THE PHONOLOGICAL DEVELOPMENT OF ADULT JAPANESE LEARNERS OF ENGLISH: A LONGITUDINAL STUDY OF PERCEPTION AND PRODUCTION

MAMIKO AKITA

Ph.D. THESIS **2001**

The Phonological Development of Adult Japanese Learners of English: a Longitudinal Study of Perception and Production

by

Mamiko Akita

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

University of Durham

2001

Abstract

Phonology is an area in which first language transfer and incomplete attainment is more apparent than in any other component of language in adult second language acquisition. The present thesis discusses that adult L2 learners seem to be under a maturational constraint in that they *mis-perceive* the target language (TL) in terms of their native language. It is also discussed that socio-psychological factors (e.g. motivation, attitudes) have a profound negative effect on learners, as well as the tendency for learners to receive qualitatively (i.e. foreign-accented) and quantitatively insufficient input compared to first language acquisition (L1A) or child second language acquisition (L2A). Taken together, it is concluded that it seems inevitable that adult foreign language learners fall short in proficiency compared to that of native speakers.

An interesting research question then is what effect exposure to native-accented-input has on such foreign language learners. Will they benefit from this exposure and practice. Or is it too late for them to improve their phonology after being exposed to non-native-accented-input for years in a classroom setting?

A longitudinal study was carried out to examine the effect of primary linguistic data over time on learners who had received a limited amount of input, often foreign-accented. Data were collected from three Japanese learners of English (*NI*, *MO* & *SK*) starting right after they arrived in the U.K. and continued on a monthly basis for over a year.

Data from one perception and two production studies suggest that adult L2 learners who had received years of primarily foreign-accented input can improve their L2 phonology once they are in the target language setting and exposed to native-speaker-accented input.

Results from production study I show that two of the subjects (*NI* & *MO*) started to produce formant frequencies appropriate for a target schwa and one of the subjects (*NI*) seemed to have acquired the targetlessness of schwa in the second formant frequencies. Results from production study II show that in the later stage of the data collection, *NI* and *MO*'s timing of the falling pitch became target-like for English: the falling pitch started from the *first* syllable, rather then the *second* as in their native language Japanese. It was suggested that these seemingly separate processes were connected: the acquisition of vowel reduction was linked to correct timing and the acquisition of the former trigger the latter. The results from the perception and production studies indicate that perception and production is related and it is not the case that they are two separate process which are acquired independently. To conclude, the effectiveness and need for explicit instruction on prosodic features is suggested.

No part of this thesis has previously been submitted for a degree at the University of Durham or any other University.

Acknowledgements

I would like to express my heartfelt gratitude to my supervisor Martha Young-Scholten. Her support, academic and moral, has been invaluable throughout my graduate career at the University of Durham. She gave me encouragement and valuable suggestions in every step I went through.

I am much indebted to Hiroshi Matsusaka, my supervisor at Waseda University, for his unfailing support over the years. I owe to him the interest I first developed in second language acquisition.

I am also grateful to the members of staff at the University of Durham for providing a supportive and fertile academic environment - Randal Holme and Peter Grundy for their general support; Joseph Emonds for being a supportive and friendly department chair; Maggie Tallerman for her moral support and Bonnie Schwartz for her constant encouragement and guidance. Special thanks to S. J. Hannahs for his stimulating lectures in phonology and expertise in acoustic phonetics.

I would like to thank my examiners, Mike Davenport and Martha Pennington for their insightful comments and important suggestions.

I would like to thank my former colleagues at Chiba University, especially Yukio Takefuta and Chikako Nishigaki for their support and guidance, and my current colleagues at the University of Electro-Communications for their support.

I would like to sincerely thank the people who helped me with the data collection and writing up the thesis. I gratefully acknowledge *NI*, *MO* and *SK* for their co-operation throughout the process of data collection. Without them, this thesis would not have been possible. Thanks to Kaori Nishizawa for drawing the picture cards; Megan Williams and Natalie Hanchard for providing native speaker data. I am also grateful to Siobhan Casson and Melinda Whong-Barr for proof-reading my drafts.

I am much indebted to Ana Parrondo and Naomi Cross for stimulating discussions; Belma Haznedar, Miho Katsumata, Yoshie Nakazora and Ariko Komiya for their friendship. I would like to mention that Bruce Scholten fed me the most delicious pizza in the world to keep up my spirits when I was fed up with work. I would also like to thank my Ph.D. colleagues who I shared the study room with, in particular, David Stringer, Karima Soutsane, Suthee Sumdangdej, Yuet Wah Lam, Jeong-Young Kim, Siobhan Casson and Melinda Whong-Barr for their friendship. Especially, my thanks go to Julie King, Seiki Ayano and Kyoko Oga for the fun we had together when we organised the first, now annual postgraduate conference. Special thanks go to Seiki Ayano for his constant support and guidance.

My close friends in Japan deserve special mention. Yasushi Tezuka, Masaki Hirahata, Yasuko Asayama and Mayumi Okutsu have been supportive friends over the years. I wish to thank Tadashi Fukami for his friendship and companionship.

Another special person has been Kyoko Oga. I wish to thank her for her friendship, support and encouragement since we set out on the Ph.D. programme together. She is the best possible friend I can ever have.

Lastly but not the least, my greatest thanks go to my parents, Rumiko Akita and Yasuhiro Akita who have been my most solid support. This thesis is dedicated to them.

Table of contents

List of Figures

List of Tables

Chapter 1

Introduction

An overview of second language acquisition (L2A) research reveals that one area of L2A which has been largely overlooked in the early days of the field, but which has attracted a great deal of attention over the past two decades, is the area of phonology. One reason for is that phonology is an area in which incomplete attainment is more apparent than in other components of the language, e.g. syntax, lexis and pragmatics (see e.g. Ioup 1987). If we compare adult[1] L2A and child L2A, researchers are convinced that younger learners are at an advantage, especially in the area of phonology (see e.g. Asher and Garcia 1969; Major 1987; Oyama 1976; Patkowski 1990; Suter 1976; Thompson 1991). The more difficulty learners experience in acquiring a certain aspect of a language, the more likely it is that their behaviour will reveal, in the form of errors, the mechanisms involved in a second language.

Investigating the issue is equally important for practical reasons. If we can determine the nature of adult second language learning with respect to phonology, we might be able to apply the findings to classroom teaching and find out what the

[1] Adult is defined as post-puberty learners, who are over 15 years of age (Patkowski 1990).

most needed area for the learners is and how we should carry out pronunciation teaching in second language phonology especially in foreign language setting, where incomplete attainment is most often observed.

A good deal of controversy has been generated over the years as to whether complete attainment is possible in adult second language acquisition and about the reasons for its incompleteness. However, no definite conclusions have been reached.

Numerous studies point to age as a factor responsible for the incompleteness and neuro-biological constraints on adult SLA have been suggested. Lenneberg (1967) formulated the critical period hypothesis in an attempt to account for the difficulty of acquiring a first language after 9-12 years of age. He claimed that there is a biologically determined time limit during which acquisition is possible for the mother tongue. Many researchers have examined the applicability of this hypothesis to second language acquisition and it has received some support (Patkowske 1980; Scoval 1981; Lamendella 1977).

Although neuro-biological changes appears to have a large effect on adult L2A, the effect of socio-psychological factors such as motivation, attitude, language ego, language anxiety on adult L2A have also been demonstrated to be important.

Additional factor which may affect adult L2A is the input factor, especially in the foreign language setting. To begin with, the input learners receive in a foreign language setting is often qualitatively inappropriate; in other words, it is typically non-native-accented. Moreover, the input learners receive is quantitatively insufficient.

In recent years, the biological maturation effect on L2 phonology has been re-examined by focusing on the development of perceptual abilities in first language acquisition. *Mis-perception* of the target language (TL) input has been suggested as a possible cause for L2 learners' failure to achieve native proficiency.

Based on the observations that adult L2 learners suffer maturational constraints in the form of *mis-perception* of the target language; socio-psychological factors have a profound negative effect on learners; and, the tendency for learners to receive qualitatively and quantitatively insufficient input, it seems inevitable that foreign language learners fall short in proficiency compared to native speakers.

An interesting research question, then, is what effect exposure to native-accented-input has on such foreign language learners. Will they benefit from this exposure and practice? Or is it too late for them to improve their phonology after being exposed to non-native-accented-input for years?

A longitudinal study was carried out to examine the effect of primary linguistic data (PLD) on learners who had received a limited amount of input, often foreign-accented. Data were collected from three Japanese EFL (English as a foreign language) learners starting right after they arrived in the U.K. and continuing on a monthly basis for over a year. Subjects performed several tasks testing their perception and production.

The following is a brief overview of the structure of this thesis.

In Chapter 2, factors which have been suggested to be responsible for the incomplete attainment of L2 phonology are examined. I first look at studies which point to age as a factor responsible for incomplete attainment in L2 phonology (2.1.2), and then review the literature which suggests there exist neuro-biological constraints on adult L2A (2.1.3). I then look at studies which suggest a profound effect of socio-psychological factors (2.1.4) and also refer to the importance of abundant exposure to native-accented-input for successful L2A (2.1.5).

Further in 2.2, the effect of biological maturation on L2 phonology will be discussed focusing on perception. To begin with, 2.2.2 examines studies which shows that perception and production are related, and that perception precedes production. This gives ground to argue that the difficulty L2 learners experience with L2 speech may possibly have its root in their difficulty in perceiving the L2 sounds. I then discuss the development of perceptual abilities in first language acquisition and review literature which suggests *mis-perception* of target language (TL) input as a possible cause for L2 learners' failure to achieve native proficiency (2.2.3).

I then discuss its implications for pronunciation teaching in 2.3. Both that prosodic characteristics may be one of the most important features in the perception of foreign accent and that perceived foreign accent is in turn closely tied to reduced intelligibility of accented speech for native speakers is discussed (2.3.1).

Additionally, the need for explicit instruction in phonological features of the target language is suggested (2.3.2).

In Chapter 3, phonologies of English and Japanese, the target language (L2, TL) and the native language (L1, NL) of the test subjects respectively are compared. I first look at the segmental features of the two languages (3.1). To begin with, underspecification theory is outlined (3.1.2) and it is shown that English schwa is underspecified, in other words, targetless (3.1.3). We also review studies which discuss variability in vowels in English (3.1.4) and in Japanese (3.1.5). Implications of English targetless schwa for Japanese learners are then discussed (3.1.6). In 3.2, suprasegmental features of both languages are compared. I first look at syllable structure (3.2.1) and then at the rhythmic style of both languages (3.2.2), followed by a discussion of pitch-accent in English and Japanese. The implications of English suprasegmental features on Japanese learners are also outlined (3.2.3).

In Chapter 4, a longitudinal study on three adult Japanese learners of English is reported. I first outline the methodology of the study (4.2) and report on one perception study (4.3) and two production studies (4.3): one examining the learners' acquisition of English vowel reduction and the other examining the learners' acquisition of timing of the falling pitch. Discussion of the results of the perception and production studies is provided in 4.5 followed by a summary of the findings in Chapter 5.

Chapter 2

The second language acquisition of phonology

2.1 (In)complete attainment in L2 phonology

2.1.1 Introduction

A good deal of controversy has been generated over the years as to whether complete attainment is possible in adult second language acquisition and about the reasons for its incompleteness. However, the conclusions are far from definite. In this sub-section, factors which are suggested to be responsible for the incompleteness are examined. We first look at research suggesting neurological constraints on adult second language acquisition. This is followed by the arguments of those who contest the biological constraint theories and suggest that socio-psychological factors have a profound influence on the observed age-effect.

2.1.2. Age and second language acquisition

Whether or not L2 learners are capable of complete attainment, it is undoubtedly true that the majority of post-puberty learners do not reach native competence in their L2 phonology. Numerous studies point to age as a factor responsible for fossilisation. These show that later age of arrival in the TL setting makes a difference.

Asher & Garcia (1969) found that 71 Cubans who arrived in the U.S between the ages of one and six were judged closest to native-like on a sentence–repetition task. The next closest were those who arrived between the ages of seven and twelve. Those who arrive later than thirteen and later than nineteen performed the poorest.

Along the lines of Asher & Garcia's study, Oyama showed in his 1976 study that children under the age of twelve had a better chance of acquiring native competence. He tested 60 Italian born male immigrants taken from the upper educational group. Subjects were distributed along two independent variables: age at arrival in the Unites States (6-20 years) and number of years in the United States (5-18 years). Pronunciation was scored from two taped speech samples. The first was a reading of a short paragraph which was constructed to include a number of phonological variables. The second sample was a brief anecdote where subjects were asked to describe the most frightening episode in their life. For the second sample, the last 45 seconds of each recording was analysed. The speech samples were judged for degree of accent by two American-born graduate students in Linguistics on a 5-point scale ranging from no foreign accent to heavy foreign accent. Analyses of variance and correctional methods were used to examine the relationships between the two main independent variables, the questionnaire variables and the accent measures. It was found that age of arrival was a strong indicator for degree of accent, while length of study in the U.S. was not. Here it should be noted that although the arrivals before the age of twelve performed in the range of native speakers, accents were evident in some who arrived earlier than twelve.

Patkowski (1980) obtained global syntactic proficiency ratings consistent with the notion of an age limitation on the acquisition of syntax in a second language. He obtained a clear bimodal distribution among pre- and post-puberty L2 learners by examining five-minutes of transcribed speech of 67 immigrants to the U.S. Similarly, Johnson & Newport (1989) investigated whether there was a relation between age of acquisition and ultimate performance in the syntax of a second language. They found that children have an advantage over adults in acquiring a second language and age of arrival and performance were significantly correlated. Due to the large range of ages in the subjects, they were able to find a good generalisation about the shape of the relationship. They found that subjects who arrived in the U.S. before the age of seven reached native performance on the test. For arrivals after the age, there was a linear decline in performance up through puberty. Post-puberty learners performed on average much more poorly than the younger group. However, performance did not continue to decline with the increasing of age. Rather, the learners' average production performance was marked with great variability.

2.1.3. Neurological factors: critical/sensitive period

Lenneberg (1967) formulated the Critical Period Hypothesis in an attempt to account for the difficulty of acquiring a first language after 9-12 years of age. He claimed that there is a biologically determined time limit during which language acquisition of the mother tongue is possible. He set this at around puberty to coincide with the completion of the specialisation of brain function, i.e. lateralization. Prior to puberty, a critical period exists during which the brain is more plastic; and at puberty, cortical lateralization is completed during which language functions become localised in the left cerebral cortex and the former plasticity of the brain atrophies.

Many researchers have examined the applicability of this hypothesis to L2A. The term 'sensitive period' has been proposed (Patkowski 1980, Scovel 1981, Lamendella 1977) which implies that the critical period is not an abrupt criteria, but a gradual process during which ultimate level of L2 attainment becomes more difficult. Scoval argued that it appears as if language learning ability slowly declines as the human matures, and plateaus at a low level after puberty; however, the precise level of this plateau differs between individuals. Scovel further argued that the loss of plasticity has a particularly significant effect on the area of phonology, because the loss of flexibility of neuro-physiological programming of neuromuscular coordination mechanisms adversely affects an individual's ability to control the articulators necessary in pronunciation of a second language (1980:37). Here, it should be noted that this argument was based on the incompleteness in the domain of speech production: the foreign accentedness. How then can we account for the non-nativeness in perception? I discuss in detail in 2.2 that L1-based perception is one of the causes of non-native perception and production.

Although neuro-physiological changes appears to have a large effect on adult second language acquisition, it has been demonstrated that this may not be truly the *causal* factor for incomplete attainment in second language acquisition. Krashen (1973) presented evidence that the development of lateralization is complete much earlier than puberty, at around age five. Since children up to the age of twelve seem to acquire native-like competence in the second language, completion of brain lateralization can not mean the establishment of an absolute barrier to language acquisition.

However, if we concentrate on phonological acquisition, this brain lateralization does coincide with the age which divides the native competence achievers and non-achievers. Oyama's study (1976) offers very interesting data. As stated earlier, Oyama's learners who arrived before the age of twelve performed in the range of native speakers, but accents were evident in some who arrived earlier than twelve. It is possible to interpret this data as evidence supporting that the multiple critical period (see below) for phonology is age five or six. On the other hand, it is rather misleading to do so, because it is undoubtedly true that the majority of the learners did acquire native competence after the age of five or six.

Seliger (1978) and Long (1990) suggest that multiple critical periods exist for the various sub-components of language. According to Long, at the age of six the critical period ends for phonology, with successively later ages for morphology and syntax. However, if lateralization is completed by the age of five, why do the majority of children after the age of five (but not before puberty) achieve native pronunciation in a second language (maybe at least in a natural setting)?

Tahta, Wood & Lowenthal (1981) used their experimental study to propose the existence of multiple critical periods within phonology. The study examined predictors of transfer of accent from the L1 to L2 in a group of people whose acquisition of English as an L2 has begun at ages ranging from 6 to 15. The subjects in their study had all lived in the U.K. for a minimum of two years. There were 54 males and 55 females. Among them, 10 had started to acquire English at the age of 6, 10 at the age of 7, and 10 at each age through to 14. 19 had begun acquiring English at or after the age of 15. The subjects' native languages were Arabic, Cantonese Chinese, Dutch, French, German, Greek, Hindi, Japanese, Persian and Polish, among other languages. They were asked to read a paragraph of English prose and were then interviewed about their history of L2 learning. Each speech sample was listened to by three independent native judges who assigned a score of either 0 (no foreign accent), 1 (detectable but slight accent) or 2 (marked accent).

A multiple regression analysis was carried out with accent as the dependent variable and it was found that age of arrival was the strongest predictor of foreign accent. They found that if the age of arrival was after 12-13, there is invariably accent transfer, usually very heavy. In contrast, the accent transfer of to those who arrived in the country between the age of 7-11 was very slight. Thata et al. (1981) also observed that up until the age of 11, accents when present were usually very slight, and much of the transfer from L1 seemed to be of intonation. They concluded that age-related constraints begin as early as six for suprasegmental phonology, and soon after that for segmental phonology. They backed up their argument with Thata (1980)'s study in which she studied British schoolchildren's ability to replicate foreign intonation and pronunciation. She observed that the ability to replicate foreign intonation dropped rapidly from the ages of 8 to 11, while the ability to replicate segmental pronunciation declined more gradually over the whole age-range studied (5-15).

Given the difficulty in determining when a learner has reached native speaker competence, and the conflicting data, the notion of a sensitive period for second language acquisition remains in doubt. Some researchers suggest that the attainment or failure of attainment of native-like L2 performance is not exactly a proof of the continued operation of a language acquisition device operating in first language acquisition. Several researchers have proposed that attainment of Piaget's stages of formal operations have a negative effect on adult second language acquisition (Rosansky 1975; Felix 1981). They argue that child and adult second language acquisition involve different processes: children utilise a LAD (Language Acquisition Device) as in L1 acquisition, while adults employ general problem-solving abilities. However, the existence of Piaget's stages is still controversial. Furthermore, if we assume that children and adults learn an L2 differently, we would expect to see more evidence of different acquisition processes and sequences, but there is little evidence of this to date (e.g. Newport 1984; Clahsen & Muysken 1986) and there is some evidence that they are acquiring it in the same way (e.g. Bley-Vroman, Felix & Ioup 1988). Some argue that native-like behaviour in some areas of the language might be the result of different processes. For example, James (1996) speculates that poor performance by adults as opposed to children in the area of phonology might be ascribed to different language processing being involved and not directly due to brain lateralization. Successful L2 attainment in phonology such as the learner in Ioup et al's study (1994) may be the result of different language processing: older learners tend to process a signal linguistically in terms of their L1 auditory/acoustic patterns, whereas younger learners are more capable of auditory processing of a signal without reference to already developed L1 linguistic patterns (for empirical evidence see Wieden & Nemser 1991).

2.1.4 Social-psychological factors

There are several indications in the literature that language learning difficulties after puberty may be related to social and/or psychological (*motivation*, *attitude*, *language ego* and *language anxiety*) changes an individual undergoes at that age (Schumann 1975). It should be noted that researchers such as Guiora, Beit-Hallahmi, Brannon, Dull & Scoval (1972) view pronunciation ability as being most vulnerable to suggested socio-psychological negative influences.

2.1.4.1 Motivation & attitude

Some researchers have attributed the failure of adults in acquiring a second language to social and/or psychological (e.g. *motivation*, *attitude*, *language ego*, *language anxiety*) changes an individual undergoes at that age (Schumann 1975).

Gardner and his colleagues (Gardner 1980; Gardner, Clement, Smythe & Smythe

1979) claim that when a learner is motivated by utilitarian purposes, such as furthering a career or improving social status, or has an integrative motivation to become like speakers of English (i.e. a desire to blend in to the community and make friends), this leads to successful second language learning. The Attitude Motivation Index (AMI) was developed by Gardner, Clement, Smythe & Smythe (1979). The index was formed by combining scores on the following 11 measures: attitudes toward French Canadians; attitudes toward European French people; interest in foreign languages; integrative orientation; motivational intensity; desire to learn French; attitudes toward learning French; French teaching-evaluative; French course-evaluative; instrumental orientation; French class anxiety. Gardner (1980) examined 29 samples of students from various grades and regions across Canada and inspected the relation between affective variables and grades in French as a second language. He argued that the evidence indicates *motivation* is positively correlated with second language acquisition.

Furthermore, the effect of parents and peers is suggested to have effect on second language learners. Labov (1972) found that middle-class New Jersey children of New York City-born parents did not drop postvocalic /r/ as did their parents, but rather pronounced /r/ as did their peer group.

However, correlations between age, attitudinal and motivation factors are weaker than expected, especially if we look at the negative, conflicting or less easily interpreted results from second language contexts (Oller, Hudson & Liu 1977; Oller & Perkins 1978; Lukmani

1972) and foreign language contexts (Chihara & Oller 1978). For example, Oller et al. (1977) investigated the relation between various measures of attitude toward self, the native language group, the target language group, reasons for learning English as a second language (ESL), reasons for travelling to the U.S. and attained proficiency in ESL. A cloze test was constructed by deleting every seventh word from a passage of prose. The subjects were 16 female and 28 male native speakers of Chinese who had been in the U.S. for an average of 3.6 years and who were students in U.S. universities at the time of the testing. A questionnaire was designed to collect descriptive information concerning variables of sex, age, years of ESL study, whether or not the subject attended college, the degree to which English texts were used if the subject went to college and how well subjects' parents spoke English. Identity scales were constructed and the subjects were asked (1) to rate themselves on the thirty identity scales; (2) to indicate how they 'would like to be' on the same scales; (3) to rate their fellow Chinese people; and (4) to rate Americans. The subjects were also asked questions to evaluate their integrative and instrumental orientation. These fell into three categories: (a) reasons for coming to the U.S. (b) attitudes toward the U.S. (c) reasons for learning English. Questions included: whether they thought they could find a better job in the U.S. or at home; whether they would consider staying in the U.S. permanently; and how they felt about the warming of relations between the U.S. and mainland China. It turned out that several of the variables which had been expected to correlate positively with the cloze procedure did not. For example, the number of years spent studying English, the use of English texts in college, the parents' skill in English and the number of years in the U.S. did not correlate positively with the cloze score. On the whole, they found that attitude toward self, the native language group, and toward the target language group were positively correlated with attained proficiency in ESL. However, the relation between reasons for studying ESL or travelling to the U.S. and attained proficiency was contrary to previous findings. It was found, for example, that there was a significant negative correlation between desire to stay in the U.S. permanently and attained ESL proficiency in English.

One view to account for these conflicting results is to take note of socio-cultural differences between categories of studies where those factors could drastically influence the nature of the role played by affective variables in second (foreign) language acquisition. Au (1988) evaluated various researches examining the role of affective factors on second language acquisition. She attempted to give an evaluative account of Gardner's socio-psychological theory. Gardner's theory was broken down into five propositions: (1) the integrative motive hypothesis was found to lack generality and also challenged by Au as a non-unitary concept; (2) the cultural belief hypothesis, a hypothesis which maintains that cultural beliefs can influence the development of the integrative motive, was concluded to be poorly defined as regards what constitutes a cultural belief and was therefore a much untested notion; (3) the active learner hypothesis maintains that integrative motivated learners are better learners. Au argued that studies relevant to this hypothesis suffered from serious methodological flaws in that the level of L2 proficiency had never been controlled; (4) the causality hypothesis maintains that integrative motive causally affects L2 achievement, the former the cause, the latter the effect. Au argued that relevant studies were incapable of generating concrete empirical evidence concerning a casual relationship; (5) the two-process hypothesis maintains that linguistic aptitude and integrative motive are two independent factors affecting L2 achievement. Au notes that conflicting evidence exists in relation to this hypothesis.

Concerning Gardner's causality hypothesis, Burstall, Jamieson, Cohen & Hargreaves (1974) found a reverse relationship between the *cause* and *effect* of motivation/attitude and attained proficiency from Gardner's study. They reported a longitudinal study spanning a period of ten years conducted in the U.K. They found that early achievement in French affected later attitudes toward learning French and later achievement in French to a significantly greater extent than early attitudes toward learning French affected the subsequent development of either attitudes or achievement. This result suggests motivation is causally related to L2 achievement, the former the cause, the latter the effect. Such studies are much needed.

In conclusion, it seems certain that *attitude* and *motivation* does play role in second language acquisition to some extent (even if not crucial) and that negative influences are likely to find fertile soil in second language acquisition.

2.1.4.2 Language ego & language anxiety

Yet another psychologically related factor suggested is *language ego* as well as *language anxiety* (Guiora 1972; Horwitz, Horwitz & Cope 1986; MacIntyre & Gardner

1989; 1994). Guiora (1972) proposed that adult differ from children in that young children have a relatively flexible language ego, i.e. they are not inhibited nor do they have a strong identity as a speaker of a certain L1. According to Guiora, language ego is conceived as a maturational concept in that language ego directly parallels that of general ego development. He hypothesised that to learn a second language is to take on a new identity, and since pronunciation appears to be the aspect of language behaviour most resistant to change, it is therefore the most critical to self-representation. Guiora, Beit-Hallahmi, Robert, Brannon, Dull & Scovel's study (1972) deals with the experimental manipulation of what they have referred to as the flexibility of psychic processes or the permeability of ego boundaries. 87 university students were recruited and were asked to drink a cocktail presented by the experimenter. The 'cocktail' ingested by each subject contained either zero, one, one and a half, two or three ounces of ninety proof liquor. After leaving 10 minutes for the drink to take effect, subjects were given two language tests. The first test was the Standard Thai Procedure developed by the authors to test ability to approximate pronunciation in an unfamiliar foreign language. The second test was the Digit Symbol Test, which is a subset of the Wecheler Adult Intelligence Scale, a well-known measure of overall mental functioning. The Standard Thai Procedure (STP) consists of ten one-syllable contrastive items and six expansion drill series. Among the ten one-syllable contrastive items: five test for five Thai tones using the same syllable; two test for contrast of aspiration and non-aspiration; two test for contrast of vowel length; one tests for an initial velar nasal sound. As for the expansion drill series, it is composed of four items beginning with a single syllable to which a second, third and fourth syllable are added successively (number of items: 4x6). For instance, an English version of the test would be: since/since swifts/since swifts surpass/since swifts surpass swans. Six such expansion drill series generated twenty-four of the total 34 test items. The STP was an aural-oral test. Each of the thirty four test items were recorded on a master tape and the testees were required to imitate the preceding utterance. Each of the speech samples were rated independently and simultaneously by a team of three experts in the Thai language and were given a rating on a scale of one to three: (1) poor - strongly marked with foreign accent; (2) fair - passable and not likely to cause any misunderstanding although not completely native-like; (3) good - quite native-like.

Mean performance scores on each alcohol treatment indicate a highly significant effect due to the amount of alcohol ingested. Considering components of this curve separately, the one and one and a half ounce conditions were significantly higher than the control condition of no alcohol. Since the STP showed a rise and then a drop in performance, despite the fact that the Digit Symbol Test indicate the subjects were literally unaffected by small quantities of alcohol, Guiora et al. were led to argue that pronunciation ability was apparently critically influenced by a psychological variable that can be successfully isolated from the larger web of ego functions. They concluded that the lowering of inhibitions via alcohol operates to induce a state of greater permeability of ego boundaries, or the ability to partially and temporarily give up one's separateness of identity, which was evidenced in an enhancement of certain linguistic skills.

MacIntyre & Gardner (1994) define language anxiety as the feeling of tension and apprehension specifically associated with second language contexts. The subjects in their study were taking first-year credit courses in French-as-a-second-language at a large, monolingual (English) Canadian University. A total of 97 students participated in the study. The study examined some specific cognitive processes that may be involved in language acquisition in terms of a three-stage model of learning: (1) an Input stage, illustrated by the learner's first experiences with a given stimulus at a given time; (2) a Processing stage, involving cognitive operations performed on the subject matter (organisation, storage and assimilation of the material); and (3) an Output stage, involving the production of previously learned material. Their achievement in French was measured based on the following 9 sets of measurements: (1) course gradés (to provide scores for all three stages); (2) Word Span; (3) Digit Span; (4) T-Scope (to measure the input stage); (5) a Multiple Choice French Achievement Test; (6) Paragraph Translation (to obtain scores for the processing stage); (7) Thing Category Test, where participants were asked to write down as many elements of a category as they could think of, e.g. adjectives that describe people; (8) Cloze Test; (9) Self-Description Task. Language Anxiety Measures were developed based on the three stages. Each stage had a six-item scale including three positively worded items (e.g. I never feel tense when I have to speak in French [the example is taken from the Output Anxiety Scale]) and three negatively worded (e.g. I get upset when French is spoken too quickly [the example was taken from the Input Anxiety Scale]). The students were tested in two stages. First, they completed a questionnaire containing the measurement of language anxiety and with the French Achievement Test. The subjects were then tested individually with the rest of the tests/tasks. It was found that there was significant negative correlation between grades and scores on each of the Input, Processing and Output anxiety scales. In general, it was found that each of the six-item scales representing the three stages showed similar patterns of correlation with each of the specific performance measure, although some exceptions were observed, e.g. the Digit Span test and the T-Scope showed a more complex pattern of anxiety effects for the Input stage. Taking these results together, MacIntyre & Gardner (1994), concluded that the effects of language anxiety on L2 learners may be both pervasive and subtle.

2.1.4.3 Social factors

Various research has explored social distance factors that either promote or inhibit social solidarity between two groups (TL group and second language learning group) and thus affect the degree to which a second language learning group (2LL group) acquires the language of a particular TL. Shumann (1976) argued that social distance and hence a poor language learning condition exist where (1) the 2LL group is either dominant or subordinate; (2) where both groups desire preservation and high closure for the 2LL group; (3) where the 2LL group is both cohesive and large; (4) where the two cultures are not congruent; (5) where the two groups hold negative attitudes toward each other and; (6) where the 2LL group intends to remain in the target language area only for a short time.

Since the social factor does not seem to fit the foreign language situation which I am focusing on, I will not go into details (see, for example, Dennis & Scott 1975; Shermerhorn 1970; Paulston 1975).

