THE EFFECT OF PERCEPTUAL TRAINING ON THE LEARNING OF JAPANESE FRICATIVE AND AFFRICATE CONTRASTS BY NATIVE THAI LEARNERS OF JAPANESE

BY

TANPORN TRAKANTALERNGSAK

A dissertation submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy

Graduate School of Language and Culture
OSAKA UNIVERSITY

2016

日本語要旨

タイ人学習者による日本語摩擦音及び破擦音の学習における知覚トレーニングの有効性

TANPORN TRAKANTALERNGSAK

タイ人日本語学習者は、日本語の摩擦音・破擦音の発音においてしばしば両者を混同しており、習得が困難であると指摘されている。タイ人学習者による摩擦音と破擦音の混同は、聞き手の意味の混同にもつながり、コミュニケーションに大きな支障をきたすことが予想される。また、日本語母語話者にとっては幼児のような発音と受け取られかねず、マイナスの印象を与えることも予想される。

日本語教育現場における実践的音声教育の研究は未だ少ないが、英語教育における発音指導に関する研究は実践的研究も含め数多く行われている。中でも、ESL(English as a second language)の発音指導には、英語の正しい知覚範疇の形成を促すという目的の下でHVPT法の知覚トレーニング(High Variability Phonetic Training)が導入され、分節音でも超分節音でも大きな効果のあることが確認された。HVPT法の知覚トレーニングは多種多様性が重視され、さまざまな音環境と語中位置に現れるミニマルペアの音声の強制選択肢の同定と区別、多数の母語話者の音声の聴き取り、また、回答直後に即時の正誤のフィードバックを行うなどの手法が取られている。このトレーニングは発音指導における大きな効果のあること、具体的にはL2における聴取能力の向上、さらには知覚範疇の形成・産出のプロセスの正確性の向上に対して有効であることが明らかにされている。

上述の知覚トレーニングの概念に加えて、第二言語習得レベルを母語話者レベルまで引き上げることを目指す学習においては、自分自身の発音に対する意識を高めること－自己意識化－が必要であると指摘されている。これらの認識に基づき、本論文では、日本語発音指導に未導入であるHVPT知覚トレーニングと、それに加えて新たに、学習者に自分自身が発話したものを聞かせるという、自己モニタリングの概念を組み込んだ実験を行った。つまり、

本論文は、指摘されている聴取・発音の混同を改善するための日本語発音指導方法の提案を目的とする実践的研究である。

本論文では、目的達成のために次のような実験を行った。31名のタイ国在住のタイ人学習者に被験者として実験に参加してもらい、11名を知覚トレーニングをしない統制群、残りの20名を知覚トレーニングを行う2つの実験群（A群：11名、B群：9名）に分けた。実験群のみに5週間にわたり15分～20分程度のHVPT法の知覚トレーニングを計9回実施し、知覚トレーニングがもたらす効果に関して検証を行った。実験群A群およびB群に対しては、まず同定タスクを課し、B群のみに対して、自己モニタリングの区別タスクを合わせて実施した。これによって自己の発音に対する認識の促進、すなわち自己意識化がトレーニングの効果においていかなる差を生ずるのかについて検証した。トレーニング効果の測定にはプレテスト、ポストテスト、般化テスト（Generalisation test）と6ヶ月後のディレイポストテストの計4回のテストを実施した。プレテストから般化テストまではおよそ3ヶ月間に渡って実施し、その6ヶ月後にディレイポストテストを実施した。

実験結果は以下のようにまとめられる。

⑴まず、知覚テストに関しては、統制群においても実験群AとB群においても正聴率の成績に上昇が見られた。全群で上昇が見られたものの、実験群AとBのポストテストの正聴率が統制群を有意に上回った。つまり、HVPT法の知覚トレーニングをすることで日本語の摩擦音と破擦音における聴き取り能力がより効果的に向上することを示す結果となった。また、実験群AとBにおいては正聴率そのものに有意差は見られなかった。結果からは、成人学習者のL2音声に対する知覚能力の向上を確認した。これらの聴き取りにおける正確性の向上を示す結果から、トレーニングを行うことによって、成人学習者であっても知覚範疇の形成や組み変えが可能となること、また産出のプロセスの正確性の向上、不正確な知覚の改善に至ることが確認された。

⑵自己意識化をさせた実験群Bのみに見られた特徴として、新語および新しい話者が発話した音声に対してトレーニングの有効性の転移が見られた。⑶さらに、実験群AとBのみが知覚トレーニング終了から6ヶ月後にも、向上した聴き取り能力を保持していた。

⑷産出テストに関しては次のような結果となった。日本語母語話者の判断による同定評価と典型性評価をおこなった結果、知覚トレーニング後は日本語の摩擦音と破擦音の産出において正確さと明瞭度が向上したことが認められた。つまり、知覚トレーニングしかしていないにも拘わらず、それが産

出能力にも転移し、向上したことがわかる。このことから、知覚と産出の間には関係があることが示唆された。

さらに、実験を通じて観察された事象からは、知覚と産出のプロセスにおける両者の相関性をより強く認識するに至った。これにより、第二言語習得研究においては、グループ平均値のみではなく、個人差をより考慮したうえで音声習得に至る要因を分析する必要があるという考察を本論文の成果の一つとして提示したい。

本論文では、タイ人学習者のために効果的な発音指導の実践方法を提案することを目指し、HVPT知覚トレーニングの効果を検証した。その結果、般化テスト以外では両実験群でトレーニングから同等の効果が得られたことから、HVPT知覚トレーニングを用いた知覚訓練は摩擦音と破擦音の習得に対して有効であることがわかった。トレーニングを実施した実験群にみられた摩擦音と破擦音の知覚における長期的な効果の発現と産出への転移がHVPT法の知覚トレーニングの有効性を証明したと言える。さらに、実験群Bに行ったように、発話音声を自己モニタリングする自己意識化トレーニングを取れ入れることにより、HVPT法の知覚トレーニング効果に加えて新語と新しい話者の発話の定着率が高められ、発音習得によりよい効果をもたらすという概念が支持できる。

本論文の結果としては、タイ人日本語学習者における日本語の摩擦音と破擦音を区別・産出する能力の改善において、知覚トレーニングが一定の効果をあげるものであると述べる。また、学習者に自らの発音を認識させる自己意識化の重要性を提案するものである。本論文では全体的に有効性を示す結果が得られたが、今後はさらに被験者のデータを蓄積した上で再度トレーニングを行い、詳細な分析をもとに知覚トレーニングの有効性に関するより精緻な確認を行う必要がある。研究対象として他の子音と母音を取り入れることやトレーニング手法の変更も行いつつ、より効果的な知覚トレーニング手法の構築に結び付けたい。

以下に、各章の概要を述べる。

第１章は「序論」であり、本論文の構造および各章の関連性を示す。

第２章は「先行研究」として、前半は、L2 音声学習、 SLMやPAMなどのL2音声学習モデル、知覚と算出の関係について叙述する。また、ここでHVPT法の知覚トレーニング方法について一概する。先行研究で効果があると評価されたトレーニング方法のうち数種類は、本実験でも取り入れていることをふまえて概説する。後半は先行研究で指摘されたタイ人学習者の摩擦音・破

擦音の発音および聞き取りの問題点や日本語教育における音声教育の現状についてその概要をまとめ、最後に、先行研究の結果や改善すべき点などを参考にしながら、本研究の目的と研究課題を示す。

第３章は「実験方法」で、先行研究をもとに、本実験で取り扱う調査方法を示す。

第４章は「結果」であり、章の前半部には計4回のテスト結果および、これらの実験から明らかになったことを提示する。章の後半は、知覚と産出の相関関係の分析結果と、合わせて実施した被験者へのアンケートとフォローアップインタビュー結果である。

第５章は「考察」とし、課題に対する解答を提示しながら、本論文の研究結果を明らかにする。これをもとに日本語音声教育方法の具体的な提案を行い、HVPT知覚トレーニングの応用を提言する。

第６章は「まとめ」と題して、本論文を総括し、最後に、残された問題点や今後の課題について論じる。

Abstract

THE EFFECT OF PERCEPTUAL TRAINING ON THE LEARNING OF JAPANESE FRICATIVE AND AFFRICATE CONTRASTS BY NATIVE THAI LEARNERS OF JAPANESE

TANPORN TRAKANTALERNGSAK

Perceptual training using a high variability method (HVPT perceptual training) has been shown to be an effective tool for improving the perception and production of L2 contrasts. The ultimate aim of this study is to investigate the effects of HVPT perceptual training on L2 contrast learning. This study examined whether Thai learners can be trained to better learn the Japanese fricative and affricate contrasts and to what extent perceptual training improved their perception and production learning ability. Moreover, it is acknowledged as part of this study that consciousness and awareness-raising are important in second language acquisition. The training approach used in this study differs from previous training studies in that self-monitoring was also applied to see whether self-awareness raising using a self-monitoring task would induce better learning.

Thirty-one adult native Thai speakers, 11 in the control group and 20 in the trained groups, participated in the study. Nine sessions of HVPT perceptual training were given to the 20 Thai native speaking learners of Japanese, who were randomly distributed into two groups: 11 trained with standardised HVPT perceptual training using an identification task (Trained Group A) and nine learners trained with a modified version of HVPT perceptual training which used both an identification task and a self-monitoring task (Trained Group B). Both groups were trained using the identification task with two-alternative forced-choices and immediate feedback. However, to investigate whether the effect of self-monitoring lead to

better learning or not, Trained Group B participants were asked to complete an identification task as well as an additional self-monitoring discrimination task.

The effect of HVPT perceptual training was measured using pre-tests and post-tests, generalisation tests and a delayed post-test. Overall, it can be concluded that HVPT perceptual training is beneficial for training Thai learners of Japanese. The results showed that 1) both trained groups significantly improved in their ability to perceive the target contrasts, 2) long-term retention of perceptual learning was observed, 3) a transfer of the perceptual learning to production was observed. However, there was reliable evidence of a transfer or generalisation to new words and a new talker observed only in Trained Group B.

These findings provide evidence that HVPT perceptual training can serve as an effective tool to improve Thai learners' abilities in the perception and production of Japanese fricative and affricate contrasts. Moreover, this study contributes to the growing literature showing that HVPT perceptual training leads to improvements in L2 adult learners' perception and production. Interestingly, the results also show that conducting HVPT perceptual training together with a self-monitoring task might yield better learning outcomes.

The findings of this study have relevance to theoretical issues in cross-linguistic speech perception, L2 speech training, and the relationship between L2 speech perception and production. Moreover, individual differences are also considered and discussed. Implications for Japanese language teaching and learning, limitations of this study as well as directions for future studies are also discussed.

Acknowledgements

There are many people who I would like to thank for helping me in the completion of this dissertation. First of all, I would like to thank my supervisor, Professor Yasuo Iwai, for his unending support throughout the course of both my Master's and doctoral degrees. He has always supervised my course of study, providing wisdom and insight since I first came to Osaka University in 2008 as an exchange student hoping to learn about Japanese linguistics. He introduced me to the study of phonetics of which I was not so fond at first! He then taught me how to overcome my initial worries and succeed. For his help and guidance I will always be grateful. Secondly, I would like to thank Professor Marasri Miyamoto for her inspiration and tips about living in Japan. I will always remember the laughs we had. I would also like to thank Professor Shiro Kori and Professor Masayoshi Kakudo for their willingness to provide guidance, feedback and support all of which has truly impacted and strengthened my work. Furthermore, I would like to thank MEXT (the Japan Ministry of Education, Culture, Sports, Science and Technology) for funding my degrees from the time of my exchange year, through to my Master's and to this doctoral degree. And of course, Assistant Professor Chomnard Setisarn who recommended me for a MEXT scholarship and inspired me to attempt to enter into the academic world to begin with.

I would like to dedicate this dissertation to Andrew Gledhill, who has shown a great understanding of my research along the way and has made this doctoral dissertation possible. A “thank you” alone might not be enough for what he has done for me during the course of this study. He has been better to me than I could have ever asked for. Without his support and warm encouragement I do not think I could have accomplished what I wanted, I truly would have been lost.

Importantly, I would like to highlight the contribution of Salita or Tak, a PhD student from the Department of Engineering, Osaka University. Tak was the person who helped me design the perceptual training program used in the training phase of this study. Without her help, this experiment could not have been completed. Thank you for working with me along the course of this experiment. I would also like to thank all of those who participated in my research in Japan and in Thailand over these three years. And I would not have been able to finish my degree if I had not had the help of my aunt, Associate Professor Dr. Suda Trakantalerngsak. She imparted crucial knowledge to me about statistics and more logical ways of conducting research. Yasuyo Wada and Kubota Ikumi who helped correct my Japanese and have showed great friendship along the course and until now. From Waranya Mahajuntakarn, who helped me from the beginning by allowing me to use her class for the experiment and who helped me again at the last stage with editing my references, to P.pao who is the most valuable friend I made here in Japan. Thank you for always looking after me and being by my side whenever I needed you and for all the nice advice you gave. To my friends (Li Kin, Pika, Bokuto, Kin Li, etc.) on campus who shared experiences and opinions about research and entertained me every Monday in class - I owe you all a great debt of gratitude.

Next, to my best friends (Kwang and Pat) in Thailand who are always by my side, whenever I need to talk or cry, and who have supported me throughout this testing chapter of my life. And finally, to my family, who have always been invaluable to me throughout this whole time. Even though they have been busy with their own lives they have never shut their door to me. There have been many things going on during the time I have been here in Japan, however, the links I have with these people has been shown to be made of strong stuff and I am forever grateful for what has been done for me. Thanks.

Table of Contents

List of Figures

List of Tables

Chapter 1 Introduction

L2 speech perception models such as the Speech Learning Model (SLM; Flege, 1995) hold that difficulties in the perception and production of non-native sounds arise from the influence of the L1 phonological structure on the perception of L2 sounds. The L1 acts as a filter through which all the sounds of the L2 are perceived and they are thus classified into L1 phonological categories which can result in a non-native perception and accented production of particular sounds in the L2. However, the SLM suggests that perceptual ability in categorising non-native speech sounds can be modified throughout the life span of an individual. Recently, a great number of cross-linguistic perception studies have employed perceptual training using a high-variability phonetic training method (hereafter referred to as “HVPT perceptual training”). Using this high-variability method, learners are exposed to a wide variety of tokens produced by multiple L2 native speakers in various phonetic contexts and positions with the aim to create robust novel phoneme categories by exposing L2 learners to acceptable variations within each category (Lively, Logan, & Pisoni, 1993; Logan, Lively, & Pisoni, 1991; Thomson, 2011). This HVPT perceptual training has shown to be the most effective tool in improving learners’ ability to accurately perceive L2 consonants (Bradlow, Akahane-Yamada, Pisoni, & Tohkura, 1999; Bradlow, Pisoni, Akahane-Yamada, & Tohkura, 1997; Iverson, Hazan, & Bannister, 2005; Lively et al., 1993; Lively, Pisoni, Yamada, & Tohkura, 1994; Logan et al., 1991; Lopez-Soto & Kewley-Port, 2009), vowels (Iverson, Pinet, & Evans, 2012; Iverson & Evans, 2009; Lambacher, Martens, Kakehi, Marasinghe, & Molholt, 2005; Nishi & Kewley-Port, 2007, 2008; Shin, 2014; Wong, 2013) and suprasegmentals such as pitch and tone (Hirata, 2004; Huensch & Tremblay, 2015; Perrachione, Lee, Ha, & Wong, 2011; Shin, 2014; Thomsom, 2011;

Wang, Spence, Jongman, & Serene, 1999; Waylang & Guion, 2004) and has also been shown to lead to an improvement in production (Bradlow, Akahane-Yamada, 1999; Bradlow, Pisoni et al., 1997; Hardison, 2003; Lambacher et al., 2005; Rochet, 1995). Furthermore, the improvement gained from this type of training has also been shown to have generalised to new tokens and new talkers (Bradlow, Pisoni et al., 1997; Yamada, Tohkura, Bradlow, & Pisoni, 1996), and these improvements were retained in the long-term memory of participants after the completion of their training (Wang et al., 1999; Wang & Munro, 2004). These findings are highlighted in order to show support for the view that HVPT perceptual training can lead to successful L2 contrast learning even for the most difficult cases and that perceptual ability in categorising non-native speech sounds can be attained by adult learners throughout their life span (Flege, 1995). In other words, it can be said that if learners are given sufficiently robust native speaker input, they will eventually be able to perceive and produce L2 contrasts more accurately.

However, most HVPT perceptual training studies undertaken thus far have used only an identification task in their training. Recently, growing attention has been focused on developing and advancing other types of training paradigms in order to explore the full potential of HVPT perceptual training such as the use of the visual, the audio and the audiovisual (Pereira & Hazan, 2013). Moreover, there is a rising undercurrent that emphasises that consciousness and self-awareness raising are important in Second Language Acquisition (e.g., Schmidt, 1990; Sharwood, 1981). Recently, a growing body of research is investigating the role of attention in learning by adopting a self-monitoring task to see whether this can improve learners' abilities in L2 speech learning (e.g., Couper, 2011; Hirano-Cook, 2011). Several studies have demonstrated evidence that using self-monitoring in L2 speech instruction can lead to better learning (e.g., Couper, 2003; Ingles, 2011; Nagamine, 2011; Sardegna, 2011). Considering studies such as these, it is worth exploring whether promoting self-awareness using self-monitoring would aid better speech and perceptual learning.

This study will investigate the effects of HVPT perceptual training on Thai learners of Japanese in the learning of Japanese fricative and affricate contrasts. The reason that these particular elements of Japanese speech have been chosen is because these contrasts have been proven to be difficult for Thais and are the area in which the most frequent errors are made by Thai learners both in perception and production, even after a long period of experience with the language (Kawano, 2014; Sukegawa, 1993; Trakantalerngsak, 2013; Yamakawa, Chisaki, & Usagawa, 2005). The HVPT perceptual training approach used in this study differs from previous training studies in that an additional self-monitoring task is applied, to see whether raising self-awareness by conducting a self-monitoring task enhances better speech learning. It should be noted that this study does not aim to test any theoretical hypotheses but will use them to understand more of the L2 speech learning process. Instead, the ultimate aim of this study is to enhance the efficacy of HVPT perceptual training for improving the perception and production ability of non-native contrasts.

This dissertation will consist of six chapters. Chapter Two will present a summary of background research relevant to the main aim of this dissertation, starting by introducing a brief overview of L2 speech learning, L2 speech perception models such as the Perceptual Assimilation Model and the Speech Learning Model, and the relationship between perception and production. Next, some previous L2 perceptual training studies as well as methodologies used in perceptual training will be reviewed. The chapter also includes a short brief of the Japanese and Thai fricative and affricate contrasts, followed by an examination of the difficulty of learning Japanese fricative and affricate contrasts for Thai learners. Following that, some Japanese pronunciation textbooks will be reviewed. Lastly, a research overview, questions and aims to be addressed will be given.

Chapter Three will describe a design summary of HVPT perceptual training and the detailed methodologies used to carry out the perceptual training in this study.

Chapter Four will present the results of the perceptual training as carried out in this experiment; first, the effect on perception; then, any generalisation of improvements to new words and talkers; and next, any long-term retention that remains; followed finally by an examination of any effect that the training might have had on production. The chapter will then explore the correlation between perception and production and ends with the results of a questionnaire which participants were asked to undertake.

Chapter Five will present the main findings of this research, aiming to answer the research questions and to address general discussions based on findings from Chapter Four, as well as discussing the pedagogical implications for Japanese pronunciation teaching and learning.

Chapter Six will draw conclusions from the various discussions and results obtained in this study. Limitations and future directions will also be summarised.

Chapter 2 Background

2.1 Introduction

In order to detail previous studies that have set the foundation for this study, this chapter will introduce an overview of previous research into L2 speech perception as well as prior L2 speech training studies followed by examining some methodological issues that need to be taken into consideration when training L2 speech perception. Next, a short description of the Thai and Japanese fricative and affricate consonant systems, as well as previous studies investigating the perception and production of Japanese fricative and affricate contrasts by Thai learners, will be presented. Following that, some Japanese pronunciation textbooks will be reviewed. Lastly, the research overview, aims and questions of this study will be presented in the final part of this chapter.

2.2 An overview of L2 speech perception

It is widely accepted that mastering the phonology of a foreign language is particularly difficult in adulthood even after years of exposure to the L2 - adult L2 learners continue to produce L2 sounds with a foreign accent (Flege, Yeni-Komshian, & Liu, 1999; Piske, Mackay, & Flege, 2001). Not only is the production of L2 sounds affected, its perception often remains non-native since adult L2 learners hear the L2 sounds through their native language filter (Flege, 1995; 2003). Lenneberg's Critical Period Hypothesis (1967) claims that learners must be exposed to L2 input before a critical age, which ends at about the age of puberty, to attain native-like language proficiency. According to this hypothesis, due to age-related decline in neural plasticity, any L2 acquisition which takes place after the age of puberty is fundamentally different in some respects and native-like attainment is impossible, resulting in foreign-accented speech. However, recently, many researchers have provided evidence that attempts to refute the Critical Period Hypothesis. Strange (1995) states that "perceptual difficulties are not due to a loss of sensory capabilities, but rather reflect perceptual attunement to phonetic information that is phonologically relevant in their native language" (p.79). Moreover, there exists evidence that adults can achieve near native-like degrees of pronunciation of a language learned after the putative critical period (Birdsong, 1992; Bongaerts, Van Summeren, Planken, & Schils, 1997; Moyer, 1999). In addition, more recent research also shows that L2 speech training can result in learners exhibiting more native-like performance (Huensch & Tremblay, 2015; Lengeris, 2008; Wang, Jongman, & Sereno, 2003). This indicates that adult learners' neural pathways are malleable when provided with sufficient input and the capability of making new L2 category formations remains intact over the life span of an individual (Flege, 1995). It must be noted, however that there are factors which may still limit attainment. These are through things such as the relationship between learners' L1 and L2 phonological systems, the age of learning and the amount of experience that learners have with their L2 (Flege, Frieda, & Nozawa, 1997; Piske et al, 2001) as well as individual attitudes and motivation (Moyer, 1999).

2.2.1 L2 speech perception models

There are three main influential theoretical frameworks that have largely influenced research in L2 speech perception: 1) The Perceptual Assimilation Model (PAM) (Best, 1995), 2) the Speech Learning Model (SLM) (Flege 1995, 2003), 3) the Native Language Magnet (NLM) model (Kuhl & Iverson, 1995). A broad summary of these models will be briefly given below.

The Perceptual Assimilation Model (PAM)

The Perceptual Assimilation Model (PAM) developed by Best (1995) makes predictions on the basis of a comparison of L1 and L2 contrasts. According to Derwing & Munro (2015), PAM was not originally developed as a model of L2 learning. It was rather intended to account for the ways in which native monolingual listeners would perceive non-native speech sounds having had no prior experience with the non-native language. PAM predicts assimilation at the initial stage but does not describe language development. PAM posits that most non-native contrasts perceptually are perceived in terms of their gestural similarity to similar ones from the native language. According to PAM, discrimination of a non-native contrast depends on how each member of the contrast is assimilated to native categories. There are several possible assimilation types and for each assimilation type there is a specific discrimination prediction. First, two-category assimilation (TC) where each L2 contrast assimilates to a different native category, allowing easy discrimination of the L2 contrasts; second, single-category assimilation (SC) where two non-native contrasts assimilate to a single L1 category and discrimination is predicted to be poor; third, category-goodness assimilation (CG) wherein two non-native contrasts assimilate to the same native category but with one member being a better match to that category than the other and discrimination is predicted to be moderate to very good; forth, uncategorised-categorised assimilation (UC) wherein one non-native phone is categorised while the other is uncategorised and discrimination is predicted to be very good; fifth, uncategorised-uncategorised assimilation (UU) where two non-native phones are uncategorised and discrimination is predicted to vary from fair to good according to how similar these sounds are to each other and to native categories; last, non-assimilable (NA) where two non-native phones are perceived as non-speech sounds and discrimination is predicted to be very good (Best, 1995, p. 194-195).

The Speech Learning Model (SLM)

The Speech Learning Model (SLM) developed by Flege and colleagues (Flege,1992, 1995, 2002; Flege, Yeni-Komshian, & Liu, 1999) is aimed at explaining L2 learning, particularly pertaining to experienced L2 learners. L2 categories are initially classified in terms of L1 categories based on the perceived phonetic similarity or dissimilarity between the L1 and the L2 categories, the so-called the "*equivalence classification*". SLM distinguishes between "new" and "similar" sounds and claims that the greatest difficulty for L2 learners can be expected in the category of "similar" sounds, rather than "new" sounds. L2 learners will not create a new phonetic category for sounds that are similar to L1 phonemes, while for L2 sounds that do not exhibit a categorical overlap with L1 phonemes, a new category is created. SLM argues that many of the speech production errors in non-native speech arise from an incorrect perceptual representation of the properties that specify L2 sounds. However, the mechanisms of the native language sound system remain intact over a life span, indicating that it is possible for adult L2 learners to establish new non-native phonetic categories. In order to maximise attainment, a greater quantity and a higher quality of input are needed (Flege, 1995; Flege, Yeni-Komshian, & Liu, 1999).

Native Language Magnet (NLM)

The Native Language Magnet Model (NLM, Kuhl & Iverson, 1995) claims that perception of L2 sounds occurs within the L1 frame so that L1 speech sounds act as magnets attracting perceptually similar L2 sounds and blocking category formation. In other words, L2 acquisition constraints are caused by prior linguistic experience not by lack of brain plasticity. This model argues that linguistic experience alters phonetic perception such that perceived sounds are distorted due to a "magnet effect". According to this model, there is the tendency that L2 sounds similar to L1 sounds are interpreted as bad exemplars of L1 sounds and are thus grafted on to mental representations of L1 sounds. Hence, the NLM posits that L2 sounds which are similar to L1 categories will be more difficult to discriminate accurately because they have associated target L2 sounds with their own L1 archetype. On the contrary, L2 sounds that are not similar to L1 prototypes are not affected by this magnet effect, thus making discrimination easier.

Taken together, these three models capture that the L2 learners' difficulties in L2 perception and production are caused by non-native perception and that the native language of learners plays a role in L2 speech acquisition.

2.2.2 The relationship between perception and production

The Speech Learning Model hypothesises that there exists a moderate correlation between L2 perception and production accuracies and suggests that the degree of L2 perception accuracy influences how accurately L2 segments are produced. Hence, an improvement in perception might help to improve production (Flege, 1995, 2003; Rochet, 1995). Several empirical studies have been shown to support such a prediction (e.g., Flege, Frieda, & Nozawa, 1997; Flege, Mackay, & Meador, 1999; Hattori, 2009; Wang et al., 2003). For example, Flege, Mackay, & Meador (1999) investigated whether English vowel production and discrimination by Italian speakers are related. The results demonstrated that the production accuracy of the Italians as judged by English speakers and their discrimination accuracy of target vowels were related. Furthermore, other studies that investigated both the perception and production of L2 contrasts also indicated that L2 sounds that are perceptually difficult to acquire also pose production difficulties (e.g., Rochet, 1995; Wang & Munro, 2004).

Many studies claim that perception precedes or is a pre-requisite for production (e.g., Flege, 1991; Llisterri, 1995). Despite the existence of empirical evidence supporting these studies, other studies have suggested that L2 production accuracy might conversely precede L2 perception. Goto

(1971) and Sheldon & Strange (1982) found that Japanese learners were more successful at producing the English /r/ and /l/ accurately than perceiving those contrasts. However, Thomson (2008) states that "the ability might have resulted from prior pronunciation instruction, coupled with a reading task where speakers could more easily attend to articulation" (p. 134). This instruction possibly enabled them to produce the English /r/ and /l/ contrasts more accurately than to perceive them. In addition to this effect, Hattori (2010, p. 80) states that these two studies did not aim to directly examine the correlations between L2 perception and L2 production, therefore it is difficult to draw a firm conclusion regarding the relationship between the two modalities from Goto (1971) and Sheldon & Strange (1982). More recently, Zampini (1998) investigated English-speaking Spanish learners' perception and production of the Spanish /p/-/b/ contrast. The results showed that learners' perceptual boundaries of the VOT (Voice Onset Time) of /p/ and /b/ did not correlate with the VOT they produced for those phones in most target items. However, recently, the results of L2 speech perceptual training studies provide strong evidence for a link between these two modalities. For example, Bradlow et al. (1997) trained Japanese learners in the English /r/ and /l/ contrasts. The results of Bradlow's study showed that training was not only effective in perceptual learning but the improvement was also transferred to production. Wang (2003) also reports a similar result in that perceptual learning gained from perceptual training can be transferred to improvements in production. Results from these training studies have provided evidence that L2 speech perception and production are closely linked and strongly supports the finding that an improvement in perception might help to improve production.

Although there is strong evidence that there is a close relationship between perception and production, empirical studies of these links in L2 speech learning are still very limited in nature and the results investigating these links have always been inconsistent. A number of laboratory-based training studies have suggested that intensive perception training clearly helped participants to perceive L2 contrasts better as well as improving their production of target sounds (e.g., Bradlow, Akahane-Yamada et al., 1999; Bradlow, Pisoni et al., 1997; Rochet, 1995). However, more research is clearly needed to understand the influence of training on the perception and production of learners.

2.3 Previous perceptual training studies

Next, an overview of earlier research on L2 perceptual training and particular methodological issues will be presented in this subsection focusing mainly on L2 segmental training studies.

2.3.1 Perceptual training using a high-variability method

The training conducted by Strange & Dittmann[1] (1984) shows the ineffectiveness of their training in that the modification of perception is slow and effortful possibly because the training used stimuli produced by only a single speaker and in a single context. They suggested that a wider range of stimuli with different phonetic contrasts and different stimulus materials should be covered in future training. In addition, Jamieson & Morosan (1989) proposed that training should also use a wide range of sounds, selected to represent the range of acoustic variability which occurs in the categories being learned. Logan et al. (1991) and Lively et al. (1993) were the first to show that increasing the variability of the input - the so-called “high-variability phonetic training method (later called “HVPT perceptual training”)” - results in greater and more generalisable gains in L2 speech perception. This HVPT perceptual training provides a wide range of variations in training stimuli in terms of the numbers of native speakers’ input, multiple phonetic contexts and

[1] Strange & Dittmann (1984) trained native Japanese speakers learning English using a discrimination task with immediate feedback on English /r/ and /l/.

multiple word positions (initial singleton, initial consonant cluster, intervocalic, word-final consonant clusters, and final singleton positions) of the L2 contrasts. Immediate feedback to promote the formation of robust new phonetic categories was also given during the training (Jamieson & Mososan, 1986; Lively et al., 1993). According to Thomson (2011), “exposing learners to high variability input will allow them to begin recognizing the difference between meaningful phonetic cues associated with particular sound categories, and cues that are irrelevant to category identity. It is believed that this knowledge will then transfer to the perception of novel speech tokens” (p. 749). HVPT perceptual training has become more widely used and has been proven to promote robust learning of L2 perceptual categories in both perception and production both at segmental (Bradlow, Akahane-Yamada et al., 1999; Bradlow, Pisoni et al., 1997; Iverson & Evans, 2009; Wang, 2002; Wang & Munro, 2004) and suprasegmental levels (Shin, 2014; Thomson, 2011; Wang et al., 1999; Wayland & Guion, 2004). Moreover, HVPT perceptual training yields generalisations to new untrained tokens and untrained talkers and also demonstrated long-term retention for a period of three months or even more (Bradlow et al., 1997; Lively et al., 1994). One reason why variability in training materials may strengthen the process of category formation is that it may help learners to develop abstract representations that can accommodate a wider range of examples (Logan et al., 1991). According to this evidence from previous studies, if adult learners are given sufficient and robust native speaker input, eventually they will be able to perceive and produce L2 contrasts more accurately.

Next, the effects of HVPT perceptual training and methodologies used in previous HVPT perceptual training will be discussed in detail below.

2.3.1.1 The effects of HVPT perceptual training on *perception*

According to Hazan, Sennema, Iba, & Faulkner (2005), most HVPT perceptual training studies have mainly investigated L2 consonants and vowels, with a strong focus on the English /l/-/r/ for Japanese learners (e.g., Bradlow, Akahane-Yamada et al., 1999; Bradlow, Pisoni et al., 1997; Hardison, 2003; Iverson et al., 2005; Lively et al., 1993). Previous findings clearly show that trained L2 learners benefit from perceptual training with improvements in identification ability increasing by 10% to 20% (Bradlow et al., 1997; Lively, Logan et al., 1993; Lively, Pisoni et al., 1994; Lopez-Soto & Kewley-Port, 2009; Wong, 2013).

Logan et al. (1991) first developed a method of HVPT perceptual training which emphasises the variability of the training stimuli in terms of speaker differences and phonetic environments to train six native speakers of Japanese to better identify the English /r/-/l/ contrasts. Training consisted of 15 sessions using a two-alternative forced-choice identification task with immediate feedback. The training stimuli were 68 minimal pairs that contrasted /r/ and /l/ in multiple word positions produced by five talkers. A pre-test, a post-test and two types of generalisation test - (1) using new words spoken by an old (familiar) talker and (2) using new words spoken by a new talker - were conducted. Results showed a significant improvement of an 8% increase from pre-test to post-test as well as transference of perceptual learning in both tests of generalisation, indicating that HVPT perceptual training can be used to modify Japanese learners' perception of the English /r/ and /l/.

Next, Pruitt, Jenkins, & Strange (2006) trained 20 American English speakers and 20 Japanese to perceive the Hindi dental and retroflex stops. Participants received 12 training sessions which lasted about 30 to 50 minutes each session. They were given a total of 8400 tokens produced by six talkers with immediate feedback. The results demonstrated an overall improvement from pre-test to post-test with an increase of 20% in identification accuracy in both language groups. These results suggested that new phoneme contrasts can be learned through HVPT perceptual training and there is a considerable capacity to learn new speech contrasts even in adulthood.

To summarise, perceptual training using a high-variability method has been proven to be effective in improving the identification of L2 contrasts. Results found in previous studies support the claim that perceptual learning can occur and that speech perception of L2 adult learners can be modified through laboratory training (Bradlow et al., 1997; Iverson & Evans, 2009; Iverson & Hazan, 2005; Lively et al., 1993; Logan et al 1991; Nishi & Kewley-Port, 2008; Pruitt et al, 2006; Rochet, 1995).

2.3.1.2 The effects of HVPT perceptual training on *production*

Previous studies state that L2 learners' speech production errors occur due to inaccurate perception (Flege, 1995, 2003; Rochet, 1995). Thus, an improvement in perception might help to improve production. Recently, improvements in perception due to perceptual training have been shown to also lead to an improvement in production even without any specific production training (Bradlow et al., 1997; Hazan et al., 2005; Lambacher et al., 2005; Rochet, 1995). However, there are still only a small number of L2 training studies that have investigated the transfer of perceptual learning to production (Bradlow et al.1997; Rochet, 1995).

Bradlow et al. (1997) were the first to investigate the effects of HVPT perceptual training on production. Eleven native speakers of Japanese were trained to perceive the English /r/-/l/ contrasts in 45 sessions. The results showed that improvements transferred to production without explicit instruction on pronunciation, as judged by English native speakers. However, improvements on perception due to perceptual training have been shown to only partially transfer to production. There was a 7% improvement in production of words found in Bradlow et al. (1997). A later study (Bradlow et al., 1999) replicated the results mentioned above and also found an improvement in production.

Next, after six 30-minute sessions of training in which 12 native Mandarin learners were trained to perceive the synthetic French stops (/pu/-/bu/), Rochet (1995) reported that the Mandarin learners shifted their VOT to a French native-like VOT when they listened to the synthetic stimuli. The results also demonstrated that the improvement was transferred to the production domain in that they produced more native-like VOT in both the target sounds.

Wang et al. (2003) also demonstrated that HVPT perceptual training can improve production ability by an increase as large 18% points even though there was no specific production training.

On the contrary, there have been some studies which have reported that improvements in perception do not necessarily result in an improvement in production. Wang (2002) demonstrated that training Mandarin and Cantonese learners to perceive English vowels failed to show such a transfer to production.

Taken together, there is some evidence showing that an improvement in perception will help to improve production despite no explicit production training, however, these results are yet to be demonstrated consistently.

2.3.1.3 The generalisation of perceptual learning

Logan & Pruitt (1995) define the generalisation of learning as the ability to transfer knowledge gained to multiple dimensions such as new words, new talkers, or new contrasts and suggest that if generalisation occurs, it is to be believed that robust learning has also occurred. Several previous training studies have reported a generalisation of perceptual learning to new words and new talkers (Bradlow et al., 1997; Sadakata & McQueen, 2013; Thomson, 2012; Wang et al., 1999; Wang & Munro, 2004; Yamada et al., 1996).

Eight American native speakers were trained by Wang et al. (1999) in eight sessions to perceive Mandarin tones. The results from this study showed that training had made a robust effect through a substantial improvement in perception. Moreover, the improvement gained in the training was also generalised to new words produced by a familiar talker used in the training, with an 18% increase in perceptual accuracy and a 25% increase in the perceptual accuracy of new words produced by a hitherto unfamiliar talker.

Furthermore, Yamada et al. (1996) and Bradlow et al. (1997) also reported a generalisation of /r-l/ perceptual learning to new words produced by a familiar talker and to new words produced by an unfamiliar talker.

2.3.1.4 The long-term effects of HVPT perceptual training

As Thomson & Derwing (2014) suggest, "the assessment should also include a delayed post-test to determine whether the intervention had a lasting effect" (p. 2). Some previous studies showed that a group of trained participants maintained their improved level of identification ability three months after (e.g., Bradlow et al., 1999; Iverson & Evans, 2009; Lively, Logan et al., 1993; Lively, Pisoni et al., 1994; Wang, 2002; Wang & Munro, 2004; Yamada et al., 1996) or even six months after their perceptual training had been completed (e.g., Lively et al., 1994; Wang, 1999; Yamada et al., 1996).

Lively et al. (1994) tested the long-term retention of perceptual learning by conducting two retention tests; a test three months after and a test six-months after training was completed. The results showed that a significant improvement was still evident three months after training and the effect still partially persisted at the time of the six-month test.

Moreover, results from Bradlow et al. (1999) also demonstrated that three months after the completion of perceptual training, trained Japanese learners maintained their improved levels of identification performance. In addition, a three-month follow-up speech production test also demonstrated that trained participants retained their long-term improvements in the general quality and overall intelligibility of their English /r/-/l/ productions.

To summarise, perceptual training using a high-variability method has proven to be effective in improving the identification of sounds and also has benefit for participants' production of L2 contrasts. Results found in previous studies support the claim that perceptual learning can occur and L2 adult learners speech perception can be modified through laboratory training. An improvement in perception can help to improve production despite no explicit production training. Furthermore, the improvement gained from the training has been generalised to new tokens and new talkers, and was retained in the participants' long-term memory.

2.3.1.5 Methodologies used in HVPT perceptual training

The use of appropriate methodologies needs to be carefully considered when using perceptual training, since the choice of appropriate tasks to meet the goals of the training can be complicated by interactions among the stimulus variables, task variables and subject characteristics (Wang, 2002, p. 23). The following section presents some elements of training methodologies used in previous L2 training studies:

Input

In HVPT perceptual training studies undertaken thus far a wide range of variations in training stimuli has been used. These have taken the shape of employing a great number of native speaker inputs, multiple phonetic contexts and multiple syllable positions of L2 contrasts (initial singleton, initial consonant cluster, intervocalic, word-final consonant clusters, and final singleton positions) (Lively et al., 1993; Logan et al., 1991; Thomson, 2011).

Task types

Previous L2 training studies have used both discrimination tasks (in which trainees must decide whether pairs of L2 segments are the same or different) and identification tasks (in which trainees must identify the speech sound they have actually heard). Identification tasks have become more widely used in recent studies since they encourage learners to classify contrasts into categories and facilitate the establishment of new phonetic categories (Bradlow et al., 1997; Lambacher et al., 2005; Logan et al., 1991; Logan & Pruitt, 1995; Rochet, 1995). In addition, the efficacy of identification tasks has been found to be superior to that of discrimination tasks. Evidence from previous studies conducted by Jamieson & Morosan (1986, 1989) and Logan & Pruitt (1995) indicates that a discrimination task is likely to have an undesired effect on phonetic training since the task focuses the participants' attention on within-category variation, rather than on between-category variation, which is necessary for the formation of new phonetic categories.

Training materials

Synthetic and natural stimuli have both been commonly used in prior training studies. Natural stimuli make it possible to represent a suitable range of difference, since the ultimate goal of perceptual learning is to enable learners to perceive target contrasts in real speech (Wang, 2002; Wang, 2003). The use of synthetic stimuli in training has shown to improve the perception of novel contrasts only for the same type of input but has failed to generalise to new natural tokens (Strange & Dittmann, 1984 as cited in Pereira, 2013, p. 30). This clearly demonstrates that natural input is preferable for improving natural L2 speech perception. However, selecting between the two types of stimuli also depends on the training task that is being used. According to Logan & Pruitt (1995), a fading task maybe more easily employed using synthetic stimuli than natural stimuli. This is because using synthetic stimuli allows the simple manipulation of key acoustic differences such as exaggerating or shrinking differences along a continuum (Wang, 2002, p. 26).

Training sessions

According to Wang (2002, p. 27), there is no fixed standard for the number of sessions or hours required for a period of training. Some studies have conducted only a short-term training period of one session (Pisoni, Aslin, Perey, & Hennessy, 1982), whereas other studies have conducted training over long-term periods (e.g., Rochet, 1995; Wang, 2002; Yamada, 1993), hence it is difficult to draw a firm conclusion about what constitutes a suitable amount of training sessions. Logan & Pruitt (1995) state that long-term training that occurs over several days or weeks, or ranges from 6 to 45 sessions is generally more characteristic of cross-language training studies but a modal number seems to be around 15 sessions which are spread over three weeks. Moreover, according to Iverson et al. (2012), five to ten sessions of HVPT perceptual training tends to show the most improvement in identification accuracy. Logan & Pruitt (1995) state that "the reason that the duration of training does not typically exceed 15 sessions is more likely because of practical considerations, such as problems in retaining subjects and providing the infrastructure necessary for training over an extended period of time" (p. 365). Also, it has been found that an increment in performance would occur early in training and that subsequent improvements would be smaller as a function of additional training. Several studies have found that learners' performance improved most during the first 10 training sessions (Lively et al., 1993; Logan et al., 1991; Yamada, 1993). However, Thomson & Derwing (2014, p. 11) state that finding an appropriate length of training is related to the number of features targeted or the scope of instruction and that global improvement in intelligibility and comprehensibility requires weeks or even months of instruction. Hence, it is difficult to draw a firm standard about what constitutes the right amount of training sessions.

Feedback

According to Logan & Pruitt (1995, p. 362), feedback is an important element of training since it enables learners to determine whether what they are doing is appropriate or not. If the feedback indicates that the learner is responding adequately then the learner continues to respond in the same way as before. If the feedback indicates that the learner is not responding appropriately, then the learner understands that they must modify their response. The importance of feedback appears even more obvious for learning L2 pronunciation, because many errors produced by L2 learners can be attributed to unconscious interference phenomena from the L1 built-in phonological representations (Flege, 1995). Immediate feedback is the most frequently used method of providing feedback in perceptual training studies. This immediate feedback allows learners to rehear the trial stimulus that was incorrectly identified as well as giving an opportunity to hear that stimulus associated with its correct category label (Logan & Pruitt, 1995, p. 363).

