterns were influenced both from L1 and L2 (Sypiańska, 2013). Here, their L3 VOT values were intermediate between L1 and L2, demonstrating L1 and L2 combined effect. In addition, the previous study of VOT production patterns in stressed onset positions by trilinguals with L1 Polish, L2 English and L3 French/German, their L3 VOT values were intermediate between the L1 and L2 mean VOT values, suggesting the combined effect of L1 and L2 on L3 phonetic acquisition (Wrembel, 2014).

However, the languages used in most of the previous studies are typologically similar, belonging to Indo-European language family, or they share the similar linguistic patterns. As I discuss later in 2.2.2, language distance has been reported to affect the performance of L3 learners: when one language is genetically or formally closer to the target L3 language than the other language, the former has a greater impact on L3 performance. It has not been fully studied whether and how much a language distance has an impact on L3 performance. In order to investigate the effects of the languages which are distant from each other in perceiving or producing L3, the present study deals with the three languages which are not genetically close to each other.

2.2.1.4 Regressive transfer

Language transfer reported by the majority of previous studies on L3 phonetic acquisition is progressive transfer from L1 and/or L2 to L3 performance. However, as discussed in the literature review of L2 phonetic acquisition in 2.1, there can be also regressive transfer from L3 to L1 and/or L2 (Wrembel, 2011). Regressive transfer is a cross-linguistic influence where a more recently acquired language influences an existing linguistic system. In fact, regressive transfer was observed in the production experiment by native speakers of Polish whose L2 was English and L3 was French (Cabrelli Amaro & Rothman, 2010; Wrembel, 2011). Here, the target phonemes were voiceless plosives /p, t, k/ in onset positions, which are produced with short-lag VOT in Polish and French and with long-lag VOT in English. The author found not only the combined effect on L3 VOT from L1 and L2, but also regressive effect from L2 to L1, making their L1 VOT longer than that of Polish monolinguals. Similarly, the study on VOT of word-initial /p, t, k/ by Polish native speakers with L2 Danish and L3 English also found regressive transfer in addition to the combined effect of L1 and L2: their VOT values of Polish and Danish were longer than those of Polish-Danish bilinguals due to the influence from their English long-lag VOT. Regressive transfer was also reported in the longitudinal study of sequential/simultaneous English-Spanish bilinguals with L3 Brazilian Portuguese (Cabrelli Amaro & Rothman, 2010). In this study, the authors concluded that the successive bilinguals showed greater regressive transfer from L3 to L2 as their L3 proficiency developed.

2.2.2 Language distance

Language distance is a factor which is investigated in the studies of various kinds of area, and is called with a different term in a different previous study: while "language distance" is used in Ringbom (1987), for example, "psychotypology" or "typological proximity" is used in Kellerman (1977). According to a previous study, language distance has three different definitions: (a) genetic relatedness (i.e. language family), (b) formal similarity, and (c) perceived similarity of languages by learners (Falk & Bardel, 2010). Wrembel (2015) also claims that language distance can be seen either as "an objective formal measure of a genetic relationship between language families or as learners' subjective perception of that language distance, i.e. psychotypology" (Wrembel, 2015, p. 42). In his investigation of the a polysemous word, Kellerman (1987) claimed that transferability, that is, how one language is likely to transfer to the other, is conditioned by perceived similarity and prototypicality: that is, the more a feature in a language is similarly and prototypically perceived, the more likely the effect of language transfer would be. If this can be applied to phonetic acquisition, too, it follows that perceived similarity and prototypicality of a sound or a feature have an impact on L3 phonetic acquisition.

A comparative number of previous studies on phonetic acquisition of L3 speakers have demonstrated the effect of language distance: that is, the language which is genetically closer or similarly perceived to the L3 exercises more influence on L3 performance. Wrembel (2015) compared the results of her previous study on L3 VOT of L1 Polish/L2 German/L3 English speakers (Wrembel, 2010) and that of L1 Polish/L2 English/L3 French speakers (Wrembel, 2012a), where trilinguals' L3 performance was assessed in terms of the degree of the perceived foreign accent by expert raters of English native speakers. Here, she categorized the four languages into the two groups: Polish and French as “voicing” type (i.e. the language in which stop contrast is made by the presence/absence of voicing), while English and German as “aspirating” type (i.e. the language in which stop contrast is made by the presence/absence of aspiration). What she found was that L1 transfer was stronger for the trilinguals whose L1 is Polish, L2 English and L3 French, while L2 transfer was more salient for the trilinguals whose L1 is Polish, L2 English and L3 German. The results show that language distance (here, formal similarity of the features) had as a stronger impact on L3 production regardless of the sequential order of language acquisition. The author concluded that what determined the strength of cross-linguistic influence was language distance rather than language transfer.

The three languages in the current study, namely Mandarin, Japanese, and English, are genetic typologically unrelated for they all belong to the different language families. As for formal similarity, on the other hand, the three languages cannot be categorized straightforwardly as Wrembel (2015) did: while Mandarin can be allocated to “aspirating” type and Japanese would belong to “voicing” type, English can belong to neither because English stop contrast is made by voicing at an underlying level but by aspiration at a surface level. Here, in terms of formal similarity, the three languages are not related to each other. Thus, in the present study, no effect of language distance can be expected. Instead, language transfer effects including L1, L2 and combined transfers may be stronger, depending on the bilinguals’ proficiency of each language. In the next subsection, the limitations and the gaps of the previous studies reviewed in 2.2.1 and 2.2.2 are discussed.

2.2.3 Limitation of the previous studies

In the previous subsections, I briefly introduced studies on L3 phonetic acquisition in terms of two influential factors: language transfer and language distance. Although these studies show that L3 learners are distinct from L2 learners due to the different number of languages in process, further studies are still required in terms of two points.

First, most previous studies on L3 phonetic acquisition focused on speech production only, "ignoring a fundamental component of non-native phonological acquisition: speech perception" (Cabrelli Amaro, 2012). The influential L2 phonetic acquisition models such as SLM and PAM which will be reviewed in the next section are based on L2 perceptual process positing that the difficulty of L2 phonetic acquisition is determined by the perceived similarity of acoustic cues. Therefore, not only production studies but also perception studies are required to test the hypotheses and the predictions of existing L2 models and to understand L3 phonetic acquisition as a whole.

Second, several previous studies report only the L3 performance of L3 speakers without testing their L1 and L2. In these studies, the comparison of their L3 with L1 and L2 are done utilizing the data from monolinguals speakers of L1 and L2. In fact, among the recent previous studies on L3 phonetic acquisition, Tremblay (2007) is one of the very few studies that reported the data of the three languages (English, French, and Japanese) by trilingual participants. However, L3 speakers are definitely distinct from monolingual speakers and therefore their L1/L2 speech cannot be represented by the speech of L1/L2 monolinguals. Moreover, L3 speakers tend to have wider inter-individual variability in their speech due to their variety of linguistic background. Thus, for more precise investigation, it is necessary for L3 researchers to observe L1, L2 and L3 of L3 speakers.

In an attempt to fill the gaps in past research, the present study measured the production of VOT stops in all three languages (Mandarin, Japanese and English) by Mandarin-Japanese bilingual learners of English, Mandarin monolinguals, Japanese monolinguals and English monolinguals. The current study also evaluated the perception of an English stop contrast along a VOT continuum by speakers of Mandarin, Japanese, and English in addition to the bilingual speakers. The production and perception experiments were designed to investigate whether knowing two languages facilitates or impedes the acquisition of an additional language, when the three languages are not typologically related. The following research questions were specifically investigated in the current study:

1) Which previously learnt language, L1 or L2, has a stronger impact on L3 phonological acquisition? Or will both L1 and L2 have no impact on L3 phonological acquisition? In particular, which of Mandarin or Japanese affects the results of English production experiment by Mandarin-Japanese bilinguals? Will the results of English perception experiment by bilinguals be closer to the results by Mandarin and Japanese monolinguals?
2) Will the bilinguals' L1 and L2 production be distinct from that of monolinguals? Will the results of production experiments in Mandarin and Japanese by Mandarin-Japanese bilinguals be similar to those of Mandarin monolinguals and Japanese monolinguals?
3) Will the bilinguals' age of arrival in an L2 country affect their L3 phonological acquisition? In particular, does the bilinguals' age of arrival in an L2 country have an influence on the results of production experiment and perception experiment in L3?

Here, in order to exclude the effect from language distance reviewed in 2.2.2, I focused on the language which are typologically unrelated, Mandarin, Japanese, and English. The experimental group in the current study consists of Mandarin-Japanese bilinguals who began to learn Japanese as an L2 between the age of 0 and 15, and English as an L3 between the age of 2 and 13.

In the next section, before the hypotheses for the current research questions are made, three phonetic acquisition models are briefly reviewed. These three models are compared and the model for the prediction of the present study is chosen.

2.3 L2 phonetic acquisition models

In this section, two major phonetic acquisition models, that is, Speech Learning Model (SLM) proposed by Flege (1995) and Perceptual Assimilation Model (PAM) advocated by Best (1995) are introduced in 2.3.1 and 2.3.2 respectively. Then, in the 2.3.3, I will briefly review the BLIP model proposed by Grenon (2010), which is used for the better understanding of the results of the present study.

2.3.1 Speech Learning Model

The Speech Learning Model (SLM) is one of the most popular and widely-used model of L2 phonetic acquisition. It posits that "an L2 phonetic category may dissimilate from a neighboring L1 vowel category in order to preserve phonetic contrast among the elements of the L1 and L2 subsystem, which are said to exist in a common phonological space" (Flege et al, 2003, p.2003). According to the prediction by SLM, the development of a new L2 sound category would more likely to proceed when L2 sound is perceptually distant from the closest L1 categories, rather than when it is perceptually close.

In SLM, the perception of L2 sounds is not based on counterparts in the L1 phoneme inventories, but on the phonetic realizations of speech sounds in particular contexts, in other words, the acoustic property of the L2 sound. The importance of phonetic properties, not phonological ones, has been proved in the previous studies comparing the effects of transfer from L1 phonological and phonetic properties. For example, the perception of English approximant pairs /r-l/ and /w-j/ by Danish and German learners of English was affected by phonetic properties more significantly than those of phonology (Bohn & Best, 2012). Here, although Danish lacks /w/ in the inventory, Danish learners of English even outperformed English monolinguals because Danish has a rich vowel inventory using lip rounding contrastively for vowels: Danish learners were able to distinguish English /w-j/ using this lip rounding, which is the defining feature of English /w-j/ contrast.

Thus, according to SLM, an L2 learner is able to establish a new phonetic category for an L2 sound when the learner is able to recognize phonetic differences. If the learner fails to recognize them, on the other hand, category formation for the L2 sound may be blocked even after many years of L2 experience, and he/she would classify the L2 sound and its L1 counterpart into a single phonetic category.

The scope of SLM is not limited to the acquisition of L2 perception. It predicts that when a new L2 category is not established, production of this sound will be inaccurate. In fact, previous studies show that the L2 sound and its L1 counterpart merged (Flege, 1987; Flege & Hillenbrand, 1984).

As possible results of the present study, SLM predicts that when a bilingual successfully recognizes the phonetic difference between L1, L2 and L3, and establishes a separate category for sounds in each language, he/she is able to perceive and produce L3 sounds accurately. On the other hand, in the case the bilingual fails to create a new category for L3 and merges it to either L1 or L2 category, his or her perception and production would be L1- or L2-accented, that is, not native-like.

2.3.2 Perceptual Assimilation Model

Another influential perception model which has often been referred to parallel with SLM is the Perceptual Assimilation Model. PAM was designed to explain the possible patterns in which a listener perceives a new sound, and make predictions of difficulties encountered by him/her in each pattern. PAM, like SLM, predicts that L2 sounds are perceived according to their phonetic similarities to L1 sounds. The assimilation patterns are divided into three types: Single-Category Assimilation (SC), Two-Category Assimilation (TC), and Category-Goodness Assimilation (CG). SC occurs when two L2 sounds are mapped onto a single L1 sound, leading to poor discrimination. In the case of TC, on the other hand, two L2 sounds are mapped onto two different L1 sound categories, which allows a good discrimination. As for the case of CG, two L2 sounds are mapped onto a single L1 sound, but with different perceived degree of similarities to it. What should be noted here is that perceived L2 properties are based on its articulatory gestures, not a phonemic status or phonetic similarities. Thus, the L2 discrimination by a listener should be most accurate in the cases of TC, less accurate in CG cases, and the least accurate in SC cases.

These predictions by PAM are supported by many previous studies. For example, when French monolinguals discriminated English /d-ð/, some identified English /d/ as French

/d/, and /ð/ as a non-native sound (TC), while others assimilated both /d/ and /ð/ to French /d/ (SC), or with a different perceived similarities (CG) (Polka et al., 2001). In this case, as predicted, the French with TC pattern showed good discrimination, whereas those with SC pattern displayed poor discrimination. Those with CG pattern also exhibited good discrimination, intermediate between the SC and TC cases.

According to the predictions by PAM, the results of the present study may be predicted as follow: bilinguals are supposed to have four labial stop categories of their L1 and L2, /p^h/ and /p/ in L1 Mandarin, and /p/ and /b/ in L2 Japanese. When a bilingual perceives an L3 stop contrast /b-p/ in terms of VOT, bilinguals have to assimilate the two sounds into one of the four categories. Here, they should perform very well because English /p/ would be identified as Mandarin /p^h/ and English /b/ as Japanese or Mandarin /p/ (a TC pattern) because English /p/ is perceptually similar to Mandarin /p^h/ in terms of VOT, and English /b/ is similar to Japanese or Mandarin /p/.

There are three major differences between SLM and PAM. First, SLM primarily explain L2 production, whereas the PAM focuses on the cross-language perception of naive listeners. Second, SLM focuses on individual phonetic categories, whereas PAM focuses on pairwise phonological contrasts. Due to these differences, it may not be appropriate to compare SLM and PAM.

In fact, the predictions of SLM and PAM are supported by various previous studies and therefore the two models have been the most influential in the studies of L2 phonetic acquisition for over the past two decades and more. The two models, however, have one common precondition hindering the prediction of the present study. As noted above, the predictions of both SLM and PAM are based on “perceptual similarity”, which is difficult and complicated to assess. For example, PAM describes phonetic similarity in terms of gestural features such as the degree and the place of the tongue, lip postures, and velar gestures, but it is technically difficult to determine the exact position of these articulators for the similarity judgment. Similarly, several previous studies working within SLM and PAM have assessed perceptual similarity using either informal L1 transcriptions by non-native listeners or a cross-language identification, such as rating the category goodness of an L2 sound as an exemplar of the L1 category with the scale of 1 being very foreign sounding to 7 being very native sounding. It is inevitable, however, for the results of these techniques to be subjective and intuitive allowing high inter-subject variability in one's judgment. Furthermore, since bilinguals have both an L1 and L2, it is questionable to which language they will compare the L3 categories when asked to rate the "nativeness" of the L3 sounds. In order to avoid these problems, the present study will use the BLIP model proposed by Grenon (2010) for the predictions of the experiments, although the predictions made by SLM and PAM are also mentioned in the Discussion, in Chapter 5.

2.3.3 Bi-Level Input Processing (BLIP) model

The BLIP model was designed "to identify the exact source of the difficulty and to extend this knowledge to predict the perception of any non-native contrast in any other language" (Grenon, 2010, p. 158), and its aim is to "account for the categorical processing of speech sound contrasts in perception" (Grenon, 2010, p.74) with the approach focusing on neural mapping. It attempts to link neural processing with phonetic categories, especially features and allophones. In the BLIP model, acoustic cues such as noise bursts, formants, and periodicity are assumed to be processed by neurons especially tuned to these components.

The set of neurons that process the variations of an acoustic cue into one category is called a neural map. Table 2.1 shows neuron types activated by each acoustic cue related to the processing of possible speech sounds or contrasts. Note that one neural map can be associated to only one feature. According to this correspondence, for example, Japanese stop voicing contrast /p, t, k/ vs. /b, d, g/, where the former is generally produced with short-lag VOT and the latter is produced with prevoicing (i.e. with periodicity prior to the burst release), are processed by Phase or Time-locked neurons or combination-sensitive neurons that encode the onset of periodicity relative to the onset of the burst signal. That is, a Japanese native speaker has neurons sensitive to variations in Time-locked neurons or combination-sensitive neurons, and they are activated when he/she listens to Japanese stops.

Table 2.1. *Correspondence between acoustic cue, linguistic percept and type of neural response proposed in Grenon (2010).*

Acoustic cues	**E.g. Speech sound/contrast**	**Neuron type**
	Spectral components	
Noise bursts	Stop place contrast	Center frequency and bandwidth of noise burst (NB)
Spectral peaks (e.g. formants)	Vowels, liquids. fricatives, glides	Constant frequency components (CF)
Modulated spectral peaks (e.g. formant transitions)	Stop place and voicing contrast	Frequency-modulated components (FM)
	Temporal or synchronous components	
Timing cues	Durational contrasts, VOT	Phase or Time-locked (discharges) and/or combination-sensitive neurons
Periodicity	Voicing contrasts	Amplitude-modulated components (AM)
Pitch correlates (e.g. F0)	Intonation, lexical tone, stress, and accent	Amplitude-modulated components (AM)

Note. Reprinted from *The Bi-Level Input Processing Model of First and Second Language Perception* (p.61), by I. Grenon, 2010.

In the BLIP model, speech sounds are processed through two levels, a Neural mapping level and a Phonological level. When a listener hears a sound, its acoustic cue triggers firing of the corresponding neurons at the Neural mapping level. Then, at the Phonological level, the neurons which encode the corresponding phonological feature are activated, allowing the listener to identify the sound. In other words, at the Neural mapping level, a group of neurons corresponding to a sound contrast in the L1 are activated by the acoustic cue in the input, and then at the Phonological level these neurons are associated with the relevant abstract feature, which enables a listener to contrast the sounds. Grenon illustrates the processing of speech sounds in the BLIP model as Figure 2.1.

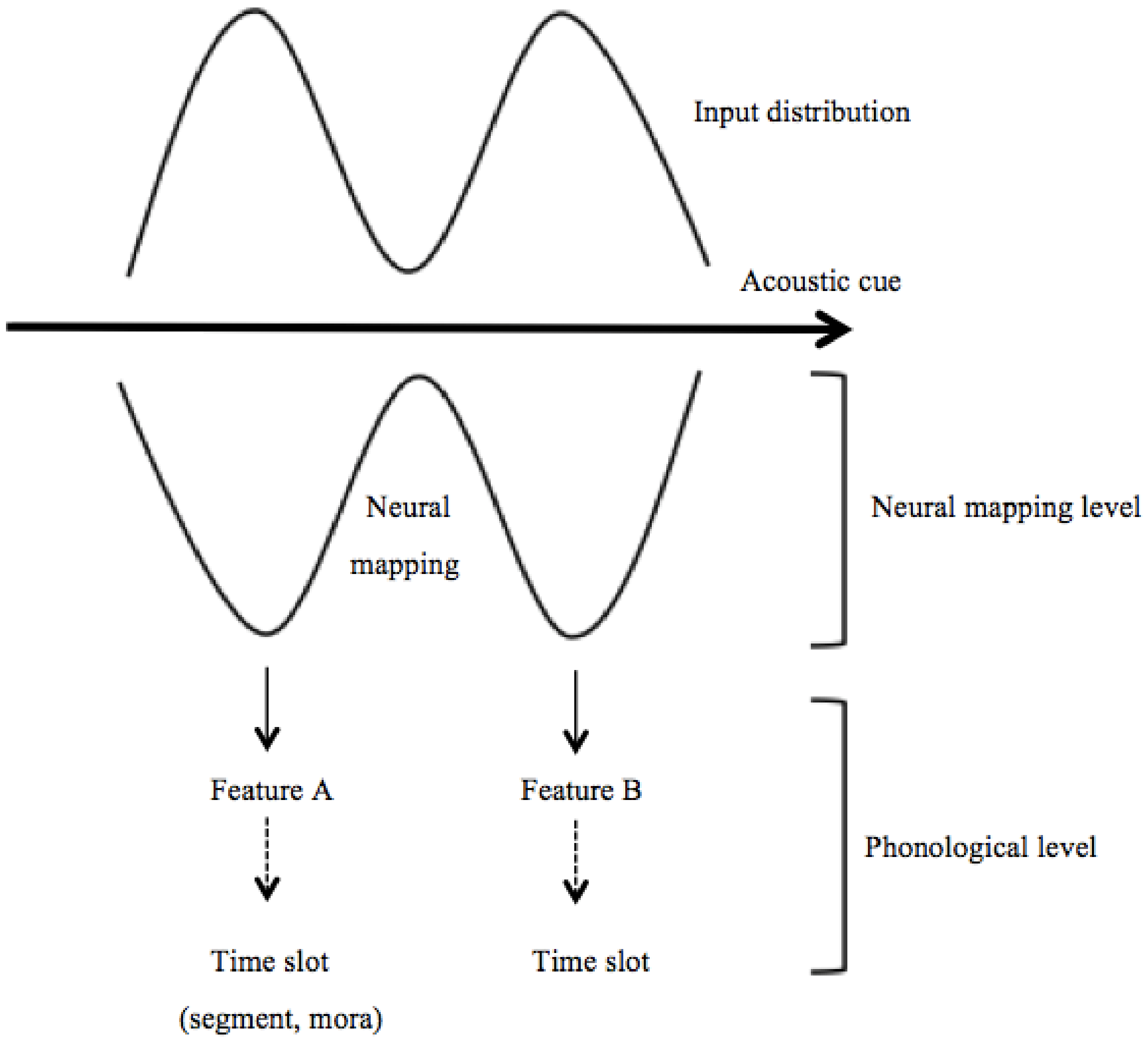

Figure 2.1. Processing of speech contrasts according to the BLIP model. Reprinted from *The Bi-Level Input Processing Model of First and Second Language Perception* (p.79), by I. Grenon, 2010.

Grenon (2010) demonstrate examples of speech processing of various speech contrasts in many languages, including stop voicing contrast. As for stop voicing contrast, she illustrated the distinguishing process with the cases in French and General American English.

In French, voiced stops /b, d, g/ are distinguished from voiceless stops /p, t, k/ by the presence of prevoicing. On the other hand, stop sounds in initial position in General American are discriminated not by the presence of prevoicing but by length of VOT: that is, voiced stops have short-lag VOT while voiceless stops have long-lag VOT. Thus, although voiced stops in General American are transcribed as "b d g" orthographically, they are acoustically voiceless in initial position. This fact implies that a voicing contrast in General American is not based on the presence or absence of periodicity during the stop closure in initial position (i.e. prevoicing) which is processed by Phase or Time-locked neurons or combination-sensitive neurons. Nevertheless, Grenon (2010) argues that "given that English speakers are sensitive to the presence of periodicity for the voicing contrast in fricative they should be able to perceive stop voicing contrasts given the proper testing conditions" (Grenon, 2010, p.113). That is, although English speakers do not have neural mapping for periodicity in initial voiced stops because they are acoustically voiceless in English, they can actually perceive periodicity even in word-initial position by using the neural mapping for periodicity that they use in other phonetic contexts.

The current study adopts the BLIP model for the prediction of the results for two reasons. First, other models on L2 phonetic acquisition such as Best (1995)'s Perceptual Assimilation Model (PAM) and Flege (1995)'s Speech Learning Model (SLM), have one common precondition hindering the prediction of the present study. The predictions of these models are based on "perceptual similarity", which is difficult to assess objectively, and which may not work well with bilinguals, since the concept of "nativeness" may be ambiguous for them. For example, PAM describes phonetic similarity in terms of gestural features such as the degree and the place of the tongue, lip postures, and velar gestures, but it is technically difficult to determine the exact position of these articulators for the similarity judgment. Similarly, several previous studies working within SLM and PAM have assessed perceptual similarity using either informal L1 transcriptions by non-native listeners or a cross-language identification, such as rating the category goodness of an L2 sound as an exemplar of the L1 category with the scale of 1 being very foreign sounding to 7 being very native sounding. It is inevitable, however, for the results of these techniques to be subjective and intuitive allowing high inter-subject variability in one's judgment. The BLIP model enables to avoid these problems.

Second, the present study decided to use the BLIP model for the prediction of the results because the aim of the BLIP model is more suited to the current study compared to other L2 acquisition models. While other models, such as the PAM and SLM models, aim to predict the difficulty of L2 speech contrasts, while the purpose of the BLIP model is to identify the source of the difficulty as mentioned at the beginning of this subsection. Similarly, the present study intends to determine what makes the perception of bilinguals easy or difficult compared to that of monolinguals, rather than to investigate which sound contrast is more difficult than the others. For these two reasons, the present study will use the BLIP model for the predictions of the experiments.

In 2.4, as a premise, stop contrasts and VOT in Mandarin, Japanese and English are briefly explained in 2.4.1, followed by the definition of the terms used in this study (2.4.2). After that, I will discuss what the BLIP model would predict on the research question of the current study with the hypotheses for perception results (2.4.3) and for production results (2.4.4).

2.4 Hypothesis

The present study examined the production and perception of English VOT by Mandarin-Japanese bilinguals, comparing them with those of monolingual speakers of Mandarin, Japanese and English. In this section, the characteristics of VOT as well as the stop contrasts in Mandarin, Japanese, and English are introduced first in 2.4.1, followed by the definition of the terms used in this paper in section 2.4.2. Then, the predictions based on the current study are inferred from the view of the BLIP model (Grenon, 2010) for perception (2.4.3) and production (2.4.4).