Although the factors discussed above seem to play an important role in L2A, can they account for the observed age-related differences as well? As we have come to expect, we can raise some doubts for each of these possible explanations. On what grounds can we say that children are not subject to some of the same inhibitions as adults? Hatch (1978) and Kubota (1987) reported that some Japanese children in the U.S. acquiring English as a second language had psychological difficulties blending into the TL community. How can one say that adults on the whole have a negative attitude toward the TL speakers? If adults do have that, don't they influence the children as parents, and children will end up having negative attitudes as well? On what grounds can we say that adults have lower motivation to learn the language than do children?

2.1.4.4 Concern for pronunciation accuracy

One last factor I will discuss is a factor specific to phonology: L2 learners' concern for pronunciation accuracy. In a study on intermediate learners of Spanish, Elliott (1995) examined a range of exposure variables and found that among several attitudinal variables concern for pronunciation accuracy was most significant for developing accuracy. Along this line, Purcell & Suter (1980) reanalysed Suter's (1976) study. Suter (1980) studied 61 non-native speakers of English whose native language was Arabic, Japanese, Persian or Thai. Subjects were asked to describe a holiday or celebration which exists in their native country and the speech samples were rated by 14 native speaker judges. It was found that predictors accounting for the variability of subjects' pronunciation accuracy scores were: 1) first language; 2) the time learners had resided with a native speaker of English; 3) whether they were concerned about the accuracy of their pronunciation of English among 20 tested variables. The predictor variables tested were: (1) age at which the speaker first resided in an English-speaking country; (2) age at which the speaker was first able to converse meaningfully English; (3) number of years the speaker had lived in English-speaking countries; (4) amount of conversation at home which was carried on in English with native speakers of English; (5) amount of conversation at work and at school which was carried on in English with native speakers of English; (6) residence with native speakers of English; (7) total amount of formal classroom training in English; (8) amount of intensive formal classroom training in English; (9) amount of formal classroom training dedicated to specifically the pronunciation of English; (10) amount of formal training in English carried out under teachers who were themselves native speakers of English; (11) speaker's native language; (12) number of languages the speaker is able to converse in; (13) sex; (14) economic motivation; (15) social prestige motivation; (16) integrative orientation; (17) cultural allegiance; (18) strength of the speaker's concern about pronunciation of English; (19) innate aptitude for oral mimicry; (20) extroversion-introversion.

Whether or not they had much or little formal instruction in English or English pronunciation was found to be of no importance. Therefore, a theoretically superior pronouncer will be a person who resides in the TL country living with a native speaker of the TL and with a high concern for her/his pronunciation.

A recent study by Bongaerts, Summeren, Planken & Schils (1997) shows that a high motivation to speak the second language without an accent was correlated with ultimate attainment in L2 phonology. They had three groups of subjects: (1) 10 native speakers of standard English; (2) 11 native speakers of Dutch who were highly successful learners of English; (3) 20 native speakers of Dutch who had widely different levels of proficiency in English. All subjects read aloud six sentences (e.g. "Arthur will finish his thesis within three weeks") a total of three times. The test sentences were picked because they contained phones ranging from very similar to very different from Dutch phones. The speech samples were judged by 13 native speakers of British English who spoke standard British English without a regional accent. The judges rated all speech samples for accent on a 5-point scale. The results show that the highly successful learners were given extremely high scores, but it was found that the difference between the native speaker group and the group of highly successful learners was statistically significant. However, it should be noted that there were 5 subjects in group 2 who met the criterion as natives. They reported that those subjects were highly motivated and had plenty of exposure to native-accented input, and input enhancement through instruction seemed to play a important role for their learners' success.

2.1.4.5 Interplay of factors

Although various explanations concerning socio-psychological factors and age-effects in second language acquisition have been proposed, no one single factor seems to account for all the acquisition patterns (success or failure) observed. It might be the case that no one single factor is responsible for failure, but in fact the interplay of factors snow-balls the negative outcome of adult second language acquisition, the characteristics of which are concomitant with social-psychological maturation.

In summary, while various explanations of the age relationship have been proposed, the idea that there exist neuro-biological constraints on SLA seems most tenable at present. To be more precise, Patkowski's (1990) explanation, which acknowledges *both* the biological and socio-cultural factors (also referring to the importance of abundant exposure to TL input), seems most convincing after 20 years of extensive research and discussion. He argues that native-like proficiency in all aspects of language is impossible to attain for adult learners; however, it does not hold that extremely high, quasi-native levels cannot be attained in one or more areas, such as the Polish-born English writer Joseph Conrad who attained native proficiency in his writing but still had a thick accent. Reasons for incomplete attainment do not deny a possibly large effect of socio-cultural, attitudinal and motivational factors on second language acquisition, but nevertheless, hold that such factors operate within the constraints of a genetically-based sensitive period. Patkowski insists that what is referred to is the eventual level of proficiency attained after a sufficient period of exposure to, and immersion in, the target language under optimal sociolinguistic and affective conditions.

2.1.5 The Input factor

Here, I will discuss one other factor, a factor Patkowski (1980) also mentions, namely, the input factor. I argue that the input factor has a profound effect on second language acquisition, especially in a foreign language setting. I will first suggest that the input learners receive is often qualitatively inappropriate, in other words, it is typically non-native-accented. Equally importantly, the input learners receive is quantitatively insufficient. In classrooms, learners usually get only several hours of instruction per week, and the linguistic environment outside of the classroom lacks sources of native TL input. As will be discussed later, it has been argued that linguistic exposure outside of the classroom is essential for learners to improve their L2 phonology. If we construct the psychological profile of a typical L2 learner learning the target language in a foreign language setting, it is most likely that, because of their negative state of mind toward acquiring L2 phonology, they will not push themselves to try to be exposed to native-speaker-accented input.

2.1.5.1 'Inappropriate' and 'insufficient' input

Learners often receive inappropriate input, i.e. non-native-accented (Flege 1991; Young-Scholten 1995): Young-Scholten refers to this input as "positive evidence with a *negative* effect" (1995:110). Deviant input is typically from the following two sources: (1) aural input from teachers; (2) aural input from classmates. A non-native teacher with first language (L1) accented pronunciation transmits his/her foreign accent to the pupils. Learners' classmates are another source of 'distorted input.' L2 learners are inevitably exposed to their peers' pronunciation, such as when their classmates are told to read the textbook aloud. In communicative classrooms, the amount of peer input learners receive increases. This issue is of some importance as deviant peer input has been reported from immersion programmes as a possible cause of persistent output errors even when teachers do provide appropriate TL input (Plann 1977). Moreover, pre-puberty children learning a second language sequentially (i.e. not simultaneously with their L1) and in the absence of fellow native-speaker peers have shown typical interlanguage behaviour rather than child L1. This seems to imply it is essential to be exposed to input from native speakers of each language and not to other young learners (such as in immersion education).

Not only is the input learners receive qualitatively inappropriate but learners are also typically exposed to much less phonological input than the child learning his/her L1 or the child learning his/her L2 in a natural environment. Typically, L2 learners are in an environment in which: (1) there are only several hours of instruction per week in the foreign language, (2) few authentic materials exist on TV programmes or in films (i.e. programmes are predominantly dubbed, rather than being subtitled) (3) there are few people using English in daily conversation (4) there is a lack of available native speakers.

2.1.5.2 Amount of input affects L2A

Hamayan, Genesee & Tucker (1977) reported very interesting findings showing that the amount of exposure and practice does have a statistically significant effect on second language acquisition. They examined language exposure factors associated with subjects learning French as a second language who were enrolled in different second language programmes. They looked at three groups of Grade Seven students: (1) early immersion (EI) group; 2) late immersion (LI) group; and 3) non-immersion, conventional second language learner group (English control group, EC). Four tests were used to obtain measures of the students' competence in French, including a oral proficiency test, along with a questionnaire to assess a number of predictor variables such as use of French with strangers, use of English with acquaintances, conformity, shyness and other affective factors. Overall, the EI group performed better than students in the LI group, and both immersion groups performed better than the EC group. For all students, a high frequency of use of English with acquaintances was negatively related to French oral proficiency while a high frequency of use of French with strangers was positively related to oral proficiency. The effect of these two language exposure factors was most pronounced for the EC group. They speculate that since the EC student's programme affords them so little opportunity to practice French in school, any additional use of the language was beneficial for individuals, while immersion students already have considerable opportunity to practice French in school, so extracurricular use of French was less influential.

The implications of Hamayan et al.'s study for my study are: (1) the amount of exposure and interaction does have a significant effect on second language acquisition; (2) when learners have little exposure to the TL in the classroom, as in a foreign language context, they may benefit more from moving into a TL setting and thus be able to improve their L2 phonology.

2.1.5.3 Informal exposure is important

Moreover, Suter (1976) found that informal exposure is more important than formal classroom training in the development of certain second language skills. He found that Japanese learners of English scored low among learners of various mother tongues in his study. 61 non-native speakers of English were measured on 20 variables suspected of displaying significant relationships to pronunciation accuracy. The production of the subjects were then rated by 14 native English speakers. The variables which proved to be most strongly related to pronunciation accuracy were: 1) native language 2) strength of the speaker's concern about his/her pronunciation 3) amount of conversation carried on at work and school with native speakers of English. A negligible factor was amount of formal classroom training (not necessarily pronunciation training) the learners received in their home country.

It was found that "Far East" languages (Japanese & Thai) were almost always rated low by the judges, while speakers from the "Middle East" (Arabic & Persian) were nearly always given high or middle range ratings. Suter argued that conceivably some cultural or personality trait shared by Middle Eastern peoples facilitated the development of oral second language skills, at least in English. Or, perhaps there was some trait common to Far Eastern people which retarded the development of such skills: maybe Arabs and Persians were better because they speak an Indo-European language other than English (i.e. French) and the Persian language itself is Indo-European. Also, along highly speculative lines, he argued that perhaps the Persians and Arabs had fewer problems pronouncing English in the U.S. because they were less distracted by cultural shock. Arabs had undoubtedly experienced "Western" culture as part of their experience with the French language, and is it possible that the Persians, being Indo-European, were more "Western" to start with? I am skeptical about these lines of reasoning, but might one say that "speaking" English (or any foreign language) was not an unfamiliar thing in those countries? One must wonder also whether Japanese and Thai schools downplay oral skills when they teach English and concentrate on reading and vocabulary skills. The findings in this study give some support to the idea that informal exposure is more important than formal classroom training in the development of certain second language skills that Japanese learners totally lack. A good deal of further research is needed in order to more fully understand the dimensions of these factors.

It should also be added that not only input but also *output* is very important for successful second language acquisition. Swain (1985) looked at children whose first language is English and were learning French in an immersion programme. She examined input-output relationships at the level of their language proficiency traits, especially the traits of grammatical, discourse, and sociolinguistic competence. She suggested the reason why those students fell short of native speaker competence was because output was generally missing in the typical classroom setting. The claim is that immersion students do not demonstrate native-speaker productive competence, not because their comprehensible input is limited, but because their comprehensible *output* is limited. Can this be applied to L2 phonology as well? What if we enhance the chances of output of the learners, will it be any good for them?

2.1.5.4 Summary

To sum up so far, I have tried to make the point that the input foreign language learners receive is often qualitatively inappropriate and quantitatively insufficient. In particular, they lack input and practice outside the classroom, which has been shown to be essential for learners to improve their L2 phonology. However, their misfortunes as language learners do not stop here. If we look at the psychological profile of a typical foreign language learner, it is most likely they have a negative state of mind for language learning. As has been examined earlier, various socio-psychological factors have been suggested to affect acquisition. I would like to point out that the negative influence of those factors impede the learners' exposure to the target language input. Being in a linguistic environment where access to native-accented-input is difficult, with such a negative state of mind, we have little hope that learners will be able to get themselves exposed to sufficient and appropriate input to promote acquisition.

Adults may avoid socialising with native speakers due to high inhibition and low empathy; this affects acquisition because learners receive far less linguistic input than children do, or adults in a second language context do who are motivated to mingle with the natives (in fact, it is very difficult to find native speakers to talk to in some foreign language contexts such as Japan). TV and film do not help since programmes are predominantly dubbed, rather than being subtitled, and do not provide native-accented-input. Therefore, it is very likely that learners will be exposed to very little input outside of the classroom.

It should be pointed out that with respect to motivation towards language learning, someone studying/learning a target language as a foreign language would be less likely to have a strong urge to learn the language for integrative reasons such as a desire to blend into the TL community, than someone who has been immersed in the target language community in a second language context.

I believe that the influence of peers' attitudes should not be taken lightly. In foreign language classrooms, this effect can be vital. In a foreign language context, language learning takes place with peers being present the majority of the time: learners may start to prefer to speak accented L2 speech to identify them as a speaker of a particular L1, or to be more precise, as a member of their L1 accented peers. Recall the data given by Elias-Olivares (1976) and Shuy, Wolfram & Riley (1967), who found a negative influence of peers on second language learners.

In conclusion, having a negative attitude towards acquiring the TL variety of pronunciation due to peer pressure, lacking the motivation to learn the language (having little utilitarian or integrative motivation), and suffering from high inhibition and low empathy is certainly not the best state of mind for a learner to acquire L2 pronunciation. Naturally, there is little chance that learners will be exposed to input outside of the classroom or make the most of the input they get in the classroom (note however, classroom input is mostly distorted and will do no good for the learners anyway). If that is the case, is it all together impossible for learners to improve L2 pronunciation in a foreign language setting?

Given the foregoing discussion, Japanese speakers learning English as a foreign language, who perfectly fit the linguistic profile of a typical foreign language learner which has just been provided, are unlikely they will achieve high levels of pronunciation. However, we can look at what happens when they stay in a TL setting for a period of time, where they can get plenty of native input and output practice. The research question is, will they benefit from this exposure and practice? Or is it too late for them to improve their phonology after being exposed to L1-accented-input for years? Lightbown (1983) made an observation that "once fossilisation occurs, continued exposure is quite ineffective in changing language behaviour'. Are these learners in fact fossilised and no longer able to improve their phonology?

2.2 Perception in L2 phonology

2.2.1 Introduction

In the previous section, the effects of socio-cultural, attitudinal and motivational factors in second language phonology have been discussed. I also referred to the importance of abundant TL input for successful acquisition and the possible negative effect of deviant input (i.e. native-language-accented) in a foreign language setting. I also discussed that such factors operate within the constraints of a genetically-based sensitive period.

In this section, I will further examine how biological maturation might affect interlanguage phonology by focusing on the development of perceptual abilities in first language acquisition. *Mis-perception* of target language input is suggested as a possible cause for L2 learners' failure to achieve native proficiency.

To begin with, section 2.2.1 examines studies which shows that production and perception are related and that perception precedes production. This gives grounds to argue that the difficulty second language learners experience with L2 speech may possibly have its root in their difficulty perceiving the L2 sounds.

Next, section 2.2.2 discusses how perceptual ability develops in first language acquisition and the impact of L1-based perception in acquiring a second language. Studies have shown that in second language acquisition, L1-tuned phonology *filters out* acoustic differences in the input that are not phonemically relevant in the L1 and *assimilates* non-native phones, in a sort of *magnet* effect, to that of the native *prototype*, which results in *mis-perception* of the target sound, which in turn results in incorrect production of the target sound.

2.2.2 How perception is related to production in L2 phonology

In reviewing the literature, there are some studies which indicate a dissociation between production and perception, as well as ones which suggest that learners produce phonetic forms which they cannot discriminate or identify (Sheldon & Strange 1982). For example, Sheldon & Strange (1982) found that Japanese learners' production of English /r/ and /l/ contrast was more native-like in their production than their perception. Such a finding may be explained by either the effect of conscious training or by the effect of the testing procedure, i.e. mimicking rather than acquiring the target form. Bever (1981) hypothesised speech perception and production develop independently in second language acquisition. According to Bever, in first language acquisition a *psychogrammar* 'equilibrates' (or aligns) perception and production, in other words, they should work in tandem. Bever further claims that once production and perception are brought into alignment, and there is no ongoing learning activity when the role of psychogrammar ceases, perception and production progressively become independent and the critical period for speech learning ends. This results in the difficulty post-puberty language learners experience. This hypothesis is rejected by Flege because if this is correct, as he argues, "one would not expect to observe correlations between measures of post-critical period L2 learners' production and perception of L2 vowels and consonants" (1999:113). However, "the results obtained in recent studies do show significant, albeit modest, correlations" (1999:113).

The majority of literature on both first and second language acquisition suggests that perception and production ability is indeed related (Dreher & Larkins 1980, Pimsleur 1963).

In first language acquisition, Vihman (1992) and Werker & Pegg (1992) claimed that there is a relationship between perception and production, given that language-specific influences on perception seem to parallel that seen for production at around the age of 6-12 months. Werker (1993) states that the direction of this influence is not yet known, and that it could be the case that the influence is only unidirectional, that perceptual retuning directs vocal production. But she also considers the possibility that the influence could be bi-directional, given that the language-specific influences on perception of prosody in early infancy could direct the child's attempts to reproduce prosodic elements such as intonation, and later on babbling (Best, Roberts & Sithole 1988; Kuhl, Williams, Lacerda, Stevens & Lindblom 1992, for discussion see also Werker 1993).

Although the relationship between perception and production is complex and evidence pointing to the direction of influence of the two abilities is not easily obtained, a number of studies in second language acquisition support (either directly or indirectly) the view that perception and production is related and that perception precedes production.

Major, in his earlier work in 1987, hypothesised that learners with good perceptual abilities of the L2 have a mental representation (near) identical to that of native and they will gradually proceed to produce forms closer to the TL. On the other hand, learners with poor perception need to first improve their perception and then production. He examined the perception and production of English vowels by Brazilians whose overall pronunciation of the TL language was relatively good or poor and found that /æ/ tokens produced by students with mild accents were more intelligible than the /æ/s produced by students with stronger accents.

Flege (1981) reported a study in which he found that the perception of L2 sounds may be more accurate than learners' production, which is consistent with the hypothesis that perception 'leads' production in second language acquisition, in other words, production may lag behind perception. Flege tested three groups of subjects: monolingual speakers of English and Taiwanese English speakers who had lived in the U.S. for an average of 1.5 years, and 5.3 years. The subjects rated the foreignness of English sentences produced with various degrees of foreign accent. It was found that the subject group which had lived in the U.S. longer were more sensitive perceptually to the phonetic characteristics of English. The subjects' production was also evaluated and it was found that the two non-native groups spoke with an equal foreign accent. Taking together the results of the perception and production experiment, Flege concluded that production may lag behind perception.

Studies by Oyama (1973, 1982a,b); Meador, Flege & MacKay (1997); Flege, Bohn & Jang (1997); Flege & Schmidt (1995); Schmidt & Flege (1995) all found that perceptual ability and productive ability were related in that the better the subjects perceived English, the better they were at producing it.

Oyama (1973) tested sixty U.S. immigrants from Italy. They had lived in the States ranging from five to eight years and their ages of arrival were between six and twenty years old. English native speakers were asked to assess the subjects' degree of foreign accent in paragraph-length speech samples. To test sentence perception, she had her subjects repeat as many words as possible in a set of English sentences presented in noise. She found that the more accurately the native Italian English speakers were able to produce English sounds, the larger the number of words they were able to repeat in noise.

Meador, Flege & MacKay (1997) carried out a study which extends Oyama's study. They had 54 Italian immigrants to Canada (mean age of 48) who arrived in the country between the ages of three and twenty-three and had lived in Canada for an average of thirty -four years. Semantically unpredictable English sentences (e.g. 'The blond dentist ate the heavy bread') having a single syntactic form (NP-V-NP) were examined. The participants were asked to repeat the words of a sentence presented. They found that the more accurately the participants pronounced English sentences (which was rated by native English-speaking judges), the larger the number of words they were able to repeat. In other words, they found that perceptual ability and productive ability were related in that the better the subjects pronounced English, the better they were at comprehending English sentences (r = 0.646). Meador et al. note that the correlation reported may actually have underestimated the relationship between the participants' ability to perceive and produce the vowels and consonants in the sentences, because individual differences in phonological short term memory (PSTM) may have affected the results. They then measured participants' PSTM by having them repeat non words formed by concatenating two to five Italian CV syllables. When the variation in the PSTM was partialled out they found a even stronger correlation between perception and production (r = 0.734).

Flege, Bohn & Jang (1997) obtained similar results in their vowel study with late bilinguals. They had twenty speakers each of German, Spanish, Mandarin, and Korean who were assigned to relatively experienced or inexperienced subgroups based on their length of residence in the US. Target vowels were embedded in a list of consonant-vowel-consonant (CVC) English words containing one of /i, ɪ, ɛ, æ/. The eighty subjects' accuracy in producing English vowels was assessed with acoustic measurement and also by English native speakers. First of all, vowel production accuracy was measured by the size of the spectral (F1, F2) difference that the participants produced between /i/-/ɪ/, /ɛ/-/æ/. Next, native judges were asked to identify which vowels had been spoken. To test the subjects' perception, they were asked to identify vowels in synthetic *beat-bit* and *bat-bet* continua. The perception test was a two-alternative forced-choice identification test. Results showed that relatively experienced non-native subjects both produced and perceived English vowels more accurately than inexperienced subjects. Multiple regression analyses confirmed that the non-native subjects' accuracy in producing English vowels was related to their accuracy in perceiving the same English vowels.

Flege & Schmidt (1995) and Schmidt & Flege (1995) examined the voice onset time (VOT) dimension in production and perception of word-initial English stop consonants. They examined forty native speakers of Spanish who had arrived in the U.S. as young adults. They had the subjects assess the goodness of the synthetic continua with varying degrees of VOT. It was found that the correlation between the VOT values produced by the proficient participants and their perceptually preferred VOT values were significantly correlated. This finding suggests again that as non-native adults become proficient in a second language, their production and perception align.

Champagne-Muzar, Schneiderman & Bourdages (1988) examined whether perceptual training induces better production performance. Tests carried out included discrimination and production of French phones, rhythm and prosody by learners of various L1 backgrounds. There were two subject groups: one group of twelve subjects who underwent a phonetic training session and a control group consisting of seven subjects who did not get the training. The training programme consisted of twelve one-hour, tape-recorded lessons and an accompanying workbook. In the first six lessons, the learners were introduced (with the help of graphic representations of intonation and rhythmic contours) to segmental and suprasegmental aspects of French through a series of discrimination exercises. The training did not include oral production. While the training session was administered to the test subjects, the control group spent their laboratory periods doing listening comprehension exercises. Testing took place prior to the start of the French course, in which both groups were enrolled simultaneously, and at the end of the semester immediately after they had finished the course. Testing included three types of discrimination tests and two production tests which measured the ability to imitate French phones. Investigators rated the production data with five or four point scales with the endpoints labelled as native and non-native. The results loosely supported the authors' predictions that 1) there is a positive association between discrimination and production ability and 2) improved discrimination abilities will result in more native-like production.

Similarly, Bradlow, Pisoni, Yamada & Tohkura (1996) demonstrated that perceptual training induce more accurate production of the L2 contrasts in the absence of speech production training. The evidence supports the view that correct perception of the TL is necessary condition for successful production of a second language. They examined acquisition of English /r-l/ contrast by 11 adult Japanese learners who had gone through a perceptual training programme for the contrast. They also had 12 Japanese speakers as controls who did not undergo the treatment. Both groups were tested for their perception and production before and after the /r-l/ identification training using a high-variability presentation format. The training programme had 45 sessions over a period of 3-4 weeks of perceptual identification with feedback. The post-test phase included a perceptual identification test (identical to the pre-test) and two tests of generalisation. The tests of generalisation consisted of a minimal word pair identification task with novel words spoken by a new speaker and novel words produced by one of the speakers used in creating the training stimuli. The experimental group showed significant perceptual learning as a result of the programme (from 65% correct identification to 83% correct identification) and this perceptual learning generalised to novel items spoken by new talkers. Improvement in the subjects' production as a consequence of perceptual training was evaluated using two tests. The first was a paired-comparison of the pre-test and post-test tokens. Native-speaking judges were asked to decide which version of the target word sounded better using a 7 scaling technique. The results of the first judgement test by native speakers showed significant improvement in the test subjects' perceived rating of /r/ and /l/ production as a consequence of perceptual training. This method was selected because it was expected to be sensitive to small differences in articulation. The second test was a minimal-pair identification task which provides information about a change in speech intelligibility specifically related to improved /r/ and /l/ articulation. It was found that the post-test productions were more accurately identified by English listeners than the pre-test production in a two-alternative minimal-pair identification test. Bradlow et al. (1996) concluded that the knowledge gained during perceptual training for /r-l/ contrast transferred to the production domain.

Taken together, these studies suggest it is safe to conclude that in second language acquisition, learners' overall pronunciation of their second language and their perception ability are related, and perception precedes production in L2 phonology.

2.2.3 Developmental changes in perceptual abilities and its effect on L2 phonology

Findings from infant head-turning studies suggest that human beings begin life with language-universal/general abilities for discriminating the possible segmental phonetic contrasts used in the world's languages. In other words, infants come into the world equipped with broadly tuned perceptual and attentional mechanisms. The support for this theory comes from studies showing that infants have a capability to discriminate minimal pairs which include both native and non-native phonetic contrasts (see for review, Jusczyk 1985, Kuhl 1987; Mehler 1985). For example, English-learning infants have been shown to be able to discriminate pre-voiced vs. voiced distinction even though it is not used in their ambient language (Aslin, Pisoni, Hennessy & Perey 1981). English-learning infants were also found to be able to discriminate non-native differences in place-of-articulation as found in Hindi retroflex/dental contrasts /t/ vs. /ɖ/ (Werker & Lalonde 1988, see also Werker & Polka 1993). Furthermore, it was found that, under certain testing conditions, even some contrasts that are said to be difficult for young infants (e.g. /f/ vs. /θ/ and /s/ vs. /ʃ)/ have been shown to be discriminable (Kuhl 1979, Eilers, Gavin & Oller 1982; Levitt, Jusczyk, Murray & Carden 1988).

As has been outlined, at birth all infants seem to be capable of discriminating many, if not all, of the possible phonetic contrasts present in the world's languages. These findings have demonstrated that human beings have the sensory and perceptual prerequisites to eventually acquire any spoken language (Jusczyk 1992). If so, when and how does language transform our language-general abilities to language-specific ones?

By the second half-year of life, listening experience with the native language begins to influence the perception of contrasts that are non-native. For example, Werker showed that by 10-12 months of age, infants demonstrate a failure to discriminate foreign contrasts they once discriminated (Werker 1989, and see also: Werker & Tees 1984; Werker & Lalonde 1988).

The absence of exposure to contrasts that are not distinctive is said to result in a selective loss of the ability to discriminate those contrasts. The mechanisms responsible for this loss of sensitivity may be neural, attentional, or both. Based on these infant studies, Kuhl (1979, 1987, 1993) proposed that '*prototypes*' (good exemplars) play an important role in speech perception. The Native Language Magnet (NLM) theory describes how experience with a specific language alters the initial state. As early as six months of age, infants develop stored representation of speech information based on the language they are exposed to, and these representations of native-language sounds constitute the beginning of language-specific perception. According to the model, infants are born equipped with the ability to partition the sound stream into gross categories separated by natural boundaries. This allows infants to separate phonetic units into rough categories that conform to the language input they get. Kuhl, Williams, Lacerda, Stevens & Lindblom (1992) supported their arguments with data from non-human animals. The boundaries reflect the infant's phonetic category prototypes and function like perceptual magnets for other stimuli (Kuhl 1993), attracting nearby members of the category. For example, Kuhl, Williams, Lacerda, Stevens & Linblom (1992) found that American and Swedish infants show magnet effects for, and only for, their native-language prototypes as early as 6 months of age. This induces certain perpetual distinctions to be minimised (i.e. those near the magnets themselves) and certain perceptual distinctions to be maximised (i.e. those far from the native language). The effect of magnet acquisition on the boundaries that divide the underlying phonetic space by infants from three languages (English, Swedish, Japanese) is shown in Figure 2.1. The boundaries, shown here as division in a two-formant vowel space are formed after half a year of exposure to the ambient language. The distributional properties of vowels are schematically illustrated by the dots.

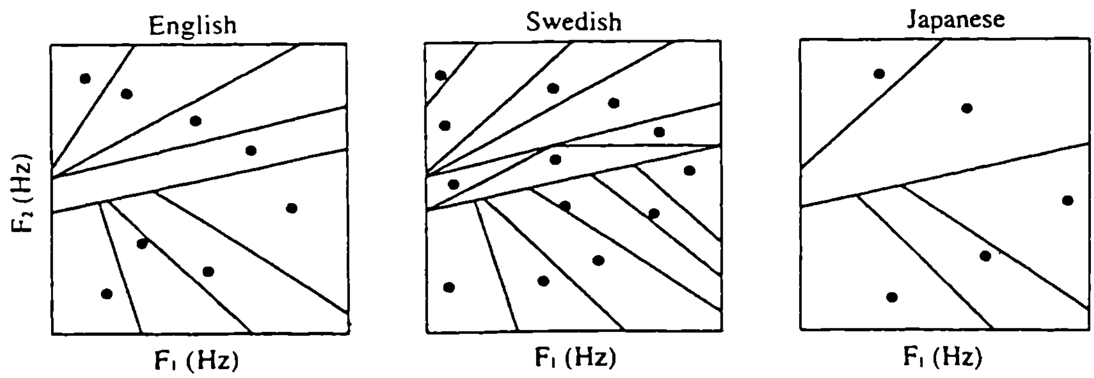

Figure 2.1 NLM theory (taken from Kuhl 1993:130)

According to this acquisition model, infants in the latter half of their first-year and eventually as adults are incapable of discriminating non-native sounds due to the magnet effect. Its implication for second language acquisition in general is that once the language-specific magnet is in place and the phonetic boundaries are fixed, foreign sounds tend to be pulled to a single magnet, making them no longer discriminable. By this account, foreign sounds which are similar to those in the native language are particularly vulnerable to the magnet effect, and sounds that are not similar to the native language category are relatively easy to distinguish from the native language sound.

Similarly, Best and her colleagues developed a 'Phonological Assimilation Model' (Best, Roberts & Sithole 1988, Fower, Best & McRoberts 1990) and later a model called the 'Perceptual Assimilation Model' (PAM) (Best 1994). They hypothesised that the influence of the native phonological system on infants and adult listeners entails the perceptual assimilation of non-native phones to that of native ones with which those non-native phonemes share the greatest similarity in phonetic characteristics. They assessed English-speaking adults' and infants' (6-8, 8-10, 10-12, 12-14 months) discrimination of the Zulu apical versus lateral click contrast plus /a/ vowel. This non-native contrast was expected to be non-assimilable to any English phonemes because the suction-release gesture used in them is not employed in English. They also had a control contrast English /ba/-/da/. In each test, the subject was conditioned to fixate on a colourful slide to hear repetitions of the multiple natural tokens of the habituation syllable, which terminated whenever the infant looked away from the slide. To assess for discrimination, mean looking times were computed for the two trials immediately preceding the stimulus shift (habituation level) and for the first two post-shift trials (response recovery). For comparison, they also had adult listeners take the test as well. The results for the study supported the prediction that discrimination would remain high across all ages. They then ran a new test including Zulu clicks and Salish ejectives (a contrast which may be assimilated to a single native phone) and an English control contrast. They again used a visual fixation procedure with new groups of 6- to 8- vs. 10- 12-month-olds, twelve per age group. They found that both age groups discriminated the Zulu contrast and the English control, whereas only the younger infants discriminated the Salish ejectives. Therefore, the developmental difference between the two non-native contrasts could be attributed to differences in perceptual reorganisation for those types of contrast.