Assessment

According to Logan & Pruitt (1995) and Wong (2013), implementing both a pre-test and a post-test is the most well-established evaluation procedure to assess change in participants' performance due to training. Several studies have also included a control group as a comparison to see the effect of training using more detailed comparisons. When a study also investigates the transfer of perceptual learning to the production domain, using acoustic measurements or native speakers' judgements are also commonly adopted. Additionally, according to Wang (2002), it is common to also include generalisation tests to see whether perceptual learning has transferred to sounds in new phonetic contexts or new talkers who were not used during the training period. Some studies have also assessed long-term retention by conducting a delayed post-test after the completion of training (e.g., Lively et al., 1994; Wang, 1999; Yamada et al., 1996).

The following table presents a summary of previous HVPT perceptual training studies:

Table 2.1 Overview of HVPT perceptual training.

Study	Research design	Training target	Training duration	Main findings
Logan et al. (1991)	- Six Japanese speakers. - Pre/post-test of perception. - No control group.	- Natural stimuli of English /r/-/l/ produced by six talkers.	15 sessions (40 mins) of 272 trials over three weeks.	- Improvement in perception (+8%). - Transferred to new words and talkers.
Lively et al. (1994)	- Nineteen Japanese speakers. - Pre/post-test of perception. - Three-month and six-month retention test.	- English /r/-/l/.	15 sessions over three weeks.	- Immediate improvement in perception. - Improvement was retained at the three-month stage but had gradually reduced by the 6-month test.
Rochet (1995)	- 12 Mandarin speakers - Pre/post-test of perception and production. - No control group.	- Synthetic stimuli of voiced and voiceless contrasts of French stops.	Six sessions (30 mins).	- Improvement in perception. - Perceptual learning was transferred to production.
Yamada et al. (1996)	- 23 Japanese speakers (Trained: 11 and control: 12). - Pre/post-test of perception and production. - Generalisation test. - Three-month and six-month retention test.	- English /r/-/l/ produced by five talkers.	45 sessions (30-40 mins) over two weeks.	- Significant improvement in perception. - The improvement generalised to new talkers and words. - Perceptual learning was transferred to production. - The production improvement was maintained even after three or six months.

Study	Research design	Training target	Training duration	Main findings
Bradlow et al. (1997)	- 11 Japanese speakers. - Pre/post-test of perception and production. - Generalisation test. - Three-month delayed post-test.	- 68 English / r/-/l/minimal pairs by five talkers.	45 sessions (20-30 mins) over three to four weeks.	- Improvement in perception (+16%) and production. - Perceptual learning was generalised to new words and talkers. - The improvement decreased after three months but was comparable to post-test.
Wang et al. (1999)	- Eight American speakers. - Pre/post-test. - Generalisation test. - Six-month retention test.	- Mandarin tones. - Natural stimuli by four talkers.	Eight sessions (40 mins) over two weeks.	- 21% improvement in perception. - Improvement generalised to new words and talkers. - Improvement was retained after six months.
Wang & Munro (2004)	- 21 Mandarin and Cantonese speakers(Trained : 16 and control: 5). - Pre/post-test. - Generalisation-test. - Three-month retention test. - Participants could control their own training schedules.	- Three English vowels. - Synthetic and natural stimuli.	Two months (two to three times per week).	- Substantial improvement in perceptual performance was observed. - Transferred to new contexts. - Improvement was maintained in three-month retention test.
Lambacher et al. (2005)	- 34 Japanese speakers.	- American mid/low vowels.	Six sessions (20 mins) over six weeks.	- 16% improvement in perception. - Production performance also improved.

Study	Research design	Training target	Training duration	Main findings
Nobre-Oliveira (2007)	- 36 Brazilian learners of English (Trained: 29 and control: 7). - 15 trained with natural stimuli and 14 trained with synthetic stimuli.	- Six English vowels. - Natural and synthetic stimuli.	Three sessions.	- Both group significantly improved but there was more improvement in the group who received synthetic stimuli. - Improvement generalised to new words and syllable structures. - A transfer to production was partially observed. - Improvement was maintained one month after the training.
Iverson et al. (2012)	- 36 French speakers. - Pre/post-test of perception and production. - No control group.	- English vowels produced by eight British speakers. - Natural stimuli.	Eight sessions.	- An improvement in perception and production was observed.
Thomson (2012)	- 26 Mandarin speakers. - Pre/post-test. - Generalisation-test.	10 English vowels by 20 talkers.	Eight sessions over three weeks.	- Learners improved their ability to identify the target contrasts. - An improvement was maintained a month after training.
Sadakata & McQueen (2013)	- 30 Dutch speakers. - Every session following training, identification tests were conducted. - Generalisation test. - Compare HVPT and LVPT.	- Japanese geminate-singleton fricative contrast. - Natural and synthesized stimuli.	Five sessions.	- HVPT led to superior performance in identification tests and better gegenralisation of learning than LVPT training.

Other training paradigms

Although HVPT perceptual training has been proven to be effective in L2 speech learning, some studies have found that input variability makes learning more difficult (e.g., Wade, Jongman, & Sereno, 2007) or showed no learning outcome (e.g., Iverson et al., 2005). The ultimate goal of training is to produce as much improvement in performance as possible, hence, it is necessary to look for some modifications of HVPT perceptual training or to develop other effective training paradigms (Iverson et al., 2005). Huensch (2013, p. 179) also states that one issue with perceptual training relates to the interest level of users. Participants in her research often commented on the "less-than-exciting" or "boring" nature of the perceptual training tasks. She suggests overcoming this issue by using methods other than forced-choice identification tasks. Moreover, recent perceptual training studies have developed and advanced different training paradigms with an aim to present new and encouraging tools that garner more beneficial outcomes for L2 learners. Samples of those modified paradigms are described below:

- A combination of perceptual training and production training (Aliaga-Garcia & Mora, 2009; Delvaux, Huet, Piccaluga, & Harmegnies, 2013; Wong, 2013). Aliaga-Garcia & Mora (2009) concluded that the approach of combining both production and perception training is largely effective in improving L2 production accuracy.
- Some studies compared the use of auditory input versus audiovisual input on perceptual training, the results showed that the audiovisual group gained greater improvements than the auditory group (Hardison, 2003; Hazan et al. 2005; Ortega-Llebaria, Faulkner, & Hazan, 2001).

Taken together, results from the studies described above suggest that adult L2 learners possibly gain better learning if the perceptual training paradigm is well modified to be more suitable for the target learners.

2.4 An element of self-monitoring

According to Ingles (2011, p. 6), “self-monitoring” is defined as training in which learners are able to listen critically to their own L2 productions and this listening is then used to help learners focus their attention on the target pronunciation features.

Some research suggests that L2 errors resulting in a foreign accent occur because learners are not able to detect the difference between their own productions and those of native speakers of the target language (Colantoni, Steele, & Escudero, 2015; Dlaska & Krekeler, 2008). There is thus a rising current that proposes that consciousness-, noticing- and self-awareness-raising are important in Second Language Acquisition (e.g., Schmidt, 1990, 2001; Sharwood, 1981). Recently, a growing body of L2 speech studies are investigating the role of attention, examining the extent to which self-monitoring can improve learners’ abilities in speech learning (e.g., Couper, 2011; Ingles, 2011; Nagamine, 2011; Sardegna, 2011). Some of the L2 studies that have exploited the use of self-awareness instruction will be introduced below.

Sardegna (2011) used the Covert Rehearsal Model[2] (CRM) developed by Dickerson (1989, 1994, 2000) using strategies such as “speech monitoring (evaluating the accuracy and fluency of production)” and “comparing the performance with other models (compare their own production to a recording of the original text)” to instruct 38 learners of English from differing nations to practice their ability to link sounds within words for 50 minutes three times a week for four months. Results revealed that all participants showed short-term and long-term improvements in linking, providing evidence in support of the

[2] The Covert Rehearsal Model (Dickerson, 1989, 1994, 2000) is an approach designed to teach learners how to self-direct their learning, or how to become their own teachers. Teachers should teach students to empower students with strategic skills to self-monitor and self-correct their own mistakes. CRM comprises six key components, namely: 1. Privacy 2. Oral practice 3. Speech monitoring 4. Comparing performance with other models 5. Changing performance to match the models 6. Practicing the changed performance until fluent.

effectiveness of pronunciation instruction and supporting Dickerson's claim in favour of empowering students with explicit pronunciation rules that learners can use to self-correct and self-monitor their own speech production.

Smith & Beckmann (2005) have also examined the technique, taking learners through a series of steps including listening to and analysing their own speech according to specific phonetic features, and then comparing their pronunciation to that of a model using their own judgements as a guide - this is labelled "noticing the gap[3]". According to the study, this method proved to be an effective technique for teaching pronunciation to advanced learners of English and concluded that "noticing the gap" as part of self-analysis facilitates improvement in pronunciation.

Couper (2003) created a pronunciation syllabus aimed at raising each individual learners' awareness of their difficulties with pronunciation and of the main features of spoken English. One of the tasks featured in the pronunciation syllabus is having students develop their ability to monitor their own pronunciation. The participants were from different L1 background such as Chinese, Korean, Japanese and Arabic. Participants were asked to record themselves after listening to a model and then listen again and compare their recording with that of the model. This task encouraged an increased awareness of specific problems and emphasised the value of improving the ability to self-monitor. The results of this study showed that all participants made clear gains in pronunciation accuracy and, hence, these results support the value of explicit attention to pronunciation in the classroom.

Apart from the studies described above, other prior studies have also shown that noticing facilitates improvement in pronunciation (e.g., Derwing & Munro, 2005; Smith & Beckmann, 2005). However, the use of a self-

[3] "Noticing the gap" is a noticing technique first coined by Thornbury (1997). In this study this "noticing the gap" is referred to the action that learners compare their own pronunciation with the model pronunciation, using their analyses as a guide (Smith & Beckmann, 2005; Yule, Hoffman, & Damico, 1987).

monitoring strategy in L2 training studies has still received limited attention. Next, training studies which have used self-monitoring will be presented.

Borden (1983) trained Korean speakers in the perception and production of the English /r-l/ contrasts. Apart from training that included an identification task, he also conducted an additional self-perception task that consisted of asking subjects to listen to a taped model example and the subjects' own previously recorded sound and immediately ask them to judge whether they agreed that their production matched that of the taped model or not. The results of the study suggest that speakers make better self-judgements when they monitor themselves, and improvements in self-perception may be prerequisite to improvements in speech production.

Next, Hirano-Cook (2011) conducted six sessions of Japanese pitch accent training on 31 American learners of Japanese. This training was designed with the aim of raising American learners' awareness of Japanese pitch accents and improving their self-monitoring. The procedures employed in this study were Peer Learning[4] and instructor feedback on participants' productions. While Peer Learning was used in an attempt to practise production, the reason instructor feedback was used was to improve learners' self-monitoring skill. Regarding the instructor feedback, participants were asked to pronounce the target words, then the instructor provided a correct model. Participants were then asked to explain verbally how their pronunciation differed from the model sounds. The results showed that the participants were able to significantly improve their ability to perceive and produce Japanese pitch accents through the training.

Taken together, a review of the studies above implies that a more meaningful input of the target, brought about by raising learners' attention through a method such as self-monitoring, is likely to lead to learning (Wong, 2013, p. 33). Thus, it is clearly worth exploring whether exploiting self-

---

[4] According to Ikeda & Tateoka (2007), Peer Learning is a method through which learners cooperate with classmates to make significant contributions to what they learn.

monitoring in HVPT perceptual training would aid better speech perceptual learning.

2.5 Individual learning differences

Most L2 training studies concentrate on group performance and overall training effects without investigating the individual differences that occur into learning. However, recent studies suggest that taking individual differences into consideration is an important part of determining the success of perceptual training (e.g., Hattori, 2009; Lengeris & Hazan, 2010). Several L2 speech training studies have shown that the performance of individuals differs considerably. While some learners are good at identifying non-native contrasts, others perform less successfully, even when training and feedback are provided (Bradlow, 2008; Bradlow et al., 1997; Hazan et al., 2005; Lengeris, 2009; Pruitt, Strange, Polka, & Aguilar, 1990).

Bradlow et al. (1997) constituted a homogeneous L2 group in terms of L1 background, age and experience with written and spoken English. However, the pre-test performance in both perception and production varied considerably across individuals. Moreover, although participants improved significantly in both domains after undertaking perceptual training, the improvement in perception and production did not significantly correlate with individual participants. Hazan et al. (2005) also report a similarly uneven outcome. They state that the differences in improvement of /r/-/l/ identification scores originating from perceptual training ranged from -5% to 48% across individuals. Similarly, the differences in improvement in production ranged from -11% to 20% across individuals. Hanulikova, Deidu, Fang, Basnakova, & Huettig (2012) have looked into individual-specific factors that may contribute to differences in perceptual learning. They demonstrated how individual differences in pre-training measurements of speech and language-learning aptitudes interact with the design of training paradigms to the benefit or detriment of learning outcomes. In particular, they found that high variability perceptual training typically promoted learning among individuals with strong perceptual abilities. Learners with weaker perceptual abilities were impaired by high variability training relative to a low variability condition.

To summarise, there is strong evidence that large individual differences in learning gains can be observed under laboratory training conditions. Among learners with similar language backgrounds, some are able to perceive L2 sounds in adulthood, or to acquire them quickly through laboratory training, while for others, learning to hear L2 sounds is more slow and effortful (Golestani & Zatorre, 2009). Previous research has identified several factors which may contribute to these variations in success in the acquisition of L2 speech, such as motivation, quality of training, amount of L2 input, socio-psychological factors, personality, general intelligence, and age of acquisition (Birdsong, 2006; Moyer, 1999; 2004). However, even at the onset of L2 speech acquisition, when several factors are controlled, large individual differences are still often observed. The causes of these individual differences are still unclear (Ellis, 2004). Moreover, according to Lengeris (2009, p. 7-16), factors that are related to the learners' backgrounds and that may affect success in acquiring a second language can be assigned to three broad categories: 1) factors concerned with the learners' first or second language experience such as age of L2 learning, relationship between the L1 and L2 sound inventories, length of residence in an L2-speaking environment and amount of ongoing L1 use; 2) factors concerned with the learner's language aptitude such as phonological memory and working/short-term memory; 3) factors concerned with the learner's attitudes towards language learning such as motivation.

As has been explained, there are many factors that influence L2 speech learning outcomes, it is, however, unclear as to what specific factors determine these individual differences. Hence, taking individual difference factors in speech learning into consideration and seeing how these factors relate to the degree of improvement may allow for the development of training paradigms that will maximally benefit all learners.

2.6 Thai learners' perception and production learning of Japanese fricative and affricate contrasts

This section will briefly describe the phonetic features of fricatives and affricates belonging to both Thai and Japanese, as well as reviewing previous studies into the perception and production of Japanese fricatives and affricates by Thai learners.

2.6.1 A comparison between Japanese and Thai fricative and affricate inventories

In order to provide a clear comparison between the two languages, Japanese and Thai fricatives and affricates are shown in Tables 2.2. and 2.3.

Table 2.2 Japanese fricative and affricate contrasts.

	Labial	Alveolar	Alveolopalatal	Palatal	Glottal
Fricative	ɸ	s z	ɕ ʑ	ç	h
Affricate		ts dz	tɕ dʑ		

(Adapted from Vance, 2008; Labrune, 2012)

Table 2.3 Thai fricative and affricate contrasts.

	Lab-dental	Alveolar	Post-alveolar	Glottal
Fricative	f	s		h
Affricate			tɕ $tɕ^h$	

(Adapted from Tingsabadh & Abramson, 1993)

Japanese has a larger inventory of fricatives and affricates in comparison with Thai. Japanese comprises a set of seven voiceless and voiced fricatives, namely [ɸ, s, z, ɕ, ʑ, ç, h] and four voiceless and voiced affricates, namely [ts, dz, tɕ, dʑ]. In Thai, however, there are just three fricatives and two affricates with no voicing contrasts. That is, /f/ is labiodental; /s/ is articulated with the tip of the tongue making partial contact behind the upper teeth; /h/ is glottal. For affricates, there are only two affricates in Thai with aspiration contrasts, namely a voiceless aspirated post-alveolar affricate [tɕʰ] and a voiceless post-alveolar affricate [tɕ]. As shown in the figures above, Japanese distinguishes voiced and voiceless contrasts, but Thai distinguishes between aspirated and unaspirated contrasts. Moreover, the [tɕ] sound of both languages are transcribed with the same phonetic symbols, however, in terms of articulation, the Japanese [tɕ] is said to be more posterior[5] than the Thai [tɕʰ] and [tɕ] (Konishi, 2005; Trakantalerngsak, 2015).

The SLM model (Flege, 1995) makes predictions about the difficulty of speech learning. The model distinguishes between “new” and “similar” sounds and claims that the greatest difficulty for L2 learners can be expected in the category of “similar” sounds, rather than “new” sounds. L2 learners will not create a new phonetic category for sounds that are similar to L1 phonemes, while for L2 sounds that do not exhibit a categorical overlap with L1 phonemes, a new category is created. For the charts above, if using IPA labels as indicators of identical, similar or new phonetic properties, then the Japanese sounds [(d)z, ts, (d)ʑ] would be considered as “new” sounds for Thai learners and easy to acquire since these sounds do not exist in Thai. In contrast, [tɕ] is considered as a “similar” sound (hence making it difficult to acquire for Thai learners) since they share the same IPA labels in both languages and Thai learners possibly assimilate the Japanese [tɕ] to their Thai [tɕ] category.

[5] The articulation of palato-alveolar is further forward than alveolo-palatal sounds (Ladefoged & Maddieson, 1995).

2.6.2 Thai learners' perception and production of Japanese fricative and affricate contrasts

Previous studies have proved that Thai learners both distinguish and produce Japanese fricative and affricate contrasts poorly. This section will present previous studies that have examined the learning difficulties presented by the Japanese fricative and affricate contrasts in perception and production by Thai learners. The explanation will focus on [(d)z], [ts], [tɕ], [(d)ʑ] since these contrasts will be the main target of the current study and they were proven to be the most problematic sounds to acquire for Thai learners, as investigated in the pilot study to this research.

Table 2.4 illustrates the error patterns in both perception and production made by Thai learners as reported in previous studies. An explanation of error types by each contrast will be given.

[(d)z]

The Japanese [(d)z] is highly problematic for Thai learners because there is no equivalent sound in the Thai sound system. (e.g., Higashi, 1986; Kawano, 2014; Onishi, 1976; Sukegawa, 1993). The Japanese [(d)z] is typically substituted with the voiceless sound [s]. This substitution is fully consistent in most previous studies.

[ts]

It is commonly stated that [ts] presents a challenge and is one of the most difficult sounds for many Japanese learners to produce (e.g., Yamakawa et al., 2005). As shown in the table, Thai learners confuse [ts] most commonly with the voiceless fricative [s], followed by the voiced [z]. Previous studies report results from acoustic analysis that [ts] as uttered by Thai learners has no burst spike in the beginning of the sound, which is a prominent acoustic feature of an affricate sound, showing evidence that Thai learners substitute the target sound with an [s] (Trakantalerngsak, 2015; Yamakawa et al., 2005).

Table 2.4 Error patterns in perception and production made by Thai learners of Japanese.

Japanese		Error patterns	
Error type for [(d)z]			
[z]	➤	**[s]**	Hiraiwa (2004), Kawano (2014), Onishi (1976), Sukegawa (1993), Yamakawa et al. (2005).
Error type for [ts]			
[ts]	➤	**[s]**	Hiraiwa (2004), Kawano (2014), Onishi (1976), Sukegawa (1993), Yamakawa et al. (2005).
	➤	**[z]**	Hiraiwa (2004), Kawano (2014), Sukegawa (1993), Yamakawa et al. (2005).
Error type for [tɕ]			
[tɕ]	➤	**Thai [tɕ]**	Onishi (1976), Sukegawa (1993).
	➤	**[ɕ] (initial position)**	Hiraiwa (2004), Konishi (2005), Trakantalerngsak (2013).
	➤	**[(d)ʑ] (medival&final)**	Konishi (2005), Sukegawa (1993) Trakantalerngsak (2013), Yamakawa et al. (2005).
Error type for [(d)ʑ]			
[(d)ʑ]	➤	**[tɕ]**	Kawano (2014) Konishi (2005), Onishi (1976), Sukegawa (1993), Trakantalerngsak (2013), Yamakawa et al. (2005).
	➤	**Thai [j]**	Onishi (1976), Sukegawa (1993), Trakantalerngsak (2013).
	➤	**English [dʒ]**	Trakantalerngsak (2013).

[tɕ]

Trakantalerngsak (2013)[6] found that both the perception and production of [tɕ] is likely to be mistaken. Prior studies show that there is a high possibility that Thai learners use the Thai [tɕ] to produce the Japanese [tɕ] (Konishi, 2005; Trakantalerngsak, 2013, 2015). They also suggest that the Thai [tɕ] sounds stronger than the Japanese [tɕ] because the Thai [tɕ] is articulated at a more frontal position than the Japanese [tɕ]. An acoustic analysis of the place of articulation by Trakantalerngsak (2015) also shows evidence that Thai learners produced [tɕ] with lower spectral peak location values than Japanese native speakers, indicating that their place of articulation occurs at a more frontal position than Japanese speakers.

[(d)ʑ]

It has been reported in previous studies that Thai learners substitute the Japanese [ʑ] and [dʑ] with the Thai [tɕ] and palatal approximant [j], respectively (e.g., Onishi, 1976; Sukegawa, 1993). However, recent studies have found only error patterns substituting in the English [dʒ], the Thai [tɕ] and a sound in between [dʑ] and [tɕ] (Kawano, 2014; Trakantalerngsak, 2013).

---

[6] Trakantalerngsak (2013) investigated the production and perception acquisition of the Japanese sounds [ɕ], [tɕ], [(d)ʑ] by 30 Thai learners.

Table 2.5 Conventions of errors that Thai learners follow when perceiving the Japanese fricative and affricate contrasts (Trakantalerngsak, 2012).

Target contrast	Sound substitution
[ɕi]	- mostly as Thai [tɕʰi] - as English [ʃi] - as Japanese [ɕi]
[tɕi]	- as Thai [tɕʰi] - as Thai [tɕi] - as Japanese [tɕi]
[(d)ʑi]	- as Thai [tɕi] - as Thai [ji] - as English [dʒi]

Through conducting interviews Trakantalerngsak (2012) investigated what rules or strategies Thai learners use to perceive the Japanese [ɕi], [tɕi] and [(d)ʑi] contrasts. From the table, the majority of the results indicate that Thai learners substitute voiced contrasts with unvoiced contrasts and they tend to use their L1 to grasp the target sounds. The substitution rules from this study are relatively similar to those reported in Table 2.4. Strategies used to perceive target sounds vary between individuals but most Thai learners are shown to follow similar patterns. The table shows that these perceptual-based strategies result in inaccurate substitution errors.

To summarise, as described earlier, previous studies have shown that Thai learners tend to substitute in their L1 sounds to perceive and produce the Japanese sounds. It is for this reason that the present study aims to focus on the four Japanese fricative and affricate contrasts [(d)z, ts, tɕ, (d)ʑ] and hopes to show that HVPT perceptual training can modify Thai learners' perception to enable them to more accurately identify Japanese fricative and affricate contrasts.

2.6.3 Pronunciation instruction of Japanese as a second language

While recent studies have indicated that Japanese pronunciation instruction is necessary, unfortunately, very few studies have investigated learners' difficulties with Japanese pronunciation. Moreover, very few studies have conducted any instruction to reduce problems, even though it has been reported that learners do require and are eager to have more opportunities to improve their pronunciation learning (e.g., Kawano & Ogawara, 2009; Matsuzaki, 2001; Ogawara, 1998a,b). Ikeda (2003, p. 89-90) explains the need of CALL (Computer-Assisted Language Learning) in Japanese language education and gives a concrete example of CALL teaching materials. Concerning pronunciation teaching issues, he mentions that there is still a lack of pronunciation teaching materials that can aid effective instruction in a limited amount of time. The main purpose of developing CALL pronunciation materials is to give an opportunity for learners to practice on their own as much as possible, and to present functions such as minimal-pairs practice, feedback and procedures that will help enhance motivation towards learning. Also, inaccurate pronunciation might have been caused by either inaccurate perception or the possibility that learners themselves do not realise their own mistakes when producing L2 contrasts. Hence, it is necessary to promote self-listening or self-monitoring to enable learners to evaluate themselves whether their speech is accurate or not.

Methapisit (2014) detailed the current status and issues regarding Japanese pronunciation teaching for native Thai learners of Japanese. The study surveyed and investigated three main issues, which are 1) consciousness towards pronunciation acquisition and learners' self-evaluation of their pronunciation ability, 2) pronunciation learning methods 3) teachers' knowledge of pronunciation teaching. The results show that Thai learners have a high need to improve their intelligibility in acquiring Japanese pronunciation in order to sound less non-native and to better identify those contrasts which do not exist in the Thai phonology system. More than 70% of learners reported that they do not feel confident with their Japanese pronunciation and listening ability. Specifically, they reported that fricative, affricate and voiced contrasts are the most problematic for them. Concerning the method of pronunciation learning, the study reports that Thai learners mainly receive pronunciation instruction through “practice in the classroom[7]”, “practice through media[8]” and through “conversation practice with Japanese native speakers”. Also, the study reports that many learners practice shadowing to help improve their listening and production ability. Lastly, the study proposes that instructors should inform learners about the importance of pronunciation learning and present learning methods. It is necessary to encourage learners to practice pronunciation by themselves since this might yield better results in pronunciation learning than from receiving instruction from a teacher alone.

Next, some Japanese textbooks and workbooks which are commonly used for pronunciation[9] instruction will be reviewed.

Table 2.6 A review of textbooks containing Japanese pronunciation instruction.

Textbook	Main content
Japan Foundation (1978)	An overview of the Japanese sound system and explicit explanations of pronunciation rules. Uses visual aids such as illustrations of mouth and lips during sound articulation.

[7] “Practice in the classroom” refers to using audio sounds from textbooks for practice (Methapisit, 2014).

[8] ”Practice through media” refers to practice by presenting real speech from TV dramas, animations and music to practice pronunciation (Methapisit, 2014).

[9] Pronunciation in the present study refers to both perception and production.

Textbook	Main content
Tanaka & Kubozono (1999)	An overview of the Japanese sound system focusing on theory and explicit explanations of pronunciation rules. Uses a few visual aids such as illustrations of mouth and lips during sound articulation. A few exercises are provided.
Kawano, Tsukiji, Kushida, & Matsuzaki (2004)	Contains practice focusing on Japanese prosody.
Toda (2004)	Contains listening and production practice covering Japanese vowels, consonants, accent and intonations. A handbook for Japanese sound instruction is also provided.
Saito (2006, 2010)	Shadowing practice.
Japan Foundation (2009)	A detailed theory of the Japanese sound system covering segmental and suprasegmental features and explicit explanations of pronunciation rules.
Nakagawa & Nakamura (2010)	Mainly focuses on listen and repeat activities. Shadowing practice is also given, focusing on Japanese prosody.
Yoshiki (2010)	Listening practice focusing on vowels and consonants as well as shadowing practice.
Akagi, Uchida, & Furuichi (2010)	Practice focusing on Japanese prosody.
Miyamoto & Osaki (2011)	Listening practice with a minimal pair identification task focusing on vowels and consonants.

Textbook	Main content
Intercultural Institue of Japan (2011)	An overview of the Japanese sound system and explicit explanations of pronunciation rules and phonetic transcriptions. Uses visual aids such as illustrations of mouth and lips during sound articulation. Some listening exercises in each unit and shadowing practice are provided.
Okubo, Kamiyama, Konishi, & Fukui (2012)	Pronunciation activities focusing mainly on shadowing practice.
Nakagawa, Kihara, Akagi, & Shinohara (2015)	Production practice focusing on Japanese prosody.

Table 2.6 provides an overview of the content of textbooks and workbooks focusing on the instruction of Japanese pronunciation. Some textbooks focus on Japanese vowels and consonants (Miyamoto & Osaki, 2011) while other textbooks focus on Japanese prosody, such as elements like accent or intonation (Nakagawa & Nakamura, 2010; Nakagawa et al., 2015). Most textbooks provide a comprehensive introduction to Japanese sounds. Some also speculate about learning difficulties for some specific learners such Korean and English learners (Intercultural Institute of Japan, 2011). Moreover, some short identification task listening exercises are given in some textbooks (Tanaka & Kubozono, 1999). However, it is clear that most textbooks remain underdeveloped and ineffective. Explanations are still limited to verbal indications which do not help learners put theory into practice and do not serve to motivate learners. Even though some textbooks provide listening exercises, the volume and range of task types are still very limited. Some textbooks contain no pronunciation activities or exercise at all. It is necessary to conclude that a variety of pronunciation task types in L2 textbooks is called for since a wide range of task types would benefit students of varied learning styles (Derwing, Diepenbroek, & Foote, 2012). Based on

the analysis of these textbooks, it is clear that most of the contents and exercises are not designed to suit learners from different L1 backgrounds. Moreover, most textbooks still lack any access to instant feedback and to a variability of Japanese speakers (Miyamoto & Osaki, 2011). Feedback is important since it enables learners to determine whether what they are doing is appropriate or not. Recently, new techniques such as the use of "shadowing" are frequently introduced to help with the instruction of Japanese pronunciation (Saito, 2006, 2010; Okubo et al., 2012; Yoshiki, 2010). However, Derwing & Munro (2015) state that "shadowing has been broadly criticised for being boring and demotivating" (p.123).

After undertaking an examination of the relevant textbooks dealing with Japanese pronunciation instruction, it is clear that there is still a lot that needs to be developed since these textbooks have not sufficiently responded to learners' needs and do not motivate learners. Bowen (1972) states that motivation is a powerful factor influencing the improvement of learners' pronunciation. New and advanced tools such as web-based pronunciation teaching or HVPT perceptual training that can control for individual differences or different L1 backgrounds, provide instant feedback with more refined functions and help enhance motivation towards learning would benefit learners more. Derwing & Munro (2015) state that nowadays the advent of advanced digital technologies has greatly expanded the possibilities for pronunciation instruction. It is thereby necessary to adopt more effective and approachable learning tools that can provide more convenient and higher quality tools in pronunciation instruction. Moreover, the benefit of online-based teaching is now doubly attractive because of its cost-saving, user-friendliness, convenience and accessibility such that even learners in remote areas can benefit from immediate success (Dewing & Munro, 2015; Huffman, 2011).

2.7 Research overview, goals and questions

Based on the studies reviewed above, if L2 adult learners are given an accurate and sufficient amount of L2 input, they will be able to acquire the target sounds accurately. Previous studies have shown that perceptual training using a high-variability method can successfully improve L2 learners' perception and production with several L2 contrasts. However, as far as the author of this study has found, none of the previous research undertaken has tested the effects of HVPT perceptual training on Japanese fricative and affricate contrasts by Thai learners, which have proved to be difficult and are one of the most frequent errors made by Thai learners both in perception and production. The ultimate goal of this study is to investigate the effects of HVPT perceptual training on the perception and production of Japanese fricative and affricate contrasts by Thai learners. Moreover, while the approach of this perceptual training was based on previous research, it differs from earlier work in that this study also investigated whether introducing a self-monitoring task aimed at raising awareness in learning would yield better learning of the target sounds. In other words, "standardised HVPT perceptual training" and "a modified HVPT perceptual training including a self-monitoring task" are both key aspects of the design of this study. The research questions of this study are detailed below with the aim to contribute to the present understanding of the value of HVPT perceptual training in the teaching and learning of L2 contrasts. To that end, the pre-test, post-test, generalisation-test and delayed post-test conducted during the course of this study were conducted in order to address the following issues:

1. This study adds to the growing literature on the use of HVPT perceptual training by targeting particular Japanese fricative and affricate contrasts which are known to be difficult to acquire among Thai learners of Japanese. As has been examined by several previous studies, HVPT perceptual training can improve the performance of L2 segmental identification, and the positive

results garnered from those previous studies led to the idea that this training could be modified to use with Thai learners of Japanese. The ultimate purpose of this study is to investigate whether computer-based perceptual training using an HVPT method is effective in leading to increased accuracy in the perception of Japanese fricative and affricate contrasts by Thai learners.

2. With the aim of raising awareness in learning, this study will also adopt perceptual training with a self-monitoring task. One group will receive a course of standardised HVPT perceptual training with an identification task, while the other group was designed to be trained with HVPT perceptual training using an identification task as well as a self-monitoring task. The aim is to see whether introducing the “self-monitoring” task will lead to greater improvement in learning the target sounds. However, instead of aiming at testing the efficacy of standardised HVPT perceptual training, rather, this modification seeks to extend HVPT perceptual training’s potential in that self-monitoring might possibly contribute to greater learning.

3. Some previous studies have indicated that perceptual training has lead to an improvement in production despite no explicit production training having been provided (Bradlow et al., 1997; Rochet, 1995). However, there has not been a sufficient number of training studies which have investigated this transfer to the production domain. Question three is to investigate more thoroughly whether there is any transfer of perceptual learning to production. If there is any improvement in production ability, it will provide evidence to support the existence of a relationship between speech perception and production.

4. Generalisation has often been used as a measure of the robustness of perceptual learning (Lively et al., 1994; Hardison, 2004, Lacabex, Lecumberri, & Cooke, 2009). This study also aims to investigate whether perceptual

learning generalises to new words and new talkers that learners had not encountered during training.

5. To see whether or not any perceptual learning gained from the training would be retained for six months after training is over.

6. To provide empirical data about the relationship between L2 perception and production.

Table 2.7 Summary of objectives and hypotheses.

Issue	Objectives	Hypotheses	Theoretical background
Perception	To see whether the training given is effective in leading to increased accuracy in perception.	HVPT perceptual training is effective in improving the identification of Japanese fricative and affricate contrasts.	Bradlow et al. (1997); Lively, Logan et al. (1993); Lively, Pisoni et al. (1994); Lopez-Soto & Kewley-Port, (2009); Wong (2013)
Production	To see whether there is a transfer of perceptual learning to the production domain.	There will also be an improvement in production.	Bradlow et al. (1997); Hazan et al. (2005); Iverson & Evans (2009); Nobre-Oliveira, (2007); Rochet (1995)
Generalisation	To see whether any perceptual learning gained from the training is generalised to new words and new speakers or not.	Generalisation to new words and new talkers will occur.	Rochet (1995); Thomson (2012); Wang (2002); Wang & Munro (2004); Wang (2008)
Long-term effects	To see whether any perceptual learning gained can be retained after the training is over.	Any perceptual learning will be retained in the long-term memory of participants.	Wang (2008); Wang et al. (1999); Wang & Munro (2004); Yamada et al. (1996)
Comparing two training types	To investigate whether introducing a self-monitoring task would lead to a better learning outcome.	Raising learners' awareness by including self-monitoring in the training will yield greater learning in learners.	Borden (1983); Hirano-Cook (2011); Ingles (2011); Sardegena (2011)
The relationship between perception and production	To examine whether there is a link between perception and production.	There will be a link between these two modalities. An improvement in perception will also lead to a production improvement.	Bradlow et al. (1997); Rochet (1995); Wang (2003)

Chapter 3
Research Methodology

3.1 Introduction

The main aim of the research experiment was to investigate the effect of HVPT perceptual training on Japanese fricative and affricate contrast learning by Thai learners of Japanese. In this chapter the research design and procedures will be explained. It covers the details of the pre-test phase, the nine-session HVPT perceptual training phase, the post-test phase, the generalisation-test phase and a six-month delayed post-test phase.

3.2 Participants

A total of 38 native speakers of Thai (35 females and 3 males; mean age = 21.5, range 18-32) participated in the experiment. Two participants were then excluded because they were overqualified for the experiment as their pre-test identification scores were greater than 85%[10] (indicating that they were near or at ceiling level and were unlikely to show further improvement). Five participants dropped out due to conflicting schedules with the long-term commitment of the experiment meaning that a total of 31 participants finished the experiment and these are the subjects that form the basis for the eventual analysis of results. All participants spoke standard Thai and they were all recruited and tested in Thailand. They were paid for their participation. Their levels of Japanese proficiency varied widely at the pretest phase and all were highly motivated to improve their Japanese listening and pronunciation skill. They were informed that the experiment would involve a

[10] The selection of participants was based on prior training studies (Iverson et al., 2012).

pre-test, training, a post-test and generalisation tests which would take place over approximately 12 weeks. Thirty-one participants were randomly assigned into trained groups (Trained Group A and Trained Group B) and a control group according to their time availability; 11 served as controls and 20 as trained participants in the two trained groups. All participants performed similar pre/post tests, generalisation tests and delayed post-tests but only the trained groups went through a nine-session course of HVPT perceptual training. The 20 trained participants were divided into Trained Group A and Trained Group B which differed in the training tasks conducted. None of the participants reported having hearing problems. The participants' information and language background are detailed in Appendix A.

Table 3.1 Language background information.

Group	Average years of learning	Age
Control	5	21-26 (Mean=22)
Trained Group A	5.5	19-27 (Mean=21)
Trained Group B	6	19-32 (Mean=23)

Table 3.2 Number of participants.

Control Group	Trained Group A	Trained Group B
- 14 completed pre-test (3 dropped out) - No course of training - 11 completed post-test and generalisation test - 11 completed delayed post-test	- 13 completed pre-test (2 dropped out) - 11 subjects completed training, post-test and generalisation test - 11 completed delayed post-test	- 11 completed pre-test (2 overqualified) - 9 completed training, post-test and generalisation test - 8 completed delayed post-test

3.3 Speech materials

Speech materials included: (a) Pre/post-test stimuli that were presented to participants before and after the training; (b) training stimuli and (c) generalisation-test stimuli.

3.3.1 Pre/post/delayed post-test stimuli

Thirty-two nonsense words[11] of natural tokens consisting of Japanese fricative and affricate contrasts and 14 fillers were used to create the pre-test and post-test stimuli (listed in Appendix G). All words were presented in isolation[12] in Hiragana[13]. Each word was made up of three moras and contained either [z], [ts], [tɕ], or [(d)ʑ][14] in three possible positions: Initial (e.g., [tɕimama]); medial (e.g., [mazoma]); and final (e.g., [maʑoː]). In the Japanese language, all Japanese fricative and affricate contrasts apart from the alveolar affricate [ts] are followed by a limited selection of phonetic environments [a, i, ɯ, e, o] as shown in Table 3.3, according to the phonotactics of the language.

Table 3.3 Target sounds and their phonetic environments.

Contrasts	Phonetic environments
[(d)z]	[_a], [_ɯ], [_e], [_o]
[ts]	[_ɯ]

11 Nonsense words were used to avoid the influence of lexical knowledge and due to the limited availability of phonetically appropriate real words (Hazan et al., 2005; Huensch, 2013).

12 According to Colantoni et al. (2015), L2 perception experiments commonly include words in isolation if researchers are interested in how accurately learners can recognise the words or sounds. Thus, in this study words were presented in isolation.

13 All participants were familiar with reading Hiragana and so the Hiragana script was considered to be appropriate for representing words to participants in this experiment.

14 A pilot study on perception and production of Japanese fricative and affricate contrasts by eight Thai learners showed that [(d)z, ts, tɕ,(d)ʑ] were proven to be the most difficult to identify and produce among Japanese fricatives and affricates. [ɕ] was removed since most participants were already near ceiling.

Contrasts	Phonetic environments
[tɕ]	[_a], [_i], [_ɯ], [_o]
[(d)ʑ]	[_a], [_i], [_ɯ], [_o]

All of the pre/post-test stimuli used in the tests were produced by a male native speaker of Japanese (JS1) from Saitama prefecture who is also a phonetician and who represents standard Japanese pronunciation. JS1 was asked to read a list of words presented in Hiragana twice. The utterances were recorded using a SONY PCM-D50 recorder and a RODE NT3 microphone with a sampling rate of 44.1 kHz. The recordings of the words were then segmented into individual WAV files using Praat software (Boersma & Weenink, 2012).

In order to measure the generalisation of improvements, another perception test was conducted at the post-test phase and this perception test used sounds uttered by a new talker (JS6). This test will be later referred to as the first type of generalisation test. The recording procedure was same as conducted with JS1.

Before using these stimuli to conduct the pre/post-test, the stimuli were screened for intelligibility by five Japanese native speakers who were used as a baseline group (See details about the Japanese baseline group in Appendix D). To demonstrate reliability, all the test stimuli used in the pre-test were identified with an accuracy of 99.38% (159/160) of correct identifications among those five native Japanese speakers, indicating that the test stimuli were appropriately representative of each sound tested.

3.3.2 Training stimuli

In order to develop a word list large enough to accommodate unique training and to obtain a greater variability in training tokens, both real words and nonsense words were used in the training. Seventy-two minimal pairs (which are detailed in Appendix H) served as training stimuli. The real words were chosen from the *Meikyo Kokugo Jiten* dictionary (Kitahara, 2010). Training stimuli were grouped into three pair sets: Set ① with "[(d)z] vs. [s]", Set ② with "[ts] vs. [s]", Set ③ with "[tɕ] vs. [(d)ʑ]". For each of these contrasts, two of the minimal pairs consisted of real-word vs. real-word pairs while the other half were real-word vs. nonsense-word pairs. To best exploit the high variability method, 1) four different native Japanese speakers (two males and two females) as represented in Table 3.4 produced the training stimuli; 2) a wide variety of different phonetic environments ([a, i, ɯ, e, o]) and word locations were used (initial/medial/final). Before the training, the stimuli were screened for intelligibility by the same five native speakers used in the pre/post-test stimuli. All the stimuli used in the training were 100% accurately identified by the five native Japanese speakers.

Table 3.4 Training talkers.

Talker	Gender	Age	Hometown
JS1	M	50s	Saitama
JS2	F	20s	Nigata
JS3	M	20s	Kobe
JS4	F	30s	Osaka

3.3.3 Generalisation-test stimuli

The generalisation-test stimuli were 88 real words (64 target words and 24 fillers), selected from a word familiarity database by the NTT Database series (Amano & Kondo, 2008) to assure the frequency of their use. All stimuli were spoken in isolation. Stimuli with different levels of lexical familiarity were used to see whether lexical familiarity might have an effect on their identification ability. Lexical familiarity of stimuli starts from level one to seven (the higher the number, the higher the degree of familiarity) and was controlled by a database. Words from level one and two were chosen to create "a low lexical familiarity test" and words from level five to seven were chosen to create "a high lexical familiarity test". All words are presented in Appendix I.

Table 3.5 Generalisation test design.

	Low lexical familiarity test (level 1 to 2)		High lexical familiarity test (level 5 to 7)
1	By familiar talker (JS2)	1	By familiar talker (JS2)
2	By unfamiliar talker (JS5)	2	By unfamiliar talker (JS5)

As shown in Table 3.5, the stimuli were recorded by two Japanese native speakers; JS2 (a Japanese female speaker also used in the training session) and JS5 (a new talker whose voice was not used in the training and pre/post-test) Later JS2 is referred to as the "familiar talker" and JS5 as the "unfamiliar talker" (See Appendix E for talker information). Before the generalisation test, the stimuli were screened for intelligibility by the five native speakers. All stimuli used in the generalisation test were 100% accurately identified by the same five native Japanese speakers who identified the pre/post-test and the training stimuli.

3.4 Procedures

The procedure consisted of a pre-test phase, a training phase, a post-test phase and a generalisation test phase, which was conducted over approximately 12 weeks. A delayed post-test was then conducted approximately six months after the post-test. A more detailed explanation of each phase is given below. The pre-test, post-test and generalisation-test phases were administered individually in a quiet room. The perceptual training and a delayed post-test phase were both completed online.