2.4.1 Stop contrasts and VOT

The aim of the current study is to evaluate whether L1, L2 or both have an impact on the accuracy of L3 perception and production by bilingual speakers. To evaluate this question, I conducted experiments on English stop contrast /p, t, k/ and /b, d, g/ by Mandarin Chinese and Japanese bilinguals. The reason why the stop contrast was of particular interest for this study is that it plays an important role in many languages including English, and also it is realized differently depending on languages. Furthermore, VOT is relatively easy to measure acoustically and therefore enables researchers to observe the performance of a speaker objectively. In fact, a considerable number of previous studies on L3 phonetic acquisition focused on VOT (Tremblay, 2007; Llama et al., 2010; Wunder, 2010; Wrembel, 2011; Sypiańska, 2013; Wrembel, 2014).

Languages split between those in which the duration between the stop release and onset of periodicity plays a distinctive role (e.g. Mandarin Chinese) and those in which its duration does not have a distinctive role (e.g. Japanese). When investigating this feature, "the time interval between the burst that marks release of the stop closure and the onset of quasi-periodicity that reflects laryngeal vibration" (Lisker and Abramson, 1964, p. 422) is often measured. It is called VOT (Voice-Onset-Time). According to Lisker and Abramson (1964), stop sounds are either of three modal ranges of VOT; lead (negative value), short-lag (positive but small value), and long-lag (positive and large value). Some languages use these VOT ranges differently in order to make phonological contrasts such as /p/ vs. /b/. For example, although there is no "aspirated" consonant as an independent phoneme in English as stated above, English stops in a stressed syllable not preceded by /s/ are pronounced as "aspirated" consonants with long-lag VOT (cf. after an initial /s/, English voiceless plosives have short-lag VOT as in voiced plosives). Table 2.2 summarizes the mean VOT values in English voiced and voiceless plosives in the previous studies. Note that as shown in Table 1.1, English initial voiced plosives may have both short-lag VOT as in Harada (2007) and lead VOT as in Matsuura & Shimizu (2002). From these facts, it can be inferred that what is important for stop contrast seems to be the difference in VOT values, rather than absolute VOT values.

Table 2.2. *Summary of mean VOT (ms) for English initial voiced and voiceless stops reported in the previous studies.*

	/b/	/d/	/g/	/p/	/t/	/k/
Lisker & Abramson (1964)	1	5	21	58	70	80
Homma (1985)	-80	-68	-47	44	48	75
Matsuura & Shimizu (2002)	-88	-74	-88	68	82	85
Harada (2007)	7	19	22	68	80	88

As for Japanese, on the other hand, VOT values of Japanese plosives are much shorter than in English according to Homma (1985). Table 2.3 shows the mean VOT values of Japanese initial voiced and voiceless stops in the previous studies. The major difference between English is that Japanese has lead VOT for /b d g/. As in English, VOT in Japanese voiced plosives performs a wide range of VOT values, too, but most of them have lead VOT. Also, an increasing number of young native speakers of Japanese produce longer VOT for initial voiced plosives. Due to these weak VOT of Japanese stops, VOT is not a primary cue for the native speakers of Japanese in contrasting the two categories, but rather a secondary cue which has a subsidiary role in distinguishing stop contrasts (Sato, 1958; Wilson & Hashimoto, 2013). Here, their primary cues for voicing are reported to be prevoicing[1] and F0 or F1 transition. This means that Japanese use multiple phonetic cues for stop contrasts, and therefore when they listen to non-native stop contrasts their results would be different regardless of VOT values of tokens, which could also cause the difference of the perception experiment results between English and Japanese monolinguals. The tendency of using multiple phonetic cues in contrasting stop sounds is not limited to the case of Japanese but also other languages such as English, Korean and Spanish, although the usage pattern of these cues is different according to the languages (Oglesbee, 2008). In order to investigate only the effect of VOT, in the present study, only VOT will be manipulated while other cues such as formant transitions will be kept constant.

[1] In Sato (1958)'s study, he considers prevoicing and aspiration as two different phonetic cues.

Table 2.3. *Summary of mean VOT (ms) for Japanese initial voiced and voiceless stops reported in the previous studies.*

	/b/	/d/	/g/	/p/	/t/	/k/
Homma (1985)	-71	-78	-64	24	32	66
Shimizu (1996)	-89	-75	-75	41	30	66
Harada (2007)	-27	-34	1	24	26	42
Riney et al. (2007)				20	30	55
Takada (2011)	-50	-49	-56			

Looking at Mandarin, previous studies report long-lag VOT for its aspirated plosives /p^h, t^h, k^h/, while unaspirated plosives /p, t, k/ are realized with short-lag VOT (see Table 2.4 for details). One can see that in Mandarin no lead VOT is observed, and the values long-lag VOT are much longer than both English and Japanese. In addition, compared to English and Japanese, Mandarin shows less variation in VOT values across the previous studies.

Table 2.4. *Summary of mean VOT (ms) for Mandarin initial unaspirated and aspirated stops reported in the previous studies.*

	/b/	/d/	/g/	/p/	/t/	/k/
Rochet & Fei (1991)				100	99	110
Shimizu (1996)	7	12	19	96	98	112
Chao & Chen (2008)	14	16	27	82	81	92

The simplified distribution of VOT in each language is shown in Figure 2.2. Comparing Mandarin Chinese, English and Japanese in terms of the VOT length it can be seen that Mandarin shows the longest long-lag VOT for [p^h] followed by English [p]. As for the short-lag VOT, Mandarin [p], English [b] and Japanese [p] exhibit a similar range. The Japanese voiced consonant [b] is usually prevoiced (i.e. its VOT value is negative).

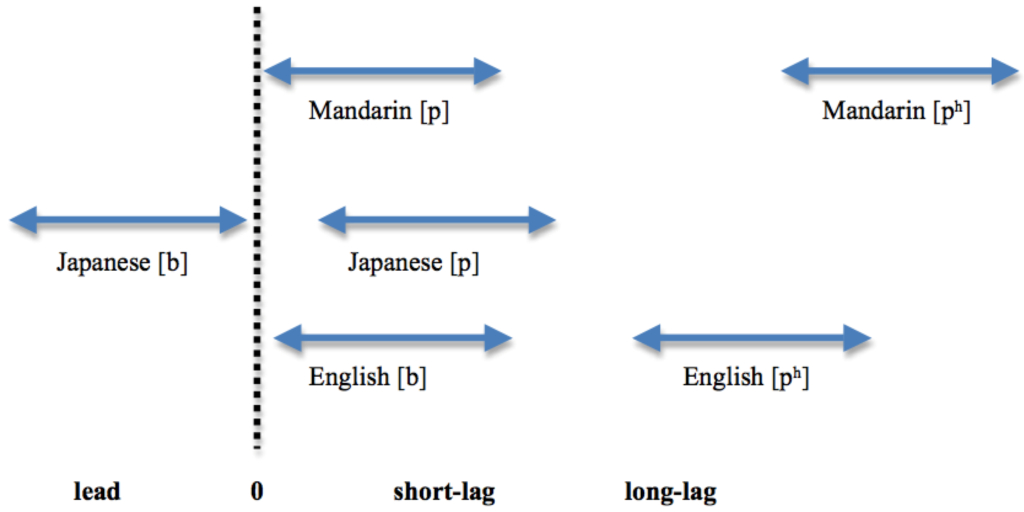

Figure 2.2. The general distribution of VOT in Mandarin Chinese, English and Japanese.

Not only VOT changes according to languages, VOT can also be affected by several factors. For example, it is generally suggested that the VOT values become longer as the place of articulation moves backward (Cho & Ladefoged, 1999). That is, the VOT values of velar stops (e.g. /k/) are the longest, while those of bilabial stops (e.g. /p/) are the shortest. As another influential factor, the effect of vowel context has been reported, claiming that stops before high vowels such as /i:/ show longer VOT values than those before mid (e.g. /e/) and low vowels (e.g. /ɑ:/) (Klatt, 1975). In order to eliminate the effect of vowel context, the target words of the present study contained only low vowels across all languages. Also, the VOT values can be affected by tones: the VOT values associated with falling tone (Fourth tone) were significantly shorter than those in rising tone (Second tone) and low tone (Third tone) in Mandarin (Liu et al., 2008). Hence, for the current production study, I used the minimal pairs with the same tones for the target words in Mandarin, mainly high tone (First tone). Due to the absence of the appropriate minimal pairs with high tone, those with mid rising tone (Second tone) and high falling tone (Fourth tone) were also employed.

In the BLIP model, timing cues such as VOT are processed by phase or time-locked neurons. At a Neural mapping level, the VOT cue prompts the firing of phase or time-locked neurons when a listener hears stop contrasts in any languages. The set of neurons associated with a given range of VOT values, called a neural map, are then are associated with an abstract feature at the Phonological level.

Languages differ in how to contrast stop sounds and which features are associated with them. First of all, languages can be divided into three groups: those that use a voicing contrast, those that use an aspiration contrast, and those that use both contrasts. For example, English has two types of stops, voiceless /p, t, k/ and voiced /b, d, g/, which form minimal pairs such as *tie* /taɪ/ and *die* /daɪ/. Thus, a native speaker of English is supposed to have features [+voice] and [-voice] for English word-initial voicing contrast. However, the phonetic differences between the two in word-initial positions are not made by whether one is voiced and the other is voiceless. Rather, the relative length of VOT after the release of an articulation makes the difference. The so-called "voiced" stops are actually voiceless stops with short-lag VOT [d̥] in initial position, such as in *die* [d̥aɪ]. As for the "voiceless" stops, they are produced as voiceless with long-lag VOT in word initial position [t^h], as in *tie* [t^haɪ]. Therefore, the native speakers of English actually do not rely on the presence or absence of periodicity (that is, lead VOT) during the stop closure to differentiate their stops in initial position, but on the duration between the stop release and onset of periodicity (short-lag versus long-lag VOT). Note that the English /b, d, g/ and /p, t, k/ contrast is still presumably associated with a [+voice] and [-voice] contrast at the underlying level, according to the BLIP model, because there are other acoustic cues associated with those sounds that suggest a voicing contrast—rather than a [-voice] and [+spread glottis] contrast as in Mandarin Chinese—such as the formant transitions (e.g. in voiced stops, the first formant is excited during the entire CV transition, while the first formant is not excited until very late in the CV transition in voiceless stops).

Now let's have a look at Japanese. In the same way as English, Japanese also has voiced and voiceless categories for stop sounds, /b, d, g/ and /p, t, k/. Therefore, Japanese native speakers are supposed to have features [+voice] and [-voice] for Japanese word-initial stop contrast, in a same way as English native speakers. Unlike English, however, the difference of the two categories [t] vs. [d] is between lead VOT and short lag VOT, such as in *dai* ("table") [daɪ] and *tai* ("sea bream") [taɪ].

As for Mandarin Chinese, it also has two types of stops, but it is not a voicing contrast as in English, because it is not associated with a [+voice] and [-voice] contrast at the phonological level. Neither is it a voicing contrast as in Japanese, because it is not contrasted by the presence/absence of periodicity. Mandarin stop contrasts are based on whether a stop is pronounced with short-lag VOT or long-lag VOT, in other words, whether a stop is "unaspirated" or "aspirated": /p, t, k/ or /p^h, t^h, k^h/ as in *tā* ("he") [t^há] and *dā* ("add to") [tá]. This means that the features associated with the stop contrasts in Mandarin are expected to be [-voice] and [+spread glottis]. The native speakers of Mandarin Chinese therefore are not attuned to an underlying voicing distinction. In fact, according to Chen (2007), the voicing difference in languages with voicing contrast such as English, Spanish or Japanese is hard to perceive for Mandarin native speakers. Wang (1995) indicates that because the lead VOT and short VOT plosives are allophones of [-voice] plosives, Mandarin native speakers have difficulty in differentiating between English voiced and voiceless stops, and "/b/ is often perceived and produced as unaspirated /p/, and /p/ as /p^h/ regardless of its context" (Wang, 1995, p.39).

Thus, Mandarin Chinese, Japanese and English have different initial stop contrasts and VOT value ranges. In the following subsection, the definition of the terms regarding bilingualism used in the present study is discussed.

2.4.2 Definition of the terms

A general meaning of the word "bilingual" is speaking two languages fluently, which is relatively vague. In fact, in the previous studies in phonetic acquisition, the term "bilingual" has been used in a wide variety of meanings. For example, in a series of studies by Flege and his colleagues, the term "early bilingual" means a native speaker of L1 who begins to learn L2 as children while a native speaker of L1 who begins to learn L2 "in late adolescence or early adulthood" (Piske et al., 2001, p.204) is referred to as "late bilingual". Here, "bilingual" does not always imply native-like fluency, for L2 performance of late bilinguals who start to learn early adulthood tends to have a strong L1 accent. Also, the actual age of "adolescence" or "adulthood" has not been clearly defined and therefore until when the early bilinguals begin to learn L2 is varied according to the studies. On the other hand, other studies such as Garcia-Sierra et al. (2011) use "bilinguals" to refer to those who begin to learn both L1 and L2 in their infancy. In this case, "bilinguals" has the same meaning as "early bilinguals" in the former case. In the present study, "bilingual" refers to a person who begins to learn two languages mainly in a natural setting during their childhood. Because they learn both L1 and L2 in a natural setting, their input of L1 and L2 is mainly from native speakers of L1 and L2. The present study also divides bilinguals into two groups according to the age of which they begin to learn their L2: that is, L1 native speakers who begin to learn L2 from 0-9 years old are categorized as early bilinguals, while those who begin to learn L2 from 10-15 years old are categorized as late bilinguals. Here, early bilinguals learn L2 either simultaneous to L1 or following the establishment of L1, and for most of the early bilinguals L2 is a dominant language. Late bilinguals, on the other hand, begin to learn L2 during/after puberty and therefore the dominance is first on L1 and switches to L2.

Similarly, the word "monolingual" generally describes someone who speaks only one language. In this study, "monolingual" speakers refer to native speakers of L1 who have learnt other languages, mainly English, during/after puberty at school. The difference between late bilinguals and monolinguals is that late bilinguals reside in an L2 speaking country receiving native L2 input, and the language dominance switches from L1 to L2, while monolingual speakers reside in an L1 speaking country with a scarce input of L2, and the language dominance remains in L1, not L2.

As for additional languages, there is no common view on how they are called, and different researchers use different terms for them. In the previous studies, it is common to use terms based on the concept of the linear order of language learning. In this case, the language is numbered in an acquisition order, that is, the language acquired first is called L1, the second acquired is L2, and the third one is L3. In other studies, an additional language is named based on whether it is learned before or after puberty, and the one acquired before puberty is L1, while the one after puberty is L2. The present study, however, do not adopt either trend, because the linear order of acquisition is hard to determine for simultaneous bilinguals. Instead, considering that the bilinguals in this study are immigrants from China to

Japan, L1 refers to Mandarin, L2 to Japanese, and L3 to English. It is worth noting that for some of the late bilinguals in this study, the sequential order of language learning differs from this: after they acquired Mandarin at home, they learned English at school in China before immigrating to Japan and learning Japanese. Therefore, the late learners' actual linear order of language learning is Mandarin, English and Japanese. However, because their dominant language after immigration is Japanese and they did not use English outside the English classes, their de facto L2 can be considered not as English but Japanese and, therefore, I consider that Mandarin is the L1, Japanese is the L2, and English is the L3 of all bilinguals in this study.

2.4.3 Hypothesis for perception

In the BLIP model, the difficulty of non-native speech contrasts is determined in terms of the number of activated neural maps in the L1 and L2, and the linkage between the neural maps and underlying features. When an L2 acoustic cue is processed by the same number of neural maps as in the L1, and when the neural maps are linked to the same features in the L1 and L2, a listener is able to perceive the contrast both at the neural and phonological levels, that is in tasks that triggers phonetic or phonological processing (e.g., an identification task with long inter-stimulus intervals is supposedly conducive to phonological processing). In this case, he/she is predicted to have the least difficulty in L2 speech perception. In both English and Japanese, there is only one contrast along the VOT dimension, and the VOT contrast is associated with the features [+voice] and [-voice] at the phonological level. As the BLIP model assumes that the neural maps are flexible and should adapt to new VOT values with sufficient input, it is expected, in the current study, to be easy for Japanese native speakers to perceive the English voicing contrast in a native-like fashion.

On the other hand, when the neural maps are linked to different features in the L1 and L2, the L2 perception is predicted to be more difficult, because the L2 contrast is perceived only at the neural mapping level. That is, there is a conflict occurring at the phonological level, when using a task conducive to phonological perception. The Mandarin learners of English applies to this pattern in perceiving initial stop contrast in English. Here, Mandarin monolinguals have neural maps corresponding to short-lag and long-lag VOT as in English. However, at the phonological level, these neurons are associated with the features [-voice] and [+spread glottis] in Mandarin, whereas they are associated with the features [+voice] and [-voice] in English. This conflict makes it difficult for Mandarin native speakers to perceive the contrast in the same way as English native speakers. Thus, the BLIP model predicts that it is more difficult for Mandarin native speakers to perceive English initial stop contrasts compared to native speakers of Japanese.

When it comes to bilinguals, Grenon (2010) does not directly refer to the prediction of bilingual speech. However, providing that they have both Mandarin-attuned neural mapping and Japanese-attuned neural mapping, the BLIP model may predict that it is difficult for the Mandarin-Japanese bilinguals to perceive English stop contrasts, because bilingual speakers have three neural maps associated to features [+voice], [-voice], and [+spread glottis] along the VOT dimension, while there are only two dimensions in English, for the features [+voice] and [-voice]. That is, the VOT dimension in bilinguals is more crowded with three neural maps, with lead VOT associated with the feature [+voice], short VOT associated with the feature [-voice], and long lag VOT associated with the feature [+spread glottis]. Whereas in English, the short-lag VOT is associated with a [+voice] feature and long-lag VOT with a [-voice] feature. How will bilinguals resolve this conflict is the focus of the current investigation? I see two possibilities here, either they can rely on Japanese-attuned neural maps, for the number of neural maps and underlying features associated with VOT cues ([+/- voice]) are more similar to English mapping at the phonological level, and accordingly show some L2 transfer by using a categorical boundary closer to the Japanese categorical boundary, that is, by having a categorical boundary shorter than the English one (shorter VOT). Or, they can rely on Chinese-attuned neural maps, because the neural maps are more similar to the English neural maps at the neural mapping level. That is, both English and Mandarin use short-lag and long-lag VOT to contrast stops at the neural mapping level. If this is the case, the categorical boundary may tend to be longer than the English categorical boundary (longer VOT) and therefore, exhibit some L1 transfer.

Thus, the prediction by the BLIP model is: perceiving English stop contrasts is difficult for Mandarin-Japanese bilinguals and Mandarin monolinguals at the phonological level. In the case of bilinguals, since Japanese and English use the same underlying contrast and I use a phonological task for my experiment, I expect that the perception of English VOT will be affected by L2 Japanese, that is, it will show some L2 transfer. This prediction agrees with the one mentioned in 2.1.2.

2.4.4 Hypothesis for production

As mentioned in 2.3.3, the aim of the BLIP model is to predict the perception of non-native contrasts, and therefore Grenon (2010) does not predict non-native production. However, I will base my predictions for production on the premises explained in the previous section. According to the BLIP model, what is different among the native speakers of the three languages is the mapping of the VOT neuron. For example, VOT mapping of the Mandarin native speakers is expected to consist of [-voice] and [+spread glottis] ranging from 0 ms to 110 ms, while in English [-voice] and [+voice] ranging from 0 ms (or optionally from -50 ms) to 80 ms, and in Japanese [-voice] and [+voice] ranging from -50 ms to 70 ms, as illustrated in Figure 2.3.

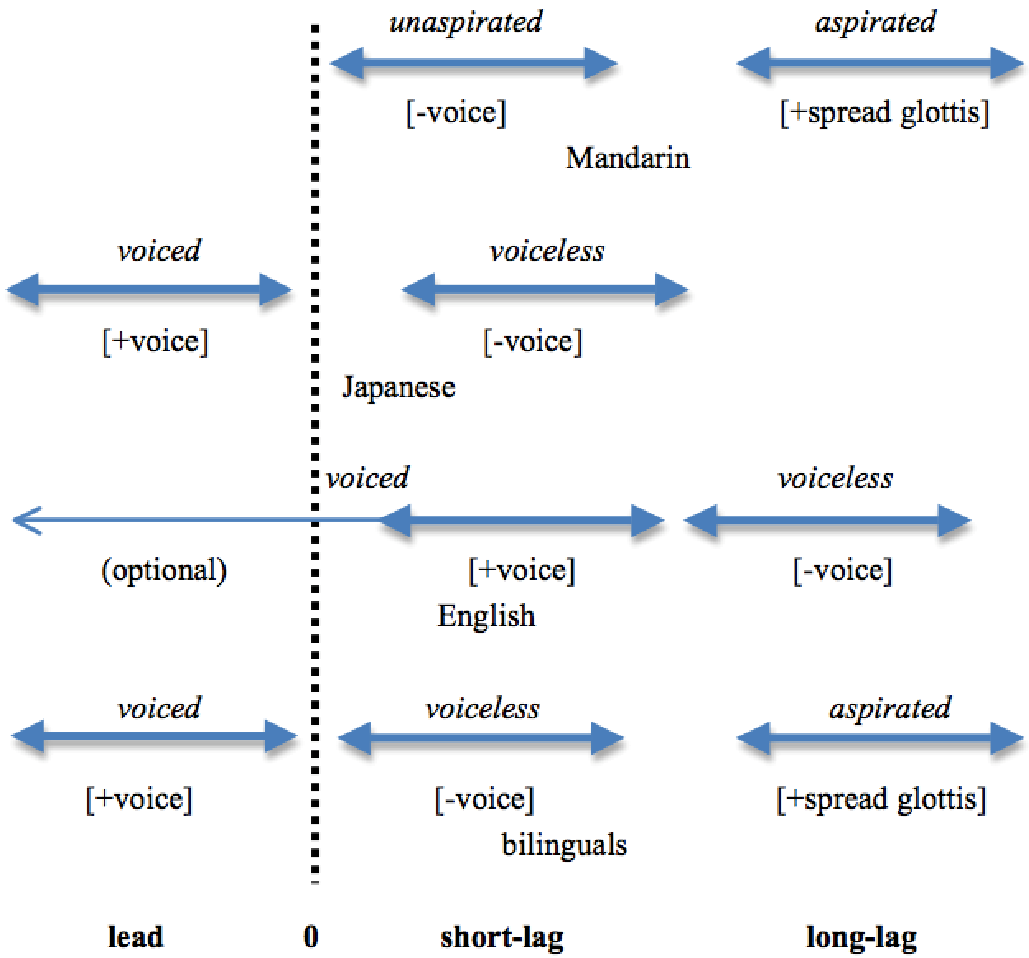

Figure 2.3. Predicted mapping of the VOT as an acoustic cue for Mandarin monolinguals, Japanese monolinguals, English monolinguals and Mandarin-Japanese bilinguals.

When it comes to Mandarin-Japanese bilinguals, it is expected that they have both groups of neurons for the two languages, namely, Mandarin and Japanese. Thus, the bilinguals have a wider mapping of VOT than monolingual speakers of Mandarin, Japanese, and English. Figure 2.3, displaying the expected ranges of VOT for Mandarin-Japanese bilinguals and monolingual speakers of the three languages, shows that the bilinguals' VOT range covers all VOT mappings for Mandarin, Japanese and English. This assumption suggests that bilinguals are sensitive to any VOT values within the ranges of the three languages. Therefore, it can be hypothesized that Mandarin-Japanese bilinguals are able to produce English stop contrasts utilizing their knowledge of the two languages at the neural mapping level, unless there is an effect of L1 transfer, L2 transfer or combined transfer. If there is an L1 transfer, the L3 VOT produced by bilinguals would be short-lag and long-lag, closer to the average value for the Mandarin VOT contrast. If, on the other hand, there is L2 transfer, bilinguals are expected to produce L3 VOT with some lead and short-lag values, like in Japanese. By contrast, if a combined effect of L1 and L2 is observed, the L3 VOT values of bilinguals would occur somewhere between those of L1 monolinguals' and L2 monolinguals' as in the previous studies such as Sypiańska (2013) and Wrembel (2014). Since the three languages in the current study are typologically (i.e. genetically) unrelated considering that none of them belong to the same language family, no effect of language typology such as mentioned in 2.2.2 can be expected.

This prediction is also applicable to L1 and L2 production of bilinguals. For example, in the production of L1 Mandarin, bilinguals are expected to exhibit native-like values of VOT if they are sensitive to Mandarin VOT. When there is a transfer from L2 Japanese, however, their Mandarin VOT values would be shorter than those of Mandarin monolinguals. Similarly, in the production of L2 Japanese, bilinguals would be able to produce Japanese VOTs in a similar manner to Japanese monolinguals given that bilinguals have a group of neurons sensitive to any Japanese VOT values. If a significant effect of L1 transfer is observed, they would produce longer VOTs in Japanese compared to Japanese monolinguals.

Now, let's review the research questions of the current study. They are followed by the predictions of these research questions. As for the third research question regarding AOA, the prediction is made based on the results of the previous studies introduced in 2.1.4.

Research questions:

1) Which previously learnt language, L1 or L2, has a stronger impact on L3 phonological acquisition? Or will both L1 and L2 have no impact on L3 phonological acquisition? In particular, which of Mandarin or Japanese affects the results of English production experiment by Mandarin-Japanese bilinguals? Will the results of English perception experiment by bilinguals be closer to the results by Mandarin and Japanese monolinguals?
2) Will the bilinguals' L1 and L2 production be distinct from that of monolinguals? Will the results of production experiments in Mandarin and Japanese by Mandarin-Japanese bilinguals be similar to those of Mandarin monolinguals and Japanese monolinguals?
3) Will the bilinguals' age of arrival in an L2 country affect their L3 phonological acquisition? In particular, does the bilinguals' age of arrival in an L2 country have an influence on the results of production experiment and perception experiment in L3?

