# An investigation of factors behind foreign accent in the L2 acquisition of Japanese lexical pitch accent by adult English speakers

Emi Sakamoto

PhD

The University of Edinburgh

2010

# Declaration

I hereby declare that this thesis has been composed by myself, and the research presented herein is my own work, except where explicitly mentioned. It has not been submitted for any other degree or professional qualification.

Emi Sakamoto

December, 2010

# Abstract

The productions of adult second language (L2) learners are often detected as having a foreign accent by native speakers of the target language. However, there is no clear answer for what kind of problems contribute to L2 learners' foreign accent. This thesis aims to investigate potential factors behind foreign accent. We intend to achieve this goal by examining cross-linguistic empirical evidence of the L2 acquisition of Japanese lexical pitch accent by English learners. L2 prosody has been found to significantly influence native speakers' auditory impression of foreign accent. L2 prosody also allows us to test cross-linguistic differences in the function of the key acoustic correlates of L2 contrasts. In this thesis we examine F0, which signals both lexical pitch accent and phrasal distinctions in Japanese, but which signals only phrasal distinctions, not lexical distinctions, in English.

For adult L2 learners to achieve target-like productions, the literature suggests that three abilities are the key factors: 1) learners' ability to differentiate the acoustic correlate of the target L2 contrasts, 2) ability to articulate the acoustic correlate of the target L2 contrasts and 3) ability to categorize the target L2 contrasts. This thesis evaluates all three of these potential factors. The main contribution of this thesis is to provide a comprehensive view of foreign accent, by investigating possible interactions between the factors and by examining the different abilities of the same learners. Another contribution is to provide empirical evidence for the nature of learners' problems with foreign accent during L2 acquisition, by testing two groups of English learners of Japanese (experienced and inexperienced) in comparison with Japanese native speakers.

The first experiment used intelligibility scores and overall F0 patterns to quantify the degree of foreign accent in the learners' productions of Japanese lexical pitch accent. The second experiment showed that the learners' ability to differentiate F0 contours in a non-speech context was equal to that of the native speakers. The third experiment showed that the learners' ability to articulate the F0 contours in a non-speech context differed from that of the native speakers. The fourth experiment showed that although learners were able to hear the phonetic differences between the target L2 contrasts, due to poor formation of the target L2 categories and poor lexical assignment ability, the inexperienced learners seem to have greater difficulty than experienced learners both in categorizing boundary items into the target L2 categories and in assigning the L2 categories to lexical items.

Overall, the foreign accent of adult L2 learners' productions is explained through a combination of articulation and categorization factors. Importantly, this cross-sectional study has indicated how learners' problems with foreign accent change as they gain L2

experience. Whereas experienced learners seem to have problems mainly in the articulation and phonetic realization of the L2 contrasts, the inexperienced learners seem to have mainly problems in phonetic and lexical-phonological representations of the target L2 categories in addition to articulation and phonetic realization. This study offers both theoretical insights for the field of L2 speech acquisition research and also practical insights for the L2 classroom.

# Acknowledgements

The long journey of my postgraduate studies would not have been accomplished without the tremendous support from so many different people in Edinburgh, Japan and abroad. First and foremost, I would like to thank my co-supervisors: Dr. Mitsuhiko Ota for his great passion, sharp thinking and inspirational comments since my MSc study; and Prof. Alice Turk for her enthusiasm, overflowing scientific knowledge, invaluable advice, and genuine heartfelt encouragements throughout my study. I am also thankful to my former supervisors, Prof. Kensaku Yoshida, who introduced me to the field of SLA/Applied Linguistics, and Prof. Antonella Sorace, who welcomed me to the international field of SLA in Edinburgh.

My special thanks go to Prof. Bob Ladd, who provided me with very insightful feedback regarding F0 studies and with the opportunity to be involved in his course. I would also like to thank Dr. Bert Remijsen for the tutorials on using the MOMEL scripts, Dr. Yi Xu for sharing his valuable F0 script and for his generous support, Dr. Satsuki Nakai and Dr. Takeshi Ishihara for the great help with the acoustic analyses, and Mr. Olaf Dabrunz for programming the scripts.

I am enormously indebted to the technical staff at Linguistics and English Language for their invaluable support; Michael Bennett for his patience and effort in solving numerous script problems, Eddie Dubourg for his endless help with programming and conducting a series of my experiments, Barry Campbell for studio support, and Cedric Macmartin for comprehensive support. Thanks also go to Katie Keltie for her administrative help.

It is beyond words how to express my gratitude to Dr. Catherine Dickie not only for patiently proof-reading my drafts, with a very limited time, but also for her moral support. I am so grateful to Saeko and Reiko for their cooperation with the recordings and with rating a massive quantity of data, in addition to our friendship.

I also wish to thank the council members of the Japan Society of Scotland and the Consulate General of Japan for unforgettable experiences and precious opportunities in Scotland. Thanks go to the staff and colleagues of Japanese Studies for job opportunities and for their help with finding participants for my experiments. I would also like to express my gratitude to the people who participated in my experiments: my former students in Japanese 2 and 4 at Edinburgh University, JET returnees and Japanese people in Edinburgh.

My warm thanks go to my dear friends, Batoul Diab and Yoko Yamauchi, for sharing tears and laughter, and especially for their support during my sergical operation. I am also indebt to admirable seniors, Prof. Kasajima, Prof. Koyanagi, and Prof. Sugawara. Thanks also go to great colleagues and my special friends in Edinburgh and elsewhere:

Angela, Kazumi, Michiko, Ian, Rosalia, Mariangeles, Andy, Susana, Takeshi, Evia, Sheila, Shinobu, Moira, Noriko, Charles, Mika, Miki, Ivan, Antje, Wouter, Kyoko, Kaori, Koji, Masaya, Christina (S & C), Christine, Hannele, Florence, Charlotte, Rhona, Clarissa, Barbora, Vinton, Tareq, Manabu, Takae, Eileen, XiXi, Frances, Anna, Sophy, Minako, Hajime, Wendy, Mariko, Oliver, Ski *Aikoukai* mates, Taichi friends, German classmates and all the others in my heart.

Sincere gratitude is expressed to my cousin, Shingo, Mr. & Mrs. Sato, Mr. & Mrs. Ishibashi, Mr. & Mrs. Fujita for their tremendous support for me and my mother during our toughest years. Thanks to Dr. Ellis, Dr. McIntyre, Dr. Gilchrist, Ms. Wilson, and Dr. Yamaguchi for their medical support.

Last but not least, I would like to express my special thanks to my dearest family, my parents, Masaji and Katsuko, Tony and Jan, my grandparents, Kenzo and Omie, Yoshie and Ei, my sisters, Jessie and Rachael, and my partner, Olaf, for their constant love, patience and effort over the years and for the joy of discovering world wonders. Without their moral and financial support, I would not have been able to go through the process. This thesis is dedicated to my family for their love, support and unforgettable and loving memories.

*Arigatou!*

# Table of Contents

# List of Diagrams

# List of Figures

# List of Tables


Table 3.1 Characteristics of English speaking participants (NEinexp and NEexp learners). 65

# 1 Introduction

## 1.1 Aims of the thesis

Learning another language often requires tremendous effort for adult learners. Learners need to acquire a wide range of linguistic features in the target language in order to master the language, including grammar, semantics, phonology, phonetics and pragmatics. One of the most challenging areas for adult second language (L2) learners, however, is known to be L2 speech production (e.g., Major, 2001; Sheldon & Strange, 1982).

It is widely observed that L2 adult learners often face great difficulty in learning some L2 contrasts. Non-target like L2 production is often detected as ‘foreign accent’ by native speakers. Even for fluent L2 speakers, foreign accent is easily detectable in their speech by native listeners of the target language. The main aim of this thesis is to investigate the factors behind foreign accent in the adult L2 speech acquisition of prosodic contrasts, using cross-linguistic empirical evidence. What could be the explanation for foreign accent?

To explain the general speech production mechanisms of a language, a number of models have been proposed (e.g., Dell, 1986; Levelt, 1989; Levelt, Roelofs & Mayer, 1999; Liberman, 1996). The models differ in many ways, such as the processing order, the processing type, the intermediate units, the number of unit levels, and whether they include feedback systems. However, these speech models seem to agree on three main components involved in the speech production mechanism: input process, output process and mental representation. In brief, the input process involves perceiving phonetic signals. The output process involves phonetically realizing the signals as speech. Then there is also the stored mental representation, which is related to the ability to associate phonetic signals with phonetic representations and the ability to retrieve relevant information from phonological or lexical representations when we implement speech. Even though these models are not proposed as a language learning model for L2 speech acquisition, the abilities that are related to these three components have also been found to be crucial for L2 speech acquisition regarding foreign accent. For adult L2 learners to achieve target-like productions, the literature suggests that three abilities are the key factors: 1) learners' ability to differentiate the acoustic correlate of the target L2 contrast, which is related to input process (e.g., Polka & Werker, 1994; Werker & Tees, 1984), 2) learners' ability to articulate the acoustic correlate of the target L2 contrasts, which is associated with output process (e.g., Esling & Wong, 1983; Kerr, 2000), and 3) learners' ability to categorize the target L2 contrasts, which is related to mental representation of the target categories (e.g., Best, 1995; Flege, 1995; Flege, Bohn & Jang, 1997).

According to the literature, one possible explanation for foreign accent focuses on the input process. Since the learners are adults who have already mastered a complete set of L1 sounds as a sound system, L2 learners may have less perceptual sensitivity to the relevant acoustic correlates of non-native sounds, and so they may have more difficulty in differentiating the acoustic dimensions which are linguistically irrelevant in their L1. Another possible explanation has a focus on output processes. L2 learners seem to have a problem in articulating target L2 contrasts, presumably due to lack of practice in using the relevant motor control skills. A further possibility focuses on mental representation. In this view, foreign accent could stem from a problem in categorizing target L2 sounds. As we will discuss later, categorization ability can be decomposed into three elements: linguistic perception, categorization and lexical assignment. Linguistic perception refers to the ability to perceive the phonetic difference in the target L2 contrasts and classify each token into the appropriate L2 category. Categorization refers to the ability to associate the phonetic patterns with the target L2 categories, something that is related to phonetic representations of the categories. Lexical assignment refers to the ability to assign L2 categories to lexical items, which is related to lexical-phonological reorientations of the categories. A problem in categorization could be due to cross-linguistic differences as well as cross-linguistic similarities. However, by learning to associate the relevant acoustic information with the target L2 categories and to associate the L2 categories with lexical items in the L2, learners may be able to form phonetic and lexical-phonological representations of the L2 target categories.

In this thesis, we aim to account for foreign accent in adult L2 learners through evaluating these three potential factors. Previous studies have focused on one of the three potential factors but not on a combination of these multiple factors. However, it is important to evaluate the potential interactions between these factors in order to deepen our understanding of the issue of foreign accent in the L2 speech acquisition. Hence, by examining different but related abilities in the same learners, we aim to evaluate these three potential factors to see how far these factors can explain the foreign accent of adult L2 learners. As a secondary aim, with a cross-sectional approach, we intend to provide empirical evidence of whether or not the nature of learners' problems with foreign accent changes as they gain L2 experience.

The following methods will be used in order to achieve these goals. This thesis will examine cross-linguistic empirical evidence of the L2 acquisition of one aspect of Japanese prosody by English learners, namely Japanese lexical pitch accent. We will also test two groups of English learners of Japanese – experienced and inexperienced – in comparison with Japanese native speakers. We will exploit L2 prosody for the insights that it can give us into the issue of foreign accent. Thus far in studies of foreign accent, the main attention has been paid to L2 segmental contrasts. However, another crucial area for the issue of foreign accent is L2 prosody. Prosody has been found to significantly influence the auditory impression which native speakers of the target language make of an L2 speaker's accent. Moreover, the area of L2 prosody allows us an especially valuable means of investigating foreign accent in L2 prosodic acquisition, since prosodic variables (such as stress, length or pitch) tend to play a cross-linguistically different role between languages while at the same time retaining significant cross-linguistic similarities. Only a limited number of studies have so far investigated L2 prosody in relation to foreign accent (e.g., Flege, Munro & Mackay, 1995; Mennen, 1998; Mennen, 1999; Mennen, 2004; Ueyama, 1997).

The key acoustic correlate of Japanese lexical pitch accent is fundamental frequency (F0). In Japanese, F0 has a lexical function, that is, it distinguishes lexical items, unlike in English, where F0 by itself does not have the function of distinguishing lexical items. In both Japanese and English, however, F0 has a phrasal function, namely, of distinguishing intonation patterns. To learn L2 intonation patterns such as for questions or statements, L2 learners need to form the phonetic categories of the intonation patterns and also make the association between the target F0 patterns with the target function. However, to learn lexical pitch accent, L2 learners need to go beyond this, and also learn the arbitrary connection between the categories of the accent types and each lexical item in the L2. Therefore, the acquisition of Japanese lexical pitch accent contrasts by English learners can provide a good testing ground to illustrate the cross-linguistic differences between learners' L1 and L2.

Specifically, in cases like Japanese lexical pitch accent, a key acoustic correlate of an L2 contrast differs from the L1 in one of its functions while sharing another function with the L1. This offers great potential for shedding light on the issue of foreign accent. In such cases, the linguistic function of the key acoustic correlate of the L2 contrasts may not be completely foreign to the L2 learners, since they also use the shared function in their L1, yet they still need to learn the different function of the acoustic correlate as it is employed in the target L2. To be able to signal the functional difference, L2 learners need to learn that the target L2 categories are acoustically similar to the equivalent L1 categories but yet play a different linguistic role. To aim at target-like productions, the learners may therefore need to be able to hear the phonetic difference of the L2 categories within the range of the L2, not within the range of their L1. However, the learners may have problems in doing so because they can hear only the phonetic difference which signals the function in their L1. In addition, target-like productions may require the L2 learners to be able to (re-)associate the relevant phonetic information with each of the target L2 categories from the association that they have for their corresponding L1 categories. Since the association between the phonetic patterns and the L1 categories are closely linked in the L1, the learners may have problems in taking the same phonetic information and associating it with the target L2 categories.

They may also need to learn new labels for the target L2 categories. Ultimately, the learners may also need to understand associations between the L2 categories and their roles in the L2 that are different from their L1, in order to achieve native-like productions. Moreover, motor control skills or phonetic implementations that are required to signal target-like L2 contrasts could be specific to the L2 function. Thus, the cross-linguistic difference of F0 function between Japanese and English allows us to shed light on potential problems for English learners of Japanese when they make this functional difference in their productions of Japanese lexical pitch accent: whether the learners have problems in perceiving the relevant phonetic signal, in phonetically realizing the signal, or in forming phonetic and phonological representations to link these two.

By examining our experimental data, this thesis addresses the following six questions: 1) whether or not native speakers of English show non-target-like patterns when their productions of Japanese pitch accent are compared with those of native speakers of Japanese; 2) whether or not native speakers of English have problems in perceptually differentiating the target F0 contours; 3) whether or not English speakers have problems in articulating the target F0 contours; 4) whether or not English speakers have problems in categorizing Japanese lexical pitch accent contrasts, with categorization considered in terms of its three components, linguistic perception, categorization and lexical assignment, 5) whether or not these three factors can explain the non-target-like production patterns of English speakers; and 6) whether or not the production problems of English speakers change as they gain L2 experience.

To answer these questions, we first examine the productions of English speakers in terms of the auditory impression of native speakers of Japanese (i.e., intelligibility scores) and overall descriptions of their F0 patterns in comparison to Japanese speakers' productions. We then move on to the evaluation of each of the three potential factors related to foreign accent. A more detailed overview of the thesis is provided in the following subsection.

## 1.2 Thesis overview

The chapters of this thesis are organized as follows.

In Chapter 2, we review the previous studies on foreign accent which motivate this study, including a presentation of how foreign accentedness has been assessed in the literature. We also review the three factors which have been suggested to lie behind foreign accent; 1) problems in differentiating the primary acoustic correlate of the target L2 acoustic contrasts, 2) problems in articulating the primary acoustic correlate of the target L2 acoustic contrasts, and 3) problems in categorizing the target L2 contrasts. We then describe the key characteristics of lexical pitch accent in Japanese (i.e., Tokyo Japanese) and also the cross-linguistic differences in F0 function between Japanese and English. In the last section of Chapter 2 we describe the methodological strategies used in this study.

There are then five chapters covering the experiments in which the same groups of learners were tested in different ways. Chapter 3 presents the production experiment, where imitations of Japanese lexical pitch accent produced by native speakers of English are compared with those produced by native speakers of Japanese. Two groups of learners are tested, differing in the amount of L2 experience (i.e., experienced and inexperienced learners). The comparisons between the groups in terms of intelligibility scores and the descriptions of overall F0 patterns allow us to examine whether or not the learners' production patterns show non-target-like (i.e., foreign accented) productions, and how they are different from native speakers' patterns. In each of the subsequent chapters, different factors of interest related to foreign accent are investigated.

Chapter 4 presents a perception experiment using non-speech stimuli (i.e., F0 contours which are extracted from canonical tokens of Japanese lexical pitch accent contrasts). This experiment helps us to evaluate the extent to which it is plausible to suggest that learners' foreign accent problems are due to difficulties in differentiating the acoustic correlate of the target L2 acoustic contrasts as part of the input process.

Chapter 5 presents a production experiment using the same non-speech stimuli as were used in the perception experiment in Chapter 4. The overall F0 patterns produced in the non-speech imitations by the English learners of Japanese are compared with those produced by native speakers of Japanese. This allows us to examine whether or not learners' problems in their speech productions are related to their ability to articulate the acoustic correlate of the target L2 acoustic contrasts as part of the output process.

Chapter 6 presents three perception experiments which aim to examine three different aspects of learners' categorization ability (i.e., linguistic perception, categorization and lexical assignment). The first perception experiment uses canonical tokens of Japanese lexical pitch accent contrasts, and allows us to diagnose whether or not the learners are able to hear the difference between the tokens and to classify the target item into the L2 category based on similarity judgments. The second perception experiment uses stimuli on the relevant continuum. It allows us to test the learners' ability to categorize the target L2 contrasts and to diagnose whether the learners have formed phonetic representations of the L2 categories by associating the phonetic information with the L2 categories. The third perception experiment involves assigning labels. It provides us with a means to examine whether or not the learners have problems in assigning the L2 categories to a lexical item and allow us to diagnose whether the learners have formed lexical-phonological representations of the L2 categories.

Chapter 7 builds on the findings of Chapter 6, where it will be shown that a subset of the learners who had problems in classifying boundary items (categorization) and also were not able to assign the L2 categories to lexical items (lexical assignment) seem to have problems with phonetic and lexical-phonological representations of the target L2 categories. These learners tended to show different patterns from the rest of the group members. Hence, Chapter 7 re-discusses the whole results that were found to be related to the learners' foreign accent after excluding these learners. This helps us to specify the learners' problems behind foreign accent and allows us to evaluate the potential interaction of the factors further.

Chapter 8 presents a summary and general discussion of the findings obtained in this study, with a focus on evaluating which factor (or factors) is best related to foreign accent. We then discuss the conclusion which this thesis comes to. Lastly, the chapter ends with some further issues for future studies.

# 2 Background

## 2.1 Introduction

In this chapter, the background information which motivates this study is presented. First, the issue of foreign accent is described. Next, three issues that could be associated with foreign accent are introduced. Then, the key characteristics of Japanese lexical pitch accent (Tokyo Japanese) are presented in terms of the function of F0, and also in terms of cross-linguistic difference between Japanese and English. Lastly, a methodological overview of the data collection of this research is provided.

## 2.2 Foreign accent

What is foreign accent? Foreign accent is usually described as the perceived degree of non-nativeness of L2 learners' speech productions, based on the overall impression of the native speakers of the target language. It has a great importance not only because it could be used as a means to measure the fluency of the learners, but also because of its relation to other factors, such as age of arrival (e.g., Flege, Birdsong, Bialystok, Mack, Sung & Tsukada, 2006), L2 experience (e.g., Flege, Bohn & Jang, 1997), sociolinguistic variation (e.g., Major, 2001), speaking rate (e.g., Hirata, 2005; Munro, 1998) or social distance (e.g., Al-Issa, 2003; Schumann, 1976). In addition, other factors that are discussed to be relevant to foreign accent; gender, motivation, length of residence (which is often used as an index for the amount of L2 experience), formal instruction in L2 classroom, the amount of L1/L2 daily use, and language learning aptitude (e.g., musical ability or mimicking ability) (see Piske, MacKay & Flege, 2001). As we can see, foreign accent has been widely discussed and investigated in the field of L2 speech studies. However, no straightforward answer to the question about the factors behind foreign accent with conclusive evidence has emerged so far. More specifically, no clear answer has been found for what is contributing to non-target-like productions of L2 learners.

Foreign accent is often measured as an overall auditory impression given by native speakers of the target language in terms of ‘intelligibility’ or ‘comprehensibility’ (e.g., Derwing & Munro, 1997; Major, Fitzmaurice, Bunta & Balasubramanian, 2002). Various other methods of evaluating L2 learners’ foreign accentedness have been also used in the literature. One method is to transcribe learners’ productions in order to compare them with the target productions (e.g., Catford & Pisoni, 1970; Neufeld, 1988; Yang, 1996). Alternatively, acoustic measures of the acoustic correlate of the target L2 contrasts can be used as a means of identifying the criterion for foreign accent. For example, in case of vowel contrasts, the movement of the first and the second formant (F1 and F2) have been measured as the acoustic correlates of vowel height and frontness in both languages (e.g., Bohn & Flege, 1997; Barry, 1989; Flege et al., 1997). In case of stop contrasts such as /t/-/d/ (e.g., Flege & Efting, 1987) or /p/-/b/ (e.g., Gass, 1984), VOT was measured to assess the degree of foreign accent as the acoustic correlate of voicedness in L1 and L2.

These methods of evaluating L2 learners' productions have had varying degrees of success. Although the transcription method could be beneficial, it requires special skill that needs to be trained. However, in everyday life, foreign accent can easily be detected by untrained listeners without such skills. Hence, the transcription method may not capture the overall auditory impression of learners' productions which native listeners of the target language experience. In addition, it has been also pointed out as a problem for this method that transcription systems are not always consistent across languages (e.g., Flege et al., 1997; Yang, 1996). For these reasons, using the transcription method for assessing foreign accent may not be appropriate. Alternative methods for assessing the degree of foreign accent in the existing studies can be generally classified into two types; intelligibility measures[1] and acoustic measures. Auditory judgements by native speakers, i.e., intelligibility ratings, are very informative as they reflect the overall impression of non-nativeness which native speakers experience when they hear learners' speech productions. This is the essence of foreign accent. However, the disadvantage of intelligibility measures is that they could

[1] By 'intelligibility, I intend to mean whether or not the target token can be identified by the native speakers of the target L2 as intended.

allow researchers to overlook distinctions which the learners may be phonetically realizing in their productions (but which may not be apparent from the judgments of native speakers). On the other hand, the advantage of acoustic measures is that they do allow researchers to capture specific details of the phonetic differences in learners' productions of the target L2 contrasts. Although acoustic measures also require trained skill for analyses, this method allows us to evaluate the degree of non-nativeness by comparing the acoustic differences that are not affected by the auditory judgment on accentedness. Nonetheless, it is possible that acoustic measures may not fully reflect native speakers' overall impression. It is possible that learners' productions are intelligible yet perceived as having foreign accent. This foreign accentedness may stem from phonetic differences in their productions that are not fully target-like. For these reasons, both intelligibility measures and acoustic measures would mutually fill their weaknesses in evaluating learners' foreign accentedness.

Some studies use only intelligibility scores for the assessment of production data (e.g., Borden, Gerber & Milsark, 1983; Rochet, 1995; Sakamoto, 2003; Sheldon & Strange, 1982) while other studies use only acoustic measurements of the relevant acoustic correlates of the target L2 contrasts (e.g., Barry, 1989; Bohn & Flege, 1997; Gass, 1984). For instance, Goto (1971) and Borden et al. (1983) used the judgments of native speakers of English as intelligibility scores for L2 learners (native speakers of Japanese in this case), whose task was to produce minimal pairs of English words containing either /r/ or /l/. The intelligibility scores were calculated from the proportion of the overall success as to whether each production of the native speaker of Japanese was identified as intended. On the other hand, Bohn and Flege (1997) acoustically measured the duration and the formant values of English vowel contrasts (/ɛ/-/æ/) produced by the L2 learners (native speakers of German) and compared the learners' values with the values produced by native speakers of English. In another study, Gass (1984) measured the VOT values produced by L2 learners with different L1 backgrounds to compare English stop consonant contrasts (/b/-/p/). These are some examples of the studies which used either intelligibility scores or acoustic measurements for evaluating the learners' productions. Only a few studies have used a combination of both types of analysis (e.g., Flege, 1993; Flege, Bohn & Jang, 1997). For example, Flege et al. (1997) calculated the intelligibility scores of the learners' productions of minimal pairs of English words containing /ɛ/-/æ/ or /i/-/ɪ/ as the criterion. In addition, they measured the acoustic values of each vowel in terms of duration, vowel height and frontness/backness.

Intelligibility scores provide indeed an important criterion for assessing learners' production data. As one of the main aims of L2 acquisition is to be able to communicate in the target language, it is crucial whether or not the utterance is identifiable by the native speakers of the L2. However, despite its importance, if we only measure intelligibility scores, we solely rely on the judgments of raters who are native speakers of the target language. These judgments have limitations including the fact that they may not be able to capture more precise details of the data. There may be some minor but systematic differences which exist in the data but which cannot be found in the results of the intelligibility scores. Therefore, although intelligibility scores may be sufficient to investigate foreign accent, by combining both types of analysis, it should be possible to explore and understand the data more accurately, as the shortcomings of the two methods will be mutually supplemented.

Previous studies have focused on L2 segmental contrasts to investigate the issue of foreign accent (e.g., Flege 1995; Llisterri 1995; Sakamoto 2003; Strange 1995). However, some studies suggest that a larger role for prosody in foreign accent perception than segments (e.g., Boula de Mreuil & Vieru-Dimulescu, 2006; Magen, 1998). Since prosodic errors are more likely than segmental errors to influence the impression of foreign accent (Kawano, 1998; Pennington & Richards, 1986), it is clear that the investigation of L2 prosodic contrasts as well as segmental contrasts is important for studies of foreign accent. What is also unique about prosody from the point of view of foreign accent is that prosodic variables such as stress, length or F0 tend to play a different role cross-linguistically. While both the L1 and L2 of the learners sometimes share the acoustic property of prosodic information, a different function may be assigned to the acoustic property in different languages. In this case, learners have to learn how to use the relevant prosodic parameter in terms of its function in L2 to produce and identify L2 contrasts. A failure to do so may result in foreign accent or poor identification ability.

This can be exemplified with F0, which is used in both the Japanese and English speech system, yet it plays a functionally different role in the two languages. In Japanese, F0 is used for lexical distinction as well as intonation patterns. Meanwhile, in English, F0 is used to distinguish intonation patterns but not lexical items. Although F0 is used in the L1 English speech system, foreign-accented speech has been observed in the Japanese lexical pitch accent contrasts produced by English speakers (e.g., Hirata, 1999; Toda, 2003). These studies showed that native speakers of English have difficulty in producing Japanese lexical pitch accent contrasts with foreign accent. Therefore, the investigation of L2 prosodic contrasts is a good testing ground for issues of foreign accent, and has the potential to help us understand the L2 speech learning mechanism more comprehensively.

What, then, could be the potential explanation for foreign accent? Let us now consider three potential factors which have been separately investigated in the literature.

### 2.3 Difficulty in differentiating the acoustic correlate of L2 contrasts

One possible explanation for L2 speech problems has been discussed in terms of loss of perceptual sensitivity, i.e., loss of the ability to differentiate sounds which are linguistically irrelevant in learners' L1. According to this view, as a problem in the input process of speech productions, foreign accent is caused by loss of the ability to differentiate the phonetic property which is relevant in the L2, i.e., employed to signal L2 contrasts. In the L1 acquisition literature, children show perceptual sensitivity to native contrasts even before they can produce these contrasts, and their perceptual sensitivity is found to be tuned into the target phonological system within the first year of life (e.g., Polka & Werker, 1994; Werker & Tees, 1984). Indeed, Werker and Tees (1984) show that there is a significant decline in infants' perceptual sensitivity, as seen in the ability to differentiate non-native contrasts, between 6 and 12 months of age. This suggests that there is some degree of loss of ability to discriminate non-native sounds as a function of age in terms of acoustic perception. If learners are not able to distinguish the relevant phonetic property triggering the target L2 contrasts, they may not be able to translate the phonetic differences between the contrasts regarding the phonetic property into linguistically meaningful units. Hence, it is possible that adult L2 learners cannot make linguistic distinctions between the target contrasts regarding the phonetic property in their productions of the target L2 sounds. Alternatively, it is also possible that the ability to perceive differences regarding the acoustic correlate of the L2 contrasts and the ability to produce these differences are independent from each other.

It has also been argued that, due to biological reasons related to brain plasticity, once we have lost perceptual sensitivity to non-native sounds, we cannot regain it, at least after a certain age around puberty (Lenneberg, 1964). According to the Critical Period Hypothesis (which hypothesizes that there is a critical period for learning speech (Scovel, 1969; Patwoski, 1989)), the chance of adults being able to learn L2 sounds is considerably low, as they have already lost the perceptual sensitivity to non-native sounds. For example, Goto (1971) reports poor discrimination ability in adult Japanese speakers whose task was to distinguish two English words containing /r/ and /l/. The question therefore arises of whether or not loss of perceptual sensitivity could be a good indicator of poor pronunciation ability reflected in the foreign accent of adult L2 learners. Miyawaki, Strange, Verbrugge, Lieberman, Jenkins & Fujimura (1975) found that Japanese listeners who were unable to discriminate /ra/-/la/ syllables, were able to discriminate isolated F3 stimuli taken from these syllables. Iverson, Kuhl, Akahane-Yamada, Diesh, Tohkura, Ketterman and Siebert (2003) found that Japanese listeners were more sensitive to differences in F2, which is not the crucial acoustic property, than the relevant differences in F3 for the English /r/-/l/ contrast. These studies suggest that even though learners have not completely lost sensitivity to the relevant acoustic correlate of the target contrast and are able to hear the differences between the acoustic properties, sensitivity to a certain acoustic property could be influenced by their L1s and they could be more attentive to a linguistically irrelevant acoustic property rather than the actual acoustic correlate of the L2 contrast.

Werker & Tees (1984) found that native speakers of English could discriminate extracted consonant parts of Hindi dental and retroflex stops (i.e., the ejective portion consisting of the burst and the beginning of the transition), although not the original CV syllables. Best, McRoberts & Sithole (1988) also found good discrimination ability of Zulu clicks by native speakers of English. These findings indicate that L2 learners have not completely lost perceptual sensitivity to non-native phonetic properties but that instead a phonological problem may be involved in the L2 learners' difficulty. In other words, it is possible that adult L2 learners are not "acoustically deaf" to the relevant phonetic property but "phonologically deaf". This implies that learners may have a problem in phonologically associating the acoustic information and the relevant linguistic information in the target L2 since the information of the relevant acoustic property is masked by other linguistic information in a linguistic context. Although learners may be able to distinguish acoustic differences between the target L2 contrasts regarding the relevant acoustic property in a non-linguistic context where they can focus on that property. However, they may not be able to hear the differences once the relevance acoustic differences are embedded in a linguistic context, where it may be difficult for learners to focus solely on the relevant acoustic property. If L2 learners maintain perceptual sensitivity to non-native sounds in the L2, the perceptual sensitivity account seems less likely to provide a sufficient account of the foreign accent of L2 learners. It is less plausible when both the L1 and L2 share the prosodic property which signals L2 contrasts while employing the same prosodic property in a functionally different way. When learners are linguistically familiar with the acoustic correlate of the relevant prosodic property in their L1 speech system, it is possible that learners would be sensitive to the acoustic correlate of the target L2 contrasts regardless of the degree to which they have acquired the L2. However, there is no guarantee that learners have good acoustic perception ability.

In the case of Japanese lexical pitch accent acquisition by English speakers, for example, it is not likely that native speakers of English have lost their sensitivity to F0 differences since F0 is also used in English. If this is correct, the problem in the ability to differentiate the acoustic correlate of L2 contrasts as part of the input process would not be a plausible explanation for foreign accent. It is possible that English speakers are not as equally sensitive to the acoustic correlate of Japanese lexical pitch accent, i.e., F0, as Japanese speakers. Alternatively, it is also possible that even though English speakers are sensitive to phonetic differences of F0 in Japanese lexical pitch accent contrasts, these phonetic differences may be masked by the linguistic information in a linguistic context. Some L1 acquisition studies show the evidence that English-learning infants become less sensitive to certain types of F0 contours by 9 months in comparison to infants learning tone languages such as Chinese or Yoruba as their L1s (e.g., Harrison, 2000; Mattock, Molnar, Polka & Burnham, 2008). Hence, it seems profitable to investigate learners' ability to differentiate the acoustic correlate of L2 contrasts, i.e., F0, where we could put the linguistic information aside. If the learners still show equally good ability to differentiate F0 differences as native speakers in a non-linguistic context, then we could discard the possibility that the difficulty is due to lack of ability to differentiate the acoustic correlate of L2 contrasts.

## 2.4 Difficulty in articulating the acoustic correlate of L2 contrasts

As part of the output process of speech productions, it has been also proposed that a factor behind non-target like productions of L2 contrasts may be articulatory or motor processes (e.g., Esling & Wong, 1983; Kerr, 2000; Mennen, Scobbie, De Leeuw, Schaeffler & Schaeffler, 2010; Strange, 2007). According to this claim, L2 learners may have difficulties in the production of L2 sounds, not because they are physiologically unable to produce them, but because they have not had enough practice or experience in using the oral/nasal space or adjusting articulators such as tongue, lips or jaws in the way that is required to produce the target L2 contrast. Hence, it could also be possible that when English speaking learners produce Japanese lexical pitch accent contrasts which diverge from the native patterns, this may be due to imperfect mastery of the articulatory or motor processes that are relevant for producing Japanese lexical pitch accent contrasts, and which might be independent of the linguistic function of F0.

Catford & Pisoni (1970) looked at non-native sounds (such as glottal stops or close back unrounded vowels) learned by native speakers of English. After providing the learners with training for articulatory postures and movements, they found that the learners showed a significant improvement in the degree of target-like production (73% accuracy) while another group, which was not given such training, did not show such good accuracy in producing these sounds (32% accuracy).[2] The importance of these results is partly that they indicate that it is not impossible for adult L2 learners to learn to produce the acoustic characteristics of non-native sounds when they are given articulatory practice which is relevant to the target sounds. Just as importantly, however, these results suggest that L2 learners were nevertheless having a problem in articulating L2 sounds, since their accuracy

[2] This group had an auditory training instead, where the learners had to listen to the target sounds repeatedly. For this reason, these learners also seem to have learned how to produce them through the training, since the accuracy level was not completely zero.

was no higher than 73%. It seems that the articulation problem shown by the learners manifested itself as divergence from the target norm in their production results (which were assessed based on how target-like each production was).

More recent studies also provide evidence in support of this argument. Thoren (2006) investigated phonological quantity differences in Swedish acquired by a native speaker of Polish (Polish does not exhibit such distinctions). Thoren's results showed that the duration of long and short vowels produced by the native speaker of Polish was similar to that produced by a native speaker of Swedish. However, the duration of long and short consonants produced by the Polish speaker showed deviations from those produced by the Swedish speaker. Moreover, the Polish speaker produced a greater durational vowel ratio compared to the Swedish speaker. In addition to these acoustic differences, the timing and the magnitude of jaw and lip movements used by the Polish speaker tended to be different from those used by the Swedish speaker when articulating the target quantity distinctions. This indicates that the divergent profile of the relevant acoustic correlate of the target contrasts could be due to learner's difficulties in controlling the relevant articulators which contribute significantly to signalling the target L2 contrasts. We should also note that this study does not provide any evidence that the result did not stem from the learner's perceptual ability. Therefore, it is still possible that the learner also had a perceptual problem as well as an articulatory problem.

Lowie and Bultena (2007) also showed that articulatory settings for individual articulators are difficult to acquire even for advanced L2 learners. They investigated vowel formants (i.e., F1 and F2) as a reflection of the articulatory settings of tongue position and lip shape, in the non-native productions of Dutch advanced learners of English. Their results showed that the advanced L2 learners exhibited productions which were rated as non-native-like by native speakers of English. Interestingly, among those learners who showed foreign accent, there were some who were also unable to achieve the target-like articulatory control that was measured in their study. On the other hand, other learners who showed little foreign accent used articulatory settings which were similar to the target norms. This indicates that even as advanced learners, the Dutch speakers were having problems with the articulatory settings for English vowels, and these learners also show divergent profile from the target-norm. Hence, their results imply that the learners' non-native-like patterns in their productions, perceived as foreign accent, might be due to their articulatory problems. Another important point is that Lowie and Bultena suggest that acoustic analysis of the acoustic correlate of the target L2 contrasts can be used as a way of measuring articulatory settings. Again, though, this study does not provide evidence that the result was not due to the learner's perceptual problems, and it could therefore be possible that the learners had perceptual problems in addition to their articulatory problems.

With the recent development in technology, a few studies investigated the language-specific phonetic settings by applying ultrasound images which allows us to examine actual articulatory settings such as tongue positions or lip protrusion (e.g., Gick, Wilson, Koch & Cook, 2004; Wilson, 2006). Wilson (2006) showed that articulatory settings such as lip protrusion and lip narrowing differ across English monolinguals, French monolinguals and English-French bilinguals (Wilson, Horiguchi & Gick, 2007). This implies that L2 learners may have problems in achieving the target-like articulatory settings and thus, not be able to produce native-like tokens.

Taking these pieces of evidence into consideration, it would appear that articulatory or motor control problems may also be a source of the difficulties, as well as perceptual problems, which appear to be reflected in the productions of English speakers whose F0 patterns diverge from the native norm when producing Japanese lexical pitch accent. However, these findings were limited to a small number of participants and also apply only in limited cases. Therefore, the question still remains as to whether or not we can generalize the suggestion that articulatory/motor control processes are a factor which contributes to the non-native patterns seen in the acoustic correlates of the target L2 contrasts in the learners' productions.

A related question of interest is to consider what happens when learners may be familiar with the articulatory demands of the L2 through the articulatory demands of their L1. For example, it is not clear whether or not native speakers of English are having a problem in articulating the different F0 patterns of the different F0 contours of Japanese lexical pitch accent contrasts. Native speakers of English are certainly familiar with producing F0 movements at the phrasal (post-lexical) level through intonation patterns in English, something which, from the motor control perspective, may be similar to the production of Japanese lexical pitch accent contrasts. Therefore, English-speaking learners may not be having a problem in articulating F0 patterns in their productions of Japanese lexical pitch accent contrasts. If we could investigate their articulatory ability to make F0 movements by somehow removing the linguistic function of these F0 movements, and if it happens to be the case that they have a problem in articulating these movements, it would be possible to explain the non-native-like F0 profile of English-speaking learners by an articulatory problem as part of the output process of speech productions.

However, producing F0 movements to signal lexical items may yet require slightly different articulatory/motor processes from producing them to signal intonation patterns. For instance, lexical level F0 movements may require the articulators to be controlled or moved within a shorter period of time than those at the phrasal level. This implies that this may also require different ability to understand the connection between lexical items and phonetic patterns. It is also possible that the problems of English-speaking learners may be entirely independent from such cross-linguistic difference in the F0 function between Japanese and English. Taking these possibilities into account, it seems reasonable to first investigate whether or not the English-speaking learners are able to articulate F0 movements of Japanese lexical pitch accent when we put aside the linguistic function of F0 in each language. If the learners are having problems in articulating F0 movements, this in turn suggests the possibility that the divergent F0 profile of English-speaking learners can be attributed to their articulatory problem.

### 2.5 Difficulty in categorizing L2 contrasts

Another important component of speech production is the internal mechanism that connects the input process and the output process. Some have argued that it is not possible for adult language learners to acquire L2 sounds after a critical period (e.g., Patwoski, 1989; Scovel, 1969). In their view, it is not possible for adult L2 learners to form the target L2 categories. Meanwhile, others maintain that the capacity of adult L2 learners to learn non-native sounds has not been completely lost but rather it remains modifiable (e.g., Best & Strange, 1992;

Flege, 1995). In other words, they argue that even adult learners are still capable of learning L2 speech to some extent. Among the researchers who support this view, the problem of categorization ability has been discussed as the determining factor for the non-native-like productions of L2 learners. That is, according to this view, learners' non-target-like patterns in L2 speech productions can be explained in terms of their ability to categorize the target L2 contrasts. Hence, according to this view, foreign accent is the reflection of learners' poor categorization ability.

To explain non-native-like productions in L2 speech acquisition, Flege proposed a working model of L2 speech learning, i.e., the Speech Learning Model (SLM) (Flege, 1995). This model was developed through a series of empirical studies which he conducted with his colleagues. This model aims at capturing the L2 learning mechanism on the basis of the perceived similarity between L1 categories and L2 categories. It predicts that learners assimilate their L2 inventory (i.e., category membership) to their L1 inventory, especially in the early stage of learning, when they have little L2 experience. The more similar L2 phones are perceived to be to L1 categories by a learner, the more likely these items in the L2 inventory will be assimilated into the L1 inventory. According to this model, the problem of categorization is due to this assimilation process. As a result, L2 learners may consider two categories in the L2 to be one category. Because of this assimilation, learners are not able to categorize L2 contrasts into two (or more) different categories. If they are not able to differentiate the L2 categories, learners may show poor production ability in terms of achieving native-like norms for the target contrasts. As a result, this poor performance in the learners' production would be detected as foreign accent. However, this model also assumes that the perceived relation between L1 and L2 sounds does not remain static and may potentially change as a function of L2 experience. Once learners become capable of perceptually categorizing the target L2 contrasts, they will also be able to produce the contrasts categorically. Therefore, the ability to categorize L2 contrasts seems to be a good indicator for non-native-like productions of L2 learners.

Evidence in support of this view comes, for example, from Rochet (1997). Rochet tested the ability to produce and categorize French high vowel contrasts, /i/-/y/-/u/, in Portuguese and English speakers as L2 learners of French. Whereas there are phonologically three high vowel categories in French, in both Portuguese and English there are only two high vowel categories (i.e., /i/ and /u/ but not /y/). The interesting difference between the two categories in Portuguese and English was in the average F2 value of /y/. The average F2 value of the French /y/ happens to fall within the range of the Portuguese /i/ category, whereas in English, it happens to fall within the range of the English /u/ category. Rochet's results showed that Portuguese speakers tended to imitate and categorize French /y/ as /i/, and English speakers tended to treat French /y/ as /u/, in both cases like their own L1s. This can be taken as evidence that non-native-like L2 vowel production may stem from the learners' categorization ability (i.e., the ability to hear the difference between the target L2 contrasts and classify the target stimuli into the target categories). In addition, this also suggests that learners' categorization ability may be influenced by how the target sound is perceived in terms of their L1 categories. For this reason, Rochet's conclusion was that foreign accented productions of L2 learners are perceptually motivated, i.e., motivated by the problem of learners' phonological representation of the target L2 categories.

To investigate the connection between foreign accent and phonological problems, Flege, Bohn and Jang (1997) also investigated learners' production and perceptual categorization ability for the English vowel contrasts /ɛ/-/æ/ and /i/-/ɪ/. They tested L2 learners of English with different L1 backgrounds (native speakers of German, Spanish, Mandarin and Korean). These four languages differ in how native speakers use formant movement and duration as acoustic cues to distinguish vowels. For example, Spanish vowels show less formant movement compared with the closest English vowels, while duration does not play a role in distinguishing Spanish vowels. Korean has phonemic durational contrasts between long and short vowels. German speakers use less formant movement but more duration variation than English speakers in differentiating vowels. If inexperienced learners classify L2 sounds as instances of the closest L1 sound category, the use of the relevant acoustic cues in the learners' L1 will play an important role in how learners perceive English vowel categories in terms of their L1 vowel categories. To test categorization ability, Flege and colleagues conducted an identification task by using a synthetic continuum for each vowel pair, where they manipulated the relevant acoustic correlates, i.e., vowel height and frontness/backness. They found that L2 learners' perceptual categorization ability and their production ability were indeed related. More specifically, the results showed that the non-target-like vowel height and frontness/backness values of English vowels in learners' speech could be attributed to the learners' poor perceptual categorization for the target contrasts. The accuracy of the learners' perception accounted for 29.8% of the /i/-/ɪ/ and 33.5% of the /ɛ/-/æ/ production data on average. In other words, the learners' perceptual categorization ability could partially explain their non-native-like production patterns. This also suggests that the categorization problem may be a good indicator of foreign accent exhibited by L2 learners. Flege et al. also found some cross-linguistic differences in the learners' performance which could have been due to cross-language differences in the perceived difference between their L1s and English. For example, their inexperienced German speakers performed much more poorly in producing and perceiving the difference between the English /ɛ/-/æ/ contrast compared to their inexperienced Spanish speakers. Flege et al. reasoned that German speakers tended to identify realizations of English /ɛ/ and /æ/ as a single L1 vowel category such as German /ɛ/ whereas Spanish speakers tended to identify the same English vowels as two different Spanish vowels (i.e., Spanish /e/ and /a/).

On the other hand, the problem with SLM is that (as Flege et al. (1997) point out) it is difficult to test empirically how learners actually perceive the relation between their L1 categories and the target L2 categories. It is also difficult to predict which L1 category the learners may perceive to be the closest to the target L2 inventories. Perhaps it is relatively straightforward to predict the L1 candidate which potentially falls within the target L2 values in the case of vowels and consonants where the relevant acoustic correlate(s) seem to be used in both the leaner's L1 and the target L2. For instance, in case of vowel contrasts such as in Flege et al. (1997)'s study, the movement of the first and the second formant (F1 and F2) are the acoustic correlates in both languages. Or, in case of stop contrasts such as /t/-/d/ (e.g., Flege & Efting, 1987) or /p/-/b/ (e.g., Gass, 1984), VOT is the acoustic correlate used in both the L1 and the L2. The categorization problem seems to be one plausible factor for foreign accent based on the previous findings. However, in the case of prosodic contrasts like Japanese lexical pitch accent, it is not so simple to predict the comparable L1 categories, since although prosody is employed in many languages, the relevant acoustic cues vary widely from language to language. Hence, testing a reliable method itself requires further investigation. One possible way could be to make use of cross-linguistic differences between the learners' L1 and the L2 in terms of the function of the key acoustic correlate of the target prosodic contrasts. Moreover, it seems also feasible to test learners' categorization ability in terms of how the learners group L2 sounds into corresponding L2 categories. If learners do have a problem in categorizing the target L2 contrasts, it could potentially be a problem which may contribute to their non-target-like productions.

While the categorization ability account of L2 production difficulty has been well supported by experimental findings on L2 segmental contrasts, it has not gone unchallenged. Several studies have reported that even though L2 learners' categorization ability was poor, the learners nonetheless produced target L2 contrasts without foreign accent, which may seem to contradict the prediction above. Goto (1971) and Sheldon and Strange (1982) tested learners' production based on results from their performance on an identification task where the stimuli were speaker productions, and learners' perception skills were based on their ability to identify canonical tokens of both native speakers' productions and their own productions. The finding of these studies was that the learners were able to produce canonical tokens of English words with /r/-/l/ contrasts which were identified by the native speakers as intended. Meanwhile, the same learners had difficulty in perceptually categorizing the canonical tokens of the target contrasts. As Sheldon and Strange (1982) point out, this result may be restricted to advanced L2 learners. In both studies, the Japanese speakers whose production performance was better than their perceptual categorization performance were actually fluent learners of English. In other words, the learners in Goto (1971) and Sheldon and Strange (1982) did not show a significant amount of foreign accent in producing target contrasts despite having categorization problems. Hence, these learners did not seem to have a foreign accent problem. However, many studies report poor production ability for the English /r/-/l/ contrast by Japanese speakers (e.g., Kinnaird & Zapf, 2004; Komaki, Akahane-Yamada & Katagiri, 2002) and also perceptual categorization difficulty (e.g., Miyawaki et al., 1975; Takagi & Mann, 1995). The categorization ability of L2 learners might be dynamic in nature and may develop during the L2 acquisition process along with L2 experience. In this case, we may find that experienced learners may be better able to categorize the target contrasts than the inexperienced learners, due to their L2 experience. Hence, Flege and his colleagues have proposed that the performance of inexperienced learners who exhibit poor production might be explained by their perceptual categorization ability. Unfortunately, the studies by Goto (1971) and Sheldon and Strange (1982) did not test inexperienced learners or less advanced learners. Moreover, the results of these two studies may imply that once perceptual categorization ability reaches a certain level in adult L2 learners, there may be a limit beyond which it does not develop further. It might be that learners' poor production can be explained by perceptual categorization ability up to this level. Meanwhile, as learners' production approaches the native norms, it might be that their perceptual categorization ability can no longer explain their production ability, as their production ability may continue to develop while their categorization ability stops improving. Therefore, this leaves room for investigation, in order to conclude whether or not these results are indeed counter-evidence for the categorization ability account for non-native like productions. In addition, this also highlights the importance of examining the same learners longitudinally, or else cross-sectional groups of L2 learners.

Another implication of these results is that we need to test two different kind of learners' categorization ability for better understanding of learners' performance in categorization tasks. Goto (1971) and Sheldon and Strange (1982) used canonical tokens of the target L2 contrasts, whereas studies like Rochet (1997) and Flege et al. (1997) used synthetic stimuli on a continuum of some relevant acoustic correlate of the target L2 contrasts. With a perception task using canonical tokens, we can test learners' ability to hear the difference between the target L2 tokens and classify the target item into the corresponding L2 category based on similarity judgments (i.e., linguistic perception). On the other hand, since learners need to predict membership of the target L2 categories, a perception task using stimuli on a continuum requires categorization ability that may require phonetic representations of the categories (i.e., categorization). In addition to being able to hear the difference between the stimuli, learners need to associate the phonetic patterns with the target L2 categories. This will be discussed more fully below in 2.7.

## 2.6 The key characteristics of lexical pitch accent in Tokyo Japanese

### 2.6.1 F0 parameters characterizing lexical pitch accent in Japanese

To specify what we need to look out for in the comparison between productions of learners and native speakers, let us now overview the key acoustic characteristics of Japanese lexical pitch accent that have been found in the literature. What is described in this thesis as 'Japanese' refers to Tokyo Japanese, the variety which is considered to be standard Japanese, and the one which is usually taught as 'Japanese' in the L2 classroom.

In Tokyo Japanese, the number of accent types is the number of the syllables consisting of a word, plus one type (n + 1) (Kubozono, 2001). A word can have one pitch accent, which could be on any syllable, or it can be unaccented. For instance, for a disyllabic word, there are always three accent types (i.e., pitch accent located on the first syllable, pitch accent located on the second syllable, and the unaccented type). The unaccented type and the accent type with the pitch accent located on the last syllable are clearly distinguished when words are accompanied by a post-positional particle such as *ga* (subject-marker) or *mo* (also). The key acoustic correlate of Japanese lexical pitch accent is fundamental frequency (F0) and pitch is phonetically realized by F0 movement (Beckman, 1986; Sugito, 1982). Previous studies indicate that what characterizes F0 patterns in native speakers' production of Japanese lexical pitch accent is an F0 peak followed by a rapid fall (Hasegawa & Hata, 1992; Sugito, 1982). We can decompose this into two elements: where the location of the F0 peak of a word is (i.e., F0 peak alignment) and what the shape of the F0 contour is (i.e., the degree of the F0 fall and F0 range).

It has been reported that F0 peaks in Japanese words are aligned relative to a landmark such as the onset or the offset of consonant closure, or the onset or offset of the vowel. For instance, Sugito (1982) observed that the F0 peak of a pitch accent in the productions of native speakers of Japanese tended to occur not on the associated tone-bearing unit but slightly after the associated tone-bearing unit. Ishihara (2006) also showed that in Tokyo Japanese, the F0 peak of words was aligned with specific segmental landmarks not randomly but consistently. According to his results, the segmental landmark for initially accented words with CVCV structure was the vowel of the second syllable, and the aligned point of the F0 peak was on average 0.3ms from the onset of the second vowel where voicing starts. Meanwhile, when the first vowel is long, the F0 peak alignment point was located at approximately 70% of the duration of the vowel of the first syllable. Hasegawa and Hata

(1992) also confirmed that manipulating F0 peak location affects Japanese speakers' perception of pitch accent. These studies indicate that the F0 peak of Japanese lexical pitch accent is systematically organized in native speakers' speech and characterizes native speakers' F0 patterns.

Another F0 parameter that characterizes Japanese lexical pitch accent is the speed of the F0 movement after the F0 peak, i.e., the speed of the falling F0 movement. Sugito (1980) found that in native speakers' productions, Japanese pitch accented words are accompanied by a steep F0 fall after the F0 peak. Hasegawa and Hata (1988) also found that F0 fall rate is one of the important characteristics in the productions of native speakers of Japanese. They reported that there is a correlation between the F0 peak location and the rate of the F0 fall after the peak in Japanese speakers' productions of pitch accent. That is, the later the F0 peak occurs, the greater the F0 fall rate. Based on their production results, Hasegawa and Hata (1992) also showed that F0 fall rate also affects Japanese speakers' perception of lexical pitch accent. This is very important when we consider foreign accent that reflects native speakers' auditory impression. They discovered a correlation between the F0 peak location and the F0 fall rate after the peak. These studies suggest that the steep F0 falling movement is also systematically organized in native speakers' speech of Japanese lexical pitch accent and characterizes native speakers' F0 patterns.