Table 3.6 Research experiment design.

<table>
<tr><th>Experiment phase</th><th>Mode</th><th>Tasks</th><th>Participants</th></tr>
<tr><td rowspan="2">Pre-test</td><td>Perception</td><td>- Identification tasks (no feedback)
- Stimuli: 32 nonsense words in Hiragana</td><td>- Trained Group A (11)
- Trained Group B (9)
- Control Group (11)</td></tr>
<tr><td>Production</td><td>- Reading-aloud task
- Stimuli: 32 nonsense words in Hiragana</td><td>- Trained Group A (11)
- Trained Group B (9)
- Control Group (11)</td></tr>
<tr><td rowspan="2">Training
(9 sessions)</td><td>Type 1</td><td>- Identification tasks with immediate feedback
- Stimuli: 480 nonsense and real words in Hiragana</td><td>- Trained Group A (11)</td></tr>
<tr><td>Type 2</td><td>- Identification tasks with immediate feedback
- Discrimination tasks (self-monitoring)
- Stimuli: 480 nonsense and real words in Hiragana</td><td>- Trained Group B (9)</td></tr>
<tr><td rowspan="2">Post-test</td><td rowspan="2">Perception</td><td>Test 1: Same as perception pre-test</td><td>- Trained Group A (11)
- Trained Group B (9)
- Control Group (11)</td></tr>
<tr><td>Test 2: Same as perception pre-test but tokens were produced by an unfamiliar talker (JS6)</td><td>- Trained Group A (11)
- Trained Group B (9)</td></tr>
</table>

Experiment phase	Mode	Tasks	Participants
	Production	Same as pre-test	- Trained Group A (11) - Trained Group B (9) - Control Group (11)
Generalisation test	Perception	- Identification tasks (no feedback) - Two tests (High familiarity and Low familiarity uttered by familiar talker and unfamiliar talker)	- Trained Group A (11) - Trained Group B (9) - Control Group (11)
Delayed post-test (6-month)	Perception	Same as perception pre/ post-test	- Trained Group A (11) - Trained Group B (8) - Control Group (11)

3.4.1 Pre-test phase

All participants (the control group, Trained Group A and Trained Group B) were asked to complete a brief language background questionnaire and a consent form (Appendix B) and then to perform the pre-test. The pre-test consisted of a perception test and a production test. They were given a printed paper with clear instructions. They were asked to perform an identification task included in the perception test, followed by the production test. The completion of the perception test and the production test took approximately 15 to 20 minutes altogether.

Participants assigned to the control group were informed that they would be asked to take a test again in about 12 weeks time. Participants who were assigned to the trained groups were given a sheet of paper which detailed the training schedule and instructions.

The pre-test procedures are described in more detail below.

Table 3.7 Pre-test details.

Pre-test phase	
Perception pre-test	Production pre-test
Participants identified the consonants from two-alternative forced-choices of three minimal pairs ([(d)z]-[s], [ts]-[s], [tɕ]-[(d)ʑ]).	Participants recorded the target words which were used in the perception test.

Perception Test

The perception test took the form of a two-alternative forced-choice identification task. In the perception test, 32 target sounds and 14 fillers were presented in isolation, all arranged in random order. The pre-test and the post-test included a total of 46 trials. All stimuli were produced by a male native speaker of Japanese (JS1). The word list used is shown in Appendix G. Participants were presented with minimal pairs and they were told that they were going to listen to some Japanese words and, after hearing each word, their task was to identify the word that they had heard by circling it on the answer sheet. The two-alternative forced-choices available formed a minimal pair relating to the target word token (e.g., after hearing "つねに" the participants were asked to identify it as "すねに" or "つねに"). The reason that these particular minimal pairs ([ts] vs. [s], [(d)z] vs. [s], [tɕ] vs. [(d)ʑ]) were chosen as the multiple choices is because these pairs of sounds, in particular, present a problem for Thai learners of Japanese (Konishi, 2005; Trakantalerngsak, 2013; Yamakawa et al., 2005). They received no feedback and were not able to replay the stimulus. Test stimuli were presented through Sennheiser headphones connected to a laptop in a quiet room. The perception test lasted about five to seven minutes.

Production Test

The production test was completed after the perception test. It consisted of a reading-aloud task[15]. All participants were given a list of the words that had been heard in the perception test and asked to read the 42 stimuli out loud twice. The productions were recorded in a quiet room using a SONY PCM-D50 recorder and a RODE NT3 microphone at 44.1 kHz. The participants recorded all 42 tokens in one session and were allowed a short rest after every nine tokens. They could also pause and resume recording at their own pace. The production test took about five to ten minutes in total. The recordings were conducted separately, participant by participant. Because the Trained Group B participants will later conduct a self-monitoring task during the training, after the production test they were asked to stay five to ten minutes more to record training stimuli which will be used in the training task.

[15] According to Thomson & Derwing (2014), "reading-aloud tasks were by far the most common assessment method of L2 pronunciation studies" (p. 6).

3.4.2 Training phase

The details below will summarise the training procedures.

Table 3.8 Training details.

Training Phase		
Control	Trained Group A	Trained Group B
No training	Participants identified the word they heard from two-alternative forced choices with immediate feedback.	
		Participants compared their own pronunciation to native Japanese speaker sounds to see whether they are same or not with immediate feedback.

There were nine[16] sessions of HVPT perceptual training. The training phase took place approximately four weeks after the pre-test phase. The training was conducted using a computer program which was designed by the research author and coded by a professional programmer. The participants of Trained Groups A and B were instructed to complete the training online at the following website "*http://128.199.227.231/perceptual*"[17]. Prior to the first session of training, the participants were given a username and password to login to their own accounts so that their work progress could be verified and it could be determined whether they had completed all sessions or not. When they logged into the site, a welcome screen appeared showing them which

[16] According to Iverson et al. (2012, p. 145-146), five to ten sessions of high-variability training tends to improve the identification of the L2 contrasts and improvement are retained for a significant time.

[17] Due to the server renting service, the site was activate until March, 2016. After that only the researcher could use through by her personal computer.

sessions they had left to complete and how many sessions they had already completed, as presented in Figure 3.1 and 3.2. Instructions given in the training were either in Thai or English but the target words were presented in Hiragana script.

Figure 3.1 A screenshot from the opening page of the training site ("*http://128.199.227.231/perceptual*").

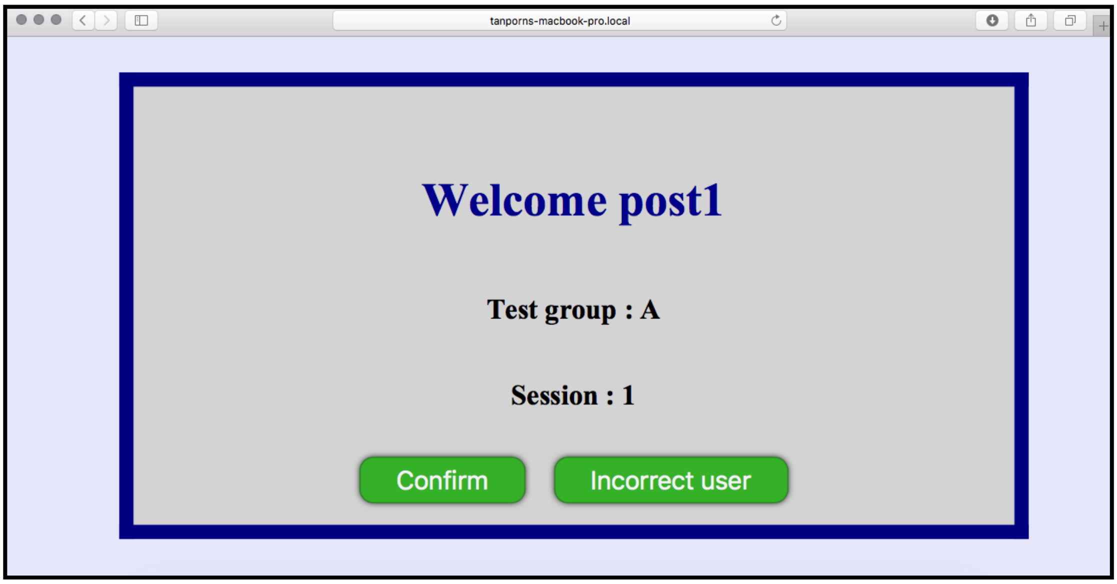


Figure 3.2 A screenshot of the second page shown after logging in.

An identification task[18] consisting of two-alternative forced-choices was mainly adopted in the training. The objective of the experiment was to investigate whether the effect of one condition on one group lead to better learning with than effect of a different condition on the other group. Therefore, 20 participants were divided into two groups; Trained Group A (*n*=11) and Trained Group B (*n*=9). In Trained Group A, participants were trained using only a two-alternative, forced-choice identification task. In Trained Group B, participants were trained using a two-alternative, forced-choice identification task and an AX[19] discrimination task (self-monitoring task) which was conducted only in sessions one to six. The procedures of each task will be described in detail later below. In order to balance exposure to stimuli across groups, while Trained Group B participants received eight tokens of self-

---

[18] A forced-choice identification task is widely used in the training studies since there is an evidence that it encourages learners to classify a new contrast into categories and develop phonetic categories (e.g. Bradlow et al., 1997; Jamieson & Morosan, 1986, 1989; Huensch, 2013).

[19] In an AX discrimination task, listeners are presented with two sounds and have to decide whether they are the 'same' or 'different' (Colantoni et al., 2015).

auditory recordings, Trained Group A participants were also presented with an identification task consisting of eight tokens produced by Japanese speakers. Both Trained Group A and Trained Group B participated in nine sessions of training. The procedure for the training sessions was identical to the procedure described for the pre/post-test sessions, except that participants received feedback during their training sessions.

Figure 3.3 Experimental interface for the two-alternative forced-choice identification task showing the two alternatives, [mazɯma] and [masɯma].

Figure 3.4 A screenshot from a training session showing an incorrect answer has been selected.

Figure 3.5 A screenshot from a training session showing how feedback is provided after a selection has been made. The dark green button shows the correct answer whereas the red button is incorrect.

In the identification task, all trained participants listened to isolated target word tokens and were then asked to identify each word they heard by clicking a button on a computer screen. The two choices available formed a minimal pair relating to the target word token (e.g., “まずま [mazɯma]”; identify as “まずま [mazɯma]” or “ますま [masɯma]”) as shown in Figure 3.3. Immediate feedback[20] was presented using text displays of “Correct” or “Incorrect” as shown in Figure 3.4. In addition, the correct response was then highlighted in green and the incorrect response was highlighted in red. When the answer was incorrect another window appeared automatically and participants were prompted to listen to the correct and incorrect stimulus for a maximum of three times each, or to proceed to the next trial.

In the AX discrimination task or self-monitoring task (performed only by Trained Group B), participants listened to native Japanese model speech and their own recorded utterances. They then responded as to whether those sounds were the “same” or “different”. As soon as they pressed the button, an immediate feedback message was presented using text displays. In this task, there were no “right” or “wrong” answers but they were presented feedback on how many degrees from nativeness their productions were, as judged by two native Japanese speakers (R1 and R3, See details in Appendix F) as shown in Figure 3.6. There were three degrees of feedback as judged by two native Japanese speakers as follows; 1) if two Japanese speakers judged that the Thai learners’ production was accurate then the feedback would be “100% native like”; 2) if only one speaker judged that the production was accurate then feedback of “50% native like” was given; 3) if both of the Japanese speakers judged the target production as wrong then the feedback would be “0% native like”. Through this task, it is hoped that Trained Group B

[20] Immediate feedback is information about whether the subject’s response was correct or incorrect and is the most frequently used type of feedback in the training research (Logan et al., 1991; Logan & Pruitt, 1995).

participants will be able to notice the differences between their own production and that of the model native speaker.

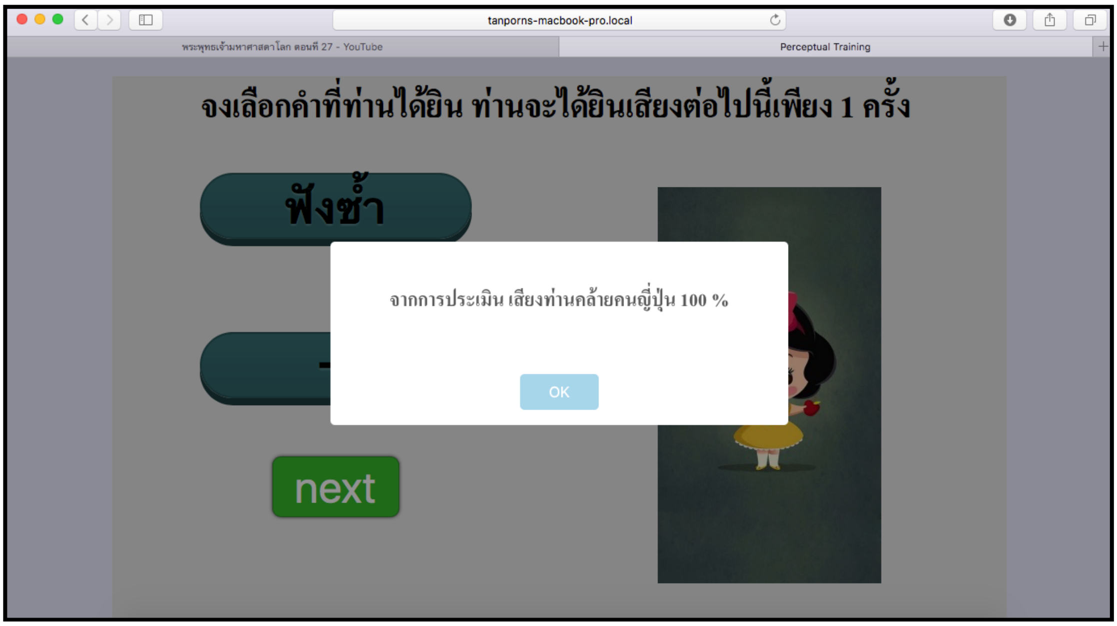


Figure 3.6 A screenshot from a training session wherein feedback is provided during the discrimination task.

For both training tasks, there was no time limit given to participants to input their responses. For sessions one to six, 56 tokens were presented to the participants during each session of training. Following this, during sessions seven, eight and nine, 48 tokens were presented, giving a total of 480 tokens for the nine training sessions. At the end of each session, their total percentage scores of accuracy were shown in order for them to keep track of their training progress. The training sessions took approximately 15 to 20 minutes, with two subject-controlled breaks between blocks. By the end of the session, the program automatically provided the participants' accuracy score given as a percentage on the screen. Participants were instructed to wear headphones during their sessions and complete all the training sessions in a quiet environment. The nine training sessions took place over three to four weeks. Each participant performed three sessions per week. Their sessions were scheduled to occur within a three-day time limit. For example, session one was due to be made available to participants during the period of 25th to 27th of May, 2015.

The table below provides a summary of the training procedures.

Table 3.9 Nine-session training contents.

Session	Contents	
	Trained Group A	**Trained Group B**
1	*Part 1*: 48 tokens of [(d)z] vs. [s] contrast by JS1 and JS2 *Part 2*: 8 tokens of [(d)z] vs. [s] contrast produced by JS1 (56 tokens in total)	*Part 1*: 48 tokens of [(d)z] vs. [s] contrast by JS1 and JS2 *Part 2*: 8 tokens of self-monitoring on [(d)z] vs. [s] contrasts (56 tokens in total)
2	*Part 1*: 48 tokens of [ts] vs. [s] contrast by JS3 and JS4 *Part 2*: 8 tokens of [ts] vs. [s] contrast produced by JS3 (56 tokens in total)	*Part 1*: 48 tokens of [ts] vs. [s] contrast by JS3 and JS4 *Part 2*: 8 tokens of self-monitoring on [ts] vs. [s] contrasts (56 tokens in total)
3	*Part 1*: 48 tokens of [tɕ] vs. [(d)ʑ] contrast by JS1 and JS2 *Part 2*: 8 tokens of [tɕ] vs. [(d)ʑ] contrast produced by JS2 (56 tokens in total)	*Part 1*: 48 tokens of [tɕ] vs. [(d)ʑ] contrast by JS1 and JS2 *Part 2*: 8 tokens of self-monitoring on [tɕ] vs. [(d)ʑ] contrasts (56 tokens in total)
4	*Part 1*: 48 tokens of [(d)z] vs. [s] contrast by JS3 and JS4 *Part 2*: 8 tokens of [(d)z] vs. [s] contrast produced by JS4 (56 tokens in total)	*Part 1*: 48 tokens of [(d)z] vs. [s] contrast by JS3 and JS4 *Part 2*: 8 tokens of self-monitoring on [(d)z] vs. [s] contrasts (56 tokens in total)

Session	Contents	
	Trained Group A	Trained Group B
5	*Part 1*: 48 tokens of [ts] vs. [s] contrast by JS1 and JS2 *Part 2*: 8 tokens of [ts] vs. [s] contrast produced by JS1 (56 tokens in total)	*Part 1*: 48 tokens of [ts] vs. [s] contrast by JS1 and JS2 *Part 2*: 8 tokens of self-monitoring on [ts] vs. [s] contrasts (56 tokens in total)
6	*Part 1*: [tɕ] vs. [(d)ʑ] contrast by JS3 and JS4 *Part 2*: 8 tokens of [tɕ] vs. [(d)ʑ] contrast produced by JS3 (56 tokens in total)	*Part 1*: [tɕ] vs. [(d)ʑ] contrast by JS3 and JS4 *Part 2*: 8 tokens of self-monitoring on [tɕ] vs. [(d)ʑ] contrasts (56 tokens in total)
7	48 tokens of [(d)z] vs. [s] contrast produced by JS1, JS2, JS3 and JS4 (48 tokens in total)	48 tokens of [(d)z] vs. [s] contrast produced by JS1, JS2, JS3 and JS4 (48 tokens in total)
8	48 tokens of [ts] vs. [s] contrast produced by JS1, JS2, JS3 and JS4 (48 tokens in total)	48 tokens of [ts] vs. [s] contrast produced by JS1, JS2, JS3 and JS4 (48 tokens in total)
9	48 tokens of [tɕ] vs. [(d)ʑ] contrast produced by JS1, JS2, JS3 and JS4 (48 tokens in total)	48 tokens of [tɕ] vs. [(d)ʑ] contrast produced by JS1, JS2, JS3 and JS4 (48 tokens in total)

3.4.3 Post-test phase

Immediately after the training was completed, the 31 participants were told to perform a post-test. The details of the post-test were identical to the pre-test. The pre-test and the post-test were conducted with a 12-week interval in between. To see whether generalisation of improvements to new talkers had taken place, two perception tests were conducted in the post-test phase (the first test was uttered by a familiar talker (JS1) and the second test uttered by an unfamiliar talker (JS6)).

Table 3.10 Post-test details

Post-test phase	
Perception post-test	Production post-test
Two perception tests identical to the pre-test (one utilised tokens produced by a familiar talker (JS1) whereas the other used tokens produced by an unfamiliar talker(JS6)).	Identical to production pre-test

3.4.4 Generalisation test phase

After the completion of the post-test, both of the trained groups and the control group were presented with two additional tests of generalisation (this will be referred to as the second type of generalisation test), which included new words not used in the training uttered by an old talker and a new talker, to see whether the effects of perceptual training had been generalised to new words and new talkers.

The generalisation stimuli were recorded by two Japanese native speakers; JS2 (a Japanese female speaker used in the training) and JS5 (a new talker) - later JS2 is referred to as “familiar talker” and JS5 as “unfamiliar talker”. The materials were recorded on a SONY PCM-D50 recorder and a RODE NT3 microphone at 44.1 kHz.

There were two tests included in the second type of generalisation test; a “high lexical familiarity test” and a “low lexical familiarity test”. In the high lexical familiarity test, all participants from the trained groups and the control group heard 88 tokens (32 target words X 2 talkers + 24 fillers) which were produced by a familiar talker and an unfamiliar talker. For the low lexical familiarity test, the participants were given 88 tokens (32 target words X 2 talkers + 24 fillers). All words used in the generalisation tests had not appeared either in the pre-test/post-test phases nor in the training phase. The procedures followed for the generalisation tests were the same as those administered in the perception test of the pre and post-tests.

3.4.5 Delayed post-test phase

In order to assess the retention of the improvement gained from the training, six months after the completion of the post-test a delayed post-test was conducted. Due to the difficulty in scheduling meetings with some participants, one participant from Trained Group B was omitted from the delayed post-test. The details and procedures of the delayed post-test were identical to the pre-test and post-test. However, only a perception test was performed in this phase. The delayed post-test was conducted on the internet using a Google form (https://docs.google.com/forms/d/1gfR3tRdg5DZa8V_mjAB6EdWIHVbesK8-eGLGb9WTpHY/viewform?embedded=true).

3.5 Analysis method

Perception test

The results of the pre-test and post-test were compared to assess the effects of HVPT perceptual training. Perceptual accuracy was assessed through an analysis of participants' mean percentage of correct identification scores. First, the mean average of correct identification scores of the pre-test and the post-test for each group were analysed by conducting a paired sample *t*-test. The results of the *t*-test analysis showed whether each group made a significant improvement in their perception from pre-test to post-test. Next, a further analysis of covariance (ANCOVA) was carried out to see whether there was a significant difference in the post-test across groups after controlling for the effect of pre-test scores.

Production test

Production accuracy was assessed through the identification judgements of five Japanese native speakers (see details in Appendix F) and then through goodness evaluation ratings. The purpose of the identification judgements and goodness ratings was to investigate whether Japanese native listeners could correctly identify the target sounds and whether intelligibility and pronunciation improved as a result of the HVPT perceptual training. All the raters reported having normal hearing. The method of the identification accuracy evaluation was based on prior training research (Bradlow, Akahane-Yamada et al., 1997; Bradlow, Pisoni et al., 1999) and similar to the identification task which Thai learners completed in the perception test. The same statistical analyses were made for both identification judgements and goodness ratings with the same purposes in mind.

The identification task was conducted to see whether Thai learners had accurately produced the target sounds. Five native Japanese raters heard a word produced by Thai learners, then they indicated the word they had heard by identifying the word from one of two choices (similar to the perception task that the Thai learners completed). They were instructed to pay attention to the target contrasts while assessing the words they heard. In some cases that raters could not identify the sounds from the two choices, those productions were considered to be inaccurate. All raters were presented with the pre-test and post-test utterances from the production test, resulting in 1,984 tokens (2 tests x 32 words x 31 Thai learners = 1984 tokens). Raters could listen to the word as much as they needed to. Raters carried out the task one by one in a quiet room and each rating session lasted about three hours in total. Here, Kendall‘s Coefficient of Concordance was performed to see whether all five raters identified participants’ productions reliably and consistently or not.

The purpose of the goodness rating evaluation was to investigate whether intelligibility and pronunciation improved as a result of perceptual training. The procedure of the goodness rating evaluation task was based on prior L2 research (e.g., Hazan & Sennema, 2007; Hirano-Cook, 2011; Southwood & Flege, 1999). The rating procedure was explained in detail to all raters to make sure that they understood the task. The five native Japanese raters[21] rated the consonants in terms of how good an exemplar of the intended consonants they were. They were instructed to pay attention to target contrasts only, while assessing the words they heard. Five raters rated each word on a Likert 5-point scale[22]. '1' represented a poor exemplar of the target consonant or "heavily accented or not Japanese native-like", '5' represented a good, Japanese-sounding attempt or "native-like or not accented at all". All raters were presented with the pre-test and post-test utterances from the production test, resulting in 1,984 tokens (2 tests x 32 words x 31 participants = 1984 tokens). Raters could hear the word as much as they needed to. The rating task was performed individually in a quiet room and lasted about three hours in total.

Both evaluation tasks were performed by participants in isolation using Sennheiser headphones on a laptop in a quiet room, which took approximately six hours in total.

3.6 Questionnaire and follow-up interview

Individual questionnaires and follow-up interviews were conducted with each of the participants after they finished the generalisation tests. The

---

[21] Recent research found no significant difference between experienced raters and inexperienced raters. Both experienced and inexperienced raters provided reliable L2 comprehensibility, accentedness and fluency judgements on ratings (Bongaerts et al., 1997; Derwing, Rossiter, Munro, & Thomson, 2004; Isaacs & Thomson, 2013). In this study, four inexperienced and one experienced raters were used to judge Thai learners' productions.

[22] Although over the past several decades a few empirical studies have suggested that 9-point rating scales are more practical, usable across contexts and sufficiently reliable for L2 pronunciation research purposes, recent research has shown that raters have difficulty managing 9-point scales. Also Isaacs & Thomson's (2013) study shows there is no different significance between 5-point scales' and 9-point scales' efficacy. Hence, the 5-point rating scale was used in this study.

questionnaires were intended to investigate the attitude of trained participants toward the perceptual training. The questionnaire questions were given as detailed in Appendix J. Each participant was asked to comment on areas of difficulty for them regarding the perception and production of Japanese fricative and affricate contrasts. They were also asked for overall comments on the training procedure.

Chapter 4 Results

4.1 Introduction

In this study, the HVPT perceptual training, which has proven effective in yielding perceptual improvements in L2 learners of a different languages in previous studies, was carried out on Thai learners of Japanese with the aim of focusing on Japanese fricative and affricate contrasts. This chapter will provide evidence to answer the research questions as described below. Results will be presented, beginning with the perception test results, followed by the generalisation test results, the delayed post-test results and ending with the production test results. In addition, the chapter will present the results of the correlation between perception and production and will end with the results of a questionnaire which participants were asked to undertake This study was designed to answer the following specific research questions and purposes.

Research questions and purposes

Perception

1. Can HVPT perceptual training of Japanese fricative and affricate contrasts improve Thai learners' identification accuracy of Japanese fricative and affricate contrasts?

2. If an improvement occurs, will there be a difference in learning between the trained groups and the untrained group? And furthermore, will there be a difference between the trained group which received standardised HVPT perceptual training and the group which received modified HVPT perceptual training and also undertook a self-monitoring task?

3. To investigate whether the degree of improvement varies across different contrasts.

4. To investigate whether participants in this study improve to an equal extent or whether there is a variance in the degree of improvement between individuals?

Generalisation

1. Does the perceptual learning gained from perceptual training of Japanese fricative and affricate contrasts enable the generalisation of improvements to new words and new talkers?

2. Will lexical familiarity have an effect on the generalisation test? That is, will words of differing lexical familiarity be generalisable to the same extent?

Delayed Post-Test

1. Will improvements in perception gained from the perceptual training be retained six months after the training has been completed?

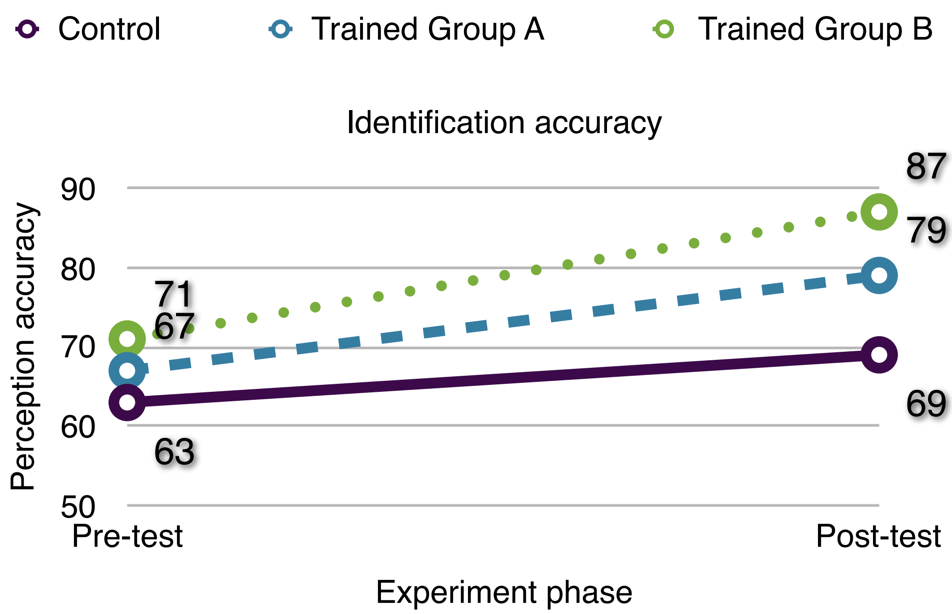
Production

1. Can the HVPT perceptual learning gained from perceptual training of Japanese fricative and affricate contrasts be transferred to production in terms of increased production accuracy scores?

2. If an improvement occurs, will there be a difference in learning effect between the trained groups and the untrained group; and furthermore, the group which received standardised HVPT perceptual training and the modified HVPT perceptual training group who also received the self-monitoring task?

3. To investigate whether the degree of improvement will vary between contrasts.

4. To investigate whether participants in this study improve to an equal extent or is there is a differing degree of improvement seen between individuals?



The relationship between perception and production

1. Is there a correlation between improvements in the perception and production of Japanese fricative and affricate contrasts before and after the training?

2. Are individual gains in perception related to production?

4.2 Perception results

4.2.1 Overall pre-test and post-test performance

The first question which will be addressed in the research experiment concerns the effects of the HVPT method of perceptual training and whether it can lead to increased accuracy in the perception of Japanese fricative and affricate contrasts.

The identification performance of the two trained groups, along with the control group, was assessed before and after the four-week period of HVPT perceptual training using two-alternative forced-choice identification task perception tests. The aim was to see whether or not there was a significant improvement from pre-test to post-test in each group, and whether the trained groups' identification accuracies were significantly higher than the untrained group or not. First, the overall perception results of pre-test and post-test scores for each group are presented in Figure 4.1.

Figure 4.1 Mean percentage of correct identification scores in Japanese fricative and affricate contrasts by Control group, Trained Group A and Trained Group B.

The mean percentage of correct identification scores for the pre-test and the post-test for each group for all four target consonants are displayed in Figure 4.1. For the control group, the mean percentage of correct identification scores for the pre-test was 63% and 69% for the post-test. For Trained Group A, the mean identification scores increased from an average of 67% for the pre-test to 79% for the post-test. Lastly, for Trained Group B, the mean identification scores increased from 71% at the time of the pre-test to 87% at the post-test.

Next, in order to get a better idea of the effects of the training, a paired samples *t*-test was performed to examine whether there was a significant change from pre-test to post-test in each group. The results of this analysis indicated that the average post-test scores across the three groups were significantly higher than the average of the pre-test scores. A significant improvement of +6% was found in the control group ($t=2.316$, $p<.05$); +12% in Trained Group A ($t=3.510$, $p \leq .001$); and +16% in Trained Group B ($t=5.161$,

$p<.001$). The highest improvement was observed in Trained Group B, followed by Trained Group A and then lastly the control group. These results show that the post-test identification performance of all groups improved significantly from the pre-test period, indicating that all participants (including the untrained participants) were able to identify the target contrasts in the post-test to a degree that was more significantly accurate than at the time of the pre-test.

A lack of time and limited numbers of participants made it difficult to strictly control the participants' proficiency level. As can be seen from the results displayed in the figure above, there seems to be a difference in average pre-test scores among each group. In order to thus investigate whether there was a significant difference among groups, a one-way ANOVA was conducted. The result of this analysis revealed that there was no significant difference between the control group, Trained Group A and Trained Group B [$F=1.242$, $p>.05$]. This result suggests that the perception performance of the three groups did not differ significantly prior to training and that other differences found among the three groups at the time of the post-test can be attributed to training.

Next, to further explore whether the mean identification accuracy of post-test scores differed in each group or not, an ANCOVA[23] was performed using the pre-test score as the independent variable, differing groups and sounds as the factor and the post-test as the dependent variable. This analysis revealed that group had a significant effect on the average post-test score after controlling the effect of the pre-test [$F(2,117)=9.634$, $p<.001$] and a significant effect on sound [$F(3,117)=2.732$, $p<.05$]. In other words, the post-test identification accuracy scores were significantly different across the three groups and the identification accuracy scores of each four contrasts were not equivalent. Since all groups did not perform equally in identifying the target

[23] An analysis of covariance (ANCOVA) on post-test scores, with pre-test scores as co-variate usually provides a more appropriate and informative analysis for pre-test/post-test studies (Dugard & Todman, 1995).

sounds, follow-up Scheffe multiple comparisons were further performed to see which groups showed a difference in identification performance at the post-test. The results showed that there was a significant difference between the control group and Trained Group A [*p*<.05]; the control group and Trained Group B [*p*<.001]; but not between Trained Group A and Trained Group B [*p*>.05]. These results indicate that post-test identification accuracy scores of the trained groups were equal and, after the training, both of the two trained groups significantly outperformed the control group when identifying Japanese fricative and affricate contrasts (control group: 69%, Trained Group A: 79%; Trained Group B: 87%).

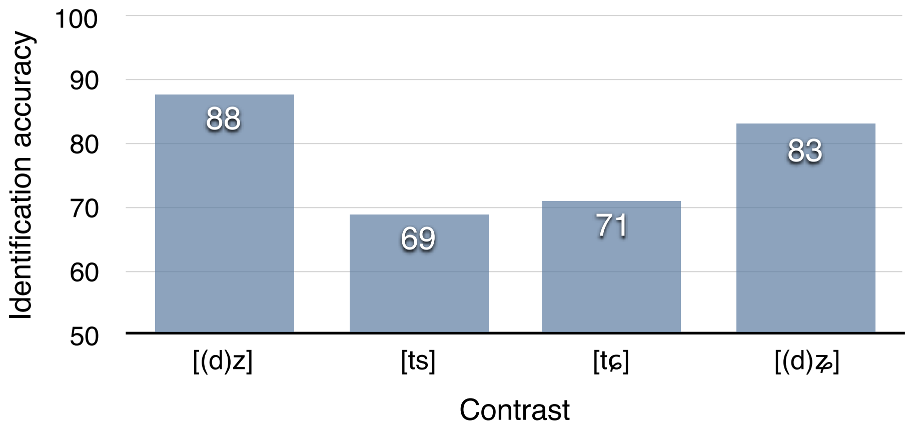


Figure 4.2 Mean percentage of correct identification scores of Japanese fricative and affricate contrasts as classified by contrasts at the *post-test.*

Next, since sound was also found to have a significant effect in the ANCOVA analysis mentioned above, follow-up Scheffe multiple comparisons on sounds were conducted to see which contrast showed a difference in identification accuracy during the post-test. The results revealed that in the post-test [(d)z] was scored with an average of 88% accuracy and [ts] with an average of 69% accuracy; [(d)z] with an average of 88% accuracy and [tɕ] with an average of 71% accuracy; [ts] with an average of 69% accuracy and [(d)ʑ] with an average of 83% accuracy differed significantly [*p*<.05]. Figure 4.2 indicates that [(d)z] appeared to be the easiest sound to perceive, when compared to the other sounds with a mean average of 88% identification accuracy, followed by [(d)ʑ] with a mean average of 83% identification accuracy. On the contrary, [ts] and [tɕ], with mean accuracy scores of 69% and 71% respectively, together showed the lowest mean averages in accuracy and are the most confusing contrasts for Thai learners to identify at post-test.

In summary, all groups were significantly better at identifying the target contrasts. However, Trained Groups A and B exhibited significantly higher average post-test scores than the control group (*p*<.000). And the untrained group improved significantly to a much lesser extent when compared to the trained groups. A difference in identification performance between Trained Groups A and B was not found, indicating that the trained participants could identify the target sounds to the same extent after the training regardless of their group, or to be specific, their training type. Moreover, regarding the average post-test scores as seen by contrast, [(d)z] and [(d)ʑ] showed relatively high accuracy scores. On the other hand, [ts] and [tɕ] seemed to still be challenging sounds to accurately identify for Thai learners even after the training.

4.2.2 Training effects as analysed by consonant contrasts

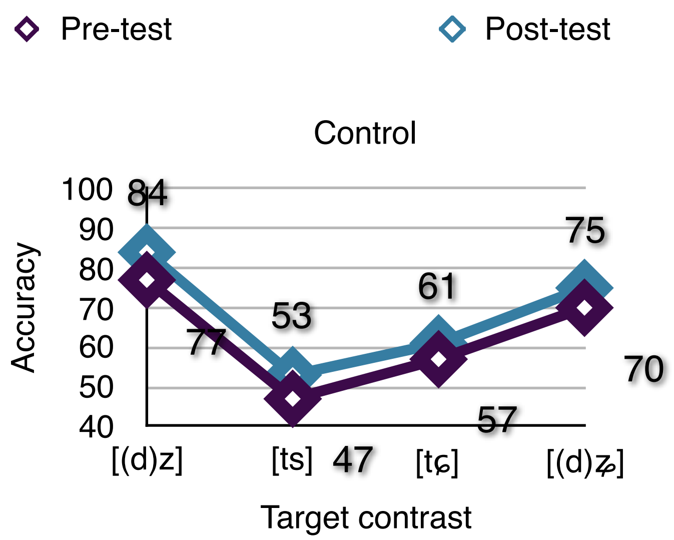
This section will analyse participants' mean percentage of correct identification scores separately for each consonant contrast with the aim to further investigate the degree of how participants performed differently across different contrasts. The mean accuracy scores of the pre-test and post-test of each of the three groups have been broken down to four individual sounds and are presented in Table 4.1.

Table 4.1 Paired samples *t*-test results of within-group comparisons between pre-test and post-test for each contrast.

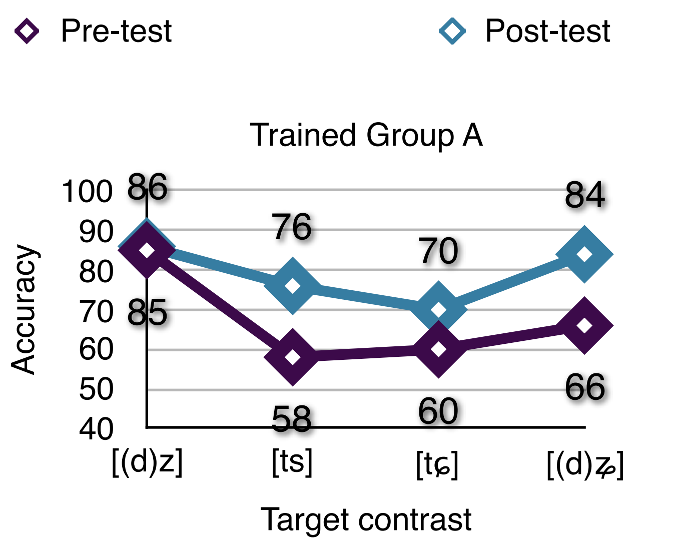
Group	Contrast	Pre(SD)	Post(SD)	%Difference	*t* (Sig)
Control (*n*=11)	[(d)z]	77 (28)	84 (23)	+7	1.604 (*p*=.140)
	[ts]	47 (22)	53 (21)	+6	1.491 (*p*=.167)
	[tɕ]	57 (20)	61 (17)	+4	1.000 (*p*=.341)
	[(d)ʑ]	70 (15)	75 (20)	+5	.690 (*p*=.506)
Trained Group A (*n*=11)	[(d)z]	85 (15)	86 (17)	+1	.289 (*p=.779*)
	*[ts]**	58 (16)	76 (17)	+18	3.975 (*p*=.003)
	[tɕ]	60 (24)	70 (14)	+10	1.218 (*p*=.251)
	[(d)ʑ]	66 (26)	84 (23)	+18	2.142 (*p*=.058)
Trained Group B (*n*=9)	[(d)z]	82 (26)	93 (11)	+11	1.955 (*p*=.086)
	[ts]	81 (15)	81 (11)	+0	.000 (*p*=1)
	*[tɕ]***	49 (13)	82 (14)	+33	6.532 (*p*=.000)
	*[(d)ʑ]***	72 (12)	92 (13)	+20	5.292 (*p*=.001)

* Means level of significance: * = $p<.05$, **= $p\leq.001$

(a) Control Group



(b) Trained Group A



(c) Trained Group B

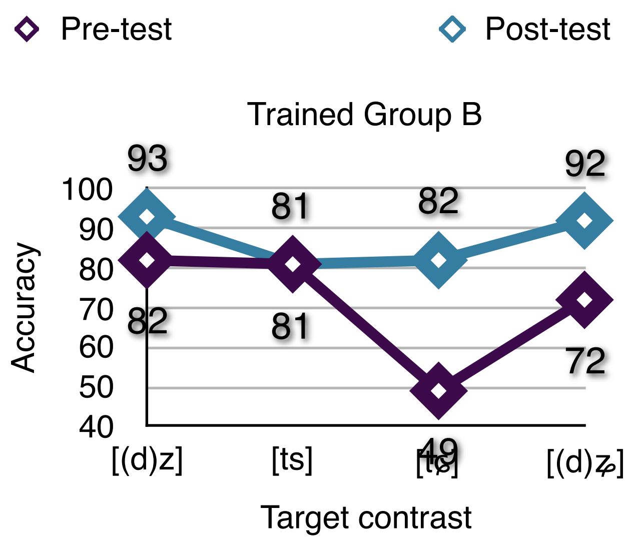


Figure 4.3 Mean percentage of correct identification scores for four contrasts [(d)z, ts, tɕ, (d)ʑ] of (a) Control Group, (b) Trained Group A and (c) Trained Group B.

A series of paired *t*-tests were performed on each individual contrast for each group. First, the analysis of the two trained groups will be discussed and then the control group. The results of the *t*-test demonstrated that the greatest improvement in Trained Group A was seen in the [ts] and [(d)ʑ] contrasts. However, there was a significant improvement observed only in [ts]. Initially, [ts] had the lowest mean correct scores when compared with other contrasts with a mean of 58% in the pre-test. However, this score increased to 76% accuracy after receiving training ($p<.05$). In comparison to [ts], there were no significant improvements observed in the identification of [(d)z], [tɕ] and [(d)ʑ]. However, despite a lack of significant improvement, it is worth looking closely at each contrast. Firstly, for the [(d)ʑ] contrast, an 18% increase occurred, which, while impressive, was not considered a significant improvement in itself. However, a significant trend of improvement was observed ($p<.1$). The results show that this is because of the great variations in individual performance, as is visible in the high standard deviation values showing that some participants might have improved greatly whereas other participants showed the opposite trend. Next, for [tɕ], participants made relatively good gains in [tɕ] identification with a 10% increase. Lastly, for [(d)z], there was nearly no improvement made from pre-test to post-test. In summary, the results indicate that Trained Group A participants' identification ability significantly improved only in identifying the [ts] contrast but also showed a relatively high improvement in [(d)ʑ] and [tɕ]. In contrast with other sounds, the identification of [(d)z] showed no improvement.

Next, for Trained Group B, there was a significant improvement found in [tɕ] and [(d)ʑ] ($p \leq .001$). The greatest improvement from the pre-test to post-test was observed especially in [tɕ] with a +33% increase followed by a noticeable improvement in [(d)ʑ] with a 20% increase. For the [(d)z] sound, there was a significant trend of improvement found with an increase of +11% ($p < .1$). In contrast to other sounds, there was no improvement seen in [ts]. One explanation for the lack of gains made in [ts] may be due to a ceiling effect. At pre-test, Trained Group B accurately identified [ts] at a good level when compared with other contrasts with an average of 81% accuracy so it might be difficult to see any improvement consequent to that initial high level of accuracy. Overall, these results clearly indicate that Trained Group B improved their ability to better identify [tɕ], [(d)ʑ] and [(d)z] after they participated in the training. The average post-test scores showed that Trained Group B's participants performed well at identifying all contrasts since the identification scores exceeded 80% accuracy in all contrasts.