Predictions for the answers of the research questions

1) Perceiving L3 stop contrasts is difficult for bilinguals, and they are expected to be affected by L2 rather than L1. On the other hand, producing L3 stop contrasts is predicted to be in-line with English VOT values, due to no/less effect of language transfer in production.
2) The pattern of bilinguals' L1 and L2 production is predicted to be identical to that of monolinguals.
3) For both perception and production of L3, bilinguals who started to live in Japan at the younger age are expected to perform more accurately than those with older age of arrival in Japan.

Here, the predictions for L3 perception and production are different: in perception bilinguals are predicted to be influenced by L2, while in production they are not expected to be affected by L1 nor L2. This is because the tasks in perception and production are different, an identification task in perception along a VOT continuum, which looks specifically at how listeners identify the categorical boundary between the English VOT contrast, and a wordlist reading task in production, which evaluates the average VOT values produced for each contrast of interest. In the following chapter, the methodology and the results of the production experiments in Mandarin, Japanese, and English, and the perception experiment in English are described in detail.

3. VOT Production in L1, L2 and L3

The production experiment was designed to observe the characteristics of English VOT production by Mandarin-Japanese bilingual speakers who were born to Chinese parent(s) and immigrated from China to Japan before the age of 15. The production test was conducted not only for English but also for Mandarin Chinese and Japanese to compare the bilinguals' production patterns also in L1 and L2 with the production patterns by monolingual native speakers of each language. As an analysis, I measured the mean VOT for each group and compared the average VOT values of bilinguals with native speakers in the three languages. The overall time line of the production experiments is shown in Table 3.1.

Table 3.1. *The time line of the three production experiments for each experimental group*.

Bilinguals	Mandarin monolinguals	Japanese monolinguals	English monolinguals
Language questionnaire	Language questionnaire	Language questionnaire	Language questionnaire
⇩	⇩		
Mandarin experiment	Mandarin experiment	⇩	⇩
⇩			
Japanese experiment	⇩	Japanese experiment	⇩
⇩		⇩	
English experiment	English experiment	English experiment	English experiment

Section 3.1 introduces the Mandarin VOT production experiment, as produced by Mandarin-Japanese bilinguals and Mandarin native speakers. Section 3.2 presents the Japanese VOT production experiment, as produced by bilinguals and Japanese native speakers. Finally, section 3.3 presents the results of the English VOT production experiment

as produced by bilinguals and native speakers of each language, Mandarin, Japanese, and English. Section 3.4 summarizes the results of the three production experiments.

3.1 Experiment I: Production of VOT in Mandarin

In order to observe whether bilinguals perform in the same or similar way as monolinguals speakers, the current study examines the production of VOT in Mandarin Chinese by Chinese-Japanese bilinguals and native speakers of Mandarin Chinese.

3.1.1 Participants

Ten native speakers of Mandarin Chinese with no or fairly limited Japanese experience, the monolingual Mandarin group, took part in this experiment: 4 were male and 6 were female. Ages ranged from 19 to 22 years (M = 20.9, SD = 1.14 years). See Table 3.2 for the summary of their background. Their L1 was Modern Standard Mandarin which is based on the Beijing dialect, while their L2 was English. They had received the "standard" English language training in the mainland China: they started to learn English as a foreign language at the age of 6 or 10. They never lived in a country where English is overwhelmingly spoken, including Hong Kong, Taiwan, Macau, and Singapore, for more than a month. Monolingual participants were asked to fill out a language background questionnaire in order to inquire about their linguistic background and English ability (see Appendix A for the format).

Table 3.2. *Background information for Mandarin monolingual participants in the current research.*

ID	Gender	Age	AOL Eng	SelfEval Eng
NC01	M	20	11	InterH
NC02	M	20	9	InterL
NC03	F	20	12	InterH
NC04	F	22	10	InterH
NC05	M	22	5	InterH
NC06	F	22	9	InterL
NC07	M	22	13	InterL
NC08	F	22	8	InterH
NC09	F	20	6	InterH
NC10	F	19	7	Adv

Note: Age = age at the time of the experiment (in years); AOLEng = the age of first exposure to English; SelfEval = self-evaluation of the proficiency of English based on five levels (Beg = Beginner, InterL = Lower intermediate, InterH = Higher intermediate, Adv = Advanced, Native = Native/native-like).

The experimental group of this study consisted of 16 Mandarin Chinese native speakers who immigrated to Japan before or around puberty. This means that all participants in the experimental group were Mandarin-Japanese bilinguals. There were 6 male speakers and 10 females. Ages ranged from 15 to 24 years (M = 19.31, SD = 1.86 years). All bilinguals resided in Japan and were undergraduate or high school students at the time of the experiment. The bilinguals were divided into two groups in terms of their age of arrival (AOA) in Japan, in order to investigate the effect of AOA in the L1 production. The participants of the first group, who are referred to as early bilinguals in this study (EB), arrived in Japan between the age of 0 and 6 years old. The AOA of the second group was, on the other hand, between 10 and 15 years old, and the participants in this group are referred to as the late bilinguals (LB). As for early bilinguals, they learned English after their acquisition of Mandarin and Japanese, which means that English was their L3. As for late bilinguals, on the other hand, they learned English at school in China before they immigrated to Japan since English education in China usually starts around grade 3 (age 8). This means that their L3 was not English but Japanese in terms of their linear order of language learning. However, considering the fact that their knowledge of English was quite limited (i.e. the elementary level knowledge such as the Roman alphabet and basic vocabulary), while their knowledge of Japanese was generally more advanced than their knowledge of English or reported as their current dominant language, their de facto L3 will be considered English, not Japanese. The actual quantity and quality of English input the bilinguals received in their English education were not clarified in the questionnaire. Therefore, there is no telling how much L3 input they received in the classroom and whether their L3 input was from native speakers of English or from native speakers of Japanese. A summary of the background information of our bilingual participants in this study is provided in Table 3.3 and 3.4. As a collective term for bilinguals including both EB and LB, BL is used when allocating ID (see the leftmost column),

Table 3.3. *Background information for bilingual participants in the current research.*

ID	Group	Gender	Age	AOA	LOR (years)	Best Lg	Father Lg	Mother Lg
BL01	EB	F	20	0	20.1	Jp	Ch, Jp	Ch, Jp
BL02	EB	M	20	0	13.2	Jp	Ch	Ch
BL03	EB	M	20	0	20	Jp	Ch	Jp
BL04	EB	F	19	6	9	Jp	Ch	Ch
BL05	EB	M	18	4	14	Jp	Ch	Ch
BL06	EB	F	15	0	9	Ch	Ch	
BL07	EB	F	19	0	18.75	Jp	Ch	Ch
BL08	EB	M	19	0	17	Jp	Ch, Jp	Ch, Jp
BL09	EB	M	19	2	13	Jp	Jp	Ch
BL10	EB	M	21	0	11	Jp	Ch, Jp	Ch, Jp
BL11	EB	F	20	0	17	Jp	Ch, Jp	Ch, Jp
BL12	LB	F	20	10	10.3	Jp	Ch	Ch
BL13	LB	F	24	12	12	Ch	Ch, Jp	Ch, Jp
BL14	LB	F	20	14	6.5	Jp	Ch, Jp	Ch, Jp
BL15	LB	F	18	15	2.5	Ch	Ch	Ch
BL16	LB	F	17	15	1.4	Ch	Ch	Ch, Jp

Note: Age = age at the time of the experiment (in years); AOA = age of arrival in Japan (in years); LOR = length of residence in Japan; BestLg = self-reported language at which the participant is the most proficient; FatherLg = language(s) spoken by the father of a participant; MotherLg = language(s) spoken by the mother of a participant.

Table 3.4. *Language background information for bilingual participants in the current research.*

ID	Group	JpSchool	AOL Eng	SelfEval Ch	Jp	Eng	DomLg
BL01	EB	No	9	InterH	Native	InterH	Jp
BL02	EB	No	6	Native	Native	Adv	Jp
BL03	EB	No	13	Adv	Native	InterL	Ch, Jp
BL04	EB	No	12	InterH	Native	InterL	Jp
BL05	EB	No	12	InterH	Native	InterL	Jp
BL06	EB	No	7	Adv	InterH	InterL	Ch
BL07	EB	No	13	Adv	Native	InterH	Jp
BL08	EB	No	2	InterH	Native	InterL	Jp
BL09	EB	No	6	Native	Native	Adv	Jp
BL10	EB	No	11	InterL	Native	InterH	Jp
BL11	EB	No	8	Adv	Native	InterL	Jp
BL12	LB	Yes	6	Native	Native	InterH	Jp
BL13	LB	No	9	Native	Native	InterL	Ch, Jp
BL14	LB	No	6	Native	Native	InterH	Jp
BL15	LB	Yes	7	Native	InterH	InterH	Jp
BL16	LB	Yes	13	Native	Beginner	Beginner	Ch

Note: JpSchool = experience of attending Japanese language schools; AOLEng = the age of first exposure to English; SelfEval = self-evaluation of the proficiency of Mandarin Chinese, Japanese and English based on five levels (Beg = Beginner, InterL = Lower intermediate, InterH = Higher intermediate, Adv = Advanced, Native = Native/native-like); DomLg = participants' dominant language determined by their language use in 14 daily activities.

As shown in Table 3.3 and 3.4, all bilinguals except four (BL06, BL13, BL15 and BL16) reported that their most proficient language was Japanese. Similarly, all bilinguals except four (BL03, BL06, BL13 and BL16) recognized their dominant language as Japanese. In addition, three of five late bilinguals (BL12, BL15 and BL16) had an experience of attending a Japanese language supplementary school in Japan in order not to fall behind on their classwork. Although most bilinguals were in Mandarin-Japanese bilingual communities (i.e. at university they participated in the Chinese course for Mandarin-Japanese bilingual students or foreign students from China), they conducted their daily conversations in Japanese and seldom used Mandarin.

There were two bilingual speakers whose LOR was less than 3 years (BL15 and BL16 in Table 3.3). Although there were concerns that they were not eligible to represent "bilingualism," I included them in the analyses because after their immigration to Japan they used exclusively Japanese in the community in their daily life, such as at their school or at their Japanese language supplementary school, and used Mandarin only when they talked with their parents. Both of them also passed the same entrance exams for as other Japanese native speakers in order to enter a Japanese high school.

Depending on AOA of EB and LB, the proficiency, the frequency of use, and the pedagogical experience of Mandarin and Japanese were varied. For example, while the participants in EB tended to be able to perceive and produce Mandarin sounds without any difficulty due to their persistent use of Mandarin at home, they did not have sufficient Mandarin reading and writing ability because the education they had received was all conducted in Japanese and they did not have a chance to read or write Mandarin. For this reason, all bilingual participants completed a language background questionnaire which asked them, for instance, about their AOA, the length of residence (LOR) in Japan, the age at which they began to learn Mandarin/Japanese/English, the frequency of Mandarin/Japanese/English use, their experience of staying in English-speaking countries, and self-estimate of their ability in the three languages (see Appendix B for the format). The summary of the results in the language questionnaire for bilingual participants can be found in Table 3.4.

All participants, namely early and late Mandarin-Japanese bilinguals and Mandarin monolingual speakers, reported having no known hearing or speech impairments, and all received a monetary compensation for participating in the experiment.

3.1.2 Stimuli

In order to make their production as natural as possible, only real words with a high frequency of occurrence in Mandarin were used. The 24 target words consisted of one syllable, with initial aspirated/unaspirated plosives, because in Mandarin initial stop contrasts are made with the existence/absence of aspiration. These initial plosives were followed by low vowels with the same tone category, which created 12 minimal pairs varying in VOT. The target words are shown in Table 3.5.

Table 3.5. *Mandarin Chinese target words used in the current experiment.*

Unaspirated IPA	Mandarin	Meaning	Aspirated IPA	Mandarin	Meaning
[paɪ̯˧˥]	白	“white”	[pʰaɪ̯˧˥]	牌	“(sign)board”
[paʊ̯˥]	包	“wrap”	[pʰaʊ̯˥]	泡	“bubble”
[pan˥]	班	“class”	[pʰan˥]	潘	“Pan (surname)”
[pɑŋ˥˩]	棒	“stick”	[pʰɑŋ˥˩]	胖	“fat”
[taɪ̯˥˩]	带	“have”	[tʰaɪ̯˥˩]	太	“very/too”
[taʊ̯˥]	刀	“knife”	[tʰaʊ̯˥]	掏	“take out”
[tan˥]	单	“single”	[tʰan˥]	贪	“greedy”
[tɑŋ˥]	当	“just”	[tʰɑŋ˥]	汤	“soup”
[kaɪ̯˥]	该	“should”	[kʰaɪ̯˥]	开	“open”
[kaʊ̯˥˩]	告	“report”	[kʰaʊ̯˥˩]	靠	“rely”
[kan˥]	干	“dry”	[kʰan˥]	刊	“publication”
[kɑŋ˥]	钢	“steel”	[kʰɑŋ˥]	康	“peaceful”

3.1.3 Procedure

First, the participants filled in the language background questionnaire which was introduced in the previous section. Before they started the recording session, they watched a 2-minute-long animated movie in Mandarin Chinese, so that it would encourage the bilingual speakers to switch to Mandarin.

Recordings took place in a sound-proof recording booth at the University of Tokyo. The speech was recorded with a Marantz PMD660 digital recorder (24bit, 48kHz) and a SHURE SM58 microphone, a unidirectional dynamic vocal microphone. The participants were asked to read aloud the 24 target words embedded in the carrier sentence *wǒ lái shuō* ___ *zhè ge cí* ("I say the word ___.") at a natural speech rate. These sentences were printed on a paper both in Chinese characters and in pinyin. The order of sentences was randomized for each participant. The experimenter required them to read a sentence again when the participants mispronounced a target word. All sentences were repeated three times, which made 24 × 3 = 72 speech samples for analysis.

3.1.4 Results and discussion for Experiment I

The speech samples were analyzed acoustically to examine the VOT values of initial plosives in Mandarin words produced by the bilinguals and Mandarin native speakers. The VOT values were measured with the aid of wide-band spectrogram and waveform in Praat version 5.4.16 (Boersma & Weenink, 2015). Positive VOT were measured from the beginning of the burst to the onset of periodicity of the following vowel, while the length of negative VOT was determined by measuring the time between the beginning of periodic striations and the beginning of the burst. Here, however, all measured VOT values were found to be positive. Table 3.6 shows the mean VOT values for early bilinguals, late bilinguals and the native speakers of Mandarin Chinese.

Table 3.6. *Mean VOT values from data recorded in the Mandarin production experiment.*

		Mandarin monolinguals	Early bilinguals	Late bilinguals
unaspirated				
	labial	17.54	14.77	17.54
	alveolar	17.47	18.57	16.53
	velar	27.55	28.69	28.83
aspirated				
	labial	107.22	101.74	87.77
	alveolar	104.21	99.92	96.85
	velar	107.23	107.63	107.73

Note: All measurements are in milliseconds (ms).

For each participant, the mean VOTs for velar plosives were larger than for alveolar and labial plosives, although there were little differences in the VOT values of /p^h/, /t^h/ and /k^h/ in the production of Mandarin monolinguals and early bilinguals (/p^h/ = 107.22 ms, /t^h/ = 104.21 ms, /k^h/ = 107.23 ms for Mandarin monolinguals, and /p^h/ = 101.74 ms, /t^h/ = 99.92 ms, /k^h/ = 107.63 ms for early bilinguals).

Measured VOT values of each token were then compared and interpreted in terms of the participant group and subject (nested by the participant group) by statistical tests in R version 3.2.0 (R Core Team, 2015). A mixed-effects ANOVA evaluating the effects of Group (early bilinguals, late bilinguals, Mandarin monolinguals), VOT (unaspirated, aspirated), the interaction of Group × Participants, and the interaction of Group × VOT was conducted for the Mandarin VOT of labial, alveolar and velar plosives. Here, what is important is the result in the interaction of Group × VOT, because it shows whether there is a difference between bilinguals and monolinguals in their production of unaspirated and aspirated plosives, which is what I was looking for.

As for labial plosives, the mixed-effects ANOVA reported a significant effect of the interaction of Group × VOT when comparing late bilinguals and Mandarin monolinguals, $F(1, 103) = 5.87$, $p = 0.017$, and early bilinguals and late bilinguals $F(1, 110) = 4.75$, $p = 0.031$. When early bilinguals and Mandarin monolinguals were compared, however, there was no statistical significance, $F(1, 145) = 0.15$, $p = 0.695$ (See Figure 3.1). The mixed-effects ANOVA for alveolar plosives, on the other hand, did not show a significant effect of Group × VOT: $F(1, 145) = 0.44$, $p = 0.51$ for the comparison of early bilinguals and Mandarin monolinguals; $F(1, 103) = 0.85$, $p = 0.36$ for late bilinguals and Mandarin monolinguals; $F(1, 110) = 0.01$, $p = 0.92$ for early bilinguals and late bilinguals. Similarly, the significant effect of Group × VOT was not observed for velar plosives: $F(1, 145) = 0.0097$, $p = 0.92$ for the comparison of early bilinguals and Mandarin monolinguals; $F(1, 103) = 0.01$, $p = 0.92$ for the comparison of late bilinguals and Mandarin monolinguals; $F(1, 110) = 0.00$, $p = 0.1$ for the comparison of early bilinguals and late bilinguals. The ANOVA tables for all analyses are shown in Appendix C1.

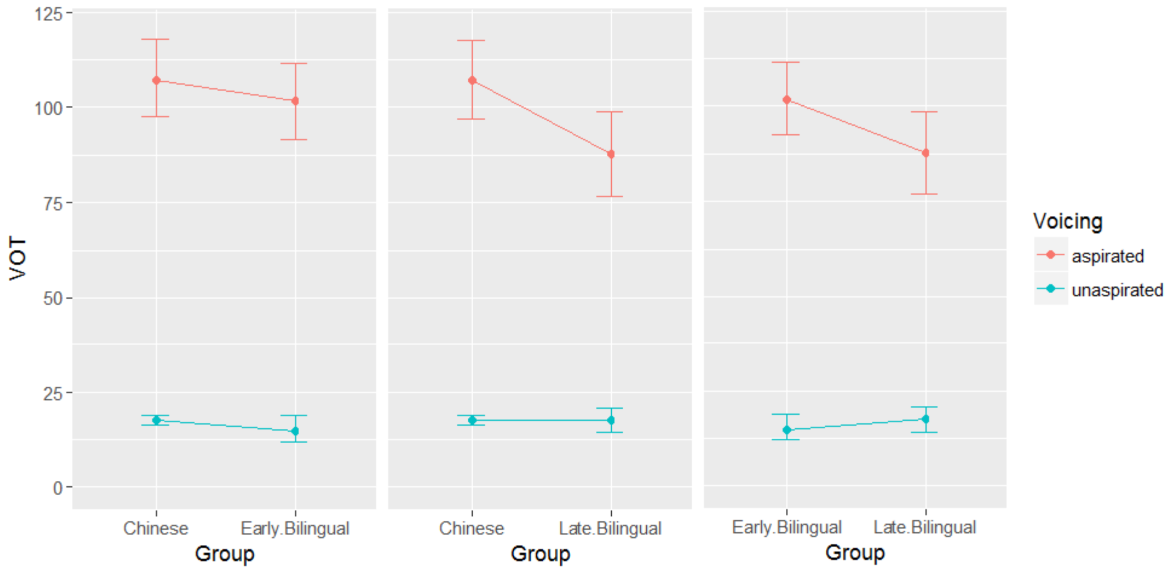

Figure 3.1. The mean VOT values of the unaspirated (blue) and aspirated (red) labial plosives in Mandarin Chinese production. The figure is comparing Mandarin monolinguals and early bilinguals (left), Mandarin monolinguals and late bilinguals (middle), early bilinguals and late bilinguals (right).[2]

These results show that early bilinguals, late bilinguals and Mandarin monolinguals produced Mandarin VOT without any significant differences except for the production of labial plosives. Considering the significant differences were not found in alveolar and velar plosive production, this suggests that AOA in Japan do not have significant effect on the Mandarin VOT values, and that both early and late bilinguals produced Mandarin VOT in a way similar to the native speakers of Mandarin Chinese in most cases. Although late bilinguals performed differently from Mandarin monolinguals in the production of labial plosives due to their shorter VOT of aspirated plosives, our bilinguals, therefore, maintained monolingual-like VOT values in their L1, in other words, there was no L2 (or possibly L3) transfer on L1, even though they all have been living in the L2-speaking country for a while. These results differ from those found in previous studies reporting that bilinguals' L1

[2] The figures in this section were drawn with the ggplot2 package (Wickham, 2009) in the R environment (R Core Team, 2015).

production patterns were different from those of monolinguals (Caramazza, Yeni-Komshan, Zurif & Carbone, 1973; Flege, 1987; Flege & Eefting, 1988; Major, 1992; Sancier & Fowler, 1997; Bullock, Toribio, González & Dalola, 2006). The reason may be that in this case, the Mandarin-Japanese bilinguals have created a new neural map for the [+voice] category in Japanese, which remains distinct from the [-voice] and [+spread glottis] neural maps in Mandarin, hence there is no reason for L2 transfer. If this is the case, we can also expect native-like performance on the Japanese contrast, especially since the Mandarin [-voice] neural map is very close, in terms of average VOT values, to the Japanese [-voice] neural map.

3.2 Experiment II: Production of VOT in Japanese

3.2.1 Participants

Two groups of subjects participated in Experiment II. The first group was composed of 10 Japanese speakers (NJ) who had no Mandarin experience: 5 were male and 5 were female. Ages ranged from 19 to 30 years (M = 21.5, SD = 4.3 years). All Japanese monolinguals were undergraduate or graduate students at the time of the experiment. They were from the Kanto region, and spoke the so-called standard Japanese dialect. None of them reported any known hearing or speech impairments. The summary of the background information for Japanese monolinguals is shown in Table 3.7.

Table 3.7. *Background information for Japanese monolingual participants in the current research.*

ID	Gender	Age	AOLEng	SelfEval Eng
NJ01	F	30	9	Adv
NJ02	M	30	13	InterH
NJ03	M	21	12	InterH
NJ04	F	19	5	InterH
NJ05	M	19	6	InterH
NJ06	F	19	12	InterL
NJ07	F	19	12	InterL
NJ08	F	19	12	InterL
NJ09	M	19	3	InterH
NJ10	M	20	5	InterH

Note: Age = age at the time of the experiment (in years); AOLEng = the age of first exposure to English; SelfEval = self-evaluation of the proficiency of English based on five levels (Beg = Beginner, InterL = Lower intermediate, InterH = Higher intermediate, Adv = Advanced, Native = Native/native-like).

Japanese monolinguals had learnt English as a foreign language at school, and had received the "standard" English language training in Japan: their English education consisted of 6 years of instruction in junior and senior high schools, and some language classes at the university, although the first English class for a number of Japanese monolingual participants started when they were in early elementary grades. They had never lived in an English-speaking country for more than a month.

The second participant group in Experiment II was early and late Mandarin-Japanese bilinguals: the same 16 bilingual speakers as in Experiment I. All participants received a monetary compensation for their participation in this experiment.

3.2.2 Stimuli

The Japanese word list contained 12 minimal pairs of CVV, CVC or CVCV structures, with initial voiced/voiceless plosives followed by the low vowel [a]. The two words in each pair had the same pitch accent: a falling pattern or a flat pattern. The Japanese word lists can be seen in Table 3.8.

Table 3.8. *Japanese target words used in the current research.*

Voiced IPA	Japanese	Meaning	Voiceless IPA	Japanese	Meaning
[bai]	バイ	“bye”	[pai]	パイ	“pie”
[basɯ̥]	バス	“bus”	[pasɯ̥]	パス	“pass”
[baɾi]	バリ	“Bali”	[ˈpaɾi]	パリ	“Paris”
[baɴ]	バン	“van”	[ˈpaɴ]	パン	“bread”
[dai]	台	“table”	[tai]	鯛	“sea bream”
[dakɯ]	抱く	“hold”	[takɯ]	焚く	“burn”
[daɕi]	出汁	“broth”	[taɕi]	足し	“complement”
[dani]	ダニ	“tick”	[tani]	谷	“valley”
[gai]	害	“damage”	[kai]	回	“inning”
[gake]	崖	“cliff”	[kake]	賭け	“gambling”
[gasɯ̥]	ガス	“gas”	[kasɯ̥]	カス	“scum”
[gamɯ]	ガム	“gum”	[kamɯ]	噛む	“bite”

Note: All target words beginning with /p/ are loanwords, because /p/ can only appear as a geminate, after nasal /ɴ/, in loanwords, and in onomatopoeia due to the sound change history in Old Japanese, in which /p/ changed into /ɸ/ and then /h/ or /w/.

3.2.3 Procedure

Before the experiment, Japanese monolinguals filled in the language background questionnaire which was introduced in the previous section. Early bilinguals and late bilinguals were exempt from answering the questionnaire for they had already answered it in the previous experiment. A short animated movie in Japanese was played before the recording for approximately 2 minutes, in an attempt to make the participants, especially the bilinguals, in a Japanese-speaking mode.

After the movie, the participants read aloud the 24 target words in the carrier sentence *mata* ___ *to itte kudasai* ("Please say ___ again.") at a natural speech rate. The recordings were made in the same sound-proof recording booth with the same devices introduced in 3.1.3. The order of sentences was again randomized for each participant, and they were asked to repeat any mispronounced or misread sentence. All sentences were repeated three times.