One other important F0 parameter is F0 range. F0 range is generally one of the characteristics of the F0 patterns which have been used to evaluate the shape of F0 contours (e.g., Beckman, 1986; Ladd, 1997). As F0 range tends to be widely diverse among speakers, previous studies have proposed that normalized F0 range values should be used (Earl, 1975; Ladd, 1997; Rose 1987). This measure can potentially add a useful dimension to the F0 profile as a means to assess foreign accent, especially because the main purpose of the analysis is to compare the productions of learners with those of native speakers of Japanese. We do not know how wide/narrow the F0 contour will be in utterances produced by native speakers of English. Not having fully learned what to do or how to produce the accurate F0 patterns for Japanese lexical pitch accent, native speakers of English may over-exaggerate the rising or the falling of the F0 movement, which could yield extremely wide F0 range. On the other hand, it is equally possible that they would not produce enough F0 movement compared to native speakers, which could yield very small F0 range.

In the literature, other acoustic measures were used for evaluating L2 learners' productions of Japanese lexical pitch accent but not regarding these three F0 parameters. For example, Hirata (1999) compared the mean F0 values of the two vowels in disyllabic words with those of native speakers to evaluate the data. Similarly, Toda (2003) compared the mean F0 values of the two syllables in each disyllabic word with those of native speakers.

However, Sugito (1980) points out that the mean F0 of the accent-bearing syllable in disyllabic words is not necessarily higher than that of the other syllable in the productions of native speakers of Japanese based on her data. For this reason, approaches which rely on solely on mean F0 the accent-bearing syllable do not seem to provide with us an entirely satisfactory way to capture the characteristics of the F0 patterns. Hence, they do not seem to be the best criterion for comparing the learners' productions with the native speakers' productions. To assess foreign accent, however, it is worth while investigating the overall shape of F0 contours in learners' productions of Japanese lexical pitch accent in terms of F0 peak location, F0 fall and F0 range.

With this background, the question which we will go on to consider is: What will happen to these characteristics of overall F0 patterns in the productions of native speakers of English when they produce Japanese lexical pitch accent contrasts? Productions of English speakers of Japanese lexical pitch accent contrasts are investigated in Chapter 3.

### 2.6.2 Cross-linguistic differences in F0 function between Japanese and English

In both Japanese and English, one of the functions of F0 is to distinguish intonation patterns such as affirmatives or questions. Affirmative sentences are marked by a relatively flat overall gradual falling of the F0 contour while question sentences are marked by a clear rising of F0 contour towards the end of the sentence. However, in Japanese, but not in English, F0 has an additional lexical function: F0 is used in Japanese to distinguish lexical items. As described in the previous section (2.6.1), in Japanese lexical pitch accent, pitch (which is phonetically realized by F0 movement) is the only key systematic acoustic correlate, and other acoustic properties (such as duration and amplitude, etc.) do not change systematically. On the other hand, in English stress, pitch is not the only acoustic correlate but one of several acoustic correlates which distinguish lexical items (Beckman 1986; Ladd, 1997). Other acoustic correlates such as intensity, duration, and vowel quality, all are found to play essential roles as well as pitch. Moreover, it has been found that for non-focused minimal pairs of English words, pitch has minor or little cue-value (Sluijter & Van Heuven, 1996a; Slujiter & Van Heuven, 1996b).

We can illustrate this cross-linguistic difference between Japanese and English in terms of the function of F0 using simple examples cited in Kubozono (2001, p82).

(1) a. *piza*

b. pizza

(2) a. *piza*?

b. pizza?

The two words in (1) are ‘pizza’ in Japanese (1a) and in English (1b). The curve indicates the F0 movement of each word when they are pronounced as a word in isolation. As we can see in (1), the word has a falling F0 movement in both languages. The curve in (2) indicates the F0 movement of each word when they are pronounced as a question. The overall Japanese lexical pitch, the falling F0 contour, remains unaffected even in a question sentence and only the end of the phrase has a rising contour, which is independent of the lexical pitch (2a). Meanwhile, in the English question, the pitch of the word is affected by the intonation pattern and absorbed into a question intonation shown by the rising F0 contour (2b). This example shows how, in the case of Japanese, pitch is lexically assigned to a word and thus, F0 has a lexical function. On the other hand, in the case of English, pitch is one of several available acoustic correlates but it is not lexically assigned: F0 does not play a lexical role. Hence, although F0 has the function of marking linguistic intonation differences both in English and in Japanese, F0 does not have a lexical function in English to correspond to its lexical function in Japanese lexical pitch accent. Moreover, to acquire Japanese lexical pitch accent, English speakers need to learn to form phonetic representations of the categories by associating the target F0 patterns with the corresponding categories. In addition, the learners have to learn to form lexical-phonological representations of Japanese lexical pitch accent by associating the categories with Japanese words.

How, then, does this cross-linguistic difference relate to the abilities discussed earlier as potential factors behind foreign accent? Since F0 plays a linguistic role in English, it is possible that English learners of Japanese may able to hear differences between the F0 contours in a non-linguistic context, at least as inexperienced learners. In a linguistic context, one possibility could be that due to the presence of other linguistic information, learners may have problems in hearing differences between Japanese lexical pitch accent contrasts. Another possibility could be that although English learners of Japanese are able to hear the difference between Japanese lexical pitch accent tokens, they may not be able to categorize the tokens into the target L2 categories. This is because English speakers may have problems in associating the F0 patterns with Japanese lexical pitch accent categories. It is also possible that the learners may have problems in assigning L2 categories to Japanese words because they are not able to learn the connection between the lexical pitch accent categories and the lexical items in Japanese. For inexperienced learners, this may be particularly difficult due to poor phonetic representations of the L2 categories for categorization and poor lexical-phonological representations of the L2 categories for lexical assignment. For experienced learners, due to their greater L2 experience, it is possible that they might have formed sound Japanese lexical pitch accent categories, and may thus be able to categorize Japanese words into the target pitch accent categories to some extent. However, it might be still difficult for experienced learners to associate Japanese lexical pitch accent with a word as they have to learn the association one by one.

For the same reason as the ability to hear the difference between the F0 contours, it is also possible that English learners of Japanese will be able to articulate the differences between the F0 contours when the linguistic information is stripped away. However, the differences produced by learners might be different from those produced by native speakers of Japanese. Moreover, since F0 is not the key acoustic correlate in English to mark lexical items, the difficulty which English speakers have in producing Japanese lexical pitch accent could lie in their ability to lexically associate Japanese lexical pitch accent categories with a Japanese word in their phonological representations. As a result, it is possible that this difficulty is related to the foreign accentedness of learners' productions of Japanese lexical pitch accent. These possibilities are investigated through experimental data in the following chapters.

## 2.7 Methodological strategies

This section provides a methodological overview of the data collection of this research. Four experiments were conducted to investigate potential explanations for foreign accent in the L2 prosodic contrasts being acquired by L2 adult learners. The first experiment addresses the question of whether or not L2 learners show non-native-like productions of the target contrasts. Then, each of the remaining three experiments addressed a different question: whether the foreign accent of adult L2 learners is related to, 1) perceptual sensitivity, 2) articulatory/motor processes, or 3) categorization ability. To investigate possible interactions between these factors, the same groups of participants were tested, namely, two groups of adult native speakers of English (one experienced and one inexperienced) who were studying Japanese as the L2, and a group of Japanese native speakers for the comparison (the details of the participants are described in Chapter 3). This cross-sectional design was used in order to capture the nature of the L2 learners' problems associated with foreign accent at different stages of acquisition.

We will now consider each of the four experiments undertaken in this thesis, but for ease of exposition, we will discuss Experiment 1 and Experiment 3 (production tasks) first, before turning to Experiment 2 and Experiment 4 (perception tasks). First of all, Experiment

1 was conducted to demonstrate whether or not the productions of English-speaking learners are target-like. In this experiment, we compare the learners' productions data of Japanese pitch accent contrasts with those produced by native speakers of Japanese. Comparisons are made in terms of intelligibility scores and overall acoustic patterns of F0 contours. To elicit the productions from the learners, we employed an imitation task. There is no conventional way of annotating Japanese pitch accent. A pilot study, which looked at introducing novel diacritic systems, showed that English learners of Japanese seem to have problems in learning diacritics for the different pitch accents, whereas native speakers of Japanese seem to be able to learn the introduced diacritics after a short practice (the details of this pilot study are described in Chapter 6 in relation to one of the perception tasks, i.e., a label assigning task). In order to exclude a possible effect of learning a new diacritic system on the learners' production patterns, we chose instead an imitation method, where participants repeated the audio stimuli rather than reading aloud a list of target items presented as visual stimuli. In addition, to avoid any familiarity effect due the groups' different vocabularies, nonce words such as *mene* or *noma* were selected as test items instead of real lexical items.

In Experiment 3, a non-speech imitation task was designed to examine the issue of articulatory/motor processes. The experimental question was whether or not native speakers of English could articulate the F0 contours of Japanese lexical pitch accent. Participants were required to imitate audio stimuli which consisted of F0 contours. These contours were extracted from natural tokens of disyllabic nonce words such as *mene* or *noma* accompanied by a post-positional particle, i.e., *mo* (also). The resulting audio stimuli sounded like a buzzing noise. We employed an imitation method, the same method as in the speech elicitation task, since for a non-speech production task, there was no option of a read aloud task.

Let us now turn to some methodological considerations in relation to the tasks which are used to test perceptual sensitivity (Experiment 2) and L2 learners' ability to categorize target contrasts (Experiment 4). A commonly used paradigm in the literature on L2 perception studies is the ABX task (e.g., Curtin, Goad & Pater, 1998; Dupoux, Pallier, Sebastian & Mehlar, 1997; Gottfried, 1984; Fox, 1998). Another commonly used alternative is the AXB task (e.g., Dommelen & Husby, 2007; Guion & Pederson, 2007; Mora, 2008). It has been argued that AXB tasks reduce memory load compared to ABX tasks because the target token (X) is presented at an equal durational distance from the two category representatives (A and B) (e.g., Levy & Strange, 2008; Strange & Shafer, 2008). However, ABX tasks seem to be simpler than AXB tasks for the purposes of explaining a task to learners in an experiment. In addition to this practical consideration, we assume that the distance issue can be resolved by randomizing the presentation order of A and B.

ABX tasks can vary in terms of the materials which participants are asked to classify, and therefore taps different aspects of the ability to categorize the L2 contrasts. One option is to use the ABX task to categorize canonical tokens. This allows us to diagnose the learners' ability to hear the phonetic difference between the target L2 categories and to classify an item into the corresponding categories. In this task, a minimal pair of canonical tokens (i.e., A and B) is presented as two representative members of the target L2 categories, and the target sound (i.e., X) is identical either to A or to B. The task is to classify X as an instance either of the category A or the category B. Hence, in order to do that, learners have to be able to hear the phonetic difference between A tokens and B tokens of L2 contrasts and classify whether the X token is matched with the A or the B token.

The other option is when the ABX task is used to categorize step-wise stimuli drawn from a continuum. This task provides us a way to test whether the learners have formed the phonetic representations of the target L2 categories. In this task, the minimal pair presented is the two endpoint stimuli on the continuum, as two representative members of the target categories (i.e., A and B), and the target sound (i.e., X) is one of the step stimuli on the same continuum. The task is to classify the target X token into either the category A or the category B.

These two perception tasks are similar in that in both, the representative members of the target categories are presented. However, there is a crucial difference between them. In the case of a canonical-token ABX task, the target sound is always the same as one of the two sounds provided as the category representatives. Thus, it may be possible for learners to directly compare the sound similarities between a token and the representative members when they try to match the target token with the target category. However, in the case of a step-wise ABX task, the target sound is one of the sounds on a continuum which is set up to be related in terms of one acoustic dimension but slightly different from the presented representative of the target categories, apart from when it is one of the endpoints which are equivalent to the representative members. The learners need to predict which L2 category the boundary items belong to. Hence, this task allows us to diagnose whether the learners have formed the L2 categories as phonetic representations. Therefore, a canonical ABX task can be used for testing linguistic perception ability whereas a step-wise ABX task can be used for testing categorization ability; both of these are elements of the general categorization ability.

Another component of the general categorization ability which we aim to investigate is lexical assignment. The two types of ABX tasks do not allow us to investigate the learners' ability to assign the L2 category to a lexical item, something which seems to involve lexical-phonological representations of the target L2 categories. Hence, in order to test the learners' lexical assignment ability, we also need another tool. One possible way could be to provide only one lexical item and no other items to represent target L2 categories to compare with, and ask the learners to assign the item to its L2 category. This would remove the option for the learner to classify the target stimulus into the L2 category by comparing the similarities based on the presented stimuli, and allows us to test whether the learners have lexical-phonological representations of the target L2 categories. A common paradigm for this kind of perception task is generally called an 'identification task'. However, the design used in each study varies from study to study, both in terms of the types of stimuli used and the types of the categories that need to be selected. Some studies have used stimuli on a continuum such as vowel or VOT values (e.g., Bohn & Flege, 1997; Borden et al., 1983) while others used canonical tokens (e.g., Elseendoorn, 1984; Sheldon & Strange, 1989). Meanwhile, some studies used vowels or consonants alone as the target L2 categories (e.g., Goto, 1971; Rochet, 1995) while others used target sounds embedded in words either by changing the values of the target vowel or consonant value or by using similar words with the same segmental structure while keeping the target vowel or consonant (e.g., Flege et al., 1997; Gass, 1984). However, for our purpose, it was an important part of the ability we aimed to test, that the participants would learn a new diacritic system to associate with the lexical items. This is also a different aim from other studies. For this reason, we designed a label assigning task in order to test the learners' lexical assignment ability.

Let us now come back to our methodological strategy for Experiment 2. This experiment targeted the issue of perceptual sensitivity. To test whether or not English speakers could hear the difference between F0 contours in pairs, a non-speech ABX task was carried out using audio stimuli which consisted of F0 contour extracts. Non-speech stimuli were used in order to exclude the linguistic information from natural word tokens. Participants were required to hear the differences between audio stimuli which consist of F0 contour extracts from natural tokens of disyllabic nonce words such as *mene* or *noma* along with a post-positional particle, i.e., *mo* (also).

Experiment 4 tested whether or not English speakers' ability to categorize Japanese lexical pitch contrasts could explain foreign accentedness in their productions of these contrasts. To investigate their ability to categorize Japanese lexical pitch contrasts, we decomposed the categorization ability into three key elements, namely, linguistic perception, categorization, and lexical assignment. We therefore conducted three different perceptual tasks; two ABX tasks and a label assigning task. In order to diagnose the learners' ability to perceive the phonetic difference between Japanese lexical pitch accent contrasts, we used an

ABX task with canonical tokens of the categories. In order to examine whether or not the learners have formed the phonetic categories of Japanese lexical pitch accent, we used a different ABX task with 7-step stimuli on a F0 peak alignment continuum between accent on the first syllable and on the second syllable. Lastly, a label assigning task was employed where learners were introduced to a novel diacritic system for Japanese pitch accent and were asked to assign a label to each token. This task was designed to test whether or not English-speaking learners were able to assign the L2 categories of Japanese lexical pitch accent to a lexical item. Learning a diacritic system requires the ability to hear the differences between the minimal pairs and also the ability to assign labels to each canonical token by lexically associating a token with the target category. To be able to assign labels correctly, the learners must have formed phonetic categories of Japanese lexical pitch accent contrasts and they also needed to understand the connection between the categories and the lexical items in Japanese. We assume that the more taget-like the phonetic categories become and the more developed their lexical-phonological representations of the target categories, the easier it will be for participants to learn and assign labels.

As the general procedure of the whole study, the experiments were administered individually in a sound-treated recording studio at the Linguistics laboratory at the University of Edinburgh. The whole battery of experiments took approximately one hour for each participant. Participants were allowed to proceed through the tasks at their own pace and to take a rest at any point during the experiments. Each participant took part in the speech tasks (i.e., Experiments 1 and 4) prior to the non-speech tasks (i.e., Experiments 2 and 3) so that participants would not pay particular attention to pitch contours during the speech tasks after having been administered the non-speech tasks. In addition, to counter-balance the task effect, among the speech and non-speech tasks, half of the participants took the perception tasks first and the production tasks next, and the other half took the tasks in the reverse order. The same participants were tested throughout the four experiments to explore foreign accent issue through more comprehensive approach by looking at within-subject effects. The details of each individual experiment are described in the following chapters.

## 2.8 Summary

This chapter has presented the background and methodological strategies of this thesis. Foreign accent is one of the main issues in adult L2 acquisition. So far, there has been no clear answer for what is contributing to the non-target-like productions of L2 speakers. Three factors behind foreign accent have been discussed in the available studies on L2 speech acquisition; learners' difficulty in differentiating acoustic correlates of L2 contrasts, their difficulty in articulating acoustic correlates of L2 contrasts, and their difficulty in categorizing L2 contrasts. However, while these factors have been investigated separately, they have not been studied in combination in the same learners. L2 prosodic contrasts provide us with a good testing ground where there are cross-linguistic differences regarding the key acoustic correlate between learners' L1 and the target L2, while the L1 and L2 nevertheless share some similarities. In this study, the L2 acquisition of Japanese lexical pitch accent by native speakers of English was examined through four experiments to investigate the factors behind foreign accent. This is described in the following chapters.

# 3 Experiment 1: The ability to produce Japanese lexical pitch accent

The aim of this thesis is to investigate potential factors behind foreign accent in L2 speech acquisition. To discuss the possible factors, we first need to address whether or not the productions of the L2 learners are non-native-like (i.e., show foreign accentedness). If their productions are indeed non-native-like, then we need to address how they are different from the productions of the native speakers.

F0 has a phrasal function in both English and Japanese; that is, in both languages, intonation is used to differentiate between affirmatives and questions and so on. In Japanese, but not in English, F0 also plays an important role in distinguishing word meanings. Japanese lexical pitch accent is phonetically realized by the movement of F0. As we have seen earlier (see Chapter 2.6), previous studies suggest that lexical pitch accent in Japanese is characterised by an F0 peak (aligned with respect to a segmental landmark in the pitch bearing unit) followed by a rapid fall (Hasegawa & Hata 1992; Sugito 1982). As L2 learners of Japanese, native speakers of English have to learn these acoustic characteristics of F0 systematically, both in terms of the lexical function of F0 in Japanese as well as its phrasal function. Learning Japanese pitch accent contrasts (the lexical function of F0) can however be problematic for native speakers of English. Considering these issues, there are two questions that need to be addressed. The first question is whether or not F0 profiles in lexical items produced by English-speaking learners diverge from those produced by native speakers of Japanese. If their productions do diverge from the F0 patterns of native speakers, the second question arises as to what the potential factors behind such differences during L2 speech acquisition can be.

In this chapter, we aim to provide an answer for the first question, regarding the F0 profiles in lexical items produced by learners in comparison to those produced by native speakers. To answer this question, productions of Japanese lexical pitch accent contrasts were examined in English-speaking learners and native speakers of Japanese. As Japanese does not mark pitch accent orthographically,[1] an imitation task was used to elicit the intended canonical tokens of target pitch accent types. Once the data was collected, both intelligibility scores and acoustic characteristics for descriptive analysis were adopted as a means of comparing the productions of the learners with those of the native speakers. The criterion of intelligibility scores is very important as it captures the 'foreign accentedness' in the learners' productions. Productions by two learner groups and a native speaker group were compared in terms of intelligibility scores based on the auditory impression of two native

[1] We initially attempted to train participants to learn a novel diacritic system that was designed for this study to elicit their productions. However, it turned out that this learning process itself was difficult for the learners. This point will be discussed in detail later in Chapter 6.

speakers. The overall F0 patterns of each production were also qualitatively analysed in order to make a comparison among the three groups, to investigate whether or not the overall F0 patterns in the productions of English-speaking learners diverge from those of the native speakers of Japanese. The three acoustic parameters used as the criteria were F0 peak alignment (the location in a word of the F0 peak), F0 fall (the degree of the slope) and F0 range (the width of the F0 contour).

These three acoustic parameters were selected on the following grounds. As we have seen earlier, previous studies suggest that the F0 peak alignment relative to the segmental landmarks is the key acoustic correlate of Japanese pitch accent (Ishihara, 2006; Sugito, 1982). We do not know whether English-speaking learners are able to produce the F0 contours with a peak at all, or alternatively, whether their contours may have several peaks instead of one. Even if they can produce F0 patterns with the appropriate peak, we do not know whether the learners can achieve the systematic F0 peak alignment relative to the segmental landmarks like native speakers do. We can only find out about this by observing the overall F0 patterns in the learners' productions. To observe the F0 peak alignment point of the speech imitations, the time-normalized F0 peak point relative to segmental landmarks was detected for each utterance.

Additionally, it has been said that the F0 peak of Japanese pitch accent is followed by a rapid fall in F0 (Hasgawa & Hata, 1988, 1992; Sugito 1980). We don't know whether we can find any such rapid F0 fall in the productions of native speakers of English, and if any, how rapid the F0 fall is. We also do not know whether or not the F0 contour patterns in English speakers' productions will show a greater falling movement than a rising movement or vice versa. Alternatively, there may not be any consistent patterns in the overall movement of F0. The investigation of the overall F0 contours in the learners' productions will provide the answers for these questions.

In addition, F0 range is another parameter which has been used to evaluate the shape of the F0 contours not only in Japanese but also in other languages (e.g., Beckman, 1986; Ladd, 1997). This measure can potentially add another useful dimension to the F0 profile as we compare the productions of learners with those of native speakers of Japanese. We do not know how wide/narrow the overall F0 contour will be in utterances produced by native speakers of English. Not having fully learned what to do or how to produce accurate F0 patterns for Japanese pitch accent, native speakers of English may over-exaggerate the rising or the falling of the F0 movement. On the other hand, it is also possible that they do not produce enough F0 movement compared to native speakers. By observing F0 range we will be able to answer these questions. To compare the shape of the F0 contours, the normalized F0 range was measured.

Considering these measures, the following questions are raised. Are Japanese productions of native speakers of English perceived as non-target-like (i.e., foreign accented) when native speakers of Japanese hear their productions? What will happen to the overall patterns of the three F0 parameters in the productions of native speakers of English when they try to produce a set of minimal pairs of Japanese pitch accent contrasts? Do the overall F0 shapes in the learners' productions diverge from those in the native speakers' productions regarding F0 peak location, F0 fall and F0 range? If the learners' F0 profiles are found to be different, how are they different from the native speakers' patterns? Does the pattern change during L2 acquisition as a function of L2 experience? To investigate these matters, an imitation task was carried out to obtain the production data.

Firstly, the details of the materials and the procedures will be described in the following section. Then, the details of the intelligibility scores and the acoustic observations of the overall F0 patterns will be explained. Lastly, the results of the imitation data will be reported and discussed.

### 3.1. Participants

There were 48 participants in total. None of them self-reported any speaking/hearing disabilities. The participants recruited for this study were 32 adult native speakers of English

(16 male, 16 female) who were studying Japanese at the University of Edinburgh at the time of experimentation. Based on their L2 experience (i.e., length of residence in Japan and length of studying Japanese (shown in Table 3.1, and details described below)), learners were divided into two groups: inexperienced (NEinexp) and experienced (NEexp) learners. In addition, 16 adult native speakers of Japanese (NJ) participated as a control group (8 male, 8 female). The mean age of NEinexp learners was 21.3 years[2] (*SD* = 5.9), that of NEexp learners was 21.4 years (*SD* = 0.8), and that of NJs was 29.6 years (*SD* = 5.9). The inexperienced learners included students who were towards the end of the second year of the degree course of the university and those who had been studying Japanese for at least one and a half years on average. Most of the NEinexp learners had never stayed in Japan. According to the questionnaire completed by the participants (see Appendix A), for three NEinexp learners who had stayed in Japan before, their stay was between 4 to 8 weeks either for holiday or teaching English rather than to have formal instruction in Japanese. The experienced learners, on the other hand, included students who were towards the end of the fourth year of the degree course at the same university, and who had been studying Japanese for approximately three and a half years on average. They had also stayed in Japan for about

[2] Excluding Participant 205, who was a mature student, the average age of NEinexp learners group was 19.9 years old.

one year (as part of the degree programme, learners are supposed to stay in Japan during the third year to study Japanese at a university of their choice). Hence, the NEinexp learners were not planning to leave for Japan until a few months after the experiments for their year abroad, while NEexp learners had already come back from their one year stay in Japan. The control group of native speakers of Japanese mostly consisted of either exchange students from Japanese universities or postgraduate students at Edinburgh University.[3] All of the participants were living in Edinburgh at the time of the experiments. For the data analysis, perception data from all 48 participants was used, but for the production analysis, a representative sub-set of the production data was used from each group (i.e., recordings from 8 participants were randomly selected from each group to act as the representatives of each group).

Table 3.1 provides detailed information for each English-speaking participant: chronological age, age when their first formal instruction of Japanese began, length of formal education of Japanese (LOJ), length of residence (LOR) in Japan, and self-estimated percentage daily use of Japanese. The length of residence (LOR) in Japan between inexperienced and experienced English speaking participants was statistically different ($t$(30)

---

[3] One NJ was a post-doctoral fellow, one had just graduated from a postgraduate programme and one had been staying in Edinburgh with her family for one and a half years.

= 74.68, $p < 0.001$). The experienced learners had spent longer in Japan than the inexperienced learners as a group ($M = 0.03$ years, $SD = 0.05$ for NEinexp vs. $M = 1.00$ years, $SD = 0.00$ for NEexp). The length of formal education of Japanese (LOJ) between two English speaking learner groups was also significantly different ($t(30) = 9.99$, $p < 0.01$). The experienced learners had studied Japanese longer than the inexperienced learners as a group ($M = 2.01$ years, $SD = 0.60$ for NEinexp vs. $M = 3.66$ years, $SD = 0.27$ for NEexp). On the other hand, the self-estimated daily use of Japanese at the time of the data collection tended to be similar between the two groups ($M = 12.94\%$, $SD = 10.80$ for NEinexp vs. $M = 18.06\%$, $SD = 22.90$ for NEexp), although many of NEexp learners informally reported after the experiments that their daily use of Japanese during the stay in Japan was much higher then than in general.

**Table 3.1 Characteristics of English speaking participants (NEinexp and NEexp learners)**

| L2 experience | Participant | Age | Age of starting Japanese | Res. in Japan[a] | Length of study[b] | % Use[c] |
|---|---|---|---|---|---|---|
| NEinexp | 2 | 19 | 15 | 4w | 2.7y(1y evening class) | 10 |
| | 3 | 19 | 17 | None | 1.7y | 10 |
| | 4 | 19 | 17 | None | 1.7y | 5 |
| | 5 | 42 | 41 | None | 1.7y | 15 |
| | 6 | 19 | 17 | None | 2.7y(1y evening class) | 15 |
| | 8 | 20 | 16 | None | 3.7y(2y grammar school) | 10 |
| | 9 | 20 | 18 | None | 1.7y | 15 |
| | 10 | 27 | 25 | 5w | 1.7y | 10 |
| | 11 | 19 | 18 | 8w | 1.7y | 10 |
| | 12 | 20 | 17 | None | 2.7y (1y evening class) | 5 |
| | 13 | 20 | 17 | None | 1.7y | 2 |
| | 14 | 20 | 18 | None | 1.7y | 10 |
| | 15 | 20 | 17 | 6w | 1.7y | 50 |
| | 16 | 19 | 18 | None | 1.7y | 20 |
| | 17 | 19 | 18 | None | 1.7y | 10 |
| | 18 | 19 | 18 | None | 1.7y | 10 |
| NEexp | 1 | 22 | 19 | 1y | 3.7y | 2 |
| | 2 | 21 | 18 | 1y | 3.7y | 5 |
| | 3 | 21 | 18 | 1y | 3.7y | 0 |
| | 4 | 23 | 19 | 1y | 3.7y | 15 |
| | 5 | 21 | 18 | 1y | 3.7y | 85 |
| | 6 | 21 | 18 | 1y | 3.7y | 2 |
| | 9 | 21 | 18 | 1y | 3.7y | 10 |
| | 10 | 21 | 18 | 1y | 3.7y | 20 |
| | 11 | 20 | 17 | 1y | 3.7y | 60 |
| | 13 | 21 | 18 | 1y | 3.7y | 5 |
| | 14 | 22 | 18 | 1y | 3.7y | 20 |
| | 15 | 22 | 18 | 1y | 3.7y | 5 |
| | 16 | 22 | 19 | 1y | 3.7y | 10 |
| | 17 | 22 | 19 | 1y | 4.0y | 25 |
| | 18 | 22 | 18 | 1y | 3.7y | 15 |
| | 21 | 20 | 18 | 1y | 2.7y | 10 |

[a] Res. in Japan: Length of residence in Japan, in weeks (w) and in years (y).

[b] Length of Study: Length of studying Japanese under formal instruction, in years (y).

[c] % Use: Self-estimated percentage daily use of Japanese.

### 3.2. Materials

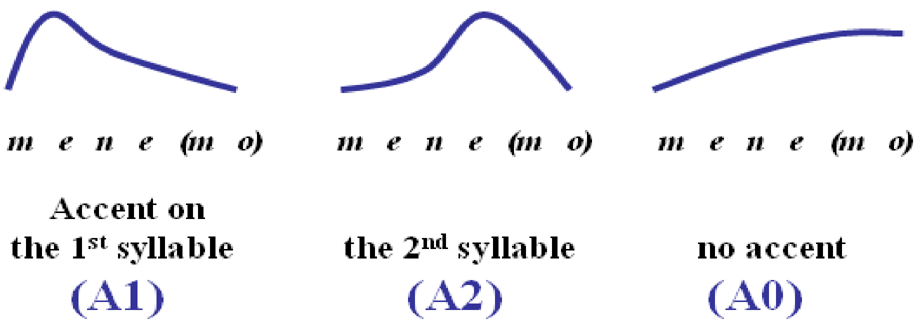
Audio materials for Experiment 1 were created from the recording of 12 disyllabic nonce words (4 words (*mene, mani*, *nime* & *noma*) x 3 accent types), each followed by a post-positional particle *mo* (also), which were produced by a female Tokyo Japanese. Disyllabic words in Japanese are lexically assigned one of three possible pitch accent types: accent on the first syllable (A1), accent on the second syllable (A2) or unaccented (A0). Hence, each test item in the experiments was prepared in such a way that each word would carry one of the three accent types. The test items consisted of sets of triplets of four disyllabic nonce words in CVCV structure, as shown in Table 3.2. There were 3 test items for each of the nonce words, giving 12 items in total. Nonce words were used rather than real words in order to avoid possible effects of word familiarity among the L2 learners, as more experienced learners are assumed to have larger L2 vocabularies than inexperienced learners.

Additionally, the test items were designed to consist only of vowels and nasals, so that F0 contours could be observed relatively easily with less variation in the F0 trajectory when participants' production data is analyzed in the imitation tasks. Due to their characteristics, accent on the second syllable (A2) and unaccented type (A0) can be distinguished when they are accompanied by a post-positional particle as the difference between their F0 contours are realized with a following sound (see Diagram 3.1). For this reason, a Japanese particle '*mo* (also)' was attached after each test item and comprised target stimuli.

**Table 3.2 Test items (A1 indicates accent on the first syllable, A2 indicates accent on the second syllable and A0 indicates unaccented)**

| | A1 | A2 | None (A0) |
|---|---|---|---|
| *mene* | *mene* | *mene* | *mene* |
| *noma* | *noma* | *noma* | *noma* |
| *mani* | *mani* | *mani* | *mani* |
| *nime* | *nime* | *nime* | *nime* |

**Diagram 3.1 Schematic illustration of three accent types for a disyllabic word using *mene(mo)***
Each line indicates the F0 contour of the word. The left: a word with pitch accent on the first syllable. The middle: a word with pitch accent on the first syllable. The right: an unaccented word.




Audio materials were used in all experiments. For the audio stimuli, the productions of a female native speaker of Tokyo Japanese were recorded in a sound-treated recording studio at the Linguistics laboratory at the University of Edinburgh. The audio interface was MOTU 828 mkll Firewire using a hypercardoid microphone (AKG CK 98) and digitalized at a sampling rate of 48 kHz with a resolution of 16 bits. Sonar 4 Studio Edition was used as a recording software programme. A list of the 12 test items was read aloud in a carrier phrase (i.e., "*Sumimasen* ____*mo kudasai.* (Excuse me but please give me ____ as well.)") three times. Subsequently, one of three repetitions of each item including the particle *mo* 'also' was edited to form audio stimuli.

In addition to these test items, 30 basic Japanese words were carefully selected for a warm-up before the participants attempted the main recording (see Table 3.3). Basic words were selected from Japanese textbooks which the learners were using in the course (ICU, 1996; Kano, Shimizu, Takenaka & Ishii, 1999). The warm-up words were also presented in Japanese orthography, *Hiragana*, one at a time on a computer monitor along with the Romanized equivalent and its meaning in English. All the audio materials were converted to a sampling rate of 22 kHz with a resolution of 16 bits using CoolEdit Pro 2.1 so that it matched the Sound Device Object Property in E-Prime.

**Table 3.3 Warm-up items (30 words)**

| 1st syllable (A1) | 2nd syllable (A2) | Unaccented (A0) |
|---|---|---|
| *asa* (morning) | *hiru* (noon) | *ocha* (tea) |
| *ame* (rain) | *heya* (room) | *hima* (free) |
| *chichi* (father) | *hana* (flower) | *isu* (chair) |
| *haha* (mother) | *inu* (dog) | *kuni* (country) |
| *ima* (now) | *mise* (shop) | *michi* (road) |
| *jisho* (dictionary) | *netsu* (fever) | *migi* (right) |
| *kasa* (umbrella) | *uta* (song) | *mizu* (water) |
| *mado* (window) | *yama* (mountain) | *sake* (sake) |
| *mae* (in front of) | *yuki* (snow) | *uchi* (house) |
| *neko* (cat) | *kagi* (key) | *ue* (above) |

### 3.3. Procedure

Participants were tested individually in a sound-treated recording studio. The participants heard two words with minimally different pitch accent types (i.e., the three accent types) in a sequence through a pair of headphones. They were asked to imitate the first speech stimulus as closely as possible. Two audio stimuli were presented so that this task would have a parallel structure to the other perception tasks which will be reported later in this study, i.e., a non-speech ABX task in Experiment 2 and the speech ABX tasks in Experiment 4. They were allowed to make more than one attempt to imitate the target stimulus if they were not satisfied with their first attempt. If they did make more than one attempt, the final imitation was used for the analysis. Thirty-six pairs of recorded stimuli were randomly presented over headphones (4 disyllabic nonce words x 3 accent types x 3 presentation orders (AB, BA, and either AA or BB)).

The audio interface was MOTU 828 mkll Firewire using a hypercardoid microphone (AKG CK 98) and digitized at a sampling rate of 48 kHz with resolution of 16 bits. Sonar 4 Studio Edition was used as a recording software programme. Participants listened to the audio materials using a pair of Sennheiser eH2270 headphones. A computer keyboard was used to proceed to the next stimulus by pressing the return key. All the instructions were presented with E-Prime using a computer monitor facing the participants in the recording booth.

Participants were given a practice session prior to the main test, in order for them to familiarize themselves with imitating the stimuli and to allow them more time to ask any questions. The stimuli for the practice session were disyllabic nonce words which were similar to the target stimuli.

Every time the participants produced their imitation of the first stimulus, they were asked to press the space bar on the keyboard in order to proceed to the next pair to be presented. This allowed the participants to proceed through the task at their own pace.

## 3.4. Analysis

The recording of each participant was digitized at a sampling rate of 48 kHz with resolution of 16 bits. The recording of each participant was segmented into smaller sound files to analyze each token of the target stimuli. This yielded 36 sound files per participant, each of which contained one of 36 imitations of the target items and lasting approximately 1.0 sec.

To assess the participants' production data, two types of analysis were carried out. One used intelligibility scores based on native speakers' auditory judgments on each utterance. The other used descriptive analysis of the overall F0 contours of each utterance: more specifically, the location of the F0 peak, the F0 fall and the F0 range for each utterance were qualitatively analyzed based on the overall F0 shape of the contours.

In the following subsections, first, we will explain the details of the intelligibility scores. Next, we will describe the segmentation criteria which we used in the analysis to capture the phonetic patterns in the productions.

### 3.4.1. Intelligibility scores

The intelligibility scores are the accuracy rate (in percentages) as to how many of the imitations of Japanese pitch accent were identified by native speakers of Japanese as being what the speaker intended. This is the key measurement to assess learners' foreign accentedness. Each utterance of the test items was identified by two raters who were female native speakers of Japanese. They were asked to identify which accent type a speaker was producing in each utterance and their ratings were provided with multiple-choice among the three accent types (A0, A1 and A2). They were also requested to guess if unsure (a forced choice paradigm). The entire data set was rated by these two raters. The mean percentage of times that Japanese pitch accent produced by the participants were identified as intended was calculated as the intelligibility score. As it turned out to be common that the learners' productions were misidentified by the two raters, the misidentification patterns were also classified into groups for further analysis.

### 3.4.2. Segmentation criteria

The segmentation criteria used in the analyses were those proposed by Turk, Nakai and Sugahara (2006). The spectrogram setting in PRAAT (Boersma & Weenink, 2005) was set with 'view range' of 0 – 8000 (Hz) and 'dynamic range' of 55 (dB). Approximately a 1-sec window was used for placing general boundaries and the waveform was zoomed into 5-10 ms for more fine-grained segmentation. The segmentation points were located at the zero-crossing point of the waveforms.

The segmentation was conducted manually by looking at the visual display of the spectrogram and the oscillogram, and was based on the criteria below. With the aid of a

PRAAT script written by Xu (2008) with modifications of the algorithm by Dabrunz (2009)[4], the overall F0 patterns, which are described below, were obtained after segmenting each audio file.

The segmental boundary was placed based on spectral characteristics which can be easily detected in spectrograms. All the imitations of the speech stimuli consisted of a test item (a disyllabic nonce word) accompanied by a postpositional particle *mo* (also). This yields a CVCVCV structure consisting of nasals and vowels to be segmented. The onset of the initial nasal was marked where the closure occurred or the regular cycle of voicing started. When irregular cycles were observed in the beginning of an utterance, these parts were excluded from the nasal onset to be analyzed (Figure 3.1). The reason behind this is that these parts often tend to cause errors in depicting the F0 trajectory of the data and finding the velocity values, resulting in extremely large or small values in the subsequent acoustic analysis. The utterance-medial nasals were segmented where the closure occurs as the onset of the nasal and the release as the offset. The utterance-medial vowels were marked at the beginning of a clear formant pattern of F1 and F2 as the onset and at the end of a clear formant pattern as the offset. The offset of the last vowel of each utterance, i.e., /o/, tended to end in creaky phonation (i.e., where glottal pulses are spaced irregularly with a

[4] The modification is provided in Appendix B.

very low F0). When the utterance ended in creaky phonation, the beginning of a clear glottal cycle of the creaky phonation on the waveform was detected and the boundary was marked on its zero-crossing point (Figure 3.2). The first clear glottal cycle was chosen rather than the last because including the glottal cycles between the first and the last also tends to cause errors in calculating the F0 values. When the offset of /o/ did not end in creaky phonation, the boundary was placed on the zero-crossing point of the glottal cycle of the waveform which corresponded to the end of strong F2 (Figure 3.3). Some of the utterances also ended in breathy phonation, (i.e., where voiceless formant structures could be observed directly after the vowel formant structure). In these cases, the offset of /o/ was marked at the end of voicing on the waveform and the voiceless formant structure was excluded (Figure 3.4).

**Figure 3.1. Onset of the initial nasal**

Four PRAAT Text Grid tiers were used for the annotation: 1 word tier, 2 segmental tiers and 1 landmark tier. The word tier was used to annotate the target word (with an accent type and the particle *mo*) (e.g. *mene1mo*). On the second tier, the C1V1C2V2 string was marked to indicate the segmental structure such as C1 for the first consonant or V2 for the second vowel. On the third tier, the string of the corresponding segmental composition was marked, along with a number to indicate the syllable number, e.g., *m*1*e*1*n*2*e*2*mo*. The fourth tier was used to mark the point where creaky phonation or breathy phonation started (creaky voice was marked as "c" and breathy voice was marked as "v").

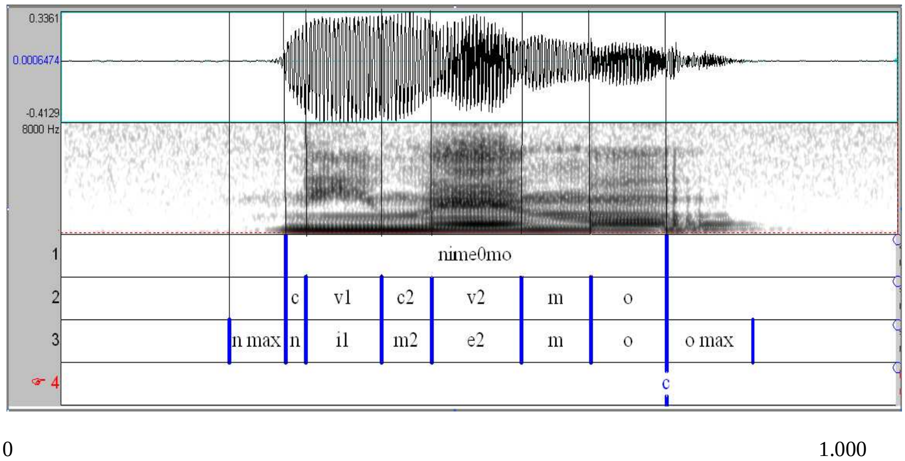



**Figure 3.2 Creaky phonation at the end of utterance /o/**

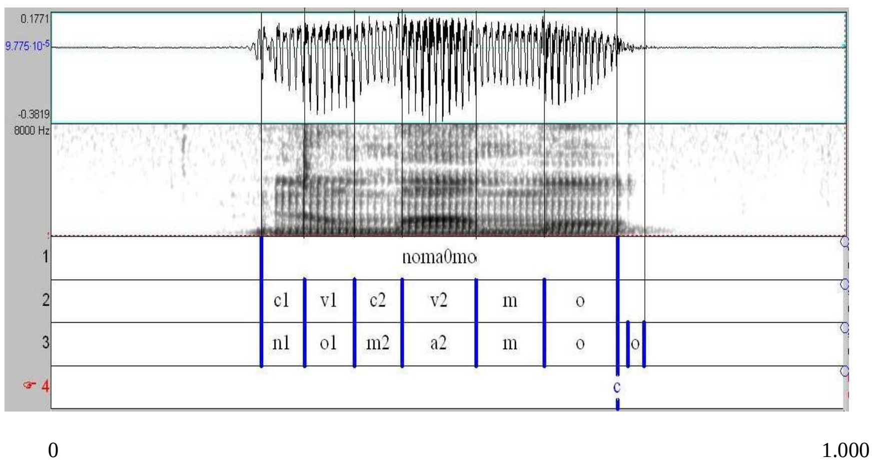


**Figure 3.3 Example of utterance final /o/ without creaky phonation**

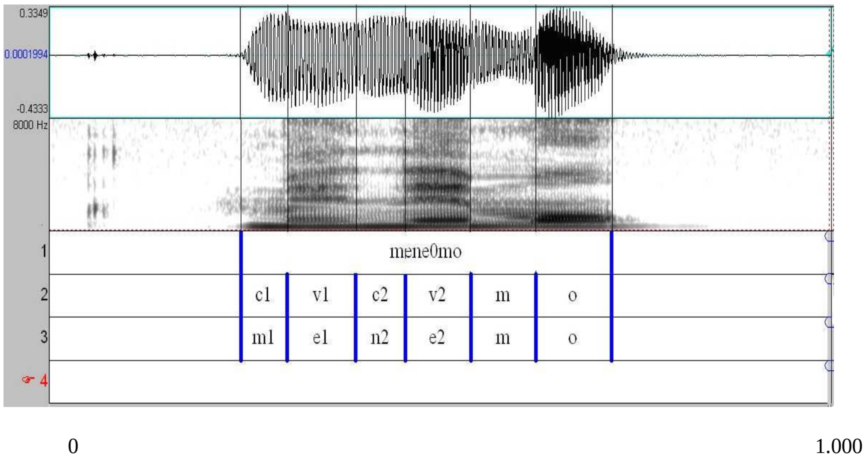
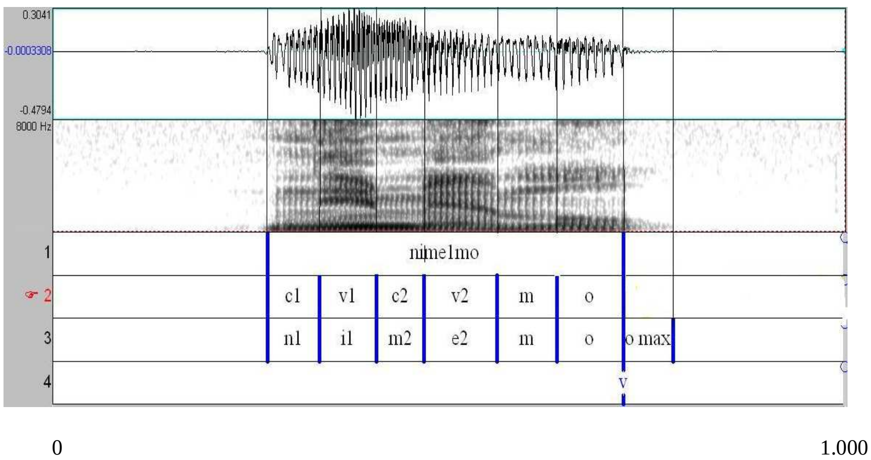



**Figure 3.4 Breathy phonation at the end of utterance /o/**




### 3.4.3. Overall F0 patterns

Having observed that the learners' productions of Japanese lexical pitch accent indicated foreign accentedness, the overall F0 patterns in their productions were also analyzed to investigate whether their F0 patterns also show non-target-like patterns. To qualitatively analyze the overall F0 patterns, we focus on the overall patterns, such as the shape of the contour where the F0 peak located, whether the F0 contour contains a sharp fall after the peak, or how wide the F0 contour is.

The F0 peak alignment value identifies where the F0 peak is located in relation to the segmental landmarks in each speech imitation. In order to observe the F0 peak location relative to the segments, smoothed F0 values in each segmental interval were divided into 10 time-normalized points. The utterances in the speech imitation task each consisted of six segments, which yield six corresponding intervals (see Table 3.4). Thus, there were 60 points for each utterance. The location of the highest F0 in the overall F0 contours among these 60 points was described as the F0 peak point. This allows us to make comparison between the locations of F0 in the speech imitations and these points were used as the reference points to describe where the F0 occurred. If the F0 peak is located at a point between the 1st point and the 20th point, this means that the F0 peak aligned within the first syllable of the target disyllabic word. If the F0 peak occurs at a point between the 21st and 40th point, this indicates that it aligned within the second syllable. Similarly, if the F0 peak is located at a point after the 41st point and up to the last point, it aligned within the last syllable (the interval which corresponds to the particle *mo*). To detect the time-normalized

F0 values, Xu (2008)'s PRAAT script was used which automatically saves the trimmed F0 values.

**Table 3.4. Schematic illustration of the location of the F0 peak alignment**

| Structure | **C1** | **V1** | **C2** | **V2** | **C3** | **V3** |
|---|---|---|---|---|---|---|
| Segment | **m** | **E** | **n** | **E** | **M** | **o** |
| Time-normalized point | 1-----10 | 11---20 | 21---30 | 31---40 | 41---50 | 51---60 |
| | 10 | 10 | 10 | 10 | 10 | 10 |

What we expect to see in the native Japanese speakers' productions in terms of the F0 peak alignment is as follows. In words with the accent on the initial syllable (A1), we expect to see the F0 peak occurring early in the utterance and aligned with or in the vicinity of the first syllable. For words with the accent on the second syllable (A2), the F0 peak is expected to occur later in the utterance and aligned with or in the vicinity of the second syllable. In the case of the unaccented words (A0), the F0 maximum value was considered to be the F0 peak value of A0 productions, as no sharp F0 fall is expected in the productions of A0 (unlike the accented types, i.e., A1 and A2). Meanwhile, in the learners' productions, it is possible that there may be no consistent F0 peak regarding the accent types, there may be more than one F0 peak, or the location of the F0 peak may occur unsystematically.

Next, we describe the F0 fall of the contours in their productions. As F0 patterns in the productions of native speakers of Japanese tend to show a rapid fall after the F0 peak, what we are interested in here is the degree of falling movement after the F0 peak. Thus, we looked at the F0 fall patterns after the F0 peak in the overall shape. More specifically, as native speakers of Japanese tend to produce a rapid fall after the F0 peak, we expect to see the largest fall in an utterance of the accented words (A1 and A2). On the other hand, we do not expect to see such a large fall in the native speakers' productions of unaccented words (A0) as unaccented words are not expected to show a sharp fall unlike accented words. In the case of the learners, we may not see a rapid fall at all in their productions as in the native speakers' productions. It is also possible that we may see a rather extreme fall instead, even for the unaccented words.

Thirdly, to capture the overall F0 shape in the learners' productions, F0 range (i.e., the general width of the F0 contours) is compared with native speakers. To make comparisons among speakers possible, the F0 values (Hz) on the time-normalized scale were converted into normalized F0 values. In this study, Z-score transforms were used to normalize the F0 values of each contour. Two normalization methods have been used in the literature: fraction of range transforms (e.g., Earle, 1975) and Z-score transforms (e.g., Rose, 1987). The method of fraction of range transforms is a way of defining the F0 value relative to the highest and lowest F0 points of each speaker. Using Z-score transforms allows us to describe each data point based on the statistical distribution of each utterance of an individual speaker. Although between-speaker variance is considerably reduced by both strategies, Rose (1987) concluded that the Z-score transform is more accurate than the fraction of range transform for normalizing linguistic tone data. For this reason, we considered that the strategy of Z-score transforms was sufficient for the purpose of this study. The formula described in Rose (1987) was used: $F_{norm} = (F0 - F0_{mean})/s$, where $s$ indicates one standard deviation of the mean F0 of all data sets produced by each speaker (i.e., $F0_{mean}$). In the productions of the native speakers of Japanese, we expect to see a relatively large F0 range for the accented words (i.e., A1 and A2) as words with pitch accent tend to have a sharp fall after the F0 peak in Japanese. Meanwhile, the unaccented words (i.e., A0) will presumably have a smaller F0 range without such a steep drop as part of the overall F0 shape. In the case of learners, the overall F0 range could be relatively wide, or very flat regardless of the accent types.[5]

---

[5] To further investigate the characteristics of the production data, attempts to quantify the three acoustic parameters (i.e., F0 peak location, degree of F0 fall, and F0 range) were made. However, when the data was smoothed in order to allow comparisons across the large number of speakers studied here, some unresolved artifacts or distortions were introduced into the data. Since these artifacts require further investigation to be resolved, the results are not presented here. Nonetheless, it is still possible that such details of attempts can be a useful contribution to the field. For this reason, the difficulties including the details of the attempted data analyses are illustrated in Appendix C.

### 3.5. Results

Each of the four data sets (*mene, noma, mani* and *nime*) showed similar overall F0 contour shapes within each group. For this reason, the *mene* set was randomly selected from the four sets as the representative subset for acoustically analyzing the participants' production patterns in terms of the selected acoustic measures. It should be noted that one Japanese speaker's *mene* (A1) production was excluded from the analysis because it contained a great deal of creaky voice in the middle of the utterance, which greatly distorted the F0 contour shape. In addition, there were three cases where a NEinexp speaker altered the segmental make-up to a different one such as */meni/*. These three cases were also excluded from further analyses. This resulted in a dataset consisting of 71 items produced by the 8 speakers in NJ group; 72 items produced by the 8 speakers in NEexp group; and 69 items produced by the 8 speakers in NEinexp group.

In what follows, we will first describe the mean intelligibility scores to show the learners' foreign accentedness and report misidentification types among each group using the overall F0 patterns to make comparisons between the native speakers and the learner groups. Secondly, we will present the overall F0 patterns of each participant to view general patterns of each group in terms of the three accent type to examine whether the overall shape of their

F0 contours also shows non-target-like phonetic patterns.

### 3.5.1. Intelligibility scores

First, the mean intelligibility scores are described. Then, misidentification patterns are analyzed using the overall F0 patterns.