By contrast, mean identification accuracy scores for the control group increased only slightly across consonant contrasts with an increase of +4% to +7%. All contrasts failed to yield significant improvements from pre-test to post-test. In other words, the untrained participants' identification ability did not change significantly over time. Although the results from 4.2.1 showed that there was a significant improvement in the control group, when further analysed by looking closely at each contrast there was no significant improvement in any contrast made by the control group. In other words, these results indicate that the untrained participants performed similarly in identifying each contrast at around the same level of accuracy from the pre-test through to the post-test.

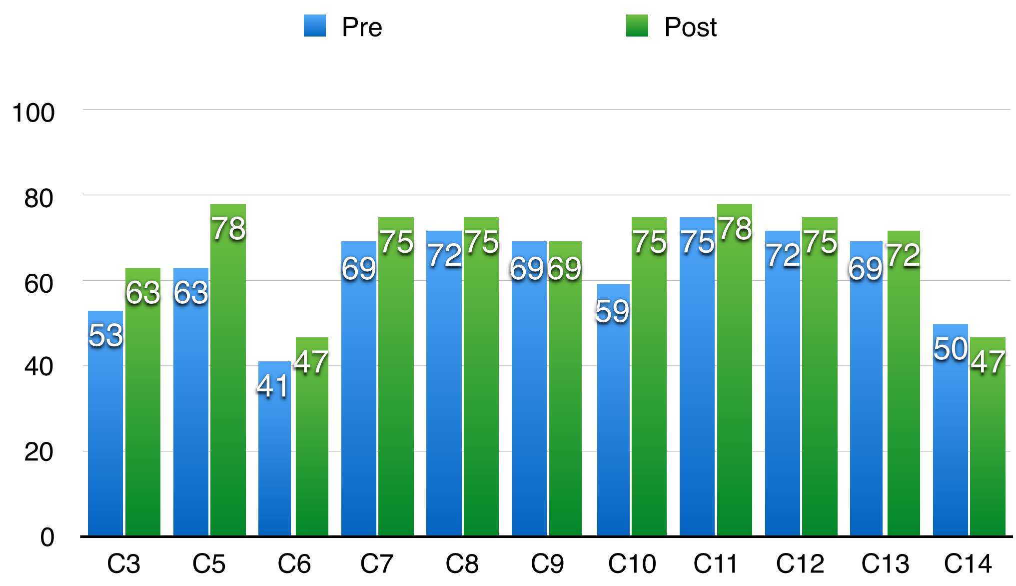
From the results overall, it can be said that, firstly, the trained groups achieved significant improvements in identifying certain sounds but not others. It must be noted that the improvements made across contrasts were not consistent for both groups i.e. Trained Group A improved significantly in their perception of some contrasts whereas Trained Group B improved significantly in their perception of separate contrasts. Nevertheless, it seems that some contrasts are easier to perceive and to improve than other contrasts. To be specific, those trained participants assigned to Trained Group A made gains in identifying [ts], [tɕ] and [(d)ʑ] but not [dz]. On the other hand, participants in Trained Group B made gains in identifying [(d)z], [tɕ] and [(d)ʑ] but not [ts]. A possible explanation of these differences in outcome could be as stated in Golestani & Zatorre (2009). They state that observed individual differences in non-native speech sound learning are “due to differences in performing the different tasks which differentially contribute to the identification task, discrimination task and training. For example, it is possible that individual differences in working memory capacity differentially influence performance on identification and discrimination tasks (p. 65)”. Nonetheless, it is possible that the difference in training type or the self-monitoring discrimination task which was undertaken only by Trained Group B may also have had an additional effect on identification ability. A more detailed explanation about individual differences in learning will be discussed again in Chapter 5. Secondly, whereas the trained groups showed some significant improvement in identifying specific contrasts, the control group made no significant improvements in identifying any contrast.

4.2.3 The individual participants' perceptual performance

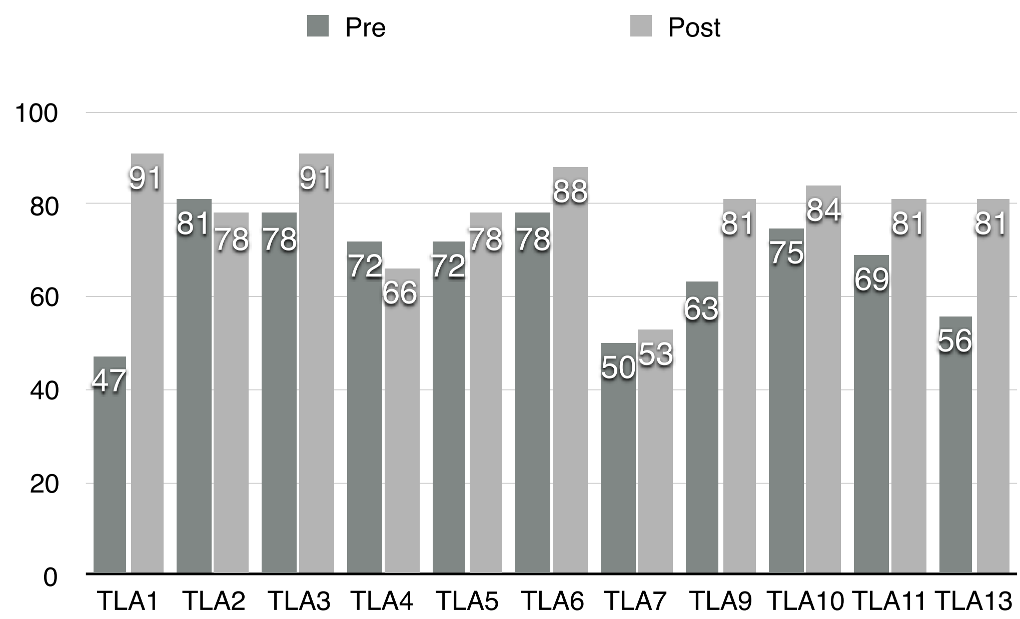
Since the number of participants was rather small, and as Thomson & Derwing (2014) state that "qualitative analyses should be conducted to provide insights into individual differences in learning" (p. 3), it is worth looking at the results for each of the individuals in the study. The next three figures (Figure 4.4 (a)-(c)) show the pre-test and post-test mean percentage scores of correct identification for each individual participant by group.

Figure 4.4 Individual participants' mean percentage of correct identification scores for the pre-test and post-test for (a) Control group, (b) Trained Group A and (c) Trained Group B.

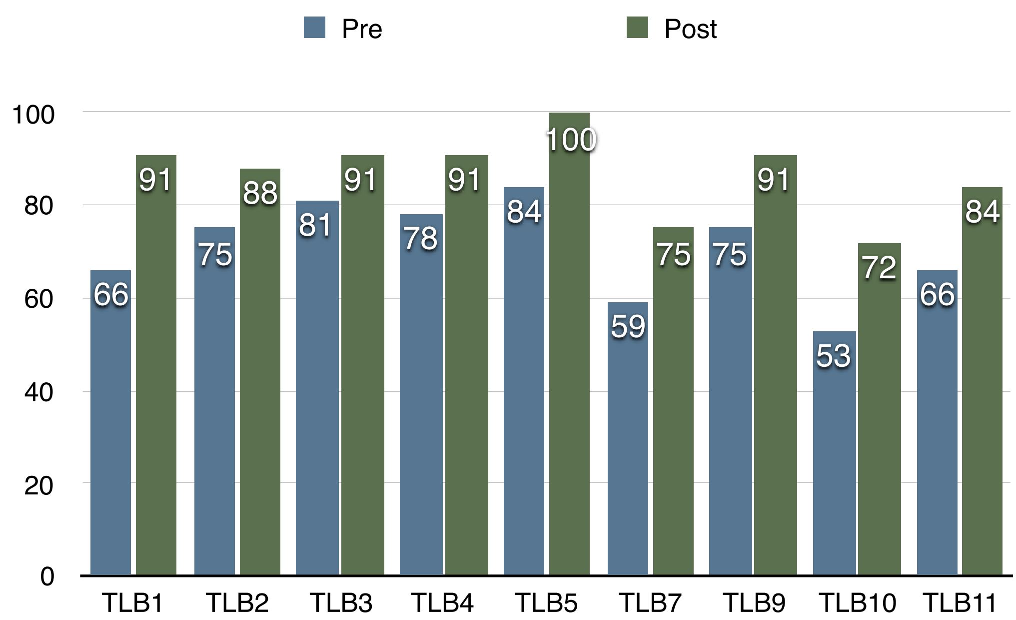
(a) Control Group



(b) Trained Group A



(C) Trained Group B



As seen in Figure 4.4, the degree of improvement from the pre-test to the post-test of each participant is clearly different. Details are explained below, group by group.

- Control Group -

As seen in Figure 4.4 (a), for the control group, the difference between the pre-test and post-test scores ranged from -3% to +16% with a mean of 6% change across all participants. A high degree of improvement was seen only in participants C3, C5 and C10 who managed more than a +10% increase. All other eight participants' identification scores slightly increased at about +6% or less. C19's identification performance remained the same and C14 demonstrated a decrease in identification scores.

- Trained Group A -

The greatest variability of individual differences in participants' improvement in identification accuracy was observed in Trained Group A. As seen in Figure 4.4 (b) the size of the improvement varied from -6% to +44% with a mean change of +12%. Although the figure shows that, overall, the majority of participants of Trained Group A improved in identifying the target contrasts, TLA2 and TLA4 did not make any improvements after the training. In fact, TLA2 and TLA4 showed a decrease in identification scores of -3% and -6% respectively in the post-test. A ceiling effect may have influenced the lack of improvement for them, because both TLA2 and TLA4 had relatively high pre-test scores - TLA2's identification accuracy was the highest in the group (TLA2: 81%; TLA4: 72%). Some other participants made notably large improvements from pre-test to post-test; TLA1's identification accuracies increased greatly by about +44%, TLA13 by +25% and TLA9 by +18%. The large gains made by TLA1 and TLA13 can possibly be accredited to the fact that they had relatively low pre-test scores when compared with other participants, and so they had more room for improvement (TLA1 scored 47%

and TLA13 scored 56% at pre-test). Participants TLA6, TLA10 and TLA11 showed improvements ranging from +9 to +12%. The rest of the participants of Trained Group A (TLA5 and TLA7) showed similarly little gain from the training. This wide range of individual performance is consistent with previous training studies which reported that the gains of training tend to be spread unevenly across participants (Goto, 1971; Sheldon & Strange, 1982; Yamada, 1994; Bradlow et al., 1997; Burnham, 2013).

- Trained Group B-

A close examination of individual participants' data in Figure 4.4 (c) showed that the difference between pre-test and post-test scores for individual learners in Trained Group B ranged from +10% to +25% with a mean of a +16% increase. All participants in Trained Group B showed a steady and consistent improvement and no decreases were found in identification accuracy when compared to Trained Group A and the control group. In other words, these results indicate that all participants in Trained Group B benefited from the training. Even though some participants had relatively high pre-test scores with an average of more than 70% (such as TLB2, TLB3, TLB4, TLB5 and TLB9) their identification accuracies still increased relatively highly by more than +10%. Five participants (TLB1, TLB3, TLB4, TLB5 and TLB9) showed higher than 91% identification accuracy, most notably TLB5 who reached 100% identification accuracy after the training, indicating that there are learners who can achieve native-like perception accuracy.

In summary, from the analysis of individuals it is clear that the degree of improvement through the course of testing and training varied across the participants. The results from Trained Group A indicate that perceptual training did not result in increased perception performance for all participants. There was a relatively high degree of individual difference observed in Trained

Group A. On the contrary, Trained Group B showed more steady and consistent improvements in identification accuracy, indicating that all participants from Trained Group B clearly benefitted from the training. The explanation for the individual differences in gains may be found in the differences of specific task components (Golestani & Zatorre, 2009). These results imply that the difference in training type used in this study may result in different learning effects. Judging from the results discussed here, it could be posited that the training conducted on Trained Group B might have contributed to more equal gains across the participants. However, it is still too early to make such conclusions since there are many factors, such as motivation or phonological memory, that could also have affected their individual performances (Moyer, 1999; Flege, 2003; Hu et al., 2013; Colantoni et al., 2015). Regardless, further study is required to identify what factors caused this phenomenon.

4.3 Generalisation test results

One further goal of this study is to examine the effects of perceptual training, specifically concerning whether any improvement gained in perception has been transferred to untrained words and untrained talkers.

4.3.1 Differences between a familiar talker and an unfamiliar talker for the same post-test

The first piece of analysis undertaken concerning the generalisation of perceptual learning was to look at the post-test, which was performed again using a talker hitherto unfamiliar to the participants, to see whether the trained participants' performance when listening to an unfamiliar talker who was not included in the training was comparable to their performance when listening to a previously familiar talker.

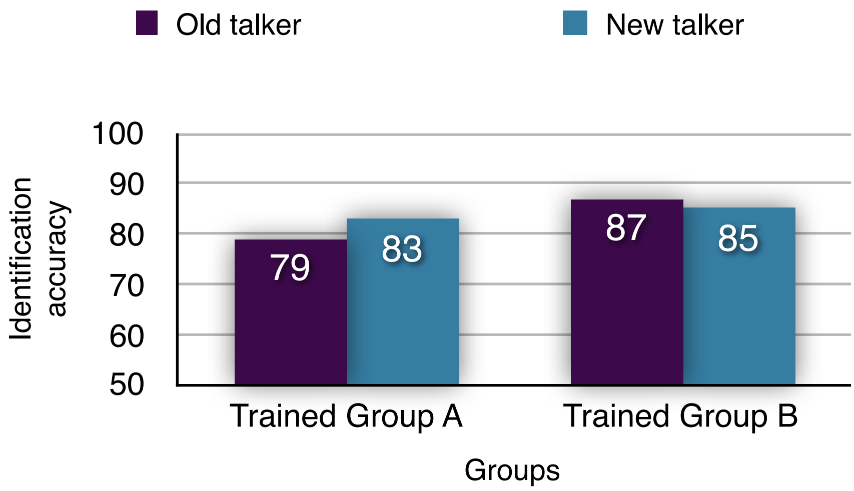


Figure 4.5 A comparison of mean percentages of correct identification scores from the post-test conducted with trained groups when stimuli were given by a familiar talker and an unfamiliar talker.

Figure 4.5 displays the mean percentages of correct identification scores from the post-test when productions are given by a familiar talker versus an unfamiliar talker. In order to investigate whether there is a significant difference between the familiar talker whose voice was used in the training and the unfamiliar talker whose voice was not used in the training, a paired samples *t*-test on the mean scores of the trained participants' post-tests was conducted. The results revealed that there was no significant difference between the talkers (t=.519, p>.05), indicating that the perceptual accuracy of the trained participants when listening to the familiar talker did not differ from their accuracy when listening to the unfamiliar talker who was not used in the training. The mean accuracies are detailed as follows; for Trained Group A - familiar talker = 79%, unfamiliar talker = 83%; and for Trained Group B - familiar talker = 87%; unfamiliar talker = 85%. Overall, the results indicate that the perceptual ability of the trained participants to identify the pre/post-test tokens produced by a new talker was the same as their ability to perceive tokens produced by the familiar talker who was used in the training.

4.3.2 Generalisation tests of new words and an unfamiliar talker

In order to thoroughly investigate whether the perceptual accuracy gained from the training had been transferred to new words and an unfamiliar talker that were not presented during training, an additional two generalisation tests were conducted. The two generalisation tests consisted of; a test using words with a low level of lexical familiarity produced by a familiar talker and an unfamiliar talker; a test using words with a high level of lexical familiarity produced by a familiar talker and an unfamiliar talker.

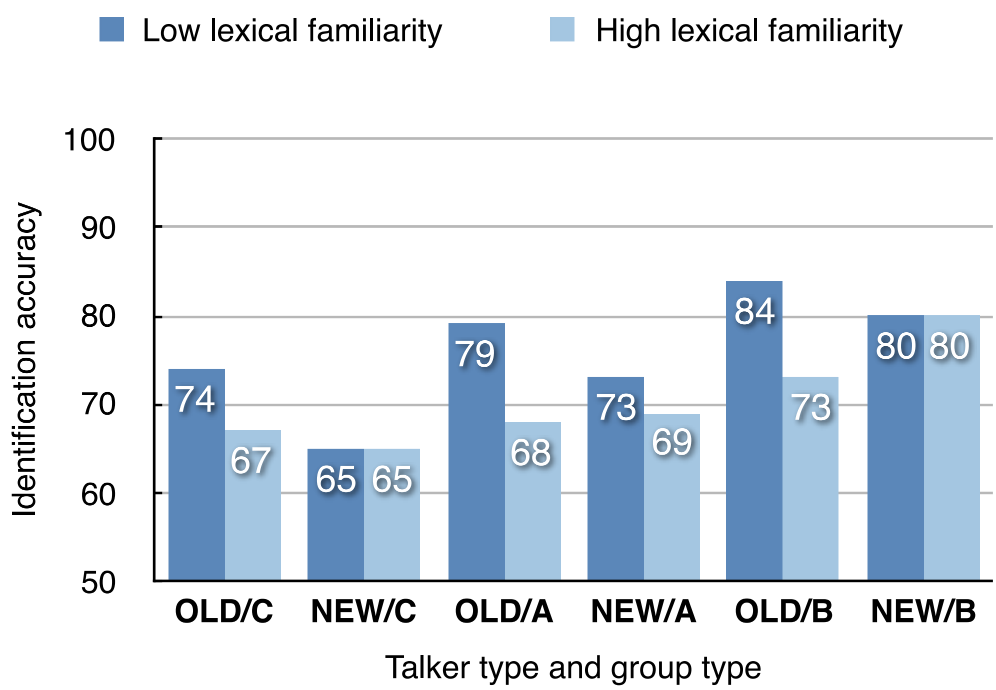


Figure 4.6 Mean percentage of correct identification scores of Japanese fricative and affricate contrasts from four generalisation tests undertaken by Control Group (C), Trained Group A (A) and Trained Group B (B). "Old" refers to words spoken by a familiar talker used in the training; "new" refers to words spoken by an unfamiliar talker not used in the training.

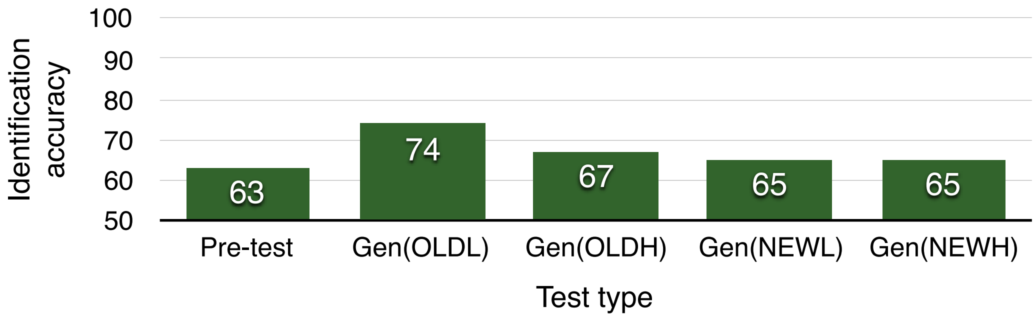
Figure 4.6 displays the mean identification scores which were broken down into four generalisation types: Mean identification scores of new words of low lexical familiarity spoken by a familiar talker (1) and by an unfamiliar talker (2); mean identification scores of words of high lexical familiarity spoken by a familiar talker (3) and by an unfamiliar talker (4). Each of the four types were tested on the trained groups and the control group. As can be seen from Figure 4.6, overall, the control group's identification scores are lower than the trained groups for all tests. To see whether or not there was any significant difference among each group, test and talker an analysis of variance (ANOVA) with group (control group, Trained Group A and Trained Group B), test (two tests of low and high lexical familiarity) and talker (a familiar and an unfamiliar talker) as fixed factors was carried out. The results of this analysis revealed that group had a significant effect (F=12.395, p<.001) as did test (F=8.228, p<.05), indicating that each group performed differently and that different tests yielded different accuracies. Also, there was also a significant interaction between test and talker (F=5.459, p<.05).

Next, a Scheffe analysis was performed to see which group posed different mean scores. The results showed that there was a significant difference between the control group and Trained Group B (p<.001); and Trained Group A and Trained Group B (p<.05), suggesting that Trained Group B significantly outperformed both the control group and Trained Group A on the generalisation tests. Trained Group A and the control group performed similarly across the generalisation tests (p>.05).

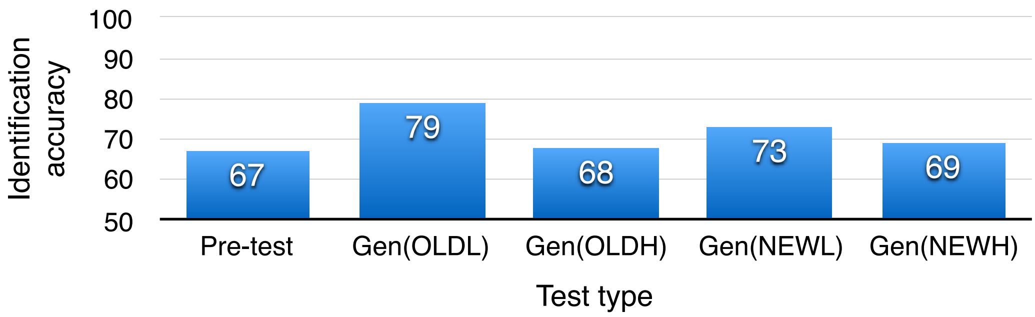
For the test, as can be seen from the figure, the low lexical familiarity test scores of each group were higher than the high lexical familiarity test (p<.05), indicating that words of low lexical familiarity are easier to identify than words of high lexical familiarity test for all groups. Regarding the difference in talkers, there was no significant difference found (p>.05), indicating that they performed equally well at identifying words produced by either a familiar or unfamiliar talker. However, there was a significant interaction between test and talker (F=5.459, p<.05). This can be seen in that all groups performed better when identifying the words of low lexical familiarity produced by a familiar talker rather than when produced by an unfamiliar talker.

The results showed that Trained Group B performed relatively well in all tests with a mean identification accuracy average of over 73%. This is especially true when considering the test that used words of low lexical familiarity given by the old talker, which was identified with a mean accuracy of 84%. Moreover, the mean accuracy scores of Trained Group B were significantly higher than the untrained control group and Trained Group A. Hence, it can be said that the perceptual learning of Trained Group B generalised through to the untrained words and to the unfamiliar talker. By contrast, in Trained Group A, the results showed that Trained Group A could not perform better in identifying generalisation test stimuli to a greater extent than the untrained control group, indicating that the perceptual learning of Trained Group A had not been generalised to new words and a new talker.

(a) Control Group



(b) Trained Group A



(c) Trained Group B

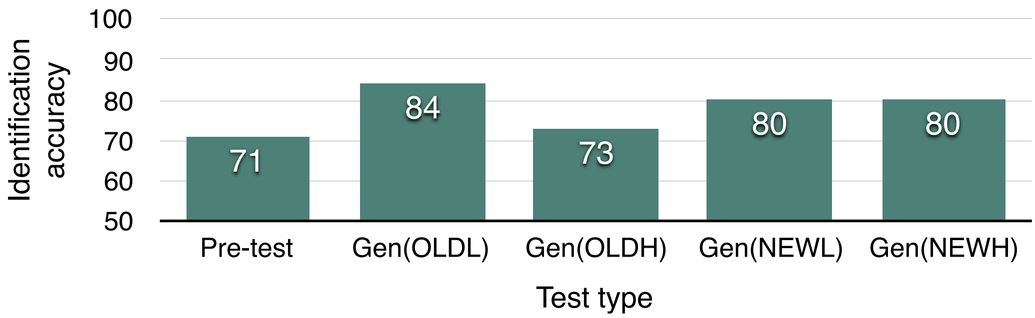


Figure 4.7 Mean percentage of correct identification of the Japanese fricative and affricate contrasts at pre-test and during four generalisation tests (low-familiarity words by a familiar talker (OLDL); high-familiarity words by a familiar talker (OLDH); low-familiarity words by unfamiliar talker (NEWL); high-familiarity words by an unfamiliar talker (NEWH)) for (a) Control Group, (b) Trained Group A and (c) Trained Group B.

Table 4.7 illustrates the change in identification accuracy scores from the pre-test to the four types of generalisation tests for each group. Generalisation of perceptual learning was assumed to occur when the generalisation test scores were significantly higher than the pre-test score. Accordingly, in order to further see whether perceptual learning from the training transferred to a new speaker and to new words or not using a different analysis, here, identification accuracies of the pre-test will be compared to the generalisation test scores of each group by conducting paired samples *t*-tests. The analysis of the paired samples *t*-tests comparing the pre-test scores to each of the four generalisation tests scores revealed that the control group and Trained Group A showed significantly higher scores in their generalisation tests than the pre-test only in the test of words of low lexical familiarity produced by a familiar talker (control: t=3.093, p<.05; Trained Group A: t=2.702, p<.05) but showed no significant difference in other tests (p>.05). By contrast, Trained Group B significantly showed higher mean scores in three out of four generalisation tests; the test of words of low lexical familiarity spoken by a familiar talker (t=4.636, p<.05), words of low lexical familiarity spoken by an unfamiliar talker (t=4.773, p≤.001) and words of high lexical familiarity spoken by an unfamiliar talker (t=5.234, p≤.001). In summary, apart from the case of words of high lexical familiarity spoken by a familiar talker, Trained Group B showed a significantly higher identification accuracy score than at pre-test in most of the generalisation tests, which suggests that the perceptual ability of the Japanese fricative and affricate contrasts of Trained Group B transferred to new words and an unfamiliar talker, while Trained Group A and the control group showed insufficient evidence to support such a generalisation.

In summary, as seen from the results above, it has been revealed that Trained Group B consistently outperformed all other groups, showing evidence that there was a transfer of perceptual improvement to untrained words and an untrained talker observed only in Trained Group B. For Trained Group A and the control group, apart from the test of words of low lexical familiarity produced by a familiar talker, the overall mean averages of identification accuracy from the generalisation tests were not significantly different to the pre-test, implying that perceptual learning had not been generalised to new words and new talkers in Trained Group A. Moreover, concerning the level of familiarity, the degree of word familiarity has been shown to influence the degree to which participants perceive accurately. All groups performed better at identifying words of low lexical familiarity than words of high lexical familiarity, and especially words of low lexical familiarity spoken by a familiar talker. However, it is not known what factors have affected these results. An explanation of lexical familiarity will be explored in detail later in Chapter Five.

4.4 Six-month-delayed post-test results

To examine whether the observed perceptual improvements of Japanese fricative and affricate contrasts from the given training were maintained without further training six months after the training had ended, a delayed post-test was presented to a total of thirty participants from all three groups. The contents of the delayed post-test were identical to the pre-test and post-test.

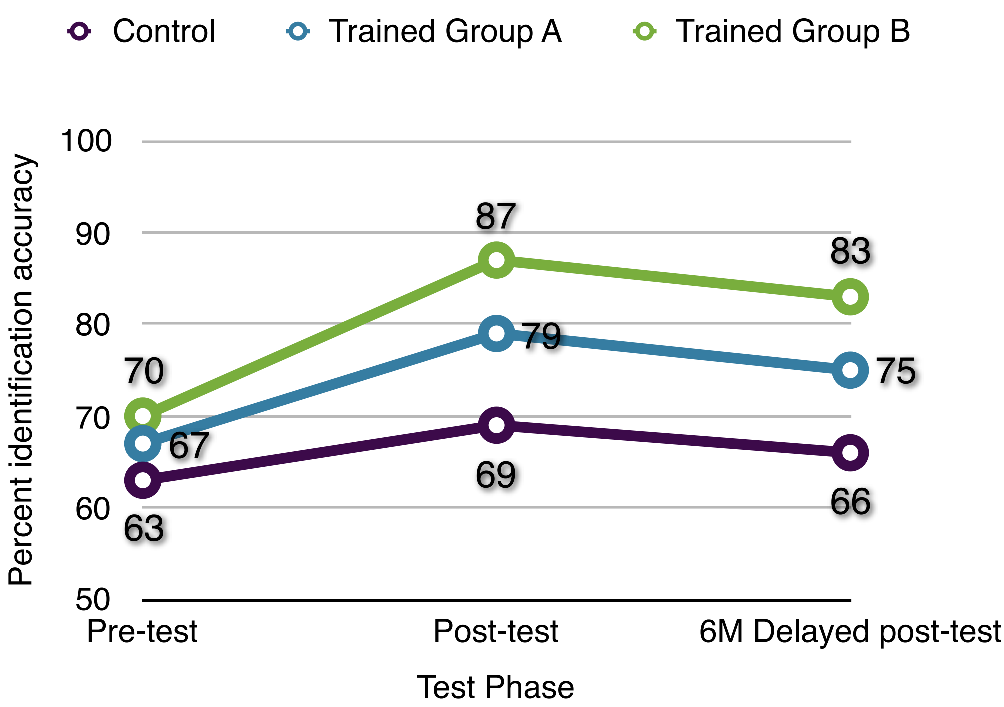


Figure 4.8 A comparison of the mean percentages of identification accuracy scores of Japanese fricative and affricate contrasts between group from pre-test, post-test and six-month follow-up delayed post-test.

The mean percentages of correct identification scores from the pre-test, the post-test and the six-month delayed post-test for each group are displayed in Figure 4.8. To compare the participants' perception performance from the six-month delayed post-test with their performance at pre-test and post-test, paired samples *t*-tests were conducted for each group.

- Control group -

This analysis revealed that there was a significant difference between the pre-test and the post-test (t=3.166, p<.05) but not between the pre-test and the delayed post-test; nor the post-test and the delayed post-test (p>.05). In other words, although there was an improvement seen between pre-test and post-test, the untrained participants' perception scores between pre-test and the delayed post-test, and the post-test and delayed post-test were not significantly different.

- Trained Group A -

For Trained Group A, there was a significant difference between the pre-test and the post-test (t=2.859, p<.05) and between the pre-test and the delayed post-test (t=2.600, p<.05), but no significant difference between the post-test and the delayed post-test (p>.05). These results indicate that identification ability differed significantly after receiving the training and the learning was maintained six months after even without further training.

- Trained Group B -

For Trained Group B, there was a significant difference between the pre-test and the post-test (t=10.673, p<.001) and between the pre-test and the delayed post-test (t=4.781, p<.05), but no significant difference between the post-test and the delayed post-test (p>.05). These results again indicate that identification ability differed significantly after receiving the training and that the learning gained was maintained six months after training.

In summary, as this figure shows, the mean identification accuracy of all three groups dropped slightly after six months from the post-test scores but still appeared to be comparable to the post-test in the trained groups (Trained

Group A: post-test (79%), the delayed post-test (75%); Trained Group B: post-test (87%), the delayed post-test (83%)). Moreover, the results also showed that the delayed post-test's mean scores were significantly higher than pre-test for both trained groups (Trained Group A: pre-test (67%), the delayed post-test (75%); Trained Group B: pre-test (70%), the delayed post-test (83%)), clearly indicating that there was a significant improvement maintained after the completion of the training. The control group showed no change in perceptual identification accuracy from the pre-test to the six-month delayed post-test. In other words, for the control group the identification ability during the delayed post-test went back to the same level as when they performed the pre-test (pre-test: 63%; the delayed post-test: 66%). To conclude, it can be posited that not only was the perceptual improvement of the trained participants visible after training but that this improvement was maintained at the post-test level even six months after the completion of perceptual training. These findings provide evidence that perceptual training produced long-term modifications in perception.

4.5 Production results

Another goal of the current study, to be examined in this section, was to explore the effects of perceptual training regarding whether any gained perceptual learning had been transferred to the production domain without any explicit training in production. Participants' production utterances were assessed in two ways: 1) A perceptual identification task in which each utterance produced by each participant was identified by five native Japanese speakers, 2) A goodness rating score in which each utterance was scored using a Likert 5-point scale. The production results will be divided into three parts as follows; 1) identification judgement; 2) goodness rating judgement and; 3) waveform and spectrogram inspection.

4.5.1 Production accuracy by an identification judgement

4.5.1.1 Judgement reliability

Five native Japanese speakers (see Appendix F) judged all participants' production utterances using two-alternative forced-choice identification tasks. Kendall's Coefficient of Concordance was performed for the productions of the pre-test and post-test to see whether all five Japanese native raters identified participants' productions reliably and consistently or not. The first analysis of the pre-test showed that the five raters tended to agree consistently with one another in their assessment of the target contrasts (W=.933, p<.001). The second analysis of the post-test also showed that all five raters evaluated to a consistently similar extent (W=.925, p<.001).

4.5.1.2 Overall pre-test and post-test performance

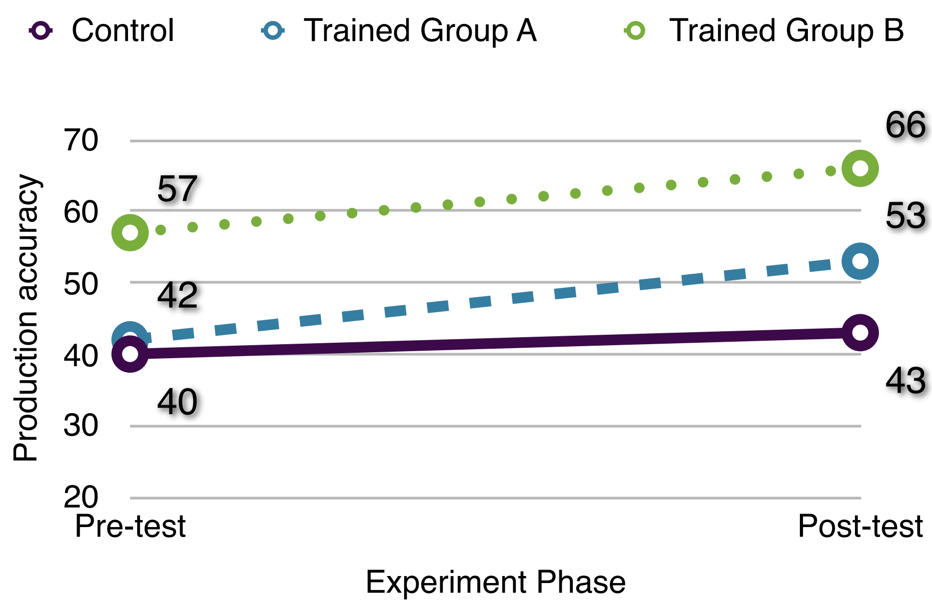


Figure 4.9 Mean percentages of production accuracy scores of Japanese fricative and affricate contrasts produced by Control Group, Trained Group A and Trained Group B.

Figure 4.9 displays the production accuracy of the Japanese fricative and affricate contrasts produced by all participants before and after the training, as judged by the five Japanese native raters. As the figure shows, overall, for the control group, the average pre-test scores were 40% and 43% for the post-test which showed a relatively modest +3% increase. For Trained Group A, the mean production accuracy scores increased from an average of 42% at the pre-test to 53% at the post-test. Lastly, for Trained Group B, the average accuracy scores increased from 57% at the time of the pre-test to 66% at the post-test. In other words, the highest improvement was observed in Trained Group A, followed by Trained Group B and then the control group.

Next, in order to get a better idea of the effects that the training had on production, a paired samples *t*-test was performed to examine whether there was a significant change in production accuracy from pre-test to post-test in each group. The results of the analysis indicated that average post-test scores were significantly higher than average pre-test scores only in the trained groups but not in the control group (t=1.176, p>.05). Specifically, a significant improvement of +11% was found in Trained Group A (t=3.257, p<.05); and a change of +9% in Trained Group B (t=2.626, p<.05). In other words, the results suggest that after nine sessions of training the trained participants made significantly better productions in their accuracy of Japanese fricative and affricate contrasts. By contrast, the production accuracy scores produced by the control group do not seem to have changed from pre-test to post-test.

A lack of time and numbers of participants made it difficult to strictly control the participants' proficiency level, and, as can be seen in Figure 4.9, there seems to be a difference among groups at pre-test. To see whether there was a significant difference among groups, a one-way ANOVA was conducted. The results of this analysis revealed that there was no significant difference between the control group, Trained Group A and Trained Group B (F=2.723, p>.05). In other words, these results suggest that the mean production performance of the three groups did not differ significantly prior to training and that other differences found among the three groups at post-test can be attributed to training.

Next, to further determine whether the mean production accuracy of post-test scores differed in each group or not, an ANCOVA was performed using the pre-test score as the independent variable, the group and sound as the factor and the post-test as the dependent variable. The analysis revealed that group had a significant effect on the average post-test score after controlling the effect of the pre-test ($F(2,116)=3.081$, $p\leq.05$) but sound did not ($F(3,116)=2.350$, $p>.05$), indicating that the average post-test production accuracies of the three groups were significantly different. Regarding the effect of sound, the production accuracies of the four target contrasts were equivalent across each sound. Since all groups did not perform equally in producing the target sounds, follow-up Scheffe multiple comparisons were further performed to see which groups posed a difference for production accuracy at post-test. The results showed that there was a significant difference in production accuracies only between the control group and Trained Group B ($p<.05$) but not between the control group and Trained Group A ($p>.05$) and also between Trained Groups A and B ($p>.05$), clearly demonstrating that Trained Group B's production accuracies (66%) at the post-test were significantly higher than the untrained control group (43%).

To summarise, the results clearly indicate that there was evidence showing that the perceptual learning gained through perceptual training was transferred to improvements in the production domain as judged by five Japanese native speakers.

4.5.1.3 Training effects as analysed by consonant contrasts

This section will analyse participants' mean production accuracy scores separately for each consonant contrast. The results of the ANCOVA above showed that differing sounds had no significant effect. It is, however, worth investigating the degree to which participants performed differently across the contrasts. As such, the three groups mean percentage of production accuracy from the pre-test and post-test scores, as judged by the five Japanese native speakers, were broken down into four individual contrasts and are presented in Table 4.2.

Table 4.2 Paired samples *t*-test results of within-group comparisons between pre-test and post-test for each contrast.

Group	Contrast	Pre(SD)	Post(SD)	%Difference	*t* (Sig)
Control (*n*=11)	[(d)z]	58 (43)	64 (44)	+6	1.093 (*p*=.300)
	[ts]	27 (43)	27 (41)	+0	.256 (*p*=.803)
	[tɕ]	38 (25)	38 (28)	+0	.126 (*p*=.902)
	[(d)ʑ]	42 (25)	47 (27)	+5	.709 (*p*=.495)
Trained Group A (*n*=11)	[(d)z]	51 (37)	67 (35)	+16	1.776 (*p*=.106)
	*[ts]**	19 (25)	34 (30)	+15	3.975 (*p*=.003)
	[tɕ]	38 (21)	41 (22)	+3	.404 (*p*=.695)
	[(d)ʑ]	61 (28)	71 (24)	+10	1.624 (*p*=.136)
Trained Group B (*n*=9)	[(d)z]	74 (37)	84 (29)	+10	1.348 (*p*=.215)
	[ts]	45 (41)	54 (36)	+9	1.393 (*p*=.201)
	[tɕ]	46 (26)	53 (30)	+7	.769 (*p*=.464)
	[(d)ʑ]	63 (23)	73 (23)	+10	1.998 (*p*=.081)

* means level of significance: * = p<.05, **= p<.001

A series of paired *t*-tests were performed for each group on each individual contrast to see whether there was a significant change in production accuracy from pre-test to post-test for each contrast. As the figure shows, first, for the control group, the results of the *t*-test demonstrated that there was no significant improvement in production accuracy from the pre-test observed in any contrast. In particular, [ts] and [tɕ] contrasts showed no change in production improvement at all. For the [(d)z] and [(d)ʑ] contrasts, there was a modest improvement found with an increase of +5 to +6%.

Next, for Trained Group A, a significant improvement from the pre-test to post-test was found significantly for [ts] with a 15% increase (t=3.975, p<. 05). This contrast showed the lowest average pre-test production accuracy but showed the greatest improvement from 19% to 34% in the post-test. Trained Group A participants made the greatest improvement of a +16% gain in the production of [(d)z] but there was no significant difference observed. The lack of significant improvement found in [(d)z] is possibly because of great variations in individual performance as seen from the high standard deviations values. In other words, some participants might have improved greatly whereas other participants showed the opposite trend or made no gains at all. Next, an improvement of +10% was found in [(d)ʑ]. In contrast with other sounds, [tɕ] showed only a tiny improvement with only a +3% increase from pre-test.

For Trained Group B, there was no significant improvement in the pronunciation of the fricative and affricate contrasts. However, it is worth examining the results closely even though there was no significant improvement found. First, for the [(d)ʑ] contrast, a large improvement of 10% points occurred but only a significant trend of improvement was observed (p<. 1). In addition, although there was no significant improvement found, the results showed that overall production accuracies of other contrasts improved to a similarly relatively large extent and very little variation across contrasts was found with an increase of +10% for [(d)z], 9% for [ts], and +7% for [tɕ].

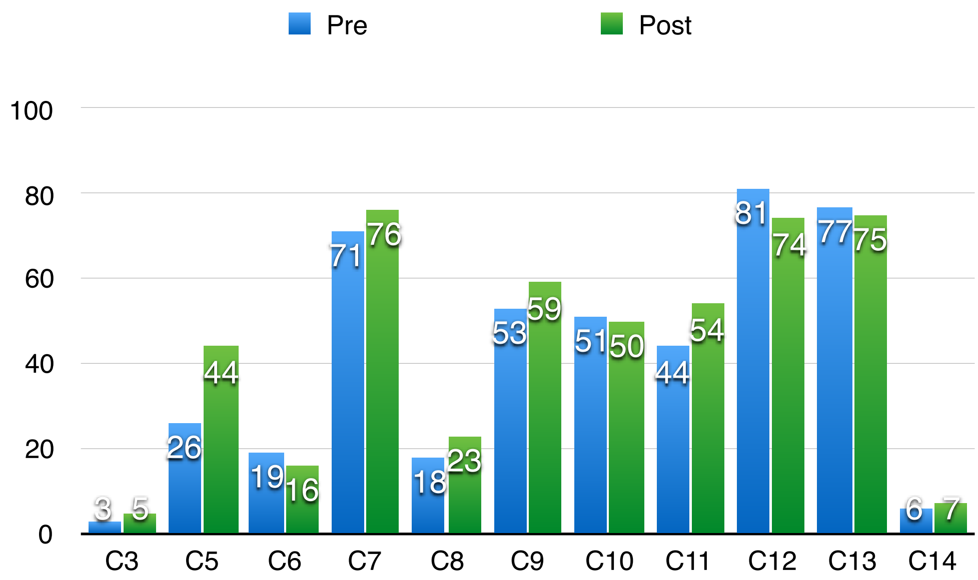
Overall, from the results it can be said that 1) there was a significant improvement only in Trained Group A for [ts] and a significant trend of improvement in Trained Group B for [(d)ʑ]. However, 2) regardless of the fact that there was a lack of significant improvements for specific contrasts, the trained groups overall showed higher improvements than the control group in producing the target contrasts.