3.2.4 Results and discussion for Experiment II

For each participant, VOT values of 24 target words × 3 repetitions = 72 test tokens were acoustically analyzed by the same method explained in the previous experiment. Table 3.9 shows the mean VOT values for each plosive for Japanese monolinguals, early bilinguals and late bilinguals. Although the distribution of the mean VOT in voiceless plosives was similar for all groups, late bilinguals showed a different pattern from Japanese monolinguals and early bilinguals for voiced plosives: while Japanese monolinguals and early bilinguals produced negative VOT for all voiced plosives, late bilinguals exhibited prevoicing only for voiced labial plosives in Japanese. In addition, it is worth mentioning that prevoicing of voiced labial /b/ produced by early bilinguals was approximately 25 ms longer (so longer lead VOT) than the other voiced plosives. This result suggests that late bilinguals might not have fully created a new neural map for the [+voice] category in Japanese associated with lead VOT, while early bilinguals have.

Table 3.9. *Mean VOT values from data recorded in the Japanese production experiment.*

		Japanese monolinguals	Early bilinguals	Late bilinguals
voiced				
	labial	-12.04	-38.28	-1.05
	alveolar	-13.57	-13.69	8.34
	velar	-4.16	-14.88	2.36
voiceless				
	labial	43.78	48.67	56.49
	alveolar	45.03	53.83	49.29
	velar	57.47	62.04	61.33

Note: All measurements are in milliseconds (ms).

A mixed-effects ANOVA evaluating the effects of Group (early bilinguals, late bilinguals, Japanese monolinguals), Voicing (voiced, voiceless), the interaction of Group × Participants, and the interaction of Group × Voicing was conducted for the Japanese VOT of labial, alveolar and velar plosives. Here, I focus on the interaction of Group × Voicing, as in the analysis of Mandarin production experiment in Experiment I. The ANOVA tables for all analyses are shown in Appendix C2.

According to the labial production result of the mixed-effects ANOVA, the interactions of Group × Voicing were revealed to have a significant effect for the pair of early bilinguals and Japanese monolinguals ($F(1, 145) = 6.63$, $p = 0.012$), and that of early bilinguals and late bilinguals ($F(1, 110) = 4.72$, $p = 0.032$). Figure 3.2 illustrates this result of Japanese labial production. As for the alveolar production, the interaction of Group × Voicing had a marginally significant effect for the pair of early bilinguals and late bilinguals ($F(1, 110) = 3.39$, $p = 0.068$), while not for the other pairs ($F(1, 145) = 0.58$, $p = 0.45$ for the pair of early bilinguals and Japanese monolinguals, ($F(1, 103) = 2.44$, $p = 0.12$ for the pair of late bilinguals and Japanese monolinguals). Regarding velar plosives, no pair exhibited a significant effect of Group × Voicing ($F(1, 145) = 1.54$, $p = 0.22$ for the pair of early bilinguals and Japanese monolinguals, $F(1, 110) = 1.23$, $p = 0.27$ for early and late bilinguals, $F(1, 103) = 0.04$, $p = 0.83$ for late bilinguals and Japanese monolinguals).

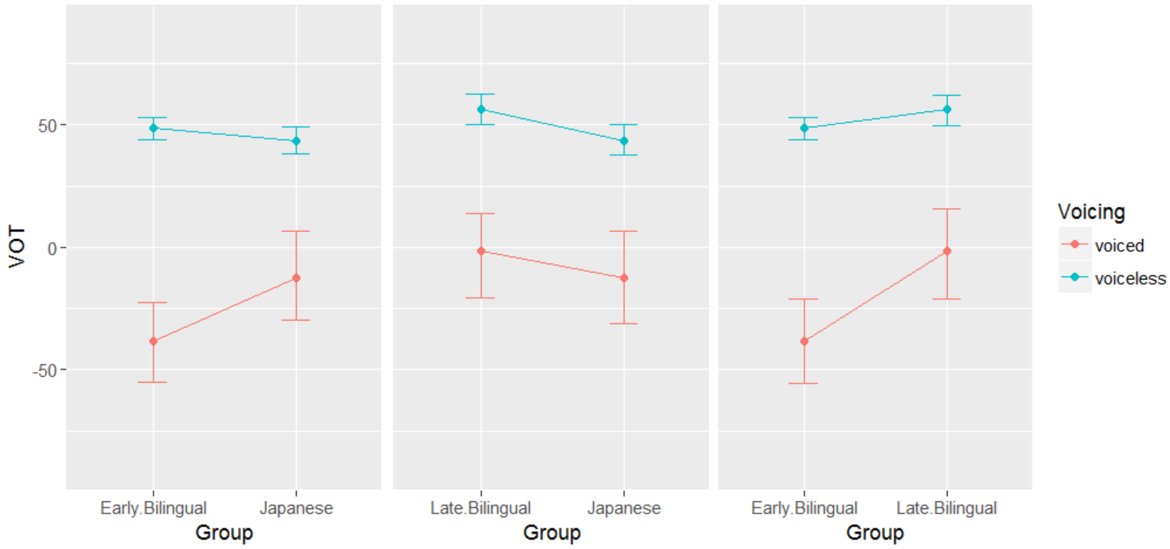

Figure 3.2. The mean VOT values of the voiceless (blue) and voiced (red) labial plosives in Japanese production. The figure is comparing Japanese monolinguals and early bilinguals (left), Japanese monolinguals and late bilinguals (middle), early bilinguals and late bilinguals (right).

This suggests that VOT patterns differed depending on AOA in Japan at least for labial plosives, and marginally for alveolar plosives. Also, although early bilinguals produced Japanese labial plosives differently from Japanese monolinguals, mainly due to the longer prevoicing of early bilinguals, there was no significant difference between monolinguals' and bilinguals' production of L2 Japanese in most cases. Hence, these results indicate that bilinguals were not heavily affected by L1 (and possibly L3) transfer, and the effect of AOA on the production of their L2 VOT was seen only in the length of prevoicing in voiced labial and alveolar plosives. In addition, in the production of labial plosives early bilinguals exaggerated the duration differences between the labial voiced and voiceless plosives compared to Japanese monolinguals. It can be predicted that early bilinguals' neural maps for Japanese [+voice] and [-voice] are created at more extreme positions than the native-like neural maps, but they are clearly separated.

3.3 Experiment III: Production of VOT in English

3.3.1 Participants

All the participants in Experiment I and Experiment II took part in this experiment, that is, 16 Mandarin-Japanese bilinguals including both early and late bilinguals, 10 native Chinese speakers, and 10 native Japanese speakers, who had never lived in an area where English was the native language more than a month. In addition to them, to serve as baseline, 8 native speakers of English who had never learnt either Mandarin or Japanese participated in the experiment. All English native speakers were male, and their ages ranged from 38 to 50 years (M = 42.25, SD = 4.09 years). They were graduate students or university lecturers from England, the United States, and Australia, all of whom spoke without a strong regional accent, and all of whom were living in Japan at the time of the experiment. Their background information is summarized in Table 3.10.

Table 3.10. *Background information for English monolingual participants in the current research.*

ID	Gender	Age	Birthplace
NE01	M	39	British
NE02	M	40	Australian
NE03	M	50	American
NE04	M	41	British
NE05	M	38	British
NE06	M	40	British
NE07	M	42	Australian
NE08	M	48	American

Note: Age = age at the time of the experiment (in years)

All participants reported normal hearing and no language impairments. A monetary compensation was offered to early and late bilinguals, Mandarin monolinguals and Japanese monolinguals, while English native speakers voluntarily participated in the experiment.

3.3.2 Stimuli

The English word list consists of 12 CVC minimal pairs using real English words, with initial voiceless/voiced plosives followed by low vowels such as [aɪ̯], [aʊ̯], [æ], [ʌ] or [ɑː] (see Table 3.11). Although it was inevitable to include words with relatively low frequency of usage, such as *dime* or *gab*, in order to make minimal pairs with these linguistic limitations, most target words were frequently appearing words in Standard English.

Table 3.11. *English target words used in the current research.*

Voiced		Voiceless (Aspirated)	
IPA	Spelling	IPA	Spelling
[b̥aɪ]	bye	[pʰaɪ]	pie
[b̥æt]	bat	[pʰæt]	pat
[b̥æk]	back	[pʰæk]	pack
[b̥æn]	ban	[pʰæn]	pan
[d̥aɪ]	die	[tʰaɪ]	tie
[d̥aʊn]	down	[tʰaʊn]	town
[d̥æn]	Dan	[tʰæn]	tan
[d̥aɪm]	dime	[tʰaɪm]	time
[g̊æp]	gap	[kʰæp]	cap
[g̊ɑːɚd]	guard	[kʰɑːɚd]	card
[g̊æb]	gab	[kʰæb]	cab
[g̊ʌm]	gum	[kʰʌm]	come

3.3.3 Procedure

As in Experiments I and II, the participants answered the language questionnaire before the recording session. Here, however, the participants of early and late bilinguals, Chinese native speakers and Japanese native speakers, who had already participated in the previous experiments, did not complete the questionnaire form because the questionnaire used in the current experiment was identical to those used in Experiment I and II.

For the recording session, the same procedure as in the previous experiments above was adopted. The carrier sentence for English was *I say ___ to my friend*. In case that the participants did not know the correct pronunciation of a target word, the experimenter presented a model pronunciation in General American accent. Again, the sentences were repeated three times for a total of 72 test tokens for analysis.

3.3.4 Results and discussion for Experiment III

Table 3.12 shows the mean English VOT values of our five groups. As this table shows, late bilinguals and Japanese monolinguals generally produced shorter VOT than the other three groups for voiceless plosives. With regard to voiced plosives, Japanese monolinguals was the only one group which produced prevoicing.

Table 3.12. *Mean VOT values from data recorded in the English production experiment.*

		English monolinguals	Mandarin monolinguals	Japanese monolinguals	Early bilinguals	Late bilinguals
voiced						
	labial	6.15	13.48	-0.57	12.13	1.27
	alveolar	13.48	14.41	-8.6	2.84	12.58
	velar	18.42	24.08	18.59	16.74	27.11
voiceless						
	labial	71.7	74.64	42.83	75.48	60.89
	alveolar	90.18	93.46	58.54	88.26	68.47
	velar	87.66	86.8	77.73	94.14	78.83

Note: All measurements are in milliseconds (ms).

A mixed-effects ANOVA evaluating the effects of Group (early bilinguals, late bilinguals, Mandarin monolinguals, Japanese monolinguals, English monolinguals), Voicing (voiced, voiceless), the interaction of Group × Participants, and the interaction of Group × Voicing was conducted for the English VOT for labial, alveolar and velar plosives. As in Experiments I and II, the interaction of Group × Voicing was mainly observed for the answers to the research questions of this study. The ANOVA tables for all analyses are shown in Appendix C3. The overall results are illustrated in Figure 3.3.

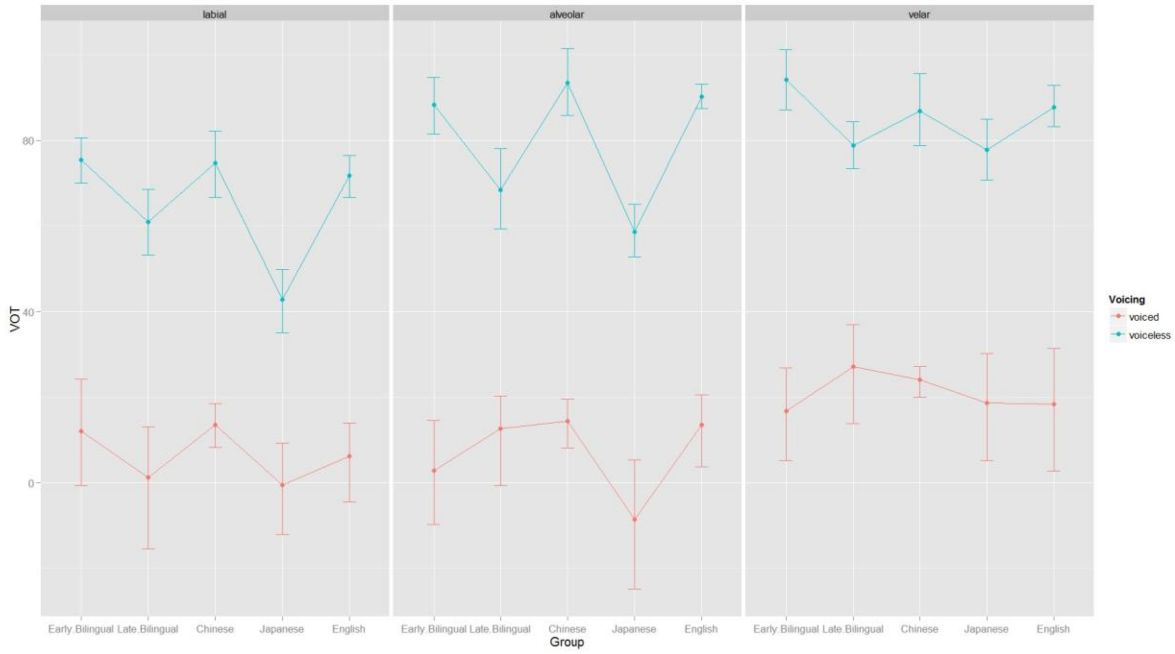

Figure 3.3. The mean VOT values of the voiceless (blue) and voiced (red) labial plosives in English production. The figure is comparing five groups for labial plosives (left), alveolar plosives (middle), velar plosives (right).

ANOVA conducted for labial plosives reports a significant effect of the interaction of Group × Voicing for the pair of early bilinguals and Japanese monolinguals ($F(1, 145) = 4.82$, $p = 0.03$), Mandarin and Japanese monolinguals ($F(1, 138) = 5.45$, $p = 0.021$), and Japanese and English monolinguals ($F(1, 124) = 7.06$, $p = 0.009$). In other group pairs, however, no significant differences were found (early bilinguals and Mandarin monolinguals, $p = 0.78$; early bilinguals and English monolinguals, $p = 0.8$; early and late bilinguals, $p = 0.74$; Mandarin monolinguals and English monolinguals, $p = 0.51$; Mandarin monolinguals and late bilinguals, $p = 0.86$; late bilinguals and English monolinguals, $p = 0.54$; late bilinguals and Japanese monolinguals, $p = 0.14$). From these results it is suggested that Japanese monolinguals produced English labial plosives differently from early bilinguals, Mandarin monolinguals, and English monolinguals. This may be caused by the fact that the

VOT values of voiceless labial plosives produced by Japanese monolinguals were much shorter than those by other participant groups, due to the influence from their L1 Japanese, as shown in Figure 3.3. Of particular interest, the effect of AOA was not significant indicating that both early and late bilinguals produced English labial plosives in a manner which was not significantly different from English monolinguals.

As for alveolar plosives, the pairs of the participant groups showing a significant effect are the following: early bilinguals and late bilinguals ($F(1, 110) = 6.43$, $p = 0.013$), late bilinguals and Mandarin monolinguals ($F(1, 103) = 7.93$, $p = 0.006$), and late bilinguals and English monolinguals ($F(1, 89) = 6.06$, $p = 0.016$). The pair of Japanese and English monolinguals also had a marginally significant effect ($F(1, 145) = 3.36$, $p = 0.069$). These results imply that the production of English alveolar plosives by late bilinguals was significantly different from that of early bilinguals, Mandarin monolinguals and English monolinguals, and AOA in Japan had an impact on the L3 VOT values. This may be because late bilinguals produced English voiceless alveolar plosives with considerably shorter VOT than early bilinguals, Mandarin and English monolinguals, which can be seen in Figure 3.3. Here, it is worth noting that a close inspection of Figure 3.3 indicates that the pattern of the mean VOT values of early bilinguals are more similar to Mandarin monolinguals, while that of late bilinguals are closer to Japanese monolinguals. In addition, the statistical analysis shows that the production of Japanese monolinguals was also different from that of English monolinguals. Again, as mentioned in the previous paragraph, this result may be due to the fact that the VOT values of both voiced and voiceless alveolar plosives by Japanese monolinguals are shorter than other groups as a result of the influence from Japanese.

Finally, ANOVA on velar plosives found that the three pairs of the participant group differed significantly each other: early bilinguals and Mandarin monolinguals ($F(1, 145) = 4.3$, $p = 0.04$), early bilinguals and Japanese monolinguals ($F(1, 145) = 4.81$, $p = 0.03$), and early bilinguals and late bilinguals ($F(1, 110) = 7.49$, $p = 0.007$). This result seems to be because early bilinguals produced longer VOT values for voiceless velar plosives and shorter VOT for voiced velar plosives, and the amount of VOT difference between voiced and voiceless plosives was greater than other groups. From this result it can be implied that AOA in Japan had an effect on the production of L3 velar plosives, and early bilinguals' production was influenced by neither L1 Mandarin monolinguals nor L2 Japanese monolinguals.

Thus, it can be generally said that bilinguals produced L3 English plosives in a way which is not significantly different from English monolinguals except for the production of velar plosives by late learners. It is also suggested that bilinguals who arrived in Japan earlier produced English sounds in a significantly different way from bilinguals who arrived later in their childhood at least for alveolar and velar plosives. In addition, there was a tendency in which L3 English production of early bilinguals were significantly distinct from their L2 Japanese monolinguals in labial and velar plosives, while that of late bilinguals were significantly different from their L1 Mandarin monolinguals in alveolar plosives. It implies that although early and late bilinguals successfully adjusted their categorical boundaries to produce English plosives, the range of VOT values associated to these categories are different according to their AOA in Japan: early bilinguals tried to avoid being influenced by their L2, which is their dominant language, while late bilinguals tried to avoid being affected by their L1 (Mandarin Chinese).

It is also worth noting that Japanese monolinguals tended to produce distinct VOT values from English monolinguals except for velar plosives. As pointed out earlier, their overall VOT values were shorter than any other groups due to the influence from their L1.

3.4 Summary of all results

In this section, the summary of the results of the production experiments in Mandarin, Japanese and English, which were discussed in the previous sections, is given. The overall result is summarized in Table 3.13.

Table 3.13. *The overall summary of the results of Experiment I, II and III.*

	Mandarin	Japanese	English
AOA effect			
- Labial	✓	✓	
- Alveolar		✓*	✓
- Velar			✓
Comparison with native speakers			
- Labial	Different (late bilinguals)	Different (early bilinguals)	
- Alveolar			Different (late bilinguals)
- Velar			
Prediction accordance	Yes (except labial of late bilinguals)	Yes (except labial of early bilinguals)	Yes (except alveolar of late bilinguals)

Note: Comparison with native speakers = whether the result of bilinguals and native speakers is same or different; Prediction accordance = whether the result of the experiment agrees with the prediction; * = statistically marginal effect.

As a whole, our bilinguals produced their L1 (Mandarin), L2 (Japanese) and L3 (English) without significant differences from the native speakers of each language. With respect to L3 production, it was found that bilinguals generally did not show a significant difference in terms of L3 English VOT from English monolinguals. This result basically agrees with the prediction of the present study in 2.3.

Observing the results more precisely, the results of early and late bilinguals varied according to the place of articulation. In fact, as shown in Table 3.13, effect of AOA was found in the production of Mandarin labial, Japanese labial and alveolar, and English alveolar and velar plosives. A close look at Figure 3.3 in the previous subsection suggests that AOA effect in English alveolar and velar plosives seems to be caused by the long VOT in voiceless alveolar and velar plosives of late bilinguals. This effect of place of articulation on non-native VOT acquisition is not something unusual: in fact, in a longitudinal study of English VOT by Japanese L2 learners, for example, their L2 VOT of alveolar and velar became longer than their L1 VOT, but L2 labial VOT was as short as L1 VOT (Nasukawa, 2010). Therefore, it might be implied that late bilinguals utilized their L1 Mandarin-attuned neural map for English voiceless alveolar and velar plosives, because VOT of alveolar and velar stops are universally longer than labial stops, as mentioned in 2.3.1.

The effect of L1 transfer and L2 transfer also differ depending on the place of articulation. As for English labial production, for example, early bilinguals exhibited a significant difference from L2 Japanese monolinguals, but not from L1 Mandarin monolinguals. On the other hand, late bilinguals' production was significantly different from neither Mandarin monolinguals nor Japanese monolinguals. This means that early bilinguals performed more similarly to L1 monolinguals than L2 monolinguals, which is not the case for late bilinguals. Similarly, in producing alveolar plosives, the result of late bilinguals was significantly different from that of L3 English monolinguals and L1 Mandarin monolinguals. This might suggest that the production pattern of early bilinguals was not distinct from neither Mandarin monolinguals nor Japanese monolinguals, while late bilinguals produced alveolar plosives in a closer way to Japanese monolinguals rather than Mandarin monolinguals. Finally, as for English velar plosives, early bilinguals' production was significantly different from that of L1 Mandarin monolingual and L2 Japanese monolinguals. This implies that early bilinguals produced velar plosives in a distinct way from both Mandarin monolinguals and Japanese monolinguals, while late bilinguals did not. To sum up, the summary of the main findings in Experiment I, II, and III is the following.

(1) In producing Mandarin Chinese VOT, Mandarin-Japanese bilinguals exhibited the similar values as native speakers of Mandarin regardless of their AOA except for labial plosives, meaning that their production of L1 was not largely influenced by their L2 Japanese or L3 English. That is, there were no observable L2 or L3 transfer on L1 production. There was a significant AOA effect in the production of labial plosives.

(2) In producing Japanese VOT, Mandarin-Japanese bilinguals exhibited the similar values as native speakers of Japanese except for labial plosives, meaning that their production of L2 was not significantly influenced by their L1 Mandarin or L3 English. There was a significant effect of AOA in the production of alveolar and velar plosives. Looking at the VOT values of each group, the performance by early bilinguals was distinct from Japanese monolinguals, showing greater difference between voiced and voiceless plosives. These results suggest that both early and late bilinguals have created a new neural map for the Japanese [+voice] category, but its associated VOT values were different according to their AOA. That is, early bilinguals' neural maps for Japanese [+voice] tend to include greater prevoicing, while the neural map associated with [-voice] has longer VOT values compared to late bilinguals.

(3) In producing English VOT, Mandarin-Japanese bilinguals basically did not show a significantly different pattern from native speakers of English, meaning that they produced L3 English without being significantly affected by L1 and L2. There was, however, some differences in their alveolar and velar production according to their AOA, in which early bilinguals tended to desist from Japanese-like production while late bilinguals from Mandarin-like production.

Putting these results all together, it can be concluded that bilinguals produced their L1 (Mandarin) and L2 (Japanese) in a similar manner to the native speakers of each language regardless of their AOA with a few exceptions. This is consistent with the results of previous studies introduced in the previous chapter claiming that L2 learners were able to establish separate phonetic categories for L2 (Bohn & Flege, 1993; Flege & Schmidt, 1995; Simon, 2010), but is not compatible with the previous studies reporting language transfer in the L1 and L2 VOT production of bilinguals (Flege, 1987; Flege & Eefting, 1987; Major, 1992; Antoniou et al., 2011). One explanation for the current results is that first, Mandarin and Japanese (as well as English) are not typologically related, which may reduce the effect of L1 and L2 transfer. And second, to process the Japanese category, the bilingual speakers had to create a novel neural map for the feature [+voice], which doesn't interfere with the Mandarin neural maps associated with the features [-voice] and [+spread glottis]. Also, the range of VOT values for the [-voice] plosives in Japanese, is very similar to the range for the [-voice] plosives in Mandarin. Hence, the bilinguals, assuming that they developed a new neural map for the Japanese [+voice] category, could produce the plosives in their L1 (Mandarin) and L2 (Japanese) with native-like VOTs. Considering that bilinguals performed differently from monolinguals in some situations, however, there may also be the case in which bilinguals associated a new L2 neural map with exaggerated VOT values in order to make these neural maps separate to their L1 neural map. For example, early bilinguals had a greater difference between voiced (unaspirated for Mandarin) and voiceless (aspirated for Mandarin) than late bilinguals. In this case, early bilinguals produce L1 plosives with native-like VOTs, but L2 (Japanese) with somewhat overemphasized VOTs.

As for L3 (English) production, the current series of production experiments indicates that bilinguals generally performed in a similar way to native speakers of their L3. This may be because their neural map for VOT including both Mandarin-attuned maps (short-lag and long-lag) and Japanese-attuned maps (lead and short-lag) spanned the possible L3 values (lead/short-lag, long-lag) in English, and therefore it was easy for them to be flexible enough to produce native-like L3. Note that the fact that the neural maps for short lag and long lag VOT in English are associated with different features—with the features [-voice] and [+spread glottis] in Mandarin, but with the features [+voice] and [-voice] in English—doesn't seem to cause difficulty for the bilinguals, at least not in production. Another important thing to note is the fact that even though the long-lag VOT values in Mandarin are generally longer than in English, the bilinguals could adjust their production to fit the expected English VOT values.

On the other hand, the L3 pattern of the bilinguals was significantly different from native speakers of their L1 or L2. More specifically, early bilinguals' L3 production was distinct from L2 monolinguals for two of the three place of articulation, while late bilinguals' L3 production was distinct from L1 monolinguals for one of the three place of articulation. This indicates that early bilinguals' L3 production was not influenced by L2, and late bilinguals' L3 production was not influenced by L1. However, there is no denying that early bilinguals were affected by L1, and late bilinguals were affected by L2. In fact, Table 3.12 and Figure 3.3 in the previous subsection suggest that according to the actual values of mean VOT, early bilinguals' values are basically quite similar to those of Mandarin monolinguals. Likewise, late bilinguals' L3 VOT values were similar to those of Japanese monolinguals at in the production of voiced labial plosives, voiceless alveolar plosives, and voiceless velar plosives they were similar to those of Mandarin monolinguals, for example, in the production of voiced alveolar or velar plosives. Therefore, although there was no statistically significant difference, it can be interpreted that the production pattern of bilinguals was similar to that of their L1 or L2, and its pattern was different according to their AOA. These results concur with previous studies on trilingualism reporting the combined effect of L1 and L2 on L3 production (Llisteri & Poch-Olivé, 1987; Ringbom, 1987; Gut, 2010; Wrembel, 2012a; Wrembel, 2012b) and from L2 to L1 (Hammarberg & Hammarberg, 2005; Gut, 2010; Llama et al., 2010; Wrembel, 2010).