Proposals by Kuhl and Best are consistent with the idea elaborated by Nosofsky (1986, 1988). He has shown that selective attention to specific stimulus dimensions can modify the underlying psychological space and change the perceived similarity relations: attributes on the attended dimensions become more dissimilar to each other and unattended dimensions become more similar to each other. This selective attention strategy serves to maximise within-category similarity among exemplars sharing the same dimension and minimise between-category similarity. Selective attention for speech perception serves to obtain the maximum effect on favouring important distinctive contrasts in a particular language and the attenuation of cues for non-contrastive distinctions. For example, Terbeek (1977) demonstrated that prior language experience affects vowel perception by modifying the perceived psychological distances between the vowels. He used a scaling technique to measure the magnitude of differences between pairs of vowels. He had native speakers of five different languages listen to the vowels and found that the perceptual distances between a pair of vowels was judged to be much larger if members of the pair contrasted phonologically in the subject's native language than pairs which were not phonologically distinctive in the language. According to this view, the apparent loss of non-native discrimination is not a sensor-based loss but a change in selective attention. In other words, a systematised restructuring of the psychological space occurs with language experience, favouring distinctive contrasts in one's native language and the attenuation of cues for non-native contrasts.

In support of this view, Flege's (1992, 1995) 'Speech Learning Model' attempts to explain how speech perception affects phonological acquisition by a learner perceptually identifying 'new' and 'similar' sounds. Flege argues that 'new' sounds are not identified with L1 sounds. On the other hand, 'similar' sounds are perceived to be the same as certain L1 sounds and such 'equivalence classification' prevents the establishment of new phonetic categories. Flege further claimed in Flege, Munro & Fox (1992) that the L1 influences the perception of a second language in that it '*filters out*' acoustic differences that are not phonemically relevant in the L1. Consequently learners mis-perceive the target sound. Multi-dimensional scaling (MDS) analyses showed that Spanish-speaking listeners used fewer dimensions in judging between vowel dissimilarity than native English-speaking listeners. Here, the finding agrees with the observation that a speaker of a language which has a small vowel inventory, as Spanish does, needs to use fewer dimensions to identify vowels than speakers whose native language has more contrasts to maintain in the language (Lindblom 1988). Gottfried & Beddor (1988) found that when American subjects were asked to identify the member of synthetic French /o/-/ɔ/ continuum, they had a tendency to rely more on duration cues than native speakers of French. This was attributed to the fact of English speakers' greater use of duration cues in their native language.

The 'Feature Competition Model' developed by Hancin-Bhatt (1994a,b), which was an expansion of earlier work by Ritchie (1968) and Michaels (1973) takes Flege's model one step further and provides an algorithm for determining how, when and why L2 sounds get mapped on to L1 categories by utilising the theory of Feature Geometry (Clements 1985; Sagey 1986). Similar to other speech models just described, the model assumes that L2 sounds are assimilated to L1 categories. Brown and Matthews carried out studies on first and second language acquisition which employ this model (Brown & Matthews 1993, 1997; Brown 2000). Here I will report on a study from Brown (2000) which attempted to capture the nature of the mechanism that maps the L2 input onto L1 phonological categories.

Languages differ with respect to their phoneme inventories and the set of phonological features they manipulate. Under Feature Geometry (Clements 1985; Sagey 1986), phonemes consist of distinctive features which are organised into a hierarchy of constituents, and each phone has a structural representation that distinguishes it from other segments in the inventory. The full set of features manipulated in the world's languages is shown in Figure 2.2.

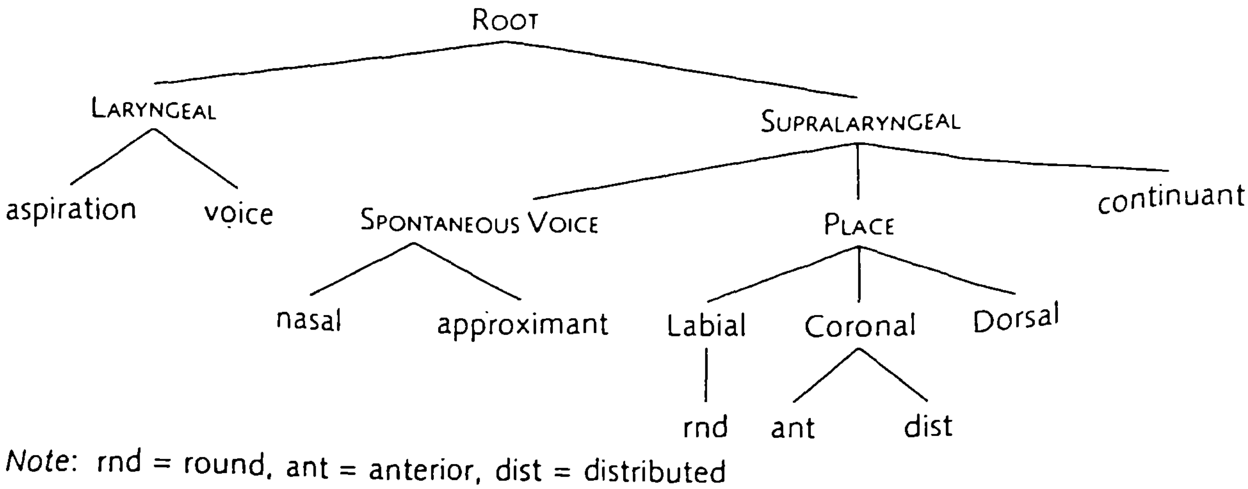

Figure 2.2 A model of Feature Geometry (taken from Brown 2000:12)

According to the theory, a child needs to determine which of the phonological features are used to contrast phonemes in the ambient language and construct the appropriate representations. Acquisition occurs as the child detects that two segments in the input are used. The phonological feature that differentiates the two segments gets added to the child's grammar. Brown & Matthews (1993; 1997) found that children will phonologically distinguish segments that require less structure to differentiate before distinguishing those segments that require highly articulated structure. They build up necessary features as they acquire the ambient language (the *Building Hypothesis*) rather than starting off from a fully elaborated Feature Geometry (which Universal Grammar provides), and subsequently truncating or pruning features which is not supported by phonological contrasts in the input (the *Pruning Hypothesis*).

Brown (2000) investigated what implications the theory has on second language acquisition. She carried out three experiments (Experiments, I, II & III, as described later) using two testing techniques which investigated how the phonologies of Japanese, Korean and Chinese speakers affect their acquisition of English contrasts.

The first test was an AX Discrimination which assessed the subjects' ability to acoustically discriminate the English contrasts. In this test, subjects heard a minimal-pair (one item containing for example, an /l/ and the other containing either /r/ or /l/) and were asked to indicate whether the words (all monosyllabic English words) were the same or different, e.g. rip/lip. The second test was a Forced Choice Picture Selection task which was carried out to assess learners' phonological competence. Subjects were presented with two pictures and a verbal cue which corresponded to one of the pictures. Prior to the test, subjects were familiarised with each of the pictures and corresponding names by means of training book.

Three types of English contrast, namely, /l-r/, /b-v/, /p-f/, /f-v/, and /s-θ/ were examined. These pairs are not contrasting in Japanese, Korean or Chinese. The internal structure of each pair is given in Figure 2.3. As you can see in the figure, each contrast is distinguished by different phonological features, e.g. the /l-r/ contrast is distinguished by the feature [coronal].

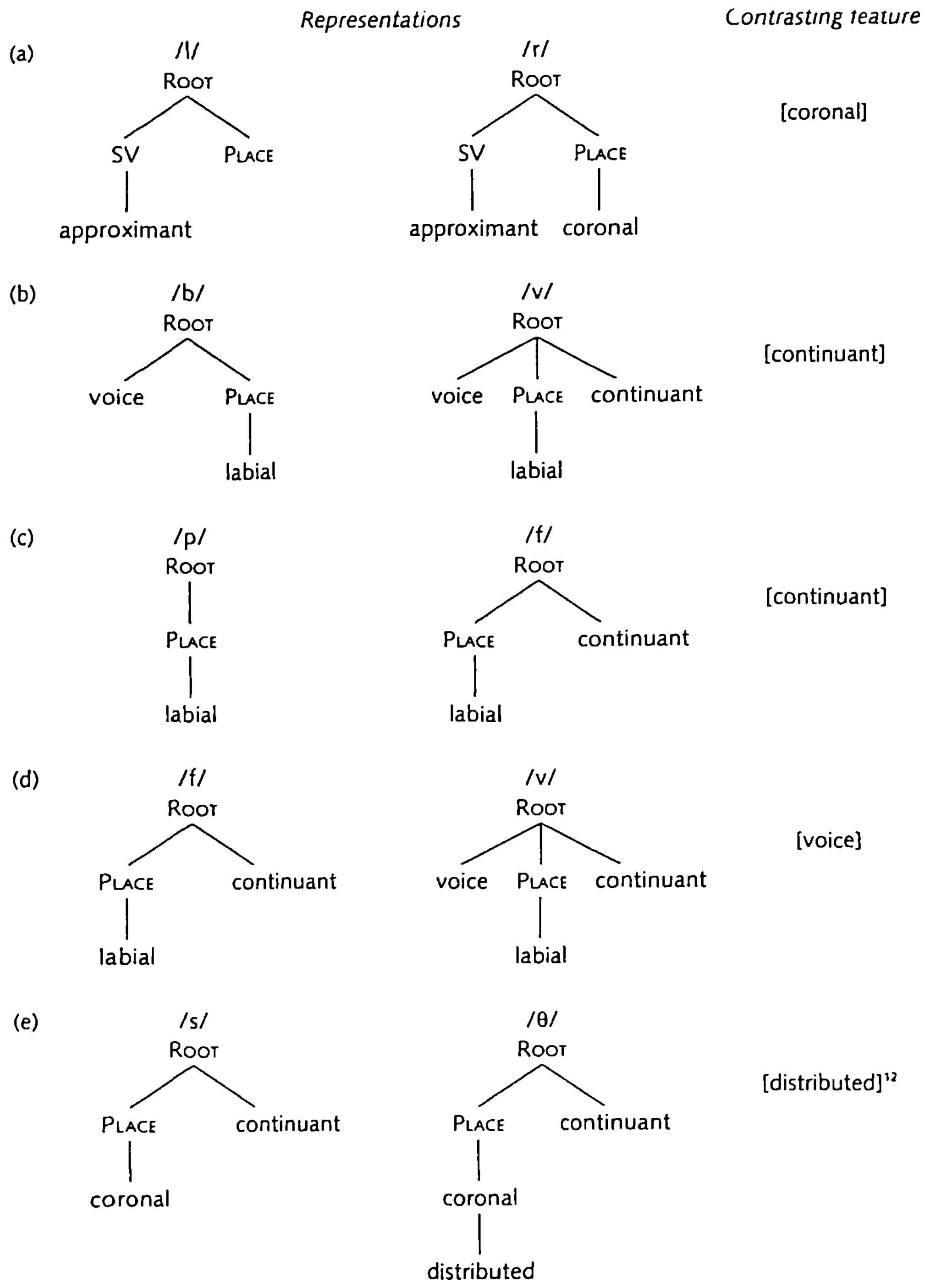

Figure 2.3 Representation of contrasts under investigation (taken from Brown 2000: 22)

These contrasts were of particular interest because, although the three languages do not have such contrasts, some of the features needed to distinguish the contrasts under investigation were present in some languages for some contrasts. Figure 2.4 gives the figure inventories of Japanese, Korean and Chinese: the features [continuant] (a feature needed to distinguish /b-v/ and /p-f/) and [voice] (needed for /f-v/) are present in the grammar of all three languages. On the other hand, the feature that distinguishes the /r-l/ contrast ([coronal]) is present only in Chinese; the feature that distinguishes the /s-θ/ contrast ([distributed]) is not present in any of the languages.

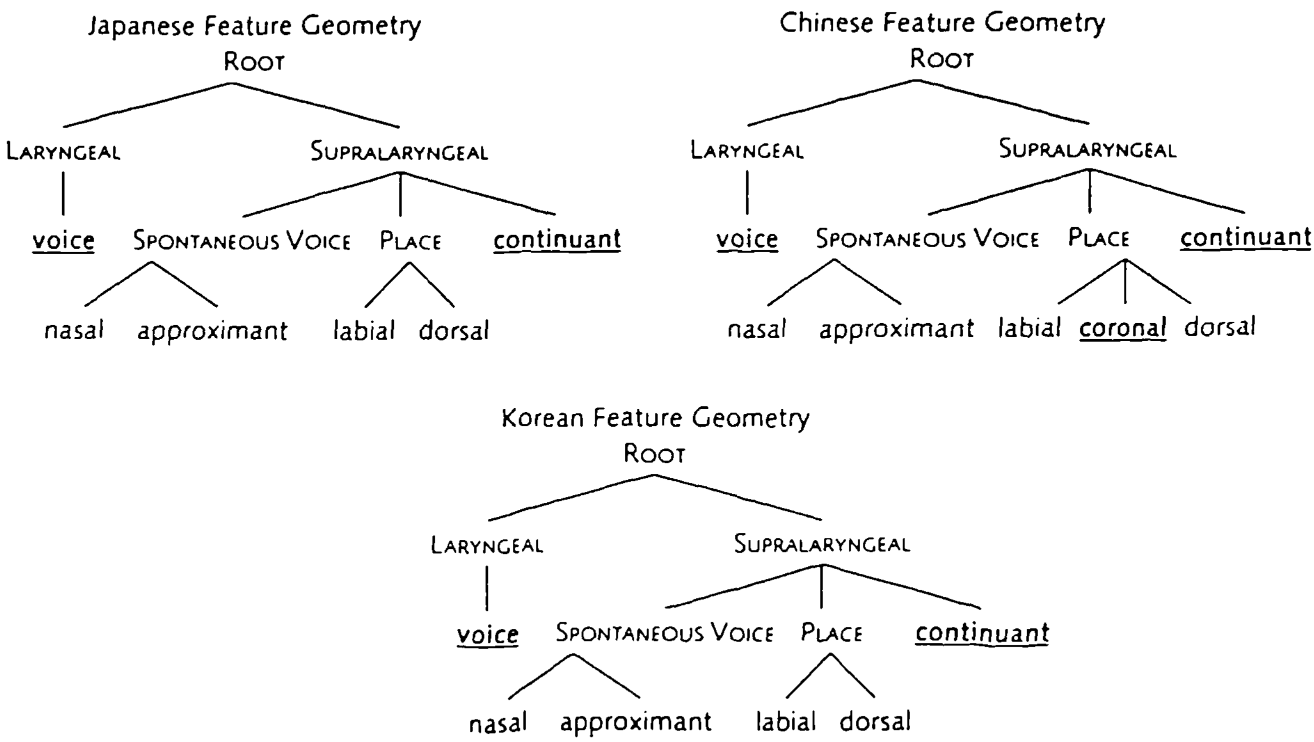

Figure 2.4 Feature inventories of L1 groups under investigation (Brown 2000:25)

Brown outlined the course of acquisition or absence of acquisition as follows. Similar to Flege et al. (1992), she predicted that at the earliest stage of acquisition, the phonemes of the L1 have a strong influence on the perception of non-native contrasts. The L2 sounds will automatically be mapped onto the L1 categories. Learners either ignore the differences they sense or do not realise the differences at all. Over the course of development, and with increased exposure, there are two paths L2 learners may take with respect to phonemes which do not contrast in their native language. If there is a relevant distinctive feature in the learner's L1 grammar which is needed to distinguish a contrast, learners will come to perceive the difference and be able to acquire those contrasts and a new phonological category will be established. As a result of the presence of the relevant contrasting feature, portions of the L2 input will not map adequately onto the L1 system. The slight mismatch between the L2 input and the L1 structures will cause perceptual reorganisation. On the other hand, if the feature that distinguishes a given non-native contrast is absent from the L1 grammar, then the L2 input will map perfectly onto an existing L1 category because the learner cannot perceive the difference, and there will be no trigger for acquisition.

Results from the three experiments were in accordance with the acquisition model just described. In Experiment I, 15 Japanese speakers who had learned English as their only second (foreign) language were tested. They ranged in age from 20 to 32 years. They went to the U.S. to study in either an undergraduate or a graduate programme and the mean residence in the U.S. was 3.5 years. Information concerning their proficiency levels in English was not given. It was found that they were able to distinguish /b-v/ and /f-v/ contrasts, despite the fact that these were both non-native contrasts, while they did poorly on the /l-r/ contrast. These results can be accounted for by the fact that there exists in Japanese phonological features needed to distinguish /b-v/ and /f-v/, [continuant] and [voice] respectively. The phonological feature needed for the /r-l/ contrast, namely [coronal], is not present at all in the learners' first language, Japanese.

Experiment II provided data of the acquisition of a particular contrast across different groups of speakers. 15 Japanese speakers, 15 Korean speakers and 15 Mandarin Chinese speakers participated in the experiment. All the experimental groups resided in Japan and were either undergraduate or graduate students in a Japanese university. Their proficiency levels in English were not mentioned in Brown's paper. In accordance with the theory of phonological interference adopted in the study, it was found that speakers from all three languages were able to distinguish /b-v/, /p-f/ and /f-v/ without problems. It was predicted that speakers of all three languages would have difficulty distinguishing the /s-θ/ contrast because the phonological feature needed to distinguish the contrast [distributed]) is not present in any of the languages; this prediction was borne out. Finally, the three language groups differed with respect to their ability to perceive and acquire the /r-l/ contrast: Chinese speakers accurately perceived the contrast and Japanese and Korean speakers failed to do so, since the former has the feature [coronal] while the latter two do not.

Experiment III investigated how the influence of the native grammar of non-native contrasts changes over time. 35 native speakers of Japanese were divided into two experimental groups: low-level (n=20) and high-level (n=15). The contrasts /r-l/ and /b-v/ were tested in the experiment. The native /p-v/ contrast was also included in the test as a control item. There was a control group consisting of 10 native speakers of English. It was found that both groups performed significantly worse than the English controls with respect to the /r-l/ contrast, with no difference between the two Japanese groups for this contrast. Brown concluded that this is an indication that the ability to discriminate the /r-l/ contrast does not change over time. As for /b-v/, the low-level group was not as accurate as the learners in the high-level group at distinguishing the /b-v/ contrast; thus there was improvement in the Japanese speakers' ability to perceive this non-native contrast.

These results confirm the prediction that learners are able to perceive and acquire those contrasts which are distinguished by a feature that their L1 employs for independent reasons (i.e. /b-v/ for the feature [continuant]) but are unable to perceive the contrast that is distinguished by a feature not utilised in the L1 (i.e. /r-l/ for the feature [coronal]). The developmental data obtained from this cross-sectional study fits nicely with the prediction Brown made with respect to the course of acquisition (or absence) of second language phonemes.

In summary, how biological maturation affects interlanguage phonology was discussed by first reviewing the development of perceptual abilities in first language acquisition. Based on the view that perceptive and productive ability is closely related in second language acquisition, *mis-perception* of a target language was suggested as a possible cause for L2 learners' failure to achieve native proficiency. It should be added that this can be expected to be reinforced in foreign language situations where learners do not receive input as rich as in first language acquisition or children learning a L2 in a naturalistic setting.

From a methodological point of view, one can say that perception sometimes plays a crucial role in the identification of both phonological and phonetic processes in interlanguage. In these cases, only a parallel study of both perception and production by a learner studied longitudinally will shed more light on the relative role of these two components in interlanguage.

2.3 Implications for pronunciation teaching

We now come to the question of what goals should be set for pronunciation teaching. While native-like pronunciation may be a goal for particular learners (e.g. learners who plan to become teachers of English), and while we should not discourage learners from setting themselves high goals, it is generally accepted that *intelligibility* is the most sensible goal for the majority of L2 learners (Kenworthy 1987). Kenworthy defines *intelligibility* as 'being understood by a listener at a given time in a given situation' and that the 'more words a listener is able to identify accurately when said by a particular speaker, the more intelligible that speaker is' (1987:13).

In this section, I will first discuss factors affecting intelligibility. It is concluded that suprasegmental/prosodic factors (which include syllable structure and lexical and phrasal stress, intonation) consistently contribute to native listeners' perception of accentedness and that perceived foreign accent results in reduced intelligibility. The implications of this discussion for pedagogical purposes is that non-native speakers would benefit most from pronunciation training focusing on that aspect of their speech to increase intelligibility.

2.3.1 Foreign accented prosody and intelligibility deficits

A number of perception studies suggest that native listeners are sensitive to various properties of non-native utterances which diverge from the native norm when they are asked to evaluate the 'accentedness' or 'acceptability' of L2 learners' productions. Results from such studies indicate that prosodic characteristics may be one of the most important features in the perception of foreign accent, and that perceived foreign accent is in turn closely tied to reduced intelligibility of accented speech for native listeners.

Anderson-Hsieh, Johnson & Koehler (1992) reported that prosodic characteristics may be more important than other features in the perception of foreign accent. Cutler and her colleagues (Cutler 1976; Cutler & Norris 1988, among others), showed that prosodic information, such as the pitch contours of an utterance, focus a listener's attention on the location in a sentence of important information: prosodic information influences word identification by directing the listener's attention to particular items in an utterance.

Cutler & Norris (1998) showed that prosody can play a role in lexical segmentation for speakers of languages in which it is a predictable cue to word boundaries. For example, English as a stress-timed language, comprises sequences of strong and weak syllables. Strong syllables are most likely to signal the beginning of a content word, while rhythm in French is syllable-timed. This difference in prosodic structure between two languages was found to reflect native speakers' segmentation strategies. Since prosodic features affect perceptual processing in a language specific way, it must be learned.

Experienced ESL teachers were asked to rate the speech produced by 60 French native speakers learning English. The authors analysed the scores with respect to three phonetic attributes: segments, syllable structure, and prosody. A multiple regression analysis revealed that all three variables played a significant role in predicting the pronunciation ratings, but the prosodic variable had the strongest regression coefficient, suggesting the relative importance of prosody.

Similarly, Magen (1998) carried out a study assessing the contribution of various phonetic and phonological factors to the perception of global foreign accent. Two fluent but heavily accented Spanish learners of English recorded English phrases containing sounds or sequences of sounds suggested to be difficult for native speakers of Spanish. Investigated factors included those affecting syllable structure (initial epenthetic schwa, non-initial epenthetic schwa); those affecting vowel quality (vowel reduction, tense-laxness); those affecting consonants (final /s/ deletion, manner /tʃ-ʃ/) fricative voicing /z-s/, stop voicing; those affecting stress (lexical and phrasal stress). Native listeners rated the degree of foreign accentedness of the Spanish-accented speech and the acoustically edited speech for each factor which was suspected to contribute to foreignness. Listeners were found to be sensitive to syllable structure, final /s/ deletion, consonant manner and lexical and phrasal stress, but were not sensitive to voicing differences. The results taken as a whole indicate that suprasegmental factors (which include syllable structure and lexical and phrasal stress) consistently contribute to native listeners' perception of accentedness.

If we concentrate on English as a target language, claims have been made that inappropriate timing of syllables and inappropriate patterns of stress alternation account for a major part of the intelligibility failure of L2 learners' productions (e.g. Adams 1979; Nelson 1982; Kenworthy 1987; Tiffen 1992). For example, Tiffen (1992) showed that the strongest cause of intelligibility deficits for Nigerian speakers' English was rhythmic/stress errors, followed by segmental, phonotactic, and lexical/syntactic errors. Suenobu, Kanzaki & Yamane (1992) reported that intelligibility of Japanese-accented English was lowest when it was analysed with respect to consonant deletion, followed by inappropriate accentuation of words. Hutchinson (1973) demonstrated that Spanish learners who maintained a greater durational contrast between English stressed and unstressed syllables were given better pronunciation ratings by native listeners than learners who spoke English with smaller durational contrasts.

However, most of the results from the aforementioned studies were obtained by native listeners' subjective ratings of overall prosody, rather than on instrumental analysis of the non-native speech. Tajima, Port & Dalby (1997) provide instrumental data which directly addressed the relationship between possible acoustic-phonetic properties of foreign accented speech and intelligibility. They employed the "speech transformation method" which was previously adopted by Osberger & Levitt 1979 and Maassen & Povel 1985. Here, speech stimuli were instrumentally modified to match normal production in specific acoustic dimensions so that their effects on intelligibility could be evaluated. Short English phrases were produced by a Chinese speaker of English and by a native English speaker. Using objective methods involving LPC-based speech coding and resynthesis (e.g. Atal & Hanauer 1971) along with dynamic time-warping (e.g. Rabiner, Rosenberg & Levinson 1978), the Chinese speaker's productions were temporally modified so that the duration of acoustic segments matched the duration of corresponding segments in the native speaker's productions, while retaining the spectral and source characteristics of the original speech data. Here, they assumed that modifications to the duration of individual acoustic segments simultaneously affect higher-order temporal features, such as syllable shape and global rhythmic properties. Therefore, their interest was in the combined effect of "temporally defined properties" on intelligibility. It should also be noted that there seems to be a generally observed tendency by the non-native speakers to reduce the durational contrast between English stressed and unstressed syllables (Hutchinson 1973). In disyllabic words, the first vocalic segment in each word (stressed) had an average of 13ms in the non-native productions while the differences were much larger in the native speaker's productions, with an average of 90ms.

On the other hand, the native speaker's speech was modified (distorted) to match the Chinese-accented productions. 36 native speakers of American English were recruited for a listening test to assess the effect of temporal modification on intelligibility. They listened to four types of productions: the original Chinese speaker version (OC), original English speaker version (OE), temporally corrected Chinese version (TCC) and the temporally distorted English version (TDE). The test was a four-alternative forced-choice identification test. The choices included: the correct phrase, plus three phonetically similar distracter phrases which were suggested by a different panel of listeners. The obtained intelligibility was quantified as percent correct recognition of the phrases in the test. To avoid ceiling effects in listeners' performances on the OE tokens, the stimuli were presented with background "cafeteria noise". Intelligibility of the OC was poor, (39% correct), but improved significantly (to 58%) after temporal modification. On the other hand, performance on the OE was high (94%), but declined significantly (83%) after temporal distortion according to the Chinese speaker's timing. These results suggest that the native listener's ability to recognise English phrases is significantly influenced by whether or not the phrases have appropriate native-like temporal properties, which provides some quantitative support for claims that have been made regarding the role of speech timing and rhythm on the intelligibility of foreign-accented utterances (Nelson 1982; Kenworthy 1987). They also suggested implications for language teaching: intelligibility of foreign language speakers may be enhanced if explicit training is provided on temporal properties of their speech.

In summary, these results indicate that native listeners are selectively responsive to non-native utterances which diverge from the native norm. It was suggested that prosodic characteristics may have important features in the perception of foreign accent and that perceived foreign accent results in reduced intelligibility. The implication of this discussion for pedagogical purposes is clear: if foreign-accentedness in prosody contributes to serious intelligibility deficits and instrumental modification of temporal properties results in enhanced intelligibility, there is a good reason to believe that non-native speakers would benefit from pronunciation training focusing on that aspect of their speech.

However, application of such research findings may not be so simple as we would hope it to be. Leather & James (1991) reported that at the segmental level, transfer is prominent only during the early stages of acquisition, while transfer persists well into advanced stages of acquisition with respect to prosodic structure. Is prosody more difficult than other areas to improve?

Recall the study by Champagne-Muzar, Schneiderman & Bourdages (1988) reported in 2.2.2. They carried out a study which measured the relationship between the comprehension and production of segmental and prosodic elements in a second language and examined what effect a period of systematic discrimination training in the second language had on this relationship. The treated group underwent twelve one-hour phonetic training session while the control group underwent listening comprehension exercises. Testing included three types of discrimination tests and two production tests, which measured the ability to imitate French phones. Results indicated that there was a positive association between discrimination and production ability, and improved discrimination abilities will result in more native-like production. What is interesting here is that the treated group made significant gains in their pre to post-test performance on all measures except discrimination test 3, which tested the perception of rhythm.

The study reported in this thesis supports this view. It was found that when adult Japanese speakers of English who had previously received mostly foreign-accented input during their early foreign language experience, but then moved to the target language setting, were able to improve their phonology. However, it was also found in the perception study that there was a tendency L1 transfer to persist with respect to prosodic structure.

Therefore, it seems to be the case that foreign-accentedness in prosody contributes to serious intelligibility deficits but at the same time this area is the most difficult to acquire. However, studies also suggest that providing L2 learners with input enhancement in the shape of perceptual (and production) training can make L2 learners' production more intelligible.

2.3.2 Explicit instruction

According to Bongaerts, Summeren, Planken & Schils (1997), the task of L2 learners can be made less difficult by providing them with input enhancement in the shape of perceptual training aimed at focusing their attention on subtle phonetic contrasts between their L1 and the target language. This can be followed by training learners in the production of L2 speech sounds in order to develop the finely tuned motor control requirements of accurate pronunciation. This is in accordance with Flege's (1987, 1995) work arguing that foreign accents are largely perceptually based, and to learn a new language one needs to shift from the L1 tuned phonetic categories to the continuous mode of perception, as L1 learners do.

In addition, Moyer (1999) found a significant relationship between outcome and type of phonological feedback received: those who were given both suprasegmental and segmental feedback scored closer to native in a predictably constant relationship. 24 highly proficient adult English learners of German were studied. Subjects were all exposed to German after the age of ten and presumed to be highly motivated given their extensive use of German through teaching, studying and research. A background questionnaire identified and grouped variables according to the following topics: (1) biological variables (age of immersion, age of instruction); (2) instructional variables (years of instruction, years of teaching, years of immersion, amount of phonological feedback, type of feedback [segmental, suprasegmental, or both]); (3) affective variables (type of motivation, self-evaluations of pronunciation, attitude towards pronunciation, self-evaluation of cultural and linguistic assimilation while immersed in the TL environment).

Subjects were asked to read (1) a list of 24 words; (2) a list of 8 sentences; (3) a paragraph of text; and also (4) participated in a free response section in which they could choose any one of five possible topics and speak freely for at least five to ten sentences. Four NS raters volunteered to listen to the speech samples for each talk and determined whether the speaker was a native German. Each subject (NNS) was rated by two judges. For each judgement, a confidence rating was requested (very confident, fairly confident, not confident). When applied to the binary judgement (NS vs. NNS) this resulted in a 6-point scale of judgement (1 = definitely not native; 2 = probably not native; 3 = maybe native; 6 = definitely non-native). These individual 6-point ratings were then averaged for both judges listening to all four speech tasks and a mean rating across tasks was assigned to each speaker based on that overall average.