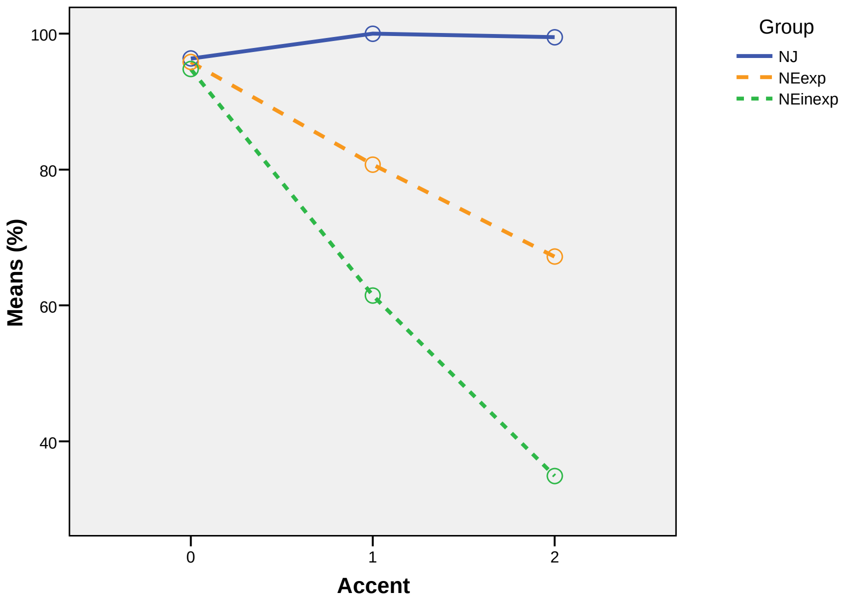
#### 3.5.1.1. Mean Intelligibility scores

As described in 3.4.1, there were two raters. They were both native speakers of Japanese. The inter-rater reliability was 0.75 (Spearman's correlation), which was acceptable. The means and the standard deviations of intelligibility scores obtained for the Japanese pitch accent contrasts are presented in Table 3.5.

**Table 3.5. The mean percentage of times that Japanese pitch accent produced by the participants was identified as intended by native judges.**

| Accent | Group | Mean | Std. Deviation |
|---|---|---|---|
| A0 | NJ | 96 | .18 |
| | NEexp | 96 | .20 |
| | NEinexp | 95 | .20 |
| A1 | NJ | 100 | .00 |
| | NEexp | 81 | .39 |
| | NEinexp | 61 | .47 |
| A2 | NJ | 99 | .05 |
| | NEexp | 67 | .45 |
| | NEinexp | 35 | .46 |

**Figure 3.5. The mean intelligibility scores of the speech productions**




These intelligibility scores were submitted to a mixed design ANOVA with Accent Type (A1, A2 and A0) as a within-subject factor and Group (NJ, NEexp and NEinexp) as a between-subject factor. There was a significant main effect of Accent Type [$F$ (2, 564) = 78.42, $p < 0.001$]. As we can see in Figure 3.5, every group obtained high intelligibility scores for A0 (nearly 100%). Meanwhile, A1 showed the next highest scores, and A2 was the lowest among the three accent types. There was also a significant interaction between Accent Type and Group [$F$ (4, 564) = 32.31, $p < 0.001$]. The overall performance of the three groups shows different patterns across the three accent types, as can be seen in Figure 3.5. Moreover, the main effect of Group was also found to be significant [$F$ (2, 282) = 67.71, $p < 0.001$]. While NJ's scores were almost perfect across the three accent types, post-hoc comparisons using the Tukey HSD test indicated that the mean intelligibility scores of NEexp were significantly lower than that of NJ. Moreover, the mean intelligibility scores of NEinexp were statistically even lower than NEexp. This tendency was also supported by the results of one-way ANOVAs. These tests were conducted to compare the three groups regarding each accent type. These results showed that there were no significant group differences in A0. However, the mean intelligibility scores of all three groups were significantly different from one another for A1 [$F$ (2, 285) = 28.68, $p < 0.001$] and A2 [$F$ (2, 285) = 71.83, $p < 0.001$].

#### 3.5.1.2. Patterns of misidentification

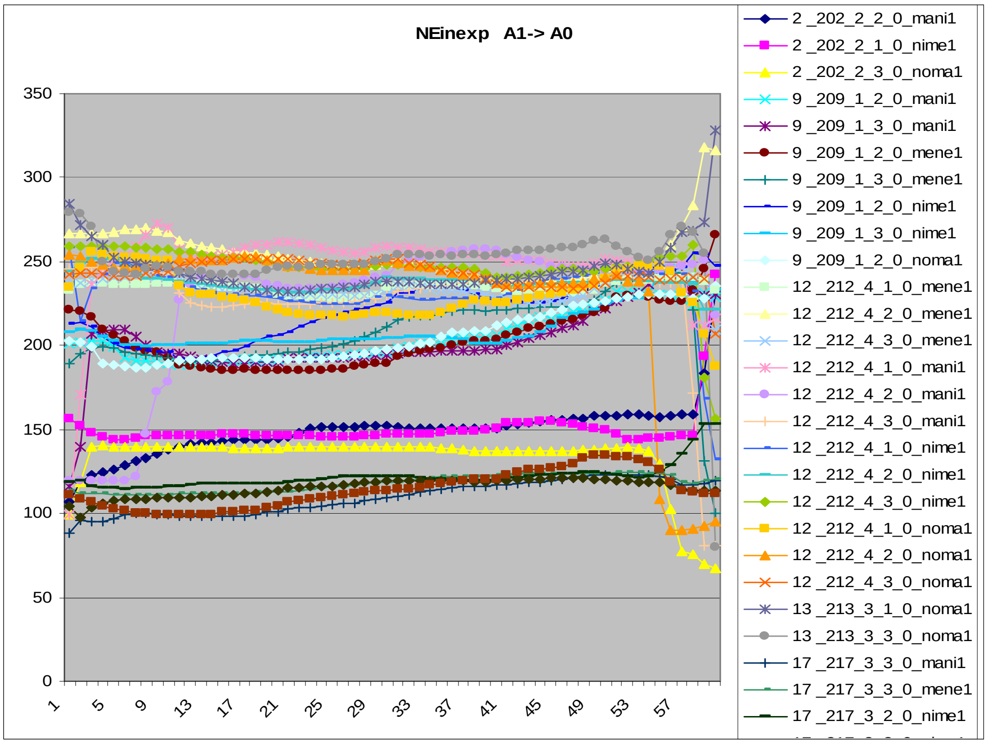
As we just saw, the lexical pitch accent produced by native speakers of Japanese was largely identified as intended, while those produced by the learners tended to be misidentified more often, particularly in the case of the pitch accented targets. To investigate further what cued the two native listeners to make their decisions, let us now look at the misidentification patterns observed in the intelligibility scores. In what follows, we will describe the classifications of the misidentification patterns, and in addition, we will demonstrate the misidentification patterns by examining the F0 contour shapes in the learners' productions.

There were in total 98 cases for NEinexp speakers and 55 cases for NEexp speakers which were misidentified by the two raters. For both learner groups, the major misidentification type was the accented targets, regardless of whether the target was A1 or A2. Both were identified by the raters as the unaccented targets, i.e., A0. This was observed in 70 out of 98 cases for NEinexp speakers and 40 out of 55 cases for NEexp speakers. When we look at the F0 patterns of these utterances in Figure 3.6 and

Figure 3.7, we see interesting patterns. It should be noted that the F0 values are presented in Hz but not as normalized values before smoothing. There is some similarity to native speakers' A0 patterns observed in the next subsection (Figure 3.1 in 3.5.2). The F0 contour shape in these misidentified utterances tends to be flat and lacking the target sequence of peak followed by fall. This suggests that the perception of the two native speakers might have been affected by these acoustic characteristics of F0 in the learners' productions when they were identifying the accent type of each utterance.

**Figure 3.6. Misidentification patterns, where NEinexp_A1 (top) & A2 (bottom) were identified as A0.**

Each line represents an utterance of a NEinexp speaker which was misidentified. The x-axis indicates the time-normalized points. The y-axis indicates the F0 value (Hz).



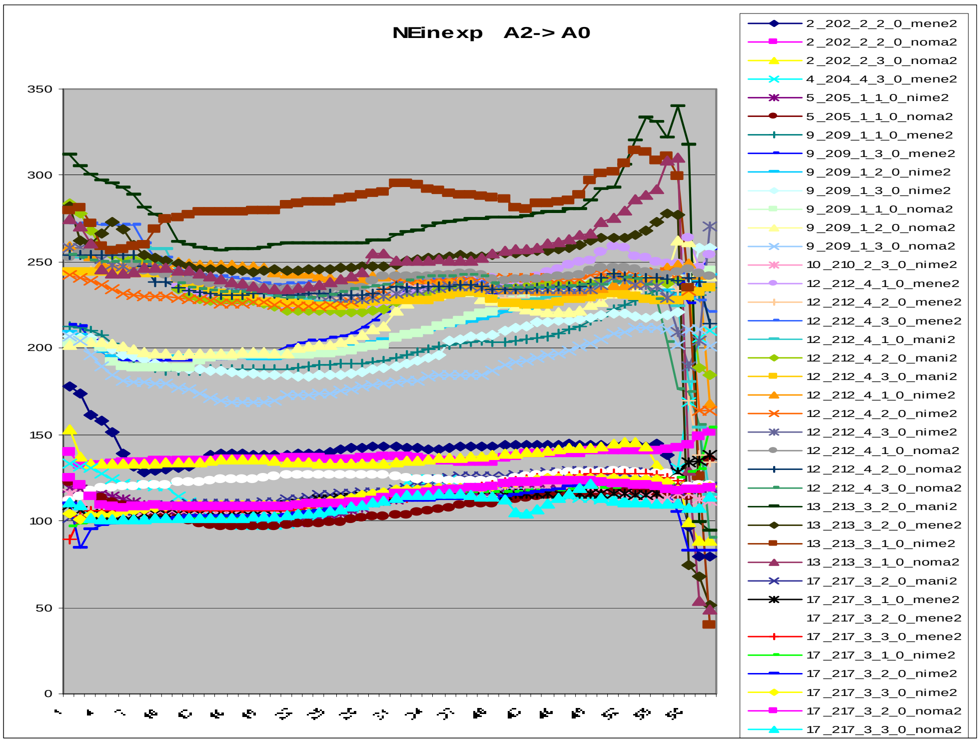
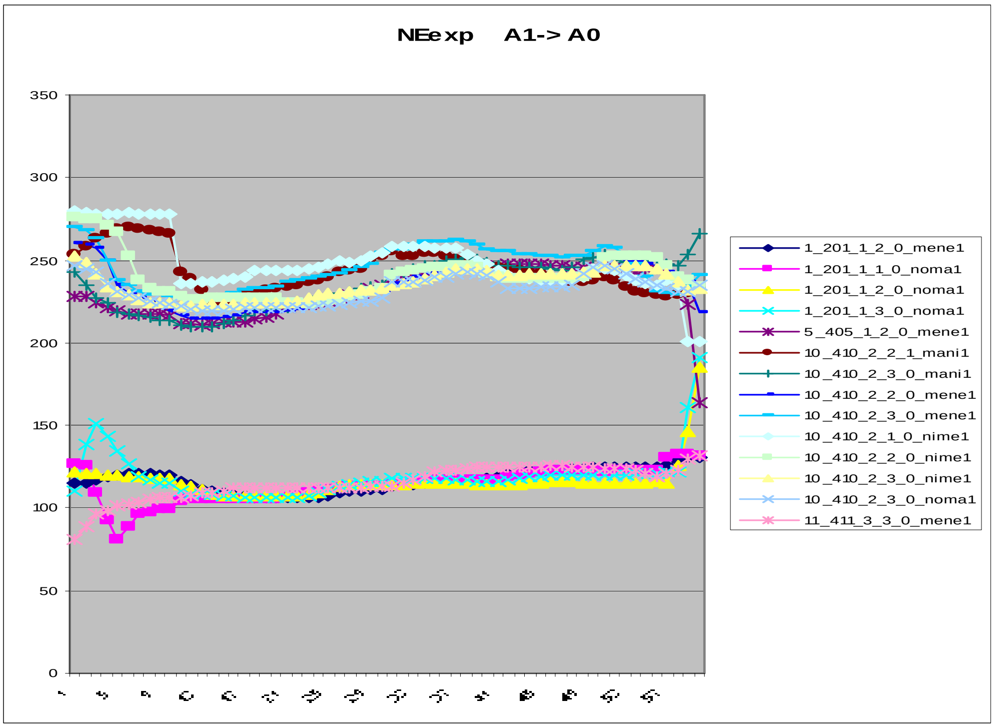
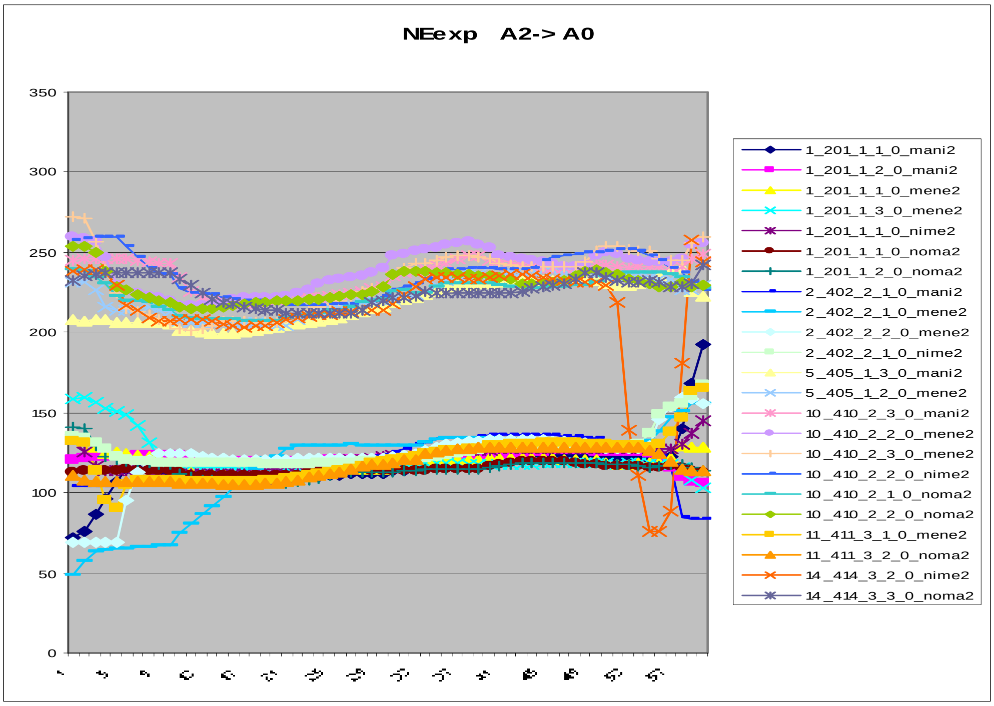



**Figure 3.7. Misidentification patterns, where NEexp_A1 (top) & A2 (bottom) were identified as A0.**

Each line represents an utterance of a NEexp speaker which was misidentified. The x-axis indicates the time-normalized points. The y-axis indicates the F0 value (Hz).



The next most common misidentification type for both learner groups was that A2 accent was misidentified as A1. There were 22 out of 98 cases of this for NEinexp speakers and 8 out of 55 cases for NEexp speakers. As we see in Figure 3.8 and Figure 3.9, the F0 peak in these misidentified utterances tended to occur earlier in the utterances, although not in every instance. This pattern resembles the F0 pattern of A1 produced by native speakers, i.e., a pattern where the F0 peak is aligned with the accent bearing unit, i.e., in the vicinity of the first vowel.

**Figure 3.8. Misidentification patterns of NEinexp_A2 as A1.**

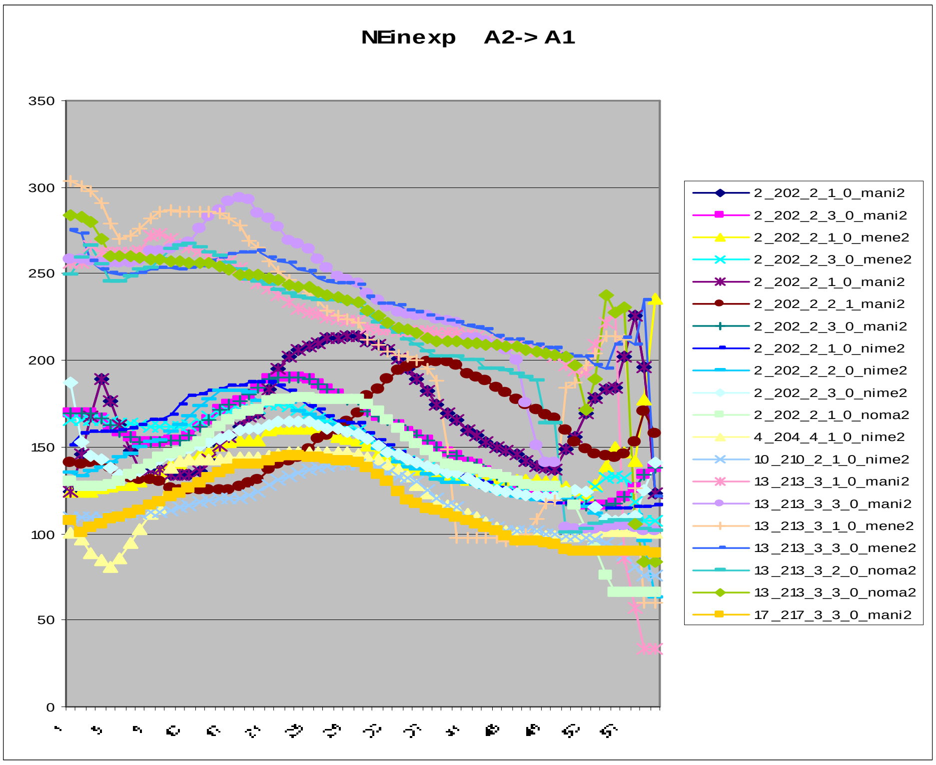
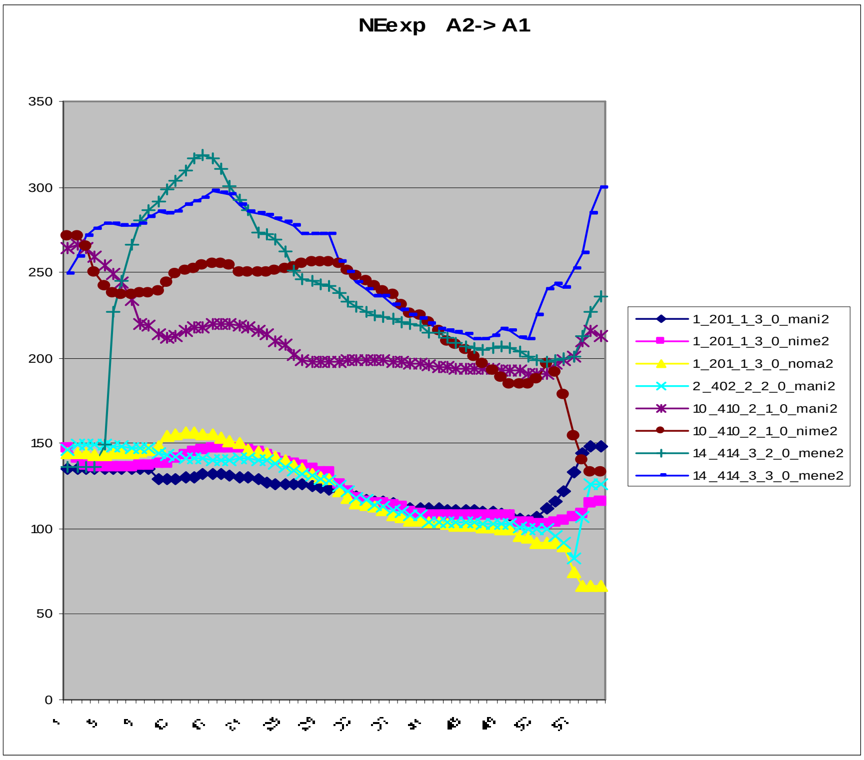


Figure 3.9. Misidentification patterns of NEexp_A2 as A1.




The remaining misidentification type was when A1 accent was misidentified as A2. There were a few cases of this in each learner group, as shown in Figure 3.10 and Figure 3.11. This time the F0 peak tends to occur later in the utterances. This pattern also resembles the F0 pattern of A2 as produced by native speakers, which shows the F0 peak aligned with the accent bearing unit, i.e., in the vicinity of the second vowel.

Figure 3.10. Misidentification pattern of NEinexp_A1 as A2.

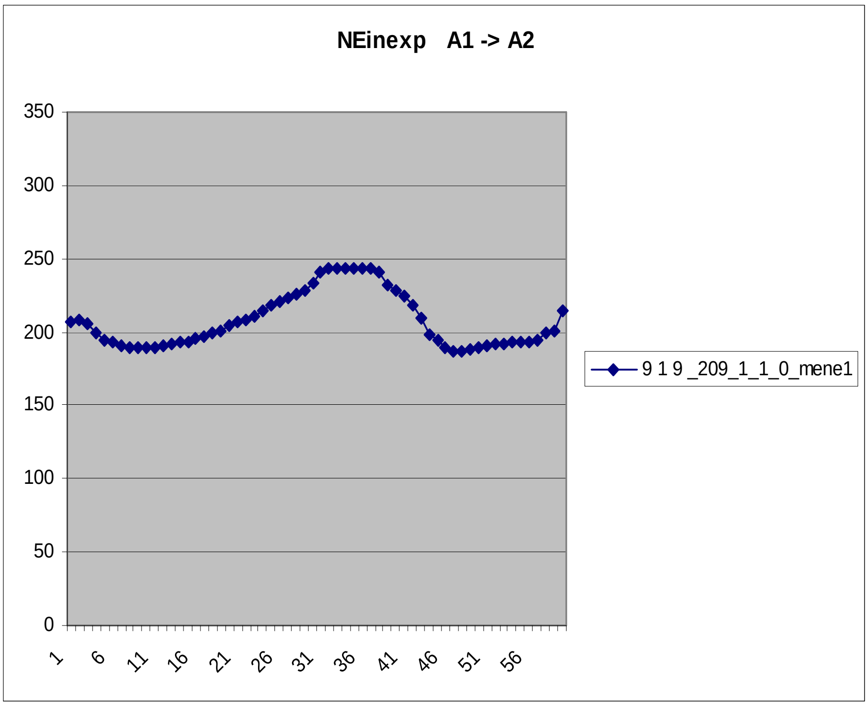
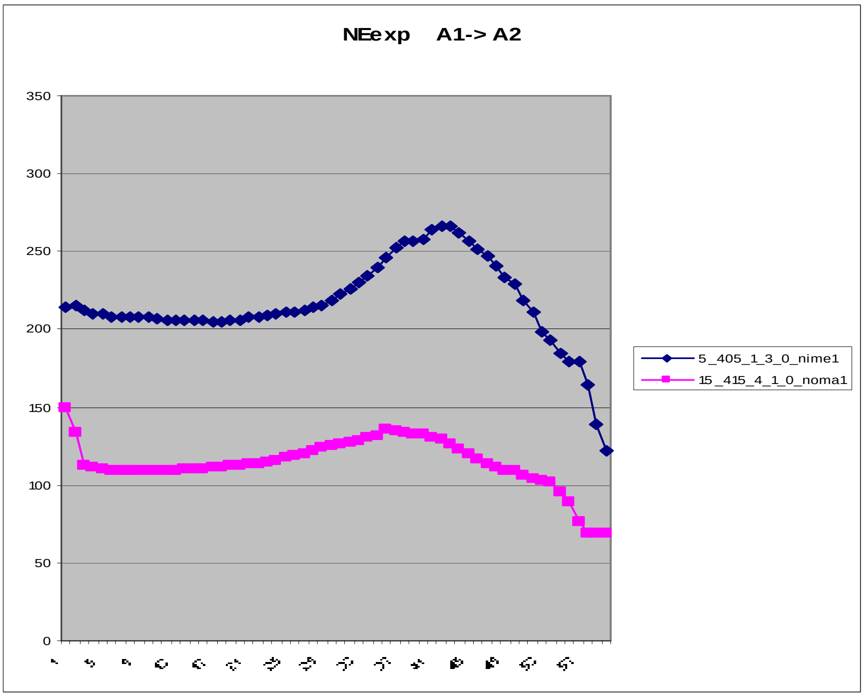



Figure 3.11. Misidentification patterns of NEexp_A1 as A2.



To summarize, having observed these patterns, what the two native listeners of Japanese were hearing in the learners' productions seems to be something which is not random but which seems to be related to some systematic acoustic patterns of the F0. More specifically, the two raters were using the information of the F0 peak location most in their decision. This further indicates that the learners did not always have control over this aspect of F0 in their productions. Hence, it is likely that this influenced the native speakers' decision on the non-target-likeness of the learners' productions. Moreover, the frequency of the misidentification pattern where the accented targets are identified as unaccented suggests that producing the F0 contour shape of the A0 accent types might have been easier for the learners compared to the accented types. Therefore, the learners might have been using the A0 pattern to mark the other two accent types because they were struggling to produce the accented patterns. To investigate in more detail as to what the participants were producing in terms of the relevant acoustic correlates of Japanese lexical pitch accent, and to compare the learners' productions with the native speakers' productions, now let us move on the results of the overall F0 patterns.

### 3.5.2. Overall F0 patterns

The following figures represent the overall F0 patterns of each participant, averaged over three repetitions of each test item. Considering the speakers' variability, the F0 values (Hz) on the time-normalized scale were converted into normalized F0 values. In this study, as discussed above, Z-score transforms were used for normalizing the F0 values of each contour. Figure 3.12 represents the F0 contours in the productions of native speakers of Japanese for the accent on the 1st syllable (A1), the accent on the 2nd syllable (A2) and the unaccented utterances (A0). Each line indicates the average F0 contour of a speaker. The F0 values in each figure are not smoothed values. By way of reminder, each segmental interval was divided into 10 time-normalized points, and thus, there were 60 points for each utterance, as shown on the x-axis. As an overall tendency, the F0 patterns of A1 produced by NJ tend to have the peak followed by a fall in the first half of the plot, which corresponds to somewhere between the first vowel and the second consonant. For A2, NJ's F0 patterns show that the peak followed by a fall tends to be located in the second half of the plot, which corresponds to somewhere between the second vowel and the third consonant. On the other hand, we do not see such a peak followed by a sharp fall in NJ's F0 patterns for A0. These overall tendencies agree with the previous findings on the productions of Japanese speakers. As for the degree of the F0 fall after the peak, in the NJ's productions, the degree of fall for the accented types (A1 and A2) seems to be observable and it seems that it is sharper for A2 than for A1. Lastly, the range of the F0 contours for A2 and A0 seems to be slightly larger than for A1 in the NJ's productions.

**Figure 3.12. The average F0 patterns of NJ individual speakers.**

Each panel represents A1 (top), A2 (middle) and A0 (bottom). The x-axis indicates time-normalized points. The y-axis indicates normalized F0 values.

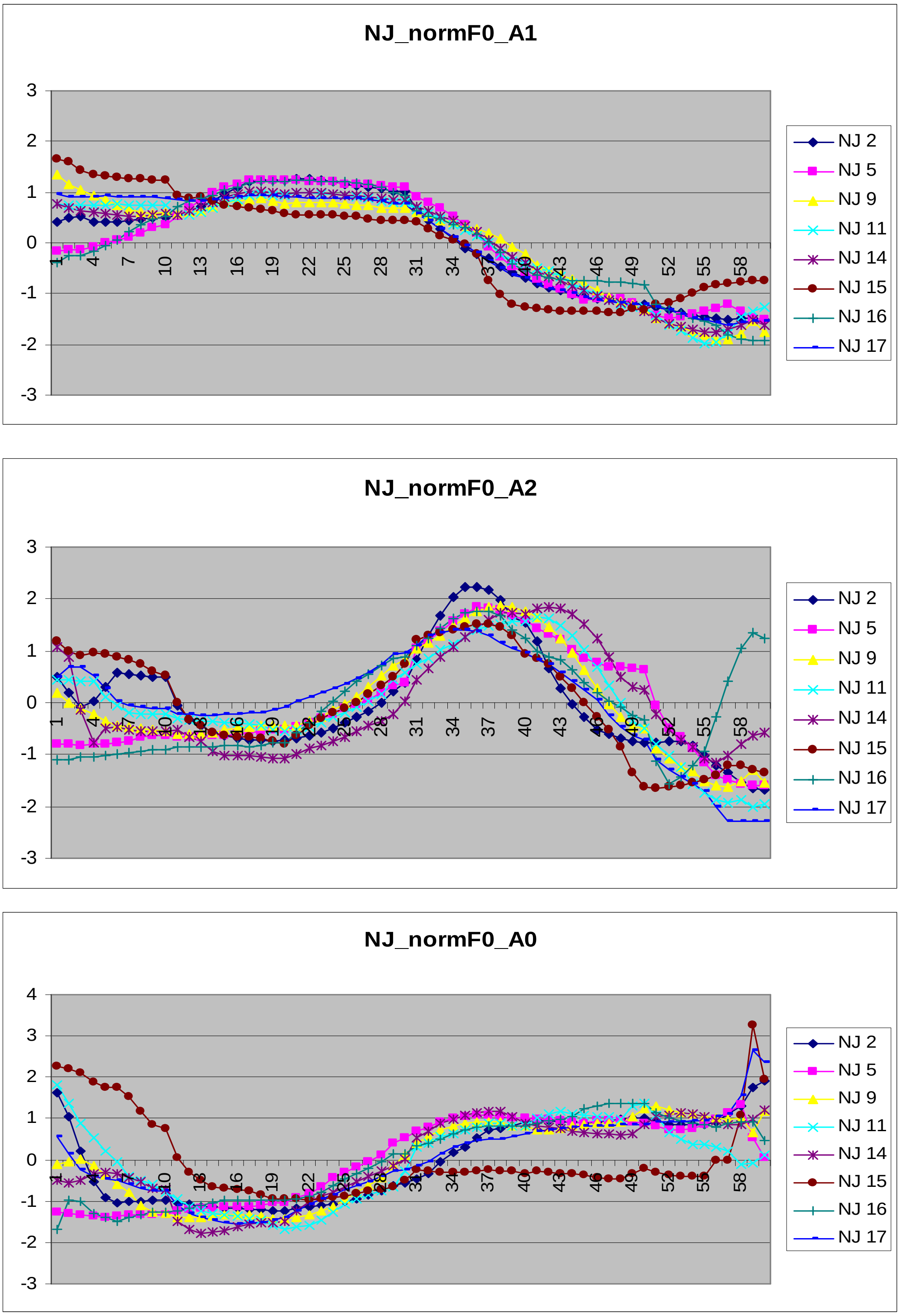
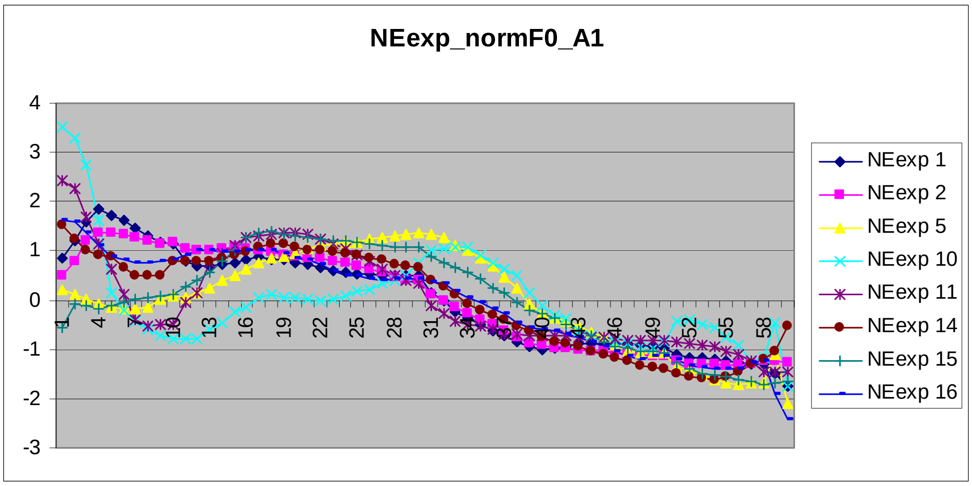
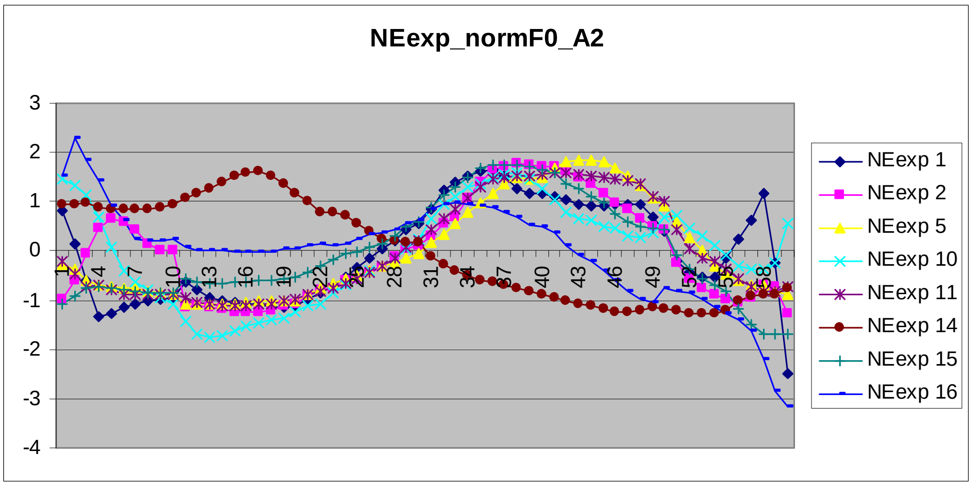


Figure 3.13 represents the F0 contours in the productions of the experienced learners (NEexp) for each accent. Overall, the F0 patterns produced by NEexp look similar to NJ's patterns, although NEexp's patterns seem to show more variability in each accent type, for A2 and A0 in particular. NEexp's A1 pattern tends to show an earlier peak relative to whole utterances while NEexp's A2 pattern seems to have a peak later peak relative to the whole utterances. The general trend of NEexp's A0 pattern is also similar to NJ's A0 pattern. Nevertheless, the A2 pattern of NEexp 14 and the A0 pattern of NEexp 1 seem to be different from the F0 patterns of other learners, i.e., they seem to be similar to the general patterns we can generally see in the A1 productions. The F0 peak locations for the accented targets (i.e., A1 and A2) in the NEexp's productions appear slightly later than the peak locations in the NJ's productions. Some NEexp show a F0 peak followed by a fall for the unaccented targets (i.e., A0). As for the degree of the F0 fall after the peak, NEexp show a similar tendency to NJ: in their productions, we can see a detectable F0 peak for the accented targets. However, the degree of fall for A1 and A2 was slightly less steep than it was in the NJ's productions. In addition, we can also detect a similar degree of fall for A0 to that for the accented targets in some of the NEexp's productions while it was relatively flat in the NJ's productions. Meanwhile, the ranges of the F0 contours for A2 and A0 were slightly wider than for A1.

**Figure 3.13. The average F0 patterns of NEexp individual speakers.**

Each panel represents A1 (top), A2 (middle) and A0 (bottom). The x-axis indicates time-normalized points. The y-axis indicates normalized F0 values.



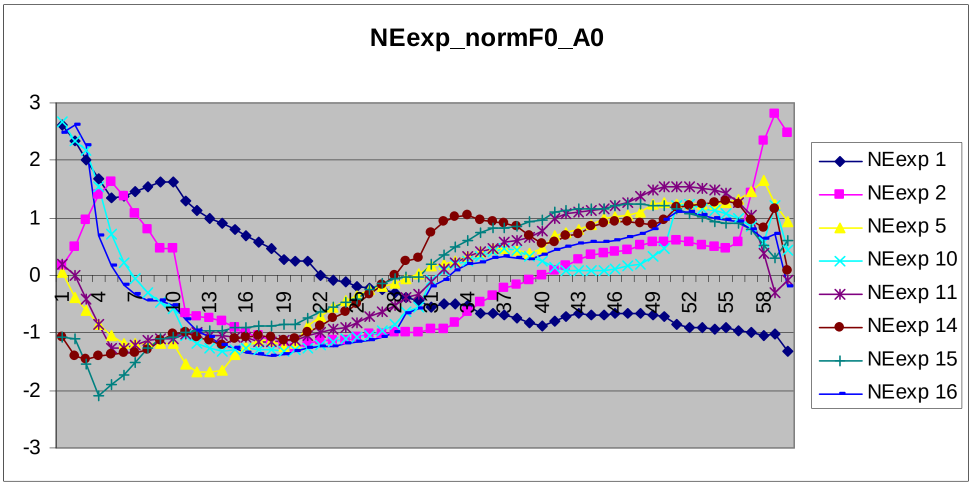
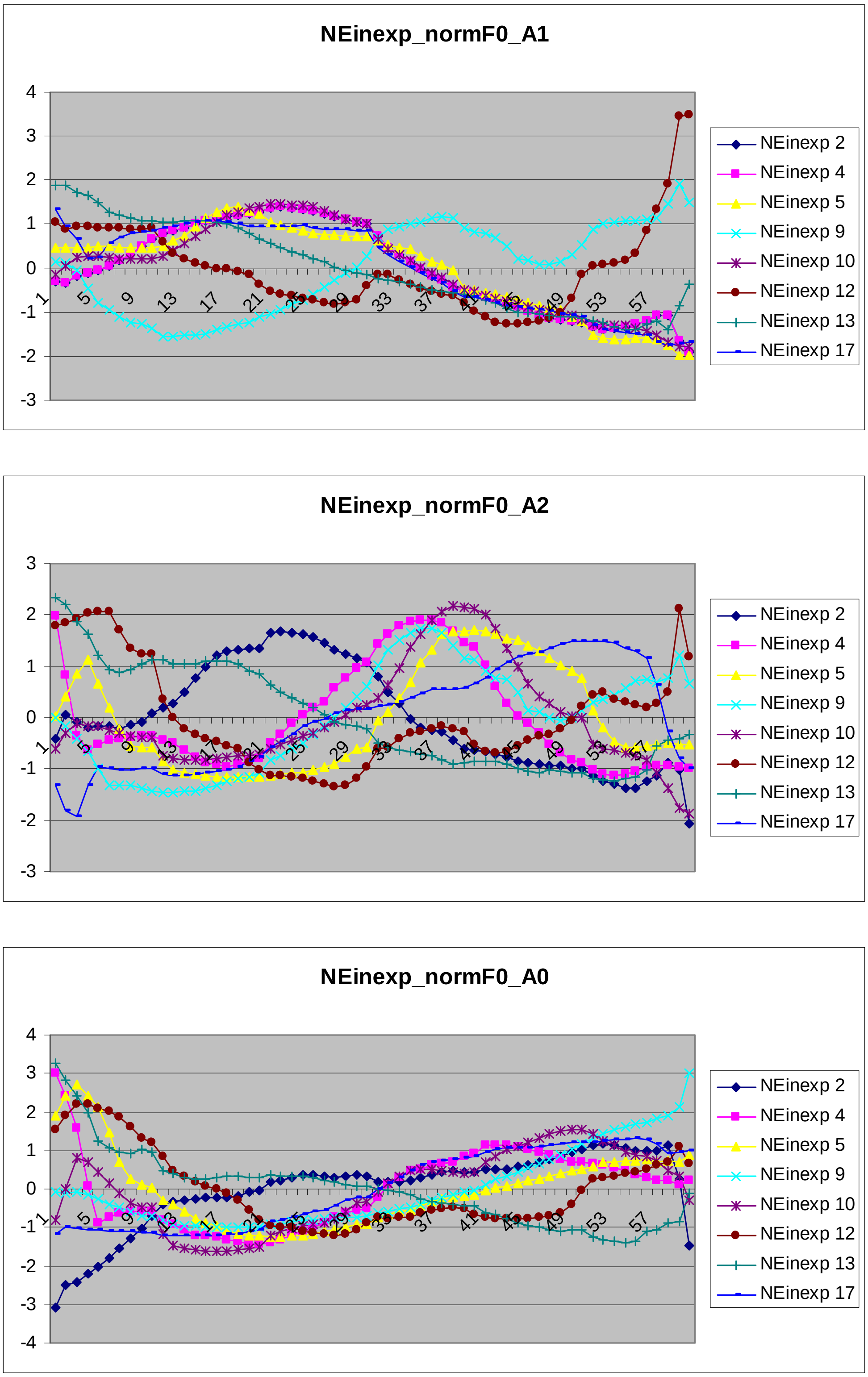



Figure 3.14 shows the F0 contours in the productions of the inexperienced learners (NEinexp) for each accent. What we see here is that the individual F0 patterns of NEinexp are much more diverse compared to NJ's patterns, and also more diverse than NEexp's patterns. However, we should also notice that the general trend of NEinexp's A0 pattern is not completely different from NJ's A0 pattern while the F0 patterns of the accented types, i.e., A1 and A2 tend to show more diversity. Additionally, the F0 patterns of some NEinexp speakers seem to resemble those of NJ speakers. In the NEinexp's productions, there do not seem to be large individual differences in the F0 peak location, particularly for A2. The F0 peak locations in the NEinexp's productions seem to be later than those in the NJ's productions, with more variance than in the NEexp's data. As for the degree of the F0 fall, we can detect some fall for the accented targets as well as for the unaccented targets in the NEinexp's data. Their ranges of the F0 contours for A2 seem slightly larger than for A1 or A0.

**Figure 3.14. The average F0 patterns of NEinexp individual speakers.**

Each panel represents A1 (top), A2 (middle) and A0 (bottom). The x-axis indicates time-normalized points. The y-axis indicates normalized F0 values.



These figures for individual speakers give us a general overview of the overall F0 patterns of each group. Overall, the learners seem to show non-target-like F0 patterns in their productions of Japanese lexical pitch accent contrasts regarding the F0 peak location and the F0 fall. In sum, native speakers of Japanese produced the F0 movements of each accent type as expected. As an overall shape, the contours contained the F0 peak followed by a fall and the peak seems to align with the pitch bearing unit in the case of accented tokens. Meanwhile, the learners' overall F0 patterns do not seem to be completely different from the native speakers' patterns. At least some of the learners seem to produce their F0 peak in a location similar to the native speakers. Moreover, the overall patterns seem to show a F0 peak for the accented targets while not having a marked F0 peak for the unaccented targets. It is also observable that the learners' patterns become more similar to native speakers' as a function of L2 experience.

### 3.6. Discussion

The question addressed in this section was whether or not the F0 profiles produced by English-speaking learners diverge from those produced by native speakers of Japanese. The data showed that the learners' F0 profiles do indeed diverge from the F0 patterns in the native speakers' productions. The learners' divergent F0 patterns in their imitations of Japanese pitch accent were compared with native speakers' norms in terms of intelligibility scores and overall F0 patterns. Both analyses revealed that native speakers of English produce significantly different F0 patterns from native speakers of Japanese. The overall results also indicate that L2 experience played a role. The more L2 experience the learners have, the more native-like their F0 profile becomes in producing pitch accent.

The results of the intelligibility scores showed that the scores of the experienced learners were lower than those of the native speakers but better than those of inexperienced learners. In other words, the two native speakers as the raters judged the productions of inexperienced learners as showing the most non-native-like patterns compared to the other two groups. Although both groups of the learners did as well as the native speakers in producing the unaccented words (A0) in terms of the intelligibility scores, the analysis of the misidentification types indicated that learners tended to overuse the unaccented pattern. The learners did worse in producing the words with the accent on the second syllable (A2) than the words with the accent on the first syllable (A1). These results imply that the learners were struggling with producing the accented patterns, particularly A2. If English-speaking learners do not know how to realize F0 curves for the accented patterns, it might have been easier for them to produce the unaccented pattern. It may have been for this reason that the learners tended to use the unaccented pattern more frequently, to compensate for their difficulty in producing the accented patterns. It is also possible that as the words with A2 accent type have the F0 peak somewhere in the middle of an utterance, this might have caused more problems for the learners to find out and/or aim for where the specific location of the F0 peak is.

The results of the acoustic observations of the overall F0 patterns showed that the native speakers of Japanese were making use of three characteristics of F0 shape (i.e., F0 peak location, F0 fall and F0 range) in producing Japanese pitch accent. On the other hand, English-speaking learners were not using these acoustic characteristics as much as the native speakers of Japanese, although it seems that the learners were trying to differentiate Japanese pitch accent in their own ways. Although some learners indicated some similarities, the overall F0 patterns in the learners' productions were non-target-like. This divergence was found to be greater when L2 experience is more limited. The learners seem to have problems in producing accented targets by phonetically realizing the location of F0 peak and the degree of F0 fall in their productions of Japanese lexical pitch accent. This suggests that if English speakers are to achieve target-like productions, they may require more advanced learning. To indicate the specific locations of the F0 peak of the target, or the rapid fall for the accented targets as well as to indicate the unaccented targets without such characteristics may have been difficult for learners to learn, as it may be not salient enough to them or it may require more fine adjustment. Hence, foreign accentedness in the learners' productions was also revealed in phonetically realizing Japanese lexical pitch accent contrasts.

To conclude, this section provided the answer to the question of whether or not F0 profiles produced by English-speaking learners diverge from those produced by native speakers of Japanese. The data showed that the learners' F0 profile was indeed different from native speakers' F0 patterns in producing Japanese pitch accent. The next question to be answered therefore was how specifically the learners' F0 patterns diverge from the native norms. This was reported in terms of both auditory measures (i.e., intelligibility scores) and descriptions of the overall F0 contours (i.e., F0 peak location, F0 fall and F0 range). The results suggested that English-speaking learners become more convergent with the native patterns as they gain experience of L2. Hence, English speakers of Japanese show foreign accentedness in the productions of Japanese lexical pitch accent. We also found that the English learners of Japanese have difficulty in phonetically realizing the F0 pattern (regarding F0 peak location and F0 fall but not so much for F0 range) which is the acoustic correlate of the contrasts. Now that we have found that the learners' F0 profiles diverge from the native speakers' profiles, we need to investigate further to answer the second question: what the potential factors behind such differences during L2 speech acquisition could be. In the next four chapters, three plausible factors are investigated in turn to provide the answer to this question.

### 3.7. Summary

In summary, this section provided evidence that native speakers of English produce F0 profiles which are different from those of native speakers of Japanese, and that they show foreign accent. They crucially differed from native norms with respect to F0 peak location and the degree of the F0 fall. Moreover, it also showed that the learners' productions of Japanese lexical pitch accent become less foreign accented and the learners' F0 patterns in their productions seem to develop during the L2 acquisition as a function of L2 experience.

# 4 Experiment 2: The ability to differentiate F0 contours

In Experiment 2, the issue of perceptual sensitivity is explored. The question addressed here is whether or not adult L2 learners are lacking in perceptual sensitivity to the relevant phonetic property which is employed to signal L2 prosodic contrasts when this phonetic property also has a linguistic function in the learners' L1. More specifically, the following question is examined: whether or not native speakers of English will be able to hear the difference between the F0 contours in pairs which are extracted from Japanese disyllabic nonce words. In a non-linguistic context, it can possibly allow us to examine learners' ability to hear the differences between the F0 contours without linguistic information. Considering that F0 has the function of differentiating intonation patterns in English, it is less likely that native speakers will have lost their perceptual sensitivity to F0 movements. In this case, we expect that English-speaking L2 learners will be able to hear the difference between sounds which consist of F0 contour. However, it is also possible that even though English speakers are sensitive to phonetic differences of F0 in Japanese lexical pitch accent contrasts, these phonetic differences may be masked by the linguistic information in a linguistic context. For this reason, it seems profitable to investigate learners' ability to differentiate the acoustic correlate of L2 contrasts, i.e., F0, where we could put the linguistic information aside. Hence, at this stage, we do not know whether it is plausible that perceptual sensitivity is a good indicator of L2 foreign accent in the case of Japanese lexical pitch accent acquired by native speakers of English. To answer this question, a non-speech ABX task was carried out using audio stimuli which consisted of F0 contour extracts.

## 4.1 Participants

Total of 48 participants were tested to examine whether or not English speakers could hear the difference between F0 contours in pairs. There were three groups: two learners groups (inexperienced and experienced) and a control group of native speakers of Japanese.

## 4.2 Materials

Audio materials for Experiment 2 were created from the recording of 12 disyllabic nonce words (4 words (*mene, noma, mani* and *nime*) x 3 accent types (A1, A2 and A0)) and a post-positional particle *mo* (also) by a female Tokyo Japanese speaker. To create audio materials for the non-speech stimuli task of this experiment, the speech stimuli were manipulated in such a way that only its F0 contour was extracted from the original recording of the speech stimuli. This was done using PRAAT. First, the visible F0 contour was extracted from each speech stimulus that was sampled at the interval of 0.01 sec. Then, the resulting extracted output was synthesized into a sound with phonation type by generating a glottal waveform from every point of the F0 contour. As a result, each final output sounded like a buzzing noise with its duration equivalent to that of the original recording.

## 4.3 Procedure

An ABX task using these non-speech stimuli, i.e., F0 extracted stimuli, was conducted. Participants were tested individually in a sound-treated recording studio. Sets of three audio stimuli were presented in a sequence at a time. The set consisted of non-speech stimuli drawn from each of the three accent types of the same word type. Two were of the same accent type, and either the first or the second stimulus would be presented again as the third stimulus. The participants' task was to judge for each set whether the last stimulus (X), was the same as the first (A) or the second (B). For example, three non-speech stimuli corresponding to the F0 contours of *mene*(A1)*mo*, *mene*(A2)*mo* and *mene*(A1)*mo* were presented, and then, participants were to indicate whether the last stimulus, *mene*(A1)*mo*, was the same as the first stimulus, *mene*(A1)*mo* or the second stimulus, *mene*(A2)*mo*.

Instructions were presented on a computer monitor in the recording booth. Participants listened to the audio materials through a pair of headphones and indicated their response by pressing the appropriate key on a keyboard. When participants thought that the last stimulus was the same as the first stimulus, they were asked to press the ‘A’ key, the leftmost key on the middle key of the keyboard, and when they thought that it was the same as the second stimulus, they were requested to press the ‘L’ key, the rightmost key. These two keys were marked with stickers so that they were visibly clear. After a practice, the main test was conducted. Forty eight sets of F0 contour combinations were randomly presented over the headphones, i.e., 4 disyllabic nonce words x 3 accent types x 4 presentation orders (ABA, ABB, BAA and BAB).

The interval between each stimulus within a set was 0.70 sec. The duration of the interval was based on findings in the literature. Pisoni (1973) found that the inter-stimulus interval affected vowel discrimination accuracy. For short vowels, between-category discrimination of short vowels was quite high and stable across the tested 5 intervals between 0 and 2.0 sec. However, it was relatively higher between 0.25 and 0.5 sec. Within-category discrimination was maximum when the interval was 0.25 sec, and beyond 0.25 sec, it decreased with increases in the interval between the stimuli. Pisoni interpreted these results suggesting that the duration indicates the processing time necessary for auditory recognition. This suggests that the interval of 0.5 sec is the minimum interval for auditory recognition when both between- and within-category stimuli are considered. Moreover, his data shows that between-category discrimination accuracy of short vowels is high regardless of the intervals. Meanwhile, within-category discrimination accuracy of short vowels falls to below 60 % when the inter-stimulus interval is between 0.5 and 1.0 sec. This also implies that not only the auditory recognition but also more phonetic processing may play a role in hearing the differences between the pair stimuli as the interval becomes longer. In our ABX task using the non-speech stimuli (Experiment 2), there was only between-category stimuli comparison. However, we also conduced a step-wise ABX task which will be described in Chapter 6 where we examined the ability of within-category stimuli comparison as well as the ability of between-category stimuli comparison. For this reason, both between- and within-category results of Pisoni were relevant to our study. Considering the interval of 0.5 sec is the minimum for the auditory recognition and the interval of 1.0 is the maximum, the intermediate value, i.e., the value of 0.70 sec was selected as the inter-stimulus interval for the task in Experiment 2 (and also the rest of the perceptual experiments and the presentation of the production stimuli reported in the other chapters). In addition, as personal communications with a few participants during the pilot study, they reported the interval of 0.5 was too quick and that of 1.0 was too long to comfortably carry on the task. Meanwhile, the interval between each response and the presentation of the next stimulus was 1.0 sec. This duration was selected which was employed in the perception task in the literature such as Flege, Bohn and Jang (1997). According to some feedback, it was a good pace for the participants who joined a pilot study of this study to carry on the task.

## 4.4 Analysis & Results

The percentage of correct responses of each participant was calculated and the means were compared among the three groups to investigate whether the performance of the two learners' groups differed from native speakers of Japanese. In general, all three groups performed quite well. The average accuracy of the group of native speakers of Japanese was 94.7% (*SD* = 0.80), that of NEinexp was 94.7% (*SD* = 0.71) and that of NEexp was 95.3% (*SD* = 0.49). A two-way mixed design ANOVA was conducted in which Accent type (A1, A2 and A0) was the within-subject factor and Group (NJ, NEinexp and NEexp) was the between-subject factor. No significant main effects or no interaction effects were found. Thus, the mean accuracy of the participants in hearing the difference between the pairs of F0 contours was not significantly different among the three groups, although a ceiling effect could have masked a difference. English speakers were able to hear the difference between F0 contours extracted from disyllabic nonce words like Japanese speakers, and they showed this ability regardless of their L2 experience.