In an attempt to observe the difference in the performance of early and late bilinguals in more detail, the production results of early and late bilinguals were investigated further. The result of the analysis on the production of early and late bilinguals is demonstrated in Figure 3.4.

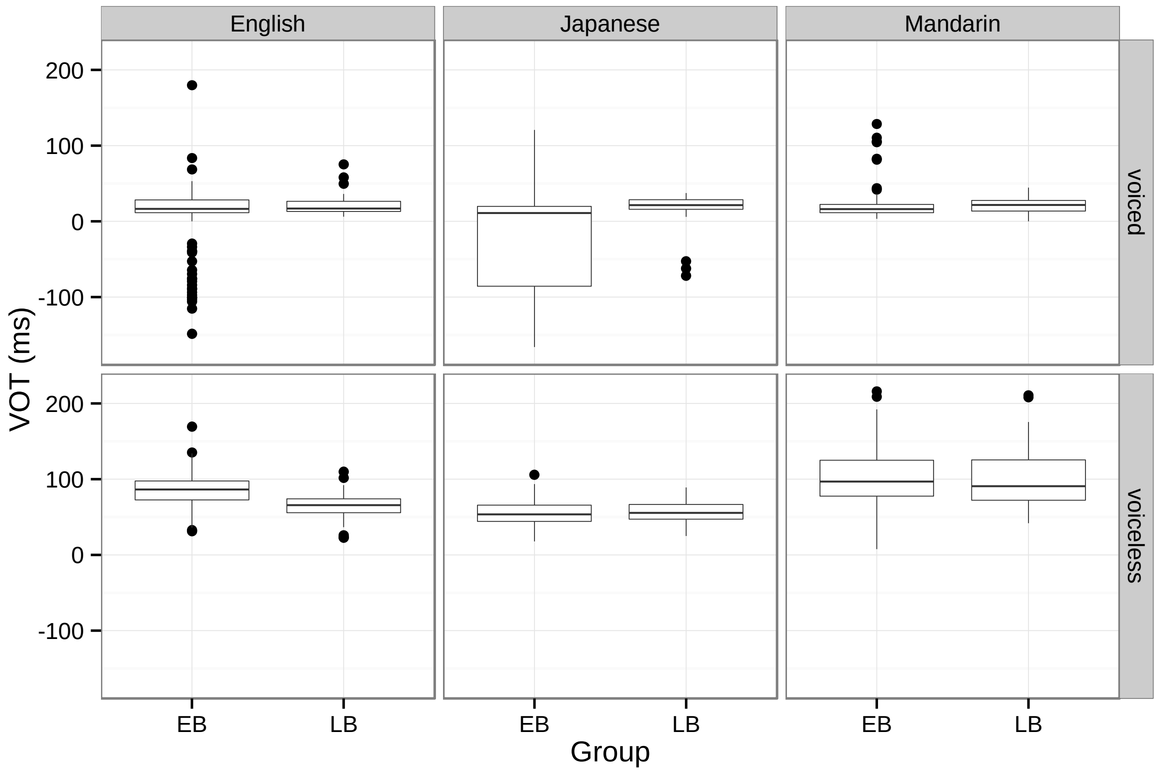

Figure 3.4. Boxplots for VOT values of the English, Japanese and Chinese produced by early bilinguals and late bilinguals.

Here, a lot of outliers are seen in the VOT values of English voiced plosives by early bilinguals. This may due to the fact that the VOT values produced by some of the early bilinguals (namely BL03, BL04, and BL16) were largely prevoiced especially for alveolar and velar plosives (M = -75.04 ms), while the rest of the early bilinguals rarely produced prevoiced plosives. The three bilinguals (BL03, BL04, BL16) who frequently produced prevoiced plosives did not have much in common except that their age of first exposure to English was later than 12 years old.

In the next chapter, I will introduce the English perception experiment, in which the same five groups of participants took part. The perception experiment was conducted right after the production experiments.

4. Experiment IV: Perception of English VOT

The current perception experiment aimed to examine the categorization of VOT values of a VOT continuum by Mandarin-Japanese bilinguals. The characteristics of L3 (English) VOT perception by bilingual speakers were explored by comparing their categorization pattern with that of native speakers of English. The performance of bilinguals was also compared with that of Mandarin and Japanese monolingual speakers in an attempt to investigate if there is a possible transfer from L1 or L2. As mentioned in 2.3.4, it is predicted that perceiving L3 stop contrasts is difficult for bilinguals, and they are expected to be affected by L2 Japanese rather than L1 Mandarin, because English and Japanese share the same underlying contrast (that is [+voice] and [-voice]). If the effect of the similarity of neural maps at the neural mapping level has a greater impact, on the other hand, bilinguals would be affected by L1 Mandarin rather than by L2 Japanese, for Mandarin neural maps are more similar to English (i.e. they cover short lag VOT and long lag VOT values). If the former is the case, the categorical boundary for bilinguals will be set at a shorter VOT value than English monolinguals, while if the latter is the case, the categorical boundary for bilinguals will be set at a longer VOT value than English monolinguals.

As discussed in 1.2, English contrasts voiced and voiceless plosives, and voiced plosives are generally produced with short-lag VOT (approximately 0 ms to +30 ms) or with prevoicing (negative VOT values), while voiceless plosives are pronounced with long-lag VOT (approximately +50 ms or more) (Lisker & Abramson, 1964; Klatt, 1975). In the current experiment, participants were required to categorize tokens of English initial bilabial plosives with various values of VOT into the “b” or “p” category.

4.1 Participants

Sixteen Mandarin-Japanese bilinguals, who were also participants in the production experiments in the previous chapter, participated in this perception experiment. As in the previous production experiments, the bilinguals were divided into two groups according to their age of arrival (AOA): early bilinguals and late bilinguals. We had, in addition, the same ten native speakers of Mandarin, Japanese and eight English speakers who participated in the production experiments. The perception experiment was conducted immediately after carrying out the English production recording reported in the preceding section.

4.2 Stimuli

The perception test required the participants to listen and distinguish a VOT continuum ranging from a VOT of -60 ms to +100 ms increasing in equal step of 20 ms, as schematized in Figure 4.1. One endpoint of the continuum (the leftmost circle in Figure 4.1) is a voiced plosive with negative VOT /b/, while the other endpoint (the rightmost circle in Figure 4.1) is a voiceless aspirated plosive with positive VOT /pʰ/.

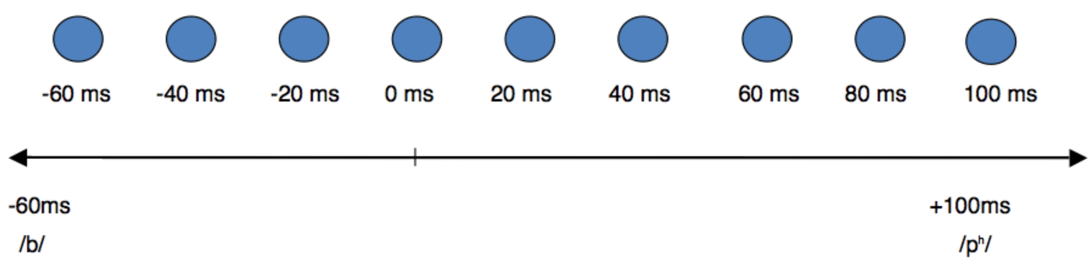

Figure 4.1. Tokens used for the perception test, which vary in terms of VOT of word-initial consonants.

In order to make the task more natural for the participants, they were asked to classify not nonce words but real words with high familiarity: *back* or *pack*. The manipulation was done with the PSOLA function in Praat version 5.4.16 (Boersma & Weenink, 2015). To make a 9 step VOT continuum, a token of *back* (VOT = 0 ms, vowel length = 210 ms vowel length, sampling rate = 11025 Hz) produced by a female native speaker of General American was used. In order to create the positive VOT tokens, a 100 ms sample of aspiration produced by the same speaker, which was taken from a token of *pack*, was inserted between the stop burst and the onset of voicing to a token of *back* whose prevoicing was removed to make a 0 ms token. The steps were created by removing 20 ms segments from the middle of the aspiration. In order to make the manipulated stimuli more natural, the release of the consonant was reproduced by adding a 10 ms friction burst at the onset of each stimulus.

The negative VOT tokens were created by taking a 20 ms sample of natural prevoicing from a token *back* produced by the same speaker. This prevoicing sample was copied and concatenated just before the point at which the voicing increase (above 0 dB) occurs, which created -20 ms, -40 ms, and -60 ms VOT tokens. The 10 ms friction burst was added at the onset of each stimulus, as in the positive VOT tokens, for the sake of naturalness.

4.3 Procedure

The participants underwent the perception test individually in a sound proof room at the University of Tokyo after they participated in the production experiments introduced in Chapter 3. They listened to 63 stimuli (9 stimuli × 7 times) in fully randomized order through Sennheiser HD25-SP headphones, allowing the volume control to be set to a comfortable level. In the experiment, the participants distinguished the presented stimuli in terms of *back* or *pack* displayed on the computer screen through a forced-choice identification task (see Figure 4.2 for the experiment screen) with inter-stimulus intervals of at least 3 seconds, to encourage processing of the tokens at a phonological level. In order to answer, they clicked on the button on which the word they heard was written. The next question appeared when they clicked the "Next" button.

Before the actual task, six extra token pairs which were different from the target stimuli (e.g. *small* vs. *never*, *fire* vs. *thin*) were presented as an exercise, where the participants learned how to handle the task. They were told to guess if they were uncertain about their judgment. The stimuli could be repeated only once. There were two break times during the test, and the participants could proceed with the test at their own pace. The number of already-judged tokens was shown in the upper-left corner of the experiment screen, which enabled the participants to know how many more stimuli they had to distinguish. All instructions given in the perception test were in English.

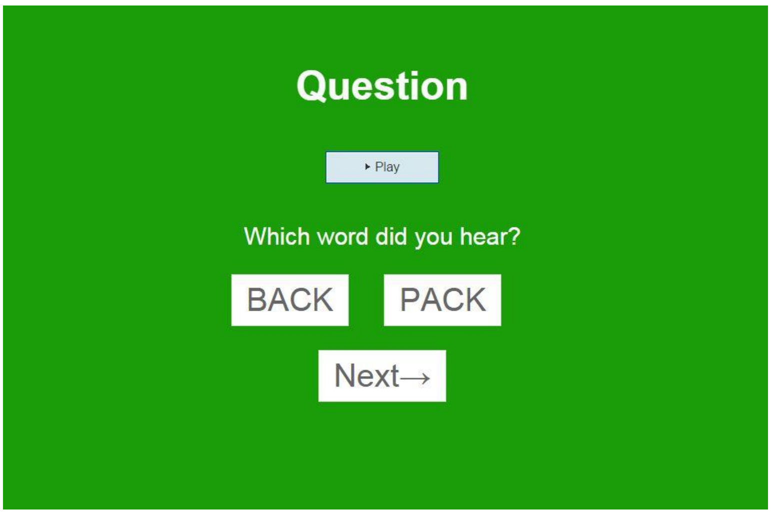

Figure 4.2. Screen shot of the forced-choice identification task in the perception test.

4.4 Results and discussion for Experiment IV

In analyzing the perception test results, a perception threshold value for each participant was determined by counting the number of times a participant recognized the stimuli as /b/. When a participant categorized a stimulus as *back* more than four times among the seven answers for each stimulus, he/she was considered to perceive the initial plosive of the stimulus as /b/. By this means, the judgment for each token made by each participant was determined. Then, the number of participants with /b/ response in each group was counted and the percentage of the participants with /b/ responses in each group was calculated for each stimulus. The results are illustrated with the line graphs in Figures 4.3 (native English listeners), Figure 4.4 (native Japanese, Mandarin and bilingual listeners) and Figure 4.5 (early bilinguals versus late bilinguals).

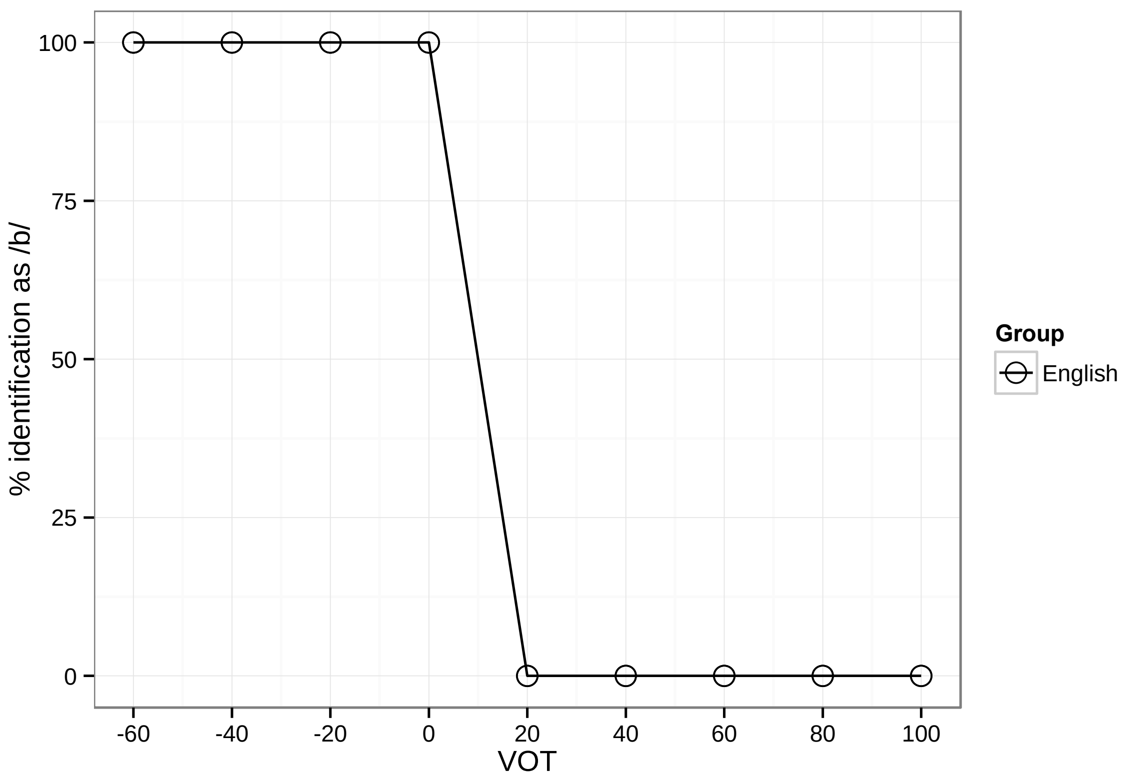

Figure 4.3. The mean identification percentage of /b/ for native speakers of English.

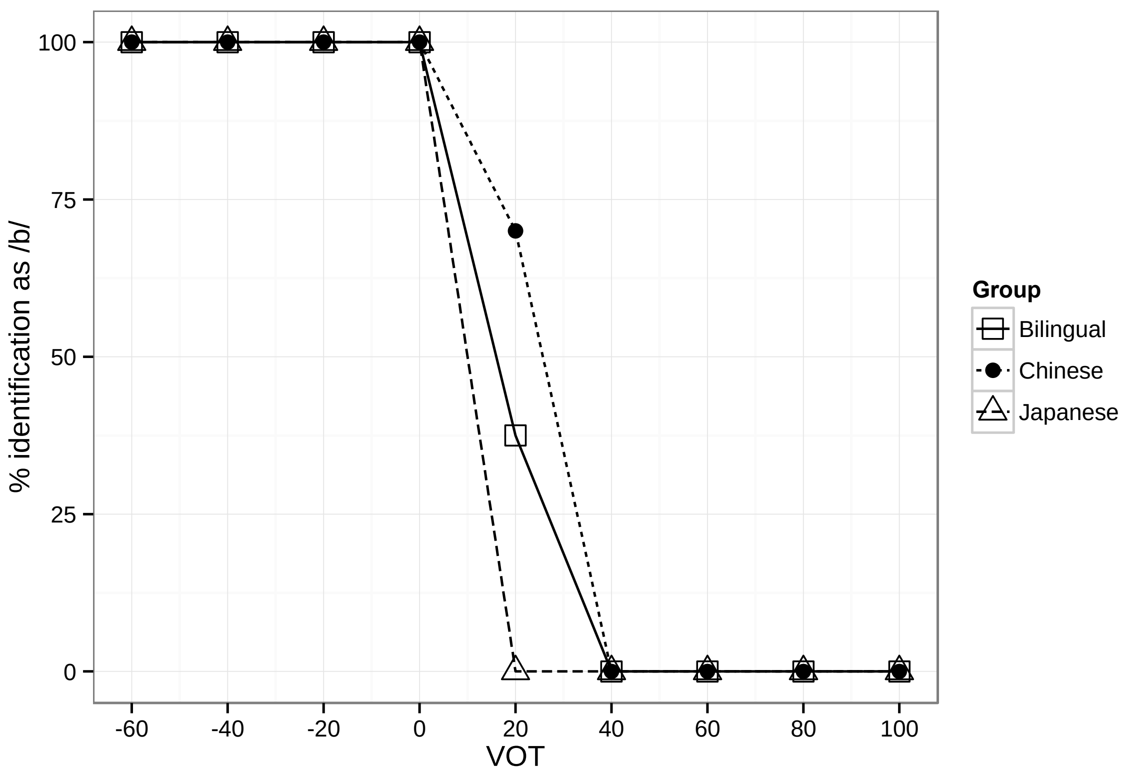

Figure 4.4. The mean identification percentage of /b/ for bilinguals, native Mandarin (Chinese) speakers, and native Japanese speakers.

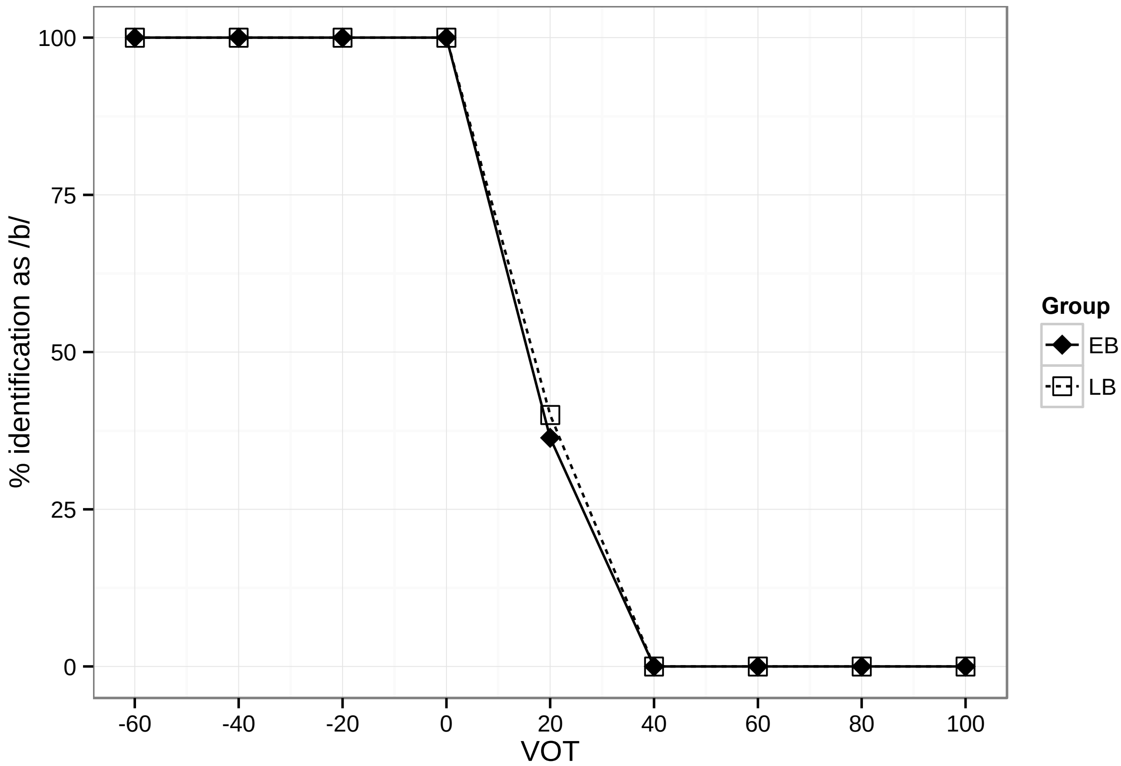

Figure 4.5. The mean identification percentage of /b/ for early bilinguals (EB) and late bilinguals (LB).

The perception pattern of English monolinguals is shown in Figure 4.3. They categorized stimuli with 0 ms VOT or less as /b/, and those with 20 ms VOT and above as /p/ without any exception. Thus, the categorical boundary for the native English listeners in this experiment is clearly between 0 and 20 ms of VOT.

As for the perception pattern of native speakers of Mandarin Chinese, their judgments varied for the 20 ms VOT stimuli (see Figure 4.4). The 20 ms VOT was discriminated as /b/ by 70% of Mandarin monolinguals.

Conversely, as shown in Figure 4.4, no Japanese monolinguals judged the 20 ms VOT as /b/, which means that Japanese monolinguals performed in the same manner as English monolinguals. The fact that Mandarin monolinguals rather than Japanese monolinguals showed a different pattern of English perception from English monolinguals is consistent with the predictions reported in section 2.4.3, where I expected that it is more difficult for Mandarin learners of English to categorize the English VOT contrast. This is because they can perceive English contrast at the neural mapping level, but at the phonological level these sounds are associated with different features in Mandarin Chinese.

The discrimination patterns of early and late bilinguals, which are shown in Figure 4.5, were nearly identical. Both early and late bilinguals perceived stimuli with 0 ms VOT or less as /b/, and those with 40 ms VOT and above as /p/. The variation among the participants was found only for the stimuli with 20 ms: 33.3% of early bilinguals and 40% of late bilinguals judged stimuli with 20 ms VOT as /b/. Thus, in terms of the mean identification percentage of /b/, the bilinguals showed an intermediate value to the values observed for Mandarin monolinguals and English monolinguals.

In order to evaluate whether there was a statistically significant difference in the patterns of the two bilingual groups, generalized linear mixed model was used, calculated with the lme4 package (Bates et al., 2015) in the R environment (R Core Team, 2015). In this model, VOT length of each token and participant group were fixed factors, and participants were specified as a random factor. Here, in an attempt to see whether there is a difference between the results of early and late bilinguals, bilinguals' test results were compared against the two baselines, that is, the result of Mandarin monolinguals and that of Japanese monolinguals. In the model in which the baseline was Mandarin monolinguals, early bilinguals showed a significant difference ($\beta = 2.54$, SE = 0.86, $p = 0.003$), while late bilinguals had a marginal significance ($\beta = 1.94$, SE = 1.004, $p = 0.053$). This result indicates that compared to Mandarin monolinguals the result of the perception experiment by early bilinguals were different, while that of late bilinguals were only marginally different from that of Mandarin monolinguals. This means that although the degree of the tendency is weakened for bilinguals with older AOA, both early and late bilinguals were not affected by Mandarin-attuned neural maps, which agrees with the prediction of the BLIP model mentioned in 2.3.4.

According to the model in which the baseline was Japanese monolinguals, on the other hand, both early and late bilinguals showed a significant difference (β = -3.79, SE = 1.28, p = 0.003 for early bilinguals, β = -4.4, SE = 1.403, p = 0.0017 for late bilinguals). This suggests that the perception test results of both early and late bilinguals were significantly different from the result of Japanese monolinguals, implying that in L3 perception bilinguals are not influenced by Japanese-attuned neural maps.

In order to see whether the results of early and late bilinguals are closer to either the result of Mandarin monolinguals or that of Japanese monolinguals, let's have a look at the overall results of the four groups illustrated in Figure 4.6. From this figure, the result of late bilinguals was found to be closer to that of Mandarin monolinguals. This differs from the prediction discussed in 2.3.3, where bilinguals' perception is predicted to be affected by L2 Japanese rather than L1 Mandarin. As for early bilinguals, on the other hand, it cannot be identified whether their perception result is closer to Mandarin or Japanese. However, as one can see that the line of early bilinguals is located in the middle of Japanese monolinguals and Mandarin monolinguals, it can be said that the result of early bilinguals is equally close to that of Japanese and Mandarin monolinguals. If this observation is correct, the result accords with the previous studies on trilingualism reporting combined transfer of L1 and L2 (Sypiańska, 2013; Wrembel, 2014).

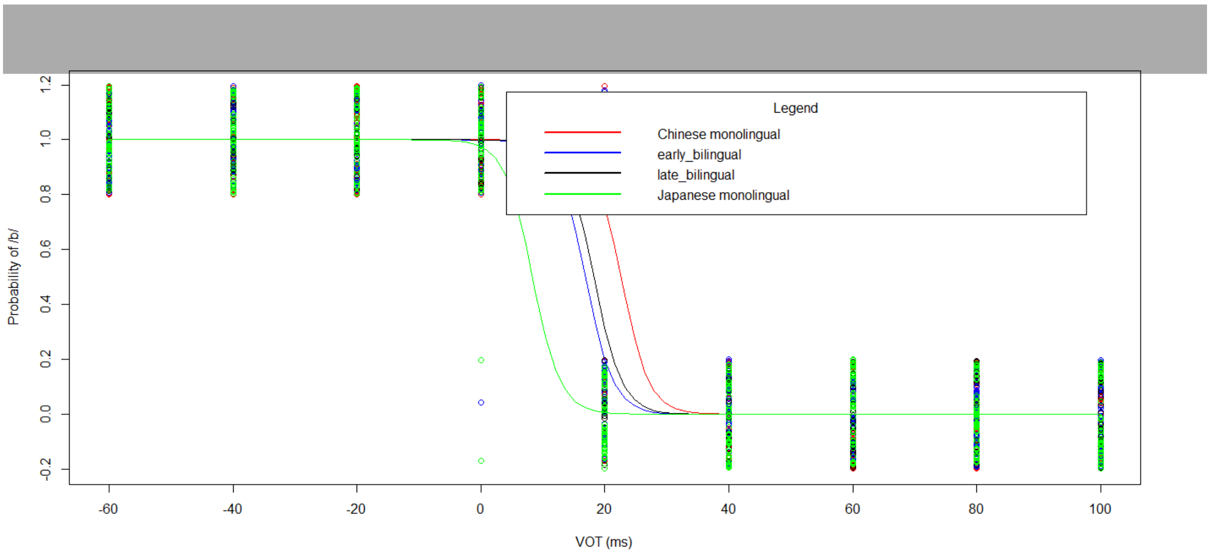

Figure 4.6. The overall results of the perception experiment of early bilinguals, late bilinguals, Mandarin (Chinese) monolinguals and Japanese monolinguals.