It was found that the higher the age of immersion and instruction, the higher the mean rating (lower accuracy). It was also found that the higher the category for phonological feedback (indicating suprasegmental as well as segmental feedback) the lower the mean rating (greater accuracy). Given that those who cited stress, rhythm and intonation training or feedback did score significantly closer to the native range than those who did not, Moyer (1999) concluded that the type of feedback more than the amount is significant for acquiring native-level pronunciation.

Pennington & Ellis (2000) examined the performance of advanced Cantonese speakers of English on recognition memory for English sentences in which prosody was the feature discriminating otherwise identical sentence pairs. Although speakers' memory for the English sentences based on prosodic information was generally poor, when learners were explicitly directed to pay attention to intonation, significant improvement in recognition was observed for sentences in which prosody cued a marked information focus ('contrastive stress') versus an unmarked one ('neutral' sentence intonation).

Thirty advanced Cantonese speakers of English completed two experiments conducted in a sequence: untutored (implicit) condition [Experiment 1] followed by a tutored (explicit) condition [Experiment 2]. For each experiment, there was a *study phase* followed by a *recognition task.*

The recognition task in the two experiments were identical, and subjects were required to listen to 48 sentences and decide whether or not each sentence they now heard was exactly the same as one of the sentences they had heard in the previous study phrase.

In Experiment 1, subjects demonstrated a high level of lexical memory in their ability to recognise previously heard sentences and identify new ones, while they were poor at identifying sentences having the same lexis as in previously heard sentences spoken with different intonation. Pennington & Ellis argued that the results show learners did not attend to key phonological information in the speech signal of the L2 and they did not have the relevant knowledge of how intonation resolves ambiguity and relied solely on lexical information. The recognition phase of Experiment 2 revealed that subjects' memory for previously heard sentences spoken with new prosody could be improved to a small extent by explicit priming of contrasting cues. Pennington & Ellis (2000) suggest that emphasis for raising L2 prosodic awareness of contrastive pairs may help the L2 learners analyse the target language prosody as a representational system on a par with that of other systems of grammar, particularly if the training is a 'focused program in isolation from other skills' and if 'the program involves perceptual training such as audio and video feedback' (Pennington 1998:328).

As has been pointed out by several researchers in the 80's, explicit training in rhythm and stress patterns has generally been ignored (Leather 1983; Pennington 1989). This suggests the need to re-evaluate such approaches and include suprasegmental training for foreign language learners (see de Bot & Mailfert 1982, Chun 1988). Unfortunately, it seems to be the case that after 20 years since those observation were made, there has not been much change in the teaching situation at least in Japan where my subjects came from.

The new 'foreign language teaching scheme' ('shin kyouiku youkou') for junior/high schools published by the Ministry of Education ('Monbusho') in 1997 emphasises the importance of improving oral skills, an area which has long been downplayed while concentrating on reading, translating and vocabulary skills. Although finally the importance of oral skills in foreign language teaching has been acknowledged, and classes dedicated to listening/speaking skills have been scheduled, teachers are very much left on their own in how to teach such classes. Teachers may be using the time for general listening-comprehension without much focus on phonological features of English.

However, on the basis of the foregoing discussion, it seems that L2 learners benefit from explicit instruction and that emphasis on raising L2 prosodic awareness by encouraging learners to concentrate on the cues are helpful for the learners to interpret spoken sentences (Pennington 2000). Moreover, since perception training has been shown to facilitate production skills (Bradlow, Pisoni, Yamada & Tohkura 1997), it seems reasonable to suggest that explicit instruction in prosody is very much called for. Furthermore, such explicit instruction is effective in another way: it may act as an facilitator to raise learners' consciousness of their pronunciation accuracy. As discussed earlier, various studies show that among a range of exposure and attitudinal variables, concern for pronunciation accuracy was one of the most significant factors for developing L2 phonology (Suter 1976)..

As to how it can be best be taught, further research is needed. We need to further investigate the nature of the phonological problems with a specific L1 encounter and suggest; what should be included in the teaching programme, how it should be presented and empirically test the methods to develop the most effective pronunciation teaching programme for learners.

Chapter 3

English vs. Japanese phonology

3.1 Segmental phonology

3.1.1 Introduction

This section provides an outline of underspecification theory in order to capture the concept of phonetic underspecification of schwa in English. It also describes the behaviour of vowels in English and Japanese. Finally, it covers the implications for the second language acquisition of English schwa by Japanese native speakers.

3.1.2 The Underspecification theory

It has been argued (e.g. Kiparsky 1985; Archiangeli 1984, 1988; Archiangeli & Pullyblank 1994) that feature values that are predictable in a given language remain unspecified at the level of underlying representation. Underspecification theories assume that these underspecified features eventually get specified by default or by rules in the course of a derivation. Underspecification has been an essential concept to explain assimilation phenomena such as vowel harmony, which is a case of feature spreading (for a survey, see van der Hulst & van de Weijer 1995), where features spread to unspecified segments.

There are two major theories of underspecification. Radical Underspecification is promoted by Kiparsky (1985), Archiangeli (1984) and Pulleyblank (1988) and Contrastive Specification is promoted by Clements (1987) and Steriade (1987), among others. Because of empirical advantages, I will base my argument on Radical Underspecification over Contrastive Specification, I will base my argument on Radical Underspecification promoted by Archiangeli (1984). The conceptual differences between the two models of underspecification are given in footnote 2.

Archiangeli (1984, 1998) assumes that the grammar of a language is simplest when the underlying representations contain the minimal amount of specification compatible with maintaining the underlying contrasts in the language. Therefore, only those features crucially relevant to the characterisation of a particular segment need to be specified and features which do not play a distinguishing part in the identification of a segment are not present at the underlying level. Values are considered predictable if either a context-free or a context-dependent rule can be formulated to insert the absent values during the course of the derivation. She illustrates a case of underspecification with a standard five vowel system using the features [High][Low][Back] and [Voice] as shown below in (1): (a) is sparsely specified (b) is fully specified, and (c) gives a default rule under which segments are defined by universal preference. For example, based on typological markedness, vowels are usually voiced (i.e. the unmarked option), while devoiced vowels are universally marked. Therefore, all the vowels unspecified for [Voice] become [+voice] by this default rule.

(1) Standard five vowel system

(a) Sparsely specified features

	ɪ	e	ɑ	o	u
high		–		–	
low			+		
back				+	+
voice					

(b) Fully specified features

	ɪ	e	ɑ	o	u
high	+	–	–	–	+
low	–	–	+	–	–
back	–	–	+	+	+
voice	+	+	+	+	+

(c) Universal Default Rules

[+low] → [–high]

[+low] → [+back]

[ ] → [–low]

[ ] → [+high]

[ ] → [–back]

[ ] → [+voice]

To account for language-particular phenomena, additional rules need to be formulated. For example, to account for a less common five vowel system such as the one in Auca, which has five vowels consisting of /i, e, o, ɑ, æ/, an additional rule to allow [æ] and one to rule out [u] are included as language-particular rules.[2]

There is an essential concept which need to be mentioned to understand the essence of underspecification: the concept of *transparency*. Transparent segments allow the spreading of a harmony feature to go right through them, in other words, transparent segments are unspecified for the feature being spread. As discussed in Archiangeli (1988:198), transparency effects involve some segment which may intervene between the trigger and the target of a rule, and yet have no effect on the application of that rule. In other words, the intervening segment can be said not to be specified for the features involved in the rule but it also does not block the application of the rule as opaque segments do. Phonetic transparency of an unspecified segment can be illustrated acoustically by the case of glottal approximant /h/, which does not

---

[2] In Contrastive Underspecification, specific values are assigned only where the feature is being used to distinguish segment in the underlying representation. Therefore, it is necessary to determine whether a feature serves to contrast two segments or not in order to determine the correct underspecified representation. The following is the necessary algorithm for determining the contrastively specified representation: (1) the fully specified segments of the language needs to be posited (2) each of the segments are compared pairwise to determine which differ by a single feature specification (3) identify all the contrastive specifications (4) remove all unmarked feature specification. In this model, segment is needed to express the possible sounds of the language. Another conceptual difference between Radical and Contrastive Underspecification concerns the issue of cross-linguistic markedness. As described in the example of the five-vowel system, consideration of markedness is directly addressed in Radical Underspecification through default rules. On the other hand, in Contrastive Underspecification, 'the feature specification rules correspond directly to surface generalisations about the segment inventory' (Archiangeli 1988:197) and the model has no internal way to directly address markedness consideration. This is one of the reason why Archiangeli argued favourably for Radical Underspecification.

have intrinsic oral feature values of its own and is often described as a voiceless version of the following vowel by phoneticians and/or phonologists. Keating (1988) argues that /h/ is not assimilated to a following vowel as a result of acquiring feature values of the following vowel, rather, /h/ is simply interpolated through by a trajectory that moves from the preceding segment to the following segments. Therefore, in the case of intervocalic /h/, the spectrogram is simply interpolated through by the preceding vowel to the following vowel.

As stated earlier, the redundant features are usually filled in later by default rules which assign values to those features not specified in the underlying representations. According to Archiangeli & Pulleyblank (1994), the values which are not present underlyingly are filled in by two sorts of redundancy rules, i.e. default rules and complement rules. Therefore, they are no longer unspecified in the output. For example, a fill-in rule may introduce the feature value [+voice] for sonorant, or the feature value [−low] for [+high] segments. In some cases, however, underspecification may persist into the surface phonetic representation. In other words, unspecified features of a segment remain unspecified at the level of phonetic representation (Pierrehumbert & Beckman 1988). 'Targetless schwa' seems to be such a case. Here, the schwa may lack a target value of its own and its phonetic value may be determined by the context alone.

3.1.3 Underspecified English schwa

Underspecification seems to play an essential role in coarticulatory phenomena (Pierrehumbert & Beckman 1988; Keating 1988; Cohn 1993), and it has been shown in numerous studies that reduced vowels (e.g. schwa) are most susceptible to such coarticulatory effects (Magan 1989; Bergem 1993; Bergem 1994; Koopmans-van Beinum 1994; Bates 1995; Kondo 1995).

In English, many stressless vowels are realised as schwa. Schwa occurs only in reduced syllables with zero stress and therefore is also referred to as a reduced vowel. Vowel reduction can be classified depending on whether the reduction is obligatory or not. Obligatory phonological reduction is independent of local context, stress and rate effects. It is part of the phonology (Bolinger 1985) and non-use of a reduced vowel will be ungrammatical. An example for such reduction within the generative tradition is an underlying full vowel reduced to surface [ə] and/or [ɪ] as in (2a).

(2a) Explain [ɛkspleɪn] *explanation* [ɛkspləneʃən] (cf. *[ɛksplenеʃən]).

In derivational pairs, (2b) is another example of such reduction.

(2b) *Photo* [foʊtoʊ] *photograph* [foʊtəgɹæf] (cf. *[foʊtoʊgɹæf])

In contrast to obligatory reduction, non-obligatory reduction is conditioned by factors such as speech rate and style. Forms tends to occur in faster speaking rates or in more casual speaking styles (Dalby 1984). Reduction also tends to occur in function words and grammatical items (Bolinger 1975). As the application of the process of reduction is optional, there will be two or more variants. In the present study our main interest is in the obligatory case.

As opposed to reduced vowels, full vowels may or may not receive stress. Current studies on schwa in English, Catalan and Dutch suggest that schwa may be phonetically underspecified (Browman & Goldstain 1992a; Browman & Goldstein 1992b; Recasens 1986; Bergem 1994; Kondo 1995). In other words, schwa may lack a target value of its own and its phonetic value may be determined by the context alone. More specifically, Bergem (1994) argues, based on Dutch data, that schwa is targetless in the F2, but targeted in the F1.[3] Bates (1995) and Kondo (1995) provide further evidence in support of the targetlessness of schwa. Let us review the literature on the precise nature of the English schwa indicating its phonetic targetlessness.

---

[3] For voiced speech, we distinguish the fundamental frequency (F0), the frequency at which the vocal cords are vibrating. In a vowel sound, the air in the vocal tract vibrates at a number of different frequencies simultaneously and certain frequencies above F0 are emphasised which are called *formants*. Vowels are largely distinguished by the first two formants above F0, namely the first formant (F1) and the second formant (F2).

3.1.3.1 Centralisation, coarticulation or underspecification?

We will first review two different views on the nature of schwa to account for schwa's observed target undershoot: is vowel reduction a case of centralisation or of contextual assimilation? In either case, vowels do not reach their target value and either result in having a value less extreme (centralisation) or a value influenced by their surrounding segment (assimilation). As we shall see, the third account, the underspecification account, reconciles elements of both of the aforementioned views.

Traditionally, vowel reduction has been considered to be a **centralisation** process (Tiffany 1959; Shearme & Holmes 1962; Delattre 1969; Koopmans van Beinum 1980; Kenyon 1946; Lehiste 1970; Gimson 1980; Crystal 1985). According to this view, unstressed vowels are generally more central in quality than stressed vowels and produced with less vocal effort, resulting in less peripheral articulatory configurations and less distinctive acoustic patterns. Tiffany (1959) compared the American English vowels /i, ɪ, e, ɛ, æ, o, ʊ, u, ɑ, ʌ, aɪ, aʊ/ produced in an /h-d/ context in isolation and in carrier sentences. Each combination was produced with and without emphatic stress. Tiffany reported that vowels seem to move toward a neutral, or at least to a central, point in the vowel space in an unstressed context, so supporting the separate specification for stressed and unstressed vowels.

An alternative account which argued against the first view was that unstressed vowels were a result of **coarticulation**. The term "coarticulation" was introduced by Menzerath & Lacerda (1933). Various models have been proposed to account for the phenomena of successive sounds articulated together and underspecification has been a central issue in these models. For example, the feature-based model (Hanke 1966) views coarticulation as feature spreading, and the coproduction model (Öhman 1966, Fowler 1981; Browman & Goldstein 1990, 1992a,b) observes that invariantly specified phonetic units may overlap in time as they are coproduced. Variability in articulatory and acoustic realisation of segments occurs because the vowel tract cannot change instantaneously from one configuration to another, resulting in interactive influence in the articulatory and acoustic transitions between segments. Lindblom (1963) argued that the reduced quality observed is a result of either an increase in rate or a decrease in stress. In other words, shorter durations result in greater overlap in the timing of commands of the articulators, which in turn results in a greater degree of target undershoot. In those cases, the articulators do not have time to reach the target configuration specified by one set of motor commands before they have to start moving in response to the next set of commands. Therefore, while speakers aim to produce a full, maximally distinct vowel, they unintentionally undershoot the formant frequencies relative to the ideal spectral target. In other words, in the production of schwa, the speaker undershoots the target due to the temporal overlap of motor commands and to contextual assimilation, i.e. coarticulatory adjustment.

Lindblom's proposal that duration is the main determinant of reduction was argued against by various researchers (Verbrugge & Shankweiler 1977; Gay 1978; Tuller, Harris & Helso 1982; Engstrand 1988). Using four-syllable nonsense words such as /əpipipə/ where either one of the medial syllable was stressed, and two-syllable nonsense words such as /pipip, pɑpɑp/, with stress placed on either the first or second syllable, Tuller et al. found that an increase in stress was always accompanied by an increase in both the duration and peak amplitude of activity, while an increase in rate did not coincide with a decrease in duration or peak amplitude. Gay (1978) confirmed Tuller et al.'s findings that speech rate had less effect on the production of schwa and that unstressed vowels are considerably reduced in F0, and somewhat reduced in overall amplitude and vowel quality with respect to their fast stressed counterparts. Engstrand (1988) also found a strong influence of stress and no amount of influence of rate in the production of reduced vowels. In recognition of these studies, Lindblom revised his original rate-dependent undershoot hypothesis and included a compensatory mechanism where speakers may override duration-dependent undershoot with increased vocal effort and more forceful opening and closing gestures (Lindblom 1990). But nonetheless, he maintains his position that vowel reduction is a product of coarticulatory effect rather than its inherent centredness in the vowel space.

Yet another account proposed, which reconciles elements of both the aforementioned views, is the **underspecification account** (Keating 1988). According to the underspecification hypothesis, unstressed vowels are inherently less fully or narrowly (see the Window Model described below) specified than their stressed counterparts, which is in accordance with the traditional view. However in contrast to the traditional view and in accordance with the coarticulation account, vowel reduction is viewed as a contextual assimilation to the surrounding segments and therefore can occupy almost any position in the vowel space depending on the context. The difference between the Lindblom account and the underspecification account lies in the fact that the underspecification account assumes a separate specification for stressed and unstressed vowels in the underlying representations.

Since studies following the main interest of Lindblom's was to examine the interaction between reduction and duration, relatively little evidence of differences in the extent of local context effects across stress condition has been found. Among those few studies, it has been reported by Magen (1984) and Fowler (1981) that unstressed vowels coarticulate more strongly with the surrounding segments than stressed vowels. Fowler (1981) observed a larger V-to-V coarticulatory effect for unstressed than for stressed /ʌ/ in VbʌbV sequences. She observed contextual variation only in the F2. It should also be noted that although there was no influence of adjacent vowels on the F1 values for /ʌ/, higher F1 values were observed for the stressed tokens than the unstressed ones. Furthermore, schwa, in particular, was observed to be most susceptible to coarticulatory effects (Magen 1989; Recasens 1986, 1991; Bergem 1994).

In line with Keating's (1988) proposals, Browman & Goldstein (1992b) using articulatory data and computer simulated data, investigated the possibility that schwa may be completely unspecified for tongue position. They concluded that schwa *does* have a target (i.e. specified) for tongue position, but that it is still colourless in that its target is the mean of all the vowels, and it is completely overlapped by the following vowel. Severe criticism of their methodology and analysis was raised by Kingston (1992) and Barry (1992) in the same volume as Browman & Goldstein's paper. I will not go into details, see Bates (1995) for discussion.

Bergem (1994) reports on a study in support of the targetless analysis. He examined Dutch schwa in 897 nonsense words with 'VCəC' with penultimate stress and 'CəCV' with final stress. The context vowels were one of /i, a:, u/ and the consonants were any two of /p, t, k, f, s, χ, m, n, ɳ, r, l, j, v/. Using three Dutch native speakers, he found very high coarticulatory effects on the schwa F2 value. 82% of the variance in the F2 of the schwa was explained with second-order polynomials, using the vowel onset, midpoint and offset as reference points. In a three-way analyses of variance, between 29% and 38% of the variance in schwa F1 value and between 72% and 79% of the variance in F2 was explained by context. Based on these results, he concluded that schwa is completely targetless and is completely assimilated with its phonemic context. In line with Browman & Goldstein (1992b), Magen (1989) reported that the F2 value moved roughly continuously from a value dominated by the preceding vowel (at onset) to one dominated by the following vowel (at offset) in formant trajectories for medial schwa in tri-syllabic sequences.

Although Browman & Goldstein (1992b) and Bergem (1994) reached different conclusions with regard to the targetlessness of schwa, it seems valid to at least conclude that both sets of data found little evidence of schwa's targetedness. And if we assume that specification can be gradient, as in the Window Model introduced below, the discrepancy seems to be explained.

3.1.3.2 The Window Model and underspecified English schwa

In her 1990 study, Keating suggested a model which incorporates underspecification into a model of coarticulation. Under this model, the nature of phonetic underspecification is considered gradient, rather than being a discrete contrast of 'specified' and 'unspecified'. According to the model, the feature value of a segment such as backness or height is associated with a range of possible articulatory or acoustic values and is characterised by 'windows' which represent the full range of contextual variability they exhibit along given articulatory dimensions. Windows are determined empirically by finding the maximum and minimum value possible for a given dimension across all contexts. It is an undifferentiated range representing the contextual variability of a feature value. For some segments this window is very narrow, reflecting little contextual variation, while for others it is very wide, reflecting large contextual variation. A segment is less specified when its window is wide and more specified when its window is narrow. Once the window widths are determined they are not themselves contextually varied: a feature value has one and only one window that characterises all contexts taken together and it does not have different windows for different contexts. Depending on the sequence of windows, the path through any given window may span the entire window width or a more limited range of values within the window. Both window width and the relative position of the windows within the specified dimension serve to determine the path through them.

Keating assumes that the path is constrained by the requirement of contour continuity and smoothness, and of minimal articulatory effort, along the lines of minimal displacements or minimal peak velocities. Therefore, depending on the particular context, a path though a segment might pass through the entire range of values in the window or span only a more limited range within the window. Figure 3.1. schematically illustrates how the model is to work (from Keating 1990:451-470). (a) illustrates a case of a segment with a narrow window between two identical segments differing from the middle one. The middle segment imposes strong constraints on the interpolation. It shows little variation across contexts and affects the interpolation in the adjacent segments. In the case of (b), a segment with a wide window between the same identical segments. The wide window segment assimilates its turning point to the context, and in some contexts, it will show no turning point at all. (c) and (d) illustrate wide and narrow windows between unlike segments. Here, the wide window allows straight interpolation between many different segment types. The narrow window more often makes its own contribution to the curve. Keating notes that the contours in the figure just described are not the only possible interpolation through these sequences. This is because even minimal curves can be moved higher or lower in sequences of wide windows. Therefore, the prediction of the model is that speakers, and repetitions of a single speaker, should vary in their trajectories through window sequences which underdetermine the interpolation (Keating 1991:455-458).

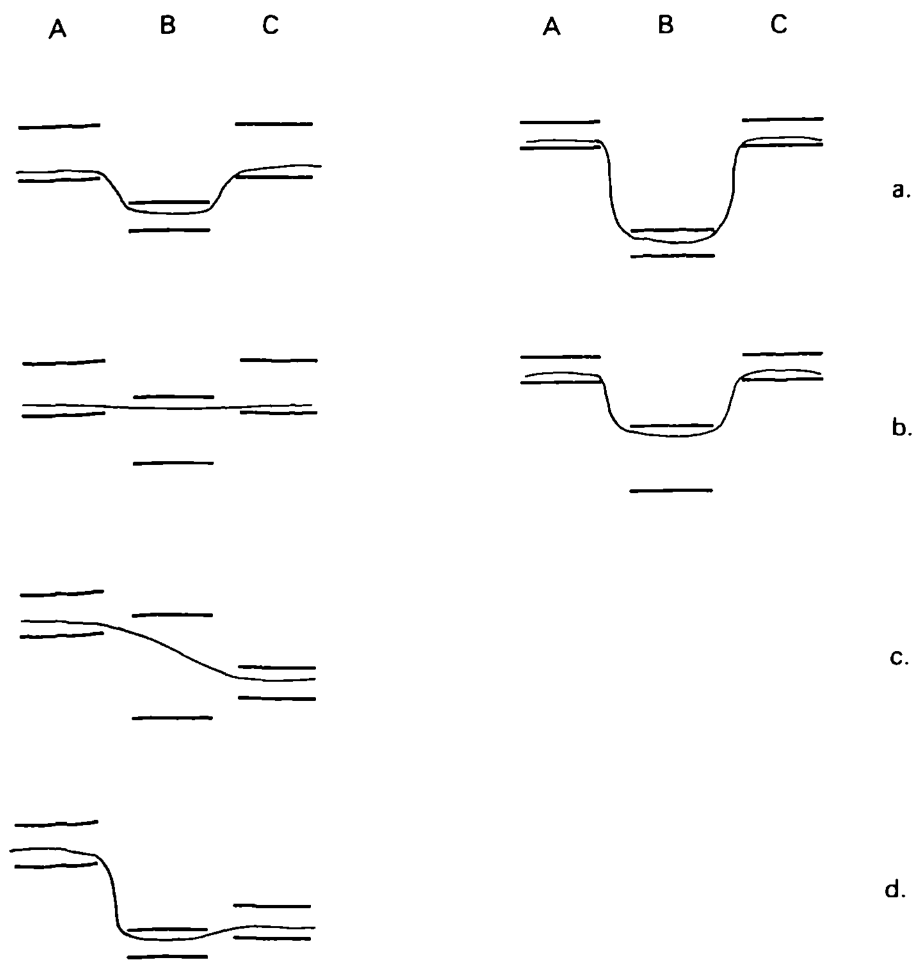

Figure 3.1 Illustration of sequences of windows of various widths

The model allows for intermediate degrees of specification and implies that there need not be a discrete alternation between schwa and full vowels. As Bates (1995) states, the Window Model framework accommodates inherent differences between vowels with respect to the degree of contextual variability they exhibit.

Bates (1995) and Kondo (1995) found more definite evidence to support the targetlessness of schwa in their experiments, which used real words instead of nonsense words, as had been used in both Browman & Goldstein (1992b) and Bergem (1994) study.

Bates investigated the relative magnitude of coarticulatory effects on schwa compared with full vowels. More specifically, she examined whether schwa can be associated with an independent phonetic target. She hypothesised that if schwa were targetless and values were determined according to phonemic context, its formant trajectories would be predictable: it should make no independent contribution to the trajectory between adjacent context segments, but is expected to be interpolated through. Her data comes from one Southern British Standard English male speaker. The results of the regression analysis provide strong support that schwa is underspecified for tongue position: 92% of explained F2 variance for schwa tokens were in the pooled data set. The linearity of the majority of schwa F2 mean trajectories and the wide spread in schwa F2 midpoint values also indicated interpolation through. However, the overall range in F2 midpoint values is subject to some constraint: the F2 mean midpoint values for schwa did not extend into the extreme upper or lower regions in F2 space (Bates 1995:112-113). It has also been shown by Bates that when compared with a series of full vowels, schwa is most susceptible to coarticulation. When it comes to F1, however, the total proportion of explained F1 variance for schwa was found to be considerably lower than the total proportion observed in F2. Interpretation given by Bates for this difference is as follows:

> One possible interpretation of this difference in results, is that lower context-dependency for F1 is reflective of an F1 target. That schwa should be specified for jaw position, albeit weakly, is to be expected given that some degree of jaw opening is inherent to its characterisation as a vowel (Bates 1995:115).

Given that the F1 variance still represents a high level of context dependency (74%), she also notes that "schwa is also highly unspecified along this formant dimension" (p117).

As was the case in F2, the full range in F1 values for schwa did not cover the maximum frequency range possible for F1 across all vowels, i.e. there were comparatively fewer tokens characterised by very high F1 values. Bates speculates that the possible reason for the lower variance in schwa in Bergem's (1994) data might be that Bergem's data came from nonsense words, which might have induced more careful production than more natural speech data (Koopmans-van Beinum 1980; Krull

1989; Lindblom & Moon 1988) because more careful production is often observed to have longer duration and less coarticulation between segments (Engstrand 1988). For more detailed discussion comparing Browman & Goldstein's and Bates' studies, see Bates (1995, Chapter 5.4.1.3.). To sum up Bates' findings, she found a linearity of schwa second formant trajectories (i.e. the wide range in schwa F2 midpoint values) and high (near-maximal) context-dependency. These indicate the phonetic transparency of schwa and interpolation through, and which strongly supports the hypothesis that schwa is underspecified for F2. Bates supports her claim by citing Choi (1995), who investigated short vowels in Marshallese. Choi reported that the short vowels in Marshallese are fully specified for tongue height but are unspecified for front/back tongue position. The acoustic data showed robust consonant-to-consonant coarticulation for these vowels, providing further evidence that the vowels lack an F2 target. In Bates' study, the data also show that schwa displays a high degree of differentiation in F2 midpoint value as a function of following consonant place of articulation in contrast to the other full vowels. As for the F1 value of schwa, Bates found little evidence showing a unique target, even conceding that some degree of jaw opening is inherent to schwa's [-consonantal, +syllabic] specification. The relative high F1 context-dependency (compared to full vowels) and the linear F1 trajectories with its contexts indicate that schwa is also highly unspecified for jaw position.

To sum up, to account for the targetlessness of schwa, in line with Keating's phonetic underspecification with reference to her Window Model, it has been argued that unstressed vowels are inherently less fully or narrowly specified than their stressed counterparts. This view is in accordance with the traditional centralisation view in respect of the inherent neutralness schwa, and also takes Lindblom's analysis into account in that schwa is unspecified for formant trajectories and can occupy any position in the vowel space depending on the context (Bates 1995, p193).

High variance in schwa midpoint in F2 (but not F1) has been also observed in Kondo's (1995) study, where she examined three British speakers' production of schwa. As she hypothesises, if schwa was targetless, it should be transparent to its surrounding sounds and its value would be determined by context alone: its articulatory and acoustic trajectory would move continuously from the preceding segment to the following without going through any target of its own (i.e. completely interpolated through; the window is said to be *wide* in the Window Model). For example, Magen (1989) reported that in trisyllabic sequences with a medial schwa [bVbəbVb], F2 moved roughly continuously from the value dominated by the preceding vowel to that dominated by the following vowel across intervocalic consonants and through the schwa.

In Kondo's doctoral thesis (1995), she investigated the acoustic variability of schwa to support the hypothesis that realisation of schwa is a function of context. She used a VCəCV sequence to investigate the effect of C-to-V as well as V-to-V coarticulation of schwa Consonants used in the test words were /p, t, k/ and the vowels were /ɪ, æ, u/. All the contexts were symmetrical, i.e. V1 and V2 were identical, as well as C1 and C2. Examples of the test sequences were: ɪpəpɪ, æpəpæ, upəpu. Each sentence was repeated 10 times by three native speakers of British English (AH, MB and DG) and schwa in function words was also studied. An example of a test sentence is: Please dip **a** pin in the solution. The hypothesis tested was whether schwa is phonetically unspecified with no target value of its own, in other words, whether formant trajectories of schwa were interpolated through the preceding segment to the following segment. The F1 and F2 values of schwa were measured and their trajectories across time (the onset, midpoint, and offset of the schwa) were measured.

The main effect of consonantal contexts was significant though small for F1 for schwa. No significant effects of the vocalic context were observed in F1 (see also Fowler (1981) who reported similar results) and the standard deviations for schwa were relatively small for F1. In this experiment, Kondo could not conclude whether schwa was targeted or targetless in F1, given that F1 trajectories do not stay level but do not seem to be aiming at a common target either (Fowler 1982:66).

On the other hand, the effect of both consonantal and vocalic context were found to be highly significant in F2 for all three speakers. The effects of consonantal context were greater than those of vocalic contexts in magnitude. A large variability was observed not only during the transitional stage but also throughout the production of schwa. The extent of variation of F2 in schwa was very large: the smallest to largest value observed for each speaker was 1249-2066 (817) Hz, 1133-2171 (1038) Hz and 1002-2101 (1099) Hz for AH, MB, and DG respectively. For the diversity of F2 trajectories as a function of context in their levelness, Kondo concludes that schwa may be targetless in F2, i.e. F2 values can be determined by context alone. It should be added that the contextual effects were observed right though the schwa and even at the vowel midpoint; the range of mean F2 values as a function of context remained about 700 Hz for schwa. Kondo asserts that this illustrates the contrast between a specified and an unspecified vowel. A clear comparison can be made with Bates (1995) where she studied formant trajectories of full vowels exhibiting divergence towards different contexts at the transition, while at the midpoint, unlike schwas in Kondo's study, they cluster into a narrower band. Kondo further asserts that schwa seems to retain a minimally open tongue height (which is correlated with the F1) in order to retain its vocalic identity. In this sense, schwa may be considered to be specified for [height] while underspecified for [backness], which is correlated with the F2. In other words, schwa may be specified as [+vocoid] by retaining vocalic height while it is unspecified for place of articulation. Kondo (1995) supports her argument by citing a language in which segments are specified only in [height]. Kabardian, a Circassian language spoken in the Northeast Caucasus, has a vertical three-vowel system, namely, /ɨ, ə, a:/ (high, mid, low). These vowels are only specified for [height] and they are redundantly distinguished by three-way durational contrast, i.e. /ɨ/ is the shortest, /ə/ is second longest and /a:/ the longest. In F2, these vowels vary in quality as a function of their adjacent consonant to a great extent, i.e. they range over 1000 Hz, and it is reported that to a casual observer it would appear as if the language had a rich range of vowel qualities (Ladefoged & Maddieson 1990; Choi 1990). Another example Kondo cites is the English /s/, which may be specified for [height] but unspecified for [backness] (see Keating 1988: 285-287).