## 4.5 Discussion

The results of Experiment 2 show that native speakers of English and native speakers of Japanese were able to hear the difference equally well between pairs of sounds created from the F0 contours which were extracted from disyllabic nonce words. In other words, learners could hear the differences between F0 contours among the three accent types as well as native speakers could when the linguistic information of the F0 function was not presented. This means that English-speaking learners were not having a problem in hearing the acoustic differences in the F0 contours of the three accent types. It is also worth mentioning that there was also no statistical difference between two learners groups as a function of L2 experience. Therefore, it is less likely that foreign accent stems from native speakers of English losing their sensitivity to detect the acoustic differences between F0 contours. Overall, it is not plausible to conclude that English speakers' perceptual sensitivity to detect F0 contour differences has been lost as adult L2 learners and that this is what is reflected in their foreign accent when producing Japanese lexical pitch accent contrasts. Therefore, this calls for further investigation in order to find an alternative account to explain what might be behind foreign accent phenomena.

## 4.6 Summary

In Experiment 2, the learners' ability of differentiating the F0 contours was investigated as a source of foreign accent. The question addressed here was whether or not native speakers of English as adult L2 learners could hear the difference between the F0 contours in pairs which were extracted from Japanese disyllabic nonce words when the cross-linguistic information of F0 function between Japanese and English is removed. Based on the results of the non-speech ABX task, native speakers of English seem to have high sensitivity to F0 contour differences just like native speakers of Japanese. Hence, it seems reasonable to conclude that the ability to differentiate the F0 contours does not provide a plausible account for foreign accent. This possibility is not tenable at least in the case of Japanese lexical pitch accent acquired by native speakers of English. Although we found that English learners of Japanese do not seem to have a problem in acoustic perception as part of the input process of speech, this does not grant that these learners are able to articulate the F0 contours in a non-linguistic context as part of the output process. It is possible that the learners are having a problem in articulating the F0 differences between the lexical items, and that difficulty is appeared as their foreign accentedness in their productions. In the next chapter, the articulatory/motor issue is explored as an alternative account for foreign accent in L2 prosodic contrasts by focusing on the learners' ability to articulate the F0 contours.

# 5 Experiment 3: The ability to articulate F0 contours

Let us now turn to the role of the learners' ability to articulate the F0 contours in the L2 productions. The result of the ABX task using non-speech F0 contours revealed that English native speakers' sensitivity to F0 contour differences is just as good as that of native speakers of Japanese. This pattern was observed in both experienced and inexperienced learners. These results suggest that English-speaking learners were not having a problem in hearing the phonetic differences between the F0 contours in pairs when the stimuli are non-speech stimuli that consist of only the target F0 contours without cross-linguistic difference between English and Japanese. Therefore, this excludes the possibility that their foreign accent is due to lack of ability in perceptually differentiating the F0 contours in a non-speech context; this ability was unrelated to English-speaking learners' divergent F0 patterns of Japanese pitch accent which was revealed as foreign accent.

If the learners are able to maintain perceptual sensitivity to a particular phonetic property (i.e., F0), what could be an alternative behind the divergent F0 patterns of English-speaking learners from Japanese speakers' norms? We can examine learners' performance more exhaustively by testing the same participants. The next step is to investigate the possibility that the English native speakers have problems regarding their ability to articulate

F0 contours.

In this experiment, we will explore the possibility that articulatory/motor processes are associated with non-native patterns of the relevant acoustic correlates. The general aim of Experiment 3 is to investigate whether the constraints of articulatory control or motor processes are making it difficult for adult L2 learners to phonetically realize the acoustic correlates of the target L2 prosodic contrasts, when this articulatory/motor aspect of the phonetic property is disentangled from its linguistic role. Here we make use of the fact that F0 patterns such as F0 peak alignment and rapid F0 fall are found to be the acoustic correlates of Japanese lexical pitch accent contrasts (e.g., Hasegawa & Hata, 1992; Sugito, 1982) whereas F0 patterns also mark intonation patterns in both Japanese and English (e.g., Beckman, 1986; Ladd, 1996). Putting aside the cross-linguistic difference of the F0 function, the more specific question addressed here is whether or not native speakers of English will be able to articulate the different target F0 movements when they imitate target F0 contours which have been extracted from Japanese nonce words. By removing the linguistic context from the Japanese nonce words, we consider that the linguistic function of F0 is excluded and thus, this allows us to examine the learners' ability to articulate the F0 movements (simply considered as a motor process rather than for their linguistic function). If the learners are able to articulate the different target F0 movements, the second question is whether or not the F0 patterns in the productions of native speakers of English will diverge from those of native speakers of Japanese.

Now, let us consider possible outcomes. Since native speakers of English use F0 linguistically in their L1 in differentiating intonation patterns, it may be that English speakers will be able to articulate the F0 movement of target contours without any problems once the linguistic context is removed from Japanese pitch accent contrasts. In addition, it is possible that their F0 patterns may show a similar pattern to native speakers of Japanese when they articulate the F0 movements without the linguistic information of pitch accent. If this was case, we would be able to assume that native speakers of English do not have articulatory difficulty in producing F0 patterns whey they only have to imitate F0 movement of the target contours. This would suggest that we can eliminate the possibility that articulatory/motor processes can account for non-native F0 patterns when native speakers of English produce Japanese pitch accent. This would also leave more room for investigating further into other potential factors behind the non-target-like F0 profile of English speakers, such as factors which may be more directly related to the linguistic function of F0.

On the other hand, it is also possible that the results may show that the F0 patterns of English speakers will differ from those of Japanese speakers. It could be that articulation of F0 movements even without linguistic context may still require specific articulatory/motor control for lexical pitch accent which could be different from the control used for phrasal differentiations (particularly because the source of the F0 contours is the lexical items). Alternatively, it could be that the target patterns are too difficult to achieve from the articulatory perspective as the articulatory requirements of F0 patterns in Japanese lexical pitch accent are very specific to Japanese pitch accent. In this case, we may see some differences between experienced and inexperienced learners as a function of L2 experience. After having more practice with the phonetic realization of the different F0 patterns in Japanese through the learning process, the experienced learners may perform more similarly to native speakers compared to the inexperienced learners. For these possible reasons, native speakers of English may or may not be able to articulate the differences of the F0 movement of target contours even in their own ways. The learners may be distinguishing the three accent types of Japanese lexical pitch accent contrasts which are still different from the means in the native speakers' productions; however, they seem to make distinctions within their means among the three accent types. In such cases, this would imply that native speakers of English are experiencing articulatory difficulties in producing F0 patterns even without linguistic context. If we could also show that such a problem was somehow related to their foreign accent, this would provide evidence for the possibility that articulatory/motor processes are a potential factor for explaining the divergence of the learners' F0 patterns from native speakers' patterns.

In order to test these possibilities, a non-speech imitation task was carried out, as described in the following sections.

## 5.1 Participants

The same 48 participants as the previous tasks were tested. Recall that there were three groups: two learner groups (inexperienced and experienced) and a control group of native speakers of Japanese. Due to time constraints, data from 24 participants (8 participants from each group) were randomly selected and analyzed.

## 5.2 Materials

Audio materials for this task were the same as those for Experiment 2: F0 contours were extracted from recordings of 12 disyllabic nonce words (4 words (*mene, noma, mani* and *nime*) x 3 accent types (pitch accent on the 1$^{st}$ syllable, pitch accent on the 2$^{nd}$ syllable and unaccented)) followed by a post-positional particle *mo* ('also'), as uttered by a female speaker of Tokyo Japanese.

As we saw in the materials for speech imitation task in 3.2.2, this speaker's productions were recorded in a sound-treated recording studio at the Linguistics laboratory at the University of Edinburgh. The audio interface was MOTU 828 mkll Firewire using a hypercardoid microphone (AKG CK 98) and digitalized at a sampling rate of 48 kHz with a resolution of 16 bits. Sonar 4 Studio Edition was used as a recording software programme. A list of the 12 test items was read aloud in a carrier phrase three times (the carrier sentence was, "*Sumimasen* ____*mo kudasai.* (Excuse me but please give me ____ as well.)"). Subsequently, one of three repetitions of each item including the particle *mo* 'also' was edited to form the audio stimuli.

In order to create audio stimuli for non-speech stimuli tasks, the speech stimuli were further manipulated in such a way that only the F0 contour was extracted from the original recording, and synthesized into sound using PRAAT functions. First, the visible F0 contour was extracted from the original sound. Then, the extracted F0 contour was synthesized into sound with the pulses algorithm. Each final output sounded like a buzzing noise with its length equivalent to that of the original recording.

All the audio materials were converted to a sampling rate of 22 kHz with a resolution of 16 bits using CoolEdit Pro 2.1 so that it matched the Sound Device Object Property in E-Prime.

## 5.3 Procedure

Participants were tested individually in a sound-treated recording studio. The participants heard paired stimuli in a sequence through a pair of headphones and were asked to imitate the F0 movement of the first non-speech stimulus. They were asked to do so by producing a Japanese vowel /a/ (the description on the monitor showed, a Japanese *hiragana* orthographic symbol, equivalent to /a/, in order to demonstrate the sound as part of the instructions). Two stimuli were presented so that this task would have a parallel structure to the non-speech task in this study, i.e., the non-speech ABX task in Experiment 2. They were allowed to make more than one attempt if they were not satisfied with their first attempt and if they did, the final imitation was used for the analysis. Thirty-six pairs of F0 contours were randomly presented over headphones (4 disyllabic nonce words x 3 accent types x 3 presentation orders (AB, BA, and either AA or BB)).

The audio interface was MOTU 828 mkll Firewire using a hypercardoid microphone (AKG CK 98) and digitalized at a sampling rate of 48 kHz with resolution of 16 bits. Sonar 4 Studio Edition was used as a recording software programme. Participants listened to the audio materials using a pair of Sennheiser eH2270 headphones. A computer keyboard was used to indicate to proceed to the next stimulus by pressing the return key. All the instructions were presented with E-Prime using a computer monitor facing the participants in the recording booth.

Participants were given a practice session prior to the main test. In order for the participants to familiarize themselves with the non-speech imitation task, and to allow them more time to ask any questions, the duration of the practice session in this task was twice as long as that of the other tasks.

After the participants produced their imitation of each stimulus, they were asked to press the space bar on the keyboard in order to proceed to the next pairs to be presented. The inter-stimulus interval was again 0.7 sec.

### 5.4 Analysis

The analysis procedures for this task are basically the same as those for Experiment 1. The recording of each participant was digitalized at a sampling rate of 48 kHz with resolution of 16 bits. The recording of each participant was segmented into smaller sound files. This yielded 36 small sound files per participant, each of which contained one of 36 imitations of the target items, each with duration approximately 1.0 sec. These sound files were then analyzed in terms of acoustic measures. Annotation and acoustic analyses were performed using PRAAT.

The segmentation was conducted manually by looking at the visual display of the spectrogram and the oscillogram based on the criteria below. With the aid of a PRAAT script written by Xu (2008), the acoustic measurements, which are described below, were obtained after segmenting each audio file. As in the analysis of the production data in Experiment 1, to capture the overall F0 patterns, the following characteristics were again of particular interest for the purpose of this study: F0 peak location with respect to the onset and the offset of each non-speech imitation, F0 fall and F0 range.

To observe the location of F0 peak in non-speech imitations, the time-normalized F0 peak point was detected for each utterance. The details of the time-normalization method are described in 3.4.3 in Chapter 3. To compare the degree of F0 fall, the fall after the peak was observed. Lastly, F0 range (i.e., the width of the normalized F0 contours) was compared among the groups. The advantage for using both F0 range and maximum velocity is that the combination of these two values can potentially provide us a way to dynamically compare whether English speakers are able to produce a rapid F0 fall after the peak when they are imitating F0 movement alone. This point is particularly relevant for analyzing the imitations of the non-speech stimuli as in these stimuli there are no segmental landmarks other than the onset and the offset of the utterance as a whole. Without segmental landmarks as a cue to phonetically realize F0 patterns, it may be more difficult for both learners and native speakers of Japanese to aim at the target F0 peak under the non-speech imitation task than the speech imitation task. For this reason, it may be difficult to compare F0 patterns produced by English-speaking learners with those by native speakers of Japanese solely based on the F0 peak location. In the following sections, we will describe the details of the acoustic interval criteria and those of each value.

### 5.4.1 Acoustic interval criteria

As in Experiment 1, the acoustic interval criteria used in the analyses followed the practical guide of Turk, Nakai and Sugahara (2006). The spectrogram setting in PRAAT was set as ‘view range’ of 0 – 8000 (Hz) and ‘dynamic range’ of 55 (dB). Approximately a 1-sec window was used for placing general boundaries and the waveform was zoomed into 5-10 ms for more fine-grained segmentation. The segmentation points were located at the nearest zero-crossing point of the waveforms.

The segmental boundary was placed based on spectral characteristics which can be easily detected in spectrograms. All the imitation of the non-speech stimuli in Experiment 3 was accompanied by the vowel /a/. The onset of /a/ was marked where a clear glottal release occurred. The offset of /a/ tended to end in creaky phonation (i.e., where glottal pulses are spaced irregularly with a very low F0). When the utterance ended in creaky phonation, the last clear glottal cycle on the waveform was detected and the boundary was marked on its zero-crossing point (Figure 5.1). When the offset of /a/ did not end in creaky phonation, the boundary was placed on the zero-crossing point of the glottal cycle of the waveform which corresponded to the end of strong F2 (Figure 5.2). Some of the utterances also ended in breathy phonation (i.e., where voiceless formant structures could be observed directly after the vowel formant structure). In these cases, the offset of /a/ was marked at the end of voicing of the waveform and the voiceless formant structure was excluded (Figure 5.3).

Figure 5.1 **Creaky phonation at the end of /a/**

Creaky voice is marked as “c”.

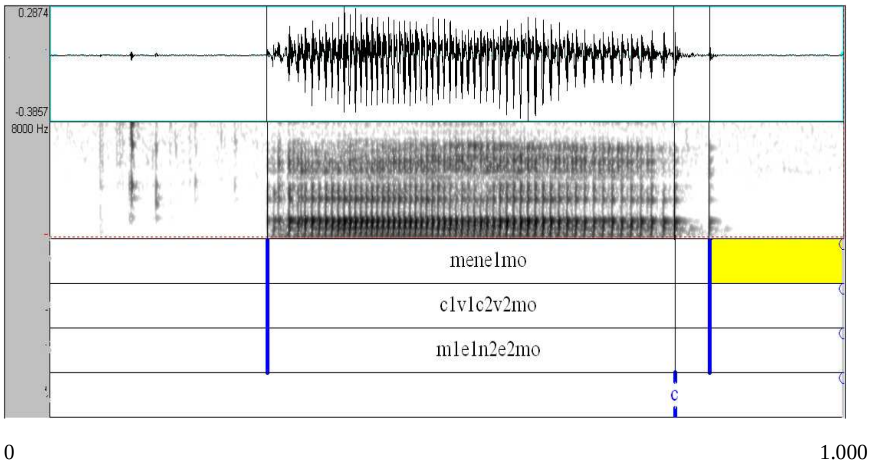


**Figure 5.2 Example of /a/ without creaky phonation ending**

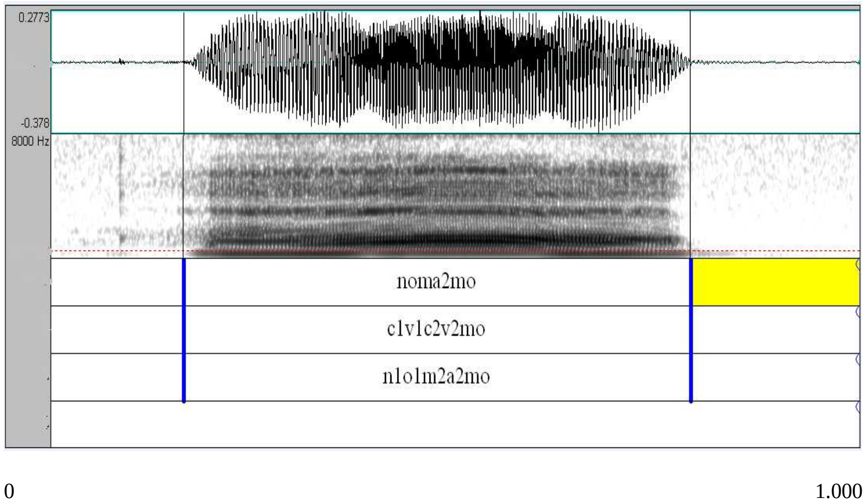
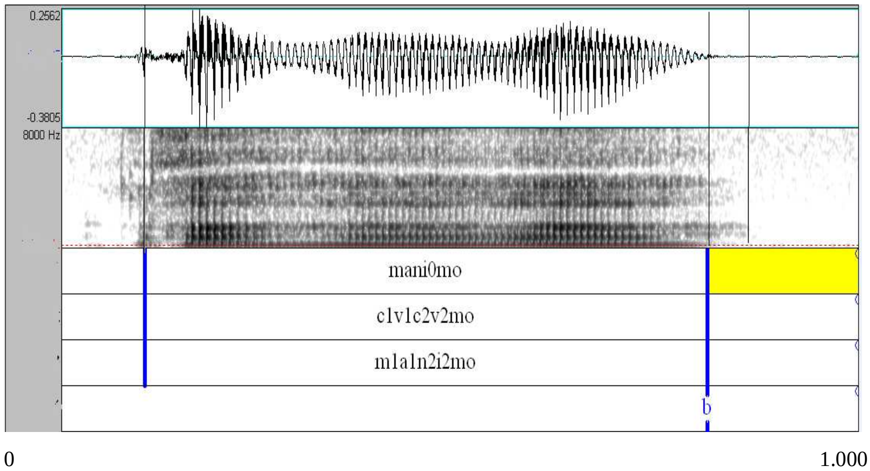



**Figure 5.3 Breathy phonation at the end of /a/**

Breathy voice was marked as “b”.




## 5.5 Results

Here we will present the overall F0 contour patterns produced by each individual speaker for each pitch accent types, focusing on F0 peak location, F0 fall and F0 range.

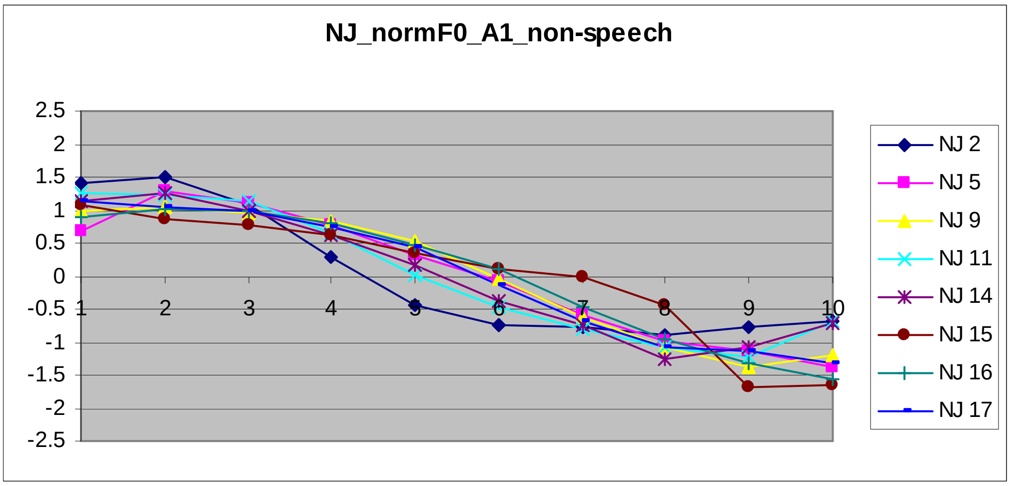
### 5.5.1 Overall F0 patterns

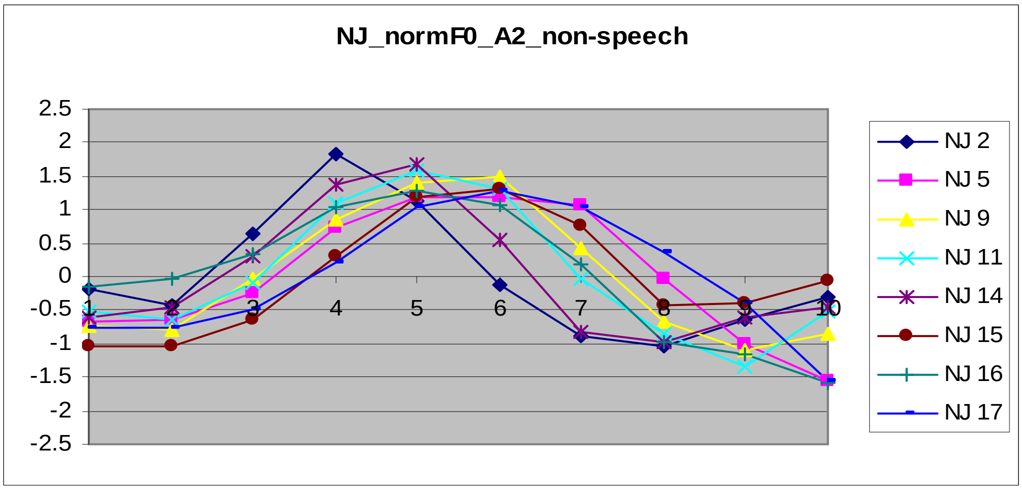
The following figures represent the overall F0 peak patterns of each participant, which were averaged over three repetitions of each test item. To reduce the speakers' variability, the F0 values (Hz) on the time-normalized scale were converted into normalized F0 values as in Experiment 1. In addition, Z-score transforms were used for normalizing the F0 values of each contour. Figure 5.4 represents the F0 contours produced by native speakers of Japanese for the non-speech stimuli created from the word with an accent on the 1st syllable (A1), the word with an accent on the 2nd syllable (A2) and the unaccented word (A0). Each line indicates the average F0 contour of a speaker.

As an overall tendency, the F0 patterns in A1 non-speech produced by the NJ tend to have a peak followed by a fall within the initial part of the utterances. The F0 patterns in A2 non-speech seem to show a peak in the middle part of the productions. Meanwhile, in the NJ's A0 non-speech we see a relatively rising or a flat F0 pattern which does not contain such a peak followed by a rapid fall, unlike their patterns of A1 or A2. In addition, the F0 ranges for A2 and A0 seem slightly wider than for A1. Hence, the F0 patterns in the NJ's non-speech data seem to show similar patterns to the patterns that we saw in their speech data in Chapter 3.

**Figure 5.4. The average F0 patterns of NJ individual speakers.**

Each figure represents the F0 patterns in the imitation of non-speech stimuli from A1 (top), A2 (middle) and A0 (bottom). The x-axis indicates time-normalized point. The y-axis indicates normalized F0 values.



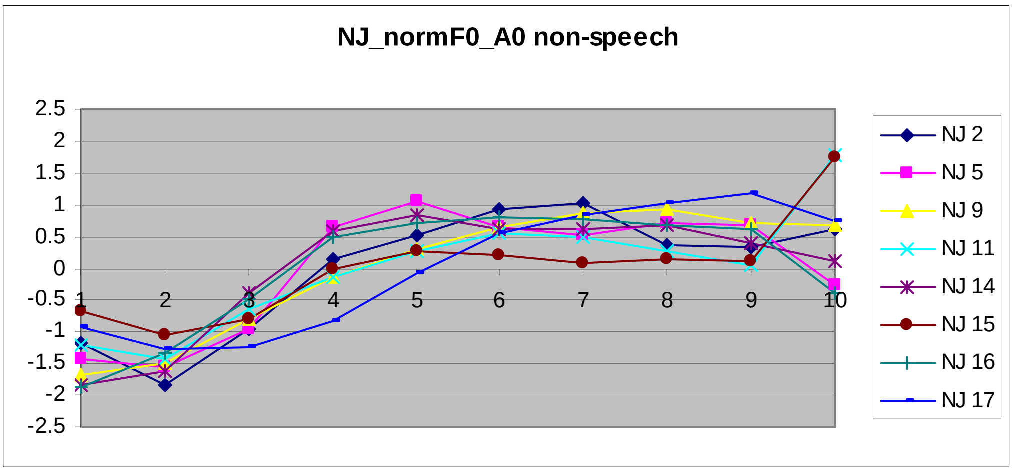
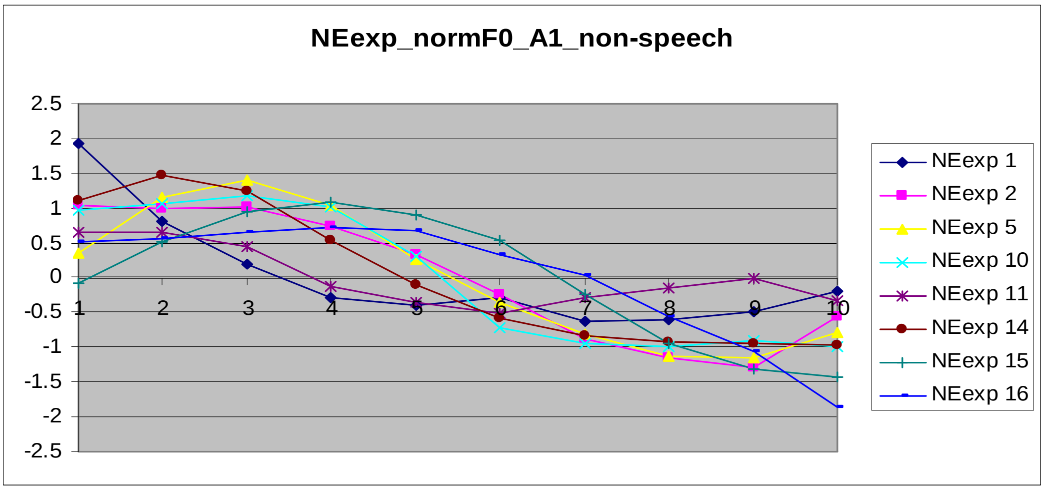
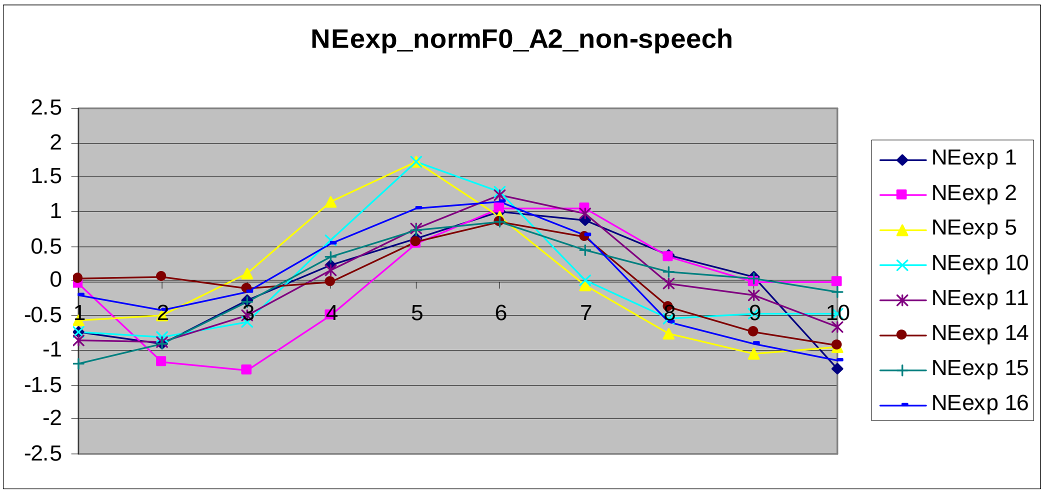



Figure 5.5 represents the F0 contours produced by the experienced learners (NEexp) for the non-speech stimuli created from the word with each accent type. Overall, the F0 patterns produced by the NEexp look very similar to the NJ's patterns, although the NEexp's patterns seem to show slightly more diversity compared with the NJ's patterns. The F0 patterns in A1 non-speech produced by the NEexp also tend to have a peak followed by a fall within the initial part of the utterances. The F0 patterns in A2 non-speech seem to show a peak in the middle part of the NEexp's productions. Meanwhile, we see a relatively rising or a flat F0 pattern which does not contain such a peak followed by a rapid fall in the NEexp's F0 patterns of A0 non-speech unlike their patterns of A1 or A2. Meanwhile, the F0 range for A2 in the NEexp's non-speech seems to be slightly narrower than in the NJ's non-speech.

**Figure 5.5. The average F0 patterns of NEexp individual speakers.**

Each figure represents the F0 patterns in the imitation of non-speech stimuli from A1 (top), A2 (middle) and A0 (bottom). The x-axis indicates time-normalized point. The y-axis indicates normalized F0 values.



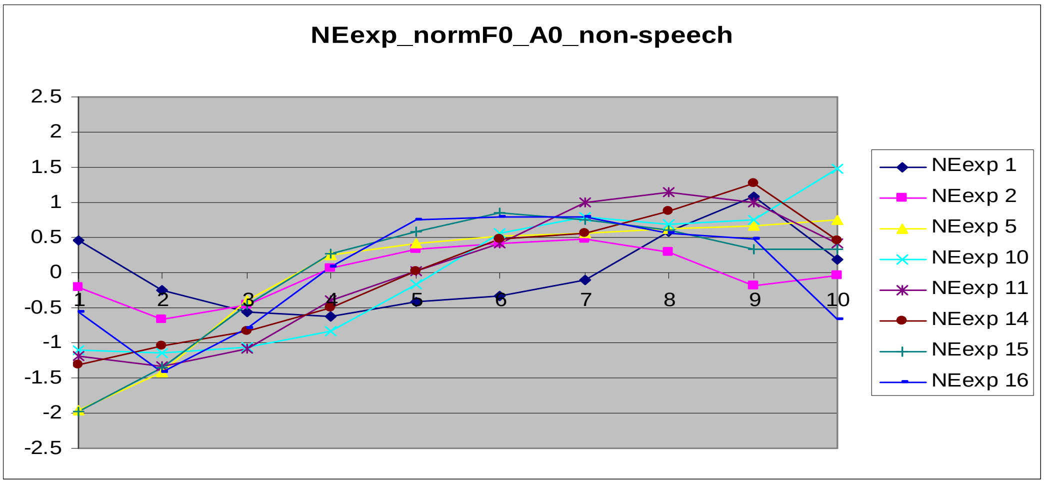
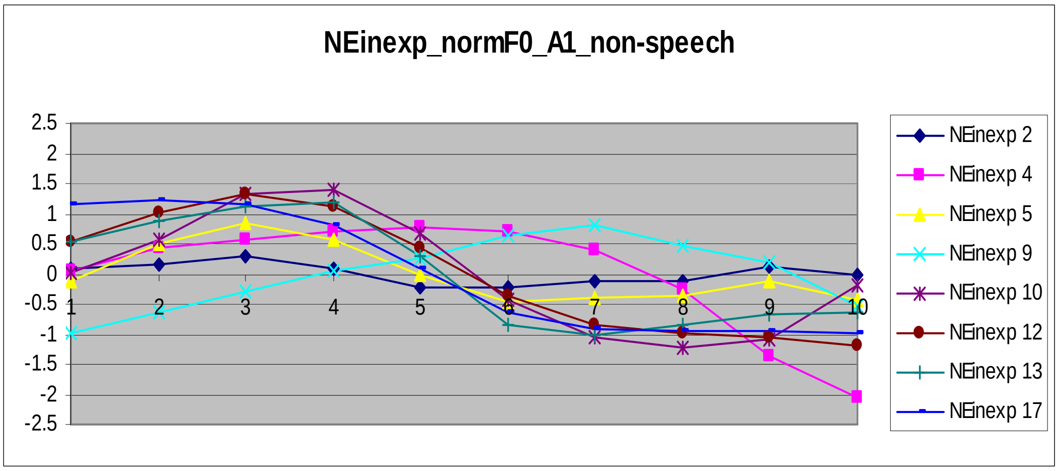
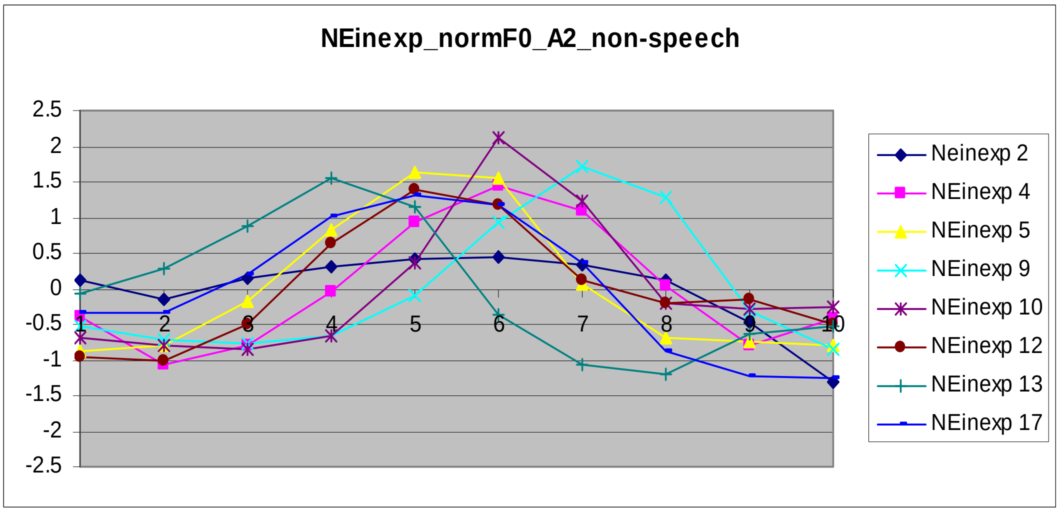



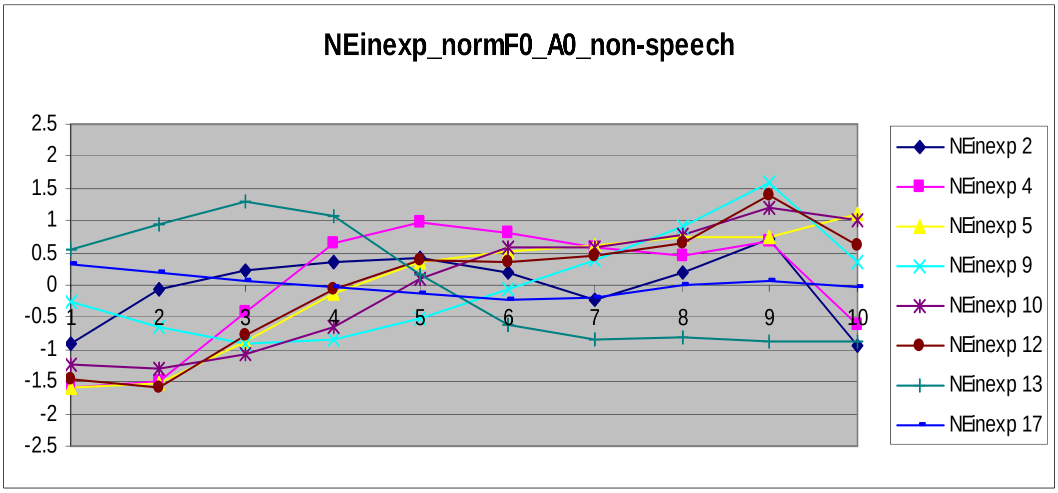
Figure 5.6 represents the F0 contours produced by the inexperienced learners (NEinexp) for the non-speech stimuli created from the word with each accent type. The individual patterns of the NEinexp also seem to show a similar tendency to those of the NJ. The overall trend of their patterns show an earlier peak followed by a fall in the A1 non-speech productions, a peak followed by a fall in the middle in the A2 non-speech utterances, and more or less flat patterns in the A0 non-speech. However, the NEinexp's patterns seem to show slightly more individual variations than the NEexp's patterns. The F0 peak locatios for the accented targets (i.e., A1 and A2) in the NEinexp's non-speech appear to be slightly later than in the NJ's non-speech. In addition, the degree of F0 fall for A1 in the NEinexp's non-speech seems to be steeper than in the NJ's non-speech.

**Figure 5.6. The average F0 patterns of NEinexp individual speakers.**

Each figure represents the F0 patterns in the imitation of non-speech stimuli from A1 (top), A2 (middle) and A0 (bottom). The x-axis indicates time-normalized point. The y-axis indicates normalized F0 values.



These figures showing the individual speakers in each group provide us with a good overview of the overall F0 patterns of each group. In sum, even in the non-speech imitations, the F0 patterns produced by the native speakers of Japanese indicated similar F0 patterns to those produced for imitation of the source words. Overall, the Japanese speakers' contours included a F0 peak followed by a fall for the non-speech imitations of F0 contours extracted from the accented word. In addition, in the Japanese group, this F0 peak and fall seems to occur within the initial part of the utterances in the A1 non-speech, as in the original A1 word with its accent on the first syllable, while in the case of the A2 non-speech imitations, the F0 peak and fall tends to occur in the middle of the productions (as in the original A2 word, with its accent on the second syllable). Meanwhile, both learner groups also seem to show this tendency in their non-speech productions. However, the learners seem to show more individual differences in their F0 patterns than the native speakers. Moreover, even in the non-speech imitations of the F0 contours, we also observed that the patterns become less diverged as a function of L2 experience, although the effect seems to be less strong than in the speech productions of Japanese pitch accent observed in Experiment 1.

## 5.6 Discussion

The results of Experiment 3 showed that native speakers of Japanese produced acoustic differences in terms of F0 peak location, F0 fall and F0 range when they were imitating the target F0 contours alone, i.e., contours taken from the three accent types while excluding the linguistic function of F0. The rationale behind the task was to remove the linguistic context so that we can put aside the cross-linguistic differences in the use of F0 between Japanese and English and test the participants' ability to produce the acoustic correlates of Japanese lexical pitch accent more directly. In the non-linguistic context (which in principle allows us to test specifically the articulation of F0 movements), the native speakers made use of all three of the F0 parameters. This result agrees with the previous findings obtained in the speech productions of pitch accent in Japanese where the F0 peak tends to occur in accent location.

In order to discuss the main question of this chapter, let us now consider in more detail how the English speakers performed in the non-speech imitation task. The first of the two questions addressed in this chapter was whether or not native speakers of English would be able to articulate (or phonetically realize) the differences target F0 movements in a non-linguistic context, i.e., where they had to imitate target F0 contours which were extracted from Japanese nonce words. As we saw in Figure 5.4, Figure 5.5 and Figure 5.6 (which give a general overview of the F0 patterns produced by the three groups), the average F0 movements of the learners seem to show similar patterns to those of the native speakers across the three accent patterns, although the inexperienced learners show more individual diversity. This seems to suggest that the learners were also making some acoustic differences in F0 patterns in imitating F0 contours depending on the accent type. Based on the overall F0 patterns in their non-speech data, like the native speakers, both the experienced and the inexperience learners were also making use of all three F0 parameters to separate the accent types in their non-speech imitations. However, the content of the F0 parameters and their magnitude in the learners' productions showed some differences relative to the native speakers' productions. The native speakers of Japanese seem to differentiate the accent types in the non-speech imitations in terms of F0 peak location and the degree of F0 fall but not so much F0 range. The experienced learners showed similar productions in this respect. Meanwhile, the inexperienced learners might have been making use of the degree of F0 fall in their non-speech to mark accent types. This may be why the F0 fall in the inexperienced learners' non-speech data for the accented target looked sharper than the native speakers' non-speech data while more F0 fall tended to be detectable in the inexperienced learners even for the unaccented targets.

The relevance of the overall patterns for the question is that they suggest it is not impossible for native speakers of English learning Japanese to be able to phonetically differentiate F0 patterns in terms of the three F0 parameters, when the linguistic function of F0 is removed. From this point, we are warranted to say that the learners do not seem to have so much difficulty in articulating F0 patterns in terms of these F0 parameters. Nevertheless, we should also note that the results also showed some differences between the learners' productions and the native speakers' productions, with more variability in the learners' data. Hence, it seems that English learners of Japanese have some problems in articulating F0 contours in Japanese lexical pitch accent, and it is possible that these differences might have contributed to the non-target-like productions of the learners.

Interestingly, even though this was a non-speech task, the results also revealed a role for L2 experience in terms of the overall F0 patterns. The results suggest that although the native speakers of English diverge in their non-speech imitations from the native speakers of Japanese in the early stage of L2 acquisition, the performance of the native speakers becomes more similar to that of the native speakers of Japanese as a function of L2 experience. This implies that some acoustic characteristics which are involved in the F0 patterns in Japanese speakers' productions are more difficult than others for native speakers of English to phonetically realize. These results imply that it may be more difficult for English-speaking learners to articulate acoustic differences between F0 patterns (such as how early or how late the F0 peak occurs, or how much F0 movement there is between the F0 peak and a valley, at least compared to how rapid the F0 peak movement is). It may perhaps require more motor control or subtle timing of articulatory movements in order to articulate target-like F0 peak location and F0 range, compared to the articulation of target like F0 maximum velocity. Moreover, the results suggest that as learners gain L2 experience, they can also learn to articulate phonetic differences between the three accent types in a way that is increasingly similar to the native speakers. This also implies that English-speaking learners may have an articulatory or motor control problem which may require more practice in imitating the F0 movements, especially in the initial stage of L2 learning. Since the linguistic function of F0 was controlled in the non-speech imitation task, the results seem to suggest that English-speaking learners have to learn to phonetically realize the target F0 patterns, which may involve articulatory demands which are specific to Japanese. Such articulatory or motor control difficulty might, in turn, partially contribute to the non-target-like F0 patterns in the learners' speech productions of Japanese pitch accent contrasts. This implies that the learners' articulatory ability is more closely related to their F0 profile in their speech productions as they gain more practice in L2. Hence, the non-target-like F0 profile in the productions of the learners can be explained by their ability to produce the F0 movements which yield the target pitch accent contrasts. At the same time, this does not exclude the possibility of other factors such as the difficulty in their ability to categorize the target L2 contrasts.

Lastly, we should also note that in the absence of segmental landmarks, it might have been difficult not only for learners but also for native speakers to imitate the F0 movement of F0 contours. This might also have led to the inexperienced learners' difficulty in targeting the native-like F0 peak location. It is also possible that measuring the F0 peak points only from 10 time-normalized points on the F0 contours might have caused some errors in comparing the group means or even individual variability within each group. This might also have been reflected in the inexperienced learners' slightly divergent F0 peak alignment patterns. Nevertheless, this seems to be not fatal to the points that have been discussed so far.

To conclude, it seems possible that when the linguistic function of F0 was discarded from the original nonce words, native speakers of English were capable of articulating the differences in F0 movement for the target F0 contours of non-speech stimuli, when observed in terms of F0 peak location, F0 fall and F0 range. However, it was also found that there are some differences between the patterns produced by the learners and the patterns produced by the native speakers of Japanese. Interestingly, the performance of the learners does seem to have been influenced by their L2 experience even for this non-speech task: the F0 patterns produced by the experienced learners were more similar to the native speakers of Japanese than were the F0 patterns produced by the inexperienced learners. What these results revealed, therefore, is that it is hard to exclude the possibility that the non-target-like F0 patterns in learners' speech productions comes from learners having articulatory or motor difficulties in producing the required F0 movements in the target contours. It seems therefore that there may be articulatory settings which are specific to the F0 parameters in the phonetic realization of Japanese pitch accent, and which therefore need to be learned during acquisition in order for the learner to be able to distinguish Japanese pitch accent contrasts. This also suggests that the learners' articulatory ability may improve with L2 experience. Importantly, the results also suggest that the learners' ability to produce the F0 movements in this articulatory or non-speech task seems to play some role in explaining the F0 profile of their speech productions. At this stage, it is still too early to draw wide-ranging conclusions, as the tasks involved articulatory ability where the cross-linguistic function of the acoustic correlate of F0 was controlled, so that the learners' articulatory ability was investigated in a non-linguistic context. Therefore, in order to investigate the issue further, we still need to examine learners' ability in a linguistic context where the ability to translate the shape into a linguistically meaningful unit. This is particularly worth pursuing, given that the overall F0 patterns in the non-speech productions of the native speakers of Japanese seem to show some resemblance to the same parameters in their speech productions, while this tendency among the native speakers of English was less so (and varied according to L2 experience).

### 5.7 Summary

In Experiment 3, the issue of articulatory/motor processes was explored as a possible means of accounting for the non-target-like F0 patterns in English-speaking learners' productions of Japanese pitch accent. The first question addressed in this task was to investigate whether native speakers of English were able to articulate or phonetically realize the F0 movement of target F0 contours of Japanese disyllabic nonce words in a non-linguistic context. The second question addressed in this task was to investigate whether F0 patterns produced by native speakers of English diverge from the patterns which were produced by native speakers of Japanese. The results of the non-speech imitation task show that English-speaking learners did tend to produce acoustic differences in terms of the three F0 parameters to distinguish the F0 patterns of the accent types. However, the patterns of how the learners use these F0 parameters did diverge from the patterns produced by native speakers of Japanese, particularly the inexperienced learners. Moreover, the F0 patterns for the non-speech productions of the experienced learners appeared to show more overlap with the F0 values for their speech productions compared to the inexperienced learner group. Such difference might be reflected as the non-target-like F0 patterns of English-speaking learners' speech imitations observed in Experiment 1. This difference may also reflect learners' difficulty when native speakers of English are acquiring Japanese lexical pitch accent contrasts. Thus, the possibility that articulatory/motor control problems may be a factor which contributes to the perception of foreign accent leaves some room for further discussion. In the next chapter, the issue of categorization ability (something which can allow us to look at an aspect of the ability in a linguistic context), is investigated as a means of understanding further the potential factors behind foreign accent during the acquisition of L2 prosodic contrasts.

# 6 Experiment 4: The ability to categorize Japanese lexical pitch accent contrasts

In Section 3.2, we observed that the F0 profiles produced for Japanese pitch accent by English-speaking learners diverged from the profiles produced by native speakers of Japanese. The F0 patterns in the learners' productions differed from the native speakers' patterns in terms of intelligibility scores and F0 peak location or the degree of the slope of F0 contours in the overall shape. To investigate the factor(s) which may lie behind such divergent F0 patterns, we first examined learners' ability to differentiate the F0 contours from canonical tokens of Japanese pitch accent contrasts as non-speech stimuli (in Chapter 4). The results showed that these English-speaking learners were able to hear the difference between the target L2 contrasts and to predict the membership of the target L2 categories equally well as the native speakers of Japanese. This tells us that the English-speaking learners are not completely lacking in perceptual sensitivity towards the key acoustic correlate of Japanese lexical pitch accent (i.e., F0), something which is required as part of the input process requirements to be able to produce target-like productions. In Chapter 5, we investigated the learners' ability to articulate the pitch contours, using a non-speech imitation task. The results showed that the English-speaking learners diverge from the Japanese native speakers in their non-speech imitations. This outcome suggests that the learners have a problem in articulating the key acoustic correlate of Japanese lexical pitch accent as part of the output process requirements to achieve target-like productions. However, we also need to take into consideration another requirement which learners must meet in order to achieve unaccented productions in their L2; the internal mechanism which connects the input process and the output process. More specifically, we are interested in learners' ability to categorize target L2 contrasts, and even more specifically, we think of categorization ability as being decomposed into three elements: linguistic perception, categorization and lexical assignment. We found in Chapters 4 and 5 that the English learners of Japanese do not have an acoustic perception problem but have an articulation problem. However, from what we found in Chapters 4 and 5, we still do not know whether or not the English learners of Japanese also have problems in the abilities involved in the internal mechanism that connects the input and the output process, and if so, whether or not this problem is related to their foreign accented production of Japanese lexical pitch accent. Therefore, these are the issues of interest in this chapter.

In the existing literature, the ability to categorize target L2 contrasts has been discussed as a potential interaction between learners' speech perception and a problem in the mental representation of L2 speech. The definition of what is meant by this categorization ability varies from study to study. However, for our purposes, it is defined here as the ability to perceptually differentiate Japanese pitch contrasts in a way which involves three abilities: (1) the ability to hear differences between the L2 contrasts (linguistic perception), (2) the ability to categorize the L2 contrasts (categorization) and (3) the ability to assign L2 categories to lexical items (lexical assignment). To be able to categorize L2 contrasts, the learners need to associate phonetic information with L2 categories. Hence, this potentially requires phonetic representation of Japanese lexical pitch accent categories. To be able to lexically assign the L2 category, the learners need to lexically associate an item with the corresponding L2 category. Thus, this potentially requires lexical-phonological representation of Japanese lexical pitch accent categories. Moreover, these three decomposed abilities are related to each other. To be able to categorize L2 contrasts, first, the learners need to be able to hear phonetic differences between the L2 contrasts. In addition, to be able to lexically assign the L2 category, the learners need to have phonetic representations of the target L2 categories.

The aim of this chapter is to address whether or not foreign accent involves a problem in the internal mechanism that connects the input and the output process of speech production. We intend to achieve this by investigating the learners' ability to categorize Japanese lexical pitch accent contrasts (where categorization is decomposed into the three abilities mentioned above, linguistic perception, categorization and lexical assignment) as a determinant factor for non-native-like productions. This will allow us to shed light on the question of whether or not the interaction between the learners' speech perception, phonetic representation and lexical-phonological representation of Japanese lexical pitch accent is a source of their foreign accentedness. Although it may not be possible for us to directly test this interaction or their mental representations directly, it is possible to test learners' perception ability, and to indirectly investigate their phonetic representation or lexical-phonological representation problems by testing related abilities. More specifically, we are interested in whether or not the problem lies 1) in hearing the phonetic difference between the representative members of the target L2 categories; 2) in categorizing tokens into the target L2 categories; or 3) in assigning the target L2 categories to lexical items. Three perception tasks were conducted: an ABX task using canonical tokens, an ABX task using a continuum, and thirdly a label assigning task.

The first perception experiment (an ABX task) uses canonical tokens of Japanese lexical pitch accent contrasts. This will allow us to diagnose their linguistic perception on whether or not the learners are able to hear the phonetic difference between the tokens and to classify the target item into the L2 category based on similarity judgments. The second perception experiment (another ABX task) uses stimuli on the relevant continuum. This experiment additionally allows us to test the learners' categorization ability on whether or not the learners have formed the target L2 categories. This could reflect their phonetic representation of the L2 categories since the learners need to learn to associate the relevant phonetic information with the L2 categories. The third perception experiment involves assigning labels to the tokens. This experiment further provides us with a means to examine their lexical assignment ability on whether or not the learners have problems in assigning categories to lexical items. This ability could reflect the lexical-phonological representations of the L2 categories since the learners need to learn to the arbitrary connection between each lexical item and the L2 categories in order to fulfill this task.

These three perception tasks importantly tap different aspects of categorization ability. In the ABX task using canonical tokens, the participants hear three representative members and judge the third token was the same as which one of the first two tokens representing the two target L2 categories. To be able to do the ABX task using canonical tokens, the learners have to be able to hear the phonetic difference between the target L2 categories. In the ABX task using a continuum, the participants hear two endpoints of a continuum which represent the two target L2 categories, followed by one of a stimulus on the continuum, and categorize the third stimulus into the L2 categories. To be able to do the ABX task using a continuum and show target-like responses, the learners need to have formed the L2 categories of the target contrasts. In the label assigning task, the participants hear one canonical token and allocate the corresponding diacritic as a label to each lexical item. In addition to hearing the phonetic difference between the L2 categories and having formed the L2 categories, to be able to do the label assigning task, the learners (as well as native speakers) need to learn a novel diacritic system and assign L2 categories to lexical items by learning the arbitrary connection between lexical items and the L2 categories.

It is possible that learners have problems in hearing the phonetic difference between the canonical tokens or the endpoint stimuli of the target contrasts that are the representative members of the target L2 categories. Even if the learners do not seem to have problems in hearing the difference between the canonical tokens, they may still have problems in categorizing the intermediate stimuli on a continuum. This might be because the learners have not formed the target L2 categories at all, or because although they have partially formed the L2 categories these categories may nevertheless overlap each other due to the relation between the L2 categories and the phonetic information not being fully learned. It is also possible that these partially formed L2 categories overlap with L1 categories. Hence, this task has a potential to provide us with a diagnosis of the learners' phonetic representation problem of the L2 categories. Meanwhile, if the results of the ABX tasks show that learners don't have such a problem, but if at the same time they have a problem in the labeling task, this may suggest that their problem lies in assigning L2 categories to the target lexical item. This could reflect that they have not learned, or have not fully learned, the association between the L2 categories and the lexical items. Alternatively, the reason why they fail to do the labeling task might be problems in learning the labels. As a further alternative, it might be a combination of both.

Since there are no explicit labels for Japanese lexical pitch accent, the labeling task is new even to the native speakers of Japanese. Despite this, a pilot study showed that Japanese speakers performed relatively well in learning the novel diacritic system and choosing the labels that matched the target category. Meanwhile, the experienced learners performed better and also closer to the performance of native speakers than the less experienced learners. As native speakers of Japanese also had to learn the diacritics, the results of the pilot study imply that the group difference could be attributed to the difficulty in allocating the L2 categories to each lexical item. Native speakers can be expected to have formed the categories of the target contrasts and to have fully developed phonetic representation and lexical-phonological representation, whereas learners have to learn to form the native-like categories and develop these representations as they learn the L2. Hence, the label assigning task has potential for testing the learners' problem which is related to lexical-phonological representation of the L2 categories.