Using generalized linear mixed model, the effect of AOA itself on the results of perception experiment of bilinguals was also investigated. Here, VOT length of each token was a fixed factor, while AOA was taken as a covariate. The result of the analysis shows that there was no significant effect of AOA (β = -0.34, SE = 0.42, p = 0.42), VOT (β = 105.47, SE = 120.1, p = 0.38), nor the interaction of AOA and VOT (β = 110.76, SE = 146.61, p = 0.45), which means that when early and late bilinguals are treated as continuous variables, the effect of AOA on the result of the perception experiment is not observable. This may due to the criteria for the categorization of early and late bilinguals: while early bilinguals are defined as bilinguals who immigrated to Japan between the age of 0 and 6, while late bilinguals came to Japan between the age of 10 and 15. AOA of the most of early bilinguals, however, is 0, which means that they arrived in Japan before the age of 1, and therefore their AOA was not evenly dispersed compared to AOA of late bilinguals.

In summary, the result of the perception experiment indicates that the Mandarin-Japanese bilinguals with younger AOA were influenced by both L1 (Mandarin Chinese) and L2 (Japanese), while those with older AOA behaved in a similar manner to the native speakers of L1 (Mandarin Chinese) rather than to the native speakers of L2 (Japanese) or L3 (English) in perceiving English plosives, although the perception pattern of Japanese native speakers was more similar to English monolinguals' than Mandarin monolinguals'. This result supports the previous studies on L2 perception by bilingual speakers reporting that bilingual speakers do not show the same categorical boundaries as monolingual speakers of L1 and L2 (Williams, 1979; Flege & Eefting, 1987; Raphael et al., 1995).

As for AOA, in this study, AOA of bilinguals had an influence on their perception pattern, suggested by the fact that the discrimination result of late bilinguals was closer to that of Mandarin monolinguals, while that of early bilinguals did not show clear proximity to any of the monolinguals' results. This result agrees with previous study showing that the effect of L1 transfer to L3 would be hindered when L2 proficiency of a bilingual is high (Gut, 2010), although the degree of the AOA effect was not profound in the present experiment. It is also consistent with the prediction of the present study, which expected that early learners whose AOA was younger performed more native-likely compared to late learners whose AOA was older.

Last but not least, it has to be noted that because Japanese and English monolinguals performed completely the same in Experiment IV, it is not possible to distinguish L2 (Japanese) transfer and L3 (English) transfer. Thus, what I mentioned as L2 transfer may be actually L3 transfer, or in other words, it may be the case that early bilinguals established L3 native-like neural maps and utilized it for native-like distinction in the experiment.

In the next chapter the general discussion includes the discussion of the results, comparison with previous studies, implications of the theoretical models and ideas for future studies.

5. General discussion

This chapter discusses the results of the experiments, the answers to the hypotheses, and implications. In 5.1 I will discuss the production result, and then I will deal with the perception result in 5.2. Finally, 5.3 argues the ideas for future studies and the shortcoming which might affect the results of the present study.

5.1 Production

In the current study, I conducted three production experiments with Mandarin-Japanese bilinguals, one for L1 (Mandarin), one for L2 (Japanese), and one for L3 (English). The results of each experiment were compared with monolingual native speakers of each respective language. For the task, the Mandarin-Japanese bilinguals and the monolinguals read target words containing initial plosives contrasting in terms of VOT (such as [b, d, g] versus [p, t, k] in Japanese), and their VOT values were measured.

From the English experiment, it was found that bilinguals generally did not show a significant difference in terms of English (L3) VOT from English monolinguals. This means that Mandarin-Japanese bilinguals were able to produce L3 in a similar way to L3 monolinguals. This conclusion agrees with the prediction made by the inference of the BLIP model in 2.4.4. That is, bilinguals were able to use their sensitivity for VOT values of both L1 and L2 in order to produce native-like L3 VOT, because their neurons for VOTs of the two languages cover the VOT range of L3 English. Being sensitive to a wide range of VOTs may have allowed bilinguals to establish new categories for L3 VOT, which led to the accurate L3 production.

Although English VOT values by bilinguals were not significantly different from

English monolinguals' in most cases, there was one exceptional situation in which a significant difference was found: in the production of alveolar plosives, late bilinguals' VOT values were significantly different from those of English monolinguals. This seems caused by the shorter VOT in voiceless plosives: the values were shorter than the target values by approximately 20 ms, and they were actually closer to the values of Japanese monolinguals, who produced short VOT throughout the experiment due to the transfer from Japanese. From these results, it can be concluded that, in the production of alveolar plosives, late bilinguals did not succeed in utilizing their L1 and L2 neurons for VOTs and allocating them to the native-like alveolar VOTs in L3.

With regard to the effect of AOA (age of arrival) in Japan, the production experiments showed that there was a significant difference between the results of early and late bilinguals. To be more precise, bilinguals who arrived in Japan earlier produced English sounds in a significantly different way from bilinguals who arrived later in their childhood at least for alveolar and velar plosives. Thus, it is suggested that early and late bilinguals successfully created new neural maps for English plosives (or were able to adjust their neural maps to accurately produce L3 VOT), but the range of VOT values associated to these maps could be different according to their AOA in Japan: the VOT range of early bilinguals for English plosives would be longer and closer to Mandarin VOT range, while that of late bilinguals would be shorter and therefore closer to Japanese VOT range. This tendency is clearly shown in Table 6.1, where the bilinguals' production of bilabial stops in Mandarin Chinese (L1), Japanese (L2) and English (L3) are compared with each other, and also with the English monolinguals for English VOT values.

Table 6.1. *Mean VOT values of bilinguals in Mandarin, Japanese, English, and those of English monolinguals in English bilabial plosive.*

		Early bilinguals	Late bilinguals	English monolinguals
Mandarin	unaspirated /b/	14.77	17.54	
	aspirated /p/	101.74	87.77	
Japanese	voiced /b/	-38.28	-1.05	
	voiceless /p/	48.67	56.49	
English	voiced /b/	12.13	1.27	13.48
	voiceless /p/	75.48	60.89	74.64

Note: All measurements are in milliseconds (ms).

As for the production of L1 (Mandarin) and L2 (Japanese) VOTs by bilinguals, they generally did not show any significant difference from L1 or L2 monolinguals. The VOT values of bilinguals were not significantly different from those of Mandarin native speakers except for labial plosive production, where late bilinguals' result was significantly different from that of Mandarin monolinguals. This result appears to be due to the late bilinguals' shorter VOT values for voiceless labial plosives than that of Mandarin monolinguals.

In L2 production, VOT values by bilinguals were not significantly different from those of Japanese monolinguals, except for the production of labial plosives, which was caused by the fact that the voiceless labial plosives of early bilinguals were much longer than other plosives. Thus, it can be concluded that Mandarin-Japanese bilinguals have created a new neural map for the [-voice] and [+voice] categories in Japanese, which remains distinct from the [-voice] and [+spread glottis] maps in Mandarin.

As for the effect of AOA in the production of L1 and L2, generally it did not have a full effect on the VOT values of bilinguals, and both of early and late bilinguals produced L1 and L2 in a similar way to the native speakers. To be more precise, AOA in Japan did not have an effect on bilinguals' VOT values except for labial plosives in both Mandarin and Japanese, which was caused mainly by the longer prevoicing of early bilinguals and shorter prevoicing of late bilinguals. From this, it can be inferred that bilinguals' neural maps for Japanese [+voice] and [-voice] are created at different positions than the native ones.

Here, it should be noted that although the VOT values of bilinguals' L1 unaspirated plosives and those of L3 voiceless plosives are similar, they were predicted to assimilate these VOT values with different features from L1 Mandarin to L3 English: the VOT categories (i.e. neural maps) which are associated to [+spread glottis] in Mandarin at the underlying level are associated to [-voice] in English. Similarly, the L3 VOT values which bilinguals seemed to associate with [-voice] in Mandarin and Japanese were associated to [+voice] in English.

Thus, the answers to the research questions on VOT production would be the following.

1) Which previously learnt language, L1 or L2, has a stronger impact on L3 phonological acquisition? Or will both L1 and L2 have no impact on L3 phonological acquisition? In particular, which of Mandarin or Japanese affects the results of English production experiment by Mandarin-Japanese bilinguals?

⇨ Both L1 and L2 had no significant impact on L3 VOT production of Mandarin-Japanese bilinguals. There was an exception, however, in the labial plosive production by late bilinguals.

2) Will the bilinguals' L1 and L2 production be distinct from that of monolinguals?

Will the results of production experiments in Mandarin and Japanese by Mandarin-Japanese bilinguals be similar to those of Mandarin monolinguals and Japanese monolinguals?

⇨ The results of bilinguals on both the Mandarin and Japanese production experiments were not significantly different from that of Mandarin and Japanese monolinguals. There were some exceptions, however, in the production of Mandarin labial plosives by late bilinguals, and the production of Japanese labial and plosives by early bilinguals.

3) Will the bilinguals' age of arrival in an L2 country affect their L3 phonological acquisition? In particular, does the bilinguals' age of arrival in an L2 country have an influence on the results of production experiment in L3?

⇨ An effect of age of arrival in Japan was observable in the production of L3 alveolar and velar plosives.

These results suggest that, as a whole, bilinguals can somewhat adapt their categories or neural maps based on VOT values to produce a third language. This is possibly due to the fact that the Mandarin-Japanese bilinguals use a wide range of VOT values to process plosives in their two languages (Japanese and Mandarin): lead VOT versus short-lag VOT in Japanese, and short-lag VOT versus long-lag VOT in Mandarin. Hence, the VOT values used in Japanese and Mandarin span the short-lag and long-lag VOT contrast used in English. This wide range of neural mapping along the VOT continuum seems to have enabled them to adjust their L3 VOT to average native-like norm. These results concur with the predictions of the current study, which were based on the premises of the Bi-Level Input Processing (BLIP) model as discussed in 2.4.3 and 2.4.4: Mandarin-Japanese bilinguals were able to produce L3 VOT without an effect of language transfer, because bilinguals are sensitive to any VOT values within the ranges of the three languages.

Now, let's compare the predictions made by the BLIP with those by the other model, that is, the Speech Learning Model (SLM) proposed by Flege (1995) which was introduced in section 2.3.1. In the assumption of SLM, a non-native sound which is perceptually distinct from native sounds is more likely to be perceived and produced better through the establishment of a new phonetic category. On the other hand, if a learner does not perceive a phonetic difference between the L1 and L2 sounds, he/she would be unable to forge a new category and possibly assimilate the non-native sounds into native categories. Here, SLM posits that if the L2 sound is close to but not identical to the L1 sound, the L2 sound cannot be produced native-likely. In the context of the present study, the phonetic difference between the L3 stops and the L1/L2 stops are not great: although both English and Mandarin have long-lag VOT, Mandarin long-lag VOTs are typically longer than the English long-lag VOTs. Therefore, SLM may predict that bilinguals will use their long-lag VOT category in Mandarin to produce the English long-lag VOT, and hence, their average VOT values would be longer. Thus, the results of the present study, which found that bilinguals produced L3 sounds accurately, do not support the outcome of the prediction by SLM.

As for the results of L1 and L2 experiments, in which bilinguals generally did not significantly differ from L1 and L2 monolinguals, they also agree with the prediction inferred from the BLIP model[3]: the pattern of bilinguals' L1 and L2 production is predicted to be

[3] It should be noted that the BLIP model does not make any predictions about early bilinguals who began to learn non-native languages before puberty. Therefore, the prediction made here is based on my

comparable to that of monolinguals, because bilinguals are expected to have both groups of neurons for Mandarin and Japanese, and therefore they are sensitive to VOT values of both languages. SLM, like the BLIP model, does not make a prediction about early bilinguals before puberty and therefore I extend the prediction of SLM to the results of L1 and L2 results of early bilinguals. From SLM, it can be inferred that bilinguals would assimilate L1 Mandarin unaspirated plosives into L2 Japanese voiceless plosives due to their phonetic similarities, which leads bilinguals to use Mandarin unaspirated category for the production of Japanese voiceless plosive, and vice versa. Hence, bilinguals are expected to produce L1 unaspirated plosives and L2 voiceless plosives in a different way from monolinguals. Compared with the results of L1 and L2 production in the present study, it can be concluded that the present study does not support the prediction inferred from SLM.

Finally, I would like to compare the result that bilinguals were able to produce L3 stops accurately with previous studies of L3 phonetic acquisition reported in section 2.2. In these studies, L3 performance was influenced by L1 (Llisteri & Poch-Olivé, 1987; Ringbom, 1987; Wrembel, 2012), L2 (Hammarberg & Hammarberg, 2005; Llama et al., 2010; Wrembel, 2010), or the combined effect of L1 and L2 (Sypiańska, 2013; Wrembel, 2014). There are four major differences between these previous studies and the present study. First, the language repertoire is limited in the previous studies. For example, the bilinguals in these studies spoke German (L1), English (L2) and Spanish (L3) (Wunder, 2010), or Polish (L1), Danish (L2) and English (L3) (Sypiańska, 2013). The fact that three languages, or at least the two of the three languages are genetically and also perceptually similar in terms of VOT would have an influence on the results of these studies, because as mentioned in 2.2.2, when

extension of the model, not the original claim.

a feature in a language is perceived more similarly than one in the other language, the feature in the former languages is more likely to be transferred (Kellerman, 1987). Second, the linguistic target of observation is different among the studies. In this study, the investigation was made based on the VOT values, which has been considered as a clear and precise acoustic phenomenon. On the other hand, Ringbom (1987) investigated intonation, for example, while Llisteri & Poch-Olivé (1987) studied vowels and fricatives. It is likely that different linguistic items can exhibit different results. Thirdly, the distribution of the linguistic features in the three languages is especially complicated in the present study, because Mandarin, Japanese, and English assimilate similar VOT values to different categorical maps, either [+voice], [-voice] or [+spread glottis]. These complex associations of acoustic cue and neural mapping in the present study may have a special impact on the results. As the last reason why the VOT values of bilinguals did not exhibit L1/L2 transfer, one may want to consider the quality and the quantity of the L3 input bilinguals had received in their English education. As claimed in Flege & MacKay (2011), in L2 phonetic acquisition the amount and quality of L2 input are one of the most significant predictors of L2 proficiency. The amount and quality of L3 English input that the bilinguals in this experiment had received were not measured specifically, but all of them belonged to the standard junior high schools and high schools in Japan, where English is generally taught by Japanese-speaking teachers and occasionally by assistant language teachers who are native speakers of English. Thus, it seems to be unlikely that bilinguals had received a considerable amount of English input from native speakers of English, and therefore the quality and quantity of L3 input may not have had a great impact on the results of the production experiments.

As noted above, however, though the VOT values of bilinguals were similar to those of L3 monolinguals, it is possible that the VOT categories are not native-like, in the sense that it is not clear with which underlying categories they are associated. For instance, is the long-lag VOT category (neural map) used to process the English initial aspirated voiceless plosive associated with a [-voice] feature or a [+spread glottis] feature at the underlying level of representation? If it is still associated with the Mandarin [+spread glottis] underlying feature, this could be considered an effect of L1 transfer, but occurring at the phonological level, not at the neural mapping (phonetic/acoustic) level.

5.2 Perception

In the perception experiment, Mandarin-Japanese bilinguals and Japanese, Mandarin and English monolinguals listened to a VOT continuum ranging from prevoiced [b] to aspirated [p^h] and had to classify the tokens heard as the English words 'back' or 'pack'. The results showed that bilinguals did not perceive the English stop contrast in a way similar to English monolinguals, regardless of their age of arrival in Japan. Especially, the late bilinguals showed a clear tendency that they performed in a way similar to Mandarin monolinguals, indicating a possible influence from L1.

Here it has to be keep in mind that Japanese monolinguals performed in exactly the same way as English monolinguals and therefore there is no distinguishing L2 transfer and L3 transfer: that is, it is unclear whether early bilinguals relied on the L2 neural map or they tried to perform L3 native-likely. The reasons of the native-like performance by Japanese monolinguals can be the following: the mean VOT categorical boundary of Japanese monolinguals was the same as that of English monolinguals since they could adjust their neural maps easily due to the similarity in number of neural maps and features. Although

English and Japanese do not share the distribution pattern of VOT values for stop contrast, their perception boundary VOT values between voiced and voiceless plosives would be similar. In fact, according to Shimizu (1996), boundary values for English /b-p/ is from 20 ms to 30 ms, while those of Japanese is from 15 ms to 20 ms, in proximity to English values. Because the perception materials for the present experiment was made in steps of 20 ms, no difference between the perception of English and Japanese monolinguals. As for Mandarin monolinguals, their categorical pattern was not similar to that of the English monolinguals, because they could not adjust their neural map perfectly in English perception. In Mandarin underlying features associated with VOT maps are [-voice] and [+spread glottis], which is different from in English [-/+voice], even though the number of the neural maps was the same.

Thus, these results suggest that when the age of arrival in Japan are categorized into two groups, early vs. late, the difference between the groups was found. However, on the other hand, when the age of arrival in Japan was treated not categorically but continuously, the effect of AOA was not found to be influential to the results of perception test. It may be caused by how bilinguals were categorized: a bilingual was categorized as early bilingual if he/she immigrated to Japan between the age of 0 and 6, while those who came to Japan at the age of between 10 and 15 were categorized as late bilinguals. Looking at the distribution of the age of arrival, most of early bilinguals came to Japan at the age of 0, which might make the data inappropriately ununiformed.

Thus, the answers for the research questions on VOT perception would be the following.

1) Which previously learnt language, L1 or L2, has a stronger impact on L3

phonological acquisition? Will the results of English perception experiment by bilinguals be closer to the results by Mandarin and Japanese monolinguals?

⇨ It depended on the age of arrival in L2 country: those with younger AOA relied slightly more on L2 Japanese, which has the same number and feature of neural maps, while those with older AOA relied more on L1 Mandarin, which has the same number of neural maps but with different features.

3) Will the bilinguals' age of arrival in an L2 country affect their L3 phonological acquisition? In particular, does the bilinguals' age of arrival in an L2 country have an influence on the results of perception experiment in L3?

⇨ As mentioned in 1), late bilinguals had a clear tendency of the reliance on Mandarin-attuned map, while early bilinguals had a marginal tendency of the reliance on Japanese-attuned map. When the age of arrival in Japan was treated as a continuous variable, the effect of age of arrival was not observed.

Now, let's compare the result of the perception experiment with the prediction of the BLIP model. As explained in 2.4.3, a listener is able to perceive the contrast both at the neural and phonological levels when an L2 acoustic cue is processed by the same number of neural maps as in the L1, and when the neural maps are linked to the same features in the L1 and L2. This prediction applies to Japanese monolinguals, because both English and Japanese have only one contrast along the VOT dimension, and the VOT contrast is associated with the features [+voice] and [-voice] at the phonological level. In this case, they are predicted to have the least difficulty in English perception, which agrees with the actual result of the experiment where Japanese monolinguals performed completely in the same way as English monolinguals. Thus, the result of Japanese monolinguals supports the prediction of the BLIP model.

When the neural maps are linked to different features in the L1 and L2, the BLIP model predicts that the L2 perception is more difficult because the L2 contrast is perceived only at the neural mapping level. Mandarin monolinguals are categorized into this case: although they have neural maps corresponding to short-lag and long-lag VOT as in English, these neurons are associated with the features which are different from their L1 at phonological level, which makes it difficult for Mandarin monolinguals to perceive English VOT. Because Mandarin monolingual exhibited the perception pattern which is the most distinct from native-like one, the result of the present study is compatible with the BLIP model.

Finally, according to the BLIP model it can be inferred that it is difficult for bilinguals to perceive English stops, because they have three neural maps associated to features [+voice], [-voice], and [+spread glottis] along the VOT dimension, while there are only two features [+voice] and [-voice] in English. Considering the number of the neural maps and the underlying features, the BLIP model expects that bilinguals rely on Japanese-attuned maps, because English and Japanese share the same number and features. Here, the present study shows that early bilinguals appeared to rely little more on Japanese, while late bilinguals on Mandarin. Late bilinguals may have chosen to rely on L1-attuned map because the VOT values are closer to the English VOT values. But why did this difference arise?

The BLIP model says that all bilinguals should have the same number of categories (i.e. neural maps) regardless of AOA, though their categorical boundary may not be exactly at the same place, because the sufficient input is required for their categorical boundary to be accurately adjusted. Here, especially our late bilinguals presumably may have not had access to the sufficient input due to their shorter length of residence in Japan than early bilinguals. Thus, the BLIP model predicts that because late bilinguals have less L2 and L3 inputs, their position of categorical boundary was vulnerable and they relied on their L1 neural maps. As for early bilinguals, they performed slightly similar to English and Japanese monolinguals, but they were also influenced by L1-attuned neural map, presumably due to the insufficient L3 input.

In addition, considering the result that L3 perception pattern of early learners was located between L1 Mandarin pattern and L2 Japanese pattern, it can also be presumed that early bilinguals used either L1 or L2 map according to the VOT values they heard: that is, for short-lag and long-lag VOT they relied on L1-attuned map, while for lead VOT they relied on L2-attuned map. If this assumption is correct, the neural maps of bilingual were supposed to be much more flexible compared to those of monolinguals. In fact, the categorical maps of bilinguals appear to be flexible: because bilinguals have both L1-attuned and L2-attuned categorical maps, the range of VOT they cover is wider than monolinguals.

Finally, I will compare the predictions of the Perceptual Assimilation Model (PAM) (Best, 1995) with that of the BLIP model, and see which one may better account for our results. PAM assumes that an L2 sound tends to be perceived according to their similarities to an L1 sound, and these similarities are determined by articulatory gestures defined by the articulatory organs, place of articulation, and manner of articulation. Also, the difficulty in acquiring L2 perception pattern depends on how the L2 sounds are categorized and are assimilated to L1 sounds. According to PAM, discrimination would be the most accurate when a native and non-native sounds are categorized into two different categories, i.e. TC (Two Category) assimilation. In the present study, where L3 English stop sounds are produced with short-lag vs. long-lag VOT, English monolinguals are expected to be applied to this type of assimilation, for they should assimilate tokens with long-lag VOT to /p/ and those with lead/short-lag VOT to /b/. The second most accurate discrimination is made when the two sounds are categorized into two categories in terms of the goodness as a sound in L1, i.e. CG assimilation. Japanese monolinguals are predicted to assimilate lead/short-lag VOT to /b/, while they should assimilate long-lag VOT as a bad example of /p/. Mandarin monolinguals are also applied to this type, for they assimilate long-lag VOT to /p/ and short-lag VOT to /b/, while assimilate tokens with lead VOT as a bad example of /b/. Finally, the least accurate discrimination, i.e. SC assimilations, follows when the two sounds are categorized into a single category. No participants are expected to show SC assimilations, because all the three languages have a stop contrast. In the case of Mandarin-Japanese bilinguals, they are supposed to have different phonetic categories of both L1 Mandarin and L2 Japanese, that is, lead/short-lag/long-lag VOT. Therefore, PAM predicts that for Mandarin-Japanese bilinguals it would not be difficult to perceive English stop contrasts, which also can be inferred that they are able to discriminate English stops accurately without transfer from L1 and L2. However, the result of the present perception experiment shows that the discriminating pattern of bilinguals was not native-like and was closer to Japanese slightly for early bilinguals and to Mandarin for late bilinguals. The results of the experiment are not fully consistent with the BLIP model nor PAM. However, the BLIP model correctly predicted for the perception of monolinguals. The comparison between the result of the present study and the predictions made by the BLIP model and PAM shows an important fact that even though a learner has a native-like categorical boundary, it does not always mean that he/she is able to assimilate a sound to a category in a native-like way. Thus, for a native-like performance, both of accurate activation of the appropriate neural maps at neural mapping level and accurate association to a linguistic feature at phonological level is required.

5.3 Summary and implication

Putting the results of the production and the perception experiments all together, bilinguals in the current study generally produced L3 English VOT in a manner which is not significantly distinct from English monolinguals, except for an exception in the production of alveolar plosives. On the other hand, bilinguals perceived L3 VOT in the different way as L3 monolinguals. In perceiving L3 VOT, bilinguals exhibited a pattern intermediate between Mandarin (L1) monolinguals and Japanese (L2)/English (L3) monolinguals.