4.5.1.4 The individual participants' production performance

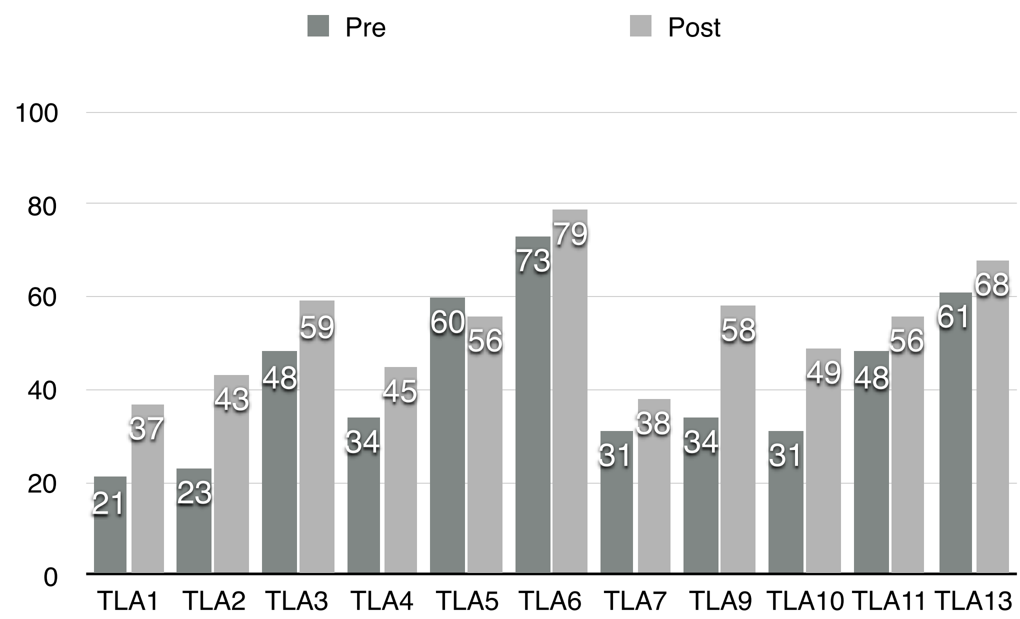
Since the number of participants was rather small, it is worth looking at the results for each of the individuals in the study. Further analyses of this data will examine productions for individual participants. Individual participants' mean percentage of correct production scores as judged by the five Japanese native speakers at pre-test and post-test are presented in Figure 4.10 (a)-(c).

Figure 4.10 Individual participants' mean percentage of correct production scores from the pre-test and post-test in (a) Control Group, (b) Trained Group A and (c) Trained Group B.

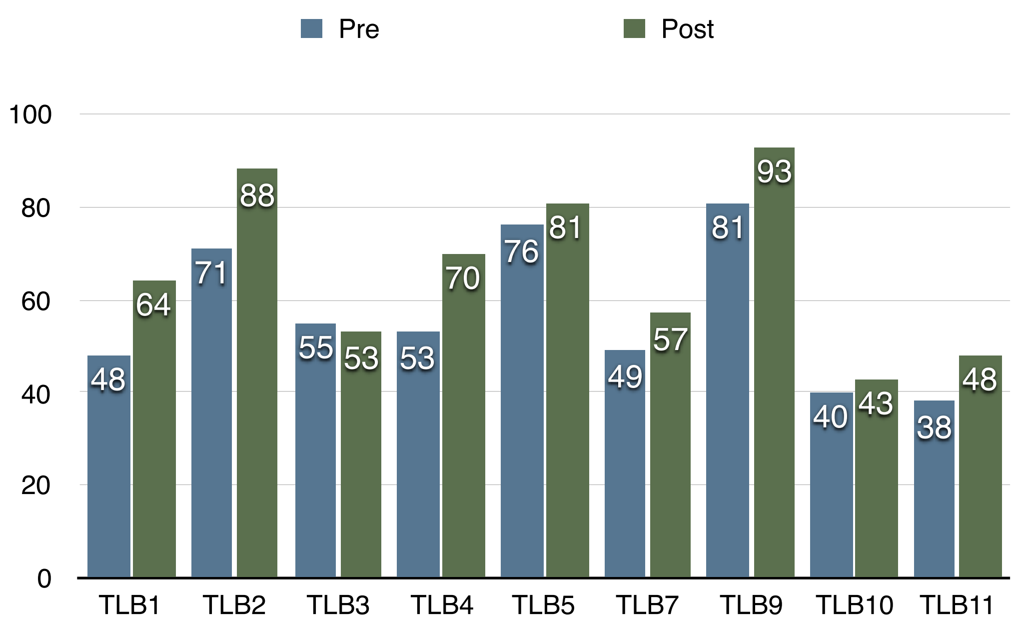
(a) Control group



(b) Trained Group A



(c) Trained Group B



As seen in Figure 4.10, the degree of improvement of the target sounds for each participant is clearly different. Details are explained by each group below.

- Control Group -

As displayed in Figure 4.10 (a), the difference between pre-test and post-test production scores in individual learners ranged from -7% to +18% with a mean of 3% change. C5 and C11 showed a relatively moderate increase of +18% and +10%, respectively. The rest of the participants in this group showed more modest improvements from about +1% to a +5% increase. Four participants showed a decrease in production accuracy.

- Trained Group A -

As seen from Figure 4.10 (b), a close examination of individual participants' data showed that the size of the gain varied from -4% to +24% with a mean score change of 11%. TLA2 and TLA9 gained noticeably large improvements from the pre-test to the post-test (TLA2: +20%; TLA9: +24%). Moderate improvements of +11% to +18% were seen in TLA1, TLA3, TLA4 and TLA10. Lesser of improvements were made by TLA6, TLA7, TLA11 and TLA13, who improved by about +6% to +8% in producing the target sounds. Contrary to other participants in the group, only TLA5 showed a decrease in accuracy of -4% in producing the target sounds from the pre-test.

- Trained Group B -

The difference between pre-test and post-test production scores in individual learners ranged from -2% to +17% with a mean of 9%. Some participants performed relatively well even before the training such as TLB2, TLB5 and TLB9. They produced the target contrasts with more than 70% accuracy, indicating that a ceiling effect may have influenced the lack of greater improvements in production performance. The improvement trend in

Trained Group B was specifically observed as follows. First, a relatively moderate +10% to +17% increase trend was observed mostly in this group (TLB1, TLB2, TLB4, TLB9 and TLB11). The rest of the participants shown an improvement ranging from a +3 to +8% increase. In contrast to other participants, TLB3 showed a decrease in production accuracy with a -2% decrease.

In summary, when analysis focused on individual differences it is clear that there was considerable individual variation across participants in the change of mean production accuracy scores from pre-test to post-test. Overall, most participants in the trained groups improved their production performance, showing a moderate change. However, some participants did not make any gains from the perceptual training (TLA5 and TLB3). Some participants such as TLA7 and TLA10 performed at similarly low levels (31%) in the pre-test; however, at post-test, TLA10 (49%) performed considerably better than TLA7 (38%). This wide range of individual performances is consistent with previous findings (e.g., Bradlow et al., 1997; Hazan et al., 2005). However, as suggested in previous studies, it is unclear what factors determine individual gains in training. More research is necessary in order to answer the questions posed by these unequal gains that have been found and in order to see what factors have caused this phenomenon.

4.5.2 Goodness rating

The aim of this subsection is to evaluate via different means whether there has been an improvement in production performance from pre-test to post-test and to see whether the results obtained in this section are in line with the production accuracy results from the previous section (4.5.1). Goodness rating data were obtained using the same five Japanese raters from the previous production accuracy judgement.

4.5.2.1 Overall pre-test and post-test rating scores

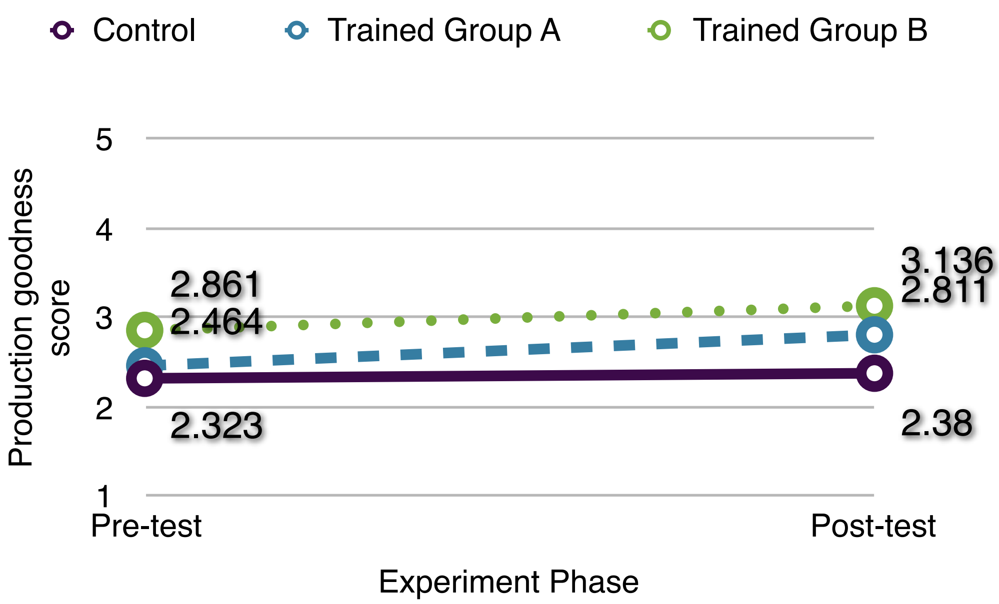


Figure 4.11 Mean of goodness rating scores evaluated by five native Japanese raters. '1' represented a poor exemplar of the target consonant or "heavily accented or not Japanese-native-like", '5' represented a good Japanese-sounding or "native-like or not accented at all" exemplar.

Figure 4.11 illustrates the average pre-test and post-test goodness rating scores for all three groups in which five Japanese native speakers directly rated all Thai participants' pre-test and post-test utterances. The rating score was conducted on a Likert 5-point scale. '1' represents a poor exemplar of the target consonant or "heavily accented or not Japanese native-like", '5' represents a good Japanese-sounding or "native-like or not accented at all" exemplar. As this figure shows, the average ratings for the control group were 2.323 at pre-test which then increased to 2.380 at the post-test, showing a tiny increase. For Trained Group A, the pre-test goodness rating scores increased from an average of 2.464 to 2.811 at the post-test. Lastly, in Trained Group B, the pre-test goodness rating score was 2.861, which then increased to 3.136 at the post test.

Next, in order to get a better idea of the effect of the training in each condition, a paired samples *t*-test was further performed to examine whether there was a significant change in production performance from pre-test to post-test in each group. The results of this analysis indicated that average post-test rating scores were significantly higher than average pre-test rating scores only in Trained Group A and B but not in the control group (control group: +0.057, t=.788, p>.05). An average improvement of +0.347 in production performance was found in Trained Group A (t=3.295, p<.05) and an increase of +0.275 in Trained Group B (t=2.713, p<.05). In other words, the results clearly indicated that Japanese fricative and affricate contrasts produced by the trained participants received significantly higher goodness rating scores after receiving the training, indicating that the trained participants had a less foreign accent than at the time of the pre-test. Although there was also a tiny increase in the untrained control group, this was to a much lesser extent and no significant improvement was found. The results obtained here are in line with the results of the production accuracy analysis reported in 4.5.1.

Once again, a lack of time and numbers of participants made it difficult to strictly control the participants' proficiency level, and, as can be seen from the results shown above, there seems to be a difference among the groups at pre-test. To investigate whether there was a significant difference among groups, a one-way ANOVA was conducted. The result of the analysis revealed that there was no significant difference between the control group, Trained Group A and Trained Group B (F=2.934, p>.05). In other words, these results suggest that the production performance of the three groups did not differ significantly prior to training.

Next, to further determine whether the mean of the goodness rating scores of the post-test results differ between each group or not, an ANCOVA was performed using the pre-test score as the independent variable, group and sound as factors and the post-test as the dependent variable. The results demonstrated that group had a significant effect on the post-test goodness rating score after controlling the effect of the pre-test ($F(2,117)=4.532$, $p<.05$) as did sound ($F(3,117)=7.160$, $p<.001$). In other words, the results indicate that the post-test production performance for the three groups differed significantly and the production goodness rating scores of each of the four contrasts were not equivalent. Since the groups did not perform equivalently in producing the target sounds a follow-up Scheffe multiple comparison was performed to see which groups posed a difference in production performance at post-test. The results showed that there was a significant difference only between the control group and Trained Group B (p<.05) but not between the control group and Trained Group A (p>.05) nor the two trained groups (p>.05). These results indicate that at post-test the production performance of Trained Group B was significantly better than the control group.

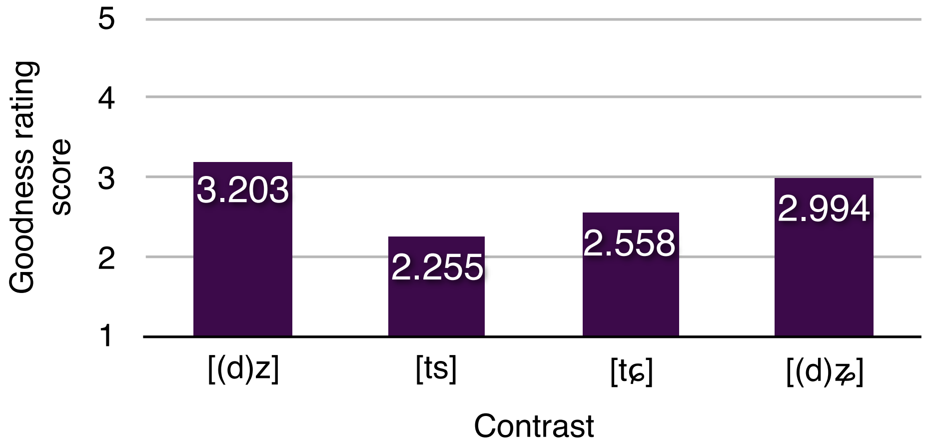


Figure 4.12 Mean goodness rating scores of Japanese fricative and affricate contrasts as classified by contrasts at post-test.

Next, since sound was shown to have a significant effect from the ANCOVA analysis above, follow-up Scheffe multiple comparisons on sounds were conducted to see which contrasts displayed a difference in goodness rating scores. The results revealed that at post-test [(d)z] and [ts] and also [ts] and [(d)ʑ] differed significantly (p<.05). Also, [(d)z] and [tɕ] showed a trend of significant difference (p<.1). Figure 4.12 indicates that [(d)z] was clearly the most intelligible sound produced by Thai learners as judged by five Japanese native raters. On the contrary, [ts] and [tɕ] are somehow similarly challenging for Thai learners to produce intelligibly.

In summary, the goodness rating results here confirmed the results reported in the previous section - that since the goodness rating scores of the post-test productions of the trained participants were significantly more highly rated than the pre-test scores, it can be said that the trained Thai learners' productions of Japanese fricatives and affricate contrasts generally became more intelligible to Japanese native listeners after the training. Lastly, regarding the post-test scores as analysed by contrast, [(d)z] was shown to be the most reliably intelligible sound to produce for Thai learners when compared to the other three contrasts.

4.5.2.2 Training effects as analysed by consonant contrasts

This section will analyse the mean goodness rating scores separately for each individual contrast. The results of the ANCOVA analysis explored above showed that differences in sounds had a significant effect, so it is worth looking at the degree to which participants performed differently across contrasts in more detail. Table 4.3 summarises the mean goodness rating scores of production utterances for the three groups at the pre-test and post-test.

Table 4.3 Paired samples *t*-test results of within-group comparisons between pre-test and post-test for each contrast.

Group	Contrast	Pre-test (SD)	Post-test (SD)	Dif	*t* (Sig)
Control (n=11)	[(d)z]	2.7(1.261)	2.891(1.316)	+0.191	1.323 (*p*=.215)
	[ts]	1.955(1.209)	1.873(1.055)	-0.082	-1.218(*p*=.215)
	[tɕ]	2.355(0.857)	2.236(0.861)	-0.119	-0.848 (*p*=.416)
	[(d)ʑ]	2.282(0.679)	2.518(0.799)	+0.236	1.267 (*p*=.234)
Trained Group A (n=11)	*[(d)z]**	2.418(1.187)	3.145(1.215)	+0.727	2.731 (*p*=.021)
	*[ts]***	1.782(0.829)	2.182(0.982)	+0.4	5.014 (*p*=.001)
	[tɕ]	2.782(0.722)	2.6(0.873)	-0.182	-1.011 (*p*=.336)
	*[(d)ʑ]**	2.873(0.970)	3.318(0.927)	+0.535	2.278 (*p*=.046)
Trained Group B (n=9)	[(d)z]	3.144(1.214)	3.656(1.112)	+0.512	2.270 (*p*=.053)
	[ts]	2.5(1.320)	2.811(1.172)	+0.311	1.483 (*p*=.176)
	[tɕ]	2.933(0.587)	2.9(0.658)	-0.033	-0.171 (*p*=.669)
	[(d)ʑ]	2.867(0.763)	3.178(0.803)	+0.311	1.863 (*p*=.100)

* means level of significance: * = p<.05, **= p≤.001

Differences between pre-test and post-test mean goodness rating scores for individual contrasts were investigated using a paired samples *t*-test. This analysis showed that in the control group, there was no significant difference found (*p*>.05). There was a little improvement observed for [(d)z] and [(d)ʑ] with an increase of +.0191 and +0.236, respectively. However, [ts] and [tɕ] showed a decrease in production intelligibility.

Next, for Trained Group A, participants in this group improved significantly in their production intelligibility for every contrast except for [tɕ]. The greatest improvement was seen in [(d)z] with an increase of +0.727 but standard deviations were relatively high and so the significant difference was not as strong as for [ts]. [ts] showed a strong significant change with an increase of +0.4 intelligibility. [(d)ʑ] showed an increase of +0.535. In contrast with the other three contrasts, [tɕ] showed a decrease in intelligibility of -0.182. These results indicate that after receiving training, the trained participants in this group produced [(d)z, [ts], [(d)ʑ] in a more native-like fashion with the exception of [tɕ].

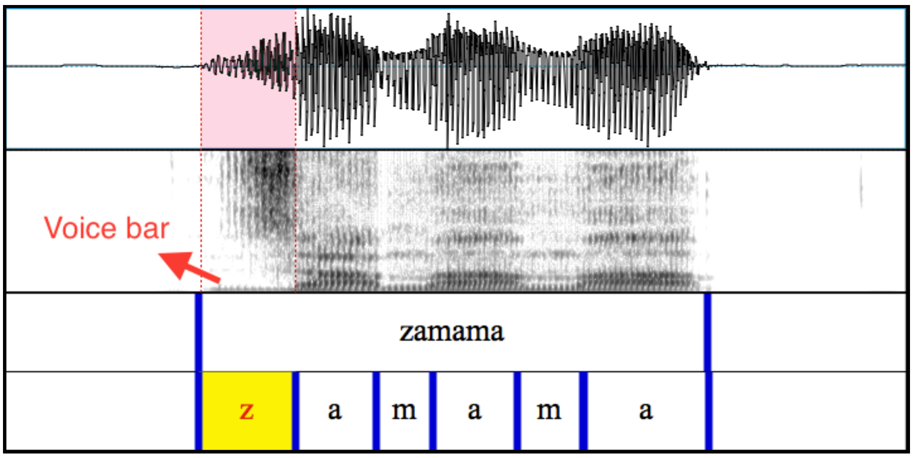
Lastly, for Trained Group B, a significant improvement was not found for any contrasts. However, it is worth looking more closely of this group even though there was no significant improvement found. For [(d)z], there was a significant trend observed with an increase of +0.512 (*p*<.1). Moreover, the results showed that overall production performance of [ts] and [(d)ʑ] similarly improved with a +0.311 increase. As in line with the results of other groups, the trained participants in this group performed in a less native-like fashion when producing [tɕ] with a -0.033 decrease in intelligibility.

Overall from the production accuracy results and the goodness rating results, it can be posited that 1) only trained groups made significant production improvements after their training, 2) the trained groups improved more than the control group in their performance of all contrasts, 3) significant improvements of specific contrasts from pre-test to post-test were seen only in Trained Group A, 4) all groups failed to make any gains in [tɕ], 5) the results of the goodness rating scores are overall in line with the previous section about production accuracy as judged by Japanese native speakers. Taken together, the two production evaluations conducted here support the claim that the "high variability method of perceptual training" has an effect on production improvements.

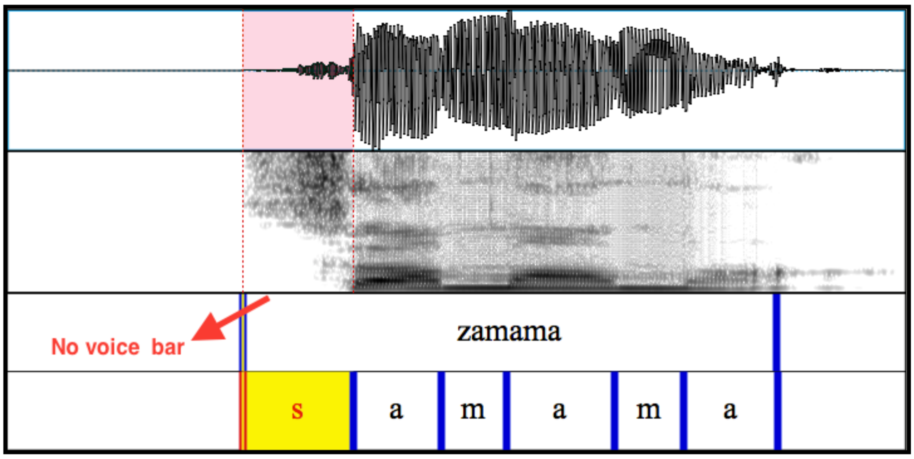
4.5.3 Waveform and spectrogram inspection

This section presents the acoustic features of the pre-training and post-training productions in order to gain more insight into the production errors made by Thai learners. These analyses, which can offer objective evidence and more comprehensive understanding of the production improvements, were conducted by inspecting waveforms and spectrograms of participants' productions compared to the productions uttered by model Japanese native speakers.

(a) [zamama] uttered by Japanese native speaker



(b) Pre-test production of [zamama] uttered by TLA2



(c) Post-test production of [zamama] uttered by TLA2

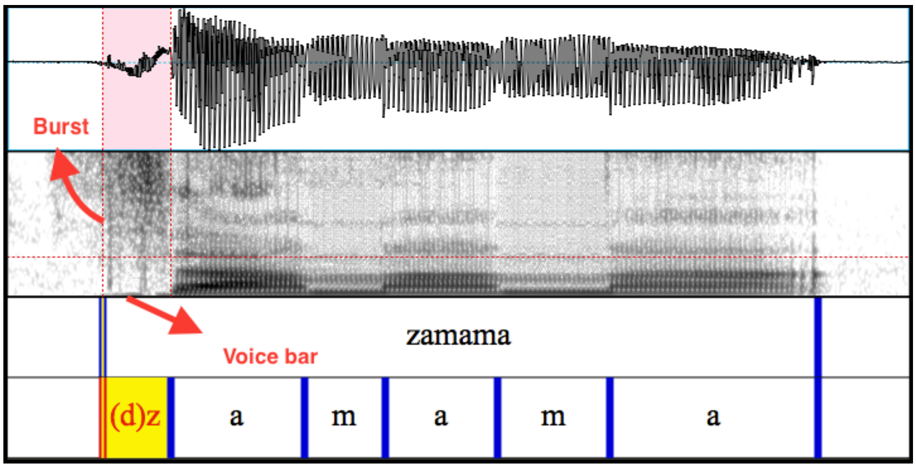
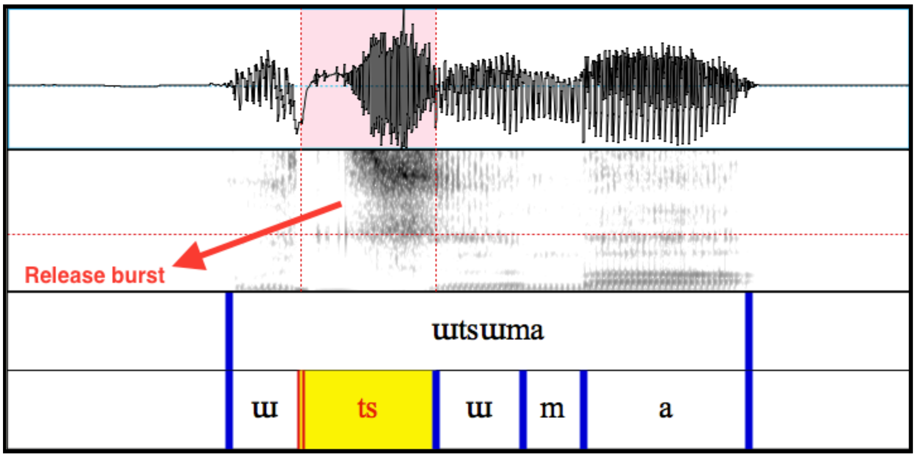


Figure 4.13 Screenshots of waveforms and spectrograms of [zamama] (a) a native-like production from a Japanese native speaker, (b) a pre-test production and (c) a post-test production from trained participant (TLA2).

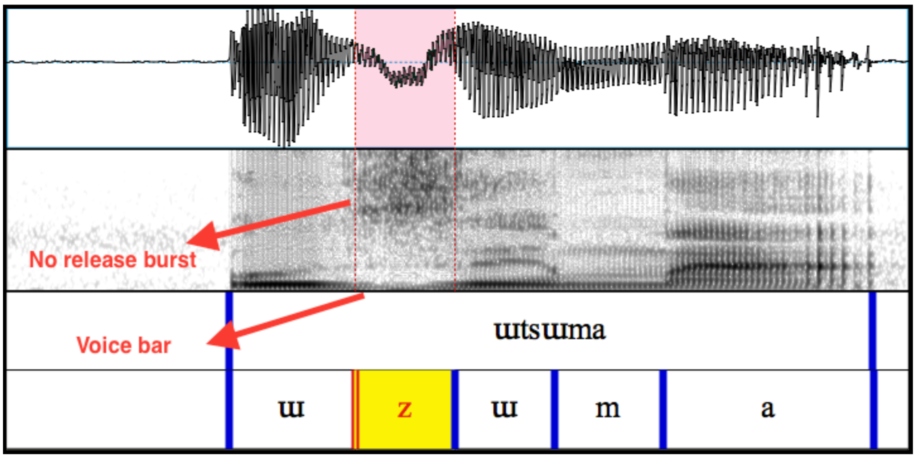
The waveforms and spectrograms in Figure 4.13 illustrate the [zamama] production as (a) a production by a Japanese native speaker; (b) a pre-test production from TLA2; (c) a post-test production from TLA2. Based on the results of the evaluation test, [z̲amama] as produced by TLA2 at pre-test was uniformly given a score of “0” (inaccurate) and perceived as [s̲amama] by all five raters. Figure 4.13 (a) shows a Japanese native speaker production of [z̲amama] with a voice bar[24] evidently presented throughout the segment in the beginning of the word and under the frication, indicating that the sound was produced in a voiced manner. [z̲amama] as attempted by TLA2 (as shown in (b)) showed no voice bar in the spectrogram as was observed in (a), suggesting that TLA2 produced the target sound in a voiceless manner, hence [z̲amama ➔ s̲amama]. The substitution pattern of [z ➔ s] has frequently been identified in prior studies (Sukegawa, 1993; Kawano, 2014). However, after undergoing the period of training, TLA2’s [zamama] tended to be accurately produced in a voiced manner, as can be seen in the appearance of the voice bar, observed in 4.13 (c). However, in the [zamama] post-test production it is possible that TLA2 produced [z̲amama] in a more affricate-like manner, i.e. as [(d)z̲amama], since a burst spike was observed.

(a) [ɯtsɯma] uttered by Japanese native speaker



[24] “Voice bar” refers to a low-frequency band at the bottom of the spectrogram indicating energy associated with voicing (Johnson, 2012).

(b) Pre-test production of [ɯtsɯma] uttered by TLB1



(c) Post-test production of [ɯtsɯma] uttered by TLB1

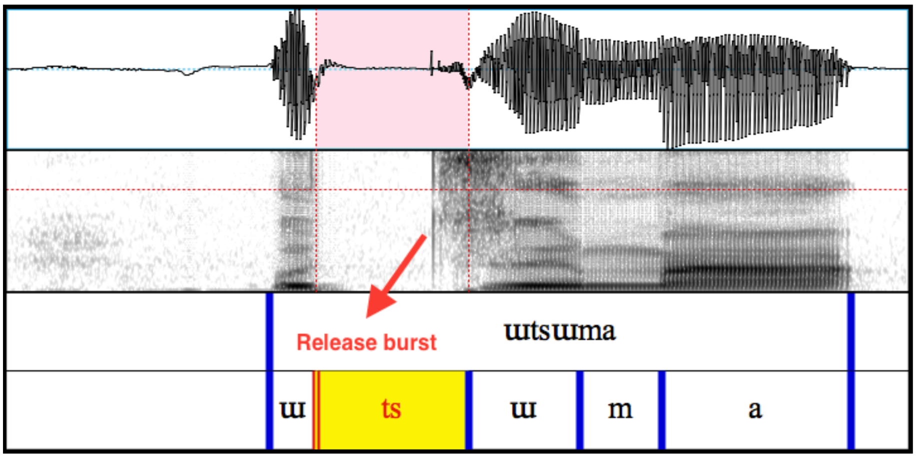
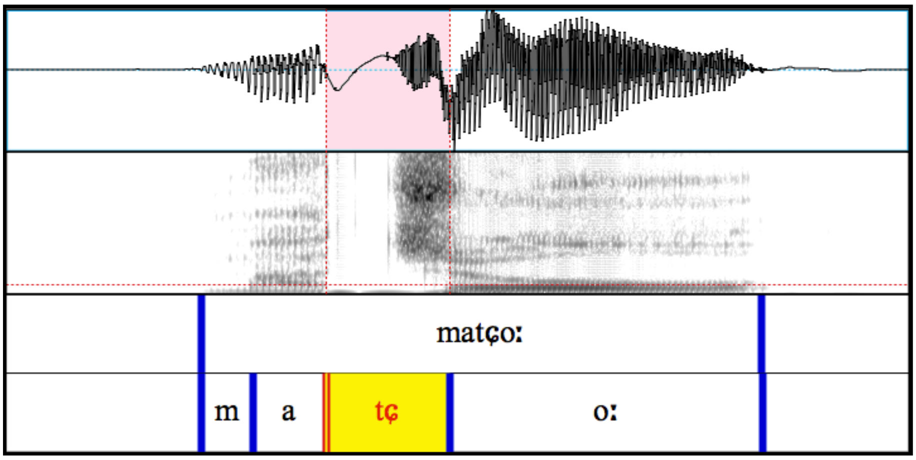


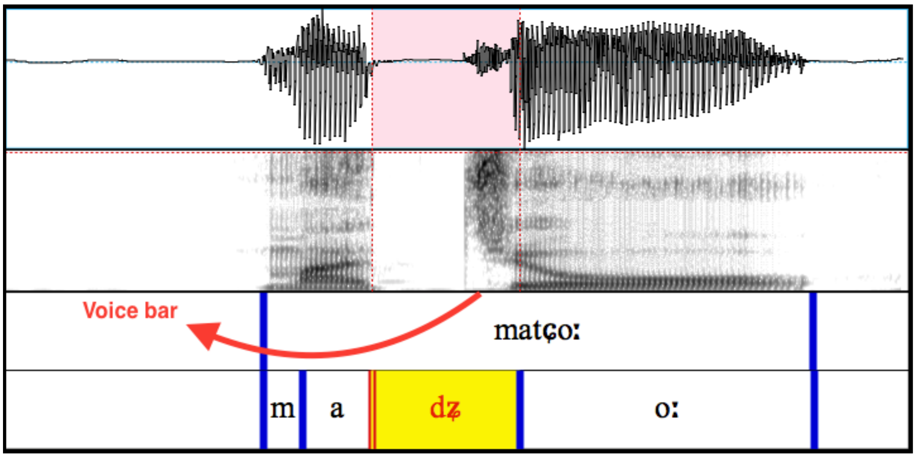
Figure 4.14 Screenshots of waveforms and spectrograms of (a) [ɯtsɯma] with a native-like production, (b) a pre-test production and (c) a post-test production by trained participant (TLB1).

Next, Figure 4.14 illustrates the waveforms and spectrograms of productions of the word [ɯtsɯma] by (a) a Japanese native speaker; (b) a pre-test production by TLB1; (c) a post-test production by TLB1. Based on the results of the evaluation test, the [ɯtsɯma] produced by TLB1 was perceived as the voiced fricative [ɯzɯma], by all five raters. It can clearly be seen from (a) that [ts] as produced by Japanese native speaker contains one brief and complete closure or a stop release burst, representing the character of an affricate sound before the frication starts. However, the [ɯtsɯma] attempted by TLB1 was acoustically different from (a) in that it shows the absence of the burst spike and there was an evident voice bar under the frication observed instead, clearly showing that TLB1 produced the target word as a voiced [ɯzɯma]. However, after undergoing training, TLB1 produced in a more native-like way. In the post-test production of the target word, after [ɯ] the waveform and spectrogram contained a complete closure and then a burst spike was shown, indicating that an affrication occurred, followed by a shorter frication, clearly showing that TLB1 produced [ɯtsɯma] accurately.

(a) [matɕoː] uttered by Japanese native speaker



(b) Pre-test production of [matɕoː] uttered by TLB9



(c) Post-test production of [matɕoː] uttered by TLB9

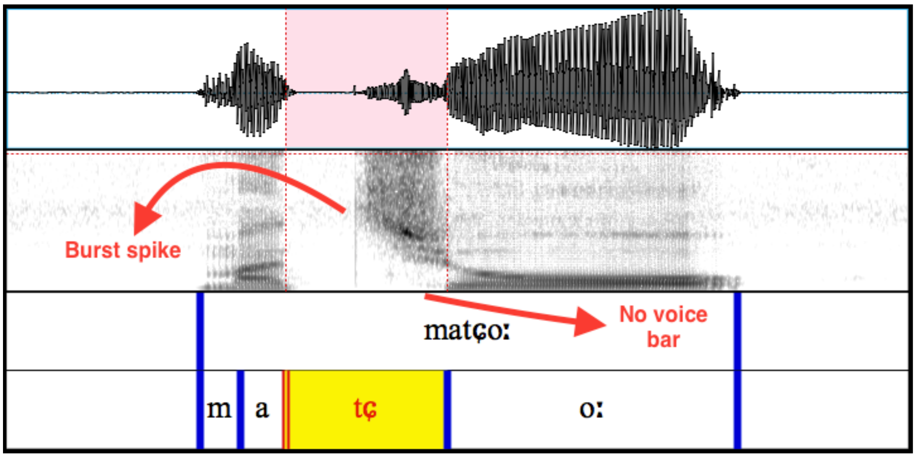
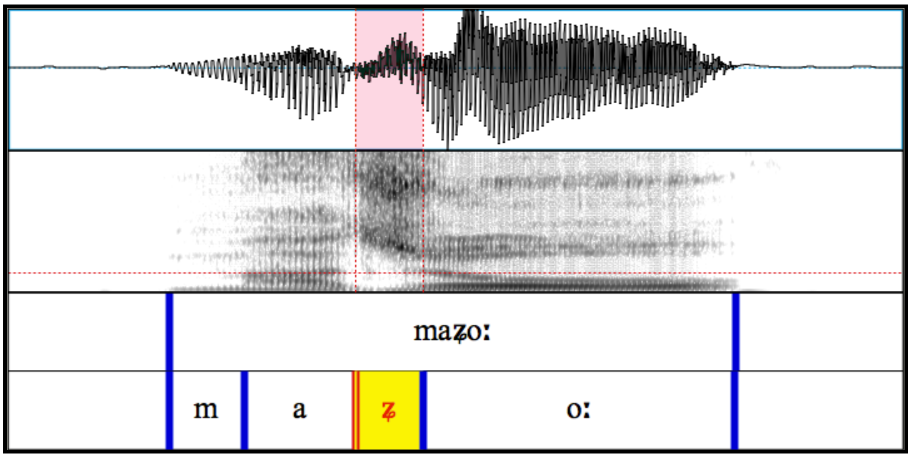


Figure 4.15 Screenshots of waveforms and spectrograms of (a) [matɕoː] with a native-like production, (b) a pre-test production and (c) a post-test production by trained participant (TLB9).

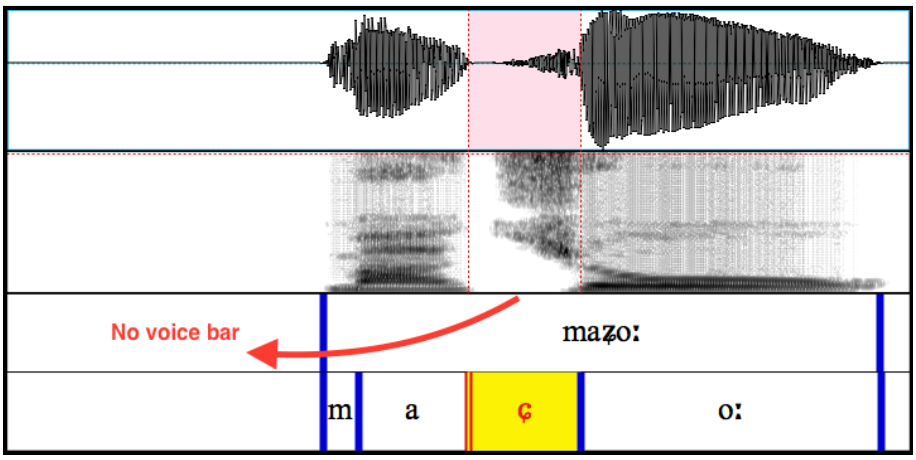
Next, the spectrograms and waveforms of Figure 4.15 illustrate productions of [matɕoː] by (a) a Japanese native speaker; (b) a pre-test production from TLB9; (c) a post-test production from TLB9. The pre-test token shown in (b) was uniformly given a score of “0” (inaccurate) by all five

Japanese raters and it was reported that they perceptually perceived it as a voiced [madʑoː]. However, the figure (b) showed an unclear detail of the pre-test production, but when looked at closely on Praat, there evidently was a voice bar observed under the frication showing that TLB9 pronounced [matɕoː ➔ madʑoː] in a voiced manner. However, after receiving the training, TLB9 pronounced with no voicing manner and yielded more similar acoustic features to (a) which was produced by a native speaker.

(a) [ma(d)ʑoː] uttered by a Japanese native speaker



(b) Pre-test production of [ma(d)ʑoː] uttered by TLA13



(c) Post-test production of [ma(d)ʑoː] uttered by TLA13

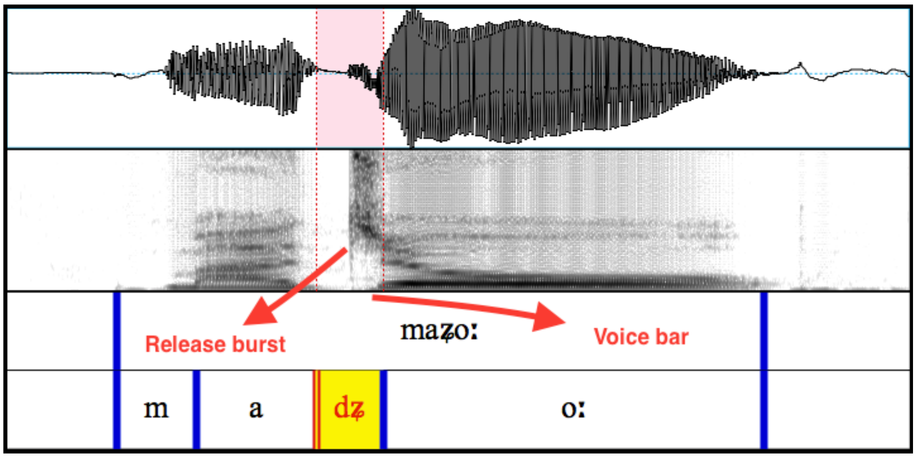


Figure 4.16 Screenshots of waveforms and spectrograms of (a) a native production of [ma(d)ʑoː], (b) a pre-test production and (c) a post-test production by trained participant (TLA13).

Lastly, Figure 4.16 illustrates the spectrograms and waveforms of productions of the word [ma(d)ʑoː] (a) by a Japanese native speaker; (b) a pre-test production by TLA13; (c) a post-test production by TLA13. Based on the evaluation test carried out by the five Japanese raters, the [ma(d)ʑoː] attempted by participant TLA13 at pre-test was uniformly given a score of "0" (inaccurate) and was perceived as [maɕoː] by all raters. As the figure shows, it is clear that TLA13 did not pronounce [ma(d)ʑoː] in a native-like way since it was acoustically different from (a), which was produced by a Japanese native speaker. In (a), a voicing bar was present throughout the frication segment of [ma(d)ʑ̲oː] in the middle of the word. In contrast, in the pre-test production in (b), [ma(d)ʑoː] produced by TLA13 shown in (b) featured a lack of a release burst and a voicing bar in the middle of the word. Also, there was a relatively large presence of frication at the segment, clearly indicating that TLA13 substituted the target word with a voiceless fricative ([ɕ]) and produced it as [maɕ̲oː]. However, according to Kawano (2014) and

Trakantalerngsak (2013), [(d)ʑ→ɕ] substitution is a relatively rare mistake for Thai learners to make when trying to produce the voiced [(d)ʑ], since normally [(d)ʑ] is mostly substituted with the voiceless [tɕ]. However, after receiving the training, the post-test production by TLA13 demonstrated a release burst followed by voicing throughout the segment as shown by the presence of a voicing bar across the spectrogram possibly produced as [madʑoː].

In conclusion, this subsection has presented some acoustic features observed in Thai learners' productions when compared with native speaker productions and production error trends made by Thai learners of Japanese. Before receiving training, participants' productions were acoustically different from the native speaker. However, after the training, they pronounced the target words in more native-like way, showing that HVPT perceptual training has a positive effect on improving learners' production ability.

4.6 Relationship between production and perception

The relationship between perception and production at group level and individual level will be examined in this section. The purpose of this discussion is to examine whether there is a relationship between the perception and production of the target contrasts and to examine the degree of improvement of individuals in both modalities. As Wang (2002) suggests, "the ultimate goal of learning the sound system of a target language is success in both perception and production, perceptual learning cannot be completely evaluated without examining its relation with production" (p. 19). Previous studies have shown that there was a transfer of perceptual training to production (Bradlow et al., 1997; Rochet, 1995), suggesting that there is a link between perception and production modalities. However, to investigate whether there is a link between perception and production, it is necessary to examine the relationship between perception and production accuracy. If there is such link, it may be possible that perceptual training promotes similarly better perception and production accuracy. This section presents correlation analyses only for the trained groups.

4.6.1 Group level correlation

Table 4.4 Accuracy of perception and production at pre-test and post-test for each contrast (Trained Group *A*).

	Pre-test (%)		*r*	**Post-test (%)**		*r*
	Perception	Production		Perception	Production	
1. [(d)z]	85	51	0.221	86	67	0.773*
2. [ts]	58	19	0.33	76	34	0.54
3. [tɕ]	60	38	0.085	70	41	-0.45
4. [(d)ʑ]	66	61	0.174	84	71	0.262

* means level of significance: * = *p*<.05, **= *p*≤.001

Table 4.5 Accuracy of perception and production at pre-test and post-test for each contrast (Trained Group *B*).