In this chapter, we aim to answer the following questions: 1) whether or not native speakers of English are able to hear the difference between canonical speech tokens of Japanese lexical pitch accent categories and to classify the target item into the correct L2 category; 2) whether or not native speakers of English are able to hear the difference between the stimuli on a F0 peak alignment continuum of Japanese lexical pitch accent contrasts and categorize each stimulus into the target L2 category; 3) whether or not native speakers of English are able to learn a novel labeling system for Japanese lexical pitch accent like native speakers of Japanese and to assign a label to canonical tokens of Japanese lexical pitch accent contrasts. By answering these three questions, we intend to address: whether or not the problem lies 1) in hearing the phonetic difference between the members of the target L2 categories; 2) in categorizing the target L2 categories by associating the phonetic information with the target L2 categories; or 3) in assigning L2 categories to lexical items by learning the connection between the category and the item.

First we will explain the details of each of the three tasks: ABX task with canonical tokens, ABX task with 7 step stimuli on a continuum, and labeling assigning task. We will then report the results of each task, and finally, we will discuss the findings of these tests of categorization ability in relation to foreign accent.

### 6.1 ABX task with canonical tokens

First, we carried out an ABX task using canonical tokens to investigate whether or not the learners are able to hear the difference between the tokens and to classify the target item into the L2 category.

#### 6.1.1 Participants

The same 48 participants who took the previous tasks were tested. They belonged to the same three groups as before: two learner groups (inexperienced and experienced) and a control group of native speakers of Japanese.

#### 6.1.2 Materials

Audio materials for this ABX task were the same as the materials used in the production task described in 3.2.2. The audio materials were created from the recording of 12 disyllabic nonce words (4 nonce words (*mene, mani, nime* & *noma*) x 3 accent types) and a post-positional particle *mo* (also), all produced by a female Tokyo Japanese speaker. There were 48 stimuli (12 nonce words x 4 presentation orders (ABA, ABB, BAA and BAB). Her productions were recorded in a sound-treated recording studio at the Linguistics laboratory at the University of Edinburgh. The audio interface was MOTU 828 mkll Firewire using a hypercardoid microphone (AKG CK 98) and digitalized at a sampling rate of 48 kHz with a resolution of 16 bits. Sonar 4 Studio Edition was used as a recording software programme. This was the case for all recordings. A list of the 12 test items was read aloud in a carrier phrase (i.e., *Sumimasen ____mo kudasai* (Excuse me but please give me ____ as well)) three times. Subsequently, one of the three repetitions of each item including the particle *mo* 'also' alone was edited to form audio stimuli. All the audio materials were converted to a sampling rate of 22 kHz with a resolution of 16 bits using CoolEdit Pro 2.1 so that it matched the sound file specification in E-Prime.

### 6.1.3 Procedure

Participants performed the task individually in a sound-treated studio. The participants heard three audio stimuli, A, B, and X, in that sequence through a pair of headphones. Stimuli A and B were always canonical speech tokens, and X was identical either to A or to B. Thus, the participants' task was to tell if the last stimulus (X) was the same as the first stimulus (A) or the second one (B) by pressing one of two keys on a keyboard. They were asked to guess if uncertain. The stimuli were presented in random order. Once the participants pressed the key on the keyboard to indicate their decision, the next set of stimuli were presented. The inter-stimulus interval was 0.7 sec as in the previous tasks.

### 6.1.4 Results

The mean percentage and standard deviation of correct responses was calculated for the three groups are given in Table 1.

**Table 6.1.The mean percentage of times that the participants matched the 3rd stimulus to the correct member of the AB pair.**

| Group | Mean (%) | SD |
|---|---|---|
| NJ | 89 | 10 |
| NEexp | 88 | 8 |
| NEinexp | 84 | 10 |

All three groups did well in this ABX task. The main effect of Group was not significant [$F$ (2, 42) = 1.20, $p$ = 0.31]. Hence, there was no evidence of any difference in hearing differences between pitch accent contrasts in either learner groups.

### 6.1.5 Discussion

The result suggests that English-speaking learners do not differ from native speakers in hearing differences of canonical tokens of Japanese lexical pitch accent contrasts. These learners are able to hear the phonetic difference between the target contrasts when the stimuli are representative members of the L2 categories. Thus, this implies that the learners' divergent F0 patterns cannot be attributed to their ability to categorize the target contrasts.

We should note that the skewness and kurtosis of the distributions of the NJ group indicated that there may have been a ceiling effect (-1.23 and 1.64 respectively). Hence, this seems to suggest that the task was too easy for the native speakers, and this may be why the difference between the native speaker group and the two learners' groups was found to be statistically non-significant.

However, it is possible these learners may have problems in a perception task where they have to hear the difference between and predict membership of more diverse stimuli regarding the relevant acoustic information that distinguishes the target L2 categories than can be captured using this task. Responses to such task could reflect whether or not the learners have learned to form the target L2 categories by associating the relevant phonetic information with the L2 categories. For this reason, the learners' categorization ability was tested further with an ABX task using 7-step stimuli where they have to categorize whether the target stimulus belongs to the category A or B, and classify each token into the corresponding category.

## 6.2 ABX task with stimuli on a continuum

To categorize the target L2 categories like native speakers, learners seems to have two options; either to compute the similarities 1) by comparing the target stimulus with the presented stimuli using their working memory, or 2) by comparing the target stimulus with their phonetic representations of the L2 categories. However, since computing similarity judgments places a heavy load on learners, which may limit their working memory capacity, they may either be forced to rely on the second option i.e., to use their phonetic representations, or else they may fail to categorize the target stimulus at all (since they may not have formed target-like L2 categories). This may possibly be because the learners have not formed the target L2 categories at all, or because although they have formed the L2 categories these categories overlap each other due to poor association between the phonetic information and the L2 categories. Hence, a step-wise ABX task seems to provide us with a way of testing the learners' phonetic representation problems.

To investigate whether or not the learners are able to hear the difference of the stimuli and categorize Japanese lexical pitch accent categories like native speakers, we compared the response patterns of the two learner groups with the pattern of the native group. First, we investigated whether or not the learners show a native-like categorical function in their responses to the step-wise ABX task. If they do not show a categorical function, this may indicate that the learners could only predict the membership of the representative members (as indicated by the results of the ABX task with canonical tokens) but not that of the stimuli in the boundary areas. Alternatively, it is possible that some learners who cannot show any indication of a categorical function. In this case, these learners may not have learned that there are Japanese lexical pitch accent categories. If they show a categorical function, we could infer that the learners are able to categorize the L2 categories based on their interlanguage categories.

Then, to compare the patterns of the learners with the native speakers further, we will analyze the location of the categorical boundary and the degree of slope in the pattern of a categorical function. The categorical boundary indicates the boundary area between the categories. In the case of native speakers, the categorical boundary is located somewhere between the two endpoint stimuli of the response pattern. Since the learners are learning the association between the relevant phonetic information and the L2 categories, it is possible that the categorical boundary in the learners' responses may occur in a different location from the native speakers' responses. The degree of the slope suggests how sharp the distinction between two categories is. The sharper the slope in a categorical function is, the more seperate the two categories are. For the same reason as the categorical boundary, learners may show a shallower slope, which would indicate that the categories that they may have formed are overlapping each other.

### 6.2.1 Participants

The same 48 participants who took the previous tasks were tested, in the same three groups: two learner groups (inexperienced and experienced) and a control group of native speakers of Japanese.

### 6.2.2 Materials

Audio materials for this ABX task were created by using one recording of *mene* (A1) tokens accompanied by a post-positional particle *mo* (also), produced by a female speaker of Tokyo Japanese. This recording was one of the canonical tokens used in the ABX task with canonical tokens. By manipulating the F0 peak alignment of the *menemo* (A1) token, 7 step stimuli on a F0 peak continuum were created. This time, the first two stimuli among the three were always either one of the end point stimuli on the *mene* (A1)-*mene* (A2) continuum followed by one of the seven stimuli on the continuum. There were 70 stimuli (7-step stimuli x 5 repetitions x 2 presentation orders (ABX and BAX). The details of the F0 peak manipulation are as follows.

A 7-step F0 peak continuum was created by shifting the location of the F0 peak alignment of the *menemo*(A1) token, where each end has its F0 peak either at the midpoint (50%) of the first vowel (V1) or the midpoint of the second vowel (V2). These values were selected based on a pilot study which showed that in the initial accented disyllabic words, the F0 peak occurred somewhere between 42% of V1 and the offset of C2/the onset of V2: in the case of words with pitch accent on the second syllable, it occurred somewhere in the range from 57% of V2 and up to 29% of the following syllable (the Japanese particle "*mo*")[1]. Then, 7 steps calculated from the log values were created by shifting the pitch contour of the A1 token horizontally towards the A2 target based on the calculated F0 peak location while fixing both ends of the F0 contour. Overall, the stimuli created in this manner sounded more natural than the reverse case, where the F0 contour of the A2 token was shifted horizontally towards the A1 target (see Diagram 6.1). Generally speaking, the stimuli created by shifting the entire F0 contour without fixing the ends and without changing the contour itself also sounded less natural compared to the selected method. When shifting F0 peak location, all the stylized points apart from the two end points were shifted rightward horizontally. In addition, a few stylized points next to the end

---

[1] The recordings of the pilot study were segmented and the F0 trajectory was time-normalized using the MOMEL script (Hirst & Espesser, 1991) with a modification of the Praat implementation by Remijsen (2004). The MOMEL script automatically detects the F0 turning point of the contour. Thus, it allows us to find where the F0 peak alignment occurs in each utterance relative to the segmental landmarks.

points were adjusted if necessary so that the contour would fit within the two end points and did not create another peak. During this process, the degree of the slope between the F0 peak and the following valley was preserved in its original shape as much as possible and modified as little as possible. Hence, it was not only the location of the F0 peak which was manipulated in the stimuli used in this task, but also the degree of the slope after the F0 peak was slightly modified. Each log value and absolute value of F0 pitch peak location referring to the absolute time scale is provided below (see Table 6.2).

**Diagram 6.1 Schematic illustration of the 7-step stimuli.**

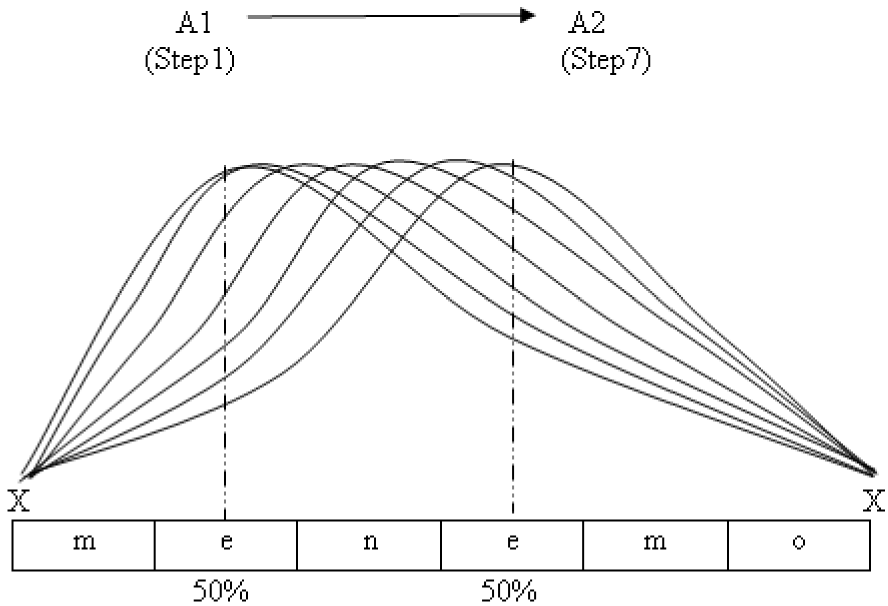


**Table 6.2. Log value and absolute value of the location of the F0 peak alignment for each step.**
The absolute time "0" indicates the onset of each sound file.

| | Step1 | step2 | step3 | step4 | step5 | step6 | Step7 |
|---|---|---|---|---|---|---|---|
| Log | -1.715 | -1.503 | -1.397 | -1.291 | -1.185 | -1.079 | -0.973 |
| Abs | 0.180 | 0.223 | 0.247 | 0.275 | 0.306 | 0.340 | 0.378 |

Once the F0 peak location was manipulated, the stimuli were synthesized with the PSOLA method (Pitch-Synchronous Overlap and Add) in Praat. The naturalness of the audio stimuli was verified with five native speakers of Japanese, who agreed 100% that the end stimuli could be heard as the intended stimuli.

### 6.2.3 Procedure

Participants took the task individually in a sound-treated studio. The participants heard three audio stimuli, A, B, and X, in that order using a pair of headphones. Then, the task was to state whether the last stimulus (X) was the same as the first stimulus (A) or the second one (B) by pressing one of two keys on a keyboard which were clearly indicated with stickers. They were asked to guess if uncertain. The stimuli were presented in random order. Once the participants pressed the key on the keyboard to indicate their decision, the next set of the stimuli were presented. The inter-stimulus interval was 0.7 sec, the same as the previous experiment.

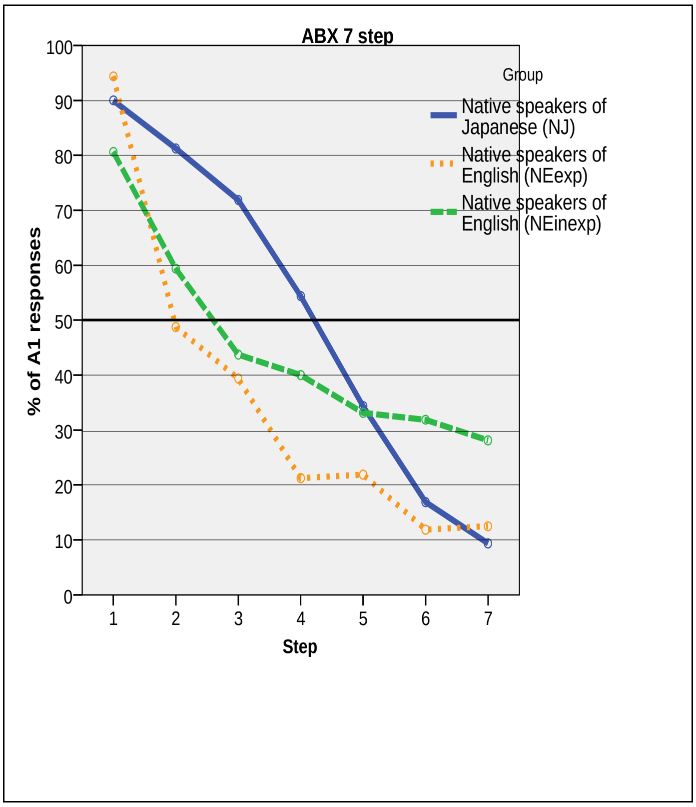
### 6.2.4 Results

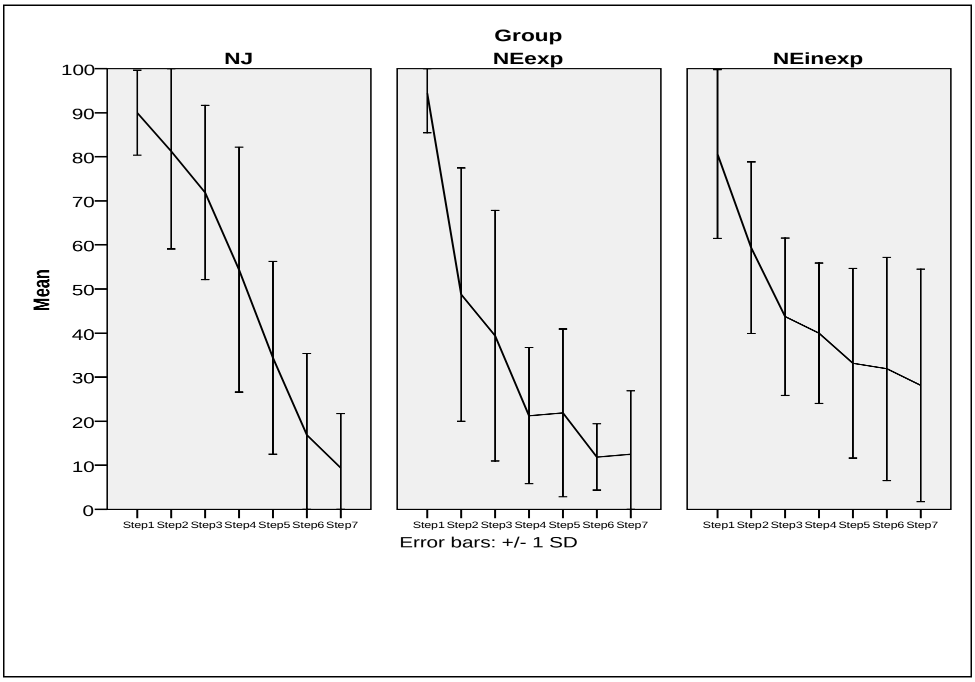
Here is a brief overview of the result of the step-wise ABX task. To investigate whether or not the learners are able to predict membership of Japanese pitch accent categories like native speakers, we compared the response patterns of the two learner groups with the pattern of the native group. First, we investigated whether or not the overall responses to the stimuli of the learners show a native-like categorical function in the responses. Since the two learner groups did show a categorical function in a similar direction to the native speakers but differences in the intermediate stimuli, we then report the location of the categorical function and the degree of the slope of their response patterns. In addition, there were eight participants (mainly from among the inexperienced learners) who did not show a categorical function. Their patterns will be reported and the results of the step-wise ABX task will be revisited without these participants. Revisiting the results after excluding these non-categorical participants resulted in the inexperienced group showing slightly more native-like categorical function than their initial pattern.

The mean percentage and the standard deviation of A1 responses over the 7-step stimuli to each step for the three groups are presented in Figure 6.1. Larger percentages indicate that the stimulus was identified more often as A1 and less often as A2. Hence, the point at the top left corner indicates that the participants classified a stimulus as A1, whereas the point at the bottom right corner indicates that they classified it as A2. The overall patterns of all three groups shift from A1 responses to A2 responses. This suggests that the learner groups also seem to show the categorical function. However, we also see some differences in the overall shape of the group mean of the two learner groups from that of the native speaker group, especially for the responses to the intermediate stimuli. We will now turn into the details of the similarities and the differences in their group patterns.

**Figure 6.1. The mean percentage of A1 responses to each step.**

The x-axis indicates the number of the step on the F0 peak continuum. The y-axis indicates the percentage of *mene* (A1) responses. Each line indicates the mean percentage of A1 responses among each group. The vertical lines in the bottom figures indicate +/- 1 standard deviation.



As can be seen in Figure 6.1, the experienced English-speaking learners were able to categorize both end stimuli of the F0 peak continuum as well as the native speakers of Japanese. This result also agrees with the result of ABX task using the canonical tokens, presented in the previous section (6.1.4), where the learners did not seem to have a problem in categorizing these tokens. Not only the experienced learners but also the inexperienced learners showed good categorization ability for canonical tokens, as we saw in the previous section (6.1.4). It seems that the inexperienced learners were also able to categorize the stimulus at one end of the continuum (i.e., Step 1) like Japanese speakers, whereas they did less so with the stimulus at the other end (i.e., Step 7). Moreover, the overall shape of the group mean of the two learner groups is different from that of the native speaker group. The inexperienced learner group seems to show a less categorical function, suggested by the shallower slope, compared to the native-speaker group and the experienced learner group. This implies that the categories of the inexperienced learners overlap each other. The experienced learner group seems to show a steeper slope, suggesting more target-like categorical function than the inexperienced learners, but their patterns are nevertheless not quite the same as the native speaker group. The categorical boundary of NJ, indicated by 50% cross-over point, is located somewhere near the Step 4 stimulus, i.e., the middle of the

A1-A2 continuum. However, the categorical boundary of both learner groups seems to fall between the Step 2 and the Step 3 stimulus, which is located closer to the A1 end of the continuum. This suggests that the learners are associating the relevant phonetic information of A1 pitch accent (F0 peak location), which the native speakers treat as A1, with the A2 pitch accent category.

In order to statistically compare the learner groups with the native speaker group in the percentage of *mene* (A1) responses, a mixed design ANOVA test was carried out. The between-subject factor was Group (3 levels) and the within-subject factor was Step (7 levels). There was a significant effect of Step [$F$ (6, 37) = 67.52, $p$ < 0.001]. The effect of Group was also significant [$F$ (2, 42) = 7.28, $p$ < 0.01]. The interaction between Step and Group was significant [$F$ (12, 74) = 4.81, $p$ < 0.001]. This indicates that the performance pattern of the groups was different over the steps. Post-hoc comparisons using the Tukey test showed that the percentage of A1 responses of NEexp ($M$ = 35.7%) was significantly smaller than NJ ($M$ = 51.2%). In other words, NEexp were more likely than native speakers to identify the stimuli on the F0 peak alignment continuum as A2. However, the mean percentage of A1 responses of NEinexp ($M$ = 45.3%) was not significantly different from that of NJ, although this may be surprising since the difference between the NEinexp and the NJ groups seemed substantial from Figure 6.1. It is possible that the reason for this is because of the large variance in the NEinexp's data. In addition, the two learner groups' means were not statistically different from one another.

In order to further examine the Group x Step interaction, one-way ANOVA tests were carried out to compare the average response patterns of the three groups on each step. Overall, what we saw in Figure 6.1 was borne out: the learners' patterns were similar to the NJ group on the stimuli at the extremes of the continuum but different on the intermediate stimuli. The results of the ANOVA tests showed that the learners' responses tended not to be significantly different in the end stimuli (and some nearer to the end) when compared with the native speakers' responses. More specifically, the average NEexp responses on Step 1 ($M$ = 94.4, $SD$ = 8.9), Step 5 ($M$ = 21.9, $SD$ = 19.1), Step 6 ($M$ = 11.9, $SD$ = 7.5) and Step 7 ($M$ = 12.5, $SD$ = 14.4) were similar to those of the NJ group (Step 1: $M$ = 90.0, $SD$ = 9.7, Step 5: $M$ = 34.4, $SD$ = 21.9, Step 6: $M$ = 16.9, $SD$ = 18.5, Step7: $M$ = 9.4, $SD$ = 12.4). However, the NEexp group categorized the intermediate stimuli, i.e., Steps 2, 3 and 4 significantly more as A2 compared with the NJ group, (Step 2: NJ: $M$ = 81.3, $SD$ = 22.2, NEexp: $M$ = 48.8, $SD$ = 28.7; Step 3: NJ: $M$ = 71.9, $SD$ = 19.7, NEexp: $M$ = 39.4, $SD$ = 28.4; and Step 4: NJ: $M$ = 54.4, $SD$ = 27.8, NEexp: $M$ = 21.3, $SD$ = 15.4).

On the other hand, the average NEinexp responses on Step 1 ($M$ = 80.6, $SD$ = 19.1), Step 4 ($M$ = 40.0, $SD$ = 15.9), Step 5 ($M$ = 33.1, $SD$ = 21.5) and Step 6 ($M$ = 31.9, $SD$ = 25.4) were similar to those of the NJ group, but this was not the case for one of the end stimuli, Step 7 (*M* = 28.1 *SD* = 26.4), or for the intermediate stimuli Step2 (*M* = 59.4 *SD* = 19.5) and Step3 (*M* = 43.8, *SD* = 17.8).

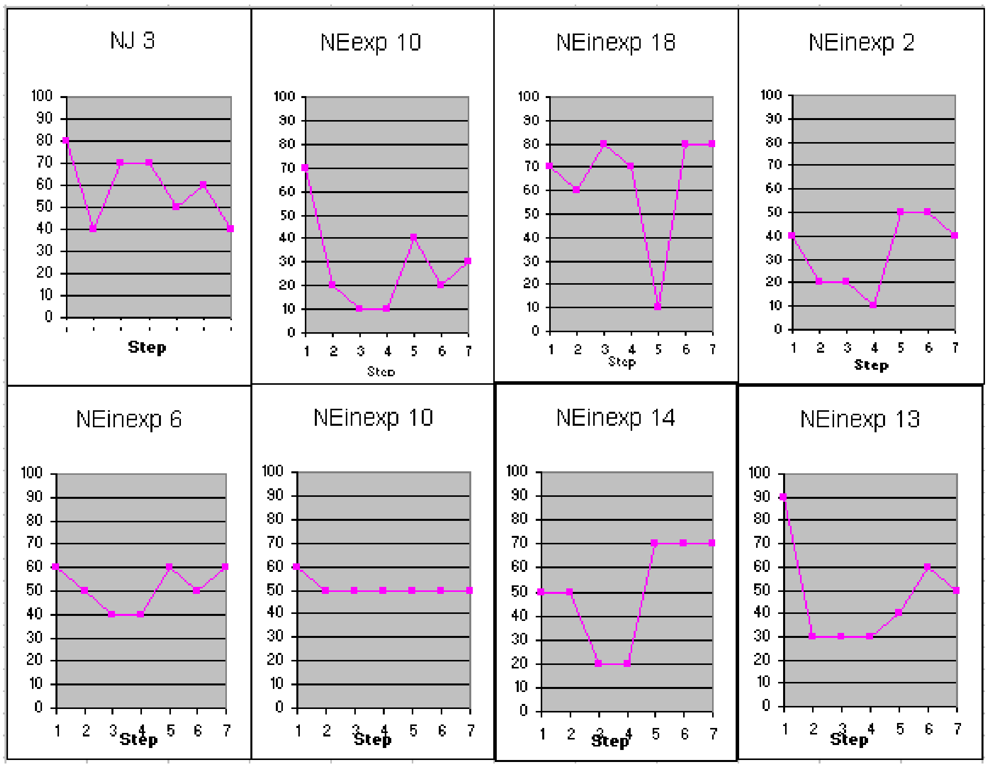
There were also differences in the patterns between the two learner groups. There were significant differences in the responses to the Step 1, Step 4 and Step 6 stimuli. The NEexp group categorized the Step 1 stimulus more as A1 compared to the NEinexp group and the Step 4 and Step 6 stimulus more as A2. In addition, the NEexp group categorized the Step 7 stimulus nearly significantly more as A2 than the NEinexp group ($p = 0.058$).

Overall, this indicates that the learner groups performed similarly to the NJ group on the stimuli at the ends of the F0 continuum but differently on the intermediate step stimuli. In particular, the learners tended to categorize as A2 the intermediate stimuli which the native speakers of Japanese categorized as A1. Moreover, the results indicate that the inexperienced learners tended to show less categorical performance compared to the native speakers and the experienced learners. Note however that this data will be revisited later in the section, although as will be explained, the overall picture will not change substantially (except in one respect, which will be pointed out).

Now let us look at the result of the category boundary and the slope of the categorical function in the response patterns. To find out statistically whether or not the learners' categories are overlapping each other and more inclined to the A2 category than the A1 category, probit analysis was conducted. Among all of 48 participants, 8 participants (1 NJ, 1 NEexp and 6 NEinexp) were excluded from the probit analysis because their performance did not show 50% categorical cross-over from one response category to the other (i.e., they showed chance-level performance, the 'reversal' pattern, or a random pattern, as shown in Figure 6.2). Most of these participants, i.e., six of the eight participants, were inexperienced learners.

**Figure 6.2. ABX individual performance for participants who were excluded from the probit analysis.**

Each panel indicates an individual participant's performance. The x-axis indicates the number of the step on the F0 peak continuum. The y-axis indicates the percentage of *mene* (A1) responses. Each line indicates the average individual percentage of A1 responses.




The means and standard deviations of the slope and the category boundary (CB) of each group are provided in Table 6.3. Once those eight participants were excluded, the probit analysis revealed that the degree of the slope of the categorical function in the performance of NEexp and NEinexp were not significantly smaller than that of NJ. Moreover, the analysis also showed that the category boundary location of the two learner groups was not after all significantly shifted to the A1 category compared to that of NJ. This suggests that if the learners show the categorical function, their performance is not statistically different to native speakers in terms of the slope and the category boundary of the function. However, it is important to note that most of the participants who were excluded from the probit analysis were the inexperienced learners. This still therefore indicates that the learners seem to have a problem with categorization that involves with phonological representation of the association between the phonetic information and the L2 categories, particularly when their L2 experience is limited.

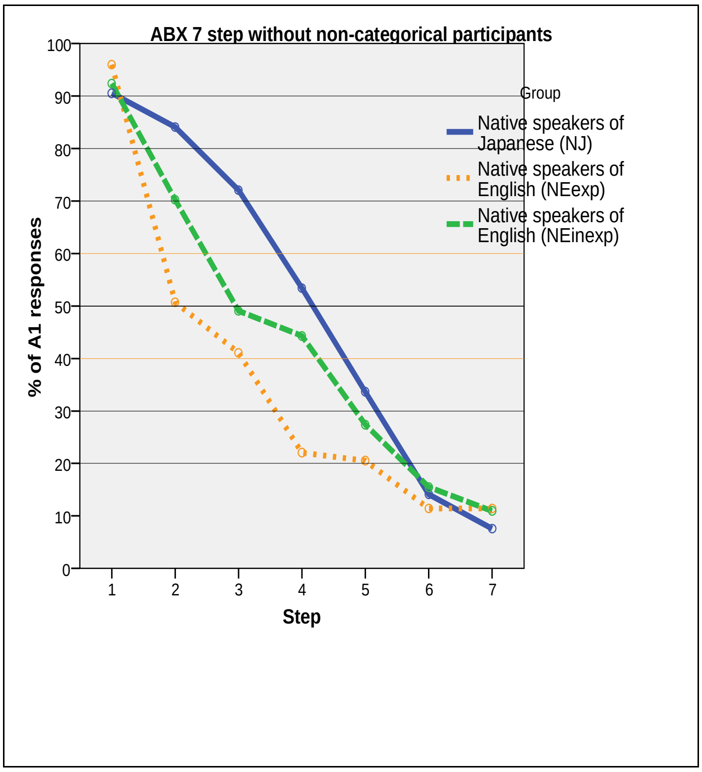
**Table 6.3. Mean and standard deviation of the slope and category boundary (CB) among the three groups.**

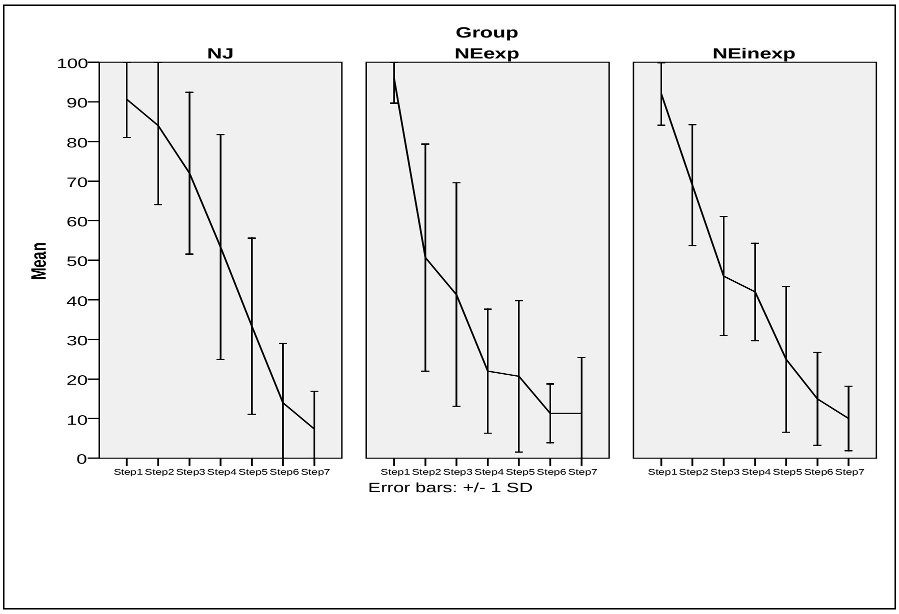
| | Slope | | CB | |
|---|---|---|---|---|
| | Mean | SD | Mean | SD |
| NJ | -0.55 | 0.45 | 3.54 | 0.92 |
| NEexp | -0.59 | 0.64 | 2.95 | 1.12 |
| Neinexp | -0.26 | 0.25 | 3.44 | 0.53 |

Following the results of the probit analysis, it is now important to investigate whether or not the results would differ in the initial analysis once these eight participants were excluded who did not show a categorical function regarding the F0 peak alignment continuum. Hence, the data from the ABX task with the 7 step stimuli were reanalyzed excluding these participants. Figure 6.3 provides the mean percentage of A1 responses among the three groups after excluding these eight participants (1 NJ, 1 NEexp and 6 NEinexp). These results suggest that the initial results were indeed influenced by the responses of these eight participants. As mentioned earlier, six out of eight participants who were excluded were the NEinexp learners. For this reason, we see the difference in the pattern of the inexperienced learner group. In Figure 6.3, we see a more categorical performance from the NEinexp than was seen in Figure 6.2 above. To test this difference statistically, a mixed design ANOVA test (Group x Step) and one-way ANOVA tests (Group) were conducted.

**Figure 6.3 Mean percentage of A1 responses to each step, excluding eight participants.**

The x-axis indicates the number of the step on the F0 peak continuum. The y-axis indicates the percentage of *mene* (A1) responses. Each line indicates the mean percentage of A1 responses from each group. The vertical lines in the bottom figure indicate +/- 1 standard deviation.



A mixed design ANOVA test, where the between-subject factor was Group (3 levels) and the within-subject factor was Step (7 levels), showed the same tendency as in the initial analysis. There was a significant effect of Step [$F$ (6, 29) = 193.89, $p < 0.001$]. The percentage of A1 responses significantly decreased as the F0 peak alignment shifted from the A1 target to the A2 target. The effect of Group was also significant [$F$ (2, 34) = 5.45, $p < 0.01$]. The interaction between Step and Group was found to be significant [$F$ (12, 58) = 2.58, $p < 0.01$]. This indicates that the performance pattern of the groups were different over the steps. Post-hoc comparisons using the Tukey test showed that the percentage of A1 responses of the NEexp ($M$ = 36.2%) was significantly smaller than NJ ($M$ = 50.8%). This means that the NEexp tended to identify the stimuli on the F0 peak alignment continuum as A2. Meanwhile, the percentage of A1 responses of NEinexp ($M$ = 44.3%) was not significantly differently from that of NJ. However, both learner groups' means were again not statistically different from one another.

On the other hand, one-way ANOVA tests on each step among the three groups revealed some differences which seem to reflect some different response patterns in the NEinexp between Figure 6.2 and Figure 6.3. The average response patterns of the NEinexp became more similar to those of the NJ once the six inexperienced learners who showed non-categorical responses were excluded. A significant difference between the NEinexp and the

NJ was only found at an intermediate stimulus, the Step 3 stimulus. The NEinexp group categorized the Step 3 stimulus more as A2 (*M* = 46.0, *SD* = 15.1) than the NJ group (*M* = 72.0, *SD* = 20.4). Meanwhile, the results of the NEexp showed the same tendency; their responses were similar to the NJ's for the end stimuli (Steps 1, 5, 6 and 7) whereas the NEexp categorized the intermediate stimuli (i.e., Step 2, 3 and 4) more as A2 than the NJ. We should also note that the difference between the two learner groups also disappeared once the eight participants were excluded. This shows that the response patterns of the inexperienced learners became closer to those of the native speakers once we excluded the outliers who did not show categorical function in the step-wise ABX task. Nevertheless, their response patterns were also not distinguishable from those of the experienced learners.

Considering the responses to the end stimuli on the F0 peak continuum, there is another important point to make. The NEexp group did as well as in the ABX task using the canonical tokens. The end stimuli on the continuum are supposed to be equivalent to canonical tokens of the A1 and A2 tokens of the Japanese pitch accent types.[2] The results of the two tasks agree in the case of the NEexp group. The NEinexp group also performed

---

[2] The stimuli were created by shifting the F0 peak location of the contour of the A1 token horizontally towards the A2 target from the midpoint (50%) of the first vowel and the midpoint of the second vowel (see 6.2.2). For this reason, strictly speaking, the two end stimuli were not exactly the same as the canonical tokens.

similarly in categorizing the Step 1 stimulus as they did in categorizing the canonical tokens. However, the NEinexp group performed differently from the NJ group in their categorization of the other end-point stimulus, i.e., Step 7. The NEinexp responded to Step 7 as A2 less than the NJ. Hence, there was a mismatch in the results between the two tasks when the NEinexp group is considered. Interestingly, once the eight participants who did not show a categorical function in the ABX task using the 7-step continuum were removed for the follow-up analysis, the difference between the NEinexp and NJ group in the responses to the Step 7 stimulus also disappeared. For this reason, it was reinvestigated whether or not the eight participants who were excluded from the follow-up analysis also performed poorly in categorizing the canonical tokens of the target stimuli (i.e., the ABX task with the canonical tokens). The individual data is provided in Table 6.4. Among these people, five participants show relatively good categorization ability of the canonical tokens (i.e., within one standard deviation from their group mean). Only three participants including one NJ did poorly on the task (NEinexp6, NEinexp18 and NJ3). It may be that the very poor performance of the two NEinexp learners influenced the NEinexp group result. Apart from these three participants, overall, most of these eight participants were not doing poorly in categorizing the canonical tokens of Japanese pitch accent contrasts. One-way ANOVA tests (3 levels) also indicated that there was no significant difference among the three groups when these eight participants were excluded. As far as the result of this ABX task is concerned, NJ3 seems to be an outlier as a native speaker. Thus, another one-way ANOVA test (3 levels) was carried out without his data. However, this also did not bring any statistical difference among the three groups [$F$ (2, 40) = 1.46, $p$ = 0.25]. These results are the same as the whole data with these participants. As this suggests that these participants were not necessarily unable to categorize the canonical tokens, the mismatch result between the two ABX tasks might have also related to at least partially the difficulty of this task.

**Table 6.4. Individual scores on the ABX task using the canonical tokens.**

| Group | NJ | NEexp | NEinexp | | | | | |
|---|---|---|---|---|---|---|---|---|
| Participant | 3 | 10 | 18 | 2 | 6 | 10 | 14 | 13 |
| % | 62.5 | 83.3 | 64.6 | 72.9 | 66.7 | 89.6 | 97.9 | 79.2 |

### 6.2.5 Discussion

The results of the previous task, the ABX task using the canonical tokens of Japanese pitch accent contrasts (6.1), showed that the ability to categorize the target canonical tokens did not seem to be a problem for the English-speaking learners. The result of the current task, the ABX task using the 7-step stimuli, also partially agreed with the result of the ABX task using the canonical tokens, in that the same learners did not have a problem in identifying the end stimuli of the F0 peak continuum in this task. The inexperienced learners initially showed a mismatch between the two tasks regarding the end stimuli. However, once the six participants who did not show a categorical function in the step-wise ABX task were excluded from the inexperienced group, the group's results agreed with those of the ABX task with canonical tokens. The follow-up analysis indicated that this might have been due either to the two inexperienced learners who performed poorly on the ABX with canonical tokens, or the task difficulty of the 7-step ABX task, considering that the inexperienced learners did not have problems in categorizing canonical tokens. As far as the endpoint stimuli are concerned, the learners in both groups did not seem to have a problem in categorizing the target tokens, although the results also revealed that some of the inexperienced learners were having a problem in categorizing the endpoint stimuli.

The results of the step-wise ABX task using the F0 peak continuum (6.2) also indicated that the English-speaking learners tend to have problems in categorizing the intermediate stimuli. This may be related to their phonetic representation of the Japanese lexical pitch accent categories. The finding that the learners tended to identify these intermediate stimuli more often as A2 than as A1 possibly suggests that the learners had associated with A2 categories an earlier F0 peak location than the native norms. Additionally, the less categorical function of the inexperienced learners compared to native speakers seems to suggest that the learners' A1 and A2 categories are overlapping each other. Interestingly, when non-categorical participants were excluded, the mean response patterns of the inexperienced learners became more similar to those of the native speakers of Japanese. However, the inexperienced learners' response patterns were not different from those of the experienced learners whose response patterns were different from the native speakers'. This suggests that those who showed a categorical function perform similarly to the native speakers of Japanese in terms of the ability to associate the F0 information with the L2 categories, while also leaving open the possibility that the inexperienced learners are still learning to develop this categorization ability. This view is further supported by the fact that the most of those who were excluded for not showing the categorical function were inexperienced learners. These results suggest that when native speakers of English are required to categorize stimuli for which they need to access phonetic representations of Japanese lexical pitch accent, they perform as well as the native speakers of Japanese, but these results also suggest that they only become capable of doing so with increased L2 experience. Nonetheless, the results also implied that it is still difficult for the learners to associate the phonetic information with the L2 categories to form the target L2 categories.

The reason why more inexperienced learners seem to have had problems with the step-wise ABX task could be that (due to lack of L2 experience) they had not fully developed the phonetic representation of Japanese lexical pitch accent categories that associates the phonetic information and the categories. Because these under-developed target L2 categories in their phonetic representation will be a poor resource for the learners to refer to, it may be that the inexperienced learners had no choice but to directly compare the similarity between the presented stimuli and the target stimulus, as this does not require them to access phonetic representations of the L2 categories. Meanwhile, the native speakers were able to access their phonetic representations more easily and to categorize the stimuli into the L2 categories (assuming that the categories of the target contrasts are fully formed as part of their phonetic representation, given that they are native speakers of the target language). This also seems to be why the learners had problems in categorizing intermediate step stimuli. As the intermediate stimuli were acoustically different from the representative tokens, it may be that this made it more difficult for the learners to do the comparison than the endpoint stimuli. This could be explained by the suggestion that they might have failed to achieve native-like categorization. Alternatively, it is possible that the learners' interlanguage categories of Japanese pitch accent, which have been formed by associating the F0 peak information and the L2 categories, largely overlap each other. For this reason, their responses to the stimuli that fall within the overlapped area were not clearly distinguishable one from the other.

What these results suggest is that the native speakers of English may still have a problem in learning the association between the F0 peak information and Japanese lexical pitch accent categories while developing the target L2 categories. From this point, we begin to have evidence in favor of the possibility that the learners' problem may lie in their categorizing ability, as an explanation for the divergent F0 patterns in their productions. In order to produce the native-like F0 profile, L2 learners may first need to form the target L2 categories, by associating the relevant acoustic information of Japanese pitch accent with the L2 categories. Therefore, this test of categorization ability can potentially be a reasonable factor to indicate foreign accent.

### 6.3 Label assigning task

The results of the two ABX tasks have showed that English-speaking learners are able to hear the differences between the target categories and to group the Japanese pitch accent contrasts. However, they are having some problems when it comes to the ability to categorize the L2 categories, as required in identifying the 7-step stimuli, where the learners have to have a formation of the target L2 categories. This ability possibly requires phonetic representation of the L2 categories since the learners need to associate the phonetic information with the L2 categories. However, what we still don't know from the results of these two tasks is whether or not this also means that the English-speaking learners have learned to be able to assign L2 categories to lexical items. For lexical assignment of Japanese lexical pitch accent categories, the learners need to learn both the connection between the target categories and the relevant phonetic information, and also the connection between the L2 categories and lexical items. If this is correct, then this presumably requires L2 experience. Even if the learners have not learned to form the categories of Japanese pitch accent, they may still be able to hear the differences between samples belonging to the different categories by directly comparing the presented stimuli. Other learners might be able to categorize Japanese lexical pitch accent categories. Nevertheless, the same learners may have a problem in assigning a label to the L2 categories, or being unable to lexically associate an item with the L2 categories due to poor lexical-phonological representation of the target categories. To investigate this further, the next task, i.e., a label assigning task, was conducted.

This task was originally designed for the purpose of eliciting productions of Japanese pitch accent contrasts, i.e., by indicating the required accent types through a system of diacritics. To elicit productions from participants, our challenge was to find a way to present Japanese pitch accent with some kind of visual cue. We selected one set of diacritics for this study based on a pilot study (see 6.3.2). As this was a new diacritic system for all participants, the label assigning task was developed through the process of designing training for participants to learn the set of diacritics so that they would match each diacritic with the target pitch accent category which it represented. Unfortunately, after piloting, we reached the conclusion that the task was not suitable for collecting learners' production data. However, it brought us a very insightful dimension as a by-product: a dimension which we now use as a measure to investigate the degree to which the L2 learners' ability to assign L2 categories to lexical items.

### 6.3.1 Participants

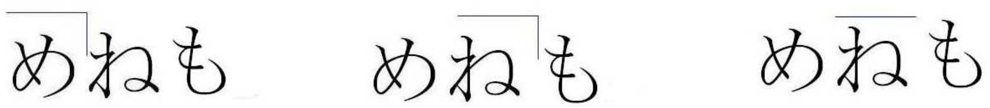
The same 48 participants who took the previous tasks were tested, in the same three groups as before: two learner groups (inexperienced and experienced) and a control group of native speakers of Japanese.

### 6.3.2 Materials

Audio materials for the labeling task were created from the recording of 12 disyllabic nonce words (4 words (*mene, mani, nime* & *noma*) x 3 accent types) and a post-positional particle *mo* (also) which were produced by a female Tokyo Japanese speaker, as described in the ABX task with canonical tokens in 6.1.2. There were 36 stimuli (12 nonce words x 3 repetitions). Each audio stimulus was embedded within a carrier sentence, i.e., *Sumimasen*___*X*___*mo kudasai* (Excuse me but please give me___X___). There were also visual materials for this task. The triplet of each test word followed by the postpositional particle *mo* (also) was written in Japanese orthography, *Hiragana*, with one of three non-conventional diacritics on top of the word. Examples of the visual stimuli with the diacritics can be seen in Diagram 6.2, using *menemo*. A square bracket (┐) is placed on top of the pitch accent bearing unit: the first syllable of the target word in the case of the accent on the first syllable (A1), and the second syllable in the case of the accent on the second syllable (A2). For an unaccented word (A0), a short straight line (‾) was placed on top of the second syllable. Although the L2 learners were supposed to be capable of reading and writing in Japanese orthography as part of their course requirements for the first year, the Romanized equivalent of each word was also provided for the benefit of the learners.

**Diagram 6.2. Visual stimuli (*menemo*) showing each diacritic on top of the word.**

*menemo* (A1) *menemo* (A2) *menemo* (A0)




This set of diacritics was selected for three reasons. First of all, this diacritic system is used in a pronunciation dictionary of Tokyo (standard) Japanese published by NHK, the Japanese national broadcasting organization (1998). Secondly, some Japanese textbooks also use this diacritic system for teaching learners how to pronounce Japanese words. The textbooks which are used for the Japanese courses at the University of Edinburgh (ICU, 1996) also use this diacritic system and therefore the learners were likely to be familiar with it. Lastly, the result of a pilot study comparing three possible diacritic systems indicated this line/bracket method to be the easiest to learn (see Appendix C)[3].

### 6.3.3 Procedure

After the two ABX tasks, participants took the label assigning task individually in a sound-treated studio. Their task was to select one of the three visual stimuli with its diacritic which matched what they heard as the audio stimulus in the middle of the carrier sentence. At the same time, they also had to learn the novel diacritic system for Japanese pitch accent. In order to learn the system, they heard one canonical token of the audio stimulus embedded in the carrier sentence through a pair of headphones and also saw the set of three visual stimuli

[3] The original purpose of this pilot study was to test which of the three diacritic systems would be the most effective for eliciting learners' productions. Thus, the performance that I describe here was specific to the participants' production. Although the learners performed similarly on the line/bracket system and the point system, to avoid confusion with the stress marker, I chose to use the line/bracket system.

on the computer monitor. They were asked to indicate their choice by pressing one of three keys on the keyboard (clearly indicated with stickers). They were asked to guess if uncertain. Feedback was given after each response to aid learning the target diacritic system. Thus they were taught to learn the diacritics that match the three Japanese pitch accent types through feedback.[4] If their response was correct, then positive feedback (a message saying 'correct!') was presented. If the response was incorrect, the correct visual stimulus was provided as the correct answer.

The three Japanese pitch accent categories were introduced using three basic Japanese words representing target pitch accent types: *ocha* (tea) as an example for an unaccented word (A0), *asa* (morning) as for a word with the accent on the first syllable (A1), and *hiru* (noon) as for a word the accent on the second syllable (A2). A post-experiment interview confirmed that all the learners were familiar with these words. Each accent type was introduced one by one within a sentence which the learners could also understand;

---

[4] There were originally two between-group conditions in terms of how Japanese pitch accent was introduced in the instruction. As briefly described as a further issue in Chapter 8, the purpose of having two conditions was to test whether or not the learners performed the task better when it was introduced as relating to intonation patterns than as a matter of lexical distinctions. Under one group condition, the diacritic system was introduced as illustrating the three pitch accent types which distinguish word meaning in Japanese. Under the other group condition, it was introduced as illustrating three different intonation patterns. Since the results obtained from the experiment were difficult to interpret, the details are not reported in this thesis.

'***Ocha*** *o nomimasu* (I drink tea.).', '***Asa*** *mo nomimasu* (I drink tea in the morning, too.).', and '***Hiru*** *mo nomimasu* (I drink tea at noon, too.).' They were advised to recall these three words as a resort if they become unsure during the task. During the instructions, three audio examples for each accent type were also presented using nonce words with the carrier phrase as examples of what they should expect to hear in the main session.

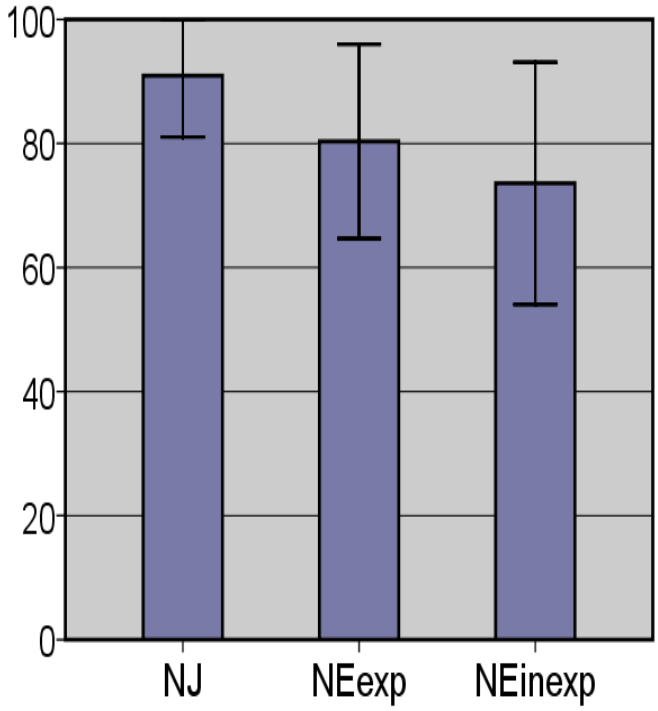
Prior to the main session, they had a practice session which was designed to be longer than the practice in the other tasks to allow them to familiarize themselves with the task. There were 9 items to practice, which made this practice session approximately twice as long as the practices in other tasks. In addition, there were 5 more extra items to practice at the beginning of the main session to allow them to practice again after they had asked any questions they had.

### 6.3.4 Results

The mean percentage of correct responses and the standard deviations for the three groups are presented in Figure 6.4. As we can see, both NEexp and NEinexp achieved better than chance level performance. However, the performance of NEexp seems to be slightly lower than that of NJ, and NEinexp scored lower than NEexp.

**Figure 6.4. Mean labeling accuracy (%) and the standard deviation of the three groups.**

Each bar represents each of the three groups. The error bars indicate +/- 1 SD.

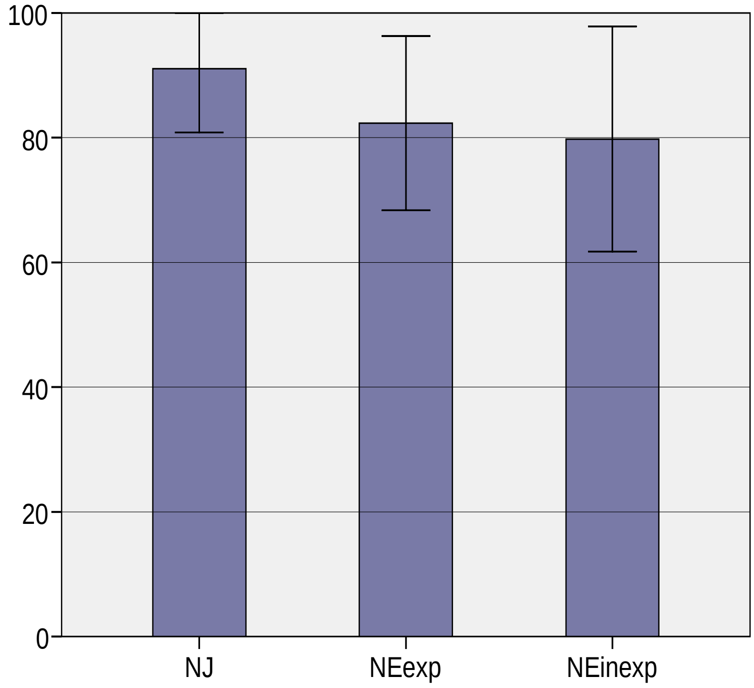



To compare the three groups' means statistically, a one-way ANOVA was carried out in which Group (NJ, NEexp and NEinexp) was the between-subject factor. The test showed that there was a significant difference among the three groups [$F$ (2, 40) = 4.23, $p < 0.05$]. Post-hoc comparisons using the Tukey test indicated that the mean score of the NEinexp group ($M$ = 73.6%, $SD$ = 20) was significantly lower than that of the NJ group ($M$ = 90.9%, $SD$ = 10) but the mean score of the NEexp group ($M$ = 79.6%, $SD$ = 17) was not statistically lower than that of the NJ group. There was no significant difference between the two learner groups. In addition, the NJ group showed a tendency towards a ceiling effect; the skewness of their distribution was -0.75 and the kurtosis was -0.91. This suggests that the label assigning task might have been too easy for the NJ group. This might have led to a smaller difference between the native speaker group and the learner groups than might otherwise be expected.