In sum, Kondo concluded that schwa is targetless in F2, in other words, schwa may be unspecified in [backness]. As for F1, both the contextual effect was small and there was low variability in F1. The F1 trajectories did not seem to converge to a common target value either, which led Kondo to suggest that schwa seems to be specified as [+syll –cons] as suggested by Durand (1990), but unspecified for place of articulation, i.e. targetless in the F1 trajectories.

3.1.4 Variability in English vowels

It has been discussed that schwa is phonetically and phonologically underspecified and lacks target in its formant trajectories. In this sub-section, it will be discussed that the transparency of schwa in English has important implications for coarticulatory strategies in the language. Additionally, it will be shown that in Japanese no such alternation between full and reduced vowels is observed and that all vowels are basically targeted in Japanese.[4]

So called stress-timed languages are characterised by phonological vowel reduction. Since the phonetic underspecification of schwa presupposes a contrast of targeted and targetless vowels, the contrast of full and reduced vowels may then be described as a contrast of targeted and targetless vowels.

In a study of VCV coarticulation, Fowler (1981) found that native speakers of English manifest a sharp contrast in the extent of vowel variability between full and reduced vowels as a function of context. She examined two lists of trisyllables of the form $V_1bV_2bV_3$. The initial and final vowels were /i/ /a/ or /u/ and the medial vowel was the reduced vowel /ʌ/. All combinations of initial and final vowels were included, i.e. there were nine different utterances in each list. The two lists differed in stress pattern. In list one, V_1 and V_3 was stressed (/ʌ/ was unstressed) while in list two, the stress pattern was reversed and the medial vowel /ʌ/ was stressed. Five American English speakers participated in the study and were asked to read both lists of trisyllables five times. The coarticulatory effects were assessed separately for F1 and F2 by two three-way analyses of variance, both with the following factors: Initial Flanking Vowel (i, a, u) and Final Flanking Vowel (i, a, u) and Stress of Medial Vowel.

[4] The existence of an unspecified vowel seems to distinguish so called stress-timed languages from syllable-timed (e.g. Italian) and mora-timed (e.g. Japanese) languages. In English, the prosodic unit stress foot is associated with the alternations between reduced and unreduced vowels. It has been shown that accent plays an important role in organising speech into the prosodic unit (Abercrombie 1964; Selkirk 1980; Beckman 1986). Furthermore, it has been shown that accent plays a minimal role in the prosodic organisation of a mora-timed language such as Japanese (Pierrehumbert & Beckman 1988; Kubozono 1989). This issue will be discussed in more detail

As for F1, neither the initial flanking vowel (carryover effect) nor the final flanking vowel (anticipatory effect) had reliable effects on medial /ʌ/ whether it was stressed or not. On the other hand, Fowler found statistically significant coarticulatory effects of a stressed vowel on a preceding or following unstressed transconsonantal vowel. It should be noted that for both stressed and unstressed medial /ʌ/, carryover effects on F2 were larger than anticipatory effects, i.e. a preceding stressed vowel exerts a more powerful influence than a following stressed vowel. Considering the fact that the coarticulatory effects were not blocked by the intervening consonants, the data indicates the transparency of consonants to the V-V effects, though the transparency of consonant to vocalic effects seems to vary from consonant to consonant. Fowler further suggested that English speakers may organise the intervals between the stressed syllables within a phrase as a basic unit of speech and this interval may possibly be the domain of coarticulation in English. In other words, this feature of vowel contrast between full and reduced plays an important role in the coarticulatory pattern of English. Therefore, the native speaker of English will for example manifest a sharp contrast in the extent of vowel variability between full and reduced vowels as a function of context. That is, while on one hand there are stressed vowels that are prominent and resistant to coarticulatory effects, there may on the other be a completely targetless vowel that is a product of coarticulation. This implies that

---

in section 3.2 on pitch-accent in English and Japanese.

English speakers need to aim only at targeted, full vowels and leave the rest of the vowels to contextual assimilation, an economical strategy for speech production.

Öhman (1966) reported similar V-to-V coarticulatory relations among non-adjacent segments, in particular among vowels that span a consonant, and also found that consonants are transparent to vocalic gestures. The formant transitions in vowel + voiced stop consonant + vowel utterances spoken by Swedish, American and Russian speakers were studied spectrographically. I will report only the methodology used for American English. A list of VCV 'words' using two voiced stops /d/ and /g/ and the vowels /i/, /a/ and /u/ were read five times by a Standard American English speaker. All possible combinations were included in the list. No examples of test words were given in Öhman's paper. Results from this American English experiment showed similar coarticulatory effects with the two Swedish experiments. He observed the effect of the first vowel at the onset of the second vowel and the effects of the second vowel at the offset of the first vowel. He also reported that the F2 trajectory seemed to have a single continuous movement from the first vowel to the second across the intervening consonant. He concluded that the consonant articulation was superimposed on the V-V sequence, which is evidence that consonants are transparent to vocalic gestures.

Kondo (1995) carried out a study to investigate whether native speakers of English manifest a sharp contrast in the extent of vowel variability in F2 between reduced vowels as a function of context. In line with Fowler's (1981) hypothesis, she predicted that the effect of the first stressed vowel may be observed at the onset of the second stressed vowel and vice versa in VCəCV utterances given that consonants and schwa are transparent to vocalic gestures. In other words, the contextual effects may pass through the schwa and the two consonants from the preceding stressed vowel to the following stressed vowel. Vb_bV sequences were used with /ə/ and the full vowel /æ/ as the middle vowel. For the contextual vowels, the vowels /ɪ/ and /ə/ were used. Therefore, the following eight sequences were tested:

(3) VCəCV: ɪbəbɪ, ɪbəbæ, æbəbɪ, æbəbæ

(4) VCæCV: ɪbæbɪ, ɪbæbə, æbæbɪ, æbæbə

Test sentences were such as:

(5) The campaign for Women's Li**b aby**smally failed. (ɪbəbɪ)

(6) We found the baby's cri**b aba**ndoned in the car park. (ɪbəbæ)

Eight male native speakers of British English participated in the experiment. She only studied the F2 values since schwa was observed to be targetless only in F2. Her data showed the transparency of schwa compared to that of the full vowel /æ/. As predicted, for schwa the effect of contextual vowels was observed right through in both directions; while for /æ/, the effect of the preceding vowel was observed only at the onset and midpoint, and the effect of the following vowel was observed only at the offset of the vowel /æ/.

Results of the experiment suggest a clear difference in the behaviour of schwa and the full vowel /æ/ in V-to-V effects. The V-to-V effect was observed right through schwa, suggesting schwa's transparency. This supports Fowler's observation that the rhythm of English is determined by the basic speech production unit in English of stressed to stressed vowels. Furthermore, as Kondo asserts, schwa's absolute transparency guarantees the coproduction of segments across a longer stretch of time, compared to a syllable-timed language in which only a few consonants may intervene between the vowels. By extension, Kondo (1995) notes that there might be yet other grounds for the targetedness of F1 in English schwa. As discussed earlier, she speculates that schwa seems to retain a minimally open tongue height (which is correlated with F1) in order to retain its vocalic identity. Here, she further argues that "it may be important that schwa retains its vocalic identities that it could still be a syllable nucleus because the existence of these weak vowels seems to essentially differentiate so call stress-timed languages from syllable-timed languages"(Kondo 1995:92).

In sum, Kondo (1995) found clear evidence that schwa is contextually dependent in its F2 trajectories and the contrast of targetedness between full and reduced vowels seems to be an important feature of the stress-timed language English.

3.1.5 Variability in Japanese vowels

Unlike in English, in a non-stress accented language such as Japanese, with no such contrast between reduced and full vowels observed, all vowels are presumably targeted. This seems to correlate with the fact that Japanese accent plays a minimal role in the prosodic organisation of the language, i.e. there is a lack of accent-dependent vowel variation. Interestingly, however, as will be discussed in more detail later, Japanese vowels in general have been found to coarticulate more strongly with contextual vowels than English *full* vowels (Magen 1984; Keating & Huffman 1984). This suggests that contextual variability is spread more or less evenly over the whole vowel in Japanese to compensate for the lack of unspecified vowels.

3.1.5.1 The nature of Japanese vowels and other related issues

The Japanese vowel inventory consists of five distinctive vowel qualities, represented as /ɪ, e, a, o, u/, which are phonetically realised as monophthongs.

Vowel length is phonologically distinctive, with contrasts between short (one-mora) and long (two-mora) vowels. For example, see (7a) and (7b).

(7a) Short vowel /a/ e.g. obasan 'aunt'

Long vowel /aa/ e.g. ob**aa**san 'old lady'

(7b) Short vowel /i/ e.g. **i** 'stomach'

Long vowel /ii/ e.g. **ii** 'good'

The five Japanese vowels form a skewed pentagon in the F1/F2 acoustic vowel space due to /u/, the high back vowel being unrounded [ɯ]. The vowel qualities of short and long vowels are described as being identical or very similar (Shibatani 1990), as opposed to the English 'tense' and 'lax' pair of, e.g. /ɪ/ or /i:/.

Secondary articulation of palatalisation is observed for the back vowels /a, o, u/ following the consonants /p, b, t, d, k, g, s, z, m, n, r, h/. For example, *gyaku* 'other way round' (cf. *gaku* 'frame') and is distinctive.

Allophonic vowel devoicing phenomena is also observed (Vance 1987), as shown in (8) and (9).

(8) *aki* [aki̥] 'autumn'

(9) *kusa* [ku̥sa] 'grass'.

The extent of vowel devoicing is held to vary from dialect to dialect. For example, the phenomenon is quite wide-spread in the Tokyo dialect, while in the Kansai area (e.g. Kyoto, Osaka), it is not common (Hirayama 1985). The vowels that most frequently undergo devoicing are the high vowels /i/ and /u/ when these high vowels are surrounded by voiceless consonants immediately preceding and following the vowel, or the word ends with the high vowel itself. However, devoicing has on occasion been observed in other vowels mostly in casual speech e.g. *kokoro* [ko̥koro]. According to Jun & Beckman (1993), vowel devoicing is a result of gestural overlap, i.e. a devoiced vowel seems to be subsumed under the strong friction of the preceding voiceless obstruent.

There is a traditional assumption that the mora is a timing unit in Japanese and that each mora is supposed to bear the same length. A mora can be realised as any following shapes: (1) V (e.g. /**i**/ 'stomach') (2) CV (e.g. /**ki**/ 'tree') (3) First part of a geminate (e.g. i**k**ki 'in one stroke') (4) 'moraic' nasal /n/ (e.g. kika**n** 'period'). The issue of mora and mora-timing will be further discussed section 3.2.2.3.

3.1.5.2 Variations in Japanese vowels

As noted above, Japanese vowels in general have been found to coarticulate more strongly with contextual vowels than English *full* vowels (Magen 1984; Keating & Huffman 1984).

Magen (1984) reported in her pilot study on Japanese VCV utterances that the magnitude of coarticulation in VbV sequences with the vowels /i/ and /a/ was observed to be greater in Japanese than in English She argued that contextual variability is spread more or less evenly over all vowels in Japanese to compensate for the lack of an unspecified vowel compared to English. The data included two possible stress patterns in English (V́bV and VbV́) and two pitch patterns in Japanese (high pitch + low pitch [HL]and low pitch + high pitch [LH]). In English, the V-to-V coarticulatory effects on F2 were stronger on unstressed vowels than on stressed vowels. When the unstressed vowels in the English and Japanese words were compared with regard to the effect of the coarticulatory effects on vowels, the effect was stronger on Japanese vowels. The Japanese data showed that V-to-V effects were observed through the duration of the vowel, while in English, the carryover effect tended to diminish towards the end of the vowel and the anticipatory effects were not yet clear at the beginning of the vowel.

Magen's suggestion that contextual variability is spread more or less evenly over all the vowels in Japanese to compensate for the lack of an unspecified vowel is in accordance with Manuel & Krakow's (1984) proposal. They argue that the extent of vowel variation observed in a language is constrained by the number and distribution of vowels in the language: vowels are less variable when they are more crowded in the acoustic vowel space in order to avoid overlap and maintain their identities/distinctiveness. For example, a language with a standard five vowel system (e.g. Swahili, Shona, and also Japanese) is more susceptible to variability than a language which is relatively rich in the number of vowels (e.g. English, German). In this regard, English full vowels need to be hyper-articulated compared to vowels in languages where a full-reduced contrast is not observed. Manuel & Krakow compared V-to-V coarticulation in Swahili, English and Shona. Swahili has a typical five vowel system /i, e, a, o, u/. A male Swahili speaker was asked to produce each of all possible vowel combinations in penultimate stressed VpV and VtV disyllables. The test words were embedded in a carrier sentence 'Nili pata VCV jana' (I received VCV yesterday). The speaker produced each token five times and the F1 and F2 for the vowels were obtained by means of Linear Predictive Coding (LPC) analysis. The values of F1 and F2 in the centre of the longest stretch of minimally varying F1 and F2 values were recorded. Manuel & Krakow found a highly significant and systematic effect of vocalic environment. It was found that anticipatory effects were large and they were statistically significant in both the F1 and F2 dimensions; carryover coarticulation was significant only for the F2 dimension. Overall, anticipatory coarticulation exceeded carryover coarticulation. Therefore, we can say that anticipatory V-to-V coarticulation in Swahili is free enough to extend into the steady portion of each of the vowels. A similar experiment was carried out with Shona, another five-vowel Bantu language, and it was found that Shona patterns like Swahili with respect to magnitude of coarticulation: F1 and F2 both show significant effects of coarticulation and anticipatory effects exceed carryover effects as was the case in Swahili. They also looked at English, which has a relatively rich number of vowels. As in the Swahili data study, English VpV syllables were stressed on the first vowel and embedded in a carrier sentence. The vowels they analysed were /i, e, a, o/ and the contextual vowels were /i, e, a, o, u/. Results of a LPC analysis of the speech produced by a male English speaker show that there were some coarticulatory effects, but these effects were not as large as those found in Swahili. Furthermore, unlike Swahili, the effect was restricted to the F2 dimension. They argued that movement is minimally restricted in the F2 dimension since relatively few vowels occupy the same horizontal plane (Manuel & Krakow 1984:77). Moreover, unlike Swahili, carryover effects of coarticulation were significantly greater than anticipatory effects. They speculated that directionality of V-to-V coarticulation might be a language-particular phenomenon.

In summary, comparative analysis of V-to-V coarticulation in Swahili, Shona and English supports the hypothesis that in general, languages with fewer vowels vary more as a function of vocalic context than languages which have larger vowel inventories.

In line with Magen (1984) and Keating & Huffman's (1984) study, Kondo (1995) also looked at V-to-V coarticulation in Japanese. She addressed the following three issues in her experiment: 1) the magnitude of V-to-V coarticulation in Japanese; 2) how far V-to-V coarticulation extends in Japanese; 3) in which direction V-to-V coarticulation in Japanese is stronger (carryover or anticipatory). I will focus on the results of the first two issues, which are relevant to this study.

Five Japanese native speakers participated in the experiment. VbabV sequences were embedded in sentences with contextual vowels of /i/ and /a/. Both symmetric and asymmetric contexts were included, using sequences such as: ibabi, ibaba, ababi, ibaba. An example of a carrier sentence is shown in (10).

(10) Ibabi: Mukashi **Babi**ronia-to-iu kuni-ga arima'shita.

'Once there was a nation named Babylonia'

Speakers' production was pooled and three–way ANOVAs were performed with F2 values of the Japanese vowel /a/ at the onset, midpoint and offset of the segment as dependent variables. The independent variables were the vowel preceding, the vowel following, and the speaker. The main effects were found to be significant for all three independent variables. The Japanese data were compared with English vowels /æ/ and /ə/ (data were from the experiment carried out with a comparable Vb_bV context).

In general, strong V-to-V effects were observed for the Japanese vowel /a/. In the symmetric Vb_bV context, the extent of V-to-V effects observed on the Japanese /a/ was intermediate in degree between that of the English /ə/ and the English full vowel /æ/, as predicted.

3.1.6 L2A and underspecified (targetless) schwa

One of the earliest systematic approaches to the acquisition of a second language undertaken within the contrastive analysis (CA) framework addressed the primary question, how is the L1 influenced by the acquisition of the L2 (Lado 1957; Weinreich 1953; Lehn & Slager 1959; Stockwell & Bowen 1965; Stockwell, Bowen & Martin 1965, among others). Lado assumed *transfer* of properties in the first language to the target language in second language acquisition and claimed that "those elements that are similar to his native language will be simple for him, and those elements that are different will be difficult" (1957:2). This approach, however, was unable to account for all aspects of the observed acquisition data: while CA correctly predicted some errors (see, for example, Dušková 1969; Chamot 1978), it sometimes under-predicted errors (e.g. Hyltenstam 1977), over-predicted errors in other cases, and some predicted errors did not materialise at all (e.g. Dulay & Burt 1974).

These disappointing findings led to the suggestion by Wardhaugh (1970), of a weak version of the CA hypothesis which involved examination of the observed learners' errors and in order determine the source of errors by taking into account the differences between the two languages in addition to universal preference and developmental factors. Since then, research has been focused on the question of under what conditions transfer is likely to occur, e.g. Eckman (1977) claimed that typological markedness plays a role in whether L1 properties are transferred to L2.

The production of schwa by Japanese speakers of English implies the need for the acquisition of a new coarticulatory strategy: Japanese speakers of English will have to learn to produce targetless schwa and thus make a contrast of targeted and targetless vowels. This involves a shift from the L1 coarticulatory patterns to the L2 coarticulatory patterns of the stress accent language.

Kondo (1995) compared schwa production by Japanese learners of English with that of native speakers of English. The interlanguage data comprised six Japanese learners of English, grouped into two learner groups by the investigator's impressionistic judgement: fluent non-native speakers (KN, MN, MT) and non-fluent non-native speakers (TT, HK, HS). This was backed up by an experienced phonetician who rated learners' fluency on a scale of 1 to 10. Their interlanguage production was also compared with native English speakers' English and native Japanese speakers' Japanese to look into whether non-native speakers exhibited any sort of 'merged' system, i.e. an intermediate interlanguage stage (see e.g. Flege & Hillenbrand 1984).

In her experiment, the vowel /a/ in Japanese was studied as a possible candidate for transfer in the production of schwa because this vowel was most similar to /ə/ auditorily among the five vowels of Japanese. The sequences used in the experiment were a VCəCV sequence. The contextual consonants were /p, t, k/ and the vowels were /ɪ, æ, u/. All the contexts were symmetrical, i.e. V1 and V2 were identical, as well as C1 and C2. Examples of the test sequences are : ɪpəpɪ, æpəpæ, upəpu. An example of a test sentence is shown in (11).

(11) Please dip **a** pin in the solution.

Before I report the interlanguage data from Kondo's interlanguage study (1995), I will refer to two studies summarised earlier in the sub-section when V-to-V coarticulation in English and Japanese was discussed. The data were obtained from two experiments, one collected from English native speakers and Japanese native speakers. The data obtained from these subjects served as reference data in Kondo's interlanguage study. I will first summarise these data.

For English native speakers' data, the extent of variation of F2 in schwa was found to be very large: the smallest to the largest values observed for each speaker were 1249-2066 (817) Hz, 1133-2171 (1038) Hz and 1002-2101 (1099) Hz for AH, MB, and DG respectively. For the diversity of F2 trajectories as a function of context in their levelness, Kondo concluded that schwa may be targetless in F2, i.e. F2 values can be determined by context alone.

As for control data for Japanese vowel production, three male speakers (MN, SO and KO) participated in the data collection. The vowel /a/ was selected as the target vowel for the contextual assimilation. The VCaCV sequences with the consonantal contexts of /p, t, k/ and the vocalic contexts of /ɪ, a, u/ were embedded in a sentence such as:

(12) kawa-ni **pa'pi**rus-no ha-ga shige't-te-iru

'Papyrus reeds are growing in the river.'

Other sequences included in the experiment were: apapa, upapa, upapu, itatʃi, atata, utatsu, ikaki, akaka, ukaku. Each sentence was repeated 10 times by each speaker in a randomised order. The extent of variability as a function of context was quite remarkable in both F1 and F2 for the vowel /a/ of Japanese. Although the F2 trajectories of the Japanese vowel /a/ do not range as widely as those of English /ə/, these trajectories are quite diverse even at the midpoint. As for F1, the trajectories cover a wide range at the midpoint for SO and KO, while for MN and SK the trajectories from the onset to the offset seem to pass through some sort of target range. Figure 3.2 shows a scatterplot of schwa tokens in the 9 VC_CV contexts for the three native speakers of English and Figure 3.3 shows those of non-native speakers' interlanguage data. Symbol * indicates an overlap between tokens belonging to different speakers. As for the native speakers' data, the spread is observed mainly in F2 values. On the other hand, the variation is observed in two different directions for non-native speakers' data: one in F1 and the other in F2.

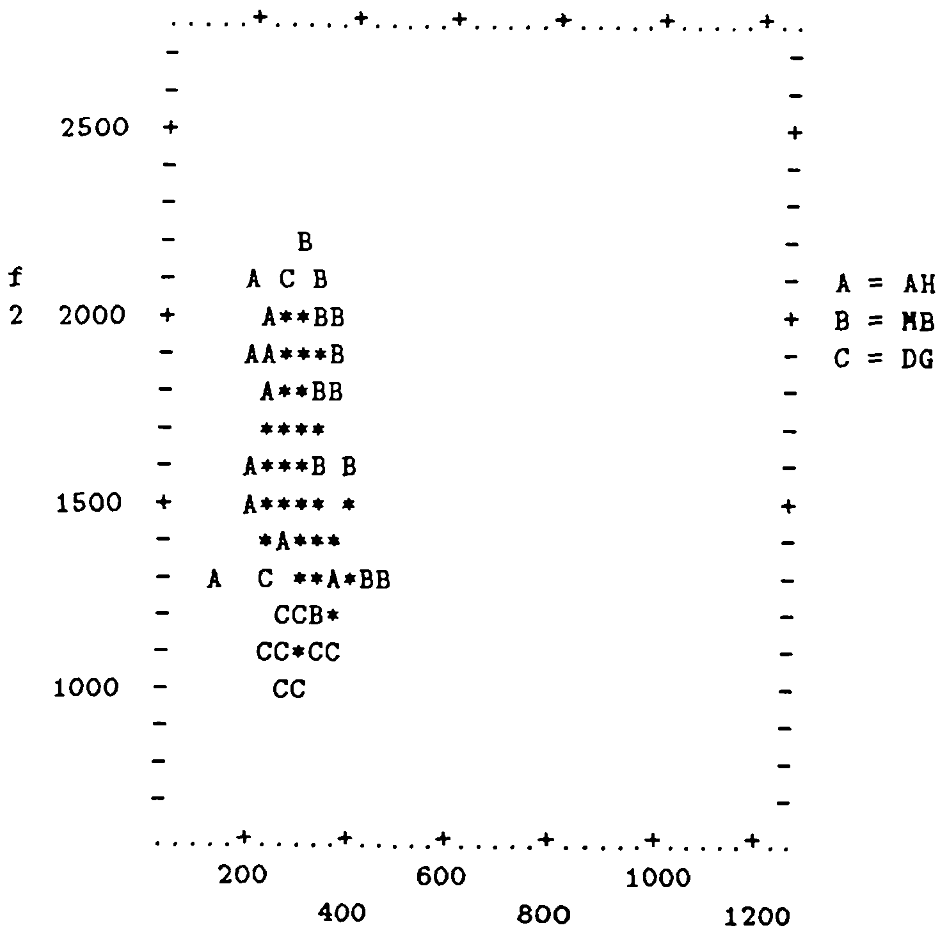

Figure 3.2 Native speakers' schwa

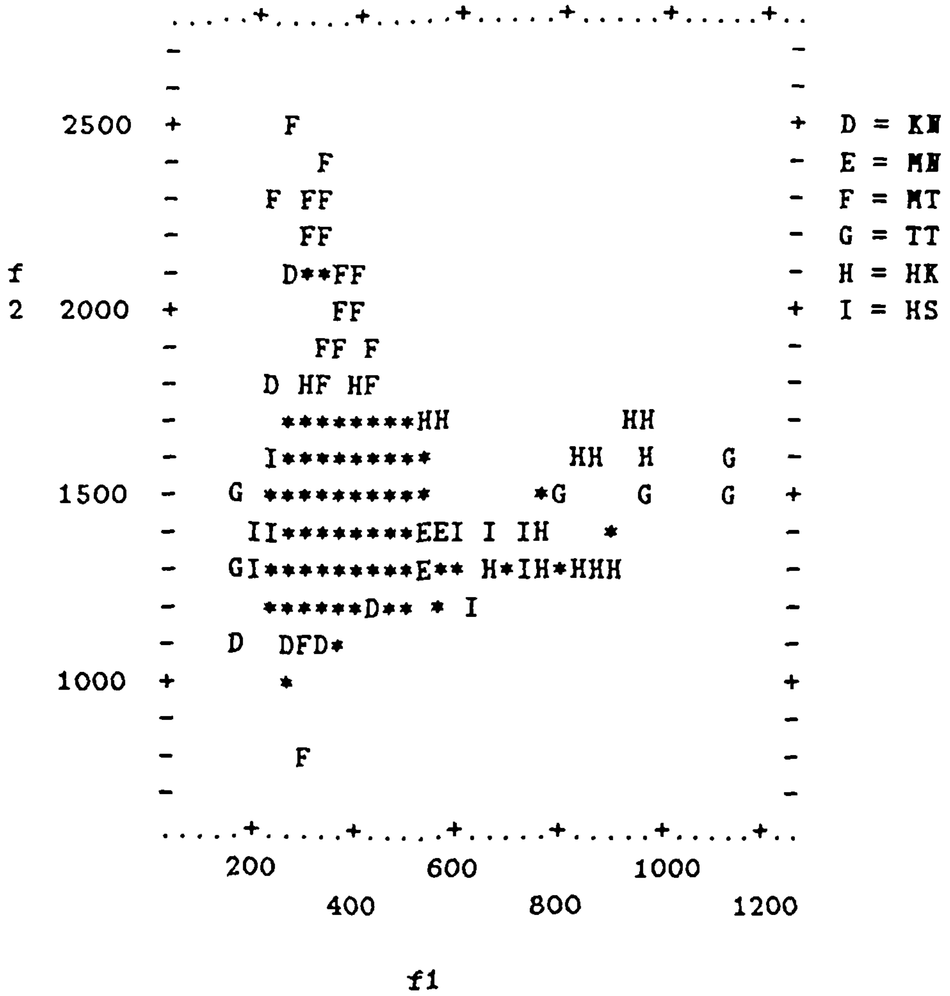

Figure 3.3 Non-native speakers' schwa

The investigator asserts that these tokens may be broken down into two groups: one group encompassing the production of KN, MN and MT and the other group encompassing the production of TT, HK and HS, which correspond to the investigator's impressionistic judgement of fluency among these non-native speakers. Fluent speakers show a spread of schwa in F2 values, whereas non-fluent speakers show a spread in F1 values as shown in Figure 3.4 (f = non-native fluent, n = non-fluent non-native, ∗ = an overlap between tokens belonging to different groups).

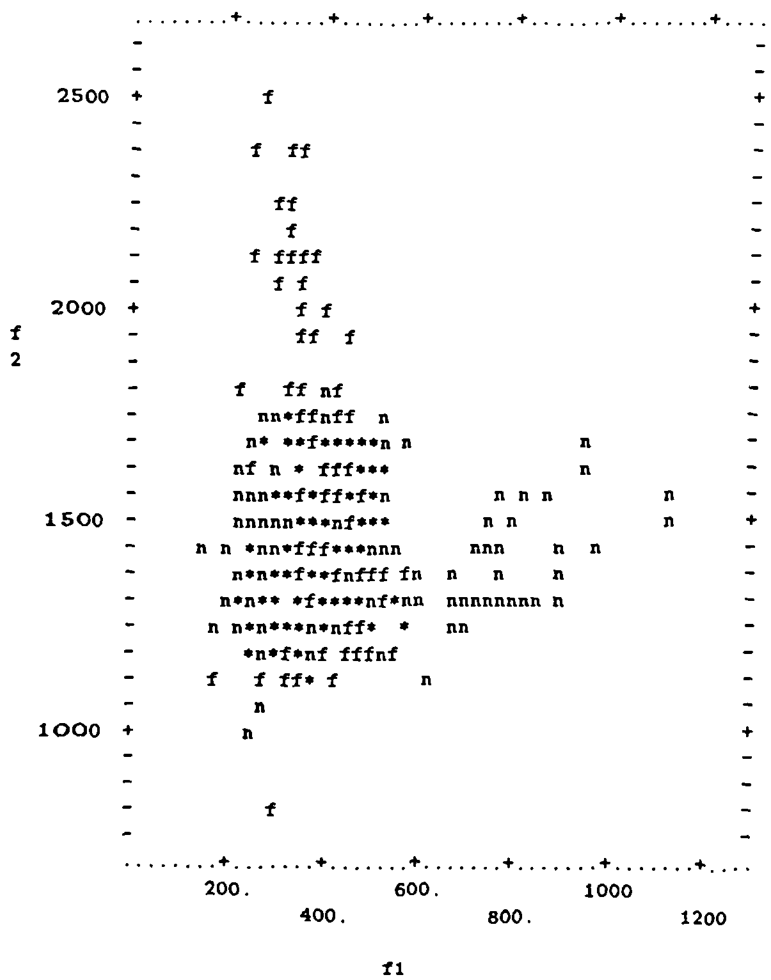

Figure 3.4 Schwa produced by fluent and non-fluent non-native speakers

There is a similarity between the native and fluent non-native speakers when the schwa token produced by the two groups are plotted on the same scatterplot (Figure 3.5): the spread in F2 values as a function of context is similar between the native and fluent non-native speakers (e = native, j= fluent non-native, * = an overlap between tokens belonging to different groups).