	Pre-test (%)		*r*	**Post-test (%)**		*r*
	Perception	Production		Perception	Production	
1. [(d)z]	82	74	0.761*	93	84	0.788*
2. [ts]	81	45	0.359	81	54	0.739*
3. [tɕ]	49	46	0.277	82	53	-0.117
4. [(d)ʑ]	72	63	0.642	92	73	0.401

* mean level of significance: * = *p*<.05, **= *p*≤.001

Mean percentages of perception and production accuracies at the pre-test and the post-test by each contrast are shown in Tables 4.4 and 4.5. As the tables show, perception performance generally exceeded production performance for all contrasts. This finding supports previous studies (Flege, 1991; Llisterri, 1995) which claim that perception precedes production. After training, the voiced sounds [(d)z] and [(d)ʑ] are relatively easy to identify and to produce for both trained groups. In contrast, the post-test results showed that [ts] and [tɕ] are somewhat easy to perceive after the training but showed a production accuracy rating of less than 50%, indicating that these contrasts are highly difficult for Thai learners from both groups to produce.

A Pearson correlation was performed on each contrast to see whether there was a correlation between perception accuracy and production accuracy before and after the training. As shown in Tables 4.4 and 4.5, the correlation analyses showed that, except for the "[(d)ʑ]" sound produced in Trained Group B, overall pre-test results showed no significant correlation, indicating that identification accuracy was not reliably related to production performance at pre-test. However, after receiving the training, for Trained Group A, the accuracy percentage of the perception and production of [(d)z] at post-test were found to be highly correlated (r=.773, p<.05). [ts] showed a trend of correlation (r=.540, p<.1). In contrast, a correlation was not found in [tɕ] and [(d)ʑ], although a higher correlation was observed in the post-test than the pre-test. For Trained Group B, the accuracy percentage of the perception and production of [(d)z] and [ts] correlated highly ([(d)z]: r=.788, p<.05; [ts]: r=.739, p<.05), indicating that the identification accuracy of some contrasts was significantly related to their production performance after the training. In other words, participants who improved in the perception of [(d)z] and [ts] were more likely to improve in the production of those contrasts, and vice versa. In contrast to other sounds, [tɕ] showed a negative correlation and a contradictory trend in both groups, revealing that while [tɕ] became relatively easy to identify after the training, the contrast was still highly difficult to produce for Thai learners. In other words, the training helped improve the perception of [tɕ] but was not effective in improving its production.

In summary, at pre-test, there was no reliable relationship observed between perception and production both in Trained Group A and B, suggesting that there were no pre-existing relationships between the two modalities before the training. In other words, there may be cases wherein perception is accurate but production is not accurate. At post-test, there was found to be an observable relationship between improvements in both trained groups. Although identification accuracy was not significantly related to production performance before the training, it was afterwards found that there was a stronger link between the two domains after the training. However, these results are still not strong to support a relationship between the two modalities. The fact that most contrasts were well perceived despite relatively inaccurate production suggests that there is no clear and consistent relationship between L2 perception and production, and that accurate L2 perception may not be sufficient for accurate production. Moreover, according to Bradlow et al. (1997, p. 2307), due to the process of learning in the two domains appearing to be distinct within individual subjects, learning in the perceptual domain is not a necessary nor sufficient condition for learning in the production domain. Here it can only be predicted that the L2 perception and production processes may share common underlying representations and perceptual training might have an effect on improving both modalities.

4.6.2 The individual level relation

Previous studies have demonstrated that there were large individual differences in benefit gained from training (Bradlow et al., 1997; Hazan et al., 2005; Lengeris, 2009), nevertheless it is worth looking into individual performance and the change in the performance of both domains in this study. Next, the degree of improvement and the relationship of the two modalities for individuals will be analysed. Mean percentages of correct identification and production scores of individual participants at pre-test and post-test are presented in Tables 4.6 and 4.7.

Table 4.6 Individuals' identification and production accuracy at pre-test and post-test for Trained Group *A.*

Subject	Perception			Production		
	pre	post	dif(%)	pre	post	dif(%)
TLA1	47	91	*44*	21	37	*16*
TLA2	81	78	*-3*	23	43	*20*
TLA3	78	91	*13*	48	59	*11*
TLA4	72	66	*-6*	34	45	*11*
TLA5	72	78	*6*	60	56	*-4*
TLA6	78	88	*10*	73	79	*6*
TLA7	50	53	*3*	31	38	*7*
TLA9	63	81	*18*	34	58	*24*
TLA10	75	84	*9*	31	49	*18*
TLA11	69	81	*12*	48	56	*8*
TLA13	56	81	*25*	61	68	*7*

Table 4.7 Individuals' identification and production accuracy at pre-test and post-test for Trained Group *B.*

Subject	Perception			Production		
	pre	post	dif(%)	pre	post	dif(%)
TLB1	66	91	*25*	48	64	*16*
TLB2	75	88	*13*	71	88	*17*
TLB3	81	91	*10*	55	53	*-2*
TLB4	78	91	*13*	53	70	*17*
TLB5	84	100	*16*	76	81	*5*
TLB7	59	75	*16*	49	57	*8*
TLB9	75	91	*16*	81	93	*12*
TLB10	53	72	*19*	40	43	*3*
TLB11	66	84	*18*	38	48	*10*

Tables 4.6 and 4.7 compare and contrast the amount of learning in perception and production of all trained participants in order to see the relationship between changes in production and perception performance. As the tables show, overall, the results show that most of the participants improved noticeably in perception and also showed some gains in production.

These results reveal considerable variability across participants. For most trained Thai learners, perception and production ability developed simultaneously. TLA1, TLA3, TLA6, TLA7, TLA9, TLA10, TLA11, TLA13, TLB1, TLB2, TLB4, TLB5, TLB7, TLB9, TLB10 and TLB11 or 16 participants out of 20 gained benefits in both modalities. However, TLA1, TLA3, TLA6, TLA11, TLA13, TLB1, TLB5, TLB7, TLB9, TLB10 and TLB11 exhibited greater improvement in perception than production. By contrast, TLA4, TLA7, TLA9, TLA10, TLB2 and TLB4 made improvements in production accuracy rather than perception accuracy. To be specific, some participants such as TLA7, TLA10, TLB2 and TLB4 gained only a little benefit in perception accuracy but showed larger gains in production accuracy after the training. However, there were some participants who did not show any improvement at all in perception accuracy and who even performed worse in perception at post-test but who showed a large gain in production accuracy at post-test, such as TLA2 and TLA4. In addition, TLA5 and TLB3 improved in the perception test but the learning did not transfer to production. Moreover, some previous studies demonstrate that weaker learners benefit more from training since they have larger room for improvement (Wang et al., 1999). As shown in the table, this study also shows that some participants with a lower initial pre-test score such as TLA1, TLA13, TLB1, TLB10 and TLB11 showed more substantial improvements in the post-test whereas training effects were much smaller for those who started with high accuracy scores in the pre-test such as TLA2, TLA4, TLA5 and TLB3.

These results indicate that perceptual training lead to perception and production improvements in most trained Thai learners. Overall, the results showed that most trained Thai learners improved their ability in both modalities. However, the results revealed considerable variability across participants. Individuals whose perception improved were not necessarily guaranteed to improve in their production, and vice versa. Some cases showed that there was no observable relationship between changes in perception and production. Improvements in perception and production do not systematically progress at equal rates within individuals (Bradlow et al., 1997; Kartushina et al., 2015).

4.7 Questionnaire responses

In addition to the experimental results described above, all trained participants were asked to assess and comment on their perceptual training by filling out a questionnaire form (See Appendix J) immediately after the generalisation test finished. The aim of this section was to understand participants' opinions regarding perceptual training. The results of the questionnaire are briefly summarised here along with some direct quotes from the participants' responses.

Results of the questionnaires:

Question 1: What is your motivation to take part in this training?

All twenty participants responded that they wanted to improve their listening and speaking ability. None responded that they felt forced to take part in the training.

Question 2: How do you feel about the training?

The following quotes represent the trainee's overall positive experiences with the training.

- Positive -

- "I found that the training was good in helping me distinguish Japanese sounds better."
- "After I attended this training, I feel like my listening skills are still not good. And I feel that I have more motivation to improve my listening skills from now on."
- "The more I do, the better score I get."
- "It's very helpful that I had a chance to practice over the summer holidays."
- "It's fun! It feels likes doing a hobby."
- "It helps my listening and pronunciation skills."
- "I now can better distinguish some sounds."
- "I found some mistakes that I always make."
- "It was very helpful."

- Negative -

- "I feel like I still have some problems in distinguishing some fricative and affricate sounds."
- "Exposure to multiple native speakers' productions make me feel more unfamiliar and confused."
- "I feel like some sessions were very difficult and I couldn't identify some sounds."

Question 3: Did you find the training helpful in listening to and understanding Japanese contrasts? (This question is a yes/no question.)

There were no negative responses to this question. All twenty participants responded "Yes", indicating that the training was helpful for learning Japanese contrasts.

Question 4: Would you like to conduct this training or similar phonetic training in your Japanese class?

All twenty participants responded they would like to take part in phonetic training in their classes.

Question 5: (Only asked of those in Trained Group B) Does listening to your own speech help you notice your production problems?

All nine trained participants of Trained Group B responded "Yes". And some participants explained that the self-listening task gave them more chances to notice and improve their production because the task let them compare their own speech with the accurate native speakers' speech. Also they directed their awareness towards listening to the model sounds for the first time. Many also commented that they had more awareness and felt it was important to hear their own production in order to improve their learning.

Question 6: Please list some advantages of this training.

The following quotes represent the trainee's overall *positive* experiences with the training.

- "I gained some new listening strategies to help distinguish some difficult sounds."
- "I did not know the difference between [ts] and [s]. But now I know the difference between these two sounds."
- "When participating in this training, I got a chance to listen to accurate Japanese speech."
- "I learned how to distinguish some similar sounds and know my own listening problems."
- "The more I practice the better I can identify accurately."
- "I do believe that my listening skill has improved."
- "I feel more motivated after attending this training."

Mostly, participants' comments were positive. However, there was one participant who explained that it was complicated to listen to many model speakers.

Question 7: Please give any suggestions that you have about the training.

- Some participants wrote that they liked the training and wanted more training sessions.
- They wished that this training had been available earlier.

Some participants gave some good suggestions for improving the training as follows:

- Technical problems should be resolved (the software was slow to respond and encountered freezing.)
- Feedback was presented only up to three times in this study. Some participants requested more than three times. Also, they requested the chance to listen to feedback sounds whether their answers were right or wrong.
- Some participants wrote that they liked the training and wanted more training sessions.

In summary, from the questionnaire results detailed here it is clear that the training program left an overall positive impression and that the majority of participants felt that the perceptual training was of benefit to them, helping them to learn more about Japanese sounds. Moreover, the training increased their awareness of the fricative and affricate contrast differences. For those who received the self-listening task, they reported that the task helped to raise their awareness about speech learning. Therefore, it seems reasonable to conclude that training tools such as the perceptual training conducted in this study have the potential to contribute towards the learning of Japanese fricative and affricate contrasts among Thai learners.

Chapter 5 Discussion

5.1 Summary of the results

The ultimate purpose of this research is to establish an effective training tool to be used in training L2 learners with the aim that it will improve learners' speech learning ability in perceiving and producing L2 contrasts. Relatively few studies in the L2 speech training canon have examined Japanese contrasts or have involved Thai learners as target participants. In this dissertation, two perceptual training approaches, a standard HVPT perceptual training method and a modified HVPT method of perceptual training which differed in the training task in that the latter also conducted a self-monitoring task, were given to Thai learners of Japanese in order to train the problematic Japanese fricative and affricate contrasts.

As presented in the previous chapter, the overall results of the current study were in general positive. Both the standardised HVPT perceptual training and the modified HVPT perceptual training yielded positive effects in modifying Thai learners' perception of the Japanese fricative and affricate contrasts. Moreover, perceptual learning gained from the training was maintained for six months after the completion of the training. The improvements were also transferred to production as well. Participants from both the two trained groups performed equally well in perceiving and producing the target contrasts, indicating that there was no difference found between the two training approaches. However, Trained Group B demonstrated more benefits since the effects of training were maximised to all learners in the perception test. Moreover, generalisations to untrained words and an untrained talker were also observed only in Trained Group B, which undertook the self-monitoring task.

The main research questions of this dissertation are answered below. Following the research questions is a discussion putting particular emphasis on the implications for models of cross-linguistic speech perception, L2 speech training studies and pedagogical recommendations for Japanese language pronunciation teaching and learning.

5.2 Answers to the research questions

I. Can HVPT perceptual training improve Thai learners' perception of Japanese fricative and affricate contrasts?

The HVPT perceptual training carried out in the study incorporated training tokens containing Japanese fricative and affricate contrasts which were produced by four different native speakers and was undertaken by Thai learners of Japanese. The first question addressed by this study was whether this HVPT perceptual training would be beneficial for improving the Thai learners' perception of Japanese fricative and affricate contrasts. The overall results showed that there was a significant improvement from pre-test to post-test observed both in the trained groups and the untrained group (Trained Group A: 67%→79%; Trained Group B: 71%→87%; Control Group: 63%→69%). Thus, while the control group showed a 6% improvement in accuracy, the trained groups showed a significantly greater degree of improvement; about 12% in Trained Group A and 16% in Trained Group B. In other words, the results clearly indicated that participants who received HVPT perceptual training outperformed those who did not in the perception of the target sounds.

In summary, the overall results provide some evidence to support the use of HVPT perceptual training in effectively improving Thai learners' perception ability of Japanese fricative and affricate contrasts. Previous L2 speech training studies have shown that L2 learners benefit from perceptual training with an improvement in identification of 10 to 20% (Bradlow et al. 1997; Lively, Logan et al., 1993; Lively, Pisoni et al., 1994; Lopez-Soto &

Kewley-Port, 2009; Wong, 2013). The results of this study are in line with the findings of previous studies, which show the effectiveness of HVPT perceptual training on improving perception. Moreover, the results also support the claim that L2 perceptual learning can occur in L2 adult learners and their speech perception can be modified through laboratory training (Bradlow et al. 1997; Iverson et al, 2005; Iverson & Evans, 2009; Lively et al., 1993; Logan et al 1991; Nishi & Kewley-Port, 2008; Rochet, 1995).

Improvements for individual contrasts

The results indicated that HVPT perceptual training had a positive effect on the perception of Japanese fricative and affricate contrasts. However, a closer investigation revealed that individual contrasts were not equivalently improved. Some contrasts proved easier to perceive and some were more difficult, even after training had taken place. That is, some sounds showed a large degree of improvement whereas other sounds showed no improvement at all. This unequal gain observed in different contrasts is also seen in previous studies (e.g., Kartushina et al., 2015; Thomson, 2008). Moreover, Trained Groups A and B did not show the same improvement trends for individual contrasts. To be specific, within groups, in Trained Group A [ts] contrast identification significantly improved and showed the greatest improvement with an 18% increase, and [(d)ʑ] showed a significant trend of improvement. However, Trained Group A did not demonstrate significant learning of the [(d)z] and [tɕ] contrasts. Conversely, in Trained Group B, [tɕ] and [(d)ʑ] were significantly improved and [(d)z] showed a significant trend of improvement. Specifically, the greatest improvement was observed in [tɕ] with a +33% point increase in accuracy. [(d)z] and [(d)ʑ] also showed a substantially high degree of improvement with a +20% and +11% increase respectively. Contrary to other contrasts (and contrary also to the results of Trained Group A), [ts] showed no improvement at all.

One possible explanation for these differences in improvements across contrasts may be that a ceiling effect is in place (Bradlow et al. 1997; Kartushina et al., 2015). Contrasts that were perceived less accurately before training began appeared to gain more benefit from the training. On the contrary, contrasts that showed considerably high accuracy scores in the pre-test did not show any change, or showed only minimal degrees of improvement, since there was little room left for improvement to occur. In this study, at pre-test Trained Group B already showed high average accuracies of more than 81% for the [(d)z] and [ts] contrasts, whereas [tɕ] showed a noticeably low average of about 49%.

Moreover, the results showed that apart from the [(d)ʑ] contrast, Trained Group A and Trained Group B did not demonstrate the same gains. The difference in improvement across contrasts within groups might be attributable to the difference in training types (Hanulikova et al., 2012). However, it is too early to conclude that different training types result in differences in learning outcomes. More research is needed to determine what factors influences the unequal learning outcomes across contrasts and groups.

Improvements in the control group

The perception test results showed that the control group also demonstrated an overall significant improvement from pre-test to post-test. The finding that untrained participants demonstrated the ability to significantly better identify the target contrasts was also observed in previous studies (e.g., Hayes-Harb, 2007; Huensch, 2013). Several factors were suggested to account for the improvements found in the control groups of previous studies. Huensch (2013, p. 161) states that the long pre/post-test intervals could have provided enough input or may have acted as a form of training for the control group. Additionally, control group participants could have become familiar with the tests' methodologies (Burnham, 2013, p. 107). Taken together, it can be inferred that the contributing factor to the improvement of perception in the control group in this study may also possibly be because of the awareness of participating in the experiment or through input gained when conducting the extensive pre-test and post-test phases. As described in the Chapter Three, the untrained participants of the control group were given a pre-test, post-test, generalisation-test and a delayed post-test which contained a large number of stimuli produced by various Japanese speakers. These exposed tests may have acted as a form of training that helped improve the perception of the control group. Although there was a significant improvement of 6% observed in overall identification accuracy, when looked at closely, the results did not show a significant change from pre-test to post-test for any specific contrast, and the post-test identification accuracy of the trained groups were significantly higher than the control group. Moreover, untrained participants improved significantly from pre-test to the post-test only in the perception test, there was no significant improvement found in any other tests, such as in the generalisation test that the control group undertook, the delayed post-test and the production-test. In other words, it can be deduced that the control participants had possibly made use of short-term memory and may have developed the ability to perceive target contrasts under certain circumstances to better identify the target contrasts, but are as yet unable to represent the contrasts and implement stored contrasts in the long-term memory or induce a transfer of learning to production. However, since the factors contributing to the improvements of the control group are not yet certain, more research is needed to investigate this issue.

II. Does HVPT perceptual training allow generalisation to new words and new talkers?

This study also examined whether the perceptual learning gained from the training generalised to new words and a new talker. Huensch (2013, p. 163) states that receiving varied input from multiple talkers will yield a greater number of exemplars stored in the memory and then allow for generalisation to new words and new talkers. There were two types of generalisation tests conducted in the current study. The first test used old tokens from the pre/post-test produced by a familiar talker also used in the training and an unfamiliar talker not used in the training. The results of this generalisation test showed that participants' abilities to identify the pre/post-test tokens produced by the untrained talker were comparable to that of tokens produced by the trained talker. This means that, although perceptual training was carried out with only four talkers, there was a transfer in identification accuracy to a fifth talker, and even though this does not guarantee accurate perception of target contrasts when given by other talkers, it at least indicates that training carried out in the study provided enough variability to achieve generalisation. These results of the first generalisation test are consistent with previous research findings (Huensch, 2013; Rochet, 1995; Wang, 2002), revealing that there was a perceptual learning transfer to a new talker.

Next, the results of the second type of generalisation tests will be discussed. In order to further investigate whether there was robust generalisation to increased accuracy in perceiving new words as well as an unfamiliar talker, an additional two generalisation tests with a large range of 64 new untrained tokens produced by a familiar talker, whose voice was used in the training, and an unfamiliar talker, whose voice was utilised for the first time, were conducted. The two additional generalisation tests refer to; 1) a test of words of low lexical familiarity uttered by a familiar and an unfamiliar talker; 2) a test of words of high lexical familiarity uttered by a familiar and an unfamiliar talker. After examining statistical analyses taken from the two tests, it was revealed that Trained Group B consistently outperformed Trained Group A and the control group in all generalisation tests. Moreover, Trained Group B showed a significantly higher score in identification accuracy for most generalisation tests (apart from the test of high lexical familiarity by a familiar talker) than at pre-test, indicating relatively strong evidence that there was a transfer of perceptual learning to both untrained words and to an untrained talker in Trained Group B. In contrast, apart from the test of words of low lexical familiarity produced by a familiar talker for both groups, Trained Group A and the control group did not perform differently to the pre-test (meaning that their accuracy remained at pre-test levels), showing that the two groups had insufficient evidence to show a transfer of perceptual learning to new words and a new talker.

Taken together, the generalisation effect observed only in Trained Group B implies that the difference in training type might have influenced the learning outcome. This finding implies that HVPT perceptual training alone was beneficial for a generalisation to a new talker producing trained words but it was not enough to show evidence of a generalisation to new, untrained words and a new talker. On the contrary, trained participants who received HVPT perceptual training as well as a self-monitoring task could show better generalisation of perceptual learning to new words uttered by an unfamiliar talker. The greater generalisation observed in Trained Group B clearly indicates the greater effectiveness of combining a self-monitoring task with HVPT perceptual training.

Moreover, concerning the level of lexical familiarity, the degree of lexical familiarity has been shown to influence the degree to which participants perceive accurately. The results obtained from the generalisation test show that Thai learners were better at identifying words of low lexical familiarity than words of high lexical familiarity. According to Pierce (2014), "lexical items that occur more often in the target language can bias the participants' performance in perceptual training tasks" (p. 63). Moreover, Flege et al. (1996) found that Japanese learners were better able to correctly identify /l/ and /r/ tokens in more common words than those in less common words. Based on this previous research, it was expected that Thai learners would be more accurate in identifying familiar words than unfamiliar words. However, interestingly, unlike findings from prior research, this study demonstrated that Thai learners were better at identifying words of low lexical familiarity than words of high lexical familiarity. The reason why unfamiliar words were easier to perceive for Thai learners might be because of the effects of perceptual training. The training stimuli used in the training included not only real words but also nonsense words. It is possible that training learners using high variability and various sources of words, such as real words and nonsense words, might have influenced learners to direct their attention to the phonetic characteristics of the target contrast. As Pereira (2013, p. 30) states, the nature of the stimulus used in training may promote different types of learning. Hence, with the training background, when learners are exposed to words of low familiarity which they have never heard, they have no lexical knowledge and phonological representation of the target words. They then draw their attention directly to the phonetic characteristics of the exposed target contrasts, making them identify words of low familiarity more accurately than words of high familiarity. As cited by Hayes-Harb (2007, p. 66), recent research has provided evidence that learners can learn to discriminate second language contrasts even without any reference to word meaning after simply listening to the target language during a training session. Schmidt (2001) also states that in communicative L2 learning environments, learners tend to focus on meaning, but when their attention is explicitly oriented toward phonetic form, learning outcomes can be improved. Furthermore, Guion & Pederson (2007) state that, “with explicit directing of attention, adult learners can better discern novel phonetic contrasts” (p. 57). In their study, English monolinguals were grouped into those who were instructed to attend to Hindi phonetic contrasts (“sound-attending group”) and those who were instructed to attend to meaning correspondences of the same stimuli (“meaning-attending group”). The results show that the sound-attending group showed better discrimination at post-test for all contrasts, whilst the other group did not improve in any contrasts. However, the meaning-attending group showed more learning than the sound-attending group in semantic tests. This study confirmed that the orienting of attention is relevant to the acquisition of novel phonetic categories, suggesting that the mechanisms of the attentional system should be an explicit component of models of second language acquisition. This finding of the current study here also shows evidence that when Thai learners were exposed to various unfamiliar words of which they had no lexical knowledge, they direct their attention solely to the phonetic characteristics, and their phonetic judgments that rely on the phonetic form, rather than the reference of the word's meaning, and tend to be more accurate than identifying words of high familiarity. The results here emphasise the importance of using nonsense words in the training. Using such words, L2 learners can truly learn to establish new phonetic categories of non-native contrasts and their phonetic judgments are accurate no matter whether the words they hear are familiar or unfamiliar. If this finding can be shown to be sufficiently reliable, then training using nonsense words may lead to better learning, since it may help L2 learners create new, strong and stable sound categories that are founded truly on the phonetic form rather than a lexical meaning base. However, the data presented here are still insufficient to draw a firm conclusion. Additional studies are needed to determine whether the lexical familiarity effects observed here are indeed the result of perceptual training using nonsense words.

III. Will the improvement be retained after six months?

Previous L2 speech training studies have shown that learners maintained their improved levels of identification performance three months (e.g., Bradlow et al., 1999; Lively et al., 1994) and six months after completion of HVPT perceptual training (e.g. Wang et al., 1999). The results of the retention test of this study also indicated that the observed improvement in perception was maintained for at least six months after the training was completed. There were significant differences in mean identification accuracy scores observed between the pre-test and the retention test for the trained groups but not the control group. Moreover, there was also no significant difference found between the post-test and the retention test in the trained groups, indicating that their perceptual ability remained comparable to the improved performance that was observed after training. This result is consistent with previous findings, suggesting that L2 learners have long-term memory representations or have established representations for new L2 sounds after training (Wang et al., 1999). In contrast, the control group showed no change in perceptual identification accuracy from the pre-test to the six-month delayed post-test, indicating that the control group's identification ability at the time of the delayed post-test went back to the same level as when they performed the pre-test. The presence of this retention effect provides further support for the efficacy of HVPT perceptual training in that it can produce long-term change in learners' ability to identify L2 contrasts (e.g., Bradlow et al., 1999; Iverson & Evans, 2009; Lively, Logan et al., 1993; Lively, Pisoni et al., 1994; Wang, 2002; Wang & Munro, 2004; Yamada et al., 1996).

IV. Does HVPT perceptual training benefit the production performance of the target sounds?

Recently, a number of previous studies have shown that HVPT perceptual training has been shown to improve production ability even though there has been no explicit production training involved in those investigations (e.g., Bradlow, Akahane-Yamada et al. 1999; Bradlow, Pisoni et al., 1997; Hazan et al., 2005; Lambacher et al., 2005; Lengeris & Hazan, 2010; Lopez-Solo & Kewley-Port, 2009; Wang et al., 2003). This study also found that a significant improvement in perception gained from HVPT perceptual training of Japanese fricative and affricate contrasts also lead to an improvement in production. Japanese fricative and affricate contrasts produced by trained learners were not only assessed with greater rates of accuracy but also received significantly higher goodness rating scores after the training as judged by five Japanese native speakers. Such an improvement was not seen in the control group. The results here support previous research findings that most pronunciation errors have a perceptual basis, thus an improvement in perception will help to improve production (e.g., Flege, 1995; Rochet, 1995). Although this study has shown that perceptual training has influenced the production domain, the mean goodness rating scores revealed that their production performance was still far from native-like. Specifically, from a native-like score of five, Trained Group A demonstrated an improvement in rating from 2.464 to 2.811 and Trained Group B showed an improvement from 2.861 to 3.136, indicating that the post-training score is still far from the ideal native-like score. This lack of a large improvement in pronunciation is due to the fact that this training was not designed to facilitate better learning in production, and so a substantial improvement was not anticipated. Next, regarding improvements in the production of specific contrasts, a significant improvement of a specific contrast was observed only in Trained Group A. Participants from Trained Group A made significant improvements in the pronunciation of [(d)z], [ts] and [(d)ʑ]. An overall significant improvement was seen in Trained Group B but there was only a significant trend seen in the improvement of the pronunciation of [(d)z] and [(d)ʑ]. The lack of significant difference in Trained Group B may be due to the initially higher pre-test scores when compared to Trained Group A, implying that there was not enough space for an improvement to be made, as previously noted regarding perception (e.g., Bradlow et al., 1997; Kartushina et al., 2015).

An observation of the spectrograms and waveforms confirmed that trained participants improved their production and their pronunciation became more native-like. In addition, based on the answers provided to the questionnaire, most participants responded that they feel like the training not only helped them improve their listening skill but also benefit their production ability since they used the exposed model voice to imitate the target sound.

Taken together, these findings support the claims of previous studies that HVPT perceptual training is effective in enhancing learners' production ability as well as perception (e.g. Bradlow et al., 1997; Hazan et al., 2005; Rochet, 1995; Wang et al., 2003). A transfer of perceptual learning to the production domain has provided support to the theory that perception and production must be connected in some way (e.g., Bradlow et al., 1999; Flege, 1995; Yamada et al., 1996). Details of the link between perception and production will be discussed in detail later.

V. Exploring whether modified HVPT perceptual training designed to raise learners' awareness resulted in greater learning than the standardised HVPT perceptual training or not.

This study adopted two training types; (a) standardised HVPT perceptual training which used only a two-alternative, forced-choice identification task (undertaken by Trained Group A), and (b) a modified version of HVPT perceptual training which used a two-alternative, forced-choice identification task and a self-monitoring task (undertaken by Trained Group B) with the aim to raise self-awareness in speech learning. This self-monitoring task has not been employed in any previous HVPT perceptual training, as far as this researcher is aware. The main purpose of this modification was to investigate whether promoting self-awareness using a self-monitoring task would yield better learning outcomes.

Based on the perception test and production test results, there were no significant differences in overall performance between the two trained groups. Both the two trained groups demonstrated similarly significant improvements in perception and production ability. However, Trained Group B showed more benefit than Trained Group A in three ways. First, Trained Group B outperformed Trained Group A (Trained Group A: +12%; Trained Group B: +16%) in perception accuracy. Second, when looking at individual performance, all participants in Trained Group B benefitted from their perceptual training whereas great variations of improvement gains were observed in Trained Group A. In Trained Group A, some participants made noticeably large improvements while others did not successfully learn the target sounds. Some made no gain and showed a decrease in identification accuracy after the training. This unequal gain was not observed in Trained Group B. Third, the generalisation tests demonstrated that sufficient evidence of a transfer of perceptual learning to new words and a new talker occurred only in Trained Group B.

A self-monitoring task has been shown to promote positive learning success in Hirano-Cook (2011). Results obtained from this study are compatible with Hirano-Cook's findings in that this study has also found that self-monitoring can raise learners' perceptual awareness and contribute to boosting L2 speech learning. Self-monitoring presented individual feedback to specific learners in order to compare their own productions with model productions so that they could understand how accurate or inaccurate their productions were. Although this study has not provided data concerning what effect self-monitoring has on the learning process in terms of mechanisms of learning, it has presented enough to evidence support the proposition that a self-monitoring task can yield better outcomes in speech learning.

To conclude, the results of this study suggest that HVPT perceptual training is effective in modifying the perception and production of Japanese fricative and affricate contrasts. Moreover, the perceptual learning gained from the training has been proven to be maintained in the long-term memory. However, conducting perceptual training together with a self-monitoring task facilitated better learning than HVPT perceptual training alone in terms of the higher degree of improvement in perception, the benefits discerned in the majority of learners, and the transfer of generalisation to new words and a new talker at least for Thai learners of Japanese fricative and affricate contrasts. The findings suggest that "self-monitoring" tasks should be combined with HVPT perceptual training to maximise the effect on L2 speech learning. However, it is premature to draw firm conclusions since more research is still required in order to make a stronger claim about the efficacy of self-monitoring in facilitating L2 speech learning.

VI. Will there be any link between perception and production?

With the hope that the findings of the present study can shed more light on the link between the perception and production, this study will now look into the mechanisms that exist to connect these two modalities. While a minority of researchers have contradicted these claims (see Chapter Two, 2.2.2), many previous studies have suggested that if there exists a close link between perception and production, then an improvement in perception would lead to improved production. This study's results are generally in line with the findings of these previous studies, which state that perceptual learning gained through HVPT perceptual training leads to improvements in production (e.g., Bradlow, Akahane-Yamada et al., 1999; Bradlow, Pisoni et al., 1997; Rochet, 1995; Yamada et al., 1996). Findings such as these seem to strengthen the claim that there may be an inherent link between speech perception and production.

There have been further strong claims made that perception necessarily precedes production (e.g., Flege, 1991; Llisterri, 1995). The results of this study also suggest that perception precedes production for all contrasts (see Tables 4.4 and 4.5), indicating that Thai learners perform better in identifying than producing the Japanese fricative and affricate contrasts.

A closer and more nuanced analysis of the relationship between perception and production will subsequently be discussed here, with reference made to the results found in this study. A correlation analysis showed that at pre-test there was no reliable relationship observed between perception and production for both trained groups. This suggests that there was no evidence of a pre-existing relationship between the two modalities before the training took place. Most contrasts were well perceived despite relatively inaccurate production. This finding seems to contradict the findings of research mentioned earlier and implies that the relationship between L2 perception and production is by no means clear and consistent, and that accurate L2 perception may not, in fact, be sufficient for accurate production - at least at the pre-test phase. Similar findings have also been reported in previous training studies (e.g., Bradlow et al., 1997; Iverson et al., 2012; Kartushina et al., 2015; Peperkamp & Bouchon, 2011; Pereira, 2013). Additional studies have also suggested that perception and production may proceed independently in L2 speech learning (e.g., Hattori & Iverson, 2010; Kartushina & Frauenfelder, 2014; Peperkamp & Bouchon, 2011).

However, it must be noted that it is possible that the relationship between the two modalities can change over a long period of time in complex ways (Baese-Berk, 2010; Strange, 1995). In this study, after the training was undertaken, post-test scores showed relatively high correlations between perception and production for certain sounds such as [(d)z] and [ts] and a trend of high correlation between overall contrasts was observed. The findings here suggest that, although identification accuracy was not significantly related to production performance before the training, afterwards, due to the effects of perceptual training, there was a stronger link observed between the two domains. The correlation results found in the post-test seem to support the view that a sound must be accurately perceived in order for it to be adequately produced (Flege, 1987; 1991). However, the findings here are still not strong enough to make a claim of the existence of a close relationship between the two modalities. It can only be predicted that L2 perception and production processes may share some common underlying representations and perceptual-based training might have an effect on improving accuracy in the production domain. However, since the relationship between L2 speech perception and production is a complex issue, more investigations looking more thoroughly at these two domains are needed in order to present further empirical evidence of the relationship between learning in perception and production (Bradlow et al., 1997; Hattori, 2009).

Additionally, the differing degrees of improvement across individuals must also be taken into consideration. Overall, examination at the individual level revealed that perceptual training lead to improvements in perception and production in most participants. However, there was a considerable variability across participants observed. Some participants demonstrated a strong link between their learning in perception and production, showing that improvement in one modality helps in the learning of the other modality. On the other hand, some participants did not show the same trend of improvement. Individuals whose perception improved were therefore not guaranteed to make improvements in their production. This finding mirrors previous findings that have suggested that improvements in perception and production do not systematically progress at equal rates between individuals and is in line with Bradlow et al. (1997) and Kartushina et al. (2015).

The findings of this study show that the relationship between perception and production is still unclear and proves to be a more complex mechanism than is sometimes assumed. However, this study provides some evidence that speech perception is somehow modifiable through training and that the perceptual learning gained was transferred to the production domain, suggesting that there may be a link between the two modalities. However, much work remains to be done regarding the relationship between these two domains and individual differences among participants also need to be taken into consideration.

5.3 Pedagogical implications

The findings obtained in this study have provided insightful ideas about the possible modification of the perception and production of Japanese fricative and affricate contrasts by Thai learners with an aim to contribute to Japanese speech learning and teaching. This section will conclude the overview of this dissertation by stating contributions and establishing future directions and key findings for Japanese second language acquisition and computer-assisted language learning.

I. Difficulty level and improvement of specific contrasts

- Difficulty level -

This part will lend some support to the L2 speech learning theories that were described in Chapter Two. Based on the Speech Learning Model (Flege, 1995), learners should demonstrate ease when identifying sounds that are new, whereas L2 sounds that have similar phonetic properties but are not the same as an L1 counterpart are predicted to be difficult to acquire. In order to gain more insight into what Japanese sounds posed difficulties for Thai learners before receiving the training, it is worth looking at the pre-training performance of the participants. Table 5.1 summarises predicted difficulties according to interpretations of the SLM model. Table 5.2 displays the pre-test mean correct identification scores as classified by contrast and group. Table 5.3 displays the pre-test mean production accuracy scores.

Table 5.1 Difficulty degree as predicted according to the SLM model.

[(d)z]	[ts]	[tɕ]	[(d)ʑ]
new-easy	new-easy	similar-difficult	new-easy

Table 5.2 Pre-test percentage of correct *identification* scores classified by contrast (degree of improvement gained after training).

	[(d)z]	[ts]	[tɕ]	[(d)ʑ]
Control	77 (+7)	47 (+6)	57 (+4)	70 (+5)
Trained A	85 (+1)	58 (+18)	60 (+10)	66 (+18)
Trained B	82 (+11)	81 (+0)	49 (+33)	72 (+20)

	[(d)z]	[ts]	[tɕ]	[(d)ʑ]
Mean	81	62	55	69

Table 5.3 Pre-test percent correct *production* scores classified by contrast (degree of improvement gained after training).

	[(d)z]	[ts]	[tɕ]	[(d)ʑ]
Control	58 (+6)	27 (+0)	38 (+0)	42 (+5)
Trained A	51 (+16)	19 (+15)	38 (+3)	61 (+10)
Trained B	74 (+10)	45 (+9)	46 (+7)	63 (+10)
Mean	61	30	41	55

As seen in Tables 5.2 and 5.3, Thai learners had the most difficulty in accurately identifying and producing [ts] and [tɕ]. It is to be noted that SLM accurately predicted that [tɕ] would be difficult or may take more time to acquire. The similarity between the Japanese and Thai sounds possibly hinders learning (the Thai [tɕ] is palato-alveolar whereas the Japanese [tɕ] is alveolo-palatal). The SLM failed to predict that they would have difficulty with [ts]. [ts] does not exist in the Thai affricate inventory and it should be easy to acquire according to the SLM model. However, apart from the high identification accuracy observed in Trained Group B, the results of testing showed that Thai learners have great difficulty in perceiving and producing [ts]. Flege (1997, p. 81) acknowledges that articulatory complexity and linguistic markedness contribute appreciably to L2 perception and production errors. One further explanation might also be relevant to the effects of markedness. According to Eckman's (1977) Markedness Differential Hypothesis as cited by Colantani et al. (2015), structures that are absent from a learner’s L1 are typologically marked and will be acquired with greater difficulty. The alveolar affricate [ts] is more marked since this sound does not occur in most languages (Yamakawa & Amano, 2013). However, when looking at the degree of improvement over time, [ts] improved greatly after the training and showed larger gains than [tɕ]. This finding is compatible with the Similarity Differential Rate Hypothesis which states that dissimilar sounds between L1 and L2 are acquired faster than similar sounds (Major & Kim, 1996). The contrast that demonstrated the highest accuracy was [(d)z] with a mean of more than 81% in perception and 61% in production at the pre-test phase. This indicates that Thai learners already perceived [(d)z] well even before they started the training. As for [(d)ʑ], it showed a high degree of difficulty in perception and production. However, after receiving training, its perception significantly improved and it was produced to a more accurate degree.

Overall, examination of the results shown above demonstrates that [ts] and [tɕ] pose considerable difficulties for Thai learners to perceive and produce. This indicates that instruction should focus and give more attention especially to [ts] and [tɕ].

The table below summarises strategies obtained from the follow-up interview used by Thai learners to grasp the Japanese fricative and affricate contrasts.

Table 5.4 Learning strategies used by Thai learners.

[(d)z]	as	[s]
	as	English [z]
[ts]	as	[s] (Most frequent trend)
	as	[z]
[tɕ]	as	Thai [tɕʰ] (Mostly used in initial position)
	as	Thai [tɕ] (Medial and final position)
[(d)ʑ]	as	Thai [tɕ]

From the chart above, it can be seen that there is a high possibility that Thai learners do not establish new sound categories for each Japanese sound, rather they tend to map those sounds onto their L1 sound system. It is clear that the L1 influences the way L2 sounds are perceived (Best, 1995; Flege, 1995). In accordance with results from prior studies, Thai learners substitute the voiced [(d)z] mostly with the voiceless [s] (Higashi, 1986; Kawano, 2014; Sukegawa, 1993). An investigation into acoustic analyses in Figure 4.13, also demonstrated the same substitution. For example, TLA2 produced [zamama] as [samama] before receiving the training. However, there were some cases showing that participants perceived the sound as similar to the English [z]. This indicates that some Thai learners do not completely lose the ability to perceive the voicing contrast. For the substitution trend for [ts], the table shows that there are two sounds, [s] and [z], used as a substitute for [ts]. However, [s] substitution is more dominant, followed by the voiced [z]. For [tɕ], there are two trends of substitution, that is; if [tɕ] appears in word-initial position, Thai learners perceive it as [tɕʰ]. By contrast, if [tɕ] appears in a medial or final position, they tend to perceive it as the Thai [tɕ]. Lastly, for the voiced [(d)ʑ], Thai learners frequently substitute the sound as the Thai voiceless [tɕ]. Most substitute trends are in line with previous studies shown in Table 2.4 (e.g., Konishi, 2005; Sukegawa, 1993). The main problems observed here are that *two* new Japanese sounds were classified as *one* single sound (the Japanese [tɕ] and [(d)ʑ] as the Thai [tɕ] or; the Japanese [(d)z] and [ts] as [s]). It is clear that some Thai learners do not form a separate category for new Japanese sounds. Moreover, they tend to assimilate multiple new sounds into one Thai sound category. There is also evidence of loanword orthography. A number of loanwords from Japanese which are originally pronounced with Japanese [tɕ] and [(d)ʑ] are adapted into Thai [tɕ], for example, "餅（もち）" [motɕi] becomes [mo*tɕ*i], and "富士（ふじ）" [ɸɯʑi] becomes [fu*tɕ*i] in Thai (Chusri, 2013; Bunnag, 2014). However, if both sounds are perceived as belonging to the same phonological category, then

L2 learners will fail to distinguish these segments (Brown, 2000). The main reason for this trend is because Thai has fewer fricatives and affricates than Japanese does so Thai learners simply try to adapt all those new sounds into the same categories as those that exist in Thai. According to PAM (Best, 1995), when two L2 contrasts are classified as equivalent to a single L1 consonant category, a so-called "single-category assimilation", learning will be the most difficult for the L2 learners to acquire. From the follow-up interviews conducted, most Thai learners reported that they have difficulty in discriminating the alveolar palatals [(d)ʑ], [tɕ] and [ɕ] of Japanese from each other. This seems to indicate that they have not formed separate categories for some Japanese sounds due to the interference of their native language.

In summary, the above shows that Thai learners assimilate some Japanese sounds into their L1 categories and these substitutions are likely to lead to learning difficulties. As Flege (1995) suggests, L1 interference seems to be a factor in the perception of L2 sounds and, perhaps, production ability. However, the ability to discriminate non-native phonemic differences are not permanently modified or lost in adults (Pisoni, 1994). In other words, the success of L2 perceptual learning depends on the degree of reorganising those phonetic categories or on the formation of new sounds. Nonetheless, if Thai learners form accurate categories for new L2 sounds then there is a high possibility that they can succeed in acquiring the L2 contrasts.

Improvement and learning gained due to the training

Strange (1995) states that, "perception of L2 contrasts may continue to improve for several years. However, some perceptual difficulties may persist even after production of non-native phonetic segments is mastered" (p. 79). In the present study, the gains from the training also show contrast-specific effects. Some contrasts achieved near native-like levels of accuracy whereas some did not make any changes in the two modalities even after the extensive training. Some contrasts improved only in perception but not in production and vice versa.

For example, in Trained Group B, [(d)z] and [ts] initially showed similar identification accuracy scores. However, while [(d)z] showed a significant trend of improvement after the training, [ts] demonstrated no change from the pre-test. In production, according to the goodness rating results, [tɕ] and [(d)ʑ] initially showed equal pre-test mean scores. However, at post-test [(d)ʑ] improved to a greater extent whereas [tɕ] was more poorly produced after the training. It is surprising that the production of [(d)ʑ] was more accurate and more native-like than [tɕ] despite the fact that the voiced [(d)ʑ] is more marked[25] for Thai learners to acquire since there is no voicing contrast of fricatives and affricates existing in Thai. An explanation for the production difficulty of [tɕ] could be that L1-based perceptual assimilation that leads to non-target-like performance (Flege, 1995). The SLM model hypothesises that the more similar sounds are to an L1 category, the more difficult it is that sounds are acquired. This may possibly explain why Thai learners are more successful at perceiving and producing [(d)z] and [(d)ʑ] than [tɕ]. From the results, it can be interpreted that trained Thai learners may be creating new categories for the voiced [(d)z] and [(d)ʑ] sounds since they started to perform much better in identifying and producing the target sounds after the training. On the contrary, [tɕ] caught the most attention since [tɕ] showed a relatively large improvement in perception but its production accuracy in both trained groups decreased. The results show that although the perception of [tɕ] is more accurate after training, participants still struggle when producing the contrast possibly because they produce it in their L1 articulatory manner. As Strange (1995) suggests, training using stimuli and tasks that emphasise "equivalence classification" can lead to significant and lasting improvement.