To examine further, the data was reanalyzed after excluding the eight participants who did not show the categorical function in the 7-step ABX task. As the label assigning task also requires the ability to hear the difference between the target canonical tokens and to match a token with the target category, the result of the labeling task might have also been influenced by these participants. Figure 6.5 shows the mean labeling accuracy (%) and the standard deviation of the three groups after excluding the eight participants. A one-way ANOVA test (3 levels) was conducted. What was found was that once these eight people were removed, the mean score of the NEinexp group ($M$ = 79.8%, $SD$ = 18.1) was no longer statistically lower than that of the NJ group ($M$ = 91.1%, $SD$ = 10.2). In addition, as in the initial analysis, the mean score of NEexp group ($M$ = 82.3%, $SD$ = 14.0) was not statistically lower than that of the NJ group. There was also no statistical difference between the means of the two learner groups. Meanwhile, when we compared the mean performance of the learners who did not show the categorical function in the 7-step ABX task with that of the native speakers, it showed that these learners performed significantly more poorly (*M* = 49.9%, *SD* =12.6) than the native speakers ($t$ (18) = 7.06, $p$ < 0.001). This interestingly suggests that even inexperienced learners, once they become capable of categorizing fine-grained tokens, are able to assign labels to each token. We can also consider NJ3, whose performance on the 7-step ABX task seemed to make him an outlier of the NJ group. His performance on the label assigning task was relatively good, with accuracy at 88.9%. A one-way ANOVA test (3 levels) was carried out without NJ3 to confirm whether or not his result had a significant impact on the group difference. The result of this test did not change from the result of the initial analysis: again, there was a significant difference between the NJ and the NEinexp but not between the NJ and the NEexp.

**Figure 6.5. Mean labeling accuracy (%) and standard deviation for the three groups without the eight non-categorical participants.**

Each bar represents one of the three groups. The error bars indicate +/- 1 SD.




### 6.3.5 Discussion

The results of the labeling task showed that the experienced learners were capable of assigning a novel label to Japanese pitch accent contrasts just as well as the native speakers. On the other hand, the results also showed that the inexperienced learners were less accurate than the native speakers in giving a label to the target lexical pitch accent contrasts. This indicates that the inexperienced learners differ from experienced learners in the extent to which they can assign the categories to lexical items, and the extent to which they can learn the labels. This may be because the learners have not fully learned to lexically associate an item with the Japanese pitch accent categories.

It is interesting that we observed this group difference in the labeling task since the native speakers also had to learn a novel diacritic system for this task just like the learners. In fact, some of the native speakers commented after the experiment that they had never previously encountered the labeling system used in the task. We should also note that the group difference seems to diminish as L2 experience increases. In addition, a significant difference between the three groups also disappeared once the non-categorical participants were excluded. The majority of these participants were inexperienced learners. These results indicated that English-speaking learners become capable of learning a novel diacritic system as a function of L2 experience by developing their lexical-phonological representation of Japanese lexical pitch accent.

Taking the results of both the ABX tasks into account, we see another interesting point. The inexperienced learners were also able to hear the difference between the canonical tokens of Japanese pitch accent contrasts as well as the native speakers of Japanese and also the experienced English-speaking learners. Nonetheless, the same inexperienced learners were not able to assign labels to these tokens as well as native speakers. As the

ABX task showed that the inexperienced learners could hear the differences between the canonical tokens, their problem in the labeling task cannot be attributed to the ability to hear the differences between the target canonical tokens. Based on the result of the step-wise ABX task, it appears that the problem for the inexperienced learners was that they have not fully formed the categories for Japanese pitch accent by understanding the relation between the target categories and the phonetic information. This implies that some inexperienced learners cannot identify the member of the target category like native speakers as they are still learning to form the categories of Japanese pitch accent. Thus, this might account for why it was more difficult for the inexperienced learners to assign a correct label to each token compared to the experienced learners who may have more developed target categories through their greater L2 experience.

Furthermore, the further analysis also showed that the label assigning performance of some of the inexperienced learners is equivalent to that of the native speakers of Japanese. Interestingly, these learners happen to be the same ones who showed the categorical function in the 7-step ABX task. In other words, those learners who showed good categorical function were also able to assign a label to each item from the target three categories. By the same token, these results also indicated that the inexperienced learners who did not show the categorical function in the step-wise ABX task also showed themselves unable to do the label assigning task like the native speakers of Japanese. Nevertheless, from the ABX task with canonical tokens, we know that all but two of these learners were able to hear the differences between the canonical tokens of Japanese pitch accent contrasts. These findings lend more support to the possibility that they may be having a problem in the phonetic representations, and lexical-phonological representations of Japanese lexical pitch accent categories, i.e., forming the target L2 categories by being able to associate the phonetic information and the L2 categories, and by being able to lexically associate a target token with the corresponding category.

If we adopt the SLM proposed by Flege (1995), we can speculate that the reason behind the inexperienced learners' problem could have been that they were assimilating L2 categories into the closest L1 category. For example, the English speakers could have perceived Japanese lexical pitch accent contrasts as being similar to English lexical stress contrasts such as English fore-stressed 'permit' (noun) and end-stressed 'permit' (verb). F0 is the key acoustic correlate for Japanese lexical pitch accent, whereas for English it is only one of several relevant acoustic cues for English lexical stress (as along with duration, intensity or vowel quality (Beckman, 1986)). Given the shared acoustic correlate, i.e., F0, the English speakers might have assumed that the L2 category of Japanese A1 pitch accent (pitch accent on the first syllable) was like the fore-stressed 'permit' (noun) category (stress on the first syllable), and equally they could have assumed that the category of Japanese A2 pitch accent (pitch accent on the second syllable) was like the end-stressed ‘permit’ (verb) category (which has the stress on the second syllable). Since English does not have unstressed words, the learners might have assimilated unaccented Japanese words to either of the English categories depending on which category they find more similar. It is also possible that the English speakers might have paid attention to the other acoustic cues which are used in English stress, such as duration or intensity. Thus, they might have perceived the two Japanese pitch accented categories to be too similar to each other, and assimilated them into just one of the English stress categories. Thus, they might have been unable to categorize pitch accented words in Japanese. However, these suggestions remain tentative as we do not know whether or not English speakers do perceive Japanese lexical pitch accent as English lexical stress among all their L1 categories. This therefore highlights the empirical challenge of how to reliably test how native speakers of English perceive the relation between Japanese pitch accent categories and English stress categories, as an issue which remains open as a topic in need of future investigation.

Taking into consideration the cross-linguistic difference between Japanese and English regarding the function of F0, there is another possible explanation for the English-speaking learners’ non-target-like F0 patterns. Although F0 is not used in English for making lexical distinctions, F0 has a significant role in distinguishing phrasal intonation patterns in both Japanese and English. For this reason, it is more likely that while the function of F0 will be *linguistically* registered for native speakers of English as well as native speakers of Japanese, it may not be *lexically* registered in case of English speakers. Thus, the native speakers of English might have been unable to associate the phonetic information and Japanese pitch accent categories, nor to associate the target categories with lexical function. We can speculate that this may be why the learners failed to form the categories of the L2 target contrasts. In the initial stage of learning, learners tend to rely on their L1 system, and thus the tendency may also be stronger in inexperienced learners than in experienced learners. But once the learners learn to understand these associations, then they may be able to form the L2 categories of Japanese pitch accent. This could also explain the difference between the experienced learners and the inexperienced learners observed in our data.

To summarize, the label assigning task showed that the experienced learners are able to learn a novel labeling system for Japanese lexical pitch accent and assign categories to lexical items like native speakers of Japanese. However, some of the inexperienced learners were having a problem in assigning a label to the categories of Japanese pitch accent, although others were not doing significantly poorly compared to the native speakers.

Interestingly enough, these learners who had difficulty with label assignment were also the learners who were unable to show the categorical function in categorization in the step-wise ABX task. This may therefore provide evidence that the inexperienced learners are having a problem in matching a target token with the L2 category in their lexical-phonological representation of Japanese lexical pitch accent, since we assume that due to lack of L2 experience their L2 categories are not fully developed by understanding the connection between the phonetic information and between the L2 categories and lexical items. Hence, it seems possible that this difficulty might have contributed to their non-target-like F0 profile in their productions of Japanese pitch accent, which may be perceived as foreign accent by native speakers. Having seen these results, we will now turn in the next subsection to investigate the relationship between categorization ability and intelligibility scores.

## 6.4 Relationship between intelligibility scores and categorization ability

Now let us consider the relation between categorization ability and the learners' non-target-like F0 patterns in their productions of Japanese pitch accent. In Chapter 3.2, we saw that the English-speaking learners produce non-target like F0 profiles which differed from the native speakers of Japanese in terms of the auditory measurements and the overall F0 patterns. The results indicated that the native speakers of Japanese responded systematically to the acoustic patterns produced by the participants and the non-native like F0 profile was perceived as foreign accent. Hence, the relationship between categorization ability and foreign accent as measured by intelligibility scores merits closer inspection.

To investigate the relationship between the intelligibility scores and categorization ability, one stepwise multiple regression analysis was carried out between the intelligibility score[5] and the three sets of categorization results (i.e., the ABX task with canonical tokens, the ABX task with 7-step stimuli and the label assigning task). The dependent variable was the average intelligibility score on each speech production for each individual speaker. There were four independent variables, drawn from the results obtained from the categorization tasks. Two of the four independent variables were the average percentage of correct responses for each participant for the ABX task with canonical tokens, and the average percentage of correct responses for each participant for the label assigning task. For the two remaining independent variables, the slope value and the category boundary value of each participant obtained from the ABX task with 7-step stimuli were used.

---

[5] As in the production analyses, the intelligibility scores of eight participants were used as the representatives of each group (24 in total). Four out of the eight participants who did not show the categorical function in the step-wise ABX task (NEexp 10, NEinexp 2, 10 and 13) happened to be also included.

The regression analysis is presented in Table 6.5. We are particularly interested in $R^2$ and $\Delta R^2$ values. The value of $R^2$ indicates how much variance in the intelligibility score is explained by the variance of the significant independent variable. The value of $\Delta R^2$ indicates the additional amount of variance of the intelligibility score explained by the variance of another significant independent variable. Interestingly, the result showed that the average intelligibility score of each participant can be mostly explained by the average percentage of correct responses of the label assigning task and the canonical ABX task. The percentage of correct responses in the label assigning task accounted for 38.6% of variance in the intelligibility score ($p$ <0.01). In addition to the variance of the label assigning task, the variance of the canonical ABX task accounted for 14.9% of the variance in the intelligibility score ($p$ <0.001). In contrast, the results also showed that neither the variance of the slope value or the category boundary value of the step-wise ABX task can meaningfully account for the variance of the intelligibility score.

**Table 6.5 Stepwise multiple regression analysis examining the relation between intelligibility and the four categorization abilities.**

The B-values indicate unstandardized coefficients. The SE B-values are standard errors of the B-values. The Beta-values indicate standardized coefficients.

| | *B* | *Std.Error B* | *Beta* |
|---|---|---|---|
| Step 1 | | | |
| Constant | 0.17 | 0.19 | |
| Label assigning | 0.78 | 0.23 | .62** |
| Step 2 | | | |
| Constant | -0.51 | 0.34 | |
| Label assigning | 0.59 | 0.23 | .47* |
| ABX canonical | 0.95 | 0.41 | .42* |

Note $R^2$ = .39 for Step 1 ($p$ <.01); $\Delta R^2$ = .15 for Step 2 ($p$ <.001). * $p$ <.05, ** $p$ <.01.

Some different patterns emerged when the eight participants who did not show categorical function in the step-wise ABX task were excluded from these analyses. In particular, as we saw above, the results of the label assigning task changed significantly when these eight participants were excluded previously. As the variance in the label assigning task was found to significantly explain the variance in the intelligibility score in the present regression analysis, it is of interest here whether or not excluding these participants also changes the result of the regression analysis. To test this, another stepwise multiple regression analysis was conducted without these eight participants. However, the result turned out to be very similar to the initial analysis and did not yield significant differences (see Table 6.6). The variance of the intelligibility score can largely be explained by the percentage of correct responses of the label assigning task, which accounted for 38.6% of variance in the intelligibility score (*p* <0.01). In addition to the variance explained by the label assigning task, the variance of the canonical ABX task accounted for 12.8% of the variance in the intelligibility score (*p* <0.01).

**Table 6.6 Stepwise multiple regression analysis without the eight non-categorical participants examining the relation between intelligibility and the four categorization abilities.**

The B-values indicate unstandardized coefficients. The SE B-values are standard errors of the B-values. The Beta-values indicate standardized coefficients.

| | *B* | *Std.Error B* | *Beta* |
|---|---|---|---|
| Step 1 | | | |
| Constant | 0.25 | 0.17 | |
| Label assigning | 0.69 | 0.20 | .62** |
| Step 2 | | | |
| Constant | -0.44 | 0.36 | |
| Label assigning | 0.50 | 0.21 | .45* |
| ABX canonical | 0.96 | 0.45 | .40* |

Note $R^2$ = .39 for Step 1 (*p* <.01); $\Delta R^2$ = .13 for Step 2 (*p* <.01). * *p* <.05, ** *p* <.01.

In addition, when only these eight participants were examined, none of the independent variables could statistically explain the variance of the intelligibility score. This result may not be surprising considering that the seven learners among those participants performed poorly on the label assigning task compared to native speaker whereas the one native speaker showed relatively good performance on the task. Moreover, when the one native speaker was excluded from the analysis, the significance of the canonical ABX task variance disappeared. This also may not be also surprising as he did perform very poorly on this task and this might have influenced the result.

It is striking that the foreign accentedness of English-speaking learners regarding Japanese pitch accent as perceived by native speakers of Japanese can be accounted for by these two types of categorization ability in the perception tasks. The significant result of the variance of the ABX task with canonical tokens shows that the learners' ability to hear the difference between the target canonical tokens and their ability to match each token to the corresponding category helps to predict how non-native like their productions sound to native speakers of that language. What is even more interesting is that the significant result of the variance of the label assigning task indicates that the learners' ability to match each target token to the category in their mental representation and their ability to learn the labels are able to a great extent to account for the foreign accentedness perceived by native speakers. Furthermore, it is very noteworthy that more than 50% of the variance in the intelligibility score can be explained by the combination of the variance of the canonical

ABX task and the label assigning task. Hence, it is possible to interpret this result as showing that it is at least potentially reasonable that the mental representation of Japanese pitch accent categories is a factor which substantially helps to account for the L2 learners' non-target-like production patterns as detected by native speakers of the target language.

On the other hand, the results also showed that the variance of the slope and of the category boundary of the step-wise ABX task were not significantly able to explain the extent to which foreign accentedness is detected by native speakers. However, the step-wise ABX task can be considered to test an ability which is intermediate between what is tested by the ABX task with canonical tokens and the labeling task. The step-wise ABX task tested categorization ability requires linguistic perception ability tested by the ABX task with canonical tokens, while the labeling task requires linguistic perception ability and categorization ability along with lexical assignment ability. For this reason, the overall results can be interpreted as showing that a range of L2 learners' categorization ability makes an important contribution to perceived foreign accent. This is noteworthy in that it is not only the phonetic aspects of the L2 learners' productions which native speakers physically hear but also their phonological aspects which native speakers apparently cannot directly see, which seems to influence how the native speakers of the target language assess the foreign accentedness of L2 learners.

## 6.5 Discussion

To investigate the factors behind foreign accent, three perception tasks were conducted. Three specific questions were addressed: 1) whether or not native speakers of English are able to hear the difference between canonical speech tokens of Japanese lexical pitch accent categories and to classify the target item into the correct L2 category; 2) whether or not native speakers of English are able to hear the difference between the stimuli on a F0 peak alignment continuum of Japanese lexical pitch accent contrasts and categorize each stimulus into the target L2 category; 3) whether or not native speakers of English are able to learn a novel labeling system for Japanese lexical pitch accent like native speakers of Japanese and to assign a label to canonical tokens of Japanese lexical pitch accent contrasts. By answering these questions, we aimed to address whether or not a problem in the interaction between learners' speech perception, phonetic representation and lexical-phonological representation of Japanese lexical pitch accent is involved in foreign accent. More specifically, we were interested in whether or not the problem lies 1) in hearing the phonetic difference between the members of the target L2 categories; 2) in categorizing the target L2 categories by associating the phonetic information with the target L2 categories; or 3) in assigning L2 categories to lexical items by learning the connection between the category and the item.

The ABX task using the canonical tokens of Japanese pitch accent revealed that both the experienced and inexperienced learners were able to hear the differences between the tokens just as well as the native speakers of Japanese. Thus, we were able to answer the first question: native speakers of English were able to hear the difference between canonical speech tokens of Japanese pitch accent contrasts and to classify the target item into the correct L2 category. This suggests therefore that English learners of Japanese do not seem to have problems in hearing the phonetic difference between the representative members of the target L2 categories.

Regarding the second question, the results of the other ABX task using the 7-step F0 peak alignment continuum showed that the learners were having more difficulty with the intermediate stimuli on the continuum, i.e. the aspect of categorization which is related to the ability to associate the phonetic information with the L2 categories. In addition, the inexperienced learners showed by their less categorical function that they had more problems than the experienced learners in categorizing these stimuli. However, after removing the participants who did not show a categorical function (8 in total, who mainly consisted of the inexperienced learners), the performance of the remaining group of inexperienced learners shows much more similarity to the native speakers' performance. This indicates that once learners become capable of associating the phonetic information with Japanese pitch accent categories and formed the target L2 categories, their performance can also become closer to the native speakers' performance. Overall, the results indicated that some native speakers of English were not able to categorize stimuli on a F0 peak alignment continuum of Japanese pitch accent contrasts like native listeners in that they seem to have a problem in categorization. However, the results also suggest that the learners become capable of making this categorization as their L2 experience increases, possibly by relating the phonetic information and the L2 categories. Hence, native speakers of English seem to have problems in associating the phonetic information with the target L2 categories until they gain enough L2 experience. This suggests that their problem may lie in their phonetic representation of the target L2 categories.

Meanwhile, as for the third question, the label assigning task showed that the experienced English-speaking learners were able to learn a novel labeling system for Japanese pitch accent like native speakers of Japanese and associate a lexical item with the L2 category, whereas some inexperienced learners were not. However, when the participants who did not show categorical function in the step-wise ABX task were excluded from the analysis, the difference between the native speakers and the inexperienced learners also disappeared. This suggested that the learners who were able to associate the key phonetic information with the L2 categories were also able to assign categories to lexical items.

Otherwise, some native speakers of English seem to have problems in lexical assignment. Hence, this suggests that the learners seem to have lexical-phonological representation of Japanese lexical pitch accent categories.

Considering these results together, they seem to suggest that some English learners of Japanese tend to have a problem in their categorization ability and lexical assignment ability, especially in the initial stage of learning to form L2 categories. The results seem to suggest that their problems are related to phonetic representation and lexical-phonological representations of Japanese lexical pitch accent categories. If the learners do not have the target L2 categories and/or are unable to lexically identify the items, they may not be able to retrieve the necessary information to implement the output. Therefore, this could result in foreign accent. In Chapter 5, where the learners imitated the non-speech stimuli, we observed that they also have a problem in articulating the key acoustic correlate of Japanese lexical pitch accent as part of the output process requirements for achieving target-like productions. Hence, in addition to phonetically realization problem in the speech, both an articulation problem and representation problems seem to be contributing to the learners' foreign accent. However, at this stage, we do not know whether the learners have an articulation problem and a phonetic realization problem as part of the output process but not a representation problem, or whether an articulation problem, a phonetic realization problem and a representation problem are all related regarding foreign accent. Since the results of the three perception tasks in this chapter indicated that phonetic representation and lexical-phonological problems for the experienced learners seems to be less problematic than for the inexperienced learners, the nature of the problem may also change as the learners gain L2 experience. Hence, we will investigate this issue further in the next chapter.

The view that the source of the foreign accent of English learners of Japanese seems to be a mental representation problem was further supported by the results obtained from the regression analyses. When we investigated the relationship between categorization ability and foreign accent as revealed in the learners' productions, it was found that L2 learners' categorization ability, i.e., linguistic perception, categorization, and lexical assignment, all contribute to predict foreign accent as detected by native speakers. Another important finding was that it was mainly a combination of the learners' ability to assign the L2 category to a lexical item and their ability to hear the difference between the tokens, that can account for foreign accentedness (approximately half of the variation in the perceived foreign accentedness score). Therefore, this indicates that the interaction between speech perception and lexical-phonological representation problems of the L2 categories seem to play a very important role in explaining learners' foreign accentedness.

## 6.6 Summary

In this chapter, we investigated whether or not the non-target-like F0 profile of the native speakers of English can be attributed to their ability to categorize Japanese pitch accent contrasts. To test the learners' ability for linguistic perception, categorization, and lexical assignment, three perception tasks were carried out: an ABX task using canonical tokens, an ABX task using a 7-step F0 peak continuum and a label assigning task.

The overall results suggest that some L2 learners have problems in categorization ability and lexical assignment ability which can be seen as a function of their L2 experience. The results also provided support for this interpretation in that a combination of the learners' categorization abilities is found to be able to explain foreign accentedness (i.e., lexical assignment and linguistic perception).

In the next chapter (Chapter 7), we further pursue the question of potential interactions, by re-analyzing the speech and non-speech productions from the point of view of mental representation problems raised in this chapter.

# 7 Production results revisited

The main aim of this thesis is to investigate possible factors behind the foreign accent observed in the productions of the L2 learners. In this chapter we further explore potential interactions between the various factors which are hypothesized to be involved in foreign accent, i.e., the nature of learners' problems. In Chapter 6 we re-analyzed earlier data from the perspective of phonetic representation problems and lexical-phonological representation problems of the L2 categories (specifically, we re-analyzed the speech productions discussed in Chapter 3, which revealed foreign accent and also indicated a phonetic realization problem, and the non-speech productions discussed in Chapter 5, which showed that an articulation problem was a possible factor)[1]. In this chapter, we aim to investigate whether the learners have an articulation problem and/or a phonetic realization problem (as part of the output process) which is independent of representation problems (part of the internal mechanism between the input and the output processes), or whether learners have a combination of these problems.

We achieve this aim by re-analyzing the data, excluding the learners who indicated

[1] Since our data showed that a problem of lack of perceptual sensitivity does not seem to be related to foreign accent (discussed in Chapter 4), this factor was excluded as a possible factor behind foreign accent, and therefore the data is not re-analyzed here.

problems with the representation of Japanese lexical pitch accent categories. We consider that participants have a phonetic representation problem if they showed a non-categorical function in the step-wise ABX task, and we consider that participants have a lexical-phonological representation problem, or a problem in representing L2 categories, if they showed difficulty in the label assigning task. We can infer from the step-wise ABX that the learners who had a non-categorical function have problems in making the association between the phonetic patterns of Japanese lexical pitch accent and the target categories. We can further infer from the label assigning task that the learners have problems in making the association between Japanese lexical pitch accent categories and Japanese words. Considering these results regarding their categorization ability, we diagnosed that the learners who showed a non-categorical function have mental representation problems.

The results of the previous chapters (Chapter 3 and 5) indicated that Japanese lexical pitch accent, as produced by inexperienced learners, was more foreign accented than that produced by the experienced learners. In addition, we showed that the inexperienced learners seem to have more articulation problems and phonetic realization problems than the experienced learners. We also provided evidence that the foreign accent of the learners is not predicted by any lack of perceptual sensitivity (discussed in Chapter 4). These results suggest that even though the learners are not completely lacking in perceptual sensitivity towards the key acoustic correlate of Japanese lexical pitch accent (i.e., F0), they still have problems in manipulating F0 to achieve target-like productions.

For the learners to achieve productions without foreign accent, we assume that there is another important mechanism: something which connects the input and the output processes. Thus, in order to understand the internal mechanism of the relation between linguistic perception, categorization and lexical assignment, we also investigated the learners' categorization ability with Japanese lexical pitch accent contrasts. This was discussed in Chapter 6, where we found that the English learners of Japanese do not seem to have a problem in hearing the differences between the target tokens of Japanese pitch accent and classifying them into the target L2 category. However, we also saw in Chapter 6 that once the learners need to predict membership of category boundary areas, or lexically assign the L2 categories to an item, they seem to have a problem. This suggests that the learners have problems in the ability to categorize Japanese lexical pitch accent contrasts, and to assign the categories to Japanese words. This crucially implies that some native speakers of English have a problem in phonetic representation and lexical-phonological representations of Japanese lexical pitch accent categories. Since the learners who did not show a categorical function were mostly the inexperienced learners, this result also indicated that the inexperienced learners have more of a representation problem than the experienced learners.

We reasoned that because of the cross-linguistic difference in F0 functions, English learners of Japanese might have problems both in associating the phonetic patterns that characterizes Japanese lexical pitch accent categories and the target categories, and in learning the link between Japanese lexical pitch accent categories and lexical items in Japanese. This makes it difficult for the learners to form the appropriate categories for Japanese lexical pitch accent. As the learners gain L2 experience, however, we predict that it is possible for them to be able to form L2 categories which will be closer to the native categories, even if they are not completely native-like. Thus, we do not intend to suggest that the learners who show a categorical function are 'problem-free' in terms of their mental representations, but only that relative to the non-categorical learners, they must be assumed to have much less of a problem in representing the target categories.

There are four possible scenarios for the nature of the learners' problems behind foreign accent that we can draw our conclusion from. First, we examine whether non-target-like patterns are observed in the speech productions of the learners who showed a categorical function. If the speech productions of these learners indicate native-like patterns, then we can conclude that the non-target-like productions that we discussed in Chapter 3 were due to the learners with representation problems, and therefore that representation problems could be a key factor for foreign accent; more specifically, phonetic representation and lexical-phonological representations of the L2 categories. Alternatively, if the speech productions of these learners still indicate non-target-like patterns, then we can conclude that they have a phonetic realization problem (and an articulation problem if they also show non-target-like non-speech productions), and therefore that these could be factors behind foreign accent independently of representation problems. A third possibility is that the learners who indicated representation problems also have a phonetic realization problem (and an articulation problem if they also show non-target-like non-speech productions); in this case we would need to examine whether representation problems are the main problem for them, or a combination of representation problems, a phonetic realization problem and an articulation problem. Lastly, in the case that the learners who indicated representation problems happen to show that they do not have problems in phonetic realization and/or articulation, this could provide further evidence that a phonetic realization and/or articulation problem could be factors behind foreign accent independently of representation problems.

It is also important to keep in mind that in the previous chapters we found that the learners who seem to have representation problems were mostly inexperienced learners, while the experienced learners seem to have relatively less of representation problems. This implies that the nature of the problem behind foreign accent can change during L2 speech acquisition.

### 7.1 Production results revisited

The results in Chapter 6 indicated that the overall patterns of the learner groups were influenced by the patterns of the individual learners who showed a non-categorical function in the step-wise ABX task. Since these non-categorical learners were mostly from the group of inexperienced learners, the tendency to show a non-categorical function seems to be particularly evident in the case of the less experienced learners. We saw the difference not only in the step-wise ABX task but also in the label assigning task. As we have been arguing, considering the nature of these tasks, this difference seems to reflect phonetic representation and lexical-phonological representation problems of Japanese lexical pitch accent categories.

In addition, the regression analysis in Chapter 6 suggested that the variance in the intelligibility scores (as a criterion for foreign accentedness) was significantly explained by the variance in the categorization ability. Interestingly, the preliminary analyses reported by Flege et al. (1997) also indicated that most learners who did not show a categorical function in their step-wise perception experiment did not produce a significant amount of difference of acoustic correlate between the target. In other words, their results indicate that learners' ability to show a categorical function seems to be related to the foreign accentedness of their productions. It is possible that the foreign accentedness in the learners' speech productions (discussed in Chapter 3) was a characteristic of those learners who showed a non-categorical function.

To find out whether or not the English-speaking learners who indicated a categorical function show target-like productions of Japanese pitch accent, we first reanalyze the learners' productions of speech imitations. Taking their categorical function as an indicator of relatively sound phonetic representations and lexical-phonological representations (even if not entirely native-like), we will be able to diagnose whether these learners do or do not have a phonetic realization problem independently of any problems with the phonetic and lexical-phonological representation of the L2 categories.

If it is the case that the learners do not have an independent phonetic realization problem, this would suggest that non-target-like productions are due to the patterns produced by the learners who did not show a categorical function, and that learners have representation problems as a factor behind foreign accent. If it is the case that the learners do seem to have an independent phonetic realization problem, then we will need to investigate further whether or not the learners who showed a non-categorical function have a phonetic problem in addition to representation problems. If the non-categorical learners also have a phonetic realization problem, this could suggest that the factors behind foreign accent are of a dynamic nature. Since these learners were mainly the inexperienced learners, we could interpret such a finding as showing that in the early stage of learning, a combination of representation problems and a phonetic realization problem is what gives rise to their foreign accentedness. However, in the later stage of learning, the learners may have a phonetic realization problem independently of representation problems. Alternatively, if the learners who showed a non-categorical function do not have a phonetic problem, this could provide evidence in support of the view that representation problems and a phonetic realization problem can be independent of each other as sources of foreign accent. This would also underscore the dynamic nature of the factors behind foreign accent, as it would suggest representation problems for the early stage of learning and a phonetic realization problem for the later stage.

Among the participants who showed a non-categorical function in the step-wise ABX task, four learners were included in the initial production analysis; one NEexp (NEexp 10) and three NEinexp learners (NEinexp 2, 10, 13). We can consider the production data of these four learners to be representative of all the learners who showed a non-categorical function in the ABX task, in order to see whether any difference emerges when we compare their results with the initial results. We therefore went back to the speech data and the non-speech data that were discussed in Chapter 3 and Chapter 5. In order to find whether the results yield any difference in the analyses with and without the participants who did not have categorical functions in the step-wise ABX task, we look at the intelligibility score and the overall F0 patterns. After excluding the non-categorical participants, we first compare each result of the three group comparison (NJ, NEexp and NEinexp) to that of the initial production results (i.e., with all participants included). Then, each result of the comparisons between only the non-categorical learners and the native speakers are compared with the initial production results.

### 7.1.1 Intelligibility scores revisited

First of all, we examined whether or not the learners who indicated a categorical function show foreign accentedness in terms of the intelligibility scores. When we reanalyze the intelligibility scores, the overall picture remains the same as was indicated in the initial analysis. The learners who showed a categorical function also indicated non-native-like patterns in their productions.

In the initial analysis of the intelligibility score, different patterns were found between the three groups across the three accent types (see 3.2.4.1). We found that the NJ performed nearly perfectly across the three accent types on average whereas the NEexp had lower scores than the NJ, and the scores of the NEinexp were even lower than those of the NEexp. In addition the learners were similar to the NJ on A0 but they performed more poorly on A1 and even more poorly on A2 compared to the NJ.

In the reanalysis, when the non-categorical participants are excluded, we find that these patterns are still observed. The mean and standard deviations of the intelligibility scores after excluding the non-categorical participants are presented in Table 7.1. According to a mixed design ANOVA (Accent (3 levels) x Group (3 levels)), there was a significant interaction ($p$ <0.001) as well as the main effects of Accent ($p$ <0.001) and Group ($p$ <0.001). The post-hoc comparisons showed that the mean intelligibility score of the NEexp was lower than that of the NJ, and that of the NEinexp was even lower than that of the NEexp. Regarding Accent, for A0, the mean intelligibility scores of the NEexp and the NEinexp were just as high as the NJ. However, on the accented types (A1 and A2) the learners had lower mean intelligibility scores than the NJ, particularly in the case of NEinexp (and lower in the case of A2). This indicates that it was more difficult for the learners to achieve target-like productions on the pitch-accented words compared to the unaccented words.

**Table 7.1 The mean percentage of times that Japanese pitch accent produced by the participants were identified as intended: intelligibility scores with non-categorical participants excluded.**

The left: provides the results obtained after excluding the non-categorical participants. The right: provides the results of the non-categorical participants.

| | | Group results after excluding the non-categorical participants | | | | Non-categorical participants | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Accent | N | Mean | SD | | Accent | N | Mean | SD |
| NJ | A0 | 96 | 96.35 | 18.13 | Non-cat | A0 | 48 | 88.54 | 31.35 |
| | A1 | 96 | 100.00 | 0.00 | | A1 | 48 | 62.50 | 45.55 |
| | A2 | 96 | 99.48 | 5.10 | | A2 | 48 | 26.04 | 42.51 |
| NEexp | A0 | 84 | 97.62 | 15.34 | | | | | |
| | A1 | 84 | 88.69 | 31.39 | | | | | |
| | A2 | 84 | 73.21 | 42.11 | | | | | |
| NEinexp | A0 | 60 | 97.50 | 10.99 | | | | | |
| | A1 | 60 | 53.33 | 49.46 | | | | | |
| | A2 | 60 | 40.00 | 47.66 | | | | | |

When we reanalyze the misidentification patterns, the overall picture also remains the same as in the initial analysis. According to the initial observation of misidentification patterns in 3.2.4.1, for both learner groups, the major misidentification type was the accented targets (i.e., A1 and A2) (70 out of 98 cases for the NEinexp and 40 out of 55 cases for the NEexp). The learners’ productions of both accented targets were typically identified as the unaccented targets (i.e., A0) by the native Japanese listeners. When the non-categorical participants are excluded, this same tendency was again observed. There were 57 cases of misidentification for the NEinexp and 26 cases for the NEexp.

In the initial data, the next most common misidentification type was that the learners' productions of A2 accent was misidentified as A1 (22 out of 98 cases for the NEinexp and 8 out of 55 cases for the NEexp). When the non-categorical participants are excluded, this remains the same: 21 cases of A2 were misidentified as A1 for the NEinexp and 6 cases for the NEexp. It is possible that the learners might have associated F0 patterns that have an earlier peak location than the native speakers' location with the A2 category of Japanese lexical pitch accent. Hence, it may have been particularly difficult for the learners to produce target-like F0 contours for A2.

Thirdly, in the initial data, misidentifications in the opposite direction (i.e., where A1 accent was misidentified as A2) occurred in only a few cases. This too remained the same when the non-categorical participants were excluded.

These patterns suggest that it was difficult for the native speakers of English to imitate F0 contours in their productions of the target Japanese pitch accent. In particular, where to aim to locate the F0 peak, and how fast to make a falling contour, seem to pose problems for the learners.

It should also be noted that this pattern of results was also the case when the non-categorical learners were compared as a group with the NJ group. The mean intelligibility score on their A2 productions was even lower than that of the NJ. The major misidentification type was also the accented targets (i.e., A1 and A2) (13 cases for the non-categorical NEinexp and 14 cases for the non-categorical NEexp). However, in the case of the non-categorical learners, there were only a few cases where A2 accent was misidentified as A1. These patterns further support the view that it was difficult for the native speakers of English to imitate F0 contour in their productions of the target Japanese pitch accent. The problems of where to locate the F0 peak and how fast to make a falling contour seem to be even more problematic for these non-categorical learners compared to the other learners.

Hence, the overall results suggest that regardless of whether or not the learners indicate categorical function in the categorization task, the degree of foreign accentedness perceived in their productions, and the misidentification patterns on how their productions were perceived, tend to remain very much the same. This means that the English-speaking learners who indicated a categorical function also show non-target-like productions of Japanese pitch accent in terms of the intelligibility scores. Therefore, we can interpret this result as suggesting that even though the learners seem to have little problem with the mental representations of L2 categories, they nevertheless seem to show foreign accentedness in their productions. This, in turn, suggests that the overall F0 patterns may also be non-target-like, not only in the productions of the learners with representation problems but also the productions of the learners with lesser representation problems. If they show non-target-like

F0 patterns, this suggests that the learners have a phonetic realization problem regardless of the learners' degree of representation problems of the L2 categories.

### 7.1.2 Overall F0 patterns revisited

Following the results of the intelligibility scores, we now examine whether the overall F0 patterns in the productions of the learners who showed a categorical function also show non-target like F0 patterns. The reanalysis described in this subsection shows that these learners did indeed show non-target-like F0 patterns in their speech productions of Japanese pitch accent.

The initial analysis of the overall F0 contours showed that three groups had different patterns across the three accent types (see 3.2.5.1). We focus on the overall patterns, such as the shape of the contour where the F0 peak is located, whether the F0 contour contains a sharp fall after the peak, or how wide the F0 contour is. Overall, for the accented targets, the learners' F0 patterns showed more variability in the location of the F0 peak, and how fast the F0 fall is; for the unaccented targets, their productions have wider F0 range compared to the native speakers' productions. These tendencies were stronger in the F0 patterns produced by the inexperienced learners.

To recap what was reported in 3.2.5.1 regarding the overall F0 patterns, in the NJ's productions, the F0 peak location for A1 occurred earlier than A2 (as expected). A similar pattern was observed in the NEexp's productions. Meanwhile, in the NEinexp's productions, the F0 peak location was not differentiated with large individual differences, particularly for A2. The F0 peak locations for the accented targets (i.e., A1 and A2) in the NEexp' productions appear slightly later than the peak locations in the NJ's productions, and the F0 peak locations for A1 in the NEinexp's productions seems later than those in the NJ's productions, with more variability than in the NEexp's data.

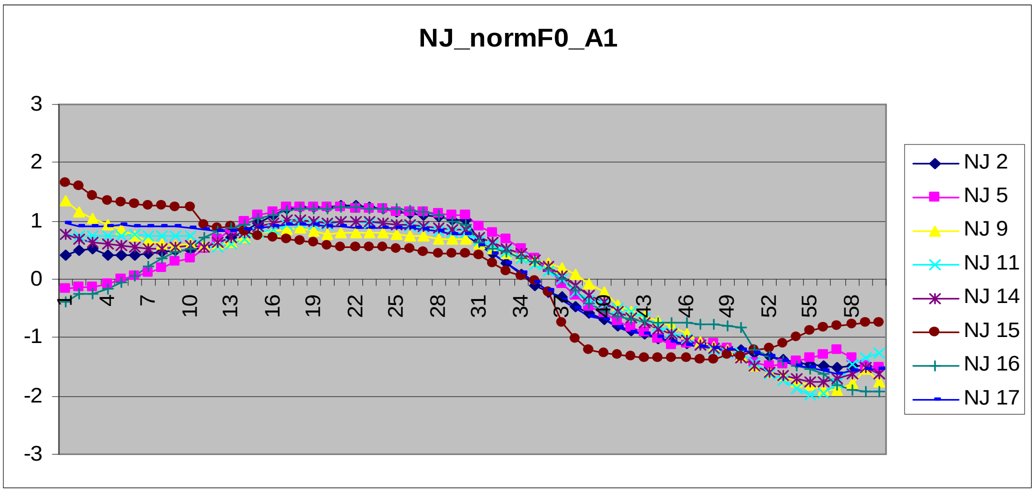
As for the degree of the F0 fall after the peak, in the NJ's productions, the degree of fall for the accented types (A1 and A2) seems to be recognizable and it was even larger for A2 than for A1. This tendency was similar in the NEexp's productions. However, in the NEexp's productions the degree of fall for A1 and A2 was slightly less steep than it was in the NJ's productions.

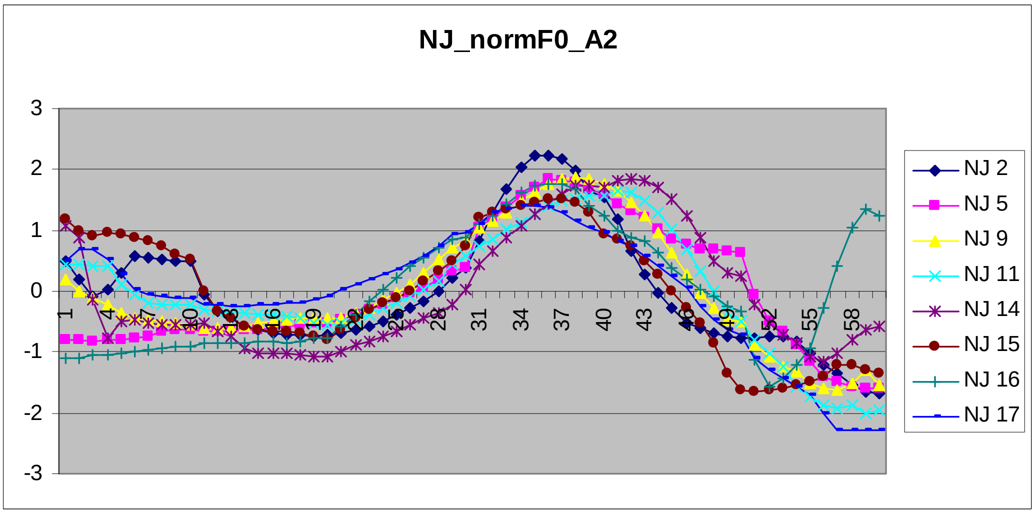
Lastly, in the NJ's productions, the ranges of the F0 contours for A2 and A0 were slightly wider than for A1; this trend was also observed in the productions of both the learner groups.

The results of the reanalysis are as follows. Figure 7.1 reproduces from Chapter 3 the F0 contours of the NJ, presented here again for comparison. The subsequent figures represent the F0 contours in the productions of the learner groups after the non-categorical participants were excluded (NEexp learners in Figure 7.2 and NEinexp learners in Figure 7.3). Since only one learner was excluded from the NEexp group for showing a non-categorical function, the overall F0 patterns of the NEexp are similar to what we saw in the initial data. Thus, here we focus on the NEinexp's data. In the case of the NEinexp, once the non-categorical participants are excluded, the individual differences seem to be reduced (particularly in the A2 productions) but the overall F0 contours still indicate more variance compared to the NJ group and also compared to the NEexp group. The F0 peak location for the accented targets (i.e., A1 and A2) still shows individual differences and the F0 peak seems to occur slightly later in the NEinexp's productions compared to the NJ's productions. The degree of F0 fall after the peak for A2 in the NEinexp's productions remains less steep than that in the NJ's productions. The steepness of the fall for A1 in a few of the NEinexp seems to be similar to what we see in the NJ's productions but overall, the NEinexp group shows wide variance among the learners. Meanwhile, the patterns of A0 show individual differences in the NEinexp's productions but the trend seems not too different from the NJ's productions. In addition, the patterns of the F0 range in the NEinexp's productions remains the same as in the initial analysis; the ranges of the F0 contours for A2 and A0 were slightly wider than for A1 in the NJ's productions. This suggests that even when extreme cases are excluded, the inexperienced learners seem to show divergent patterns in their productions regarding F0 peak location and the degree of the falling contour after the F0 peak. This indicates that regardless of whether or not the learners show a categorical function, they seem to show non-target like F0 patterns. Hence, this can be interpreted as implying that the learners seem to have a phonetic realization problem regardless of the degree of their problem in the phonetic and lexical-phonological representation of Japanese pitch accent categories.

**Figure 7.1. The average F0 patterns of NJ individual speakers.**

The panels represent A1 (top), A2 (middle) and A0 (bottom). The x-axis indicates time-normalized points. The y-axis indicates normalized F0 values.



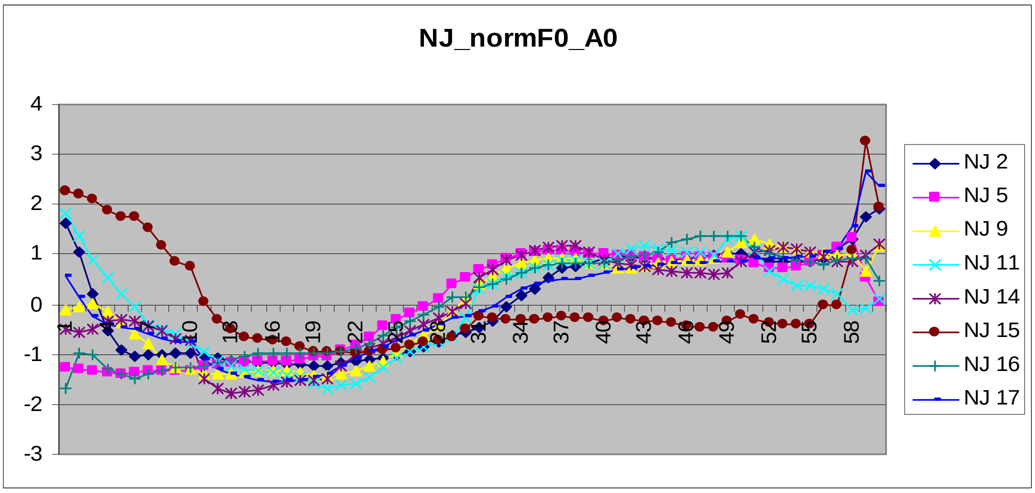



**Figure 7.2. The average F0 patterns of NEexp individual speakers excluding one non-categorical participant.**

The panels represent A1 (top), A2 (middle) and A0 (bottom). The x-axis indicates time-normalized points. The y-axis indicates normalized F0 values.

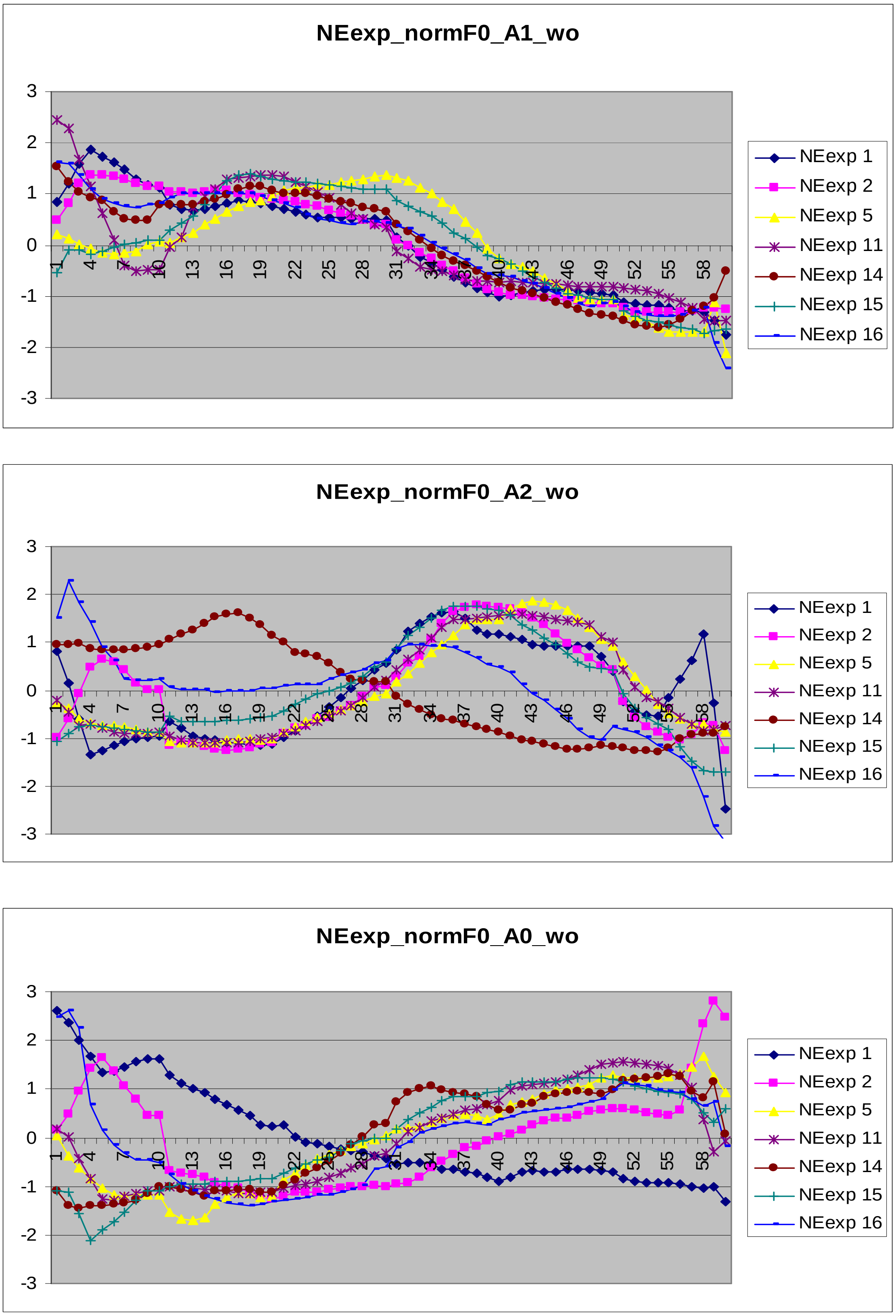
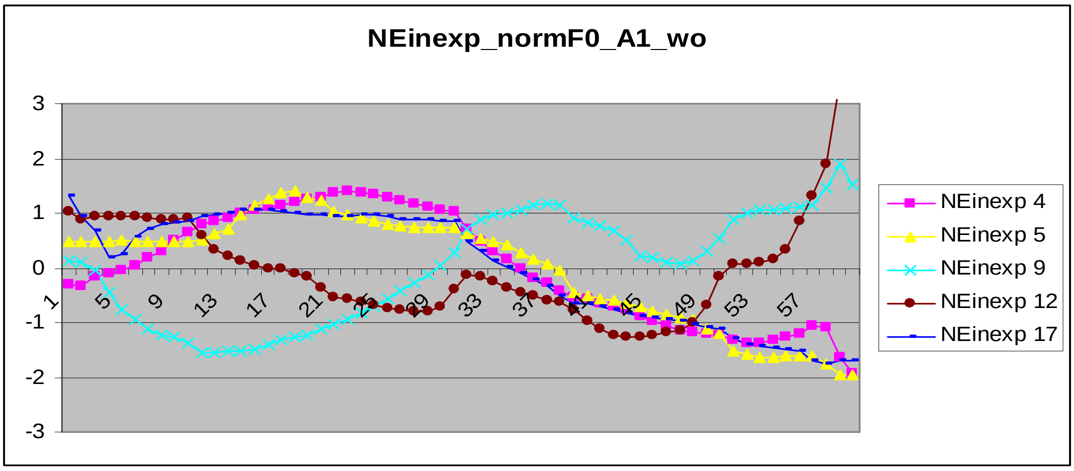
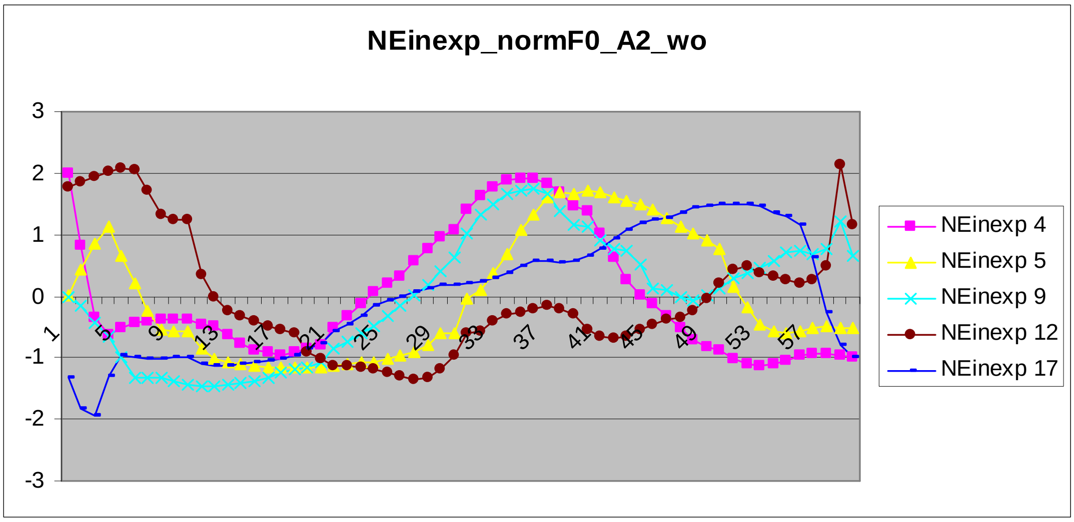
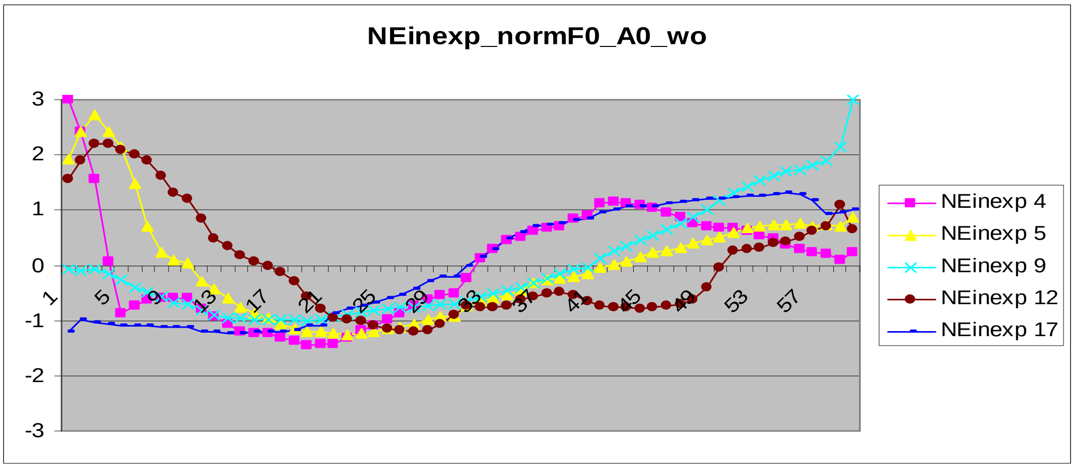


**Figure 7.3. The average F0 patterns of NEinexp individual speakers excluding three non-categorical participants.**

The panels represent A1 (top), A2 (middle) and A0 (bottom). The x-axis indicates time-normalized points. The y-axis indicates normalized F0 values.