Compared with L1 Mandarin and L2 Japanese monolinguals, L3 production and L3 perception of bilinguals, it is suggested that being bilinguals does not always have a good effect on the acquisition of additional languages. For example, while bilinguals generally performed more native-likely compared to Japanese monolinguals, they tended to produce a larger or even exaggerated difference of VOT between voiced vs. voiceless plosives in English. Similarly, bilinguals who had more neural maps (categories) which cover a wider range of VOT than monolinguals performed less native-likely than Japanese monolinguals in a discrimination task of voiced vs. voiceless labial plosives in English. In fact, they failed to associate VOT values with neural maps appropriately, for there arose a conflict at neural mapping level due to the too many neural maps. Thus, the effect of having more categories than monolinguals sometimes hinders bilinguals from the accurate performance in L3. The results of the present experiments indicate that bilinguals failed to perceive L3 accurately while they succeeded in producing L3 accurately, supporting the claim that the acquisition of production precedes that of perception. Although more previous studies on both L1 and L2 acquisition, including the L2 phonetic acquisition models such as PAM and SLM, found that perception comes before production, the tendency that the production of a sound contrast is more accurate than the perception of the contrast has been observed in several previous studies on L2 (Goto, 1971; Sheldon & Strange, 1982, Flege & Eefting, 1987; Yamada et al., 1994, Levy & Law, 2010). For example, Flege & Eefting (1987) reported that L1 Dutch speakers succeeded in producing L2 long-lag VOT with substantial differences from L1 short-lag VOT, while they failed to show the differences between L1 and L2 VOT in perception.

Eckman (2004) attributes these results to the testing method they used in their production experiments. During the production experiments of these previous studies, the participants received written input: they could read the target minimal pairs before producing the target contrasts, providing the participants with input that is not auditory. Therefore, he claimed that the production results exceed their perceptual results in some previous studies because of the visible cues obtained before the production. This may facilitate the accurate production, according to Eckman (2004), and makes it possible for production to surpass perception. On the other hand, perception ability exceeds production ability in L1 acquisition because infants usually receive auditory input only. In the production experiment of the present study, participants could read the written target minimal pairs before the recording.

Moreover, they could check whether there were unfamiliar words before producing the contrasts, as Eckman (2004) pointed out, while in the perception experiment no written input was given before the experiment. To avoid the unfairness between production and perception experiments, it would be required to check the production of spontaneous or near-spontaneous speech of participants, although it will be harder to obtain the target contrasts.

The different results in the perception and production experiments can also be attributed to the difference in task type used in perception and production. In the present study, the perception task was artificial with manipulated VOT values, focusing on category boundary, whereas the production task, looking at the average of VOT values, was more natural as the target words produced were real words, which could be produced within a range of VOT values. In fact, some authors suggest that "(p)erhaps the VOT values measured in speech production and the category boundaries obtained in forced-choice identification should not be compared directly" (Flege & Eefting, 1987, p.199). In the present paper, production and perception experiments saw completely different results: while the boundary values of VOT in L3 stop contrasts were observed and calculated in the perception experiment, the mean values of L3 VOT were measured and analyzed in the production experiments. At any rate it can be suggested that the knowledge of Mandarin and Japanese worked better on L3 English production than on L3 perception for the bilinguals in the present study.

In this study, the perception and production of L3 by bilinguals were observed and interpreted through the BLIP model, which is originally a model on the perception process of a newly acquired language. The results of the present study indicate that even if a bilingual produce a sound of an additional language which is perceptively similar to their L1 or L2, it is not strongly affected by both L1 and L2, but rather they can produce the new sound without influences from L1 and L2. This is due to the flexibility of the neural maps which they have: because the bilingual is expected to have neural maps of both L1 and L2, covering wider sound features than monolinguals, their neural maps appear to be more flexible to go back and forth along the range of acoustic inputs to get assimilated to any sound features within the range. As for perception, on the other hand, a bilingual perceives a non-native sound contrast in a different way from a native speaker. Their perception patter would be affected by their L1 or L2, depending on how similarly the sound contrast is realized in their L1 and L2. Again, it should be emphasized that having more neural maps associated to sound features in L1 and L2 sometimes makes it more difficult for the bilingual to accurately associate the neural map and the feature, because he/she has a more options that monolinguals.

These results can be extended to the L3 phonetic acquisition of a bilingual in other settings. For example, how a Korean-Mandarin bilingual makes a voicing contrast of stop consonants in English? Korean has a three-way stop contrast: lenis (produced with 0 or short-lag VOT, appeared both as a voiced plosive or a voiceless plosive), tense (voiceless aspirated), and aspirated (voiceless aspirated produced with constricted glottis). First, as for the production of L3 English, the Korean-Mandarin bilingual has five different neural maps for VOTs, three from L1 and two from L2. Here, two neural maps (i.e. Korean aspirated and Mandarin aspirated) cover the long-lag VOT, while the rest (Korean lax, Korean tense, and Mandarin unaspirated) cover the short-lag VOT. In production of L3, he/she would not have any problems: when the bilingual produces an L3 English voiceless stop, it would exhibit a long-lag VOT utilizing the neural maps for Korean aspirated or Mandarin aspirated plosive. Due to the longer VOT values of both Korean aspirated and Mandarin aspirated than those of English (Shimizu, 1996), L3 VOT of a voiceless plosive is expected to be longer than that of English monolinguals. When he/she produces an L3 English voiced stop, on the other hand, it would exhibit a short-lag VOT utilizing the neural maps for Korean tense/lax plosives or Mandarin unaspirated plosives. To sum up, L3 English production of stop contrast by the Korean-Mandarin bilingual is predicted to be native-like, with the aid of five neural maps in his/her L1 an L2. As for the L3 perception of stop contrast, the bilingual would have difficulty in associating L3 VOT values to the accurate neural map. Considering the fact that bilinguals in the present study failed to establish a new neural map for L3 English, it would be difficult for the Korean-Mandarin bilingual to perceive English stop contrast in a native-like way.

6. Conclusion

The focus of the current study was to investigate whether the knowledge of multiple languages facilitate both production and perception of the newly acquired language, and whether their L1 or L2 affect L3 performance. The current research conducted experiments evaluating L3 production and perception of bilinguals to see if having two languages facilitates or impedes their L3 stop contrast. The bilinguals in the current study were Mandarin-Japanese who learned English as an L3 with varying age of arrival (AOA) in Japan. The results of the experiments are compared against the premises and predictions of the BLIP model (Grenon, 2010).

In the production experiments of English, the VOT values in the 24 target words produced by Mandarin-Japanese bilinguals were measured and compared with those of English, Mandarin and Japanese monolinguals. The production experiments were conducted also in Mandarin and Japanese in an attempt to see whether the VOT values of the L1 and L2 of bilinguals were different from those of monolinguals. The results of these production experiments found that bilinguals produced L3 VOT in a native-like pattern, and their pattern was not influenced by either the L1 or the L2 regardless of their AOA in Japan. Thus, knowing two languages facilitated the bilinguals' production of a wide range of VOT contrasts, and therefore enabled them to produce L3 VOT accurately. The patterns of their L1 and L2 production of the bilinguals were also identical to those of Mandarin and Japanese monolinguals respectively. These results of production experiments are compatible with the prediction inferred from the BLIP model, which I supposed would predict that provided that the bilinguals have forged a new neural map and phonemic category for the Japanese [+voice] feature then they should have all the neural maps or categories to perceive the plosives contrasts in the three languages.

In the perception experiment, the bilinguals and monolingual speakers of Mandarin, Japanese and English listened to a VOT continuum of the English words *back* and *pack* ranging from lead VOT to long lag VOT and had to classify each word as corresponding to the word *back* or *pack*. The results show that there was an L1 influence on L3 perception pattern of bilinguals despite the fact that L2 perception pattern by Japanese monolinguals was more similar to L3 perception pattern than L1 perception pattern by Mandarin Chinese monolinguals. Here, it is indicated that in L3 perception, bilinguals had difficulty in associating the underlying features with VOT neural maps at the phonological level, because of the different number of underlying features between L1/L2 and L3.

Thus, the results of the series of production and perception experiments suggest that bilinguals can produce L3 with a native-like norm without L1/L2 transfer, but their L3 perception pattern was distinct from a native-like norm, and was influenced by L1. Therefore, the present study has a significant implication for the studies of not only L3 phonetic acquisition but also the relationship between production and perception in that it showed a different pattern of performance in L3 production and perception, which may be due to what each task is measuring (in production the task measures *averages* VOT values for each category, while in perception, the task measures the categorical boundary *between* categories along a VOT continuum). In addition, the present study also showed that the predictions on L3 production inferred from the BLIP model, which is originally a model of speech processing in perception with implications for the study of language acquisition, were compatible with the results of the production experiments. That is, the present study suggests that the BLIP model has the potential to be extended to predict speech processing not only in perception but also in production.

Before concluding the thesis, I would like to propose ideas for future research considering several points to be improved in the current study. First of all, the sample used in the current experiment was small. Although it was possible to get statistically significant results, a study with a larger number of participants may reveal more significantly different results, especially for the comparison between early and late bilinguals, or between bilinguals and L3 monolinguals in L3 perception. In addition, the present VOT continuum was made with equal steps of 20 ms. It is possible that by using steps of 5 ms or 10 ms a difference in the performance of English and Japanese may be detected. Finally, in the present study, what kind of English input the participants had received in their English education was not clear. If we understand the quality and quantity of the input, it may be possible for us to discuss the effect of the quality and the quantity of input in a target language. It would also be interesting to look not only at VOT but also at other acoustic cues such as F1 transitions. In fact, the effects of F1 transitions are expected to be observable in the data of L3 production of the current study, and therefore I will investigate them in the near future for a better understanding of L3 phonetic acquisition.

References

Abu-Rabia, S., & Sanitsky, E. (2010). Advantages of bilinguals over monolinguals in learning a third language. *Journal of the National Association for Bilingual Education*, *33*, 173-199.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software, 67*(1), 1-48

Best, C. T. (1995). A direct realist view of cross-language speech perception. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 171-204). Timonium, MD: York Press.

Boersma, P. & Weenink, D. (2015). Praat: doing phonetics by computer [Computer program]. Version 5.4.16, retrieved 17 Augest 2015 from http://www.praat.org/.

Bohn O.-S., Best, C. T. (2012). Native-language phonetic and phonological influences on perception of American English approximants by Danish and German listeners. *Journal of Phonetics 40*, 109-128.

Bohn, O.-S. & Flege, J. (1993). Perceptual switching in Spanish/English bilinguals: Evidence for universal factors in stop voicing judgments. *Journal of Phonetics*, *21*, 267-290.

Bullock, B. E., Toribio, A. J., González, V., & Dalola, A. (2006). Language dominance and performance outcomes in bilingual pronunciation. In M. G. O'Brien, C. Shea, & J. Archibald (Eds.), *Proceedings of the 8th Generative Approaches to Second Language Acquisition (GASLA): The Banff conference* (pp. 9–16). Somerville, MD: Cascadilla Press.

Caramazza, A., Yeni-Komshian, G., Zurif, E. & Carbone, E. (1973). The acquisition of a new phonological contrast: The case of stop consonants in French-English bilinguals. *Journal* *of the Acoustical Society of America*, *54*, 421–428.

Cabrelli Amaro, J. (2012). L3 phonology: An understudied domain. In J. Cabrelli Amaro, S. Flynn, & J. Rothman (Eds.), *Third language acquisition in adulthood* (pp. 33-60). Amsterdam: John Benjamins.

Cabrelli Amaro, J., & Rothman, J. (2010). On L3 acquisition and phonological permeability: A new test case for debates on the mental representation of non-native phonological systems. *IRAL*, *48*, 275-296.

Chao, K. Y. & Chen, L. M. (2008). A cross-linguistic study of voice onset time in stop consonant productions. *Computational Linguistics and Chinese Language Processing 13*(2), 215-232.

Chen, Y. (2007). From tone to accent: the tonal transfer strategy for Chinese L2 learners of Spanish. *Proceedings of ICPhS XVI*, 1645-8.

Cho, T., & Ladefoged, P. (1999). Variations and universals in VOT: evidence from 18 languages. *Journal of Phonetics*, *27*, 207–229.

Curtin, S., Goad, H. and Pater, J. 1998. Phonological transfer and levels of representation: The perceptual acquisition of Thai voice and aspiration by English and French speakers. *Second Language Research*, *14*, 389-405.

Eckman, F. (2004). From phonemic differences to constraint rankings: Research on second language phonology. *Studies in Second Language Acquisition 26*(4), 513-49.

Ellis, R. (1994), *The Study of Second Language Acquisition*, Oxford: Oxford University Press.

Falk, Y., & Bardel, C. (2010). The study of the role of the background languages in third language acquisition. The state of the art. *International Review of Applied Linguistics in*

Language Teaching. IRAL 48(2-3), 185-220.

Flege, J. (1987). The production of “new” and “similar” phones in a foreign language: Evidence for the effect of equivalence classification. *Journal of Phonetics*, *15*, 47-65.

Flege, J. (1991). Age of learning affects the authenticity of voice-onset time (VOT) in stop consonants produced in a second language. *Journal of the Acoustical Society of America*, *89*, 395–411.

Flege, J. (1993). Production and perception of a novel, second-language phonetic contrast. *Journal of the Acoustical Society of America*, *93*, 1589-1608.

Flege, J. (1995). Two methods for training a novel second-language phonetic contrast. *Applied Psycholinguistics*, *16*, 425-442.

Flege, J. & Eefting, W. (1987). The production and perception of English stops by Spanish speakers of English. *Journal of Phonetics*, *15*, 67-83.

Flege, J. & Eefting, W. (1988). Imitation of a VOT continuum by native speakers of English and Spanish: Evidence for phonetic category formation. *Journal of the Acoustical Society of America*, *83*, 729-740.

Flege, J. & Fletcher, K. (1992). Talker and listener effects on the perception of degree of foreign accent. *Journal of the Acoustical Society of America*, *91*, 370-389.

Flege, J., Frieda, A., & Nozawa, T. (1997). Amount of native-language (L1) use affects the pronunciation of an L2. *Journal of Phonetics*, *25*, 169-186.

Flege, J. & Hillenbrand, J. (1984). Limits on pronunciation accuracy in adult foreign language speech production. *Journal of the Acoustical Society of America*, *76*, 708-721.

Flege, J. & Hillenbrand, J. (1987a). Differential use of closure voicing and release burst as cue to stop voicing by native speakers of French and English. *Journal of Phonetics*, *15*, 203-208.

Flege, J., & Hillenbrand, J. (1987b). Limits on phonetic accuracy in foreign language speech production. In G. Ioup & S. Weinberger (Eds.), *Interlanguage phonology: The acquisition of a second language sound system* (pp. 176–203). Cambridge, MA: Newbury House.

Flege, J. & Liu, S. (2001). The effect of experience on adults' acquisition of a second language. *Studies in Second Language Acquisition*, *23*, 527-552.

Flege, J., MacKay, I., & Meador, D. (1999). Native Italian speakers' production and perception of English vowels. *Journal of the Acoustical Society of America*, 106, 2973-2987.

Flege, J., MacKay, I., & Piske, T. (2002). Assessing bilingual dominance. *Applied Psycholinguistics*, *23*, 567-598.

Flege, J., Munro, M. & MacKay, I. (1995). The effect of age of second language learning on the production of English consonants. *Speech Communication*, *16*, 1-26.

Flege, J. & MacKay, I. (2004). Perceiving vowels in a second language. *Studies in Second Language Acquisition*, *26*, 1-34.

Flege, J. & MacKay, I. (2011). What accounts for "age" effects on overall degree of foreign accent? In M. Wrembel, M. Kul and K. Dziubalska-Kołaczyk. (Eds.), *Achievements and perspectives in the acquisition of second language speech: New Sounds 2010, Vol. 2* (pp. 65-82). Bern, Switzerland: Peter Lang.

Flege, J., Munro, M., & MacKay, I. (1995). Factors affecting degree of perceived foreign accent in a second language. *Journal of the Acoustical Society of America*, *97*, 3125-3134.

Flege, J. & Schmidt, A. (1995). Native speakers of Spanish show rate-dependent processing of English stop consonants. *Phonetica*, *52*, 90-111.

Flege, J., Schirru, C., & MacKay, I. (2003). Interaction between the native and second language phonetic subsystems. *Speech Communication*, *40*, 467-491.

Flege, J., Takagi, N., & Mann, V. (1996). Lexical familiarity and English-language experience affect Japanese adults' perception of /r/ and /l/. *Journal of the Acoustical Society of America*, *99*, 1161-1173.

Garcia-Sierra, A., Rivera-Gaxiola, M., Percaccio, C. R., Conboy, B. T., Romo, H., Klarman, L., Ortiz, S., & Kuhl, P. K. (2011). Bilingual language learning: An ERP study relating early brain responses to speech, language input, and later word production. *Journal of Phonetics, 39*(4), 546-557.

Goto, H. (1971). Auditory perception by normal Japanes e adults of the sounds "l" and "r", *Neuropsychologia*, *9*, 317-323.

González-Bueno, M. (1997). The effects of formal instruction on the acquisition of Spanish stop consonants. In W. R. Glass & A. T. Pérez-Leroux (Eds.), *Contemporary perspectives on the acquisition of Spanish, vol. 2: production, processing, and comprehension* (pp. 57-75). Somerville, MA: Cascadilla Press.

Grenon, I. 2010. *The Bi-Level Input Processing Model of First and Second Language Perception*. Ph.D. dissertation. U of Victoria, Canada.

Gut, U. (2010). Cross-linguistic influence in L3 phonological acquisition. *International Journal of Multilingualism*, *7*, 19-38.

Flege, J. & Wang, C. (1990). Native-language phonotactic constraints affect how well Chinese subjects perceive the word-final English /t/-/d/ contrast. *Journal of Phonetics*, *17*, 299-315.

Hammarberg, B. and Hammarberg, B. (2005) Resetting the basis of articulation in the acquisition of new language: A third-language case study. In B. Hufeisen and R. Fouser, R. (Eds.), *Introductory Readings in L3* (pp. 11-18). Tubingen: Stauffenburg.

Harada, T. (2007). The production of voice onset time (VOT) by English-speaking children in a Japanese immersion program. *International Review of Applied Linguistics in Language Teaching (IRAL) 45*(4), 353-378.

Homma, Y. (1985). *Nichieigo-no onkyo onseigaku [acoustic phonetics in English and Japanese]*. Kyoto: Yamaguchi Shoten.

Keating, P., Linker, W. & Huffman, M. (1983). Patterns in allophone distribution for voiced and voiceless stops. *Journal of Phonetics, 11*: 277-290.

Kellerman, E. (1977). Towards a characterization of the strategy of transfer in second language learning. *Inrerlanguage Studies Bulletin Y1*, 58-145.

Klatt, D. (1975). Voice onset time, frication and aspiration in word-initial clusters. *Journal of Speech and Hearing Research*, *18*, 686–705.

Levy, E. S., & Law, F. F. I. (2010). Production of French vowels by American-English learners of French: Language experience, consonantal context, and the perception-production relationship. *Journal of the Acoustical Society of America*, *128*, 1290–1305.

Lisker, L., & Abramson, A. (1964). A cross-language study of voicing in initial stops. *Word*, *20*, 384-422.

Liu, H., Ng, M., Wan, M., Wang, S., & Y., Z. (2008). The effect of tonal changes on voice onset time in Mandarin esophageal speech. *Journal of Voice*, *22* (2), 210–218.

Llama, R., Cardoso, W., & Collins, L. (2010). The influence of language distance and language status on the acquisition of L3 phonology. *International Journal of Multilingualism*, *7*(1), 39-57.

Llisterri, J. & Poch- Olivé, D. (1987). Phonetic interference in bilingual's learning of a third language. *Proceedings of the XIth International Congress of Phonetic Sciences. Tallinn: Academy of Sciences of the Estonian SSR*, *5*, 134-147.

Lord, G. (2008). Second language acquisition and first language phonological modification. In J. Bruhn de Garavito and E. Valenzuela (Eds.), *Selected Proceedings of the 10th Hispanic Linguistics Symposium* (pp. 184-193). Somerville, MA: Cascadilla Proceedings Project.

MacKay, I., Meador, D., & Flege, J. (2001). The identification of English consonants by native speakers of Italian. *Phonetica*, *58*, 103-125.

Major, R. C. (1990). Phonological universals in second language variation. In H. Burmeister & P. Rounds (Eds.), *Variability in second language acquisition: Proceedings of the tenth meeting of the second language research forum*, Vol. II (pp. 473-488). Eugene, OR: Dept. of Linguistics, University of Oregon.

Major, R. C. (1992). Losing English as a first language. *The Modern Language Journal*, *76*, 190–208.

Matsuura, N. and Shimizu, K. (2002). *A study onvoicing contrasts of stops by Japanese learners and GA speakers*. Paper presented at JACET Chubu Chapter Meeting.

Meador, D., Flege, J., & MacKay, I. (2000). Factors affecting the recognition of words in a second language. *Bilingualism: Language & Cognition*, *3*, 55-67.

Munro, M., Flege, J., & MacKay, I. (1996). The effect of age of second-language learning on the production of English vowels. *Applied Psycholinguistics*, *17*, 313-334.

Nasukawa, K. (2010). Place-dependent VOT in L2 acquisition. *SLRF 2008: Selected proceedings of the 2008 Second Language Research Forum: exploring SLA Perspectives, positions, and practices* (pp.197-210). Somerville: Cascadilla Press.

Oglesbee, E. N. (2008). Multidimensional stop categorization in English, Spanish, Korean, Japanese, and Canadian French. Ph.D. dissertation. Indiana University.

Oyama S. (1976). A sensitive period for the acquisition of a nonnative phonological system. *Journal of Psycholinguistic Research*, Vol. 5, 261–283.

Piske, T., MacKay, I., & Flege, J. (2001). Factors affecting degree of foreign accent in an L2: A review. *Journal of Phonetics*, *29*, 191-215.

Piske, T., Flege, J., MacKay, I., & Meador, D. (2002). The production of English vowels by fluent early and late Italian-English bilinguals. *Phonetica*, *59*, 49-71.

Piske, T., Flege, J., MacKay, I., and Meador, D. (2011). Investigating native and non-native vowels produced in conversational speech. In M. Wrembel, M. Kul & K. Dziubalska-Kołaczyk (Eds.), *Achievements and perspectives in the acquisition of second language speech: New Sounds 2010*, Vol. 2 (pp. 195-205). Bern, Switzerland: Peter Lang.

Polka, L., Colantonio, C., & Sundara, M. (2001). A cross-language comparison of /d/-/ð/ perception: Evidence for a new developmental pattern. *Journal of Acoustical Society of America, 5*, 2190–2200.

Proctor, C. P., August, D., Snow, C. E., & Barr, C. (2010). The interdependence continuum: A perspective on the nature of Spanish-English bilingual reading comprehension. *Bilingual Research Journal*, *33*(1), 5-20.

R Core Team (2015). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL: http://www.R-project.org/.

Raphael, L. J., Tobin, Y., Faber, A., Most, T., Kollia, H. B., & Milstein, D. (1995). Intermediate values of voice onset time. In F. Bell-Berti & L. J. Raphael (Eds.), *Producing Speech: Contemporary Issues. For Katherine Safford Harris* (pp. 117-127). New York: AIP Press.

Riney, T. J., Takagi, N., Ota, K., & Uchida, Y. (2007). The intermediate degree of VOT in Japanese initial voiceless stops. *Journal of Phonetics*, *35*, 439-443.

Ringbom, H. (1987). *The role of the mother tongue in foreign language Learning*. Clevedon: Multilingual Matters.

Rochet, B. L., and Y. Fei. (1991). Effect of consonant and vowel context on Mandarin Chinese VOT: production and perception. *Canadian Acoustics*, *19*(4), 105-106.

Sancier, M. L. & Fowler, C. A. (1997). Gestural drift in a bilingual speaker of Brazilian Portuguese and English. *Journal of Phonetics*, *25*, 421-436.

Sato, T. (1958). On the Differences in Time Structures of Voiced and Unvoiced Stop Consonants. *The Journal of the Acoustical Society of Japan, 14*(2), 117-122.

Sheldon, A., and Strange, W. (1982). The acquisition of /r/ and /l/ by Japanese learners of English: Evidence that speech production can precede speech perception. *Applied Psycholinguistics*, *3*, 243-261.

Shimizu, K. (1995). *Eigo onseigaku (Phonetics of English)*. Tokyo: Sokei-shobo.

Shimizu, K. (1996). *A cross-language study of voicing contrasts of stop consonants in Asian languages*. Tokyo: Seibido Publishing Co.

Simon, E. 2010. Child L2 development: A longitudinal case study on voice onset times in word-initial stops. *Journal of Child Language*, *37*, 159-173.

Sypiańska, J. (2013). *Quantity and quality of language use and L1 attrition of Polish due to L2 Danish and L3 English*. Poznań: Adam Mickiewicz University. Unpublished PhD dissertation.

Takada, M. (2011). *Nihongo goto heisaon no kenkyu: VOT no kyojiteki bunpu to tsujiteki henka*. Tokyo: Kurosio.

Thornburgh, D. F. & Ryalls, J. H. (1998). Voice onset time in Spanish-English bilinguals: Early versus late learners of English. *Journal of Communicative Disorders*, *31*, 215-229.

Tremblay, M. C. (2007). *L2 influence on L3 pronunciation: Native-like VOT in the L3 Japanese of English-French bilinguals*. Paper presented at the Satellite Workshop of ICPhS XVI, Freiburg, Germany, 3-4 August 2007.

Tsukada, K., Birdsong, D., Bialystok, E., Mack, M., Sung, H., & Flege, J. (2005). A developmental study of English vowel production and perception by native English adults and children. *Journal of Phonetics*, *33*, 263-290.

Wang, C., 1995: *The acquisition of English word-final obstruents by Chinese speakers*. Ph.D.dissertation, State University of New York at Stony Brook, NY.

Wickham, H., (2009). ggplot2: elegant graphics for data analysis. Springer, New York.

Wilson, I., & Hashimoto, Y. (2013). Voice Onset Time and the Perception of Japanese Voicing Contrasts. *Journal of Nagoya Gakuin University; LANGUAGE and CULTURE, 24*(2), 15-22.

Wrembel, M. (2010). L2-accented speech in L3 production. *International Journal of Multilingualism*, *7* (1), 75-90.