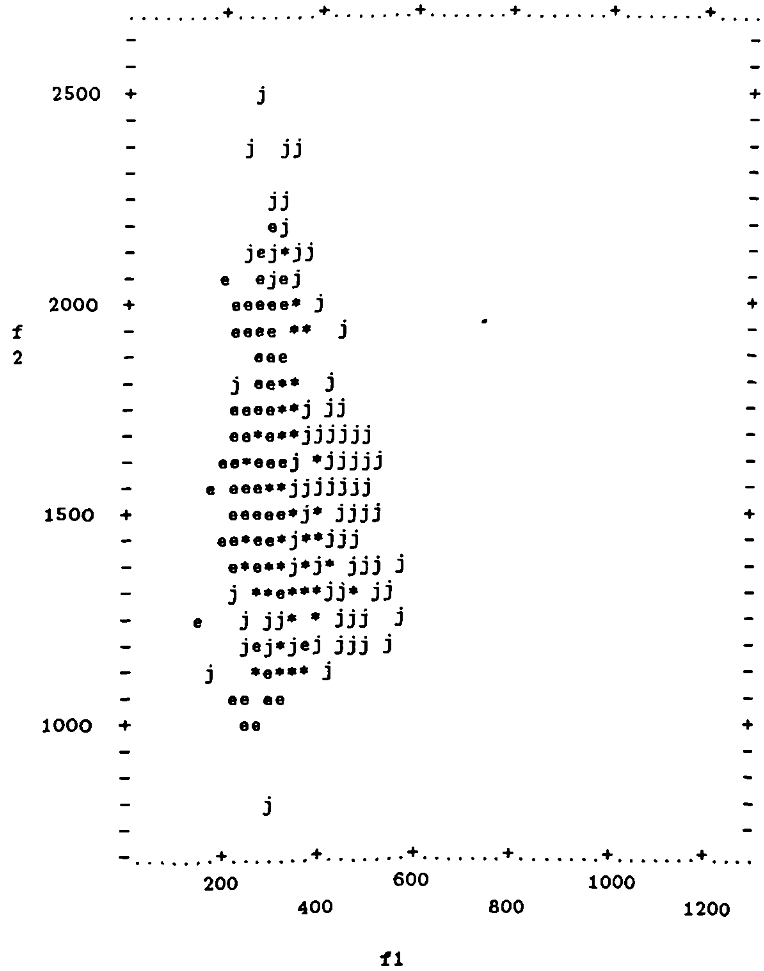

Figure 3.5 Schwa produced by native speakers and fluent non-native speakers

A difference between the two groups seems to be that the non-native speakers generally have higher F1 values than the native speakers' schwas. The mean F1 values of schwa produced by fluent non-native speakers of English was 389 Hz, while native speakers were 292 Hz. There was a significant difference in F1 values of schwa ($F(1,492) = 310.01$, $p < 0.0001$). The standard deviations for both F1 and F2 values were greater for fluent non-natives than natives; 46 Hz for natives and 73 Hz for fluent non-natives in F1; as for F2 values were 250 Hz for natives and 290 Hz for fluent non-natives. On the whole, fluent non-native speakers' schwa seems to be more variable in F1.

It has also been found that the F2 values of the non-fluent non-native speakers' schwas and the F2 values of the Japanese vowel /a/ are very much alike (Figure 3.6). (e = English schwa, j = Japanese vowel /a/, * = an overlap between tokens belonging to different groups).

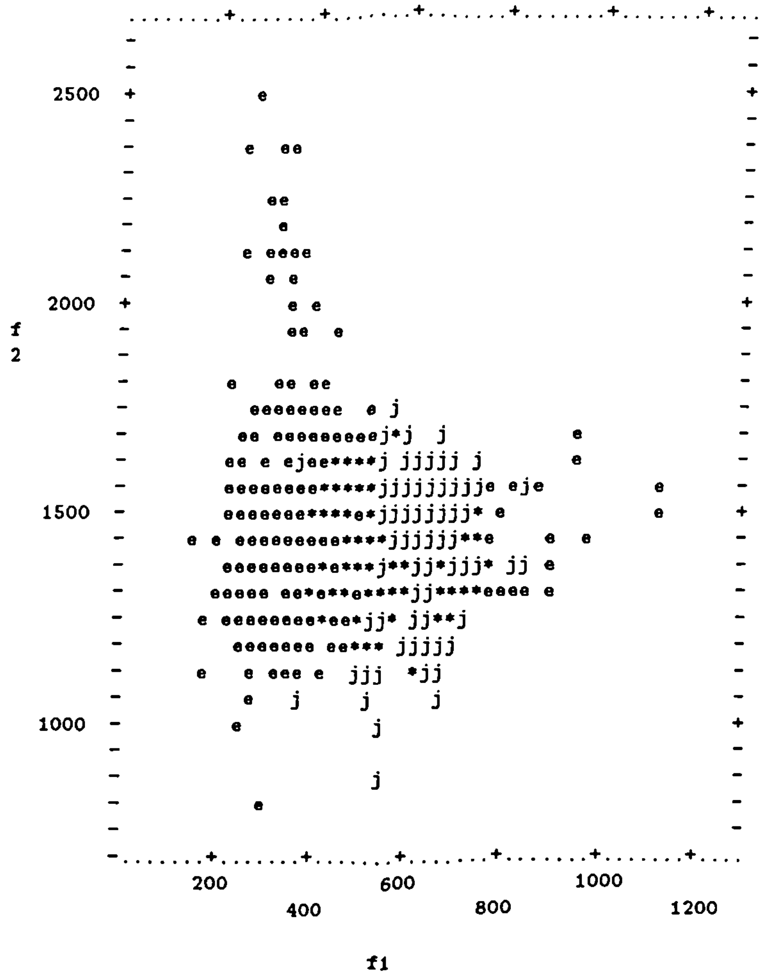

Figure 3.6 English /ə/ and Japanese vowel /a/ of non-fluent non-native speakers

In other words, there is clearly transfer from L1 to L2 production of schwa and the non-fluent non-natives haven't acquired important characteristics of schwa, i.e. its targetlessness in F2. The range of the F2 values of schwa produced by non-fluent speakers is 776 Hz (from the minimum (1007 Hz) to the maximum (1783 Hz) values). It does not show the spread in F2 values as a function of context as has been seen in English native speakers' production of schwa. On the contrary, the spread of schwa tokens in F2 was rather like that of the Japanese vowel /a/, and the spread in F1 values is far greater for their schwa than for the vowel /a/ of Japanese.

Subject MN, one of the fluent non-natives, was the only speaker who participated in both the interlanguage study and the Japanese vowel study. The mean F1 and F2 values for the Japanese vowel /a/ were 640 Hz and 1334 Hz; while those of his English schwa were 454 Hz and 1496 Hz. His schwa tokens were found to be clearly different from the Japanese vowel tokens, and they exhibited the characteristics of the schwa in the F2 values. The results of regression analyses showed differences between his English /ə/ and the Japanese /a/. The difference was most obvious for F2 and the r^2 value was 0.1920 vs. 0.6368 for Japanese and English respectively.

One other characteristic of the non-native schwas was that the duration of the schwas was longer than those of natives. Native speakers' schwas were the shortest among the three groups (35 ms.) while non-fluent non-native speakers' schwas were relatively long (70 ms). Fluent non-native speakers' schwas were somewhere in the middle with a mean value of 42 ms. The difference between the three groups was significant (F)2,800 = 471.57, $p < 0.0001$). It should be added that the duration of schwa and the Japanese vowel /a/ produced by MN did not show any significant difference, with a mean duration of 42.9 ms (SD = 12.8 ms.) for schwa and a mean duration of 46.3 ms (SD = 13.6 ms.) for the vowel /a/. This contradicts the results obtained from the values of the formant trajectories for MN. Kondo argues that this supports the view that though duration may affect the extent of contextual assimilation, the underspecification of schwa may be phonetically determined and that MN seems to have acquired the phonological features of schwa in the course of learning English.

As can be seen in Table 3.1 and Table 3.2, MN's production of schwa does not vary as extensively as native speakers' schwa in F2 and the spread in F2 values as a function of context is also smaller in magnitude, but the pattern observed for his schwa is clearly closer to the native speakers' pattern of schwa than to the pattern of the Japanese vowel /a/.

Table 3.1 shows the F1 and F2 values at the midpoint of schwa in the 3 consonant /p, t, k/ x 3 vowel /ɪ, æ, u/ contexts and the /ə/ and the Japanese /a/ produced by MN. The rightmost column of each matrix shows the mean formant frequencies for the vocalic contexts. The bottom row shows the mean formant frequencies for the consonantal contexts. The rightmost value on the bottom row is the total mean for each matrix.

Table 3.1 MN's production of English /ə/ and Japanese /a/

F1					F2				
Japanese /a/	p	t	k			p	t	k	
i	778	663	720	718	i	1331	1503	1404	1415
a	785	674	604	693	a	1209	1423	1319	1339
u	741	617	588	635	u	1226	1421	1394	1353
	773	652	644	686		1265	1450	1372	1371
English /ə/	p	t	k			p	t	k	
ɪ	498	446	439	461	ɪ	1388	1571	1684	1548
æ	508	490	442	480	æ	1368	1550	1643	1520
u	540	429	431	455	u	1279	1598	1500	1459
	511	455	437	466		1345	1573	1609	1509

Table 3.2 shows the F1 and F2 values at the midpoint of schwa for the 3 consonant /p, t, k/ x 3 vowel /ɪ, æ, u/ contexts for each group of speakers. The rightmost column of each matrix shows the mean formant frequencies for the vocalic contexts. The bottom row shows the mean formant frequencies for the consonantal contexts. The rightmost value on the bottom row is the total mean for each matrix.

Table 3.2 F1 and F2 values for each group of speakers

F_1					F_2				
Native English	p	t	k			p	t	k	
ɪ	301	276	284	287	ɪ	1391	1659	1945	1672
æ	321	283	293	296	æ	1266	1567	1850	1561
u	312	275	286	290	u	1263	1640	1562	1491
	310	278	288	291		1306	1621	1786	1573
Fluent non-native	p	t	k			p	t	k	
ɪ	403	391	361	384	ɪ	1366	1598	1858	1607
æ	390	448	353	397	æ	1306	1484	1805	1532
u	406	379	376	386	u	1211	1474	1472	1386
	399	406	362	389		1294	1512	1716	1508
Non-fluent non-native	p	t	k			p	t	k	
ɪ	323	328	403	354	ɪ	1319	1570	1501	1463
æ	350	390	341	362	æ	1269	1536	1380	1395
u	656	395	506	515	u	1297	1452	1398	1382
	458	371	414	411		1295	1519	1427	1413
Native Japanese	p	t	k			p	t	k	
i	682	620	637	645	i	1319	1567	1469	1453
a	667	552	577	596	a	1159	1484	1380	1349
u	610	603	588	600	u	1201	1453	1311	1323
	655	591	601	614		1229	1502	1386	1375

To sum up Kondo's interlanguage study, she found that schwas produced by non-native speakers of English exhibited differences between fluent and non-fluent non-native speakers of English. The production of schwa by non-fluent non-natives was characterised by transfer from their native Japanese vowel /a/ in the F2 dimension and also by a large variability in F1 which is unique to the production of the non-native /ə/. Kondo speculates that L2 speakers are aware of the difference between the non-native /ə/ and the native /a/ and try out random productions to approach the quality of /ə/. The production of schwa by fluent non-natives was characterised as being very similar to the English native speakers' production which was characterised systematically and largely as a function of context, the targetlessness in F2. She also found a merged system between the L1 and the target language for interlanguage production, i.e. the mean F1 and F2 values of schwa produced by non-natives were intermediate in value between those observed for natives' schwa and those of the Japanese vowel /a/. It should be noted that the fluent non-natives were closer to the native speakers' values.

She argues that it is unlikely that learners had access to explicit knowledge that schwa is targetless in F2, but may have had some "explicit knowledge" of the rhythmic alternation of strong and weak syllables in English, which is relatively easy to perceive. By producing schwa as a weak and short vowel, L2 learners may have succeeded in producing a more contextually variable and therefore targetless schwa. Production of targetless schwa may consequently help them to make a more prominent contrast between reduced and full vowels in English, resulting in a more English-like rhythm. This in turn may encourage learners to produce a less and less specified schwa. Here, she assumes that there should be some kind of explicit knowledge or external stimuli to trigger acquisition, which at the end helps learners to internalise the knowledge of the phonetic underspecification of schwa. Unfortunately, she does not discuss exactly what the "explicit knowledge" or "external stimuli to trigger acquisition" might be. As Kondo acknowledges, without a longitudinal study it is premature to conclude that the two different patterns observed for the fluent and non-fluent speakers of L2 in her study represent different developmental stages in the interlanguage.

The study reported in this thesis aims to further investigate the developmental stages of the acquisition of schwa and its relation to the prosodic features of English by Japanese learners. Details of the methodology and the results of the production study will be reported in Chapter 4.

3.2 Suprasegmental phonology

In this section, I will look at the suprasegmental features of English and Japanese. I will first look at the syllable structure of both languages (3.2.1). Then I will look at rhythmic typologies for language, namely 'stress-timed', 'syllable-timed', and 'mora-timed'. It will be shown that the differences between the typology of, for example, stress-timed and syllable-timed are not clear cut and that languages fall into a continuum depending on how important a role stress and other structural features play in them. I will also show that the differences in rhythmic structure are the result of differences at the structural level. Additionally, I will look at the third rhythmic typology, the mora-timing, in detail and examine whether 'mora' can exist as a timing unit in Japanese. In 3.2.3, I examine how English and Japanese share similarities when it comes to marking prominence. While there are differences between the role that accent plays in English and in Japanese, accent in both languages serves as a marker of prominence and utilises pitch.

3.2.1 Syllable structure of English and Japanese

English permits complex onset and coda and for example, up to three consonants may fill the onset position, namely, stop-glide, stop-liquid, fricative-glide, fricative-liquid, nasal-glide sequences. On the other hand, Japanese has predominantly CV syllable structure and more restricted sets of onset and coda consonant clusters are permitted. In the onset position, stop-glide, fricative-glide, nasal-glide and liquid-glide combinations are allowed and for coda position, long syllables consisting of either a consonantal mora, i.e. N or Q, or a part of the long vowel, e.g. *kan* (CVN), *kaa* (CVV), *Kyaa* (CyVV). I use /Q/ to refer to the 'moraic first part' of a long or geminate consonant and /N/ to refer to a moraic nasal (Kindaichi 1954: 162-164).

When English words are rendered in Japanese, besides phoneme substitution, they are normally modified to fit in with the predominantly CV syllable structure of Japanese. The English consonant cluster and the final consonant of a English word are altered into sequences consisting of a consonant and a vowel. The usual procedure of modification is to insert the /u/ vowel into the position breaking up the cluster, e.g. 'crispy' [krispi] will be rendered in as [kurisupi] (kVrɪsVpɪ). When the word ends with a consonant, /u/ will be added after the consonant, e.g. 'crisp' [krɪsp] will be altered into [kurisupu] (kVrisVpV). Exceptionally ,when the first consonant in the cluster is /t/ or /d/ (or the word ends with /t/ or /d/), the /o/ vowel will be inserted, e.g. 'troy' [trɔI] becomes [toroy] (tVroy). Thus, 'street' (CCCVC) is altered into [sutori:to] (sVtVri:tV).

3.2.2 The Rhythm of English and Japanese

3.2.2.1 Rhythmic typology of languages

Languages are often perceived to have distinct rhythmic styles, with variation across languages explained by typologically classifying them as either 'stressed-timed' (e.g. English, Russian, Arabic) or 'syllable-timed' (e.g. French, Spanish). These categories were traditionally thought to be mutually exclusive. A third rhythmic category was recognised more recently namely 'mora-timed', of which Japanese is the best known example (Han 1962; Hoequist 1983).

These typological labels rely on the assumption that isochrony holds either at the level of stressed syllables or at the level of individual syllables or mora, depending on the language type. Classified as 'stress-timed' are English (Classé 1939, among others), Russian (Abercrombie 1967; O'Connor 1973), Arabic (Abercrombie 1967), Thai (conversational) (Luanghongkum 1977), Brazilian Portuguese (Major 1981). In these languages, it is said that stressed syllables occur at approximately equal time intervals. On the other hand, languages such as French (Abercrombie 1967; Ladefoged 1975; Catford 1977) Spanish (Pike 1946; Hockett 1958), Telegu (Abercrombie 1967) and Yoruba (Abercrombie 1967; Corder 1973) are languages typically classified as 'syllable-timed'. For this rhythmic type, it was hypothesised that there are equal time intervals between successive individual syllables (Lehiste 1977). Japanese, on the other hand, is said to be 'mora-timed' and isochrony is said to be maintained at the level of the mora (Ladefoged 1975).

3.2.2.2 'Stress-timed' vs. 'syllable-timed' re-revisited

For the purposes of simplification, we will first concentrate only on stress-timed vs. syllable-timed languages. After Abercrombie's original claim about rhythmic typology, the search for phonetic correlates of rhythmic typology has continued; however, instrumental evidence supporting the claim of stress isochrony vs. syllable isochrony has not been very fruitful (Lehiste 1977; Bertinetto 1988).

To begin with, it must be noted that variability should not be taken as evidence against classification of rhythm types. Allen (1975) notes that generally speaking, human beings tend to act rhythmically. However, even when a person is performing a rhythmic pattern at his/her preferred rate, there is a range of variability between 3-5% of the total length of the interval, or between 7-11% if we perform the task at a less-preferred rate. Allen shows that variability found in speech timing is similar to that of other types of motor behaviour such as clapping hands, and that durational differences of inter-stress intervals tend to fall within this range (at least) in English.

Taking the example of English, the most studied stress-timed language, some have rejected isochronism in production on the basis of their data (Shen & Peterson 1962; Nakatani, O'Connor & Aston 1981), while others have claimed that a tendency toward isochronism still exists (Classé 1939; Lehiste 1977). The problem seems to have its root in the fact that there is no clear-cut definition of isochronity and investigators' claims were largely subject to the interpretation of their data (Dauer 1983: 52). For example, Abe (1967), in Allen (1975), claimed that the preponderance of interstress intervals with duration in the narrow range of 40 cs to 70 cs was evidence that English is stress-timed, while O'Connor (1965) rejected isochronity because the difference between the longest and shortest interstress intervals in a limerick was 88 ms.

Inconclusive results of physical isochrony found from instrumental studies has led some to claim that isochrony is a "perceptual illusion" (Lehiste 1977), while it has led others to consider isochrony as merely an "underlying tendency" that is rarely observable instrumentally. Furthermore, Dauer (1983) using a cross-linguistic study argued that the differences in rhythmic structures are the results of differences at the structural level, the level of phonological representation, and that the difference between 'stressed-timed' and 'syllable-timed' is gradient, not categorical or mutually exclusive.

Dauer (1983) asked native speakers of English, (conversational) Thai [classified as stress-timed languages], Spanish [traditionally classified as syllable-timed] and Italian and Greek [not yet classified] to read a 1.5 to 2.5 minute passage from a modern play or a novel in their native language. A native speaker of each language and a phonetician marked which syllables they felt were stressed. Intervals were acoustically measured from the onset of the first vowel to the onset of the next stressed vowel and so on up to the last stressed vowel before a pause.

In all the languages analysed, the durations of the majority of interstress intervals (75%) fell within the relatively narrow range of about 30 cs. Pair-wise comparisons between English speakers and Spanish, Greek or Italian speakers showed no significant differences in variance at the .05 level. In addition, the increase in average duration of an interstress interval due to the addition of another syllable was similar in all speech samples (approximately 11cs) and clearly the mean duration of interstress intervals was proportional to the number of syllables in the interval for all the languages analysed, whether the language was classified as stress-timed or syllable-timed. As Dauer notes "there was no more of a tendency for interstress intervals to clump together in English than in other languages" (1983:54). The results led Dauer to suggest that the data reflect universal properties of temporal organisation in language, which confirms Allen's (1975) suggestion that human beings tend to act rhythmically. He further claimed that "the difference between languages such as English and Spanish has to do with what goes on *within* rather than across interstress intervals" (1983:55), and that factors which contribute to the perceived difference can be attributed to a difference in language structure. He listed factors which contribute to the stress-timing nature of English:

(1) complex syllable structure

> Stress-timed languages such as English have a greater variety in permissible syllable types, and syllable length varies to a greater degree e.g. from CV to CVCC, CCVC, CVCCC, while in so called syllable-timed languages such

as Spanish[5] and French, open syllables predominate.

(2) vowel reduction

Reduction of many unstressed vowels: in stress-timed languages, there is a greater degree of vowel reduction in unstressed syllables than in syllable-timed languages and only a restricted set of segments can occupy such a position. For example, in English, /ə/, /ɪ/ or /ʊ/ or a syllabic sonorant as in 'button' [bʌtn̩], while in Spanish any of the five vowels can occupy any position and the vowel reduction process is less perceptible, as the unstressed vowel will just be phonetically realised slightly more centralised that its stressed counterpart.

To sum up, Dauer (1983) argues that the differences in rhythmic structure are a result of differences at the structural level and that the difference between 'stressed-timed' and 'syllable-timed' is gradient. In other words, he suggests that we should define the rhythm of a language as falling onto a continuum which reflects how important a role stress and the other structural features play in a language. He

[5] Confirming Dauer's claim, Borzone de Manriqe & Signorini (1983) (cited in Parrondo-rodriguez 1999) used acoustic data from Argentine Spanish to show the same rhythmic features as stressed-timed languages. They found: (1) similar duration of inter-stress intervals; (2) reduction of vowels in unstressed positions; and (3) similar syllable length depending on stress and position within the group. They concluded that the dominance of open syllables in Spanish and the particular ways in which unstressed vowel reduction is carried out might make syllables more perceptible than in other languages and this might be contribute to Spanish being perceived as syllable-timed.

proposed the continuum shown in (13), where languages can be compared to each other along the dimension of having a more or less stress-based rhythm.

(13) The degree of stress-basedness.

Stress-based →

Japanese French Spanish Greek Portuguese English

3.2.2.3 'Mora-timed'

We will now deal with the third rhythmic type, mora-timed (Bloch 1942; Han 1962; Hockett 1955; McCawley 1968; Ladefoged 1975). It has been asserted that the mora, an abstract unit of duration, holds an important role in Japanese. This means that in Japanese, speech consists of a series of timing units with a roughly equal duration. A mora has one of the following realisations, as shown in (14).

(14) a. V	e.g.	/i/ 'stomach',
b. CV	e.g.	/ki/ 'tree'
c. First part of a geminate	e.g.	ikki (iQki) 'in one stroke'
d. 'moraic' nasal /n/	e.g.	kikan (kikaN) 'period'

The first type of mora consists of one vowel such as in /i/ 'stomach'. Another type refers to a vowel preceded by a consonant as in /ki/ 'tree'. As in (c), there is a case where a long consonant, i.e. a geminate, occurs and the first part of the long consonant is considered one mora. Therefore, in /ikki/ ('in one stroke'), there are three moras (/i.k.ki/). As in (d), /n/ may stand alone as a moraic nasal. Examples of this includes, ***kikan*** 'period'. Hence, /kikan/ consists of three moras (ki.ka.n).

This can be best illustrated in a traditional verses such as a 'haiku' which consists of 5-7-5 morae. The following is an example of a haiku by Bashō Matsuo (Vance 1987: 64).

(15)	Ran no ka ya	Chō no tsubasa ni	Takimono su
	/ra-N-no-ka-ya/	/čo-o-no-cu-ba-sa-ni/	/ta-ki-mo-no-su/
	'The fragrant orchid	Into a butterfly's wings	It breathese to incense'

When the term mora-timed was introduced, isochrony was thought to be maintained at the level of individual moras, therefore, a two-mora word should be twice as long as one mora word. However, some experimental work based on durational studies questioned the validity of the assumption that the mora is a timing unit in Japanese and it was found that the moras were not isochronous at the level of individual moras (Beckman 1982; Port, Dalby & O'Dell 1987), but rather maintained at word level (Port, Al-Ani & Maeda 1980; Port, Dalby & O'Dell 1987; Han 1994), as will be seen shortly.

Traditionally, the mora has been known to the Japanese as 'onsetsu' (sound unit) or 'haku' (beat) and Japanese pedagogy explicitly states that they are of equal duration. However, the question of whether or not this perceived length has any physical basis has been the subject of much debate. Han (1962) carried out the first experimental study which looked into the inherent and contrastive durations of vowel and consonants in word-medial position. She found that they are 2.5 to 3.0 times as long as their corresponding short ones. She asserted that these extra long durations are interpreted as doubling of the preceding mora, not the preceding phoneme. Her experiment on word duration included a two-mora word at 40cs. while a three-mora word was 62 cs., which approximates the ratio of the number of moras, 2:3. She claimed that there is a strong tendency for the components of the unit to balance each other out so as to obtain equal duration with neighbouring units.

Port, Al-Ani & Maeda (1980) compared the timing control mechanisms in Arabic and Japanese and claimed that the isochronity of a mora is maintained at the word level, rather than at individual mora level. For Japanese, they used two syllable words such as /basa/, /bata/, /bada/ /bara/ and investigated the duration of each word and the total length of each word, and found that Japanese vowels on both sides of a consonant varied inversely with the consonant, and the consonant constriction duration varied inversely with the inherent vowel duration, which they called temporal compensation. Such an effect was not observed in Arabic. In other words, there was a tendency for achieving isochrony at the word level.

In line with Port et al.(1980), Homma (1981) argued that moras have isochronity at word level. He compared two-mora words (e.g. /papa/, /tata/, /gaka/, /baba/) and three-mora words which contain a long consonant (e.g. /paQpa/, /taQta/, /gaQka/, /baQba/) and found that the ratio of the total length of the word was kept at 2:3. He concluded that temporal compensation can be observed within a word, not within a syllable.

Beckman (1982) was among those who seriously questioned the validity of mora, i.e. whether the mora has any acoustic reality at all in Japanese. She rejected the notion of the mora for the following reasons: (1) when vowel deletion occurs, the resulting syllabic consonants tend to be shorter than the CV mora (in her experiment, the syllabic consonants were not consistently longer than the non-syllabic consonants); (2) phonemically long consonants (geminate consonants) were not three times as long as their single counterparts; (3) when /kaCV/ sequences with various Cs and Vs were embedded, it was not observed that the duration of moras was approximately the same, nor was there any tendency for the components of each mora to balance with other moras out to obtain equal duration with neighbouring units. She argued that mora's isochronicity may be only a historical relic maintained by traditional pedagogy, recognition of moras in the *kana* writing system,[6] or even just a perceptual unit having no phonetic basic.

---

[6] Han (1994:75) argued that there are many Japanese-Americans who speak perfect Japanese with no reading knowledge. This is a counter argument for Beckman's claim that literacy is the key for

Port, Dalby & O'Dell (1987) did not dispute Beckman's results, but attempted to salvage the mora by emphasising that isochronity is attained at word level and not at the level of individual mora, as was the case in Beckman's analysis. They carried out four experiments which followed up Port et al.'s (1980) study and demonstrated that the mora is an abstract timing unit and also has physical reality in Japanese. They found that:

(1) Words with an increasing number of moras were augmented in duration by nearly constant increments.

> A list of test words (mostly nonsense words, but plausible in Japanese) were selected so that they differed in the number of moras, for example, ra (1 mora), raku (2 moras), rakuda (3 moras), rakudaga (4 moras), rakudagashi (5 moras). There were embedded in a carrier sentence *Kore wa ____ desu.* 'This is ____'. These were read by four native speakers of Japanese at a comfortable tempo. Using spectrographic analysis, the duration of each word as well as various component segmental units were measured and it was found that each additional mora added nearly the same amount to the total word duration.

Japanese speakers recognition of moras.

(2) At two different tempos, the word durations correlated very closely with the number of moras.

The second experiment was an extension of the first experiment and aimed to test the 'isochronity at word level' independent of the change in tempo. This time a list of 28 real words was used ranging from one mora in length to six or seven. The words included a variety of vowels and geminate consonants and were embedded in a carrier sentence: *Yukkuri* ___ *to iimashita*. '*pro* said it slowly'. Five native Japanese speakers read the sentences four times in fast and slow tempos. The last two recordings were used for each tempo. Results showed that at both tempos, the word length increased by roughly equal increments for each additional mora. There was very little difference between words that had the same number of moras despite major differences in their segmental make-up.

(3) Speech timing is adjusted within words so that words with the same number of moras have about the same overall length, even though there are large universal differences in the inherent duration of various segment types such as /u/ vs. /a/. This is achieved by stretching or compressing the duration of neighbouring segments and adjacent moras.

A set of test words were used which permit examination of a single segmental change against the voicing contrast: a vowel change, a voicing change and a change in manner was used (namely, kuka, kaka, kata, kasa, kuga, kaga, kada, kaza). The eight test words were embedded in a carrier sentence: *Korewa* ___ *desu.* 'This is a ____'. Speech data were elicited from nine native speakers of Japanese (four males and five females) and their production was analysed acoustically. The means of the intervals measured revealed that regardless of the vowel in the test word (/u/ or /a/), both the initial stop in the word and the following vowel were adjusted compensatorily and achieved constant word duration.

(4) Comparison of the number of moras versus the number of syllables demonstrated that timing in Japanese was constrained by the mora rather than by the syllable durations.

The fourth experiment tested whether the tendency for word duration to come in integral steps was only a reflection of a tendency to make *syllables*, rather than moras, roughly constant in duration. Four pairs of words were included in the test: one pair had two moras and two syllables (CVCV: buku, baku), one pair had three moras and three syllables (CVCVCV: bukudo, bakudo), while the other two pairs had three moras but had only two syllables (CVVCV: buuku, baaku; CVCCV: bukku, bakku). Ten native speakers of Japanese (three male and seven female) read the test words, which were embedded in a carrier sentence: *Watashi wa ___ da to omoimasu.* 'I am ___ so think' [translation from Port et al. 1987:1582]. The results demonstrated that the overall word durations for the words nominally three-moras long were very nearly the same. Furthermore, the 'heavy' syllables (the ones with closing consonants and long vowels) did not exhibit the expected shortening of other components of the syllable, but showed lengthening in this context; this is clear evidence that timing in Japanese is constrained by an abstract temporally defined mora, not by a tendency to regularise syllable durations.

In summary, Port et al.'s (1987) four experiments lend strong support to the general view that the mora is a timing unit in Japanese. A handful of studies have confirmed Port et al.'s (1987) claim. Kubozono (1989) obtained results which in most respects confirmed Port et al.'s analysis through speech error data. Additionally, Han (1994) supported the view that a segment may be stretched or compressed to provide temporal compensation within a mora and between moras.

3.2.3 Pitch-accent in English and Japanese

In pitch-accent theory (Bolinger 1958), prominence is largely determined by a local pitch change which signals that the syllable with which it is associated is prominent. There are reasons to believe pitch accents are a part of the phonological system of both English and Japanese, even though stress is a feature of English but not of Japanese. While there are differences between the role that accent plays in English and the role that accent plays in Japanese, accent in both languages seems to serve as a marker of prominence, as will be seen shortly.