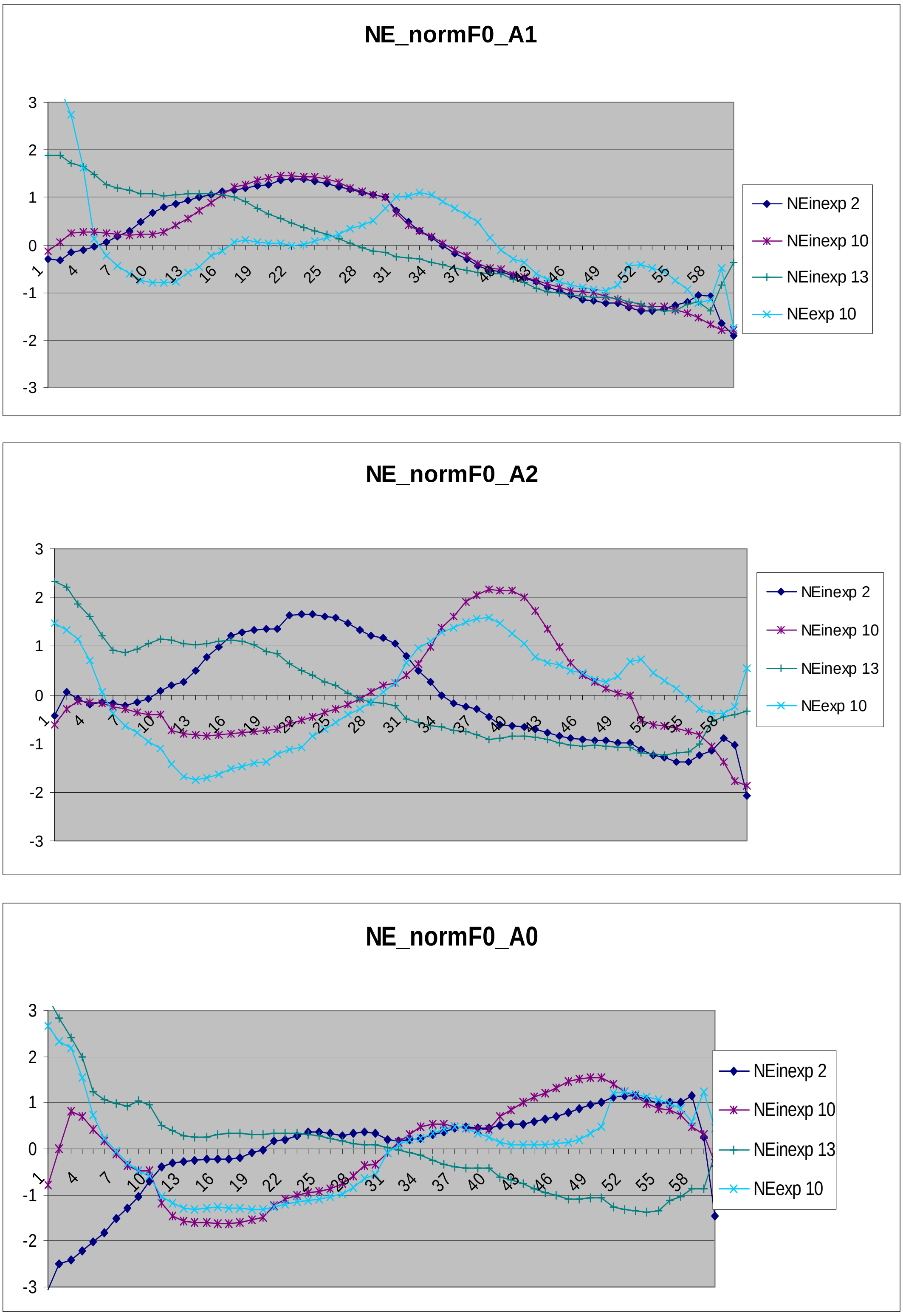



Since we now know that the learners tend to show non-target-like F0 patterns in production even if they indicated a categorical function in the step-wise ABX task, we need to investigate further whether or not the learners who showed a non-categorical function have non-target-like F0 profiles. Thus, we turn now to the non-categorical participants. Figure 7.4 represents the F0 contours in the productions of the non-categorical NE learners. When only the non-categorical learners are compared with the NJ group (as shown in Figure 7.1 above), the overall F0 patterns of these learners show individual differences. When we focus on each accent type, some learners seem to produce the shapes as if they were similar to the target norms in terms of the F0 peak location, the steepness of the F0 fall and/or the F0 range. However, when we look more closely, in the productions of two learners (NEinexp13 and NEexp10), the three accent types are not distinguished from each other in terms of the F0 shapes. In addition, in the productions of NEinexp2, the only distinction that is very evident is between accented and unaccented targets: both A1 and A2 were produced as A1, and they differed from A0. It could be that this learner has problems either in having two L2 categories instead of three, or in associating the F0 pattern with the target Japanese pitch accent category. Meanwhile, it is only in the productions of one of the non-categorical participants (NEinexp10) that the overall F0 patterns seem to resemble the shape of the target norm of each accent type (overlooking minor differences). It is possible that this participant is an exceptional case who has representation problems but is still able to show target-like F0 patterns in his speech productions. Alternatively, it is possible that he had a problem in understanding the step-wise ABX task. Although it may be difficult to generalize these four learners' patterns due to the limited number of the learners in this analysis, this tentatively shows that (apart from the one participant) the non-categorical learners tend to show non-target-like F0 profiles in their productions. There is a tendency for the non-categorical learners to show patterns which diverge from the target in terms of F0 peak alignment. Based on this tendency, we interpret that the learners who have representation problems also have a problem with phonetic realization as they attempt to implement F0 contours in the productions of Japanese pitch accent contrasts. In addition, there is an indication that representation problems of the L2 categories may be related to their phonetic realization problem.

**Figure 7.4. The average F0 patterns of non-categorical NE participants.**

The panels represent A1 (top), A2 (middle) and A0 (bottom). The x-axis indicates time-normalized points. The y-axis indicates normalized F0 values.



### 7.1.3 Non-speech production revisited

So far in this chapter, we have found that regardless of whether or not they showed a categorical function in the categorization task, the learners seem to show foreign accent. We interpreted this result as suggesting that in the early stage of learning, a combination of phonetic representation and lexical-phonological problem of the L2 categories and a phonetic realization problem can contribute to their foreign accentedness. In the later stage of learning, the learners may have a phonetic realization problem independently of representation problems. To understand the possible interactions of the problems associated with foreign accent, we now turn to investigate whether or not an articulation problem plays a role in contributing to foreign accent regardless of whether or not the learners showed a categorical function in the categorization task. According to the results of the reanalysis of this section, the learners tend to show different F0 patterns in the non-speech productions from native speakers.

As we saw in Chapter 5, the F0 patterns of the NJ group's non-speech productions resembled the characteristics of the patterns that we saw in their speech productions: earlier F0 peak for A1, later F0 peak for A2, steep falling contour after the peak for the accented targets, and relatively flat F0 contour for the unaccented target. In the initial analysis of the non-speech imitations, the overall F0 patterns produced by the English-speaking learners were much more similar to those produced by the native speakers than their speech productions (see 5.5.1). However, some differences were also revealed between the learners' productions and the native speakers' productions even though this was a non-speech imitation task where linguistic information was not involved. In addition, we saw more variance in the learners' F0 patterns than in the NJ group. To recap what we presented in 5.5.1, the F0 peak of A1 non-speech produced by the NEinexp and by the NEexp occurred later than that produced by the NJ. The F0 fall after the peak in the A2 non-speech produced by the NEexp was slightly less steep than that produced by the NJ. The shape of the A0 non-speech produced by the NEinexp appeared to be either flatter or having sharper fall than that produced by the NJ. In addition, the F0 ranges of the two accent types (A1 and A2) in the non-speech of the NEinexp were slightly narrower than those in the non-speech of the NJ.

Even when the non-categorical learners are excluded, the differences seen in the initial data continue to exist. Figure 7.5 reproduces the initial F0 contours in the productions of the NJ and is presented here again for comparison. The figures which follow it show the F0 contours in the productions of the two learner groups once the non-categorical participants were removed from the analysis (NEexp learners in Figure 7.6 and NEinexp learners in Figure 7.7). More variability in the learners' data, particularly in the NEinexp's group, can be seen.

**Figure 7.5. The average F0 patterns of NJ individual speakers in their non-speech productions.**

The panels represent A1 (top), A2 (middle) and A0 (bottom). The x-axis indicates time-normalized points. The y-axis indicates normalized F0 values.

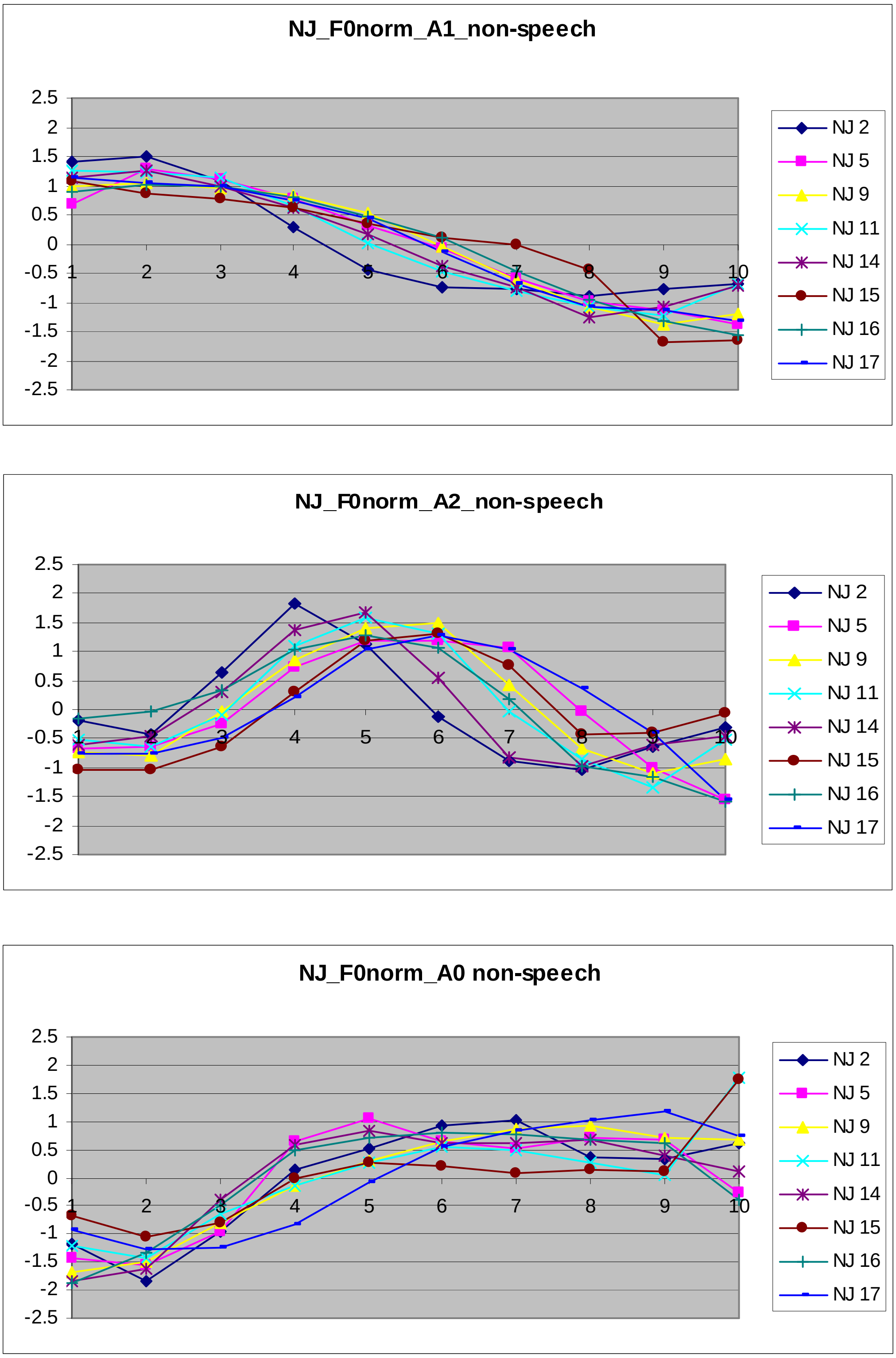
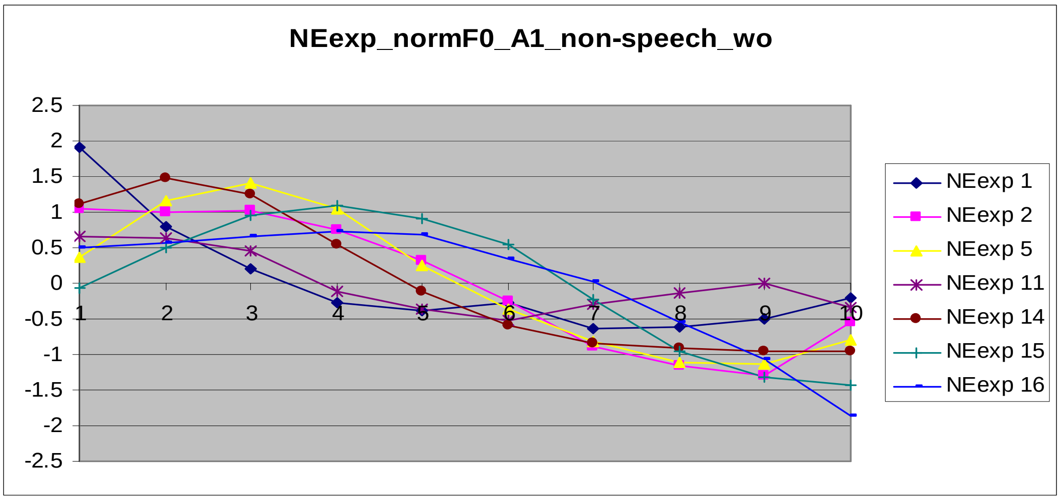
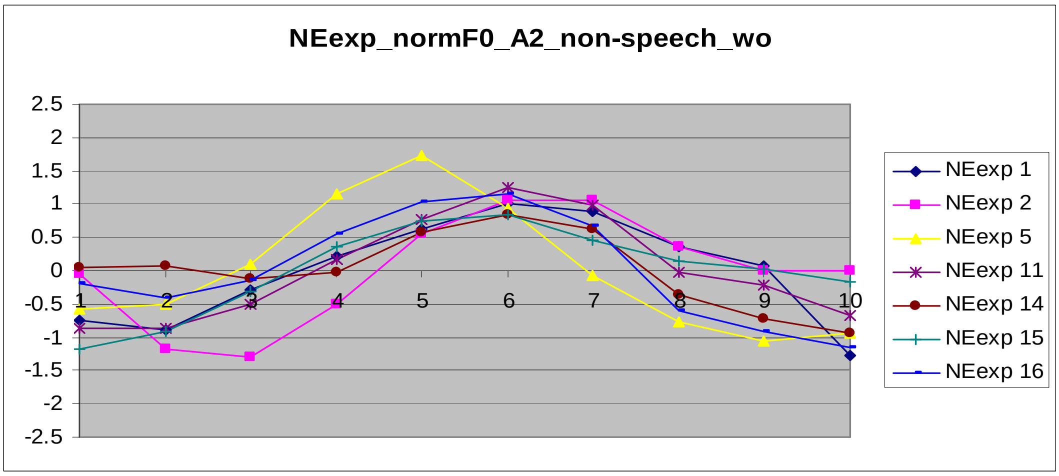


**Figure 7.6. The average F0 patterns of NEexp individual speakers excluding one non-categorical participant in their non-speech productions.**

The panels represent A1 (top), A2 (middle) and A0 (bottom). The x-axis indicates time-normalized points. The y-axis indicates normalized F0 values.



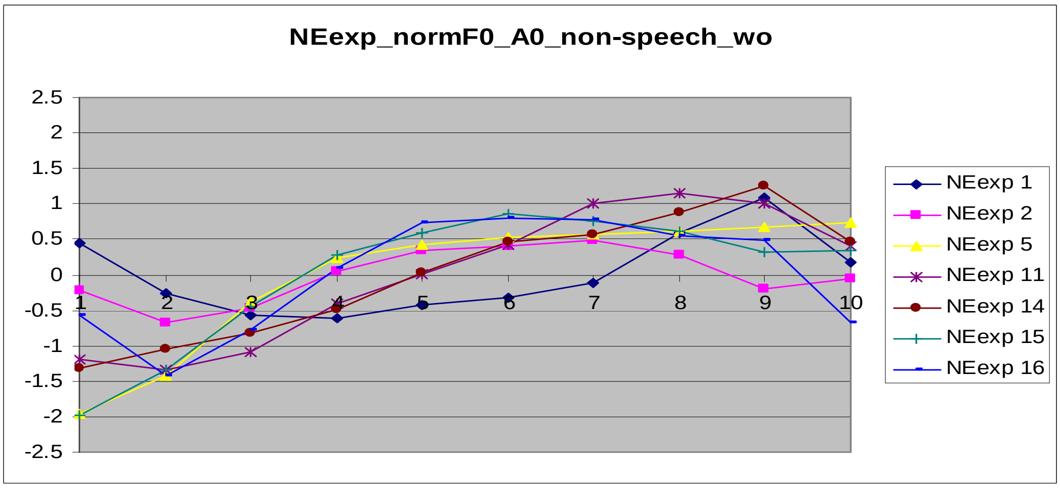



**Figure 7.7. The average F0 patterns of NEinexp individual speakers excluding three non-categorical participant in their non-speech productions.**

The panels represent A1 (top), A2 (middle) and A0 (bottom). The x-axis indicates time-normalized points. The y-axis indicates normalized F0 values.

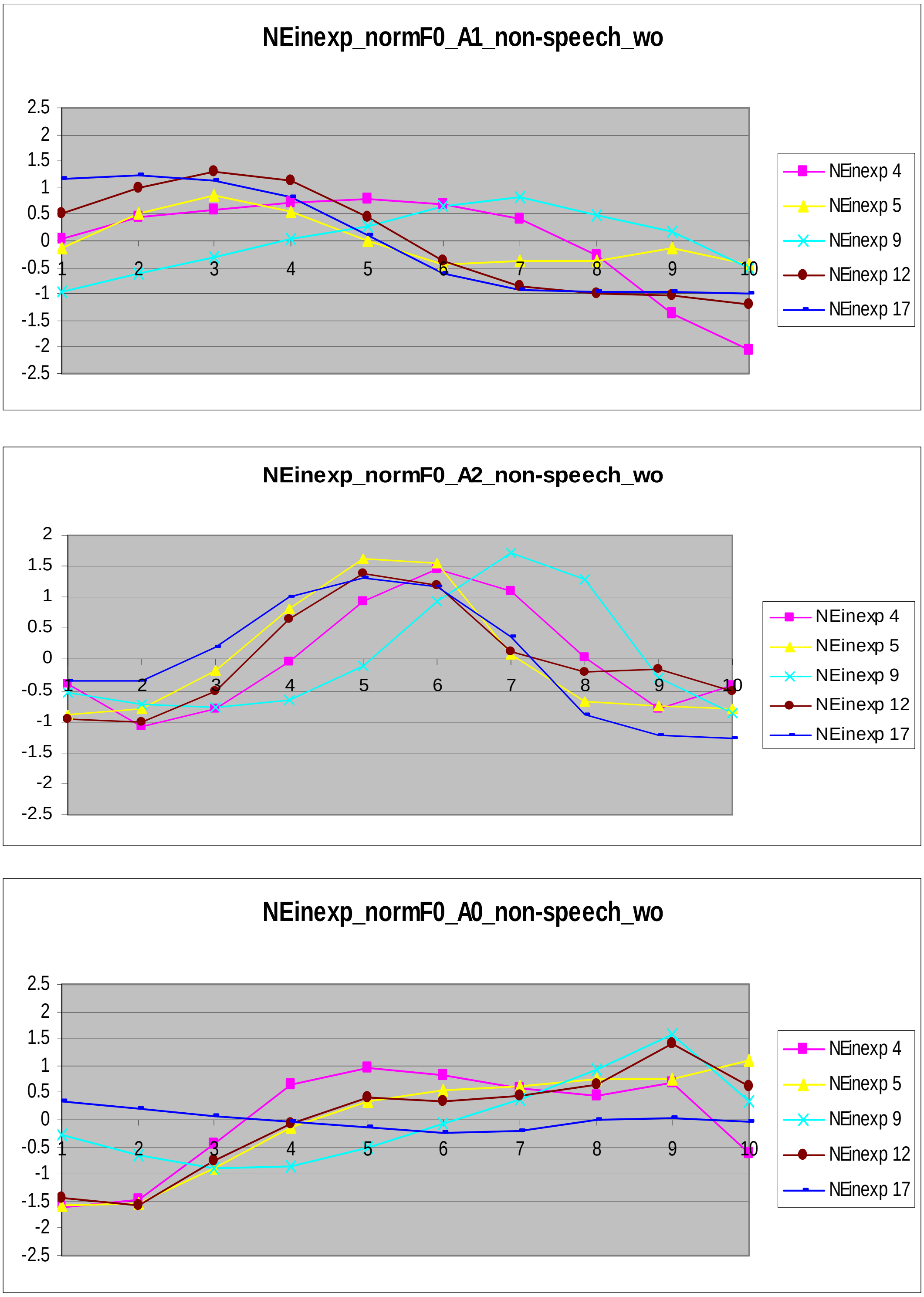


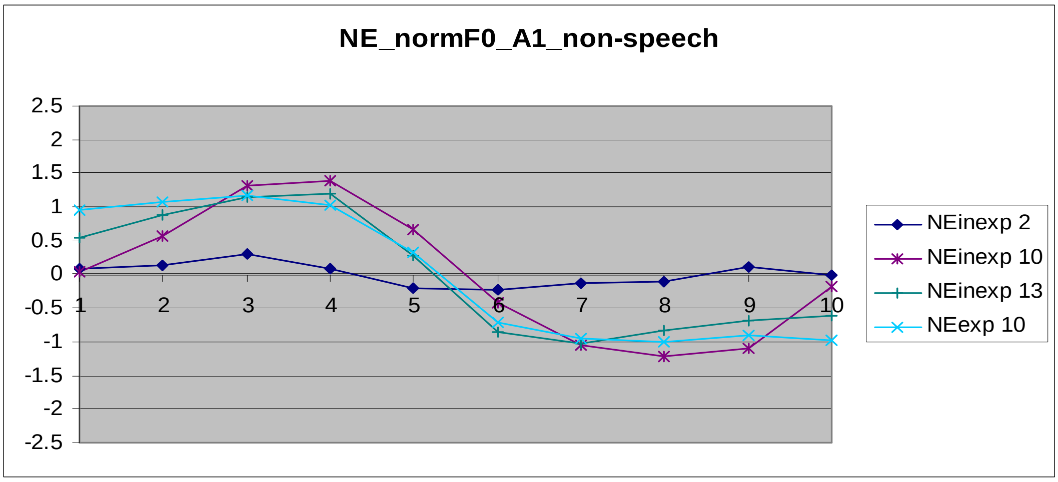
Figure 7.8 represents the F0 contours in the productions of the NE learners who did not show a categorical function. When only the non-categorical learners were compared with the NJ group (Figure 7.5), the overall F0 patterns of these learners show individual differences apart from A0 patterns. If we consider the participant labeled NEinexp13, it can be seen that, just as in the speech production task, in the non-speech task, the F0 shape of this participant is not differentiated for the three accent types. Considering NEinexp10, the overall F0 patterns of this participant's non-speech productions seem to resemble those in the non-speech productions of the NJ when minor differences are set aside. In addition to his speech data, this further supports the possibility that he may be an exceptional case, or found the step-wise ABX task too difficult.

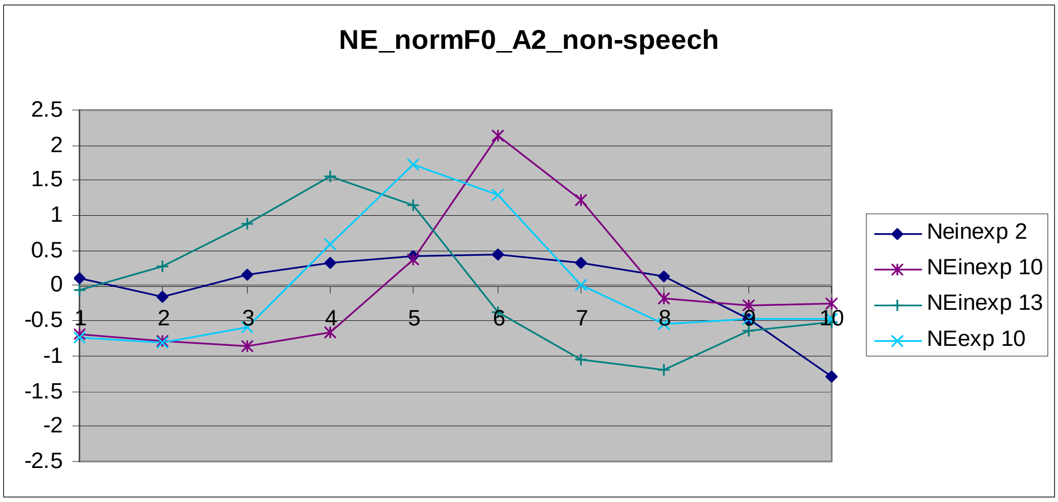
The patterns of NEexp10 and NEinexp2 are also worth commenting on. Although the F0 patterns in NEexp10's speech productions were not distinguished among the three accent types, in her non-speech productions the F0 patterns seem to be similar to the target shape. This suggests that her problem may not be in articulating the F0 contours of the target pitch accent types. Instead, because of representation problems, she may have difficulty in translating the phonetic information into speech in a linguistically meaningful way. The phonetic realization problem that we saw in her speech data could also be due to this representation problem. Meanwhile, for NEexp2, although the F0 patterns in his speech productions were distinguished, at least between the accented types and the unaccented type, in his non-speech productions the F0 patterns among the three accents seem to be no different from one another. It is possible that he might have been confused by the requirements of the non-speech task. In this case, he may not have an articulation problem. If his speech production pattern was at chance levels of performance, it is possible that he has an articulation problem. Overall, these four learners who indicated to have representation problems of the L2 categories tend to have an articulation problem. However, an articulation problem is not necessary for explaining their foreign accent. Although it is tentative, the results seem to suggest that representation problems of the target categories could be a major problem for these learners. However, this needs further research to confirm.

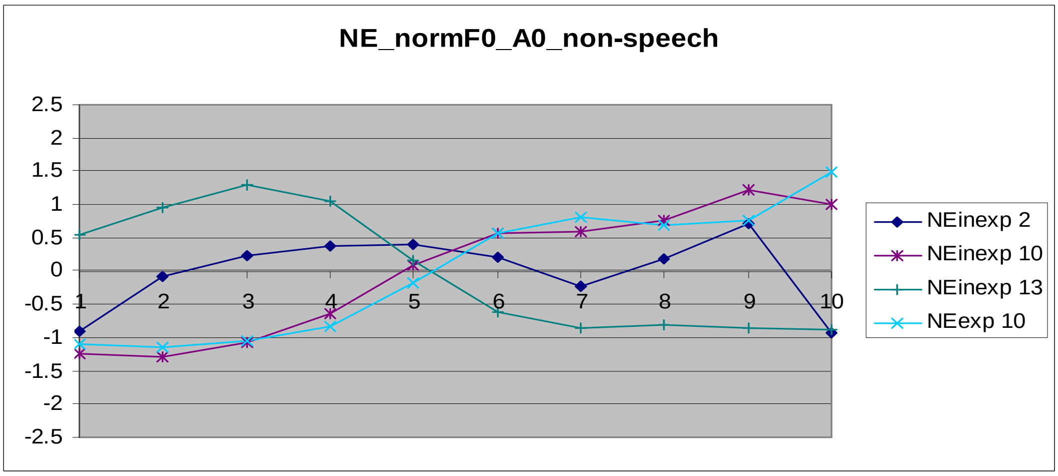
These results of the non-speech productions seem to suggest that some learners have an articulatory problem regardless of whether or not they showed a categorical function in the categorization task. The results also showed that a few learners might be capable of articulating the primary acoustic correlate of Japanese pitch accent, even though they may have representation problems. For those learners who have little representation problem, an articulation problem could be the key problem for their foreign accentedness in addition to a phonetic realization problem. Although this requires further investigation, for those learners who have representation problems, their major problem behind foreign accent could be rooted in representation problems, or in a combination of representation problems, an articulation problem and a phonetic realization problem.

**Figure 7.8. The average F0 patterns of non-categorical NE participants in their non-speech productions.**

The panels represent A1 (top), A2 (middle) and A0 (bottom). The x-axis indicates time-normalized points. The y-axis indicates normalized F0 values.



## 7.2 Discussion

The aim of this chapter was to investigate possible interactions between the factors which lie behind foreign accent in the acquisition of Japanese lexical pitch accent. More specifically, we investigated whether the learners have an articulation problem and/or a phonetic realization problem as part of the output process independent of representation problems of Japanese lexical pitch accent categories (part of the internal mechanism linking the input and the output processes), or whether learners have a combination of these problems. In order to achieve this, we re-analyzed the data of the previous chapters, excluding the learners indicated that they had problems in categorization and in lexical assignment by showing a non-categorical function in the step-wise ABX task and difficulty in the label assigning task (as discussed in Chapter 6). These learners seem to have problems in the ability to categorize Japanese lexical pitch accent categories, and to assign the categories to lexical items in Japanese. This implies that these learners have phonetic representation and lexical-phonological representation problems. We re-examined the speech production data which showed that native speakers of English show foreign accentedness (discussed in Chapter 3), and the non-speech data we showed that the learners also seem to have a problem with articulating the target F0 contours (discussed in Chapter 5). In these two previous chapters, the results indicated that Japanese lexical pitch accent produced by inexperienced learners was more foreign accented than that produced by the experienced learners and have an articulatory problem and a phonetic realization problem. Based on the findings from the categorization tasks (discussed in Chapter 6), the inexperienced learners seem to have problems with the representation of Japanese pitch accent categories, certainly relative to the experienced learners, who, although they are not perfectly identical to the native speakers, seem to have relatively few mental representation problems.

In this chapter, to understand the nature of the learners' problems, first, we examined whether non-target-like patterns are observed in the speech productions of the learners who showed a categorical function in the step-wise ABX task. Overall, the reanalysis of the speech data indicated that the English-speaking learners seem to show production patterns which diverge from the native Japanese productions, both in terms of the intelligibility score and in terms of the overall F0 patterns, regardless of whether or not they showed a categorical function. The learners generally seem to have problems with productions of the accented targets, particularly with items where the pitch accent is on the second syllable. It is possible that it is difficult for English speakers to learn where to aim at the location of the F0 peak, or how steep the F0 fall is, in order to signal Japanese pitch accent. As a group, the learners who seem to have representation problems of the L2 categories have even lower intelligibility scores than the learners who did not indicate representation problems. In addition, the productions of the learners whose categorical function was not categorical showed large individual differences in the overall F0 patterns. Hence, we cannot conclude that non-target-like productions that we discussed in Chapter 3 were due to the learners with representation problems, or that representation problems could be the fundamental factor for foreign accent. Instead, the results suggest that the learners who have few representation problems seem to have a phonetic realization problem that could be a factor behind foreign accent.

To investigate whether the learners who seem to have few representation problems also show non-target-like non-speech productions, we reanalyzed the non-speech data. The reanalysis of the non-speech data also indicated a similar tendency, although the overall F0 patterns in the learners' non-speech productions were more target-like than the patterns in their speech productions. The productions of the English-speaking learners seem to diverge from the native Japanese productions in terms of the overall F0 patterns regardless of whether or not the learners showed a categorical function. Although the experienced learners did show some divergence from the native norms, they showed much less divergence from the native target than either the inexperienced learners or the non-categorical learners. Nonetheless, the results also showed that some non-target-like patterns are to be observed, especially in the inexperienced learners' productions. Thus, the results suggest that the learners who have fewer phonetic and lexical-phonological representation problems of the L2 categories seem to have an articulation problem that could be a factor behind foreign accent.

We then examined whether the learners who indicated representation problems also have a phonetic realization problem (and an articulation problem if they also show non-target-like non-speech productions). This was in order to address whether a representation problem is the main problem for them, or whether their foreign accentedness is rather a combination of representation problems, a phonetic realization problem and an articulation problem. We found interesting patterns in the individual cases of the four learners who seem to have representation problems (manifested by their non-categorical function). NEinexp13 did not distinguish F0 patterns among the accent types either in the speech imitations or the non-speech imitations. In her case, we diagnosed that she seems to have representation problems as well as an articulation problem and a phonetic realization problem regarding foreign accent. NEexp10 also did not distinguish F0 patterns among the accent types in the speech imitations. However, she seems to distinguish F0 patterns among the accent types in the speech imitations, even though the patterns still show some divergence from the native patterns. It is possible that she was not able to interpret F0 information phonologically due to poor formation of the target L2 category, at least not enough for her to be able to associate the F0 information with Japanese pitch accent categories. Hence, this can be interpreted that her problem with foreign accent mainly stems from representation problems, and in addition she may have a phonetic realization problem. Meanwhile, NEinexp2 seems to distinguish the accented targets from the unaccented targets in his speech imitations, although he did not distinguish F0 patterns among the accent types in the non-speech imitations. This possibly implies that the target category (such as the accented and the unaccented categories) is emerging for him. If this inference is correct, his problem may also be due to a combination of factors: a phonetic realization problem and representation problems. In that case, it may be that if he did not understand the non-speech task, he would have been unable to make the distinctions in the non-speech imitations. The case of NEinexp10 needs further thought. NEinexp10 shows F0 patterns which are relatively similar to the target norms in both his speech and non-speech productions, even though these F0 patterns still show some divergence from the target patterns. Alternatively, if his speech results were merely coincidence, then this suggests that he also has an articulation problem with a combination of the two other problems. In his case, it seems that his foreign accentedness is due to representation problems. If this is correct, the results of NEinexp10 can provide evidence that an articulatory problem/a phonetic realization problem can be independent of representation problems. However, it is also possible that he is an exceptional case, since his mean intelligibility score was actually relatively good (88.8%) in spite of the fact that he does seem to have representation problems. Another possibility is that in his case, it might have been due to the task difficulty that his response pattern was showing as a non-categorical function. Overall, although this interpretation needs further investigation to generalize due to the limited number of cases, the results of these four learners seem to reflect the transitional stages between a phase where the learners have a combination of problems (i.e., an articulation problem, a phonetic realization problem and representation problems) and a phase where the learners mainly have an articulation problem, or a phonetic realization problem.

Therefore, the results from the re-analyses of the speech and the non-speech data seem to suggest that the learners who showed a categorical function may have not so much representation problems of Japanese lexical pitch accent categories (since at least they seem to have formed enough L2 categories to be able to assign the target category to an lexical item) as an articulation problem and a phonetic realization problem. This suggests that an articulation problem and a phonetic realization problem could occur independently of representation problems and give rise to foreign accent. Particularly, this seems to be the case at a later stage of L2 learning. Meanwhile, the results also indicate that the learners who showed a non-categorical function do seem to have representation problems as well as an articulation problem and a phonetic realization problem. Regardless of whether or not they showed representation problems of the target L2 categories, the learners showed foreign accentedness in their productions. We can speculate then that categorization ability and lexical assignment ability of the target L2 categories could assist the experienced learners to achieve more target-like productions than the inexperienced learners. The learners need to associate the F0 patterns that characterize the Japanese pitch accent categories and the target categories, and to associate Japanese pitch accent categories with words in Japanese. These abilities seem to require phonetic representations and lexical-phonological representations of Japanese lexical pitch accent categories. Hence, an articulation problem and a phonetic realization problem may not be completely independent of representation problems for the early stage of L2 learning.

Taking all these results into account, it seems that although both groups of learners have representation problems of Japanese lexical pitch accent categories and a phonetic realization problem and/or an articulatory problem to some extent, the experienced learners may have mainly a phonetic realization problem or an articulatory problem which might be independent of the representation problems of the L2 categories. These results also imply that the inexperienced learners have a combination of these three problems with the possibility that representation problems are the major element among them. Although all these problems seem to play a role in explaining the learners' foreign accent, the divergence between their production and categorization abilities might be more separate issues for the experienced learners, while they might be more closely related for the inexperienced learners.

## 7.3 Summary

In this chapter, in order to investigate potential interactions between the factors which are suggested to be behind foreign accent, the production data from Chapter 3 and the non-speech data from Chapter 5 were reanalyzed. We examined whether or not the learners who showed a non-categorical function in the step-wise ABX task revealed any patterns which differed from the original results, when considered separately as a group in their own right. This was particularly important as it offers to provide us with an explanation of what the learners' problems are and how these problems are related to each other. The results showed that the learners who did not indicate categorical function, who were mostly inexperienced learners, were having a problem in associating the phonetic information with the L2 categories, and in associating the L2 categories with lexical items in the L2. This seems to imply that their phonetic representations and lexical-phonological representations of the L2 categories have not fully developed, due to lack of L2 experience. The reanalysis of the production data in this chapter also showed another important finding. The learners showed patterns which diverged from the target norm, and they also showed great individual differences in both their speech productions and their non-speech productions. This tended to be shown consistently, independent of the problem with categorical function. However, the learners who did not show a categorical function seem to have more problems, although there were exceptional cases. It would seem that this difference should be interpreted as showing a role for mental representation problems of the target L2 categories. Moreover, the results also indicated that there is a function of L2 experience on the nature of the learners' problems regarding foreign accent. The important implication of this is that the experienced learners may have more phonetic realization problems or articulatory problems which might be independent of representation problems of the L2 categories, whereas the inexperienced learners may have more representation problems among the three problems.

To summarize, the reanalysis of the speech and non-speech imitations indicated that from the point of view of representation problems, the learners are also having a phonetic realization problem or articulatory problem. Considering these results, it seems to imply that not only mental representation problems but also a phonetic realization problem or articulatory problem provide an explanation for the divergent profiles of the learners, and thus can potentially explain foreign accent. Moreover, it seems that the nature of the problems regarding foreign accent may change during the L2 speech acquisition. The results also implied that the proportion of how much of these problems can explain foreign accent may change as the learners gain L2 experience.

# 8 General discussion & Conclusion

## 8.1 Summary of the findings and discussion

This section summarizes the crucial patterns regarding the acquisition of L2 pitch accent contrasts which were observed in the experimental data obtained from English learners of Japanese in order to evaluate the possible factors behind foreign accent.

As stated in Chapter 1, this thesis had two goals. The main goal was to evaluate the three factors which have been suggested as potential explanations of foreign accent in the acquisition of L2 prosodic contrasts. The three proposed factors were: 1) problems in differentiating the acoustic correlate of the target L2 contrasts, 2) problems in articulating the acoustic correlate of the target L2 contrasts, and 3) problems in categorizing the target L2 contrasts (i.e., linguistic perception, categorization and lexical assignment). The secondary goal of this thesis was to examine whether or not the nature of L2 learners' problems associated with their foreign accentedness change during L2 acquisition as a function of L2 experience.

With these goals in mind, this thesis addressed six questions: 1) whether or not the production patterns in Japanese lexical pitch accent produced by native speakers of English show non-target-like patterns when compared with those produced by native speakers of Japanese; 2) whether or not native speakers of English have problems in differentiating the target F0 contours; 3) whether or not English speakers have problems in articulating the target F0 contours; 4) whether or not English speakers have problems in categorizing Japanese lexical pitch accent contrasts in terms of linguistic perception, categorization and lexical assignment, 5) whether or not these three factors can explain non-target-like production patterns of English speakers; and 6) whether or not problems of English speakers change as they gain L2 experience.

To answer these questions, we have observed the L2 acquisition of Japanese lexical pitch accent by native speakers of English. We proposed that Japanese lexical pitch accent provides us with a good testing ground, a locus where there are cross-linguistic differences regarding the key acoustic correlate (i.e., F0) between English and Japanese while there are nevertheless similarities shared by these two languages. To learn L2 intonation patterns, L2 learners need to form the phonetic categories of the intonation patterns and also make the association between the target F0 patterns and the target function. However, to learn lexical pitch accent, L2 learners need to go beyond this, and also learn the arbitrary connection between the categories of the accent types and each lexical item in the L2. We are particularly interested in cases such as this, where a key acoustic correlate of L2 contrasts differs from the L1 in one of its functions while sharing another function with the L1, as in these cases there is great potential for shedding light on the issue of foreign accent. In these cases, the linguistic function of the key acoustic correlate of the L2 contrasts may not be completely foreign to the L2 learners, since they also use the shared function in their L1, yet they still need to learn the different function of the acoustic correlate as it is employed in the target L2. This may cause problems for the learners in learning the phonetic differences between the target L2 categories which signal the L2 function and in associating the phonetic information with each of the target L2 categories. We first investigated whether or not native speakers of English show foreign accentedness in their productions of Japanese lexical pitch accent contrasts (Chapter 3). Since the learners did show non-target-like productions, we were then able to approach the three potential factors regarding the issue of foreign accent.

Chapter 3 aimed to answer the first question: whether or not the English native speakers' production patterns in Japanese lexical pitch accent show non-target-like patterns when compared with those of native speakers of Japanese. We showed in this chapter that the inexperienced learners' productions of Japanese lexical pitch accent were perceived as more non-target-like than the experienced learners' by the two native speakers of Japanese who rated them. We also showed that the overall F0 patterns produced by the inexperienced learners showed more variance than those produced by the experienced learners. These findings provided us with evidence of the acoustic characteristics of English speakers' foreign accent that in their productions of Japanese lexical pitch accent. The productions of the English-speaking learners were non-target-like in terms of the auditory impression assessed by the native speakers, and they also showed non-native-like acoustic patterns. Hence, this also provided us with evidence that both experienced and inexperienced learners have problems in phonetically realizing the target F0 contours in their productions of Japanese lexical pitch accent contrasts. Particularly, based on the overall F0 patterns produced by the learners, they seem to have problems in phonetically realizing where the F0 peak is and how fast the falling contour is, in order to signal Japanese lexical pitch accent.

In Chapters 4, 5, and 6 we conducted a series of experiments on the L2 acquisition of Japanese lexical pitch accent by native speakers in order to evaluate the three potential factors behind foreign accent.

Chapter 4 addressed the question of whether or not the native speakers of English have problems in differentiating the target F0 contours. This was done by conducting a perception experiment using non-speech stimuli. In order to control cross-linguistic differences between Japanese and English, linguistic information was excluded from the stimuli: non-speech stimuli were created in such a way that only the F0 contour was extracted from the original sounds (canonical tokens of Japanese lexical pitch accent produced by a native speaker of Japanese). We showed that when linguistic information is controlled, English-speaking learners do not have problems in differentiating the target F0 contours, regardless of their L2 experience. This indicates that the learners' ability to hear the differences between the F0 contours is not a plausible source of foreign accent. In other words, the suggestion that foreign accent is caused by problems in differentiating the key acoustic correlate of L2 contrasts cannot explain our data. This offers evidence that in spite of their foreign accentedness, the English-speaking learners are not completely lacking in perceptual sensitivity to F0, the primary acoustic correlate of Japanese lexical pitch accent. Since F0 also has a linguistic function in English even though it does not have a lexical function, it is possible that native speakers of English retain their perceptual sensitivity towards F0 despite the different linguistic function of F0 in Japanese.

In Chapter 5, we investigated whether or not the English speakers have problems in articulating the target F0 contours. This was done by collecting non-speech productions where the participants had to imitate non-speech stimuli. We showed that the overall F0 patterns in the productions of the English-speaking learners (in their non-speech imitations) are relatively similar to those of the native speakers of Japanese when we focus only on the F0 contours. However, we also observed that the learners seem to show more individual differences in their F0 patterns than the native speakers, particularly in the case of the inexperienced learners. This suggests that the learners' ability to articulate the F0 contours may play some role in explaining foreign accent. This provides evidence in support of the proposal that the learners' problems with foreign accent can be partially associated with articulatory or motor control, or lack of L2 practice. As our data showed, both groups of native speakers of English were capable of articulating some rising or falling of F0 contours, even though the inexperienced learners showed more individual differences compared to native speakers of Japanese. Although English speakers have plenty of practice in articulating F0 contours in English (of course with different linguistic functions, such as intonation patterns), our findings imply that English-speaking learners have to learn to articulate the target F0 patterns which may be specific to the characterization of Japanese lexical pitch accent.

Chapter 6 aimed to answer the question of whether or not English speakers have problems in categorizing Japanese lexical pitch accent contrasts. This was done by carrying out three perception experiments to test three components of the learners' categorization ability. The first perception experiment used canonical tokens of Japanese lexical pitch accent contrasts to test their linguistic perception ability. This experiment showed that the English learners of Japanese are able to hear the difference between the tokens and to classify the target item into the correct L2 category based on similarity judgments. The second perception experiment used stimuli on the relevant continuum to test their categorization ability. This experiment showed that the learners have some problems in predicting category membership for items around the boundary of the L2 categories, which suggests that the learners seem to have problems in associating the phonetic patterns with the L2 categories. This implied that the learners seem not to have formed sufficient phonetic representation of the target L2 categories. We found that this tendency was stronger in the inexperienced learners than in the experienced learners. A subset of the learners, mainly the inexperienced learners (1 experienced learner and 6 inexperienced learners), were unable to categorize the intermediate stimuli on the continuum. This implied that their phonetic representations of the L2 categories largely overlap each other, or that they have not yet formed L2 categories. With the third perception experiment, which involved assigning labels to the L2 categories, we examined whether or not the learners have problems in assigning the L2 categories to a lexical item. The results implied that the learners seem to have problems in lexical-phonological representations of the L2 categories. We showed that whereas the experienced learners managed to lexically assign the corresponding L2 category to lexical items, the inexperienced learners still had problems.

The crucial finding of these three perception experiments was that the learners' ability to categorize Japanese lexical pitch accent contrasts seems to play a role as a factor which contributes to foreign accent. Interestingly, more than half of the variance in the foreign accentedness (53.5%) based on the auditory impression by the native speakers was explained by the variance in learners' categorization ability as a combination of lexical assignment and linguistic perception. The learners' lexical assignment ability was the variable which contributed most to their foreign accentedness as rated by native judges (38.6%). In addition, the learners' linguistic perception ability also explained some further variance in their foreign accentedness (14.9%). This evidence further supports the view that in order to be able to lexically assign a target L2 category to an lexical item, learners may not only need to be capable of hearing the differences, but they may also need to have formed the target L2 phonetic categories during their L2 speech acquisition. In addition, they need to learn the arbitrary link between the L2 categories and the lexical items. This might have been the reason why the inexperienced learners had more difficulties in lexical assignment than the experienced learners. Due to lack of L2 experience, the inexperienced learners may have not formed target-like L2 categories as part of their phonetic representations and may not have learned the connection between the L2 categories and lexical items as part of their lexical-phonological representations. To form target-like L2 categories, the learners may need to have representations that associate the phonetic information with the L2 categories, and to be able to assign labels to each L2 category. Meanwhile, the experienced learners may have phonetic representations and lexical-phonological representations of the L2 categories which are not fully developed but which are nevertheless more similar to target representations. Thus, we presented evidence that categorization ability is associated with foreign accent, and furthermore shed light on the specific aspects of categorization ability that are related to foreign accent in the L2 productions of Japanese lexical pitch accent.

Having observed that a subset of the learners did not show a categorical function in the data obtained in Chapter 6, and having seen that they tended to show different patterns from the rest of the group members, we turned in Chapter 7 to re-assess the learners' patterns discussed in Chapters 5 and 6 after excluding this subset of learners. We diagnosed that non-categorical performance in Chapter 6 indicates the degree of the problem that a learner has in representing the L2 categories. We addressed the question of whether the learners do or do not have a phonetic realization problem and/or an articulation problem, even with a marginal representation problem. This allowed us to specify the learners' problems and to evaluate in depth the factors which may be associated with foreign accent. After excluding the subset of non-categorical learners, the learners' group data indicated that the inexperienced learners managed to show lexical assignment ability just as well as both the experienced learners and the native speakers of Japanese. This suggests that once the learners were able to show categorical performance, even when they were still inexperienced, it is possible that they were able to form the target L2 phonological categories, even if they may not be fully target-like. This brought us to an important implication, namely that for the experienced learners, it may be mainly phonetic realization problems or articulatory problems which lie behind their foreign accent, whereas the inexperienced learners may have a combination of the three problems, with the possibility that phonetic representation and lexical-phonological representation problems are the major element. Hence, this suggests that the nature of the learners' problems behind foreign accent seem to change during the L2 acquisition of the target L2 contrasts.

## 8.2 Further improvements in the experimental design

One improvement in the experimental design could be made in the analyses of the production data (i.e., Experiment 1). In this study, to assess participants' production, two types of analysis were carried out: intelligibility scores based on native speakers' auditory judgments, and a descriptive analysis of the overall F0 contours in terms of the F0 peak, the degree of F0 fall and the F0 range. These analyses are an appropriate means of capturing and comparing the phonetic patterns in the productions of learners and native speakers, especially when there have been few studies conducted to compare productions of learners and native speakers regarding these three F0 parameters. However, these descriptive and qualitative analyses also have limitations in pointing out whether or not differences in the patterns carry any significance in more concrete terms. Therefore, an attempt to make a concrete quantitative analysis of these three F0 parameters has been carried out. However, due to difficulties in accurately quantifying these properties of the original data, the quantitative analyses of the data are not reported in this thesis. The details of the analysis attempts and the problems in quantifying the results are described in Appendix C. Further investigation of methods of quantitative methods is required. With further improvements in the accuracy of measurements, experimental design with quantitative data analyses such as those attempted here could potentially provide us with supporting evidence for the findings in this study and strengthen the arguments further.

Regarding the intelligibility scores, a forced choice paradigm was used in this study. It was important to assess whether or not the imitations of Japanese lexical pitch accent were identified by native speakers of Japanese as being what the speaker intended. It is also possible to use a rating scale to investigate the degree of foreign accentedness and to explore intermediate level of foreign accentedness. This may possibly allow us to further investigate which factors are contributing to learners' foreign accentedness during the developmental process of L2 speech acquisition.

Lastly, another improvement could be made in the label assigning task. One of the important findings of this study was that the ability to match acoustic information with lexical representations plays a significant role in the learners' foreign accentedness. However, this task was initially a by-product from a different task where the aim was to learn novel diacritics for Japanese lexical pitch accent. Hence, the methodologies for investigating this ability have room for further improvements, such as separating training and lexical assignment more precisely. To strengthen the arguments in this study, we could also investigate whether or not training on lexical assignment would reduce the foreign accentedness of the learners who showed poor lexical assignment.

## 8.3 Further issues

One issue which is not clear in our study based on the data presented here is whether or not the English learners of Japanese are able to produce Japanese lexical pitch accent like the native speakers when it is associated with intonation patterns. It is possible that the learners are unable to produce target-like productions when F0 is being used with the unfamiliar function of signaling lexical distinctions. The learners may not have completely mastery of what Japanese lexical pitch accent is due to lack of L2 experience. This tendency may be stronger in the inexperienced learners who have more limited L2 experience than the experienced learners.

The results of the three perception tasks in Chapter 6 showed that the English-speaking learners were having a problem in categorizing Japanese lexical pitch accent contrasts. Moreover, the result of the labeling task, where we examined the learners' ability to lexically map an item onto the corresponding L2 categories, further led us to an important speculation that the English speakers may be having a problem in associating the function of F0 in Japanese with Japanese lexical pitch accent categories as part of their mental representation. If the English-speaking learners were having a problem in associating the lexical function of F0 with their interlanguage categories of Japanese lexical pitch categories, it is also possible that this leads to difficulty in phonetically realizing target-like F0 in producing Japanese lexical pitch accent. To produce a target-like F0 profile in Japanese may require the learners to retrieve the information on the lexical function of F0 from their mental representation. Hence, the ability to associate the lexical function of F0 with the target L2 categories could be a good indicator of the type of non-target-like F0 patterns which native speakers of English will produce. For this reason, although not much attention has been paid to this ability as a problem of L2 speech acquisition, this could potentially shed light on a crucial aspect of L2 learners' ability that could help to explain foreign accent more fully.

The importance of the internalizing process of L2 linguistic knowledge, such as when the linguistic function is associated with the target L2 categories, has been pointed out as a key aspect of achieving native-like productions. Shehadeh (2003) argues that L2 learners are having difficulty in the process of internalizing linguistic knowledge. Swain and Lapkin (1995) suggest that noticing a problem will help learners to internalize their L2 knowledge and that this process enhances L2 proficiency in general. However, to my knowledge, there are few studies which have investigated the problem of internalizing lexical function of L2 with respect to foreign accent. Most probably this is due to difficulty in testing the internalization process empirically.