[25] It is widely believed that cross-linguistically voiced obstruents are more marked and harder to articulate than voiceless obstruents. Especially when there is no equivalent corresponding sound in L1(Eckman, 1977; Johnson, 2012).

Hence, more focus is needed for [tɕ] and its improvement may require extended training sessions. In the case of [ts], Trained Group A improved significantly in perceiving and producing after receiving the training.

Taken together, regarding these differences in learning outcomes, it is possible that the differences in improvement arose due to differences in the training types that the group received. Strange (1992) suggests that factors such as training task variables - specifically L2 learning instruction experience, types of tasks used in laboratory procedures (identification tasks and discrimination tasks), types of input, speakers' variability and sets of stimulus - may explain the great variability in the results of L2 speech perception studies. Moreover, the initial degree of difficulty when participants conducted the pre-test might also explain the complex outcomes. Contrasts that were produced more accurately at the pre-test such as [(d)z] and [ts] (observed in Trained Group B with a mean accuracy above 81%) tended to benefit less from the training than other contrasts that showed less accurately at pre-test, and vice versa. For example, [tɕ], which showed large gains of 33% increases observed in Trained Group B and benefitted most in this training, was the contrast that was pronounced least accurately prior to training when compared to other contrasts. Previous studies have similarly shown that higher pre-test score is associated with less improvement (e.g., Bradlow et al., 1997; Kartushina et al., 2015). These unequal gains of learning are also compatible with the findings of Thomson (2008) which investigated the relationship between English vowel identification and production by Mandarin learners of English and found that some sounds are easier to perceive and learn than others.

II. Individual differences in learning

Researchers in the field of SLA have long questioned to what extent various individual differences between learners predetermine their rate of learning, ultimate attainment and response to particular instructional conditions. Previous training studies have suggested that the gains of training tend to be spread unevenly across participants (e.g., Bradlow, Akahane-Yamada et al., 1999; Bradlow, Pisoni et al., 1997; Burnham, 2013; Goto, 1971; Hazan et al., 2005; Kartushina et al., 2015; Lengeris, 2009; Sheldon & Strange, 1982; Yamada et al., 1994). The results of the current study are in line with these previous studies, demonstrating that there was evidently a large variability in performance among participants both before and after training, implying that benefits gained from the training also varied according to individuals.

With the above in mind, the perception results of this study will be next discussed in further detail. The results from Trained Group A indicate that perceptual training did not result in increased perception performance for all participants. Some participants made notably large improvements with a 44% increase, while others made only very small gains or failed to successfully learn the target sounds. On the contrary, Trained Group B showed more steady and consistent improvements in identification accuracies, indicating that all participants from Trained Group B clearly benefitted from the training. Also, after a closer inspection of improvement by contrast, Trained Group A improved in certain sounds but a similar trend was not observed in Trained Group B. It could be suggested that the task variable may be an important factor in predicting learning success of the target contrasts. An explanation for the unequal gains between groups could be that the training type used in the training might not have been a good match for the participants' learning styles who did not make any gains, as stated by Hanulikova et al. (2012). The results of this study also suggest that the training type conducted in Trained Group B might contribute to the more equal gains across participants. Moreover, several previous studies have demonstrated that weaker learners benefit more from training since they have more room for improvement. This study also showed the same trend that participants who had a lower initial pre-test score benefitted from the training more than those who started with high levels of accuracy in the pre-test (e.g., Pereira, 2013; Wang, 1999; Yeon, 2004).

Next, the production results will be discussed. An examination of the production data shows that the degree of improvement in production accuracy varied across trained participants. In Trained Group A, the size of production improvement gain by individual participants varied from -4% to +24% with a mean of 11% score change. In Trained Group B, the difference between pre-test and post-test production scores by individual participants ranged from -2% to +17% with a mean of a 9% increase. Some participants made noticeably large improvements with more than 20% points of increase whereas some made very small gains with less than 5% points. Two participants did not make any gains in production accuracy (one from each of the trained groups).

In Hazan et al. (2005), it was reported that differences in improvement ranged from -5% to +48% in perception scores and -11% to +20% in production scores across individuals. Previous research suggests that differences in the effectiveness of perceptual training across individuals could also be ascribed to differences in the strategies of individual learners (e.g., Hazan et al., 2005; Thomson, 2012). Thomson (2012) investigated the effect of high variability phonetic training on 10 English vowels by Mandarin-speaking learners. The conclusions of this study inferred that learning strategies used by participants might contain individualistic variations which could thus explain variations in success rates across learners. For example, it is possible that some learners may have directed their attention more towards vowel duration during the training, while others may have focused more on things such as spectral properties. In addition, Perrachione et al. (2011) also state that training is beneficial or not depending on individual differences in the learners. They found that high variability training typically promoted learning among individuals with strong perceptual abilities or those who had high perceptual scores before the training. Learners with weaker perceptual abilities were impaired by high variability training relative to a condition of low variability. There are several factors identified by prior research which it has been suggested may contribute to variations in outcomes of the acquisition of L2 speech. These are things such as; learner's motivation to engage with the training and improve; the quality of training given; the amount of L2 input; socio-psychological factors such as personality, general intelligence and age of language acquisition; personal variables including extraversion, anxiety, linguistic insecurity and self-efficacy; cognitive characteristics including phonological memory, metalinguistic capacities (e.g., phonological awareness), analytical abilities (e.g., analytical reasoning, mimicry), etc. (Birdsong, 2006; Colantoni et al., 2015; Flege, 2002; Hu et al., 2013; Moyer, 1999; 2004).

As discussed already, there are many factors that influence learning outcomes, it is, however, as yet unclear what specific factors determine this individual difference. Even at the onset of L2 speech acquisition, when several factors are controlled, large individual differences are often observed. And the causes of these individual differences are still unclear (Ellis, 2004). Although the number of those who did not successfully learn the target sounds reported in this study is still considered to be minimal (Perception test: Two participants from Trained Group A; Production test: One from Trained Group A and one from Trained Group B), further investigation is still needed to look into what factors specifically cause this unequal effect from the training. Moreover, it seems important to consider initial performance before the training starts and look at how this relates to degrees of improvement. Taking these factors of individual differences in speech learning into consideration may allow for the development of one or more training paradigm designs that will maximally benefit all learners.

III. HVPT Perceptual Training as a tool for teaching Japanese speech

"And the mystery is why HVPT — a simple, quick, and inexpensive technique for helping adults to learn the sounds of new languages — is not widely used. In fact, as far as I can tell, it's not used at all. Over the years, I've asked many people in the language-teaching business about this, and the answer has always been the same. It's not "Oh yes, well, we tried it and it doesn't really work"; or "It works, but the problems that it solves are not very important"; or "I'd like to, but it doesn't fit into my syllabus". Rather, their answer is some form of "What's that? I've never heard of it." So I'm puzzled. As I mentioned at the beginning of this post, I've been asking language-teaching professionals about this since 1992 or so, when I first heard about the technique. And I've never run across one who's heard of the idea. Maybe in the end HVPT doesn't make enough impact on overall language-learning progress to be worth doing. But if I had to bet, I'd put my money the other way. There are many other obvious questions to ask, some of which have no doubt been answered in research that I don't know about. One that comes to mind is the role of variation due to discourse and sentence context, as opposed to variation due to phonological context and speaker differences. But for me, the biggest question is a sociological one: why the big disconnect between research and practice?"

Written by Liberman (2008) as cited by Pierce (2014)

As stated above, although HVPT perceptual training is highly efficient and successful, few have extended such findings and methods into the classroom environment (Pierce, 2014). The experiment presented in this study represents one small step towards developing an online learning tool for Thai learners of Japanese to improve perceptual learning of fricative and affricate contrasts. This study has attempted to show the effectiveness of HVPT perceptual training in improving identification ability in perceiving L2 contrasts, such as the Japanese fricative and affricate contrasts, and the transfer of perceptual training to significant production improvements despite lacking any explicit production training. Moreover, the learning gained was also maintained in the long-term memory of participants after the completion of their training. In addition, the analysis of the responses given to the questionnaire and a follow-up interview indicate that participants considered perceptual training a good means to improve their speech learning and they reported positive experiences with the training. Most participants said that they felt motivated and they wanted more training sessions. One participant said that exposure to multiple native speakers was confusing but the rest liked the opportunity to be exposed to multiple native speakers' productions. In other words, the data here highlights the effectiveness of HVPT perceptual training in improving L2 speech learning. In response to these findings, it is the opinion of this researcher that HVPT perceptual training should be implemented in training Thai learners to better learn the Japanese fricative and affricate contrasts. As Thomson (2012, p. 1253) proposes, since HVPT perceptual training will work across L1 groups, it might be time to give greater attention towards expanding HVPT applications to a greater number of learners. And because of its potential, HVPT perceptual training is a tool that should be further developed and more widely integrated into CAPT applications, or in a language teaching and learning classroom. However, further studies are needed to develop the efficacy of HVPT perceptual training. Subsequent to all of the above, it can be stated that there is now evidence to conclude that Thai learners can benefit from this training and it may be anticipated that the development of such perceptual training would be an important contribution to Japanese language teaching and learning.

After a review of the relevant textbooks in Chapter Two, it is clear that most textbooks remain underdeveloped, since an explanation of pronunciation is commonly limited to verbal indications and there is a lack of instant feedback for performance. Moreover, the results of this study showed that there exists a large variety of learning types across individuals. These individual differences cannot be managed well by only one teacher. It is thus clear that a new method of pronunciation teaching which can target a practise at specific target sounds, or at those learners who especially need to improve, is needed. In other words, a more complex tool, such as computer-based training, which can deal with a variety of learners and learning types is needed. As Thomson (2014) states, computer-based approaches may promote greater learner autonomy and most importantly, afford the possibility of individualised instruction.

In order to fulfill the full potential of HVPT perceptual training, there remain some suggestions which this researcher has observed and that might be useful for future studies. Table 5.5 summarises suggestion points for future studies to develop the efficacy of HVPT perceptual training.

Table 5.5 Suggestions for enhancing the efficacy of HVPT perceptual training.

Content	In this study	Future studies
Session	Nine sessions with a fixed schedule.	- Some conflicting schedules occurred with the participants because of the long-term commitment of the experiment, suggesting that longer sessions might not be preferable. Sessions below 10 times might be an ideal. - Extended sessions to focus on problematic contrasts might be necessary. - Flexible time schedules as conducted in Wang & Munro (2004) are advised.
Duration	10 to 15 minutes.	10-15 minutes of training may not be efficient. A longer duration of training may be preferable (20-30 minutes).
Training materials	Natural stimuli of nonsense words and real words presented in isolation.	Both isolation and sentence context is also needed.

Content	In this study	Future studies
Talkers	Four talkers.	A larger number of talkers and more variability in the age of talkers would be preferable.
Feedback	- Maximum three times. - Given only when their answer is *incorrect.*	- No limitations in listening to feedback. - Offer feedback both when the answer is *right* or *wrong*.
Technical problems during the training	- The audio sometimes could not be heard. - Frozen screen.	- The program should be improved to be more reliable. - More complex functions are needed. - To adapt some small details to make the training more entertaining such as decorations to motivate participants.

With the above in mind, and dealing first with the number of training sessions, it is difficult to draw firm conclusions about the right amount of training sessions. As Thomson (2014, p. 11) states, length of training is related to the number of features targeted or the scope of instruction, and global improvement in intelligibility and comprehensibility requires weeks or even months of instruction. Previous studies suggest that five to ten sessions of training tend to show the most improvement and a number of more than 15 sessions is not recommended (Logan & Pruitt, 1995; Iverson et al. 2012). The participants of this study were exposed to a nine-session training course and this seems to be a suitable number of training sessions in terms of sufficient gains and practicality. However, a flexible schedule might be more pleasant for participants than a fixed schedule since some participants have problems with long-term commitments. Some could not carry out the training on the scheduled date because of personal commitments to other things. Regarding the training duration, training with longer durations might amplify the magnitude of the effect. In this study, one session lasted approximately 10 to 15 minutes depending on the participants' speed of completion. Lengeris &

Hazan (2010) conducted five sessions that lasted about 45 minutes. Participants were exposed to 225 trials in each session. After the training, they demonstrated an improvement of about 20% points in identification accuracy. This implies that longer durations but lesser numbers of sessions might yield similar gains to those found in this study or perhaps even greater. To summarise, the recommendation here is to train learners not more than nine sessions but with longer durations of 20 to 30 minutes in each session.

Regarding training materials and talkers, this study used natural speech materials uttered by four Japanese speakers (two males and two females). However, using speech materials produced by more talkers than four people, or speakers of different ages, might lead to more robust learning outcomes since it might strengthen a more robust and stable representation of the target sounds. In this training, 480 tokens were presented over nine sessions of training. In Iverson et al. (2012), participants received a total of 225 tokens in each session with a total of eight sessions altogether. After the training, they demonstrated an improvement in identification accuracy of about 21% for the inexperienced group and 17% for the experienced group. This means that giving larger number of tokens in each session could possibly yield better learning outcomes. Next, all tokens were presented in isolation in the current study. However, in order to exploit a greater variety in training tasks, apart from the isolation presentation, training with the target words in carrier sentences might give participants more motivating tasks and wider forms of training. For feedback, Burnham (2013) states that, “from a pedagogical perspective, it makes sense to give subjects as much control over the feedback they receive as possible” (p. 115). This study gave feedback a maximum of three times. However, allowing participants to rehear the tokens without limitations could have allowed them to spend more time on the contrast that they want to work on. Last but not least, regarding technical problems, more development is still needed to be done to strengthen the training system, since some participants reported facing problems during the training, such as frozen screens or no sound.

IV. The importance of self-monitoring

A crucial point of focus of the present study is the exploration of the self-monitoring task conducted in addition to HVPT perceptual training. Recently, there have been many studies which have investigated modifying the training paradigm with the aim of resulting in greater learning outcomes than the standardised HVPT perceptual training. For example, Iverson et al. (2005) and Thomson (2012) modified training stimuli with the aim to draw the learners' attention to important phonetic properties that might be ignored via natural input by using synthetic stimuli. Both of them found no advantage in modifying HVPT stimuli in training L2 learners and concluded that natural variation is all that promotes learning. The explanation for why no advantage was found was possibly because these previous studies have not looked at other aspects, such as the role of attention, namely "self-monitoring". Self-monitoring and awareness-raising have been shown to promote positive successes in second language speech learning in previous studies (Couper, 2011; Ingles, 2011; Nagamine, 2011; Sardegna, 2011). This study conducted HVPT perceptual training together with a self-monitoring task with the aim to provide further enhancement of HVPT perceptual training. The task offered participants the opportunity to compare their own productions with that of model native speakers and to compare whether they are the same or not. The trained group which carried out the self-monitoring task showed greater improvements in perception and generalisation to new words and a new talker than the group which received standardised perceptual training. In addition, from the questionnaires and follow-up interviews, the participants from Trained Group B responded positively to the task and became strongly aware of the importance of the self-monitoring strategy. These results show that promoting stimuli produced by multiple native speakers together with giving a self-monitoring task might yield more positive results in learning. The results obtained in this study highlight the role of “self-monitoring” in promoting L2 speech learning. As suggested by Gilakjani & Ahmadi (2011, p. 79), L2 learners need ample opportunity to listen to their own speech in comparison with that of native speakers and to learn to distinguish the aspects of learner pronunciation that make comprehension difficult for native speakers. A major benefit of the self-monitoring task possibly is its capacity to focus participants’ attention on their own performance/production. It is highly recommended that future training studies adopting HVPT perceptual training should also combine a self-monitoring task in order to maximise the effect of perceptual learning and see whether it plays a role in L2 speech learning. If it is indeed the case that self-monitoring is effective in promoting the learning of L2 perception and production then this perceptual training should be expanded to include more advanced and more multi-faceted functions. Training could also be developed to record learners’ productions in real time and then give an opportunity to compare that with a native speaker’s model sounds. In conclusion, the findings of this study shed light on issues related to the role of self-monitoring in second language speech acquisition.

Chapter 6 Conclusion

6.1 Summary of the results

This dissertation has investigated the effects of HVPT perceptual training on the learning of Japanese fricative and affricate contrasts by twenty Thai native learners of Japanese. This chapter will conclude the main findings obtained from this study, then will end by exploring the limitations of the present study and directions for future research. It is my hope that this study will provide some evidence towards identifying the most effective method of language pedagogy for Japanese language teaching and learning.

This study has provided empirical evidence that after a nine-session period of HVPT perceptual training, the perception and production ability of the trained groups improved significantly. These results are compatible with previous training studies and demonstrate that improvements in perception and production can be achieved by L2 adult learners using HVPT perceptual training. Also, it is clear that this HVPT perceptual training works well with Thai learners of Japanese in improving their learning of Japanese fricative and affricate contrasts. The main findings obtained from this current study are summarised below.

Perception

In summary, according to the results of perception tests, HVPT perceptual training was shown to be effective in leading to improvements in the perception of Japanese fricative and affricate contrasts among Thai learners. The results indicated that all groups including the untrained participants significantly improved their ability to perceive the target sounds from pre-test to post-test (+6% in Control Group; +12% in Trained Group A; and +16% in Trained Group B). Although all groups' perceptual ability improved, both of the trained groups significantly outperformed the control group. Moreover, the two trained groups produced similarly positive results in terms of overall improvements in perception, however, the perceptual learning gained in Trained Group B was shown to be more robust. All participants in Trained Group B benefitted from the training whereas the participants from Trained Group A showed inconsistent individual gains in learning outcomes (some made large improvements and some did not make any improvement at all in learning).

To summarise, the results suggest that HVPT perceptual training resulted in overall group improvement in the identification ability to perceive Japanese fricative and affricate contrasts but there were also considerable individual differences in performance.

Generalisation

There were two types of generalisation tests conducted in this study. The first test used old tokens from the pre/post-test produced by a familiar talker used in the training and an unfamiliar talker not used in the training. The results showed that trained participants' ability to identify the pre/post-test tokens produced by an unfamiliar talker was comparable to their ability to identify the tokens produced by a familiar talker. The results of the first generalisation test thus showed that perceptual learning had transferred to a new talker producing previously trained words. The second round of generalisation tests was undertaken to see whether there were any more robust generalisations made to new words produced by an unfamiliar talker, these additional two generalisation tests were conducted using a large range of 64 new untrained tokens produced by both a familiar talker and an unfamiliar talker. However, the second round of tests of generalisation showed sufficient evidence of a transfer of perceptual learning only to new words and a new talker in Trained Group B. This finding implies that it is possible that HVPT perceptual training alone was not enough for generalisation to occur for a new talker producing untrained words. But, given the evidence of generalisation shown in Trained Group B, it is possible that combining HVPT perceptual training with a self-monitoring task may yield a greater possibility for robust generalisation to new words and a new talker.

Retention

The delayed post-test results of this study indicated that the gained improvements in perception for trained groups were maintained six months after the training was over. This suggests that trained Thai learners have made long-term memory representations for the new L2 sounds after training.

Production

A further aim of this research has been to investigate whether perceptual training has helped to improve TL's production of target sounds. Two evaluation procedures were undertaken to judge production performance; a production accuracy rating and a goodness rating. For both evaluation procedures, all participants' productions were evaluated by five native speakers. The results showed evidence that both Trained Group A and Trained Group B significantly improved their production accuracy of the Japanese fricative and affricate contrasts. Such improvement was not found in the control group. This indicates that HVPT perceptual training has also lead to an improvement in the production of Japanese fricative and affricate contrasts even when no production training is provided. Moreover, the findings also imply that there might be a close link between speech perception and production.

As summarised above, the overall results of the present study were in general positive and are in line with much of the previous research on the positive gains obtainable from learning through HVPT perceptual training. Both standardised HPVT perceptual training and the modified HVPT perceptual training yielded positive effects in modifying Thai learners' perception and production of the Japanese fricative and affricate contrasts. The perceptual learning gained was also maintained for six months after the completion of the training. Apart from the results of the generalisation tests, participants from both the two trained groups performed equally well in perceiving and producing the target sounds. In other words, it can be emphasised that HVPT perceptual training is a potent tool for enhancing better learning of the perception and production of Japanese fricative and affricate contrasts by Thai learners. However, and perhaps more importantly, modifying the perceptual training to include a self-monitoring task may lead to even better learning outcomes. These findings strongly support the view that HVPT perceptual training is suitable to be used in a classroom environment for Thai learners of Japanese.

6.2 Limitations and directions for future research

The studies presented here mark only a first step into a thorough investigation on the effects of HVPT perceptual training on Thai learners of Japanese. The limitations noted below suggest important directions for future perceptual training studies in general.

First, one major limitation of this study is that the participants' language proficiency was not controlled strictly. Recruiting a large number of participants was one of the most challenging tasks in this study. Although their initial pre-training performance was not significantly different across participants, it would still be more preferable if all participants had had more similar proficiency levels. As for the number of participants used for each training condition, for future studies it would also be preferable to have an equal number of participants with similar language experience in each group so that the effects of individual differences can be minimised. In addition, it would be advisable to aim for a larger number of participants, since it would allow for greater confidence in the results and findings. Moreover, it is of interest to see whether HVPT perceptual training would work effectively using other consonants and vowels.

For production, judgements should be evaluated not only in terms of the intelligibility of specific contrasts but also the overall intelligibility of larger segments should be examined. Moreover, delayed post-tests of production accuracy and further production generalisation tests should be conducted in future investigations in order to examine whether the participants' improved levels of production performance can be maintained long after training has ended, and whether perceptual learning can be generalised to the production of new words or not. The lack of these elements was mainly due to reasons of practicality as further retention tests would have required scheduling with participants whose calendars were already quite tight and another trip back to Thailand.

Regarding the training methodology, training may be effective to a greater extent if each session duration was made longer as was explained in Table 5.5. The amount of training may have been insufficient when using a period of only 10 to 15 minutes in each session. In addition, more functions for the website should also be developed to keep track of progress. It is necessary to further develop the training software to include a training log that can record the date, the time and the number of training blocks. Even a further function that can automatically focus on each individual's progress to encourage more usage of the training would be beneficial.

In conclusion, as a contribution to Japanese language teaching and learning, this dissertation provided a detailed examination of the effect of HVPT perceptual training on Japanese fricative and affricate contrasts among Thai learners of Japanese. HVPT perceptual training is successful in improving the perception accuracies of Japanese fricative and affricate contrasts by Thai learners and also has benefits for their production ability. They maintained their improvement in perception six months after the training. More importantly, combining a self-monitoring task together with HVPT perceptual training yielded greater benefits in perception and generalisation of learning. However, as noted above, a number of interesting issues and questions still remain. More work is necessary to further validate the efficacy of HVPT perceptual training used together with a self-monitoring task. And more time is needed to understand the influence that perceptual training and self-monitoring tasks have on learners' learning processes and how they can best be trained, and also to investigate the link between perception and production.

REFERENCES

AbuSeileek, F. A. (2007). Computer-assisted pronunciation instruction as an effective means for teaching stress. *The JALT CALL Journal*, 3(1-2), 3-14.

Akagi, H., Uchida, N., & Furuichi, Y. (2010). *Mainichi renshuu! rizumu de mi ni tsuku nihongo no hatsuon* [Everyday practice! Learn Japanese pronunciation by rhythm]. Tokyo: Suriiee Network.

Aliaga-García, C., & Mora, J. C. (2009). Assessing the effects of phonetic training on L2 sound perception and production. In M. A. Watkins, A. S. Rauber & B. O. Baptista (Eds.), *Recent research in second language phonetics/phonology: Perception and production* (pp. 2-31). Newcastle upon Tyne, UK: Cambridge Scholars Publishing.

Amano, S., & Kondo, T. (2008). *NTT database series nihongo no goi tokusei* [Lexical characteristics of Japanese]. Tokyo: Sanseido.

Baese-Bark, M. M. (2010). *An examination of the relationship between speech perception and production* (Doctoral dissertation). Northwestern University. Retrieved from ProQuest Dissertations and Theses - UK & Ireland. (AAT 8918197)

Best, C. T. (1995). A direct realist perspective on cross-language speech perception. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 171-206). Baltimore, MD: York Pr.

Best, C. T., & Tyler, M. D. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In O.-S. Bohn & M. J. Munro (Eds.), *Language experience in second language speech learning* (pp. 13-34). Amsterdam: John Benjamins.

Birdsong, D. (1992). Ultimate attainment in second language acquisition. *Language*, 68(4), 706-755.

Birdsong, D. (2006). Age and second language acquisition and processing: A selective overview. *Language Learning*, 56(s1), 9-49.

Boersma, P., & Weenink, D. (2012). Praat: Doing phonetics by computer (Version 5.3.03) [Computer program]. Retrieved from http://www.praat.org/

Bohn, O.-S., & Best, C. T. (2010). Testing PAM and SLM: Perception of American English approximants by native German listeners. In K. Dziubalska-Kołaczyk, M. Wrembel & M. Kul (Eds.), *Proceedings of the 6th International Symposium on the Acquisition of Second Language Speech, New sounds* 2010, Poznań, Poland, 1-3 May (pp. 43-48).

Bongaerts, T., Van Summeren, C., Planken, B., & Schils, E. (1997). Age and ultimate attainment in the pronunciation of a foreign language. *Studies in Second Language Acquisition*, 19, 447-465.

Borden, G., Gerber, A., & Milsark G. (1983). Production and perception of the /r/-/l/ contrast in Korean adults learning English. *Language Learning*, 33, 499-526.

Bowen, D. J. (1972). Contextualizing pronunciation teaching in the ESOL classroom. *TESOL*, 1(5), 83-97.

Bradlow, R. A. (2008). Training non-native language sound patterns. In J. G. Hansen Edwards & M. L. Zampini (Eds.), *Phonology and second language acquisition* (pp. 287-308). Amsterdam, Philadelphia: John Benjamins.

Bradlow, R. A., Akahana-Yamada, R., Pisoni, D. B., & Tohkura, Y. (1999). Training Japanese listeners to identify English /r/ and /l/: Long-term retention of learning in perception and production. *Perception & Psychophysics*, 61(5), 977-985.

Bradlow, R. A., Pisoni, D. B., Akahana-Yamada, R., & Tohkura, Y. (1995). Acquisition of the English /r/-/l/ contrast by Japanese speakers: Effects of training in perception on production. *ICPhS 1995*, *Proceedings of the 13th International Congress of Phonetics Sciences vol.4,* Stockholm, Sweden, 13-19 August (pp. 562-565).

Bradlow, R. A., Pisoni, D. B., Akahana-Yamada, R., & Tohkura, Y. (1997). Training Japanese listeners to identify English /r/ and /l/: Some effects of perceptual learning on speech production. *The Journal of the Acoustical Society of America*, 101(4), 2299-2310.

Brown, C. (2000). The interrelationship between speech perception and phonological acquisition from infant to adult. In J. Archibald (ed.), *Second language acquisition and linguistic theory* (pp. 4-63). Oxford: Blackwell.

Bunnag, A. (2014). Nihongo Taigo ni okeru gairaigo no juyou ni tsuite [Loanword in Japanese and Thai]. *Hikaka nihongaku kyouiku kenkyuu sentaa kenkyuu nenpou*, 246-254. Retrieved from http://teapot.lib.ocha.ac.jp/ocha/bitstream/10083/54967/1/p.246-254.pdf

Burnham, R. K. (2013). *Phonetic training for learners of Arabic*. (Doctoral dissertation, The University of Texas at Austin). Retrieved from https://repositories.lib.utexas.edu/bitstream/handle/2152/21607/BURNHAM-DISSERTATION-2013.pdf?sequence=1&isAllowed=y

Chusri, A. (2013). Tai go ni okeru nihongo no shakuyou go to gendai no tai go bogo wa sha no ninchido [An awareness of Japanese loanwords in Thai by Thai]. *Nichitaigengo bunka kenkyuu,* 104-118.

Colantani, L., Steele, J., & Escudero, P. (2015). *Second language speech theory and practice*. Cambridge: Cambridge University Press.

Couper, G. (2003). The value of an explicit pronunciation syllables in ESOL teaching. *Prospect*, 18(3), 53-70.

Couper, G. (2006). The short and long-term effects of pronunciation instruction. *Prospect,* 21(1), 46-66.

Couper, G. (2011). What makes pronunciation teaching work? Testing for the effect of two variables: Socially constructed metalanguage and critical listening. *Language Awareness*, 20(3), 159-182.

Delvaux, V., Huet, K., Piccaluga, M., & Harmegnies, B. (2013). Production training in second language acquisition: A comparison between objective measures and subjective judgments. *Interspeech 2013*, 1, 2375–2379.

Derwing, T. M., Diepenbroek, J. G., & Foote, J. A. (2012). How well do general-skills ESL textbooks address pronunciation?. *TESL CANADA Journal*, 20(1), 23-44.

Derwing, T. M., & Munro, M. J. (2005). Second language accent and pronunciation teaching: A research-based approach. *TESOL Quaterly*, 39(3), 379-397.

Derwing, T. M., & Munro, M. J. (2015). *Pronunciation fundamentals. Evidence-based perspectives for L2 teaching and research*. Amsterdam: John Benjamins.

Derwing, T. M., Rossiter, M. J., Munro, M. J., & Thomson, R. I. (2004). Second Language fluency: Judgements on different tasks. *Language Learning*, 54(4), 655-679.

Dickerson, W. B. (1989). *Stress in the speech stream: The rhythm of spoken English*. Urbana, University of Illinois Press.

Dickerson, W. B. (1994). Empowering students with predictive skills. In J. Morley (Ed.), *Pronunciation pedagogy and theory: New views, new directions* (pp. 17-33). Alexandria, VA: TESOL Publications.

Dickerson, W. B. (2000). *Covert rehearsal as a bridge to accurate fluency*. Paper presented at International TESOL, Vancouver, BC, Canada.

Dlaska, A., & Krekeler, C. (2008). Self-assessment of pronunciation. *System*, 36(4), 506-516.

Dlaska, A., & Krekeler, C. (2013). The short-term effects of individual corrective feedback on L2 pronunciation. *System*, 41(1), 25-37.

Dugard, P., & Todman, J. (1995). Analysis of pre-test-post-test control group designs in educational research. *Educational Psychology*. 15(2), 181-198.

Eckman, F. (1977). Markedness and the contrastive analysis hypothesis. *Language learning*, 27(2), 315-330.

Ellis, R. (2004). Individual differences in second language learning. In A. Davies & C. Elder (Eds.), *The handbook of applied linguistics* (pp. 527-547). Malden, MA: Blackwell.

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

Flege, J. E. (1991). Perception and production: The relevance of phonetic input to L2 phonological learning. In T. Heubner & C. Ferguson (Eds.), *Crosscurrents in second language acquisition and linguistic theory* (pp. 249-289). Philadelphia: John Benjamins.

Flege, J. E. (1992). Speech learning in a second language. In C. Ferguson, L. Menn & C. Stoel-Gammon (Eds.), *Phonological development: Models, research, and application* (pp. 565-604). Timonium, MD: York Press.

Flege, J. E. (1995). Second-language speech learning: Theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 233-277). Timonium, MD: York Press.

Flege, J. E. (1997). The role of phonetic category formation in second-language speech learning. In J. Leather & A. James (Eds.), *New sounds 97: Proceedings of the third international symposium on the acquisition of second-language speech* (pp. 79-89). Klagenfurt, Austria: University of Klagenfurt.

Flege, J. E. (1999). Age of learning and second-language speech, In D. Birdsong (Ed.), *Second language acquisition and the critical period hypothesis* (pp. 101-132). Hillsdale, NJ: Lawrence Erlbaum.

Flege, J. E. (2002). Interactions between the native and second-language phonetic systems. In P. Burmeister, T. Piske & A. Rohde (Eds.), *An integrated view of language development: Papers in honor of Henning Wode* (pp. 217-244). Trier: Wissenschaftlicher Verlag.

Flege, J. E. (2003). Assessing constraints on second-language segmental production and perception. In A. Meyer & N. Schiller (Eds.), *Phonetics and phonology in language comprehension and production: Differences and similarities* (pp. 319-355). Berlin: Mouton de Gruyter.

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

Flege, J. E., Takagi, N., & Mann, V. (1996). Lexical familiarity and English-language experience affect Japanese adults' perception of /r/ and /l/. *The Journal of the Acoustical Society of America*. 99(2), 1161-1173.

Flege, J. E., Yeni-Komshian, G. H., & Liu, S. (1999). Age constraints on second-language acquisition. *Journal of Memory and Language*, 41(1), 78-104.

Gilakjani, A. P., & Ahmadi, M. R. (2011). Why is pronunciation so difficult to learn?. *English Language Teaching*, 4(3), 74-83.

Golestani N., & Zatorre, J. R. (2009). Individual differences in the acquisition of second language phonology. *Brain & Language*, 109, 55-67.

Gomez Lacabex, E., & Gallardo del Puerto, F. (2014). Raising perceptual phonemic awareness in the EFL classroom. *Concordia Working Papers in Applied Linguistics*, 5, 203-215.

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

Guion, S. G., & Pederson, E. (2007). Investigating the role of attention in phonetic learning. In O.-S. Bohn & M. J. Munro (Eds.), *Language experience in second language speech learning: In honor of James Emil Flege* (pp. 55-77). Amsterdam: John Benjamins.

Handley, Z., Sharples, M., & Moore, D. (2009). *Training novel phonemic contrasts: A comparison of identification and oddity discrimination training*. Paper Presented at Speech and Language Technology in Education (SLaTE) 2009, Wroxall, England, 3-5 September 2009.

Hanulikova, A., Deidu, D., Fang, Z., Basnakova, J., & Huettig, F. (2012). Individual differences in the acquisition of a complex L2 phonology: A training study. *Language and Learning*, 62, 79-109.

Hardison, D. M. (2003). Acquisition of second-language speech: Effects of visual cues, context, and talker variability. *Applied Psycholinguistics,* 24(4), 495-522.

Hardison, D. M. (2004). Generalisation of computer-assisted prosody training: Quantitative and qualitative findings. *Language Learning and Technology*, 8(1), 34-52.

Hardison, D. M. (2005). Contextualized computer-based L2 prosody training: Evaluating the effects of discourse context and video input. *CALICO Journal,* 22(2), 175-190.

Hattori, K. (2010). *Perception and production of English /r/-/l/ by adult Japanese speakers*. (Doctoral dissertation, University College London, London, England). Retrieved from http://discovery.ucl.ac.uk/19204/1/19204.pdf

Hattori, K., & Iverson, P. (2009). English /r/-/l/ category assimilation by Japanese adults: Individual differences and the link to identification accuracy. *The Journal of the Acoustical Society of America,* 125(1), 469-479.

Hattori, K., & Iverson, P. (2010). Examination of the relationship between L2 perception and production: An investigation of English /r/-/l/ perception and production by adult Japanese speakers. In M. Nagano (Ed.), *Interspeech workshop on second language studies: Acquisition, learning, education and technology*. Tokyo: Waseda University.

Hayes-Harb, R. (2007). Lexical and statistical evidence in the acquisition of second language phonemes. *Second Language Research*, 23(1), 65-94.

Hazan, V., & Sennema, A. (2007). *The impact of visual training on the perception and production of a non-native phonetic contrast*. Paper presented at Auditory-Visual Speech Processing 2007, Hilvarenbeek, The Netherlands, 31 August - 3 September 2007.

Hazan, V., Sennema, A., Iba, M., & Faulkner, A. (2005). Effect of audiovisual perceptual training on the perception and production of consonants by Japanese learners of English. *Speech Communication*, 47(3), 360-378.

Henning, R., & Jongman, A. (2009). *Phonetics transcription, production, acoustics, and perception*. Chichester: Wiley-Blackwell.

Higashi, T. (1986). Taijin ryuugakusei no toraeta nihongo no hatsuon. [Japanese speech learning strategy by Thai learners]. *Tenri daigaku bekka nihongo katei kiyou*, 1, 25-48.

Hiraiwa, Y. (2004). Taijin nihongo gakushuusha ni taisuru hatsuon shidou no tame no kiso kenkyuu [A study for instructing Thai learners to learn Japanese pronunciation]. *Nihongo kyouiku houhou kenkyuu kaishi*, 11(1), 10-11.

Hirano-Cook, E. (2011). *Japanese pitch accent acquisition by learners of Japanese: Effects of training on Japanese accent instruction, perception, and production*. (Doctoral dissertation, University of Kansas). Retrieved from https://kuscholarworks.ku.edu/bitstream/handle/1808/8022/HiranoCook_ku_0099D_11708_DATA_1.pdf?sequence=1

Hirata, Y. (2004). Computer assisted pronunciation training for English speakers learning Japanese pitch and durational contrasts. *Computer Assisted Language Learning,* 17(3-4), 357-376.

Hu, X., Ackermann, H., Martin, J. A., Erb, M., Winkler, S., & Reiterer, S. M. (2013). Language aptitude for pronunciation in advanced second language (L2) learners: Behavioural predictors and neural substrates. *Brain and Language*, 127(3), 366–376.

Huensch, A. (2013). *The perception and production of palatal codas by Korean L2 learners of English*. (Doctoral dissertation, University of Illinois at Urbana Champaign). Retrieved from https://www.ideals.illinois.edu/bitstream/handle/2142/46739/Amanda_Huensch.pdf?sequence=1

Huensch, A., & Tremblay, A. (2015). Effects of perceptual phonetic training on the perception and production of second language syllable structure. *Journal of Phonetics,* 52, 105-120.

Huffman, S. (2011). Using mobile technologies for synchronous CMC to develop L2 oral proficiency. In J. Levis & K. Levelle (Eds.), *Proceedings of the 2nd Pronunciation in Second Language Learning and Teaching Conference*, Ames, IA (pp. 122-129). Ames, IA: Iowa State University.

Ikeda, N. (2003). *Call dou'nyuu to kaihatsu to jissen* [An introduction, development and practice of CALL - An application of computer use for Japanese language learning]. Tokyo: Kuroshio.

Ikeda, R., & Tateoka, Y. (2007). *Pia raa'ningu nyuumon: souzoutekina manabi no dezain no tameni* [Introduction to Peer learning: Design for creative learning]. Tokyo: Hitsuji Shobou.

Inceoglu, S. (2015). Audiovisual and auditory only perceptual training: Effects on the pronunciation of French nasal vowels. In J. Levis Mohammed, M. Qian & Z. Zhou (Eds.), *Proceedings of the 6th Pronunciation in Second Language Learning and Teaching Conference*, Santa Barbara, CA (pp. 104-114). Ames, IA: Iowa State University.

Ingles, S. A. (2011). *The effects of self-monitoring strategy use on the pronunciation of learners of English*. (Doctoral dissertation, University of Illinois at Urbana Champaign). Retrieved from http://hdl.handle.net/2142/26259

Intercultural Institute of Japan. (2011). *Yasashii nihongo no hatsuon toree'ningu* [Training easy Japanese pronunciation]. Tokyo: Natsume Press.

Isaacs, T., & Thomson, R. I. (2013). Rater experience, rating scale length, and judgments of L2 pronunciation: Revisiting research conventions. *Language Assessment Quarterly,* 10(2), 135-159.

Iverson, P., & Evans, G. B. (2009). Learning English vowels with difference first-language vowel systems II: Auditory training for native Spanish and

German speakers. *Journal of the Acoustical Society of America*, 126(2), 866-877.

Iverson, P., Hazan, V., & Bannister, K. (2005). Phonetic training with acoustic cue manipulations: A comparison of methods for teaching English /r/-/l/ to Japanese adults. *Journal of the Acoustical Society of America*, 118(5), 3267-3278.

Iverson, P., Pinet, M., & Evans, G. B. (2012). Auditory training for experienced and inexperienced second-language learners: Native French speakers learning English vowels. *Applied Psycholinguistics*. 33(1), 145-160.

Jamieson, D. G., & Moroson, D. E. (1986). Training non-native speech contrasts in adults: Acquisition of the English /T/-/D/ contrast by francophones. *Perception & Psychophysics*, 40(4), 205-215.

Jamieson, D. G., & Moroson, D. E. (1989). Training new non-native speech contrasts: A comparison of the prototype and perceptual fading techniques. *Canadian Journal of Psychology,* 43(1), 88-96.

Japan foundation. (1978). *Nihongo hatsuon* [Japanese pronunciation]. Tokyo: Japan foundation.

Japan foundation. (2009). *Kokusai kouryuu kikin nihongo kyoujuhou shiriizu "Onsei wo oshieru"* [A series of Japanese teaching method by Japan foundation "Teaching Japanese speech"], 2, Hitsuji.

Johnson, K. (2012). *Acoustic and auditory phonetics (3rd ed.)*. Malden, MA: Blackwell.

Kartushina, N., & Frauenfelder, U. H. (2014). On the effects of L2 perception and of individual differences on L1 production on L2 pronunciation. *Frontiers in Psychology*, 5, 1-17.

Kartushina, N., Hervais-Adelman, A., Frauenfelder, H. U., & Golestani, N. (2015). The effect of phonetic production training with visual feedback on the perception and production of foreign speech sounds. *The Journal of the Acoustical Society of America,* 138(2), 817-832.

Kawano, T. (2014). *Onsei kyouiku no jissen* [A practice of speech learning]. Tokyo: Kuroshio.

Kawano, T., & Ogawara, H. (2009). *Nihongokyouiku no kako genii mirai* [Past, present, future of Japanese language acquisition]. Tokyo: Bonjinsha

Kawano, T., Tsukiji, N., Kushida, M., & Matsuzaki, H. (2004). *Ichinichi juppun no hatsuon renshuu* [10 minutes a day of pronunciation practice]. Tokyo: Kuroshio.

Kim, Y. H., & Hazan, V. (2010). Individual variability in the perceptual learning of L2 speech sounds and its cognitive correlates. In K. Dziubalska-Kołaczyk, M. Wrembel & M. Kul (Eds.), *Proceedings of the 6th International Symposium on the Acquisition of Second Language Speech, New Sounds 2010*, Poznań, Poland, 1-3 May (pp. 251-256).

Kitahara, Y. (2010). *Meikyou kokugo jiten* [Meikyo Japanese dictionary]. Tokyo: Taishukanshoten.

Kitikanan, P., & Al-Tamimi, J. (2012). The earliest stage of voiceless fricative acquisition among Thai learners of Mandarin Chinese. *ARECLS*, 9, 91-114. Retrieved from http://research.ncl.ac.uk/ARECLS/volume_9/kitikanan_and_altamimi_vol9.pdf

Konishi, R. (2005). Taigo bogo washa no onsei shuutoku ni kansuru kousatsu -shi to chi no chikaku to seisei wo chuushin toshite- [The acquisition of accuracy in the production and perception of Japanese sounds for Thai learners -On the Japanese sounds [ɕi], [tɕi], [(d)ʑi]-]. (Master's dissertation, Osaka University, Japan).