Beckman & Pierrehumbert (1986) [henceforth, B & P] claimed that one of the most salient similarities between English and Japanese is that both have tonal phenomena that can be described in terms of the notion 'accent'. Although the phonetic nature of accent is not identical in the two languages, there is a fundamental likeness in that accent in both languages involves an association between some well-defined pitch shape (i.e. high or low tone) in the melody and a specific syllable that, by virtue of the association, is 'accented'. For example, Beckman (1986) pointed out that the accent types in Japanese and English are similar in their effect on pitch pattern. A comparison of fundamental frequency measurements of disyllabic words in Japanese and English revealed that the falling and rising intonation contours were very similar to each other in contrasts between initial vs. final stressed words in English (e.g. 'cóntrast' vs. 'contrást') and Japanese initial vs. final accented words in Japanese (e.g. *kata* 'shoulder' (HL) vs. *kata* 'form' (LH).[7] Here, 'H' and 'L' refer to high pitch and low pitch, respectively.

On a larger time scale, English and Japanese organise tone features similarly into a hierarchy of prosodic structures, from the grouping of tones into pitch-accents and at a local level to the choice of phrase-terminal tones and the manipulation of pitch range over larger domains. In a more detailed way, both languages are similar for several reasons. To begin with, in both languages, pitch accents or accentual phrase tones are limited to being one- to two-tone units, i.e. H or L. Other languages can have multiple levels, such as Chinese, which has five levels of tone, i.e. low, half-low, mid, half-high, high (Chao 1930). In both English and Japanese, the two-tone accents trigger catathesis. Catathesis seems to occur only within some intermediate level of phrasing, which is closely connected to such organisational effects as focus domain or scope of conjunction. Furthermore, larger prosodic units are marked by boundary tones, which are aligned to the edges of the unit and are not phonologically associated to any tone bearing unit in particular. I will not give details here because I will only be dealing with accents within a word and therefore they are not directly relevant to my study.

We have seen that both languages utilise pitch to mark prominence. I will now outline how accent in Japanese is determined to illusrate the precise nature of the similarity and the differences between the two languages.

---

[7] Although acknowledging the similarities between the patten of the pitch contours in the two languages, I will later show that the falling pitch in Japanese is somewhat later than in English, as

In Japanese, the location of the accent is not predictable, in other words; it is lexically marked and it must be learned separately for each word. In English, the pattern is predictable, although extremely complicated (Halle & Vergnaud 1987). In Japanese, pitch is a distinctive feature, i.e. identical words can minimally differ in pitch pattern while English does not contrast words in this manner. Consider (16) where each pair differs only in pitch pattern. It should be noted that the second example in each pairs does not have an accent, because there is no pitch fall that is a marker of accent.

(16)	a	ame	(HL)'rain'	vs.	ame	(LH)	'candy'
	b.	kan	(HL) 'can'	vs.	kan	(LH)	'sense'

In Japanese, there are very limited possible pitch patterns within a word. For example, consider (17).

(17)	a.	haru	(HL)	'spring'
	b.	saka	(LH)	'hill'
	c.	kokoro	(LHL)	'heart'
	d.	takara	(LHH)	'treasure'

Sugito (1969; 1980) argues.

The pitch pattern in Japanese is determined by the following two generalisations: (1) the accent marks the location in the word where the pitch falls and the accented syllable as well as the moras preceding it all receive high pitch; (2) in Tokyo Japanese, there is a rule called the Initial Lowering Rule (Haraguchi 1977), which states that the pitch of the first word is low unless the accent is placed on that syllable. Given these two generalisations, the pitch pattern of the entire word should be predictable no matter how many syllable there are in a word. Take the example of (17d.). In *takara* 'treasure', the accent marker is on the last syllable, which means that all the moras are realised with high pitch due to generalisation (1). However, the second generalisation, namely the Initial Lowering Rule, applies and hence the pitch pattern of the word is LHH.

While there are only four possible pitch patterns in Japanese, in English Lieberman & Pierrehumbert (1984) identify six possible tonal configurations for pitch accents: H*, L*, H*+L, H+L*, L*+H, L+H*. When some portion of an utterance is accented, it is associated with one of these six shapes by linking the starred (*) tone of the pitch accent to the accented syllable in the phrase, e.g. *an orange ball gown* H* H*+L L or H*+L H* LL or L* H*LL (example from B & P 1986: 257).

Another factor which differentiates English and Japanese is that while in Japanese there are words without an accent as seen in (4), in English there are no content words lacking stress; all non-function words have at least primary stress to indicate prominence and may also have secondary stress. While this phenomenon can be observed as a difference between the two languages, from a broader view, it can be interpreted as indicative of a fundamental similarity between the two languages (McCawley 1968; B & P 1986). In some languages, there is a limited number of possible pitch accent locations. This holds for Japanese and English, e.g. one for Japanese and maximum two for English. In other languages, every syllable in a lexical item can carry a distinctive tone, and there are no restriction on the distribution of the different tones among the syllable. Examples of such languages include Yoruba and Chinese.

Although there are fundamental similarities in the relationship between accent and the lexicon in both languages, there is another closely related aspect which differentiates the two languages. In both languages, the accentually prominent syllables are marked in the lexicon, but how they are physically realised differs (Beckman 1986; Bolinger 1958; Lieberman 1960; Fry 1958).

Beckman (1986) studied Japanese and English and drew a distinction between 'stress accent' and 'non-stress accent'. She claimed that stress accent (as in English) differs phonetically from non-stress accent (as in Japanese) in that it utilises physical parameters other than pitch, such as amplitude (intensity), vowel quality (e.g. the use of reduced vowels) and duration. Her acoustic study of Japanese and English found that along with pitch, total amplitude and duration were also reliable phonetic correlates and perceptual cues in English, while Japanese predominantly depended on pitch. It should be noted that there is some disagreement as to what extent these three parameters contribute to perceptual cues in English. For example, Lea (1977) claimed that intensity outperforms pitch (fundamental frequency) as a cue of stressed syllables.

The foregoing discussion suggests that simple distinctions between stress languages and pitch-accent languages is not clear cut and similarities between the two languages in question can be established.[8] Similarly, Sugito (1969) claimed that English and Japanese share a similarity in that 'pitch' plays a crucial role in marking accent. As has been discussed, Japanese accent is marked with pitch: it is the movement of the falling pitch (rather than the relative pitch height) which marks accent. Moreover, despite the general impression that 'stress languages' mark accent with *stress* (intensity), this is actually a mixture of at least four acoustic parameters: speech power, fundamental voice frequency, phonetic quality and duration. And, at least in English, pitch more than loudness or duration is likely to be responsible when a syllable is perceived as 'stressed' or 'emphasised' (e.g. Bolinger 1958; Lieberman 1960; Sugito 1969, 1980; Fry 1958).

Recall the example from Beckman (1986) above which illustrates that the two languages are similar in their effect on pitch pattern. A comparison of fundamental frequency measurements of disyllabic words in Japanese and English revealed that falling and rising intonation contours are very similar to each other in contrasts between initial vs. final stressed words in English (e.g. 'cóntrast' vs. 'contrást') and initial vs.

[8] McCawley (1987) also rejects a simple classification of tone languages and pitch accent languages based on a discussion of Japanese and other languages.

final accented words in Japanese (e.g. *kata* 'shoulder' (HL) vs. *kata* 'form' (LH).

Although pitch patterns are similar in English and Japanese, according to Sugito (1969; 1980) there is a difference in the actual timing of the pitch pattern Let us take an example of a bisyllabic word with penultimate accent in both languages. For example, in Figure 3.7, in a word such as *summit* (HL), the falling pitch starts from the first syllable, while in Japanese, for a word with penultimate accent, such as *kame* (HL) "turtle", the falling pitch starts from the beginning of the second syllable. The production for Japanese *kame* was by one of my test subject, namely, *MO* and the English word *summit* was produced by one of the native speaker control, who also tape-recorded the testing material for the perception task in my study. Please also note that in the figure, [s] and [t] in *summit* and [k] in *kame* do not actually show up on the pitch contour. I have added those segments in the labelling for reader-friendliness.

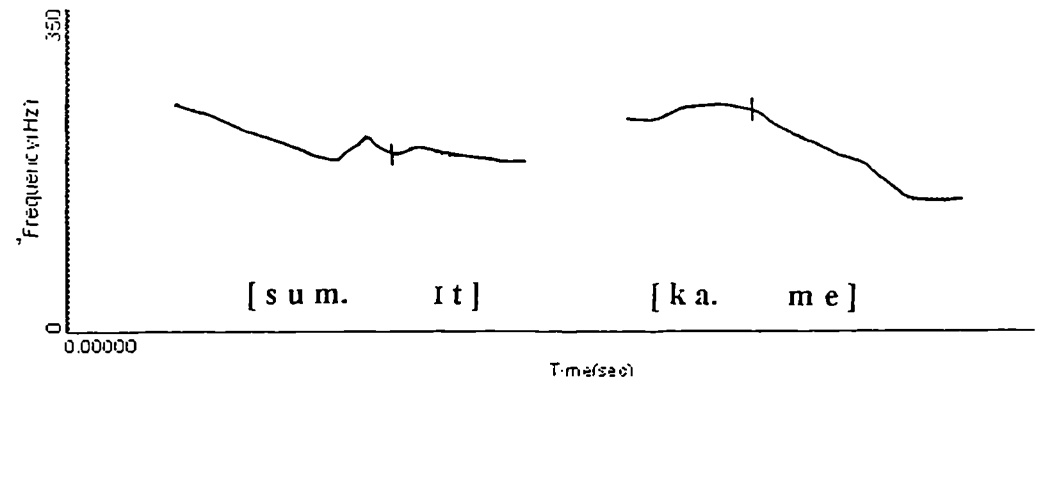

Figure 3.7 English NS *summit* vs. Japanese NS *kame*

In summary, we have seen that the differences in rhythmic structure are the result of differences at the structural level and that the differences between typology such as stress-timed and syllable-timed are not clear cut. Languages fall onto a continuum depending on how important a role stress and other structural features play them. I reviewed whether 'mora' exists as a timing unit in Japanese. I also examined how English and Japanese share similarities when it comes to marking prominence. It was concluded that, while there are differences between the role that accent plays in English and in Japanese, accent in both languages serves as a marker of prominence and utilises pitch.

Chapter 4

The study

4.1 Introduction

The objective of this study is to investigate what effect exposure to "positive evidence with a *positive* effect" would have on adult second language (L2) learners who had previously received mostly foreign-accented input during their foreign language experience. Whether L2 learners' phonology undergoes changes in perception and production was investigated based on a longitudinal study of three Japanese learners of English.

If we turn to methodological shortcomings of L2 perception research, we find that studies are typically carried out focusing on a specific aspect of phonology, and few studies have looked at the interaction of various sub-systems of phonology. We also find that most studies are based on production data only; however, this does not reveal the full picture of phonological L2A. Generally speaking, cross-sectional and group studies fail to show developmental stages, which are best seen in longitudinal studies of individual learners.

The present study was designed to address the shortcomings of previous studies. Data includes three areas of phonology, namely, segmental features, syllable structure, and prosodic features, and both production and perception data were collected longitudinally looking at the interaction of various sub-systems of phonology. The remainder of this chapter is organised as follows: in 4.2, methodology for the data collection is outlined; in 4.3, the perception study is reported; in 4.4, two production studies are reported. The first study examines learners' acquisition of vowel reduction, especially the targetlessness of schwa in F1. The second study examines learners' acquisition of the timing of falling pitch.

4.2 Methodology

4.2.1 Subjects

The subjects for this study were three Japanese females (referred to as *NI*, *MO* and *SK*) at a U.K. university. The L2 learners had studied English from six to nine years at school and university in Japan before they arrived in the U.K. None of the subjects had had native-speaking teachers, except for occasional visiting English native speaker fellows. The main method of instruction was grammar translation. The learners were first recruited from a pre-sessional intensive English programme in the U.K. where they were enrolled to improve their English to reach the university requirement of IELTS 7.0. All were at the same placement level, i.e. high intermediate, in the programme. The age of the subjects ranged between 21 and 30. *NI* was on a three-month course and *MO* and *SK* were on a two-month course. *NI* was an undergraduate student on a 12-month exchange programme. She had no previous experience of living abroad. *MO* was a postgraduate student on a 14-month exchange programme. She had lived in the U.S. for a year when she was 13 years old, but reported that her family lived in a Japanese-speaking community so had very little contact with native speakers of English, and she went to a Japanese school. *SK* was a postgraduate student on a 12-month MA course. She had lived in the U.S. for 2 years when she was 14 years old.[9]

4.2.2 Data collection procedure

The data were obtained in a series of interviews during which the subjects met individually with the investigator. Meetings took place at the investigator's home in a relaxed atmosphere. Data collection sessions were held at approximately one-month intervals for all subjects for over a year, with the first data collection session held at the earliest date the subjects were available after they arrived in the country (average two weeks). Subjects performed several tasks testing their perception and production ability in English.

---

[9] *SK*'s stay in the U.S. gave her no advantage; she turned out to be the least proficient of the three and ended up with the most Japanese-like pitch contours. This tendency was repeated in her

4.3 The Perception study

4.3.1 Methods

4.3.1.1 Segmental features

Based on a contrastive analysis of the phonemic inventories of Japanese and English, the following seven problematic segmental contrasts were chosen for investigation: vowels (æ/ʌ, a:/ə, i:/ɪ, ou/ɔ:) and consonants (b/v, θ/s, r/l). Two to three minimal pairs per contrast were selected (28 test words in total), e.g. *rice* and *lice* for the /r/ /l/ contrast. An identification task was conducted with test words presented in the form of these pairs in context-neutral statements, such as "John gave Louise a *knit* sweater" and "John gave Louise a *neat* sweater." Test words were written out on an answer sheet and subjects were asked to circle the word they thought they heard, for example:

(18) The learners heard: "John gave Louise a *knit* sweater." *or*

"John gave Louise a *neat* sweater."

answer sheet: (a) knit (b) neat

---

perception task results.

4.3.1.2 Syllable structure

Based on a contrastive analysis of Japanese and English syllable structure, 76 test words were chosen which reflected sequences of sounds not present in Japanese, i.e. involving onset clusters, coda clusters and word final-consonants. Permissible target structures ("T" forms, below) and epenthesised forms ("E" forms) of the same word (created for the test) were involved in the test. For example, for glue [glu:], an epenthesised form [g*u*lu:] (a back high vowel inserted between the CC) was created.[10] Subjects first heard a word, and then the same word, or the epenthetic version, was presented, embedded in a sentence: "I said the word ______." Thus, there were four possible combinations:

(19)	a. glue	I said the word glue.	(T-T)
	b. glue	I said the word gulue.	(T-E)
	c. gulue	I said the word glue.	(E-T)
	d. gulue	I said the word gulue.	(E-E)

Subjects were asked to mark whether the word in isolation and the word in the sentence were identical or not. Thus, if the subject heard a combination of T-E or E-T, they were expected to answer NO, and if they heard E-E or T-T, they were expected to

[10] The epenthetic vowel is consistent with Japanese phonology and a high back vowel [u] is inserted (e.g. 's*u*mell]for *smell*) , unless it is after [t] or [d] (e.g. 's*u*toreeto' for *street*; 's*u*pendo' for *spend*).

answer YES. The purpose of the task was thus not to test whether subjects could judge which form was correct, but whether they could *perceive* when there was a difference between the two.

4.3.1.3 Prosody

Based on the assumption that prosodic features (pitch) of English and Japanese systems are compatible, the source of transfer was determined by the mismatch between the metrical parameter settings of the two languages. In defining the English stress system, Dresher & Kaye's (1990) metrical parameters, which characterise accentuation patterns in terms of universal principles and parameters, are relevant in the present work. For previous studies on the acquisition of the English accentuation system using this framework, see Archibald (1992, 1993, 1994 and 1995a, b) and Pater (1993, 1997). The Japanese counterpart is taken mostly from Haraguchi (1994).

The 7 classes of English accents (stress), taken from Archibald (1992), are shown in (20).

(20) 7 classes of English stress

Class 1 (noun): penultimate stress (tense vowel in the penultimate syllable)

Class 2 (noun): penultimate stress (branching rime in the penultimate syllable)

Class 3 (noun): antepenultimate stress (neither tense vowels nor CCs to attract stress in the penult)

Class 4 (verb): final stress (tense vowel in the final syllable)

Class 5 (verb): final stress (consonant in the final syllable)

Class 6 (verb): penultimate stress (the final syllable contain neither tense vowels nor CCs)

Class 7 (noun): secondary stress.

There were three words per class; thus a total of 21 test words were involved. Words unfamiliar to the subjects were chosen. This precaution was taken because it has been argued that stress patterns in English are lexically stored and knowledge of the word might affect learners' performances. To check this effect, learners' familiarity with each test word was checked every month. Grammatical and ungrammatical forms in terms of accentuation of these words were included. This is because both being able to tell the correctness of a form and its ungrammaticality is part of native speaker competence. For example, for "synopsis", which is a class 1, penultimate accent noun, both "synOPsis" and "SYNopsis" appeared on the test. Each test word was embedded in a sentence. There were two tests involved: subjects were first asked to indicate the place where they thought the word was most strongly pronounced (*Perception Test*), and then asked to decide to what extent the way the word was presented was acceptable in English (*Acceptability Test*). The first test was to examine whether subjects could phonetically perceive prominence in the test words and the second test was to examine their internalised rules for the L2 English accentuation system. The following is a sample of the testing procedure.

(21) [tape: He wrote a SYNopsis]

subjects were expected to answer:

SYNopsis correct a little odd don't know very odd incorrect

Answers were scored as follows:

(22)

	Correct	A little odd	Don't know	Very odd	Incorrect
Grammatical items	5	4	3	2	1
Ungrammatical items	1	2	3	4	5

4.3.2 Results

For segments, as one can see in Figure 4.1, all the subjects started from more or less the same point ($\bar{x}$ = 71.1 %). However very different learning graphs were observed over the data collection period. *NI* jogged up and down, but her performance could be interpreted as an overall upward curvilinear trend. *MO* showed a mild upward curvilinear trend. On the other hand, *SK* exhibited very little movement toward native-like performance. It seems to be the case that their years in the U.S. gave *MO* and *SK* no advantage. Raw data appears in Appendix A.

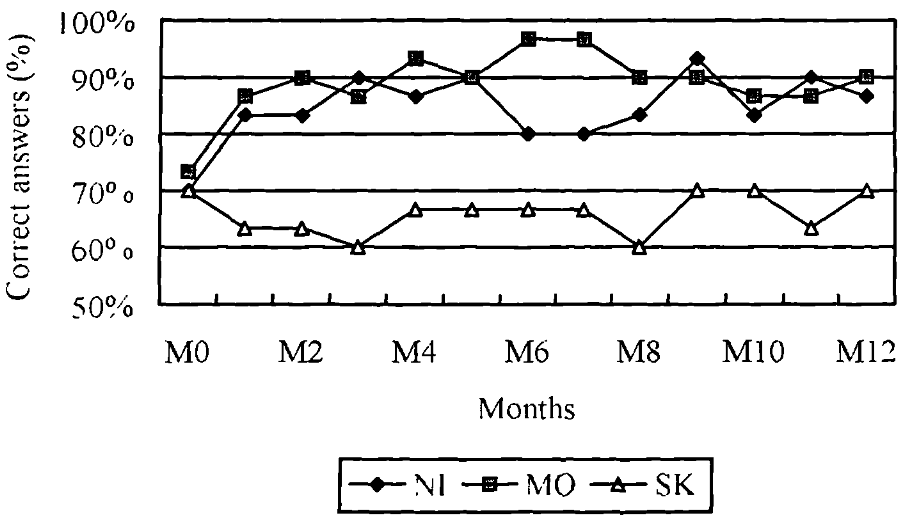

Figure 4.1 Segments: learning graphs for the three learners

For syllable structure, as seen in Figure 4.2., *NI* showed a dramatic increase in perceptual ability (from 66.3% to over 85%). *MO* started off with the highest score of the three learners, and after some improvement, plateaued at $\bar{x} = 88.7\%$. *SK* started off from a score in-between *NI* and *MO*, exhibited a mild upward curvilinear trend, and then plateaued at $\bar{x} = 84.7\%$.

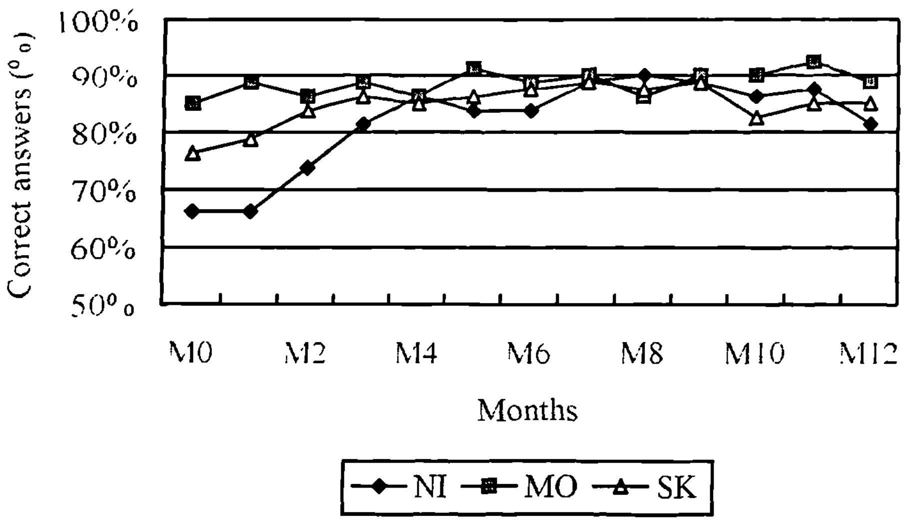

Figure 4.2 Syllables: learning graphs for the three learners

When the prosody data obtained are compared across subjects, a very different picture emerges depending on the test types: for grammatical items, all the learners had over a 90% mastery rate and no difference was observed among them. But for ungrammatical items, a clear difference was observed in the developmental pattern among learners (as can be seen in Figure 4.3).

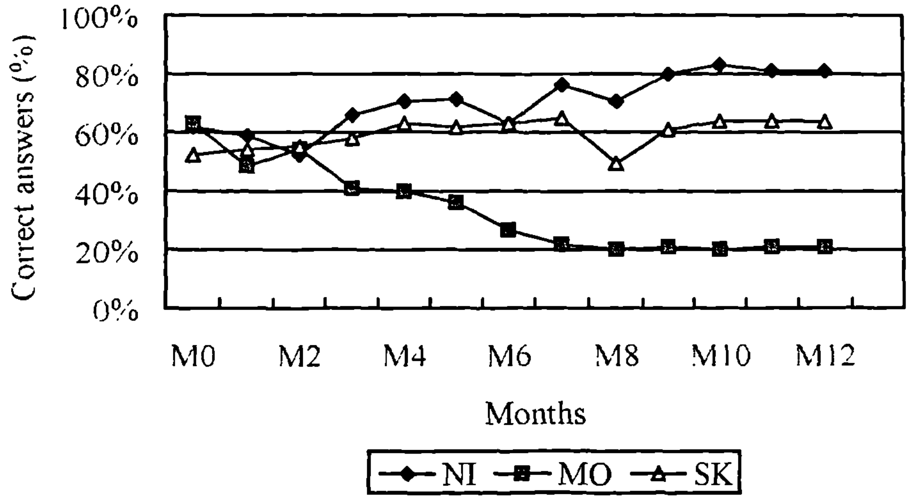

Figure 4.3 Prosody: learning graphs for the three learners (ungrammatical items)

NI exhibited a mild upward curvilinear trend towards the end of the data collection: until month 6, she exhibited only some improvement, but after this point continued to show improvement and reached over a 80% mastery rate. On the other hand, *MO* underwent a severe backslide (I will discuss this point further in the next section). SK showed little improvement.

4.4 The production studies

4.4.1 Methodology

Several factors were examined to determine the learners' production of reduced vowels, the timing of falling pitch and the marking of word prominence: 1) their pitch contours were extracted and segmented (for study II); 2) the midpoint values of the first two formants for reduced vowels were measured (for study I).

For segmentation, the criteria used in determining the beginning and the end of a vowel were similar to those described by Peterson & Lehiste (1960). In general, vowel onset and offset criteria included the initiation and cessation of voicing, friction and formant structure; aspiration was not included as part of the vowel. Syllable boundaries were determined by examining the pitch contours and spectrographic characteristics.

As for pitch contours, a computer programme called Multi Speech Model 3700 calculated the fundamental frequency of phonation between points in the waveform. For measuring vowel formants, to determine the quality of vowels, Peterson & Barney (1952) cited in Kent & Read (1992) (Figure 4.4) was used for English values. To obtain the higher female voice pitch, values increased by 20% were calculated.

	F1	F2	F3
[i]	300	2800	3300
[I]	430	2500	3100
[ɛ]	600	2350	3000
[æ]	860	2050	2850
[ɑ]	850	1200	2800
[ɔ]	590	900	2700
[ʊ]	470	1150	2700
[ʊ]	370	950	2650
[ʌ]	760	1400	2800
[ɜ]	500	1650	1950
[ə]	575	1700	2800

Figure 4.4 Vowel formant frequencies for English vowels: Peterson & Barney (1952)

Imaishi (1980) (Figure 4.5) and Sugito (1998) (Figure 4.6) were used as a reference for Japanese vowels. Figures given are for females. To measure formant frequencies using the sound analysis software Multi-Speech (Model 3700, Kay Elemetrics) and WinSAL (Version 1.2, Media Enterprise), wideband spectrograms (300 Hz filter) were first calculated and displayed with LPC formant tracks superimposed on the spectrographic display. In most cases, LPC tracks indicated accurate detection of the first two formants. When LPC analysis indicated errors, the values were estimated from the wideband. Spectral values were obtained at the midpoint of the vowels.[11]

	F1	F2	F3
[a]	978	1384	2716
[I]	381	2866	3699
[u]	390	1274	2760
[e]	510	2509	3246
[o]	567	894	3099

Figure 4.5 Vowel formant frequencies for Japanese vowels: Imaishi (1997)

	F1	F2	F3
[a]	828	1404	3084
[I]	372	2460	3648
[u]	432	1260	2736
[e]	612	2184	3048
[o]	588	1008	3192

Figure 4.6 Vowel formant frequencies for Japanese vowels: Sugito (1998)

[11] In Kondo's (1995:43) study, when one-way ANOVAs were performed between the mean values and the midpoint values of schwa F1 and F2 across speakers, the mean values of the first, middle and the last formant frequency values of a schwa segment biased the results in favour of the 'targetlessness' as the mean value across a segment includes more contextual information. Therefore, to avoid such bias, only the vowel midpoint values were recorded for analysis in this study.

4.4.2 Production study I: vowel reduction

4.4.2.1 Methods

In study I, I examined the developmental processes of the acquisition of schwa by three native speakers of Japanese (*NI, SK, MO*). Test words examined involved four nouns as in (23) and (24) which all had obligatory schwas in unstressed syllables. This meant that they could be mispronounced as Japanese [a], which was the most likely vowel to be substituted for English [ə].

(23) Parapet [ṕær.ə.pɪt] caliber [ḱæl.ɪ.bə]

(24) Cantaloup [kǽn.tə.lʊ̀:p] matador [mǽ.tə.dɔ̀:]

Data were elicited by having test subjects read index cards to compose sentences. Two tokens were obtained for each test word each month for over the period of 12 months.

I also had subjects read Japanese sentences to collect data on their native language vowels for point of reference. In the second data collection session (month 1), subjects were asked to read 20 sentences which included the vowel [a] and which had similar segmental sequences to the four English test words. For example see (25) and (26).

(25) *Nanio sureba ii desuka?* "What should I do?"

(26) Matadouzo irashite kudasai. "Please come again"

The mean formant value for vowel [a] is shown in Table 4.1.

Table 4.1 Mean formant values for Japanese vowel [a]

NI		MO		SK	
F1	F2	F1	F2	F1	F2
724	1289	832	1370	797	1367

4.4.2.2 Results

I will first discuss the general pattern of the learners' development (or non development). Figures 4.7, 4.8, 4.9 show scatterplots of schwa tokens produced by the three subjects, *NI*, *SK*, *MO* respectively. Raw data appears in Appendix A.