If the English-speaking learners are actually considering the tasks of Japanese lexical pitch accent as tasks regarding intonation patterns, we may see some difference between two conditions where the tasks are presented as a lexical task under one condition and as an intonation task under the other condition. More specifically, we would predict that the learners may perform better on tasks where Japanese lexical pitch accent is treated as a matter of intonation patterns rather than of lexical distinctions. However, this needs to be investigated further as a future issue.

## 8.4 Conclusion

The aim of this thesis was to investigate the factors behind foreign accent in L2 speech acquisition. The main goal was to account for foreign accent in adult L2 learners through evaluating these three potential factors, and the secondary goal was to provide empirical evidence of whether or not learners' problems with foreign accent change as they gain L2 experience.

Regarding the evaluation of potential factors behind foreign accent, our findings from the experiments led us to the following conclusions. In the acquisition of Japanese lexical pitch accent by native speakers of English, where the key acoustic correlate of L2 contrasts (i.e., F0) differs from their L1 in one of its functions while sharing another function with the L1, L2 learners do not have problems in differentiating the acoustic correlate of the target L2 contrasts. However, English-speaking learners do have problems in articulating the acoustic correlate of the target L2 contrasts, even when the focus is on purely the articulation of F0 contours without the linguistic information. Meanwhile, the learners' categorization ability, particularly their lexical assignment ability, which may involve the formation of lexical-phonological representations, was also found to be important to explain their foreign accentedness. Hence, this study has shown that it is not a single factor but a combination of factors (i.e., an articulation problem, a phonetic realization problem and a categorization problem) which is related to foreign accent.

This study has also shown that the problems that contribute to foreign accent may change during L2 acquisition. As learners gain more L2 experience, the main source of foreign accent seems to shift from phonological representation problems to phonetic realization or articulation problems. This suggests, importantly, that L2 learners are able to form their interlanguage phonological categories that are developing towards the target L2 phonological categories as a function of L2 experience.

There are therefore three contributions which this study offers to the field. The first contribution is that with a cross-sectional approach, we captured the part of the learning process where L2 learners' problems in terms of foreign accent emerge during L2 speech acquisition. The second contribution is that by testing the same groups of learners throughout the whole experiments in this study, we evaluated the factors associated with foreign accent in depth. We would like to emphasize that the analysis of multiple factors is our original and valuable contribution to the field. Up to now, individual studies have investigated only a single factor at a time. However, our findings show that there is a great deal of value in investigating multiple factors, as otherwise we are not able to capture the nature of foreign accent in L2 productions. The third contribution is that, by making use of the cross-linguistic difference in the function of an acoustic correlate which is in other respects shared between the learners' L1 and L2, we shed light on a valuable aspect of foreign accent in L2 prosodic acquisition. Although the linguistic function of the key acoustic correlate of the L2 contrasts is not completely foreign to the L2 learners, yet they still need to learn the different function of this acoustic correlate as it is employed in the target L2. Both the fact that the acoustic correlate shares a similar function in the learners' L1 and L2, and the process of learning the function that is specific to their L2, could potentially assist the formation of L2 phonological categories by allowing learners to make use of their phonetic and phonological representations of the equivalent L1 categories.

These three contributions have significant theoretical implications for current models of second language speech learning, including Flege (1995)'s Speech Learning Model (SLM) and Best (1995)'s Perceptual Assimilation Model (PAM). The SLM mainly focuses on experienced learners whereas PAM's targets are inexperienced learners. However, our findings suggest that L2 learners seem to have different problems in learning L2 contrasts depending on how much L2 experience they have. Hence, the current models need to be refined in such a way that they can explain the speech learning patterns both of experienced and of inexperienced L2 learners. Secondly, the current models focus on perception ability to predict L2 speech learning process. However, according to the present findings, it is not only learners' perception ability but also other abilities that seem to be relevant to foreign accent. It was shown here that articulation ability (which may involve language-specific phonetic settings) was also an important factor for foreign accent, but in current models, problems which stem from abilities other than perception cannot be predicted. This is therefore another point which needs to be incorporated into current models. Finally, the current models suggest that cross-language phonetic similarity between L1 and L2 brings perceptual problems for L2 learners (and therefore, production problems). However, our findings clearly indicate that English learners of Japanese can benefit from the cross-language phonetic similarity between English and Japanese regarding F0 in learning Japanese lexical pitch accent. This is not predicted by either SLM or PAM. This therefore also needs further investigation, along with the question of how best to define cross-language phonetic 'similarity' (e.g., Flege et al., 1997; Mennen et al., 2007). In these various ways, our findings call for current models of second language speech learning to be amended, in order to provide a more adequate account of the L2 acquisition process.

In addition to the theoretical insights for the field of L2 speech acquisition research, this study also offers practical insights for the L2 classroom. Some implications of our findings which can be put into practice include methods for assessing the productions of L2 learners, or where to put the focus in pronunciation practice in the L2 classroom in order to improve learners' productions towards target-like patterns.

Due to the complex nature of the L2 acquisition and the factors involved in speech production and perception, the issue of foreign accent requires much further investigation in many different directions. Nevertheless, this research constitutes another important step towards understanding the issue of foreign accent in the acquisition of L2 prosodic contrasts.

**Appendix A: Questionnaire for English speaking learners**

***Please fill in the following questions.*** *質問(しつもん)です。*

Name: _________ Age:__Area of study:____________
Native language(s): ______________ Age when you started learning Japanese:

Have you stayed in Japan? Yes/No
If "yes",
When how long for what

How long have you studied Japanese? Please be as specific as possible (e.g., Univ. of Edinburgh 2 years, 5 hours a week etc.).

Where Years Months hours/week

Where Years Months hours/week

Where Years Months hours/week

How often do you use Japanese? Please circle one and specify.
seldom, only in the classroom, sometimes, everyday, others:
(e.g., at work, during the meeting with customers, negotiating with them at home, watching Japanese films or listening to Japanese music regularly etc.)
Where when how
How often do you use Japanese with native Japanese speakers (circle one)?
seldom, only in the classroom, sometimes, everyday, others:

What percentage of the time do you use Japanese daily (0~100%)? _____%

Other foreign language(s) you have studied: ______how long
________how long
Do you have any certificates for Japanese language tests (e.g., *Nihongo Nooryoku Tesuto, Kanji Test* etc.)?
Test Level/Score
Are you able to participate in the 2nd session? Yes/No
If "yes", when is it convenient and when are you away from Edinburgh?
_________________Email:
Tel:_________.
Have you ever been diagnosed with a hearing impairment? Yes/No
Comments:

ありがとうございます！

**Appendix B: The non-equidistance algorithm by Dabrunz (2009)**

```
# Smoothing that does not expect equidistant points.
#
# For each point, the smoothed value is calculated by looking for points at equal distances around the
# current point and using their values for triangular smoothing. If no points can be found at those
# distances, for each point the values of the two nearest points are interpolated and used instead.
#
# The algorithm avoids the artifacts created by smoothing that is not aware of the point distances. In
# particular, the huge distortions to the derivates of the resulting curve are avoided.
#
# This algorithm applies triangular smoothing to non-equidistant points.
# The non-equidistant version is (C), devised and implemented by Olaf Dabrunz
# 11/May/2009.
#
procedure Smooth_NED curve_in$ curve_out$ width
	Create PitchTier... 'curve_out$' sampleStart sampleEnd

	for j from 1 to width				; make a triangular window of size = width
		if j < width / 2 + 0.5
			weight'j' = j
		else
			weight'j' = width - j + 1
		endif
	endfor

	select PitchTier 'curve_in$'
	smooth_end = Get number of points

	for i from 1 to smooth_end			; smooth from 1 to end
		n = 0
		smoothsample = 0.0
		select PitchTier 'curve_in$'
		sample_time = Get time from index... i
		for j from 1 to width
			; there are no points beyond the left or right boundaries of the
			; tier, need to detect that
			have_point = 1

			rawsample_time = sample_time - (round(width / 2) - j) / f0_sample_rate
		rawsample_index_left = Get low index from time... rawsample_time
		rawsample_index_right = Get high index from time... rawsample_time

			if rawsample_time < sampleStart or rawsample_time > sampleEnd or
				... rawsample_index_left < 1 or rawsample_index_right > smooth_end
				have_point = 0
			elsif rawsample_index_left = rawsample_index_right
				rawsample = Get value at index... rawsample_index_left
			else
				rawsample_left = Get value at index... rawsample_index_left
				rawsample_right = Get value at index... rawsample_index_right
				rawsample_time_left = Get time from index... rawsample_index_left
				rawsample_time_right = Get time from index... rawsample_index_right
```

```
                    # quantization rounding errors should not be a problem here
                    rawsample = rawsample_left +
                        ...     (rawsample_right     -     rawsample_left)     *
(rawsample_time - rawsample_time_left)
                        ... / (rawsample_time_right - rawsample_time_left)
                endif

                if have_point
                    smoothsample += weight'j' * rawsample
                    n += weight'j'
                endif
            endfor
            smoothsample /= n
            select PitchTier 'curve_out$'
            Add point... sample_time smoothsample
        endfor
endproc
```

## Appendix C: Difficulties in quantifying the three acoustic parameters (i.e., F0 peak location, degree of F0 fall and F0 range) in the production data and details of attempted analyses

In addition to qualitative analyses of the production data (i.e., intelligibility scores and the analyses of the overall F0 patterns reported in Chapter 3), attempts were made to quantify three F0 parameters, namely, F0 peak location, degree of F0 fall and F0 range. However, some difficulties in accurately quantifying these properties of the original data require further investigation. Hence, the quantitative analyses of the data are not reported in the main body of this thesis. Nevertheless, the details of such attempts are described here as it is still possible that such details can be a useful contribution to the field. First, the main problems in quantifying the three F0 parameters are described. Then, the details of the attempted data analyses are reported in terms of methods and results. Lastly, possible solutions which still require further research are suggested.

The aim was to capture overall patterns regarding the three F0 properties in the productions, and to compare the learners' patterns with native speakers' patterns. In order to investigate the overall F0 patterns in the productions while ignoring micro-level differences, the original data needed to be smoothed out. This involves some distortions and yields some artifacts in the data analyses. The main issues for this process are therefore how far to smooth out the original F0 contours and how to control the errors caused by smoothing algorithm (the degree of F0 fall in particular). Although some errors will remain after smoothing, at least we should be able to calculate other errors, and also to make an accurate quantification of the distortion by the algorithms if possible, in order that it can be resolved to some extent with further investigation. To avoid such artifacts from using scripts, it is in theory possible to manually detect each of the three F0 parameters for each utterance. However, in practice it is not efficient to rely solely on manual detection, considering the amount of data in this study. To quantify the three F0 parameters, the following attempts were carried out.

**Method**

**F0 peak alignment (F0 peak location)**

To quantify F0 peak location, the F0 peak was identified in relation to the segmental landmarks in each utterance. As discussed in Chapter 3 (3.4.3), in order to observe the F0 peak location relative to the segments, smoothed F0 values in each segmental interval were divided into 10 time-normalized points. The utterances in the speech imitation task each consisted of six segments, which yield six corresponding intervals (see Table 3.4 in Chapter 3). Thus, there were 60 points for each utterance. The location of the highest F0 in the overall F0 contours among these 60 points was measured as the F0 peak point. To detect the time-normalized F0 values, Xu (2008)'s PRAAT script was used; this script automatically saves the trimmed F0 values.

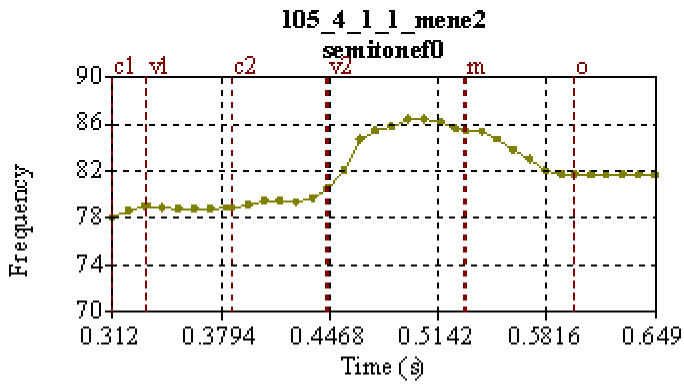
**Maximum velocity (degree of F0 fall)**

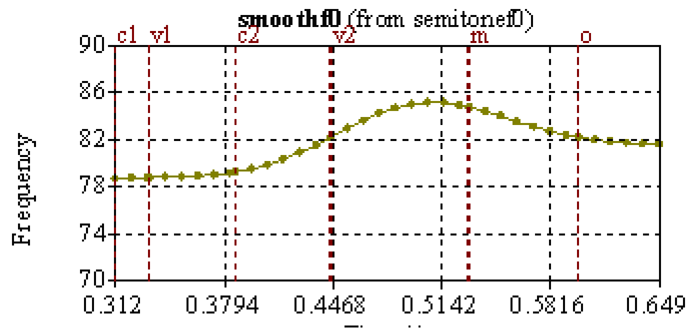
As discussed in Chapter 3 (3.4.3), what we are interested in is the degree of F0 fall after the F0 peak. For pitch accented words, F0 patterns in the productions of native speakers of Japanese tend to show a rapid fall after the F0 peak, whereas for unaccented words, F0 patterns does not show such a rapid fall.

To calculate the velocity of each F0 contour, Xu's (2008) PRAAT script was modified and used for the analysis. There were two steps. These steps yielded problems of artifacts and distortions, although some distortions were expected due to the nature of smoothing processes. First, the F0 contours were smoothed out with a smoothing value of 0.15 (window size 6.7 Hz). With this relatively large smoothing value, all the micro-level changes in F0 were removed from the F0 contours (see Figure 1) while preserving the global pattern of the original pitch data. We also found that micro-level changes in F0 tend to yield extremely high velocity values. Thus, smoothing the F0 contours allows us to capture the general shape of the contours and the degree of F0 fall. At the same time, this involved compromising the original values, as we can see in the two panels in Figure 1.

**Figure 1. An example of F0 smoothing.**

The top figure illustrates the trimmed F0 contour before smoothing. The figure below illustrates the smoothed F0 contour of the same utterance. In both figures, the x-axis represents the actual time in seconds and the y-axis represents the frequency in semitones.

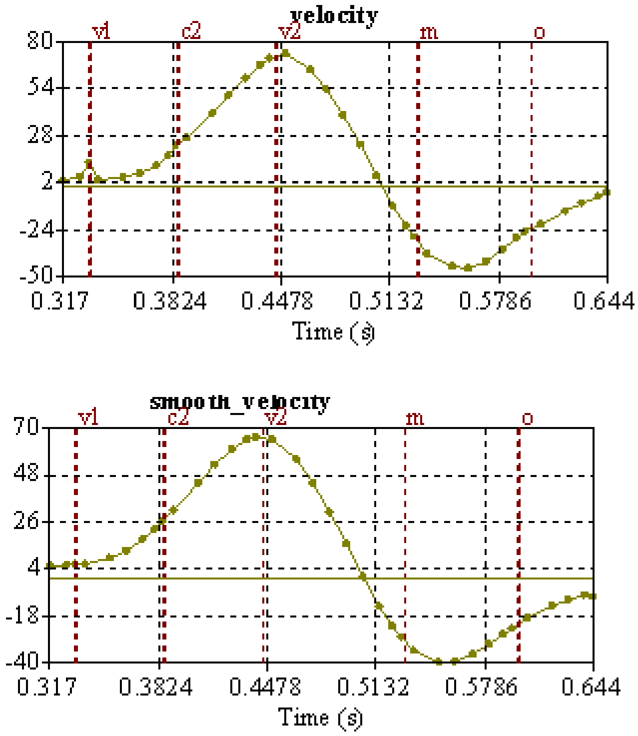



The second step of the calculation was that the maximum velocity after the F0 peak for each utterance was detected individually from the velocity values, and visually inspected to confirm that the value was correct. The triangular filter algorithm in Xu's script was used for calculating the velocity to smooth out the F0 contours with a smoothing value of 0.05 (window size 20 Hz). However, the velocity contour depicted by his script tended to show errors around each segmental boundary area in my data. This turned out to be a problem with the triangular filter algorithm in Xu's script which related to applying the algorithm to my data in combination with errors induced by other factors such as creaky phonation. Whereas the algorithm assumes equidistance between points, this is actually not the case with the points next to the boundaries, because the sampling is applied in each segmental interval. In some cases the distance between the two points near a segmental boundary is very small, and this yielded extremely large velocity values. By using the non-equidistance algorithm written by Dabrunz (2009) (see Appendix B) as a substitute for the triangular filter algorithm within Xu's script, these errors were considerably reduced and thus, Dabrunz's algorithm was applied for the velocity calculation in this analysis. Once the velocity values were calculated, each velocity contour was further smoothed out with a smoothing value of 0.075 (window size 13.3 Hz). This is illustrated in Figure 2. As mentioned earlier, there is a tradeoff between a large smoothing value and the accuracy of measurements (see Hertrich & Ackermann 1997; Xu, 1999), and therefore, the smoothing value was as low as possible while still being sufficient to provide with us the global shape of the F0 contour. However, this process again accumulates further artifacts.

**Figure 2. An example of velocity smoothing.**

The top figure illustrates the velocity contour before smoothing. The lower figure illustrates the smoothed velocity contour of the same utterance. In both figures, the x-axis represents the actual time in seconds and the y-axis represents the velocity in semitones per second.




**F0 range**

As we also reported in Chapter 3 (3.4.3), in order to make comparisons among speakers possible, the F0 values (Hz) on the time-normalized scale were converted into normalized F0 values using Z-score transforms (e.g., Earle, 1975). We quantified the width of the F0 contours (i.e., difference between the minimum and the maximum values) in each utterance. To measure F0 range by itself is not a big problem. However, the above mentioned smoothing issues also affect the F0 range values, as smoothing changes the original shape of the F0 contours.

**Results: Mean F0 values**

The sound files of the *mene* data set were analyzed as representative of the whole data in terms of their acoustic properties. There were 216 tokens analyzed (24 participants x 3 accent types x 3 repetitions) (details are discussed in 3.5 in Chapter 3). In addition, the corresponding 216 non-speech tokens in the non-speech imitations were analyzed. In this Appendix, only the speech data analyses are described. Annotation and acoustic analyses were performed using PRAAT as described in Chapter 3.

**F0 peak alignment (F0 peak location)**

The means and standard deviations of the F0 peak alignment as produced by the native Japanese speakers and the two groups of the learners are presented in Table 1. The values are based on the F0 peak location expressed as a percentage of the duration of each utterance based on the time-normalized point in relation to the segmental landmarks. To make sure that the location of the F0 peak was correctly identified, special care was taken when the F0 peak locations happened to be the beginning or the end of the utterance. If the peak alignment value is 0%, this means that the F0 peak occurs at the onset of the first consonant, i.e., the onset of the whole utterance. If it is 100%, the F0 peak occurs at the offset of the last vowel, i.e., the offset of the whole utterance. Figure 3, Figure 4 and Figure 5 represent the mean F0 contour of each accent type among the three groups for reference. These contour values are the values after trimming to remove outstanding bumps but before smoothing. Thus, in these figures, the large bumps of the F0 contours have been clipped but the contours still show bumpy curves.[1]

[1] These values, i.e., the trimmed but not smoothed F0 contours (as originally designed in Xu (2008)'s script), were used to detect the locations of F0 peak on the time-normalized scale. Xu's purpose was to detect precise phonetic values for F0 peaks which are as close as possible to the authentic values. However, for any future analyses with aims similar to this study (i.e., to capture a global view of the F0 contours), if it is technically and practically possible to detect the F0 peak alignment of the global F0 contours, it might be more beneficial to use F0 peak locations that are detected after the target F0 contours are smoothed, in order to avoid the marginal errors that are caused by bumps in unsmoothed F0 contours.

**Table 1. The mean F0 peak alignment of each group (%).**

The column "Mean (/60)" next to the mean indicates the F0 peak point on the time-normalized scale where each segment has 10 time-normalized points.

| | Accent type | Mean | Mean (/60) | Std. Deviation |
|---|---|---|---|---|
| NJ | Accent0 | 75.00 | 45.00 | 35.66 |
| | Accent1 | 25.22 | 15.13 | 15.57 |
| | Accent2 | 62.15 | 37.29 | 4.38 |
| NEexp | Accent0 | 53.51 | 32.10 | 41.54 |
| | Accent1 | 21.32 | 12.79 | 26.04 |
| | Accent2 | 57.43 | 34.46 | 32.44 |
| NEinexp | Accent0 | 51.04 | 30.63 | 45.55 |
| | Accent1 | 42.61 | 25.57 | 37.03 |
| | Accent2 | 52.46 | 31.48 | 35.39 |

**Figure 3. Mean F0 contour for A0.**

Each F0 contour represents each group's mean. F0 contours are trimmed but not smoothed.

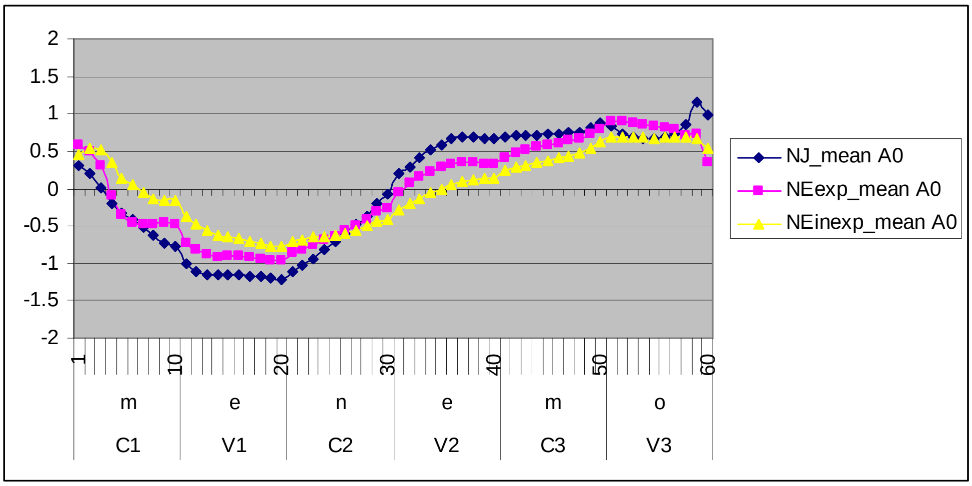


**Figure 4. Mean F0 contour for A1**

Each F0 contour represents each group's mean. F0 contours are trimmed but not smoothed.

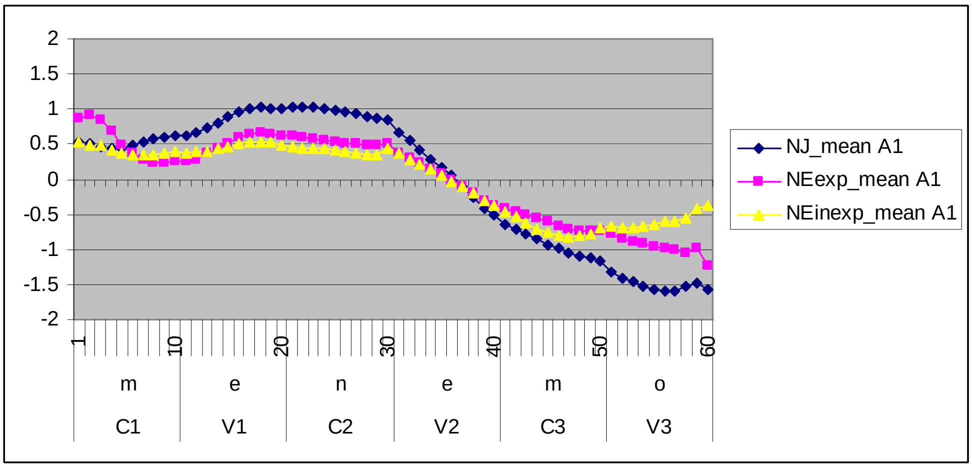
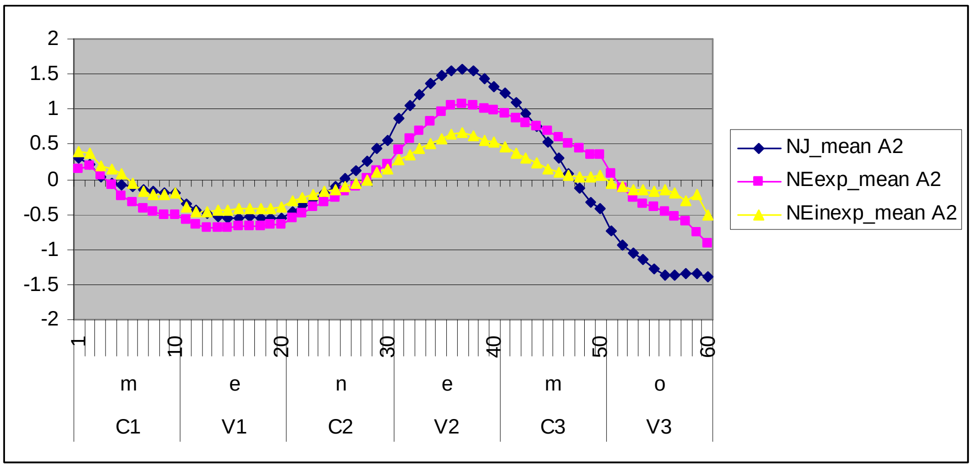



**Figure 5. Mean F0 contour for A2**

Each F0 contour represents each group's mean. F0 contours are trimmed but not smoothed.




The mean F0 peak location for A1 produced by NJ occurred at 25.2% of the utterance (using the time-normalized values). This indicates that for NJ the F0 peak occurred within the first syllable, which is the accent bearing unit of A1. More specifically, in A1, the F0 peak aligned with the $15^{th}$ time-normalized point, which is 51.3% of the first vowel (V1). In the case of A2, the mean F0 peak occurred at 62.2% in NJ's productions. This means that in A2 the F0 peak is located within the second syllable, which is the accent bearing unit of A2. More specifically, for A2, the F0 peak aligned with the $37^{th}$ time-normalized point, which is 72.9% of the second vowel (V2). In the case of A0, the mean F0 peak or the highest F0 occurred at 75.0% of the whole utterance. This is the $45^{th}$ time-normalized point, which is within the last syllable, which corresponds to the post-positional particle *mo*, which is 50% of the third consonant (C3). There are a few slight mismatches between the mean F0 peak alignment values in Table 1 and the mean F0 values in Figure 3, Figure 4 and Figure 5. The main reason for these mismatches seems to be that marginal errors in individual F0 contours (due to the bumps that remained on the F0 contours after trimming) were accumulated, and the mean F0 peak in Table 3.5 seems to have caused the F0 peak locations to be slightly shifted from the mean F0 figures shown in Figure 3, Figure 4 and Figure 5.

The analysis of the F0 peak alignment showed that three groups produced different patterns across the three accent types. According to a mixed design ANOVA (Accent (3 levels) x Group (3 levels)), there was a significant interaction effect ($p$ <0.001) as well as a main effect of Accent ($p$ <0.001), but the effect of Group was not significant. A set of one-way ANOVAs among the three accent types within each group indicated that the F0 peak for A1 was earlier than A0 and A2 in the NJ's productions. The same pattern was observed in the NEexp's productions. Meanwhile, the F0 peak location was not differentiated in the NEinexp's productions.

Another set of one-way ANOVAs was also conducted among the three groups regarding each accent. The only significant difference was that F0 peak locations for A1 in the NEinexp' productions was later than the peak locations in the NEexp's productions. However, they were not statistically different from the NJ's values. We should also note large SD values in the learners' means which indicates large individual variability in the two learner groups. Such variability was observed in the learners' individual F0 patterns in Chapter 3 (3.5.2). In addition, the mean F0 peak values in the productions of the NEexp and those of the NEinexp seem to be lower than the values in the productions of the NJ.

**Maximum velocity (degree of F0 fall)**

The means and the standard deviations of maximum velocity after the F0 peak are presented in Table 2. As it was the maximum velocity after the F0 peak which was detected, all the mean maximum velocity values were negative, indicating that the greatest speed after the F0 peak was a falling contour. When the F0 peak happens to occur at the end of the F0 contour, the velocity is zero as there is no following point. For these cases, after carefully checked the contours, the maximum velocity at the previous point was used. This sometimes showed a positive value, indicating a rising contour. As expected, the mean maximum velocity in NJ's utterances of the accented words, i.e., A1 and A2, showed relatively large negative values. In contrast, we see a much smaller maximum velocity value in NJ's productions of unaccented words, i.e., A0. Meanwhile, as can be seen in Figures 3, 4 and 5, the mean F0 patterns in the learners' productions were less curvy than those in the NJ's. This tendency appears to be stronger in the mean F0 fall in the productions of the NEinexp than those of the NEexp. This indicates that the mean degree of F0 fall in the learners' productions is less rapid compared to that in the NJ's.

**Table 2. The mean maximum velocity of each group (semitones/second)**

| | Accent type | Mean | Std. Deviation |
|---|---|---|---|
| NJ | Accent0 | -1.54 | 14.35 |
| | Accent1 | -41.05 | 12.58 |
| | Accent2 | -50.22 | 16.06 |
| NEexp | Accent0 | -13.54 | 16.13 |
| | Accent1 | -29.99 | 15.17 |
| | Accent2 | -28.72 | 18.25 |
| NEinexp | Accent0 | -12.42 | 17.01 |
| | Accent1 | -21.47 | 19.05 |
| | Accent2 | -25.09 | 17.54 |

The analysis of maximum velocity (i.e., the degree of F0 fall) also indicated that the learners' patterns were different from the native speakers' norms. According to a mixed design ANOVA (Accent (3 levels) x Group (3 levels)), there was a significant interaction between Accent and Group ($p$ <0.001) as well as a main effect of Accent ($p$ <0.001) and a main effect of Group ($p$ <0.01). Post hoc comparisons indicated that the maximum velocity produced by both learner groups was significantly different from that produced by the NJ (NEexp: $p$ < 0.05; NEinexp: $p$ <0.001) but the learner groups did not differ from each other. One-way ANOVAs were conducted among the three accent types within each group. In the NJ's productions, the maximum velocity for A0 was smaller than for A1 and A2, and the maximum velocity for A1 was smaller for A2. A similar pattern was observed in the NEexp's productions but the maximum velocity of A1 was not significantly smaller than that of A2 in their productions. In the NEinexp's case, the maximum velocity was not differentiated among the three accent types. Another set of one-way ANOVAs was also conducted among the three groups regarding each accent. The results indicated that the maximum velocity for A0 was faster ($p$ <0.05) in the learners' productions than the NJ's values while the maximum velocity for A1 ($p$ <0.001) and A2 ($p$ <0.001) were slower than the NJ's value.

**F0 range**

The means and the standard deviations of the F0 range are presented in Table 3. The mean F0 range for A1 in NJ's productions seems to be slightly narrower than that of A0 and A2. The mean F0 range of A0 and A2 happens to be the same in NJ's productions. Similar F0 range values for A0 and A2 is not surprising considering that these two accent types show similar F0 contour patterns up to the F0 peak. Meanwhile, the mean F0 range for A1 seems to be narrower in the learners' productions compared to that in the NJ's productions. The mean F0 range for A2 seems to be wider in the learners' productions compared to that in the NJ's productions.

**Table 3. The mean F0 range of each group (normalized value with Z-score transforms)**

| | Accent type | Mean | Std. Deviation |
|---|---|---|---|
| NJ | Accent0 | 3.62 | 0.82 |
| | Accent1 | 3.04 | 0.33 |
| | Accent2 | 3.62 | 0.47 |
| NEexp | Accent0 | 3.76 | 0.86 |
| | Accent1 | 3.30 | 0.48 |
| | Accent2 | 3.41 | 0.57 |
| NEinexp | Accent0 | 3.78 | 0.79 |
| | Accent1 | 3.38 | 0.67 |
| | Accent2 | 3.61 | 0.86 |

In the analysis of F0 range, a mixed design ANOVA revealed that the main effect of Accent was significant ($p < 0.001$) but there was no significant main effect of Group and no significant interaction between Accent and Group. Another set of one-way ANOVAs (3 levels) showed that it was only the NJ whose F0 range was significantly different among the three accent types; their F0 range was narrower for A1 than for A0 ($p < 0.001$) and A2 ($p < 0.001$). On the other hand, the F0 range produced by the learners did not yield significant differences among the three accent types.

Overall, regarding the three F0 parameters, the results of the quantitative analyses indicated similar results to the tendency found in the qualitative analyses of overall F0 patterns reported in Chapter 3. In short, F0 patterns in the productions of the learners are non-target like in the following ways: learners have a less rapid fall after the F0 peak compared to the NJs; learners show great variability in the mean F0 peak location values; and the learners' divergence from the NJ patterns is greater when L2 experience is more limited. This suggests that with further improvements in the accuracy of measurements, experimental design with quantitative data analyses such as those attempted here could potentially provide us with supporting evidence for the findings in this study and strengthen the arguments.

As one of possible solutions to reduce errors and increase accuracy, we could limit the number of the participants (and thus the amount data), and solely conduct the acoustic measurements manually. In this case, the limited number of participants would leave us with the issue of how far we could generalize the findings. Another possibility is that for any future analyses with aims similar to this study (i.e., to capture a global view of the F0 contours), if it is technically and practically possible to detect the F0 peak alignment of the global F0 contours, it might be more beneficial to use F0 peak locations as detected after the target F0 contours are smoothed, in order to avoid the marginal errors that are caused by bumps in unsmoothed F0 contours. Another possibility is to conduct further research on how to reduce errors from detecting overall F0 contours and from smoothing in order to increase the accuracy of calculating the speed of F0 fall. It would be particularly useful to explore methods of getting rid of the accumulated errors, and/or methods of dealing with flattening slope. However, this needs further thought and further investigation.

## Appendix D: Pilot study: three novel diacritics for eliciting Japanese pitch accent productions.

The three diacritic systems (accent on 1st syllable, accent on 2nd syllable, and unaccented)

1) line/bracket " —"/" ¬ "

めね　　めね　　めね

2) iconic object

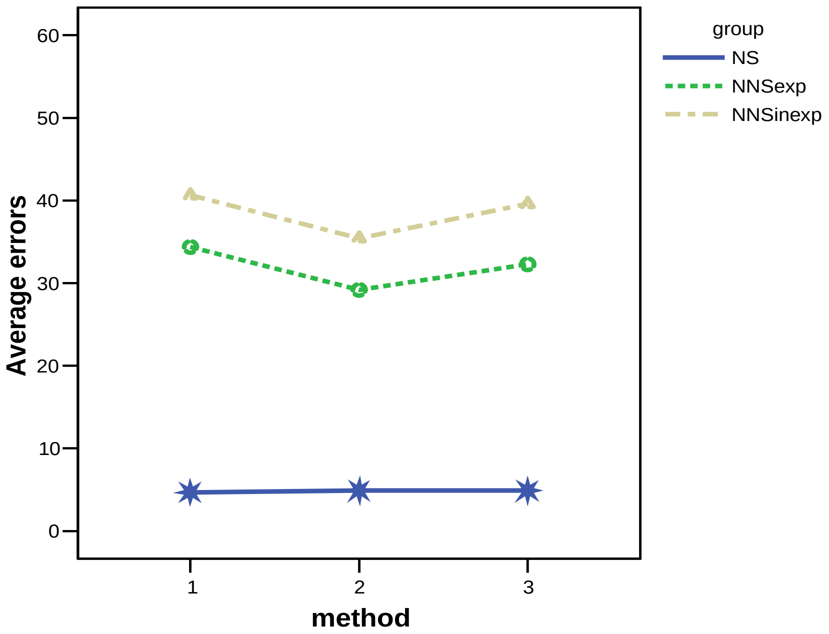
めね 　めね 　めね 

3) point " ´ "

め´ね　　めね´　　めね

**Results of production performance with three different accent markers**

**Average errors in production with three accent markers**



# Bibliography


Al-Issa, A. (2003). Sociocultural transfer in L2 speech behaviors: evidence and motivating factors. *International Journal of Intercultural Relations, 27*, 581-601.

Barry, W. J. (1989). Perception and production of English vowels by German learners: instrumental –phonetic support in language teaching. *Phonetica, 46*, 155-168.

Beckman, M. E. (1986). *Stress and Non-Stress Accent.* Dordrecht: Foris Publications Holland.

Best, C. T. (1995). A direct realist view of cross-language speech perception. In W. Strange (Ed.), *Speech Perception and Linguistic Experience: Issues in Cross-Language Research* (pp. 171-204). Baltimore: York Press.

Best, C. T., McRoberts, G. W., & Sithole, N. M. (1988). The phonological basis of perceptual loss for non-native contrasts: Maintenance of discrimination among Zulu clicks by English-speaking adults and infants. *Journal of Experiment Psychology: Human Perception and Performance, 14,* 345-60.

Best, C. T., & Strange, W. (1992). Effects of phonological and phonetic factors on cross-language perception of approximants. *Journal of Phonetics, 20,* 305-330.

Boersma, P., & David, W. (2005). Praat: doing phonetics by computer (Version 4.3.01) [Computer program]. Retrieved from http://www.praat.org/

Bohn, O.-S., & Flege, J. M. (1997). Perception and production of a new vowel category by adult second language learners. In James, A. and Leather, J. (Eds.), *Second Language Speech: Structure and Process* (pp. 53-73). Berlin: Mouton de Gruyter.

Borden, G., Gerber, A., & Milsark, G. (1983). Production and perception of the /r/-/l/ contrast in Korean adults learning English. *Language Learning 33*(3), 499-526.

Boula de Mreuil, P. G., & Vieru-Dimulescu, B. (2006). The contribution of prosody to the perception of foreign accent, *Phonetica, 63*, 247-267.

Catford, J. C., & Pisoni, D. B. (1970). Auditory vs. articulatory training in exotic sounds. *Modern Language Journal, 44*, 477-481.

Curtin, S., Goad, H., & Pater, J. V. (1998). Phonological transfer and levels of representations: perceptual acquisition of Thai voice and aspiration by English and French speakers. *Second Language Research, 14*(4), 389-405.

Dabrunz, O. (2010). http://www.dabrunz.com/ (to appear).

Dell, G. S. (1986). A Spreading-Activation Theory of Retrieval in Sentence Production. *Psychological Rev*iew, *93*(3), 283–321.

Derwing, T. M., & Munro, M. J. (1997). Accent, Intelligibility, and Comprehensibility: Evidence from four L1s. *Studies in Second Language Acquisition, 20,* 1-16.

Dupoux, E., Pallier, C., Sebastian, N., & Mehlar, J. (1997). A distressed “deafness” in French? *Journal of Memory and Language, 36,* 406-421.

Earl, M. A. (1975). An acoustic phonetic study of Northern Vietnamese tones. Speech Communications Research Laboratory (Santa Barbara).

Elseendoom, B. A. G. (1989). Production and perception of English vowel duration by Dutch speakers of English. *Proceedings of the Tenth International Congress of Phonetic Sciences,* Dordrecht: Fortis (pp. 673-676).

Esling, J. H., & Wong, R. F. (1983). Voice quality settings and the teaching of pronunciation. *TESOL Quarterly, 17*(1), 89-95.

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. 233-272). Baltimore: York Press.

Flege, J. E., Birdsong, D., Bialystok, E., Mack, M., Sung, H., & Tsukada, K. (2006). The degree of foreign accent in English sentence produced by Korean children and adults. *Journal of Phonetics, 34,* 153-175.

Flege, J. E., Bohn, O.-S., & Jang, S. (1997). Effects of experience on non-native speakers’ production and perception of English vowels. *Journal of Phonetics, 25,* 437-470.

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

Flege, J. E., Munro, M. J., & Mackay, I. R. A. (1995). Factors affecting strength of the perceived foreign accent in second language. *Journal of the Acoustical Society of America,* 97, 3125-3134.

Fox, M. M. (1998). Computer mediated input and the acquisition of L2 vowels. The fifth International Conference on Spoken Language Processing. In *Proceedings of the fifth European conference on speech communication and technology,* 5 (pp. 911)

Gass, S. (1984). Development of speech perception and speech production abilities in adults second language learners. *Applied Psycholinguistics, 5,* 51-74.

Gick, B., Wilson, I., Koch, K. & Cook, C. (2004). Language-specific articulatory settings: evidence from inter-utterance rest position. *Phonetica, 61,* 220-233.

Goto, H. (1971). Auditory perception by normal Japanese adults of the sounds “l” and “r”. *Neuropsychologia, 9,* 316-323.

Gottfried, T. L. (1984). Effects of consonant context on the perception of French vowels. *Journal of Phonetics, 12,* 91-114.

Guion, S. G., & Pederson, E. (2007). The role of orienting attention for learning novel phonetic categories. In *Proceedings of the 16th International Congress of Phonetic*

*Sciences*, Saarbrucken (pp. 1705-1708).

Harrison, P. (2000). Acquiring the phonology of lexical tone in infancy. *Lingua, 110*, 581-616.

Hasegawa, Y., & Hata, K. (1988). Delayed pitch fall in Japanese. *Journal of the Acoustical Society of America, 83,* S29.

Hasegawa, Y., & Hata, K. (1992). Fundamental frequency as an acoustic cue to accent perception. *Language and Speech 35*, 87-98.

Hertrich, I., & Ackermann, H (1997). Articulatory control of phonological vowel lengthcontrasts: Kinematic analysis of labial gestures. *Journal of the Acoustical Society of America, 102,*523-536.

Hirata, Y. (1999). *Acquisition of Japanese rhythm and pitch accent by English native speakers.* Unpublished PhD dissertation, The University of Chicago.

Hirata, Y. (2005). Effects of speaking rate on native English speakers' identification of Japanese vowel length. In *Proceedings of International Speech Communication Association Workshop on Plasticity in Speech Perception*, 173 (abstract).

Hirst, D., & Espesser, R. (1993). Automatic modeling of fundamental frequency using a quadratic spline function. Travaux de l'Institut de Phonetique d'Aix-en-Provence, *15* (pp. 75-85).

International Christian University (1996). *Japanese for College Students: Basic Vol. 1-3.* Tokyo: Koudansha International.

Ishihara, T. (2006). *Tonal alignment in Tokyo Japanese.* Unpublished PhD thesis, University of Edinburgh.

Iverson, P., Kuhl, P. K., Akahane-Yamada, R., Diesch, E., Tohkura, Y., Kettermann, A. & Siebert, C. (2003). A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition*, *87*, B47-B57.

Kawano, T. (1998). Nihongo no purosodii no shuutoku ni kansuru kenkyuu (Investigation on the L2 acquisition of Japanese prosody). *Nihongo no Syuutoku kenkyuu* (The acquisition of Japanese as a Second Language), *2*, 106-110.

Kerr, J. (2000). Articulatory setting and voice production: Issues in accent modification. *Prospect*, *15*(2), 4–15.

Kinnaird, S. K., & Zapf, J. (2004). An acoustical analysis of a Japanese speaker's production of English /r/ and /l/. *Indiana University Linguistic Club Working Papers, 4*. Retrieved from https://www.indiana.edu/~iulcwp/pdfs/04-kinnaird.pdf

Komaki, R, Akahane-Yamada, R., & Katagiri, S. (2002). Effect of native language on the production of second-language speech segments. *Acoustical Science and Technology,* *23* (3), 163-165.

Kubozono, H. (2001). *Onseigaku/Oninron [Phonetics and Phonology]*. Tokyo: Kuroshio Press.

Ladd, R. D. (1996). *Intonational phonology.* Cambridge: Cambridge University Press.

Lenneberg, E. (1964). *Biological Foundations of Language*. New York: John Wiley & Sons, Inc.

Levelt, W. J. M. (1989). *Speaking*. The MIT Press, Cambridge: MA.

Levelt, W. J. M., Roelofs, A., & Mayer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences, 22*, 1-75.

Levy, E. S., & Strange, W. (2008). Perception of French vowels by American English adults with and without French language experience. *Journal of Phonetics 36*, 141-157.

Liberman, A. M. (1996). *Speech: A Special Code.* Cambridge, Massachusetts, London, England: The MIT Press.

Llisterri, J. (1995). Relationship between speech production and speech perception in a second language. In *Proceedings of the 13th International Congress of Phonetic Science,* Vol.4, (pp. 92-99). Stockholm: KTH/Stockholm University.

Lowie, W., & Bultena, S. (2007). Articulatory settings and the dynamic settings of second language speech production. In *Proceedings of Phonetic Teaching & Learning Conference, UCL* (pp. 1-4).

Magen, H. (1998). The perception of foreign-accented speech. *Journal of Phonetics,26,* 381-400.

Major, R. C. (2001). *Foreign Accent: The Ontogeny and Phylogeny of Second Language Phonology.* Mahwah, New Jersey, London: Lawrence Erlbaum Associates, Publishers.

Major, R.C., Fitzmaurice, S. M., Bunta, F., & Balasubramanian, C. (2002). The effects of nonnative accents on listening comprehension: Implications for TOEFL. *TESOL Quarterly, 36* (2), 137-190.

Mattock, K., Molnar, M., Polka, L., & Burnham, P. (2008). The developmental course of lexical tone perception in the first year of life. *Cognition*, *106*(3), 1367-1381.

Miyawaki, K., Strange, W., Verbrugge, R., Lieberman, A. M., Jenkins, J. J., & Fujimura, O. (1975). An effect of linguistic experience: The discrimination of [r] and [l] by native speakers of Japanese and English. *Perception and Psychophysics, 18*, 331-340.

Mennen, I. (1998). Second language acquisition of intonation: The case of peak alignment. In M. C. Gruber, D. Higgins, K. Olson & T. Wysocki (Eds.), *Chicago linguistic society, 34* (2): The panels (pp. 327-341). Chicago: University of Chicago.

Mennen, I. (1998). *Second language acquisition of intonation: The case of Dutch near-*

*native speakers of Greek.* Unpublished PhD thesis, University of Edinburgh.

Mennen, I. (2004). Bi-directional interference in the intonation of Dutch speakers of Greek. *Journal of Phonetics, 32*, 543-563.

Mennen, I., Scobbie, J.M., De Leeuw, E., Schaeffler, F. & Schaeffler, S. (2010). Measuring language-specific phonetic settings. *Second Language Research, 26*(1), 191-215.

Mora, J. C. (2008). Methodological issues in assessing L2 perceptual phonological competence. In *Proceedings of Phonetics in Teaching & Learning*, London: UCL. Retrieved from http://www.phon.ucl.ac.uk/ptlc/ptlc2007_web_procindex.html

Munro, M. J. (1998). The effects of speaking rate on listener evaluation of native and foreign-accented speech. *Language Learning, 48,* 159-182.

Neufeld, G. G. (1979). Towards to a theory of language learning ability. *Language Learning, 29*, 227-241.

NHK, Nihon Housou Kyoukai [Japan Broadcasting Corporation] (1998). *NHK Nihongo Hatsukon Akusento Jiten [Japaanese pronunciation accent dictionary]*. Tokyo: NHK Press.

Patkowski, M. (1989). Age and accent in a second language: a reply to James E. Flege. *Applied Linguistics, 11*,73-89.

Pennington, M. C., & Richards, J. C. (1986). Pronunciation revisited. *TESOL Quarterly, 20*, 207-225.

Pisoni, D. B. (1973). Auditory and phonetic memory codes in the discrimination of consonants and vowels. *Perception and psychophysics, 13*, 253-260.

Piske, T., MacKay, I. R. A., & Flege, J. E. (2001). Factors affecting degree of foreign accent in an L2. *Studies in Second Language Acquisition, 17,* 191-215.

Polka, L., & Werker, J. F. (1994). Developmental changes in perception of non-native vowel contrasts. *Journal of Experimental Psychology: Human Perception and Performance, 20*, 421-435.

Remijsen, B. (2004a). Automatic detection of f0 turning points using MOMEL. Retrieved from http://www.ling.ed.ac.uk/~bert/praatscripts.html

Remijsen, B. (2004b). Producing time-normalized f0 tracks. Retrieved from http://www.ling.ed.ac.uk/~bert/praatscripts.html

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: York Press.

Rolland, G. (2000). Automatic stylization of the fundamental frequency F0 using MOMEL: documentation and translation into a PRAAT script. *ICP, November.*

Rose, P. (1987). Considerations in the normalization of the fundamental frequency of linguistic tone. *Speech Communication*, *6,* 343-351.

Sakamoto, E. (2003). *The Relationship between L2 Speech Perception & Speech Production.* Unpublished MSc dissertation, University of Edinburgh.

Schumann, J. H. (1976). Social distance as a factor in second language acquisition. *Language Learning*, *26,* 135-143.

Scovel, T. (1969). Foreign accents, language acquisition, and cerebral dominance. *Language Learning*, *19*, 245-253.

Shehadeh, A. (2003). Learner, output, hypothesis, testing and internalizing linguistic knowledge. *System*, *31*,155-171.

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

Sluijter, A. M. C., & Van Heuven, V. J. (1996a). Spectral balance as an acoustic correlate of linguistic stress. *Journal of the Acoustical Society of America*, *100*(4), 2471-2485.

Sluijter, A. M. C., & Van Heuven, V. J. (1996b). Acoustic correlates of linguistic stress and accent in Dutch and American English. In *Proceedings of the fourth International Congress of Spoken Language Processing,* Philadelphia.

Strange, W. (1995). Phonetics of second-language acquisition: Past, present, future. In *Proceedings of the 13th International Congress of Phonetic Science*, Vol.4, (pp. 76-83). Stockholm: KTH/Stockholm University.

Strange, W. (2007). Cross-language phonetic similarity of vowels: Theoretical and methodological issues. In O-S. Bohn & M. J. Munro (Eds.), *Language Experience in Second Language Speech Learning: in honorof James Emil Flege* (pp. 35-55). Amsterdam: John Benjamins.

Strange, W., & Schafer, V. L. (2008). Speech perception in second language learners: The re-education of selective perception. In J. G.. H. Edwards & M. L. Zampini (Eds.), *Phonology and Second Language Acquisition.* Amsterdam: John Benjamins Publishing Company.

Sugito, M. (1972). *Ososagari-kou: Doutai-sokutei ni yoru nihongo akusento no kenkyuu* [Delayed pitch fall: An acoustic study]. *Shoin Joshi Daigaku Ronshuu* 10. Reprinted in M. Tokkugawa (Ed.). *Akusento* [Accent] (pp. 201-229). Tokyo: Yuuseidou (1980).

Sugito, M. (1982). *Nihongo Akusento no Kenkyuu* [An investigation on Japanese Accent]. Tokyo: Sanseido.

Sugito, M. (1980). *Ososagari kou – doutai sokutei ni yoru nihongo akusento no kenkyuu*

[Peak delay – an investigation on Japanese pitch accent using acoustic measurements]. *Ronsyuu Nihongo Kenkyuu 2, Akusento [Essays on Japanese Language 2, Accent]*. Tokyo: Yuuseidou.

Swain, M., & Lapkin, S. (1995). Problems in the output and cognitive processes they generate: a step towards second language learning. *Applied Linguistics*, *16*(3), 371-391.

Takagi, N., & Mann, V. (1995). The limits of extended naturalistic exposure on perceptual mastery of English /r/ and /l/ by adult Japanese learners of English. *Applied Psycholinguistics*, *16*(4), 379-405.

Thoren, B. (2006). Phonologic demands vs. system constraints in an L2 setting. In *Proceedings of FONETIK 2006* (pp. 115-118). Lund: Lunds Universitet.

Toda, T. (2003). *Second Language Speech Perception and Production: Acquisition of Phonological Contrasts in Japanese*. Lanham, MD: University Press of America.

Turk, A., Nakai, S., & Sugahara, M. (2006). Acoustic segment durations in prosodic research: a practical guide. In S. Sudhoff, D. Lenertova, R. Meyer, S. Pappert, P. Augurzky, I. Mleinek, N. Richter & J. Schliesser (Eds.), *Methods in Empirical Prosody Research*. (pp. 1-27). Berlin: Walter de Gruyter.

Ueyama, M. (1997). The phonology and phonetics of second language intonation: the case of “Japanese-English”. In *Proceedings of the fifth European conference on speech communication and technology*, 5 (pp. 2411-2414).

Van Dommelen, W. A., & Husby, O. (2007). The perception of Norwegian word tones by Chinese and German listeners. In *New Sound: Proceedings of the fifth International Symposium on the Acquisition of Second Language Speech* (pp. 134-174).

Wilson, I. (2006). *Articulatory settings of French and English monolingual and bilingual speakers*. Unpublished PhD thesis, University of British Columbia.

Wilson, I., Horiguchi, N. & Gick, B. (2007). Japanese articulatory setting: the tongue, lips, and jaw. Paper presented at UltraFest IV, New York University.

Werker, J. F., & Tees, R. C. (1984). Phonemic and phonetic factors in adult cross-language speech perception. *Journal of the Acoustical Society of America, 75*, 1866-1878.

Xu, Y. (2008). _TimeNormalizeF0.praat, version 2.6.6. Retrieved from http://www.phon.ucl.ac.uk/home/yi/tools.html

Xu, Y. (2009). Effects of tone and focus on the formation and alignment of f0 contours. *Journal of Phonetics*, *27*,55-105.

Yang, B. (1996). A comparative study on American English and Korean vowels produced by male and female speakers. *Journal of Phonetics, 24*, 245-262.