Kuhl, P. K., & Iverson, P. (1995). Linguistic experience and the "perceptual magnet effect". In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 121-154). Timonium, MD: York Press.

Labrune, L. (2012). *The phonology of Japanese*. Oxford: Oxford university press.

Lacabex, E., Lecumberri, M., & Cooke, M. (2009). Training and generalization effects of English vowel reduction for Spanish listeners. In M. Watkins, A. Rauber & B. Baptista (Eds.), *Recent research in second language phonetics/phonology: Perception and production* (pp. 32−42). New Castle upon Tyne: Cambridge Scholars Publishing.

Ladefoged, P. (2003). *Phonetic data analysis: An introduction to fieldwork and instrumental techniques*. Oxford: Blackwell.

Ladefoged, P., & Maddieson, I. (1996). *The sounds of the world's languages*. Oxford, Cambridge, MA: Blackwell.

Lambacher, S., Martens, W., Nelson, B., & Berman, J. (2001). Identification of English voiceless fricatives by Japanese listeners: The influence of vowel context on sensitivity and response bias. *The Journal of the Acoustical Society of Americ*a, 22(5), 334-343.

Lambacher, S., Martens, W. L., Kakehi, K., Marasinghe, C. A., & Molholt, G. (2005). The effects of identification and production of American English vowels by native speakers of Japanese. *Applied Psycholinguistics*, 26(2), 227-247.

Leather, J. (1999). Second language research: An introduction. *Language Learning*. 49(s1), 1-56.

Lengeris, A. (2008). The effectiveness of auditory phonetic training on Greek native speakers' perception and production of Southern British English vowels. *Proceedings of the 2nd ISCA Workshop on Experimental Linguistics,* Athens, Greece (pp. 133-136).

Lengeris, A. (2009). *Individual differences in second-language vowel learning*. (Doctoral dissertation, University College London, London, England). Retrieved from http://discovery.ucl.ac.uk/19029/1/19029.pdf

Lengeris, A., & Hazan, V. (2010). The effect of native vowel processing ability and frequency discrimination acuity on the phonetic training of English vowels for native speakers of Greek. *The Journal of the Acoustical Society of Americ*a, 128(6), 3757-3768.

Lenneberg, E. H. (1967). *Biological foundations of language*. New York, NY: Wiley.

Li, F. (2008). *The phonetic development of voiceless sibilant fricatives in English, Japanese and Mandarin Chinese*. (Doctoral dissertation, The Ohio State University). Retrieved from https://etd.ohiolink.edu/!etd.send_file?accession=osu1228250787&disposition=inline

Liberman, M. (2008, July 6). HVPT [Web log post]. Retrieved from http://languagelog.ldc.upenn.edu/nll/?p=328

Lively, S. E., Logan, J. S., & Pisoni, D. B. (1993). Training Japanese listeners to identify English /r/ and /l/. II: The role of phonetic environment and talker variability in learning new perceptual categories. *The Journal of the Acoustical Society of Americ*a, 94, (3), 1242-1255.

Lively, S. E., Pisoni, D. B., Yamada, R. A., & Tohkura, Y. (1994). Training Japanese listeners to identify English /r/ and /l/. III. long-term retention of new phonetic categories. *The Journal of the Acoustical Society of Americ*a. 96, 2076-2087.

Llisterri, J. (1995). Relationships between speech production and speech perception in a second language. In *ICPhS 1995, Proceedings of the 13th International Congress of Phonetics Sciences vol.4,* Stockholm, Sweden, 13-19 August (pp. 92-99).

Logan, J. S., Lively, S. E., & Pisoni, D. B. (1991). Training Japanese listeners to identify English /r/ and /l/: A first report. *The Journal of the Acoustical Society of Americ*a, 89(2), 874-886.

Logan, J. S., & Pruitt, J. S. (1995). Methodological issues in training listeners to perceive non-native phonemes. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 351-377). Baltimore, MD: York Pr.

Lopez-Soto, T., & Kewley-Port, D. (2009). Relation of perception training to production of codas in English as a second language. *The Journal of the Acoustical Society of Americ*a, 125, 1-15. doi:10.1121/1.3262006

Major, R., & Kim, E. (1996). The similarity deferential rate hypothesis. *Language Learning*, 46(s1), 465-496.

Matsuzaki, H. (2001). Nihongo no onsei kyouiku [Japanese pronunciation teaching]. *Nihongo kyouikugaku shiriizu dai san ken konpyuuta*, 5, 207-258.

Methapisit, T. (2014). Taijin nihongo gakushuusha no tame no onsei kyouiku no genjou to kadai -kyoushi no kyouikuhou to gakushuusha no torikumikata wo chuushin ni [Current status and issues concerning Japanese pronunciation teaching for Thai learners of Japanese: Focus on teachers' practices and leaners' awareness], *Waseda Nihonga kyouikugaku*, 14-16, 87-104.

Mitani, S., Toshihiro, K., & Yu, S. (2006). Voiceless affricates/fricative distinction by frication duration and amplitude slope. *The Journal of the Acoustical Society of America*. 120(3), 1600-1607.

Miyamoto, T., & Osaki, N. (2011). *Gofun de dekiru nihongo oto no kikiwake toree'ningu* [Training to distinguish Japanese in 5 minutes]. Suriiee Network.

Moyer, A. (1999). Ultimate attainment in L2 phonology: The critical factors of age, motivation and instruction. *Studies in Second Language Acquisition*, 21(1), 81-108.

Moyer, A. (2004). *Age, accent, and experience in second language acquisition: An integrated approach to critical period inquiry*. Clevedon: Multilingual Matter.

Nacsakul, K. (1998). *Robob siang phasa thai* [Thai Sound System]. Bangkok: Chulalongkorn University Press.

Nagamine, T. (2011). Effects of hyper pronunciation training method on Japanese university students' pronunciation. *Asian EFL Journal Teaching Articles*, 53, 35-50.

Nakagawa, C., Kihara, I., Akagi, H., & Shinohara, A. (2015). *Tsutawaru hatsuon ga mi ni tsuku! nihongo hanashikata toreiningu* [Learn how to commucate well! Japanese speaking training]. Tokyo: Asuku Press.

Nakagawa, C., & Nakamura, N. (2010). *Shokyu bunkei de dekiru nihongo hatsuon akutibiti* [Japanese pronunciation activities]. Tokyo: Asuku.

Nakagawa, C., Nakamura, N., & Ho, S. (2013). *Sarani susundasupiichi purezen no tame no nihongo hatsuon renshucho* [A handbook of Japanese pronunciation for speech and presentation]. Tokyo: Hitsuji.

Nishi, K., & Kewley-Port, D. (2007). Training Japanese listeners to perceive American English vowels: Influence of training sets. *Journal of Speech, Langauge, and Hearing Research,* 50(6), 1496-1509.

Nishi, K., & Kewley-Port, D. (2008). Nonnative speech perception training using vowel subset: Effects of vowels in sets and order of training. *Journal of Speech, Language, and Hearing Research*. 51(6), 1480-1493.

Nobre-Oliveira, D. (2007). *The effect of perceptual training on the learning of English vowels by Brazilian Portuguese* speakers (Unpublished doctoral disseration). Federal University of Santa Catarina.

Ogawara, Y. (1998a). The effects of pronunciation learning strategies in Japanese learning. *Journal of Linguistic Science Tohoku University*, 2, 1-12.

Ogawara, Y. (1998b). *Gaikokujin nihongo gakushuusha no hatsuon gakushuu ni okeru jiko monitaa no kenkyuu* [A self-monitoring research on the learning of Japanese speech by Japanese learners] (Doctoral dissertation, Tohoku University, Japan).

Okubo, M., Kamiyama, Y., Konishi, R., & Fukui, K. (2012). *shadoingu de nihongo hatsuon ressun* [Japanese pronunciation lesson by shadowing practice]. Tokyo: Suriee Network.

Onishi, H. (1976). Taijin no hatsuon ni kansuru jakkan no kousatsu [A study of Thai learners' pronunciation]. *Kokusai gakuyuukai nihongo gakkou kiyou*, 1, 65-80.

Ortega-Llebaria, M., Faulkner, A., & Hazan, V. (2001). Auditory-visual L2 speech perception: Effect of visual cues and acoustic-phonetic context for Spanish learners of English. In D. W. Massaro, J. Light & Kristin Geraci (Eds.), *International Conference on Auditory-Visual Speech Processing, AVPS 2001*, Aalborg, Denmark, 7-9 September (pp. 149-152).

Peperkamp, S., & Bouchon, C. (2011). The relation between perception and production in L2 phonological processing. *Proceedings of the 12th Annual Conference of the International Speech Communication Association 2011, Interspeech 2011,* Florence, Italy, 28-31 August (pp. 161–164).

Pereira, Y. (2013). *Perception and production of English vowels by Chilean learners of English: Effect of auditory and visual modalities on phonetic training*. (Doctoral dissertation, University College of London, London, England). Retrieved from http://discovery.ucl.ac.uk/1417190/2/Thesis_27-12-13.pdf

Pereira, Y., & Hazan, V. (2013). *Image of different training modes on the perception and production of English vowels by L2 learners*. Paper presented at New Sounds 2013, Montreal, Canada, 17-19 May 2013.

Perrachione, T. K., Lee, J., Ha, L. Y. Y., & Wong, P. C. M. (2011). Learning a novel phonological contrast depends on interactions between individual differences and training paradigm design. *The Journal of the Acoustical Society of America*. 130(1), 461-472.

Pierce, R. L. (2014). *Learning novel vowel contrasts: Experimental methods in classroom applications*. (Doctoral dissertation, University of Illinois). Retrieved from https://www.ideals.illinois.edu/bitstream/handle/2142/49592/Lisa_Pierce.pdf?sequence=1

Piske, T., Mackay, I. R. A., & Flege, J. E. (2001). Factors affecting degree of foreign accent in an L2: A review. *Journal of Phonetics,* 29(2), 191-215.

Pisoni, D. B., Aslin, R. N., Perey, A. J., & Hennessy, B. L. (1982). Some effects of laboratory training on identification and discrimination of voicing contrasts in stop consonants. *Journal of Experimental Psychology*, 8, 297-314.

Pisoni, D. B., & Lively, S. E. (1995). Variability and invariance in speech perception: A new look at some old problems in perceptual learning. In W. Strange (Eds.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 433-459). Baltimore, MD: York Pr.

Pisoni, D. B., Lively, S. E., & Logan, J. S. (1994). Perceptual learning of nonnative speech contrasts: Implications for theories of speech perception. In J. C. Goodman & H. C. Nusbaum (Eds.), *The development of speech perception: The transition from speech sounds to spoken words* (pp. 121-166). Cambridge, MA: MIT.

Pruitt, J. S., Jenkins, J. J., & Strange, W. (2006). Training the perception of Hindi dental and retroflex stops by native speakers of American English and Japanese. *The Journal of the Acoustical Society of Americ*a, 119(3), 1684-1696.

Pruitt, J. S., Strange, W., Polka, L., & Aguilar, M. C. (1990). Effects of category knowledge and syllable truncation during auditory training on Americans' discrimination of Hindi retroflex-dental contrasts. *The Journal of the Acoustical Society of Americ*a, 87, S72.

Rato, A. (2014). Effects of perceptual training on the identification of English vowels by native speakers of European Portuguese. *Proceedings of the International Symposium on the Acquisition of Second Language Speech, Concordia Working Papers in Applied Linguistics*, 5, 529-546. Retrieved from http://doe.concordia.ca/copal/documents/34_Rato_Vol5.pdf

Rochet, B. L. (1995). Perception and production of second-language speech sounds by adults. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 379-410). Baltimore, MD: York Pr.

Roengpitya, R. (2011). An acoustic study of English and Thai fricatives produced by Thai speakers. In W.-S. Lee & E. Zee (Eds.), *Proceedings of* *the 17th International Congress of Phonetic Sciences, ICPhS XVII*, Hong Kong, 17-21 August (pp. 1698-1701).

Roengpitya, R. (2013). The perception of English and Thai fricatives and affricates by Thai learners. *The Journal of the Acoustical Society of* America, 134(5), 4248.

Ruellot, V. (2011). Computer-assisted pronunciation learning of French /u/ and /y/ at the intermediate level. In J. Levis & K. Levelle (Eds.), *Proceedings of the 2nd Pronunciation in Second Language Learning and Teaching Conference*, Ames, IA (pp. 199-213). Ames, IA: Iowa State University.

Sadakata, M., & McQueen, M. J. (2013). High stimulus variability in nonnative speech learning supports formation of abstract categories: Evidence from Japanese geminates. *The Journal of the Acoustical Society of America,* 134(2), 1324-1335.

Saito, H., Fukazawa, M., Sakai, R., Nakamura, M., & Yoshimoto, K. (2010). *Shadouingu nihongo wo hanasou chuu-jou kyuu* [Shadowing: Let's Speak Japanese! Intermediate to Advanced Level]. Tokyo: Kuroshio.

Saito, H., Yoshimoto, K., Fukazawa, M., Onoda, C., & Sakai, R. (2006). *Shadouingu nihongo wo hanasou sho-chu kyu* [Shadowing: Let's Speak Japanese! Beginner to Intermediate]. Tokyo: Kuroshio.

Saito, K. (2012). Effects of instruction on L2 pronunciation development: A synthesis of 15 quasi-experimental intervention studies. *TESOL Quarterly*, 46(4), 842-854.

Sardegna, V. G. (2011). Pronunciation learning strategies that improve ESL learners' linking. In J. Levis & K. Levelle (Eds.), *Proceedings of the 2nd Pronunciation in Second Language Learning and Teaching Conference*, Ames, IA (pp. 105-121). Ames, IA: Iowa State University.

Sargedna, V. G., & McGregor, A. (2013). Scaffolding students' self-regulated efforts for effective pronunciation practice. In J. Levis & K. LeVelle (Eds.), Proceedings of the 4th Pronunciation in Second Language Learning and

Teaching Conference, Ames, IA (pp. 182-193). Ames, IA: Iowa State University.

Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics*, 11, 129-158.

Schmidt, R. (1995). Consciousness and foreign language learning: A tutorial on the role of attention and awareness in learning. In R. Schmidt (Ed.), *Attention & awareness in foreign language learning* (pp.1-63). Honolulu, HI: Second Language Teaching & Curriculum Center.

Schmidt, R. (2001). Attention. In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 3-32). Cambridge: Cambridge University Press.

Schmidt, R. (2010). Attention, awareness, and individual differences in language learning. In W. M. Chan, S. Chi, K. N. Cin, J. Islanto, M. Nagami, J. W. Swe, T. Suthiwan & I. Walker (Eds.), *Proceedings of CLaSIC 2010*, Singapore, 2-4 December (pp. 721-737). Singapore: National University of Singapore, Centre for Language Studies.

Seferoglu, G. (2005). Improving students' pronunciation through accent reduction software. *British Journal of Education Technology,* 36(2), 303-316.

Sharwood, S. M. (1981). Consciousness-raising and the second language learner. *Applied Linguistics*, 2, 159-168.

Sharwood, S. M. (1993). Input enhancement in instructed SLA theoretical bases. *Studies in Second Language Acquisition*, 15(2), 165-179.

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

Shin, D. J. (2014). *Training Korean speakers on English vowels and prosody: Individual differences in perception, production and vowel epenthesis*. (Doctoral dissertation, University College London, London, England). Retrieved from http://discovery.ucl.ac.uk/1425728/1/Dong-Jin_Shin_PhD_THESIS_edited_version%5B1%5D_2014.pdf

Smith, J., & Beckmann, B. (2005). Improving pronunciation through noticing-reformulation tasks. *Proceedings of the Phonetics Teaching and Learning Conference 2005*. London.

Sonu, M., Kato, H., Tajima, K., Akahane, Y. R., & Sagisaka, Y. (2013). Non-native perception and learning of the phonemic length contrast in spoken Japanese: Training Korean listeners using words with geminate and singleton phonemes. *Journal of East Asian Linguistics,* 22(4), 373-398.

Southwood, M. H., & Flege, J. E. (1999). Scaling foreign accent: Direct magnitude estimation versus interval scaling. *Clinical Linguistics & Phonetics*. 13, 335-349.

Spada, N. (1997). Form-focused instruction and second language acquisition: A review of classroom and laboratory research. *Language Teaching,* 30, 73-87.

Steinlen, A. (2009). Phonetic input in second language learning: Contrastive analysis of native and non-native sounds. In T. Piske & M. Young-Schoulten (Eds.), *Input Matters in SLA* (pp. 219-237). Bristol: Multilingual Matters.

Strange, W. (1992). Learning non-native phoneme contrasts: Interactions among subject, stimulus, and task variables. In Y. Tohkura, E. Vatikiotis-Bateson & Y. Sagisaka (Eds.), *Speech perception, production, and linguistic structure* (pp. 197-219), Tokyo: OHM.

Strange, W. (1995). Phonetics of second-language acquisition: Past, present, future. *ICPhS 95*, 4, 76-83.

Strange, W., & Dittmann, S. (1984). Effects of discrimination training on the perception of /r-l/ by Japanese adults learning English. *Perception and Psychophysics*, *36*(2), 131-145.

Strange, W., & Hawles, T. (1971). Confidence ratings in speech perception research: Evaluation of an efficient technique for discrimination testing. *Perception & Psychophysics,* 9, 182-186.

Styler, W. (2013). *Using Praat for linguistic research*. Retrieved from https://phonetique.uqam.ca/upload/files/LIN2623/Styler_2013_2.pdf

Sukegawa, Y. (1993). Bogo betsu ni mita hatsuon no keikou - ankeito chousa no kekka kara- [A pronunciation trend as analysed by L1 language]. *Nihongo onsei to nihongo kyouiku Monbushou juuten ryouiki kenkyuusho kenkyuu seika houkokusho*, 187-222.

Tajima, K., Kato, H., Rothwell, A., Akahane-Yamada, R., & Munhall, K. (2008). Training English listeners to perceive phonemic length contrasts in Japanese. *The Journal of the Acoustical Society of America*,123(1), 397-413.

Tanaka, S., & Kubozono, H. (1999). *Nihongo no hatsuon kyoushitsu: riron to renshuu* [Introduction to Japanese pronunciation: Theory and practice]. Tokyo: Kuroshio.

Thomson, R. I. (2008). L2 English vowel learning by Mandarin speakers: Does perception precede production?. *Canadian Acoustics*, 36(3), 134-135.

Thomson, R. I. (2011). Computer assisted pronunciation training: Targeting second language vowel perception improves pronunciation. *CALICO Journal, 28*(3), 744-765.

Thomson, R. I. (2012). Improving L2 listeners' perception of English vowels: A computer-mediated approach. *Language learning, 62*(4), 1231-1258.

Thomson, R. I., & Derwing, T. M. (2014). The effectiveness of L2 pronunciation instruction: A narrative review. *Applied* Linguistics. Doi: 10.1093/applin/amu076

Thornbury, S. (1997). Reformulating and reconstruction: Tasks that promote "noticing". *ELT Journal*, 51(4), 326-334.

Tingsabadh, K., & Abramson, A. S. (1993). Thai. *Haskins Laboratories Status Report on Speech Research*, 131-134.

Toda, K. (2004). *komyunikeeshon no tame no nihongo hatsuon ressun* [Japanese pronucnation lesson for communication]. Tokyo: Suriiee Network.

Trakantalerngsak, T. (2012). An acquisition of Japanese pronunciation by Thai learners: An analysis based on the results of a pronunciation test. *Proceedings of The Ninth Symposium for Japanese Language Education and Japanese* Studies, Hong Kong, 24-25 November 2012. Retrieved from http://www.japanese-edu.org.hk/sympo/upload/manuscript/20121011112126.pdf

Trakantalerngsak, T. (2013). *The acquisition of accuracy in the production and perception of Japanese sounds for Thai learners*: *On the Japanese sounds [ɕi], [tɕi], [(d)ʑi].* (Unpublished master's thesis). Osaka University.

Trakantalerngsak, T. (2015). Taigo bogo washa ni yoru nihongo no hasatsuon no seisei: onkyou bunseki ni your kousatsu [A production of Japanese affricates by Thai learners: Acoustic analysis], *Nihongo ronbunshuu,* 11, 22-36.

Vance, J. T. (2008). *The sounds of Japanese*. Cambridge, New York: Cambridge University Press.

Vogt, W. P. (1993). *Dictionary of statistics and methodology: A non-technical guide for the social sciences*. Newbury Park, CA: Sage Publications.

Wade, T., Jongman, A., & Sereno, J. (2007). Effects of acoustic variability in the perceptual learning of nonnative-accented speech sounds. *Phonetica*, 64(2-3), 122-144.

Wang, X. (1997). *The acquisition of English vowels by Mandarin ESL learners: A study of production and perception* (Unpublished master's thesis). Simon Fraser University.

Wang, X. (2002). *Training Mandarin and Cantonese speakers to identify English vowel contrasts: Long-term retention and effects on production*. (Doctoral dissertation, Simon Fraser University, British Columbia, Canada). Retrieved from http://zimmer.csufresno.edu/~xinw/contents.pdf

Wang, X. (2008). Training for learning Mandarin tones. In F. Zhang & B. Barber (Eds.), *Handbook of research on computer-enhanced language acquisition and learning*, 23(2), 259-273.

Wang, X. (2013). Perception of Mandarin tones: The effect of L1 background and training. *The Modern Language Journal,* 97(1), 144-160.

Wang, X., & Munro, M. J. (2004). Computer-based training for learning English vowel contrasts. *System,* 32, 539-552.

Wang, Y., Jongman, A., & Sereno, A. (2003). Acoustic and perceptual evaluation of Mandarin tone productions before and after perceptual training. *The Journal of the Acoustical Society of America,* 113(2), 1033-1043.

Wang, Y., Spence, M., Jongman, A., & Serene, A. J. (1999). Training American listeners to perceive Mandarin tones. *The Journal of the Acoustical Society of America,* 106(6), 3649-3658.

Wayland, R. P., & Guion, S. G. (2004). Training English and Chinese listeners to perceive Thai tones: A preliminary report. *Language Learning,* 54(4), 681-712.

Wong, W. S. (2013). *Training the perception and production of English vowels /ɪ/-/iː/, /e/-/æ/, /ʊ/-/uː/ by Cantonese ESL learners in Hong Kong*. (Doctoral dissertation, The Chinese University of Hong Kong, Hong Kong). Retrieved from http://etheses.lib.cuhk.edu.hk/pdf/920150188.pdf

Yamakawa, J., Chisaki, Y., & Usagawa, T. (2005). Tai jin nihongo gakushuusha ni okeru hasatsuon no tsu no shuutoku [The acquisition of Japanese affricate contrast by Thai learners]. *Nihongo onkyo gakkai choukaku kenkyuukai shiryou*, 35(11), 663-668.

Yamakawa, K., & Amano, S. (2011). Acoustic features to discriminate between affricates and a fricative in Japanese. In W.-S. Lee & E. Zee (Eds.), *Proceedings of the 17th International Congress of Phonetic Sciences, ICPhS XVII*, Hong Kong, 17-21 August (pp. 2189-2192).

Yamakawa, K., & Amano, S. (2013). Speaking-rate effects on production boundaries between affricate [ts] and fricative [s] in Japanese native speakers. *Aichishukutoku University ronshuu-ningen jouhou gakubu hen*, 3, 21-25.

Yamakawa, K., Chisaki, Y., &, Usagawa, T. (2006). Subjective evaluation of Japanese voiceless affricate spoken by Korean. *Acoustical Science and Technology*, 27(4), 236-238.

Yamada, R. A. (1992). Perception of American English /r/ and /l/ by native speakers of Japanese. *Perception & Psychophysics*, 52(4), 376-392.

Yamada, R. A. (1993). Effects of extended training on /r/ and /l/ identification by native speakers of Japanese. *The Journal of the Acoustical Society of America,* 93, 2391.

Yamada, R. A. (1994). The intelligibility of the Japanese speakers' production of American English /r/, /l/, and /w/, as evaluated by native speakers of American English. *The 3rd International Conference on Spoken Language Processing, ICSLP 1994*, Yokohama, Japan, September 18-22, 1994

Yamada, R. A., Tohkura, Y., Bradlow, A. R., & Pisoni, D. B. (1996). Does training in speech perception modify speech production?. *Proceedings of the Fourth International Conference on Spoken Language Processing,* Philadelphia, (2), 606-609. Pennsylvania, Philadelphia: Institute of Electrical and Electronics Engineers.

Yeon, S. (2004). *Training English word-final alveolopalatals to native speakers of Korean*. (Doctoral dissertation, University of Florida). Retrieved from http://ufdc.ufl.edu/UFE0006465/00001

Yoshiki, H. (2010). *Nihongo hatsuon kantan* [Easy Japanese pronunciation]. Tokyo: Kenkyusha.

Yule, G., Hoffman, P., & Damico, J. (1987). Paying attention to pronunciation: The role of self-monitoring in perception, *TESOL Quarterly*, 21(4), 765-768.

Zampini, M. L. (1998). The relationship between the production and perception of L2 Spanish stops. *Texas papers in Foreign Language Education*, 3(3), 85-100.

Zampini, M. L. (2008). L2 speech production research: Finding, issues, and advances. In J. G. Hansen Edwards & M. L. Zampini (Eds.), *Phonology* *and Second Language Acquisition* (pp. 219-249). Amsterdam: John Benjamins.

Zhang, W. (2010). *Non-native speakers speak in phonemes: A phono-acoustic analysis of fricatives and affricates by native and Chinese speakers of English* (Doctoral dissertation). Retrieved from Purdue e-Pubs. (AAI3449421)

APPENDIX

APPENDIX A: Participants info and language background

No	Group	Sex	Age	Birthplace	Years of exposure	Japanese proficiency level
1	Control	F	21	Uthaithani	5	-
2	Control	F	21	Rachaburi	5	-
3	Control	F	21	Bangkok	5	-
4	Control	F	22	Lampang	4	N3
5	Control	F	21	Bangkok	5	-
6	Control	F	22	Nakhornratchasrima	6	3
7	Control	F	22	Nongkhai	4	N3
8	Control	M	21	Bangkok	3	N3
9	Control	F	22	Bangkok	7	N3
10	Control	F	22	Bangkok	7	N3
11	Control	F	26	Bangkok	2	-
12	A	F	19	Bangkok	4	N4
13	A	F	19	Bangkok	8	N3
14	A	F	20	Bangkok	6	N3
15	A	F	19	Bangkok	4	3
16	A	F	19	Bangkok	5	N2
17	A	M	22	Bangkok	6	N2
18	A	F	21	Trang	3	-
19	A	F	21	Bangkok	5	-
20	A	F	20	Ratchaburi	4	-

No	Group	Sex	Age	Birthplace	Years of exposure	Japanese proficiency level
21	A	F	20	Prachuabkirikhan	4	-
22	A	F	27	Bangkok	12	N2
23	B	F	19	Nonthaburi	3	N3
24	B	F	19	Nakhornratchasrima	3	N3
25	B	F	19	Bangkok	3	3
26	B	F	19	Phitsanulok	3	N4
27	B	F	19	Rayong	3	4
28	B	F	19	Chonburi	3	N3
29	B	F	29	Bangkok	14	N1
30	B	F	30	Bangkok	5	N3
31	B	M	32	Bangkok	17	1

APPENDIX B: Consent form

Consent Form

This experiment is carried out by Tanporn Trakantalerngsak, a graduate student in the Department of Language and Culture at Osaka University, under the supervision of Professor Yasuo Iwai.

Conductor	University	Contact
Tanporn Trakantalerngsak	Osaka University	t.tanporn@gmail.com

You are being asked to participate in a research experiment. This form provides you with information about the experiment. Please read the information below and ask any questions you might have before deciding whether or not to take part in the experiment. Your participation is completely *voluntary*. You may refuse to participate or discontinue participation at any time without penalty or loss of benefits. If you do change your mind about participating, please contact "t.tanporn@gmail.com" as soon as possible.

The purpose of this study is to investigate whether perceptual training is beneficial in improving L2 learners' ability in perception and production of the Japanese sounds.

If you agree to take part in this experiment, I will ask you to do the following things:

1. Fill in the questionnaire form. Provide your name and email address.
2. You will be assigned to either a control group or an experimental group and be granted access to an online perceptual training. You will be asked to conduct approximately 10 to 15 minutes of training in total of 9 times over a month.

3. You may be contacted at a later date for more information required.

The total amount of time

Control Group: The pre-test/post-test/generalisation-test/delayed post-test will be conducted.

Experimental Group: The pre-test/training/post-test/generalisation-test/delayed post-test will be conducted.

Phase Duration

- Pre-test: 15-20 minutes
- Training: 10-15 minutes per session
- Post-test: 20-30 minutes
- Delayed post-test: 20 minutes

Benefits

The possibility that your ability to perceive and produce the Japanese phonemes will improve.

Compensation

You will receive 300 Baht for participating the experiment.

Confidentiality and privacy protections

Data from this experiment will be collected for use in the conductor's doctoral dissertation. Data also may be used in journal articles and conference presentations. However, only the conductor will have an access to research results associated with your identity. In the event of publication of this research, no personally identifying information will be disclosed.

Contact and questions

If you have questions about any part of this experiment, please ask the conductor, Tanporn Trakantalerngsak, whose information is listed below.

Email: t.tanporn@gmail.com

Line ID: galipgalipgalip or add “+(81)-80-4243-5145”

Statement of consent

I certify that I have read this form and have sufficient information to make a decision about participating in this experiment. I consent to participate in the experiment.

Name: ____

Signature: ____________________________________

Date: ______________________

Summary of experiment content

1. Pre-test > Training > Post-test/Generalisation Test > Delayed post-test
2. Training Phase: 9 sessions (approx 15-20 mins)
3. Username and password will be sent to your email or line.
4. If you have any questions at any time, please email “t.tanporn@gmail.com” or call at “080-4243-5145”, Line ID: galipgalipgalip

APPENDIX C: Training instruction (Thai version)

Training handout

สวัสดีค่ะ ขอต้อนรับเข้าสู่การเทรนนิ่งการฟังเสียงภาษาญี่ปุ่น จากนี้ไปคุณจะได้เข้าร่วมเทรนนิ่งเพื่อเพิ่มทักษะทางการฟังเสียงภาษาญี่ปุ่นให้ดีขึ้น เป็นจำนวนครั้งทั้งหมด 9 ครั้ง ซึ่งจะใช้เวลาฟังในการเทรนนิ่งในแต่ละครั้ง 10-15 นาที ขอขอบคุณทุกๆท่านที่ให้ความร่วมมือด้วยนะคะ

เครื่องมือที่ต้องใช้ในการเทรนนิ่ง

1. หูฟัง
2. คอมพิวเตอร์

วิธีการเทรนนิ่ง

1. ไปที่ website "http://128.199.227.231/perceptual/"

    (สามารถเปิดได้ใน safari, Chrome และ Firefox)

2. ใส่ username และ password ที่จะส่งให้ในภายหลัง

ข้อควรระวัง

1. ตรวจสอบว่าท่านมีหูฟังหรือไม่ในการเทรนนิ่ง
2. ท่านจะต้องใช้คอมพิวเตอร์ในการทำเทรนนิ่งมิใช่แท็พเล็ต มือถือ หรือ ไอแพด
3. ท่านต้องอยู่ในที่ที่ไม่มีเสียงดังรบกวน
4. ท่านจะต้องมีสมาธิในการทำเทรนนิ่งในแต่ละครั้ง เทรนนิ่งที่จะใช้เวลาเพียงประมาณ 10 ถึง 15 นาที
5. หากมีข้อผิดพลาดประการใดสามารถติดต่อผ่าน *Line: galipgalipgalip* หรือ *"t.tanporn@gmail.com"* ได้ทุกเมื่อ

ทุกๆท่านจะได้รับการติดต่อในการทำเทรนนิ่งช่วงประมาณปลายพฤษภาคมค่ะ

ขอบคุณทุกท่านที่ให้ความร่วมมือนะคะ และขอให้โชคดีค่ะ

APPENDIX C: Training instruction (English Version).

Training handout

Hello everyone! Welcome to the Japanese sound training. From now on you will be asked to participate in a nine-session training to help you better learn Japanese sounds. Each session will take approximately 10 to 15 minutes. Thank you so much for your participation.

Thing to use in the training

1. Headphones
2. Personal computer

Training procedure

1. Go to website "http://128.199.227.231/perceptual/".

   (The site works best with Safari, Chrome and Firefox)
2. Type given "username" and "password".

Cautions

1. Before starting the training check whether you have headphones ready or not.
2. Use only a personal computer or a laptop. Small devices such as iPad, iPhone are not allowed.
3. Attend the environment in quiet environment.
4. Concentrate to the training. It will take only 10 to 15 minutes.
5. If there are any problems occurred, please feel free to contact Tanporn via *Line: galipgalipgalip* or *"t.tanporn@gmail.com"* anytime.

You will be contacted shortly in May to inform the detailed schedules.

Thank you for your cooperation Hope you enjoy the training.

APPENDIX D: Native Japanese baseline group.

	Sex	Age	Hometown
Baseline JS1	M	20s	Kyoto
Baseline JS2	F	20s	Osaka
Baseline JS3	F	30s	Osaka
Baseline JS4	M	20s	Osaka
Baseline JS5	F	20s	Hyogo

APPENDIX E: Japanese native talkers used in the pre/post-test, training and generalisation test.

	Sex	Age	Hometown
JS1	M	50s	Saitama
JS2	F	20s	Nigata
JS3	M	20s	Kobe
JS4	F	30s	Osaka
JS5 (Gen-test)	M	20s	Osaka
JS6 (Post-test)	M	20s	Osaka

APPENDIX F: Japanese native raters used for judging production accuracy and intelligibility.

	Sex	Age	Hometown
R1	F	20s	Niigata
R2	M	20s	Hyogo
R3	F	20s	Osaka
R4	F	30s	Chiba
R5	M	30s	Kyoto

APPENDIX G: List of stimuli used in pre/post test

Sound	Vowel	Initial		Medial/Final
[(d)z]	[a] [ɯ] [e] [o]	ざまま[zamama] ずまま[zɯmama] ぜまま[zemama] ぞまま[zomama]	[a] [ɯ] [e] [o]	まざま[mazama] まずま[mazɯma] まぜま[mazema] まぞま[mazoma]
[ts]	[ɯ]	つまま[tsɯmama] つまみ[tsɯmami] つまむ[tsɯmamu] つまも[tsɯmamo]	[ɯ]	あつま[atsɯma] いつま[itsɯma] うつま[utsɯma] えつま[etsɯma]
[tɕ]	[a] [i] [ɯ] [o]	ちゃあま[tɕa:ma] ちまま[tɕimama] ちゅうま[tɕɯ:ma] ちょうま[tɕo:ma]	[a] [i] [ɯ] [o]	まちゃあ[matɕa:] まちま[matɕima] まちゅう[matɕɯ:] まちょう[matɕo:]
[(d)ʑ]	[a] [i] [ɯ] [o]	じゃあま[dʑa:ma] じまま[dʑimama] じゅうま[dʑɯ:ma] じょうま[dʑo:ma]	[a] [i] [ɯ] [o]	まじゃあ[maʑa:] まじま[maʑima] まじゅう[maʑɯ:] まじょう[maʑo:]

APPENDIX H: List of stimuli used in the perceptual training

1. [(d)z] vs. [s] contrast

Vowel	Training Item	Vowel	Training Item
[a]	ざいげん vs. さいげん こうざ vs. こうさ きざみ vs. きさみ ざっぴ vs. さっぴ ざまま vs. さまま まざま vs. まさま	[u]	てずり vs. てすり けんずる vs. けんする ずあん vs. すあん はずれ vs. はすれ ずまま vs. すまま まずま vs. ますま
[e]	ぜんてい vs. せんてい かんぜい vs. かんせい めんぜん vs. めんせん ぜにん vs. せにん ぜまま vs. せまま まぜま vs. ませま	[o]	ぎぞう vs. ぎそう ぞうかん vs. そうかん のぞく vs. のそく ぞっかい vs. そっかい ぞまま vs. そまま まぞま vs. まそま

2. [ts] vs. [s] contrast

Vowel	Training Item	Vowel	Training Item
[u]	ついか vs. すいか つける vs. すける ついきゅう vs. すいきゅう つく vs. すく こつ vs. こす たつ vs. たす ふつう vs. ふすう うつ vs. うす つばき vs. すばき ついで vs. すいで つねに vs. すねに うつり vs. うすり	[u]	かつぐ vs. かすぐ きつね vs. きすね てつや vs. てすや こりつ vs. こりす つまま vs. すまま つまみ vs. すまみ つまむ vs. すまむ つまも vs. すまも あつま vs. あすま いつま vs. いすま うつま vs. うすま えつま vs. えすま

3. [tɕ] vs. [(d)ʑ] contrast

Vowel	Training Item	Vowel	Training Item
[a]	こちゃく vs. こじゃく ちゃく vs. じゃく みちゃく vs. みじゃく かんじゃ vs. かんちゃ ちゃあま vs. じゃあま まちゃあ vs. まじゃあ	[u]	ちゅうい vs. じゅうい くちゅう vs. くじゅう じゅんろ vs. ちゅんろ みち vs. みじ ちゅうま vs. じゅうま まちゅう vs. まじゅう
[i]	ちばん vs. じばん こうち vs. こうじ ほじ vs. ほち じえい vs. ちえい ちまま vs. じまま まちま vs. まじま	[o]	ちょげん vs. じょげん とちょう vs. とじょう ぶちょう vs. ぶじょう じょうよ vs. ちょうよ ちょうま vs. じょうま まちょう vs. まじょう

APPENDIX I: List of stimuli used in generalisation test

1. Low level of familiarity

Sound	Vowel	Initial		Medial/Final
[(d)z]	[a] [ɯ] [e] [o]	讒謗（ざんぼう） 瑞雲（ずいうん） 絶佳（ぜっか） 続開（ぞっかい）	[a] [ɯ] [e] [o]	小字（こあざ） 奇瑞（きずい） 党是（とうぜ） 無反り（むぞり）
[ts]	[ɯ]	痛罵（つうば） 柄寝る（つかねる） 夙に（つとに） 積ん読（つんどく）	[ɯ]	凛冽（りんれつ） むつかる 無筆（むひつ） 陸つぶ（むつぶ）
[tɕ]	[a] [i] [ɯ] [o]	着剣（ちゃっけん） 蟄居（ちっきょ） 中興（ちゅうこう） 手水（ちょうず）	[a] [i] [ɯ] [o]	飛鳥（ひちょう） 陶枕（とうちん） 蘭虫（らんちゅう） 違勅（いちょく）
[(d)ʑ]	[a] [i] [ɯ] [o]	若朽（じゃっきゅう） 塵外（じんがい） 純化（じゅんか） 助炭（じょたん）	[a] [i] [ɯ] [o]	おじゃる 狢（むじな） 隷従（れいじゅう） 身状（みじょう）

2. High level of familiarity

Sound	Vowel	Initial		Medial/Final
[(d)z]	[a] [ɯ] [e] [o]	座標（ざひょう） 図解（ずかい） 絶句（ぜっく） 増強（ぞうきょう）	[a] [ɯ] [e] [o]	無惨（むざん） 合図（あいず） 唖然（あぜん） 秘蔵（ひぞう）
[ts]	[ɯ]	墜落（ついらく） 痛感（つうかん） 尽きる（つきる） 突っ張る（つっぱる）	[ɯ]	密約（みつやく） 操る（あやつる） 猛烈（もうれつ） 遺骨（いこつ）

Sound	Vowel	Initial		Medial/Final
[tɕ]	[a] [i] [ɯ] [o]	着工（ちゃっこう） 窒息（ちっそく） 中退（ちゅうたい） 著作（ちょさく）	[a] [i] [ɯ] [o]	愛着（あいちゃく） 一塁（いちるい） 害虫（がいちゅう） 野鳥（やちょう）
[(d)ʑ]	[a] [i] [ɯ] [o]	邪道（じゃどう） 自力（じりき） 受難（じゅなん） 除夜（じょや）	[a] [i] [ɯ] [o]	大蛇（だいじゃ） 無慈悲（むじひ） 打順（だじゅん） 控除（こうじょ）

APPENDIX J: Questionnaire (Thai version)

แบบประเมิน Perceptual Training

ขอขอบคุณทุกท่านที่ให้ความร่วมมือกับการเทรนนิ่ง “Perceptual Training” ทั้งนี้ ดิฉันขอความกรุณาให้ทุกท่านกรอกแบบประเมินการเทรนนิ่งด้วยค่ะ ท่านสามารถเขียนข้อคิดเห็นหรือข้อแนะนำได้ตามอัธยาศัยค่ะ หากมีข้อผิดพลาดประการใดเกิดขึ้นในระหว่างการเทรนนิ่งนั้น ขออภัยมา ณ ที่นี้ด้วย

ทานพร ตระการเถลิงศักดิ์

1) แรงจูงในในการเข้าร่วมเทรนนิ่งคืออะไร

☐ เพื่อเพิ่มความสามารถในการฟังและการพูด

☐ มีคนบังคับให้เข้าร่วม

☐ และอื่นๆ โปรดระบุ ______________________

2) ท่านคิดอย่างไรจากการเข้าร่วมเทรนนิ่งทั้ง 9 ครั้ง

____

3) ท่านคิดว่าการเทรนนิ่งช่วยท่านให้สามารถฟังภาษาญี่ปุ่นได้ดีขึ้นหรือไม่

☐ ได้ ☐ ไม่ได้

4) ท่านอยากให้มีการเทรนนิ่งสำเนียงภาษาญี่ปุ่นของท่านในวิชาเรียนภาษาญี่ปุ่นหรือไม่

☐ อยาก ☐ ไม่อยาก

5) (เฉพาะคนที่ได้ฟังเสียงตัวเอง) ท่านคิดว่าการได้ฟังเสียงพูดของตัวเองเทียบกับเสียงคนญี่ปุ่นทำให้ท่านรู้ปัญหาในการฟังและการพูดของท่านหรือไม่

□ได้ □ไม่ได้

6) กรุณาเขียนข้อดีของการเทรนนิ่งการฟังในครั้งนี้

____

__

7) กรุณาเขียนข้อควรปรับปรุงของการเทรนนิ่งการฟังในครั้งนี้

____

__

8) ท่านคิดว่าจำนวน 9 ครั้ง ในการทำเทรนนิ่งน้อยหรือมากไป

____

__

9) ท่านมีหลักการอย่างไรในการฟังและออกเสียง し･ち･じ･つ･ザ行

____

__

APPENDIX J: Questionnaire (English version)

<u>Perceptual Training Questionnaire</u>

Thank you so much for your time in participating in this training. Lastly, I would like to ask you to fill in this questionnaire. If you have any comments or advice please feel free to write them down. Also, if you have encountered any inconvenience as a result of participating in this study then, I am deeply sorry.

Tanporn Trakantalerngsak

1) What is your motivation for participating in this training program?

☐ To improve listening and pronunciation skill

☐ Someone forced me to attend this training

☐ Others (please specify) ___________________________

2) What do you think of this nine-session training?

____

3) Do you think this training can help to enhance better listening skills?

☐ Yes ☐ No

4) Do you want to have listening and pronunciation training or instruction in your Japanese classroom?

☐ Yes ☐ No

5) (Only those who listened to their own voice “Trained Group B”) Do you think that listening to your own speech can help you to notice your perception and production problems?

☐ Yes ☐ No

6) Write down an advantage of this training.

____

7) Write down any comments or things that should be improved

____

8) Do you think that nine sessions of the training is too long? or too short?

____

9) What learning strategy do you use to perceive and produce し･ち･じ･つ･ザ行?

____