Wrembel, M. (2011). Cross-linguistic influence in third language acquisition of voice onset time. In W.-S. Lee & E. Zee (Eds.), *Proceedings of the 17th International Congress of*

Phonetic Sciences. 17-21 August 2011. Hong Kong. CD-ROM (pp. 2157-2160). Hong Kong: City University of Hong Kong.

Wrembel, M. (2012a). Foreign accent ratings in third language acquisition; the case of L3 French. In E. Waniek-Klimczak & L. Shockey (Eds.), *Teaching and researching English accents in native and non-native speakers* (pp. 29-45). Berlin/Heidelberg: Springer Verlag.

Wrembel, M. (2012b). Foreign accentedness in Third Language Acquisition; the case of L3 English. In J. Cabrelli Amaro, S. Flynn, & J. Rothman (Eds.), *Third Language Acquisition in Adulthood* (pp. 281-309). Amsterdam: John Benjamins.

Wrembel M. (2014). VOT Patterns in the Acquisition of Third Language Phonology. *Concordia Papers in Applied Linguistics (COPAL)*, *5*: 751-771.

Wunder, E. (2010). Phonological cross-linguistic influence in third or additional language acquisition. *Proceedings of 6th New Sounds, Poznan*, 566-571.

Yamada, R. A., Strange, W., Magnuson, J. S., Pruitt, J. S., & Clarke, W. D. (1994). III The intelligibility of Japanese speakers' productions of American English /r/, /l/, and /w/, as evaluated by native speakers of American English. *Proceedings of the International Conference of Spoken Language Processing* (pp. 2023–2026). Yokohama: Acoustical Society of Japan.

Yao, Y. (2009). Understanding VOT variation in spontaneous speech. *Proceedings of 18th International Congress of Linguists (CIL XVIII)* (pp. 29–43). Seoul: Korea University.

Appendix

Appendix A. Language background questionnaire for bilinguals.

Language Background Questionnaire

Please answer the following questions to the best of your knowledge. Information will remain confidential.

1. Age: __________(years old)

2. Sex: __ Male __ Female

3. Education (check the highest degree obtained or school level attended):

__ Middle School __ High School __ College __ Graduate School

Other (specify): ____

4. List the languages in order of exposure (earliest first):

1.
2.
3.
4.
5.

5. List the languages in order of proficiency (most proficient first):

1.
2.
3.
4.
5.

6. Your place of birth:

Province, Autonomous Region, City: ____

County, City: ____

7. At what age did you move to Japan? : ____________________(years old)

8. How long have you been in Japan?: __________(years) __________(months)

If you moved back and forth between China and Japan, please explain.

Period	Country
from years old to years old	
from years old to years old	
from years old to years old	

9. What language does your mother speak? : ____________________________________

10. What language does your father speak? : ______________________________________

11. How good in general do you feel you are at learning new languages? Circle the number on the scale:

Very poor Poor Fair Neutral Good Very good Excellent

1-------------2-------------3-------------4-------------5-------------6-------------7

12. Write in the box the age at which you first learned Japanese and English, and the number of years you have spent learning each language.

Language	Age first learned the language	Number of years spent learning (cumulative)
Japanese	(years old)	(years)
English	(years old)	(years)

13. How did you learn Japanese and English? Check the box of one or both of the following situations.

Language	In the classroom	Spontaneously
Japanese		
English		

14. If you have taken a standardized language proficiency test in English (e.g. TOEIC, Eiken), please indicate the name of the test, and the scores you received for each. (If you don't remember, write down a guess in the appropriate column.)

Test	Actual Score	Guessed Score

15. Please rate your linguistic ability of Mandarin, Japanese, and English. Check the box of the proficiency level for each language.

	Beginner	Intermediate		Advanced	Native or native-like
		lower	higher		
Reading					
Mandarin					
Japanese					
English					
Writing					
Mandarin					
Japanese					
English					
Speaking					
Mandarin					
Japanese					
English					
Listening					
Mandarin					
Japanese					
English					
Overall competence					
Mandarin					
Japanese					
English					

16. Have you ever spent time in an area where English was the native language? (including short trips)?

Where?	When?	How long?

17. Which language do you speak when you do the following activities? Check the box of any of Mandarin, Japanese, and English.

Activities	Mandarin	Japanese	English
Watching TV			
Reading for fun			
Reading for work			
Reading on the Internet			
Writing emails to friends			
Writing articles/papers			
Talking with family members			
Talking with friends			
Talking with colleagues at a part-time job			
Doing a calculation in your head			
Remembering numbers (e.g. telephone)			
Dreaming			
Thinking			
Expressing anger or affection			

18. Do you have (or ever had) any known hearing or speech impairment?

__ Yes __ No

18. If there is anything else that you feel is interesting or important about your language background or language use, please comment below.

Thank you very much!

Appendix B. Language background questionnaire for monolingual speakers.

Language Background Questionnaire

言語環境に関するアンケート

***Information will remain confidential**

*情報が他に公開されることはありません

A. About Yourself　あなたについて

- Sex　性別: ☐ Male　男　☐ Female 女
- Age　年齢: ______________________
- Place of Birth　出生地 (国および県) : ________________________________
- Highest level of education　最終学歴（在学中含む）:

  ☐ High school　高卒　☐ University　大卒　☐ Gradute School　大学院卒

  ☐ Other institutions　他の高等機関 ______________________________

B. First Language(s)　母国語

What is your first language?　あなたの母国語は何ですか？: ______________________

What is the first language of:

your mother?　お母さんの母国語は何ですか？: ______________________

your father?　お父さんの母国語は何ですか？: ______________________

Did you learn or were exposed to any other language before your native language?

母国語を習う前に他の言語を習いましたか？

☐ Yes　はい　☐ No　いいえ

- If you answered 'Yes', please explain.　もしそうであれば、説明していただけますか？:

How good in general do you feel you are at learning new languages? Circle the number on the scale.

あなたは、新しい言語を習得するのがどれくらい得意ですか？あてはまる数字に○をつけてください。

Very poor	Poor	Fair	Neutral	Good	Very good	Excellent
とても不得意	不得意	やや不得意	普通	やや得意	得意	とても得意

1-------------2-------------3-------------4-------------5-------------6-------------7

C. Second Languages　第二言語

	Second Languages 第二言語	
	1: English 英語	**2: ______________**
At what age did you begin to learn your second language? 第二言語を習い始めたのは、 何歳のときですか？	(years old) (才)	(years old) (才)
How many years in total have you studied this language? その言語を学んだ期間は 通算どのくらいですか？	(years) (年)	(years) (年)
How did you learn your second language? どのように第二言語を学びましたか？	☐ School 外国語の授業として ☐ Natulally 生活しているうちに自然と ☐ Other その他 ______________	☐ School 外国語の授業として ☐ Natulally 生活しているうちに自然と ☐ Other その他 ______________
Have you ever spent time in an area where this language was the native language? その第二言語を母国語とする場所で 過ごしたことはありますか？（旅行等含む）	Where?　その場所は When?　いつですか？ How long?　期間は？	Where?　その場所は When?　いつですか？ How long?　期間は？
Where do you use this language? 第二言語をどこで使っていますか。	☐ School　学校 ☐ Work 職場 ☐ Home　家庭 ☐ Social Situations その他 ______________	☐ School　学校 ☐ Work 職場 ☐ Home　家庭 ☐ Social Situations その他 ______________

D. Self Evaluation　自己評価

Please rate your linguistic ability in each of the languages you speak or have studied.

あなたが話している、もしくは習った第二言語に関して、以下の各項目の能力を自己評価し、当てはまる欄に √をつけて下さい。第二言語が英語以外の方は、「その他」の欄に具体的に記入して下さい。

	Beginner **初級**	**Intermediate** **lower** **中下級**	 **higher** **中上級**	**Advanced** **上級**	**Native or native-like** **母語話者** **または同等レベル**
Reading　読む能力					
English　英語					
Other その他 ________					
Other その他 ________					
Writing　書く能力					
English 英語					
Other その他 ________					
Other その他 ________					
Speaking　話す能力					
English 英語					
Other その他 ________					
Other その他 ________					
Listening　聞く能力					
English 英語					
Other その他 ________					
Other その他 ________					
Overall competence 総合評価					
English 英語					
Other その他 ________					
Other その他 ________					

E. Other relevant information　その他の関連情報

If you have taken a standardized language proficiency test in English (e.g. TOEFL), please indicate the name of the test and the scores you received for each. (If you don't remember, write down a guess in the appropriate column.)

TOEICや英検など、英語の言語能力テストを受けたことがあれば、そのテスト名と成績を書いてください。(正確な成績がわからない場合は、「だいたいの成績」の欄に、推定の成績を書いてください。)

Test テスト名	Actual Score 成績	Approximant Score だいたいの成績

Do you have (or ever had) any known hearing or speech impairment?

聴覚または発声に異常があると診断されたことがありますか？

☐ yes　はい　　☐ no　いいえ

• If yes, can you please specify how it affects your hearing abilities？

もし障害をお持ちの場合は、宜しければ、どの様なものか具体的に教えて頂けますか？

Please add any other comments or observations you think might be pertinent:

あなたの言語等について、今まで質問させて頂いた内容以外で、あなたが重要と思われる情報があれば、具体的に教えて下さい。

Thank you!

どうもありがとうございました！

Appendix C1. ANOVA Tables for the results of Mandarin production experiment

Labial plosives

- Comparing Mandarin monolinguals and early bilinguals

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
Group	1	713	713	1.4192	0.2355	
Voicing	1	327182	327182	651.6831	< 2.2e-16	***
Group:Participants	19	33506	1763	3.5125	8.045e-06	***
Group:Voicing	1	78	78	0.1545	0.6948	
Residuals	145	72798	502			

- Comparing Mandarin monolinguals and late bilinguals

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
Group	1	2523	2523	5.8666	0.01718	*
Voicing	1	207659	207659	482.9412	< 2.2e-16	***
Group:Participants	13	21628	1664	3.8692	4.407e-05	***
Group:Voicing	1	2526	2526	5.8743	0.01711	*
Residuals	103	44289	430			

- Comparing early and late bilinguals

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
Group	1	863	863	2.1274	0.14754	
Voicing	1	213769	213769	526.8451	< 2.2e-16	***
Group:Participants	14	25336	1810	4.4601	2.791e-06	***
Group:Voicing	1	1927	1927	4.7501	0.03143	*
Residuals	110	44633	406			

Alveolar plosives

- Comparing Mandarin monolinguals and early bilinguals

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
Group	1	107	107	0.1537	0.6956026	
Voicing	1	295747	295747	424.2741	< 2.2e-16	***
Group:Participants	19	34529	1817	2.6071	0.0006809	***
Group:Voicing	1	304	304	0.4363	0.5099677	
Residuals	145	101074	697			

- Comparing Mandarin monolinguals and late bilinguals

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
Group	1	459	459	1.4232	0.2356	
Voicing	1	214697	214697	665.1038	< 2.2e-16	***
Group:Participants	13	24414	1878	5.8178	6.279e-08	***
Group:Voicing	1	275	275	0.8511	0.3584	
Residuals	103	33249	323			

- Comparing early and late bilinguals

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
Group	1	179	179	0.2228	0.6378333	
Voicing	1	210087	210087	261.4389	< 2.2e-16	***
Group:Participants	14	37074	2648	3.2954	0.0002098	***
Group:Voicing	1	7	7	0.0091	0.9241639	
Residuals	110	88394	804			

Velar plosives

- Comparing Mandarin monolinguals and early bilinguals

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
Group	1	25	25	0.0411	0.8395936	
Voicing	1	264035	264035	433.5449	< 2.2e-16	***
Group:Participants	19	33842	1781	2.9247	0.0001454	***
Group:Voicing	1	6	6	0.0097	0.9217728	
Residuals	145	88307	609			

- Comparing Mandarin monolinguals and late bilinguals

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
Group	1	21	21	0.0532	0.8180	
Voicing	1	189226	189226	472.1212	< 2.2e-16	***
Group:Participants	13	19511	1501	3.7446	6.798e-05	***
Group:Voicing	1	4	4	0.0102	0.9199	
Residuals	103	41282	401			

- Comparing early and late bilinguals

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
Group	1	0	0	0.0005	0.982955	
Voicing	1	199311	199311	224.8984	< 2.2e-16	***
Group:Participants	14	30017	2144	2.4193	0.005397	**
Group:Voicing	1	0	0	0.0000	0.997771	
Residuals	110	97485	886			

Appendix C2. ANOVA Tables for the results of Japanese production experiment

Labial plosives

- Comparing Japanese monolinguals and early bilinguals

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
Group	1	4775	4775	3.1167	0.07960	.
Voicing	1	218515	218515	142.6193	< 2.2e-16	***
Group:Participants	19	93749	4934	3.2204	3.398e-05	***
Group:Voicing	1	10154	10154	6.6271	0.01105	*
Residuals	145	222162	1532			

- Comparing Japanese monolinguals and late bilinguals

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
Group	1	3745	3745	2.9577	0.08847	.
Voicing	1	95416	95416	75.3494	6.262e-14	***
Group:Participants	13	65945	5073	4.0059	2.742e-05	***
Group:Voicing	1	20	20	0.0156	0.90088	
Residuals	103	130430	1266			

- Comparing early and late bilinguals

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
Group	1	13955	13955	11.0699	0.001194	**
Voicing	1	193511	193511	153.5090	< 2.2e-16	***
Group:Participants	14	56378	4027	3.1945	0.000306	***
Group:Voicing	1	5947	5947	4.7177	0.032000	*

Residuals	110	138664	1261		

Alveolar plosives

- Comparing Japanese monolinguals and early bilinguals

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Group	1	789	789	0.5495	0.4597
Voicing	1	168136	168136	117.0855	< 2.2e-16 ***
Group:Participants	19	88778	4673	3.2538	2.882e-05 ***
Group:Voicing	1	836	836	0.5823	0.4467
Residuals	145	208221	1436		

- Comparing Japanese monolinguals and late bilinguals

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Group	1	4566	4566	5.3721	0.02244 *
Voicing	1	83360	83360	98.0692	< 2.2e-16 ***
Group:Participants	13	59292	4561	5.3657	2.745e-07 ***
Group:Voicing	1	2073	2073	2.4392	0.12140
Residuals	103	87551	850		

- Comparing early and late bilinguals

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Group	1	2104	2104	1.4696	0.22800
Voicing	1	112234	112234	78.4067	1.628e-14 ***
Group:Participants	14	42663	3047	2.1289	0.01527 *
Group:Voicing	1	4853	4853	3.3905	0.06827 .
Residuals	110	157457	1431		

Velar plosives

- Comparing Japanese monolinguals and early bilinguals

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Group	1	396	396	0.2496	0.6181156
Voicing	1	203696	203696	128.4252	< 2.2e-16 ***
Group:Participants	19	78105	4111	2.5917	0.0007331 ***
Group:Voicing	1	2448	2448	1.5432	0.2161430

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Residuals	145	229985	1586		

- Comparing Japanese monolinguals and late bilinguals

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Group	1	718	718	0.6793	0.4117
Voicing	1	110709	110709	104.6734	< 2.2e-16 ***
Group:Participants	13	53134	4087	3.8644	4.48e-05 ***
Group:Voicing	1	47	47	0.0446	0.8331
Residuals	103	108939	1058		

- Comparing early and late bilinguals

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Group	1	1878	1878	1.0403	0.30999
Voicing	1	162733	162733	90.1351	5.658e-16 ***
Group:Participants	14	45973	3284	1.8189	0.04434 *
Group:Voicing	1	2214	2214	1.2264	0.27053
Residuals	110	198597	1805		

Appendix C3. ANOVA Tables for the results of English production experiments

Labial plosives

- Comparing Mandarin monolinguals and early bilinguals

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Group	1	3	3	0.0040	0.94941
Voicing	1	163061	163061	244.2734	< 2e-16 ***
Group:Participants	19	23954	1261	1.8887	0.01895 *
Group:Voicing	1	50	50	0.0749	0.78467
Residuals	145	96793	668		

- Comparing English monolinguals and early bilinguals

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Group	1	882	882	1.2353	0.2684
Voicing	1	157003	157003	219.9299	<2e-16 ***
Group:Participants	17	19809	1165	1.6322	0.0647 .

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Group:Voicing	1	45	45	0.0631	0.8021
Residuals	131	93518	714		

- Comparing Japanese monolinguals and early bilinguals

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Group	1	21545	21545	24.9484	1.67e-06 ***
Voicing	1	121807	121807	141.0475	< 2.2e-16 ***
Group:Participants	19	31174	1641	1.8999	0.01805 *
Group:Voicing	1	4166	4166	4.8244	0.02965 *
Residuals	145	125221	864		

- Comparing early and late bilinguals

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Group	1	4455	4455	4.9658	0.02789 *
Voicing	1	123740	123740	137.9275	< 2e-16 ***
Group:Participants	14	14358	1026	1.1432	0.32961
Group:Voicing	1	96	96	0.1066	0.74470
Residuals	110	98685	897		

- Comparing Mandarin and English monolinguals

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Group	1	936	936	2.4119	0.1229619
Voicing	1	143409	143409	369.3470	< 2.2e-16 ***
Group:Participants	16	18365	1148	2.9561	0.0003631 ***
Group:Voicing	1	171	171	0.4412	0.5077531
Residuals	124	48146	388		

- Comparing Mandarin and Japanese monolinguals

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Group	1	21029	21029	36.3428	1.43e-08 ***
Voicing	1	109352	109352	188.9898	< 2.2e-16 ***
Group:Participants	18	29730	1652	2.8545	0.0002855 ***
Group:Voicing	1	3153	3153	5.4494	0.0210183 *
Residuals	138	79849	579		

- Mandarin monolinguals and late bilinguals

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Group	1	4494	4494	8.6820	0.003975 **
Voicing	1	110351	110351	213.1961	< 2.2e-16 ***
Group:Participants	13	12914	993	1.9192	0.036060 *
Group:Voicing	1	16	16	0.0307	0.861242
Residuals	103	53313	518		

- English and Japanese monolinguals

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Group	1	11261	11261	18.2350	3.848e-05 ***
Voicing	1	102082	102082	165.3065	< 2.2e-16 ***
Group:Participants	16	25584	1599	2.5893	0.001675 **
Group:Voicing	1	4360	4360	7.0603	0.008918 **
Residuals	124	76574	618		

- Comparing English monolinguals and late bilinguals

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Group	1	1517	1517	2.6975	0.1040
Voicing	1	104087	104087	185.1326	<2e-16 ***
Group:Participants	11	8768	797	1.4178	0.1789
Group:Voicing	1	217	217	0.3854	0.5363
Residuals	89	50039	562		

- Comparing Japanese monolinguals late bilinguals

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Group	1	2638	2638	3.3246	0.07115 .
Voicing	1	71477	71477	90.0667	1.006e-15 ***
Group:Participants	13	20133	1549	1.9515	0.03250 *
Group:Voicing	1	1752	1752	2.2081	0.14034
Residuals	103	81741	794		

Alveolar plosives

- Comparing Mandarin monolinguals and early bilinguals

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
Group	1	2945	2945	4.1920	0.0424196	*
Voicing	1	285109	285109	405.8661	< 2.2e-16	***
Group:Participants	19	36173	1904	2.7102	0.0004136	***
Group:Voicing	1	425	425	0.6051	0.4379088	
Residuals	145	101858	702			

- Comparing Mandarin and English monolinguals

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
Group	1	158	158	0.5091	0.4769	
Voicing	1	219082	219082	706.8455	< 2.2e-16	***
Group:Participants	16	23815	1488	4.8023	1.625e-07	***
Group:Voicing	1	49	49	0.1584	0.6914	
Residuals	124	38433	310			

- Comparing Japanese and Mandarin monolinguals

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
Group	1	33554	33554	48.6803	1.15e-10	***
Voicing	1	213744	213744	310.0985	< 2.2e-16	***
Group:Participants	18	50572	2810	4.0761	9.37e-07	***
Group:Voicing	1	1418	1418	2.0577	0.1537	
Residuals	138	95121	689			

- Comparing Mandarin monolinguals and late bilinguals

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
Group	1	4794	4794	10.6268	0.0015111	**
Voicing	1	152649	152649	338.4001	< 2.2e-16	***
Group:Participants	13	19019	1463	3.2432	0.0003921	***
Group:Voicing	1	3578	3578	7.9322	0.0058214	**
Residuals	103	46462	451			

- Comparing English monolinguals and early bilinguals

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
Group	1	1460	1460	2.0223	0.15739	
Voicing	1	253976	253976	351.8825	< 2e-16	***
Group:Participants	17	23048	1356	1.8784	0.02518	*
Group:Voicing	1	704	704	0.9758	0.32506	
Residuals	131	94551	722			

- Comparing Japanese monolinguals and early bilinguals

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
Group	1	17748	17748	17.0163	6.219e-05	***
Voicing	1	247211	247211	237.0138	< 2.2e-16	***
Group:Participants	19	49805	2621	2.5132	0.001069	**
Group:Voicing	1	3500	3500	3.3560	0.069012	.
Residuals	145	151238	1043			

- Comparing early and bilinguals

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
Group	1	694	694	0.7445	0.39010	
Voicing	1	185778	185778	199.2154	< 2e-16	***
Group:Participants	14	18252	1304	1.3980	0.16606	
Group:Voicing	1	5998	5998	6.4318	0.01261	*
Residuals	110	102580	933			

- Comparing Japanese and English monolinguals

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
Group	1	25645	25645	36.2133	1.849e-08	***
Voicing	1	183496	183496	259.1129	< 2.2e-16	***
Group:Participants	16	37447	2340	3.3049	8.362e-05	***
Group:Voicing	1	812	812	1.1471	0.2862	
Residuals	124	87813	708			

- Comparing English monolinguals and late bilinguals

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Group	1	3144	3144	7.1457	0.008939 **
Voicing	1	122706	122706	278.9136	< 2.2e-16 ***
Group:Participants	11	5894	536	1.2179	0.287114
Group:Voicing	1	2667	2667	6.0618	0.015740 *
Residuals	89	39155	440		

- Comparing Japanese monolinguals and late bilinguals

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Group	1	6453	6453	6.9345	0.009757 **
Voicing	1	120559	120559	129.5627	< 2.2e-16 ***
Group:Participants	13	32650	2512	2.6991	0.002615 **
Group:Voicing	1	845	845	0.9080	0.342874
Residuals	103	95842	931		

Velar plosives

- Comparing Mandarin monolinguals and early monolinguals

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Group	1	0	0	0.0000	0.99923
Voicing	1	208227	208227	396.8483	< 2.2e-16 ***
Group:Participants	19	46079	2425	4.6221	3.54e-08 ***
Group:Voicing	1	2255	2255	4.2973	0.03994 *
Residuals	145	76082	525		

- Comparing English and Mandarin monolinguals

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Group	1	204	204	0.3204	0.57237
Voicing	1	155021	155021	242.9750	< 2e-16 ***
Group:Participants	16	17424	1089	1.7069	0.05351 .
Group:Voicing	1	377	377	0.5908	0.44356
Residuals	124	79113	638		

- Comparing Mandarin and Japanese monolinguals

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
Group	1	2119	2119	3.6497	0.05816	.
Voicing	1	148514	148514	255.7900	< 2.2e-16	***
Group:Participants	18	40799	2267	3.9038	2.092e-06	***
Group:Voicing	1	129	129	0.2217	0.63850	
Residuals	138	80124	581			

- Comparing Mandarin and late bilinguals

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
Group	1	163	163	0.4055	0.52568	
Voicing	1	104638	104638	260.4889	< 2e-16	***
Group:Participants	13	10651	819	2.0395	0.02442	*
Group:Voicing	1	807	807	2.0098	0.15931	
Residuals	103	41375	402			

- Comparing English monolinguals and early bilinguals

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
Group	1	214	214	0.2684	0.6052	
Voicing	1	207877	207877	261.1953	< 2.2e-16	***
Group:Participants	17	45959	2703	3.3969	3.573e-05	***
Group:Voicing	1	617	617	0.7748	0.3804	
Residuals	131	104259	796			

- Comparing Japanese monolinguals and early bilinguals

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
Group	1	2222	2222	3.0606	0.08233	.
Voicing	1	198245	198245	273.0679	< 2.2e-16	***
Group:Participants	19	69333	3649	5.0264	5.116e-09	***
Group:Voicing	1	3492	3492	4.8106	0.02988	*
Residuals	145	105269	726			

- Comparing early and late bilinguals

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
Group	1	168	168	0.2785	0.59872	
Voicing	1	154009	154009	254.6743	< 2.2e-16	***
Group:Participants	14	39185	2799	4.6284	1.511e-06	***
Group:Voicing	1	4532	4532	7.4945	0.00722	**
Residuals	110	66520	605			

- Comparing English and Japanese monolinguals

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
Group	1	847	847	0.9697	0.3266728	
Voicing	1	145747	145747	166.8746	< 2.2e-16	***
Group:Participants	16	40679	2542	2.9110	0.0004387	***
Group:Voicing	1	907	907	1.0381	0.3102403	
Residuals	124	108300	873			

- Comparing English monolinguals and late bilinguals

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
Group	1	0	0	0.0002	0.9896	
Voicing	1	101568	101568	129.9694	<2e-16	***
Group:Participants	11	10531	957	1.2250	0.2825	
Group:Voicing	1	1888	1888	2.4161	0.1236	
Residuals	89	69552	781			

- Comparing Japanese monolinguals and late bilinugals

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
Group	1	616	616	0.8994	0.3452	
Voicing	1	96334	96334	140.6200	< 2.2e-16	***
Group:Participants	13	33905	2608	3.8070	5.47e-05	***
Group:Voicing	1	367	367	0.5353	0.4660	
Residuals	103	70562	685			