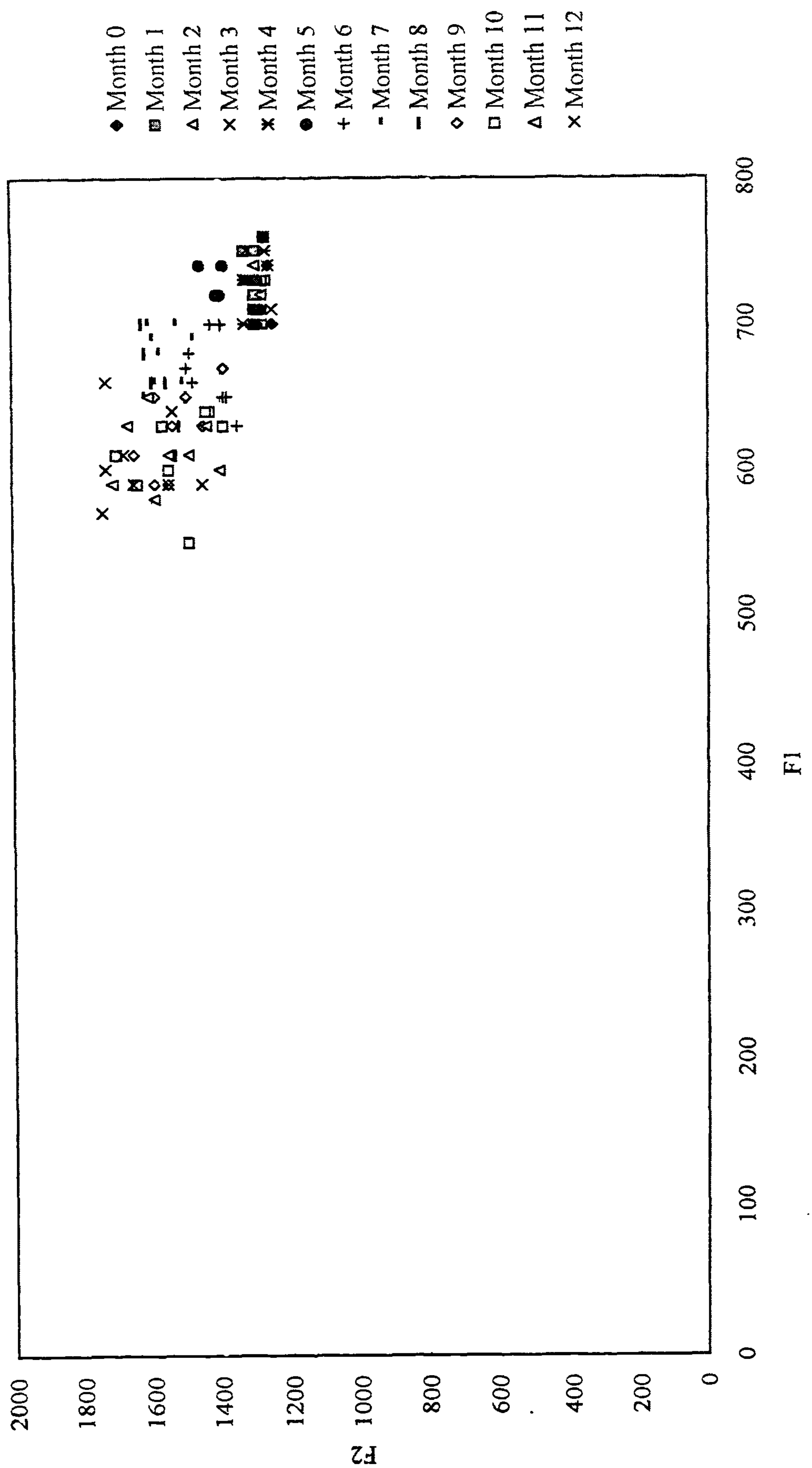

Figure 4.7 Monthly data for *NI*

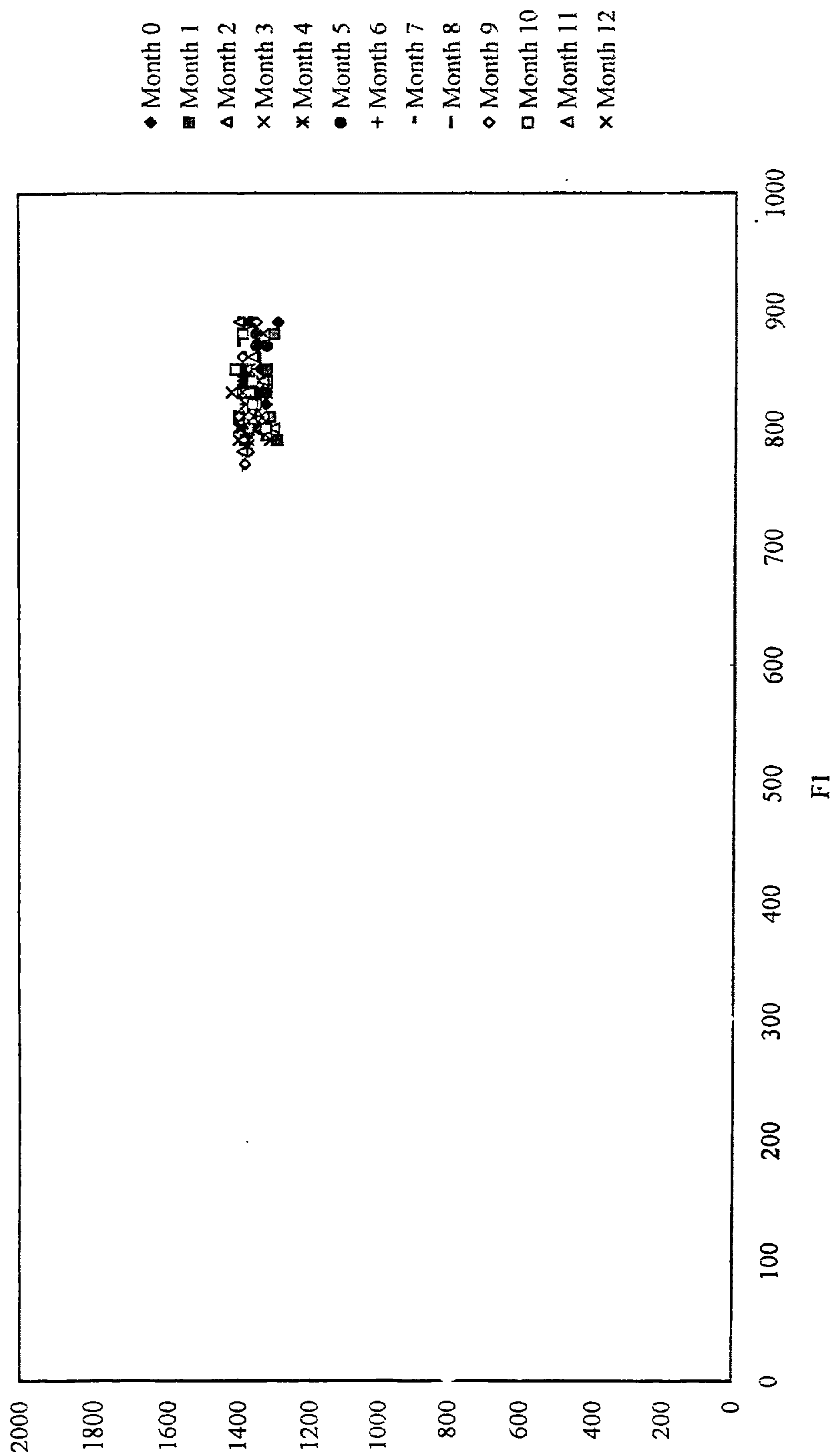

Figure 4.8 Monthly data for *SK*

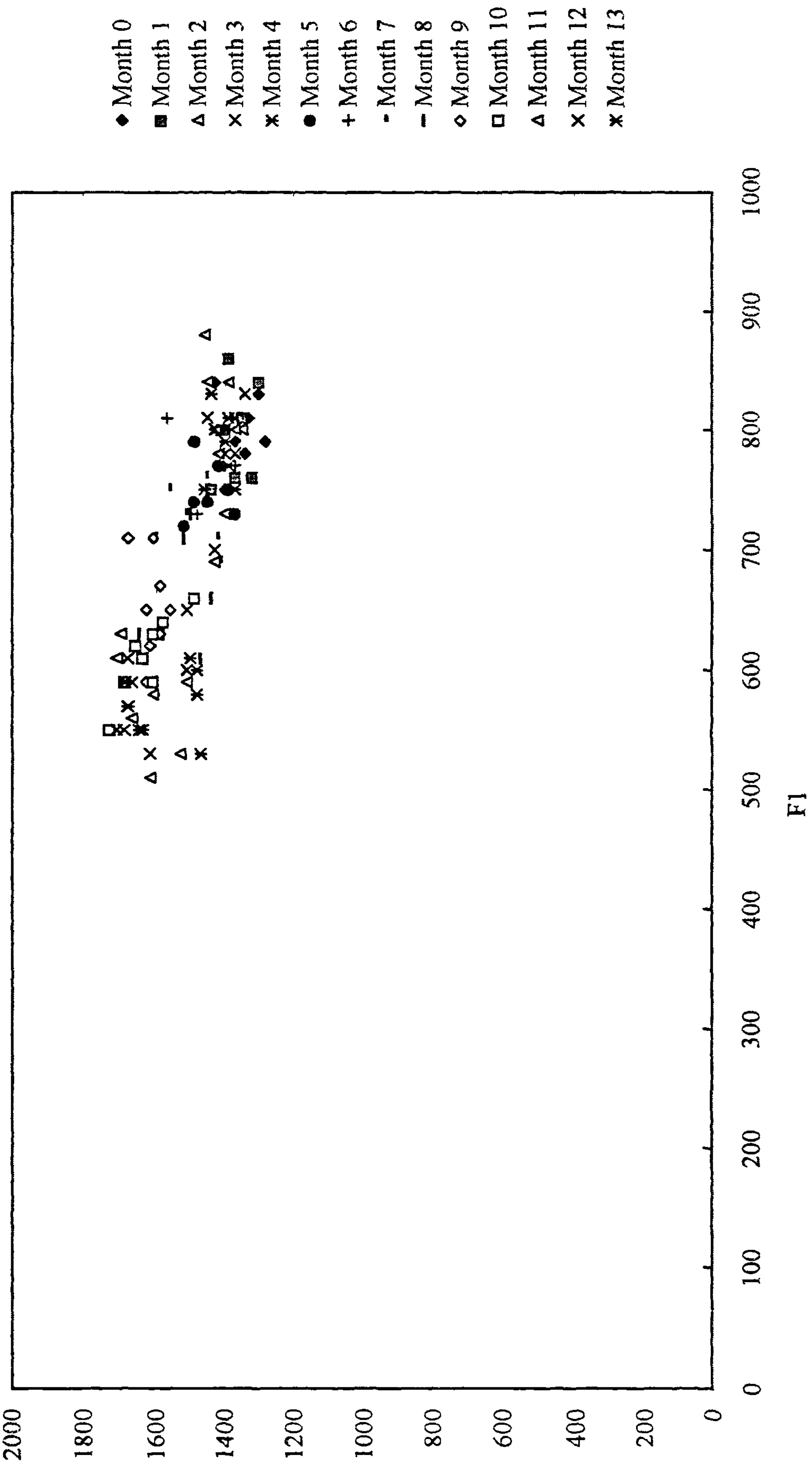

Figure 4.9 Monthly data for *MO*

As you can see from Figures 4.7 to 4.9, all the subjects started in the early months of the data collection off with formant values very similar to their Japanese vowel [a], namely between months 0-2 and months 3-5. See Table 4.2 for mean formant values for their interlanguage data and their native language data. It should be noted that for all three subjects, the mean formant values for schwa and for Japanese [a] did not show statistically significant difference with a mean formant values for F1 and F2.

Table 4.2 Mean formant values for Japanese [a] and interlanguage [ə]

	NI		SK		MO	
	F1	F2	F1	F2	F1	F2
Japanese vowel: [a]	724	1289	832	1370	797	1367
Month 0-2	725	1289	834	1348	802	1379
Month 3-5	721	1314	833	1362	781	1414
Target [ə] [From P & B (1952)]	575	1700				

Although starting off quite similar, subjects exhibited very different developmental patterns later on in the data collection; while NI and MO showed dramatic development in the later stages, SK made no progress in her acquisition of schwa, as can be observed from Figure 4.8. NI and MO showed similar developmental patterns in that their F1 moved towards a lower value while their F2 value moved upwards, to values which are typical for English schwa. See Table 4.3. for the mean formant values for each data collection period for NI and MO.

Table 4.3 Mean formant values for each data collection period

	NI		MO	
	F1	F2	F1	F2
(1) Month 0-2	725	1289	802	1379
(2) Month 3-5	721	1314	781	1414
(3) Month 6-8	666	1512	715	1508
(4) Month 9-12	614	1563	613	1610

When we look at the overall difference between NI and MO, it can be observed that MO gradually moved closer to the target schwa values between data collection periods, i.e. as the months passed, she constantly got closer to the target value. On the other hand, NI had a similar production pattern for five months and then suddenly makes rapid progress such that she has acquired the target values successfully by month 8.

Next, I will focus on the data to examine the extent of variability as a function of context in the first and second formant trajectories for NI (Figure 4.10), SK (Figure 4.11) and MO (Figure 4.12). When the Figure 4.10 and Figure 4.12 are compared, we can see that at the early stage of the data collection for MO, there was a medium-sized spread in both F1 and F2 values.

This tendency was observed throughout the data collection period for her. On the other hand, during the first three months there was very little spread in the F1 and F2 values for NI. In months three to five, this pattern repeated with some tokens produced with higher F2 values. From month six onwards, this pattern changes dramatically and NI begins to have a wide spread of formant values in F1 and F2. As discussed in 3.1.3, a wide spread of formant values in F2 is a characteristic of the reduced vowel in English.

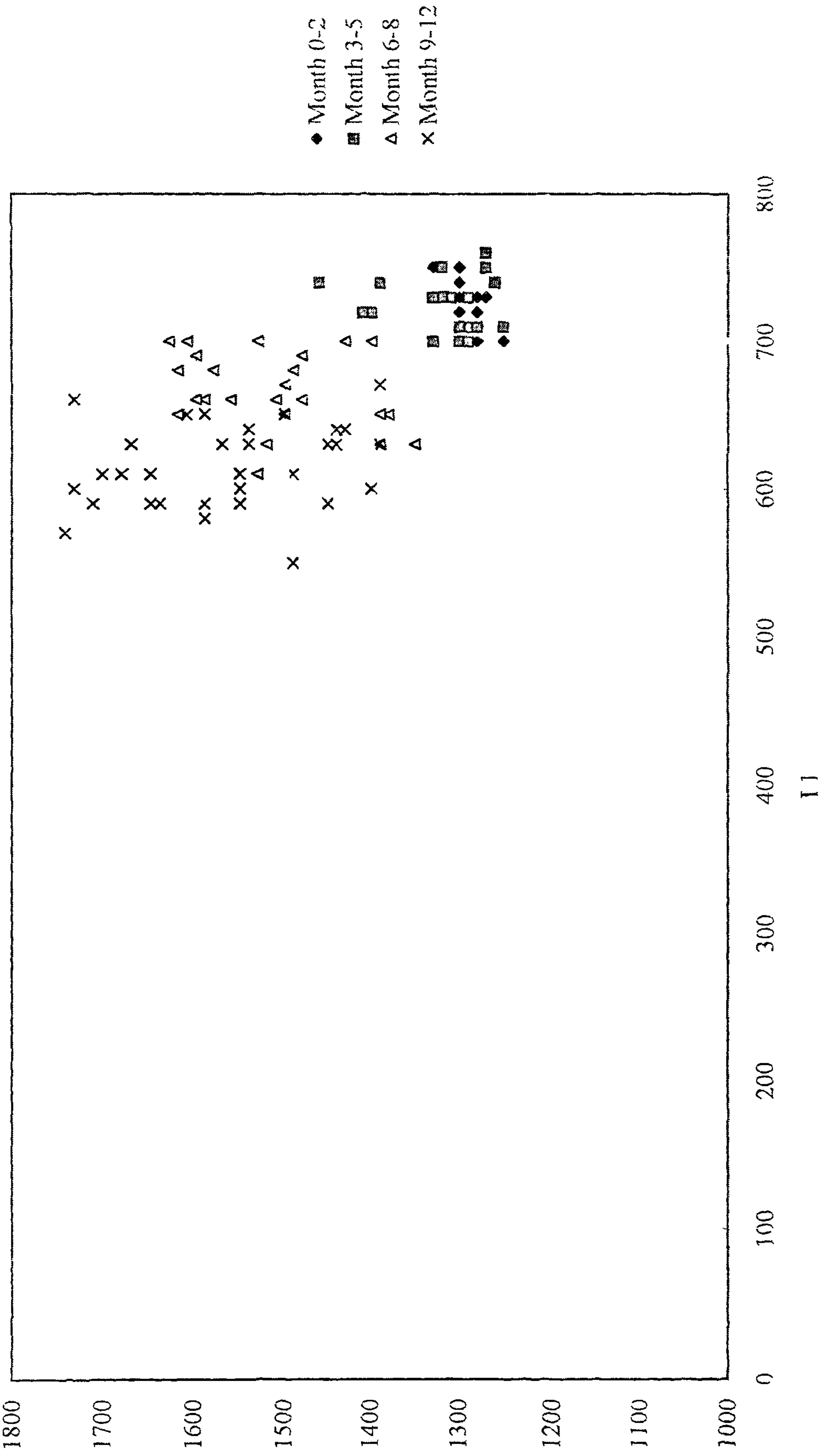

Figure 4.10 *NI*'s data (focused on)

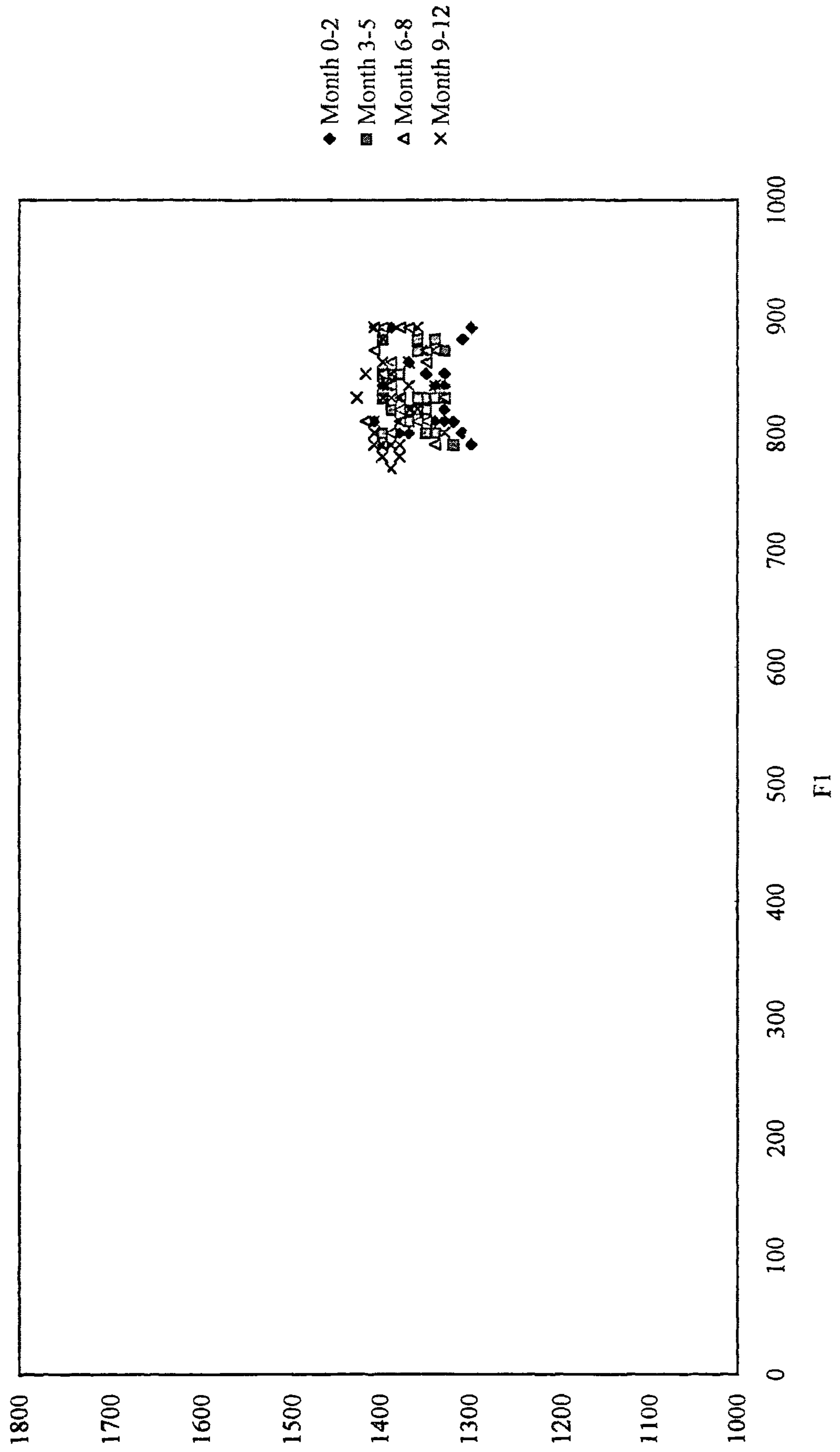

Figure 4.11 *SK*'s data (focused on)

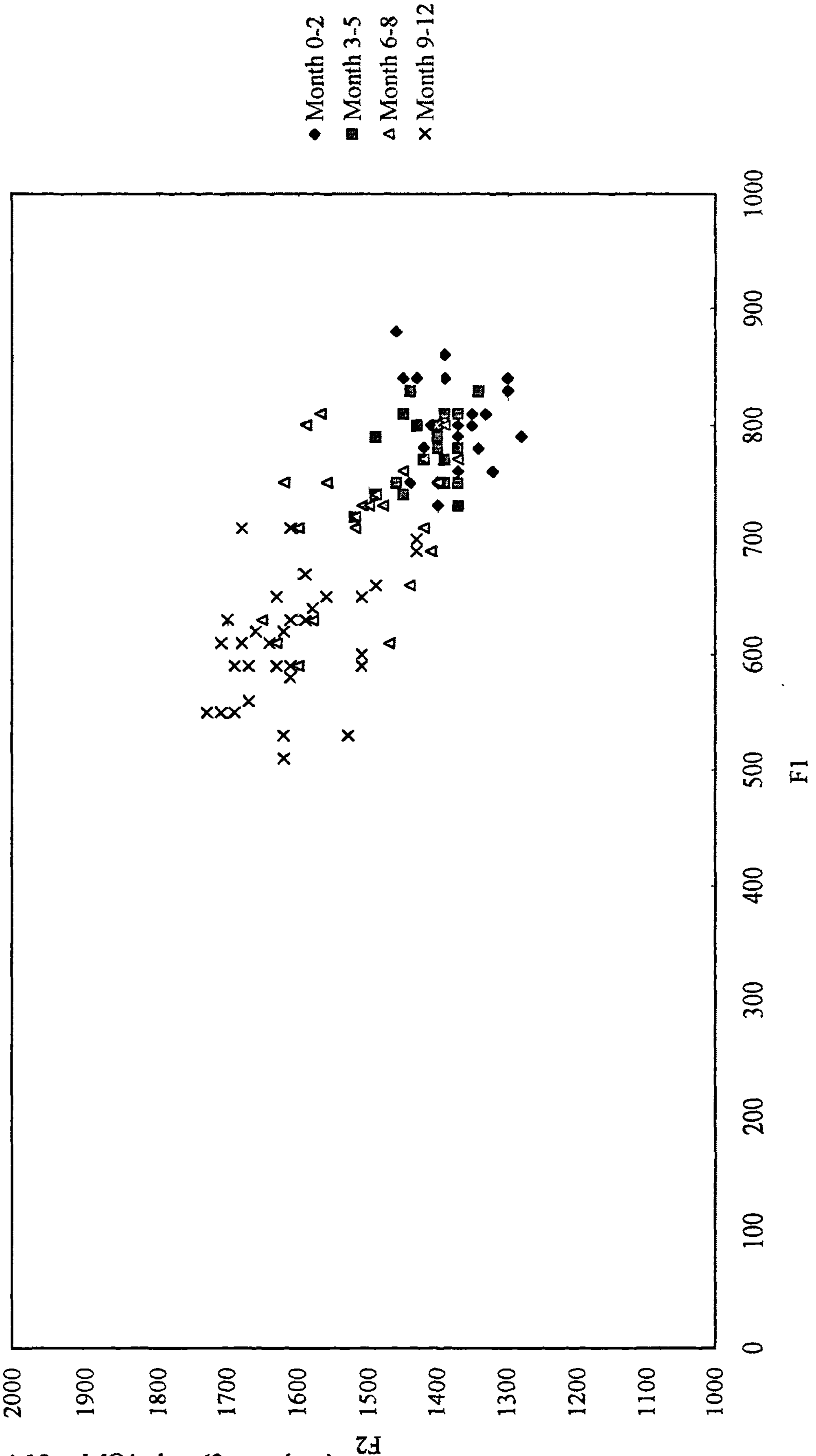

Figure 4.12 *MO*'s data (focused on)

We can have a closer examination of the extent of variation in F1 and F2 by looking at the difference between the minimum and the maximum formant values, as was done in Kondo's interlanguage study. Table 4.4 shows the difference between the minimum and maximum formant values in F1 for the three subjects, and Table 4.5. shows that of F2. As you can see from the tables, the extent of variation of F2, but not F1, in schwa was large for both NI and MO.

Table 4.4 The difference between the minimum and maximum formant values in the F1

F1	Month 6-8			Month 9-12		
	Minimum	Maximum	Difference	Minimum	Maximum	Difference
NI	610	700	90	550	670	120
SK	790	890	100	770	890	120
MO	590	810	220	510	710	200

Table 4.5 The difference between the minimum and maximum formant values in the F2

F2	Month 6-8			Month 9-12		
	Minimum	Maximum	Difference	Minimum	Maximum	Difference
NI	1350	1630	280	1390	1740	350
SK	1330	1420	90	1330	1430	100
MO	1370	1650	280	1430	1730	300

When the data obtained is compared with Kondo's data, the difference between the minimum and maximum formant values in F2 is clearly smaller in this study. In Kondo's study, the extent of variation of F2 in schwa was far larger at 817 Hz for speaker AH, 1038 Hz for speaker MB, and 1099 Hz for speaker DG. I speculate that the difference comes from the number of contexts the test words were based on. In her experiment the contextual consonants were /p, t, k/ and the vowels were /ɪ, æ, u/, therefore, 3 x 3 combinations were possible. In my study, there were only four test words. However, what can be observed from this study is *NI* shows a wide spread in F2 compared to F1.

To sum up, the results suggest that all three learners were under a strong influence of the Japanese vowel [a] at the beginning of the data collection. *SK* showed very little improvement in her acquisition of schwa in English. As for *NI*, after half a year of exposure to native-accented-input she began to produce formant values appropriate of English [ə]. Moreover, it was found that she has acquired the targetlessness of formant values in the F2 as native speakers and advanced learners were found to do in Kondo's (1995) English and interlanguage study. *MO* gradually got closer to producing F1 and F2 values appropriate for English schwa but didn't seem to have acquired the targetlessness of schwa in F2.

4.4.3 Production study II: timing of falling pitch

4.4.3.1 Methods

Recall that in 3.2.2, we saw that differences in rhythmic structure were a result of differences at the structural level. The typological differences between a stress-timed, syllable-timed and mora-timed language are not clear cut. Rather, all languages fall onto a continuum, depending on how important a role stress and the other structural features play in that language. I also examined how English and Japanese are similar when it comes to marking prominence. It was concluded that while there are differences between the role that accent plays in English and in Japanese, accent in both languages seems to serve as a marker of prominence.

As Beckman (1986) proposed, English and Japanese are similar in with respect to pitch pattern. A comparison of fundamental frequency measurements of disyllabic words in Japanese and English by Beckman revealed that the falling and rising intonation contours were very similar to each other in contrasts between initial vs. final stressed words in English (e.g. 'cóntrast' vs. 'contrást') and Japanese initial vs. final accented words in Japanese (e.g. *kata* 'shoulder' (HL) vs. *kata* 'form' (LH)).

However, although the pitch pattern is similar for both languages, a difference exists in the actual timing of the pitch pattern In a word such as *summit* (HL), the falling pitch starts from the first syllable, while in Japanese for a word with penultimate accent, such as *kame* (HL) "turtle", the falling pitch starts from the beginning of the second syllable.

The test words examined in this study involved four nouns (e.g. calibre, parapet) with a HLL accentuation pattern. The purpose of the study was to examine how learners' mark prominence, or more specifically, the timing of the falling pitch in a HLL accentuation pattern, and whether they proceed to acquire timing appropriate for English.

4.4.3.2 Results

At the beginning of the data collection, it was found that the learners' timings of falling pitch in English was late relative to English native speakers, suggesting transfer of their Japanese timing. Takefuta & Black (1977) reported that Japanese learners have a tendency to overly depend on pitch and raise/lower it; this was not observed in my data. Rather, learners tended *not* to lower the pitch enough, as well as not lowering it soon enough. This tendency was found for all test words. For example, Figure 4.13 shows *SK*'s production of *calibre* at month 0, and Figure 4.14 shows *NI*'s production of *calibre* at month 0. It can be seen that their falling pitch starts from the *second* syllable, and not from the *first* syllable. In Figures 4.13 and 4.14, the data show learners did not have aspiration in their production. This is also clearly transfer from Japanese, which has negligible aspiration in word initial voiceless obstruents.

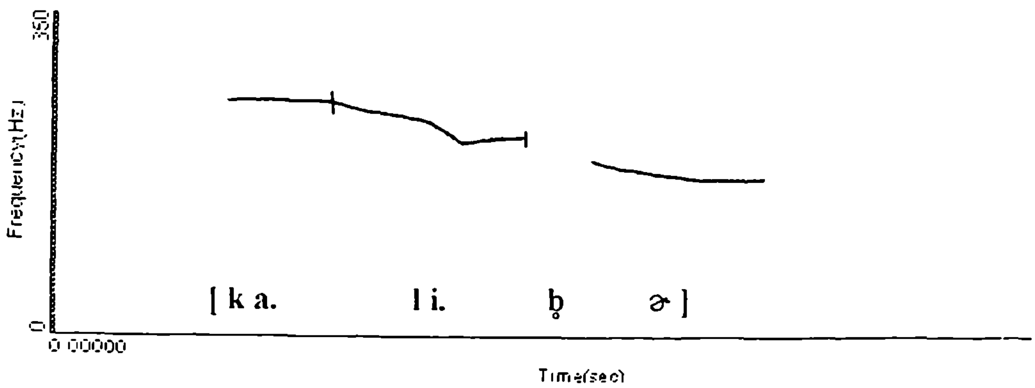

Figure 4.13 *SK* (Month 0) *calibre*

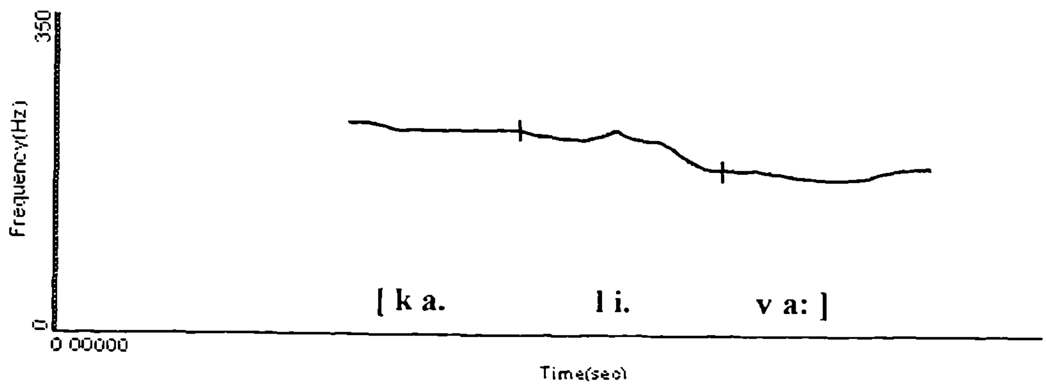

Figure 4.14 *NI* (Month 0) *calibre*

For comparison, see Figure 4.15 for an English native speaker's production of *calibre*: the falling pitch starts from the *first* syllable, while a Japanese native speaker producing a Japanese word similar to *calibre*, *camera* 'camera' (Figure 4.16), starts the falling pitch from the *second* syllable.[12]

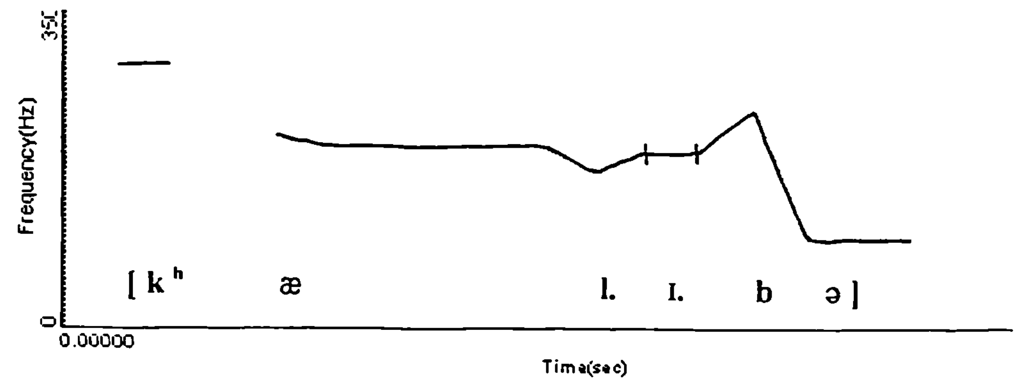

Figure 4.15 English native speaker's produciton of *calibre*

[12] I have no good explanation as to why word initial [k] in Figure 4.15 (English native speaker data) and Figure 4.17 (*MO*'s data at month 12) show voicing.

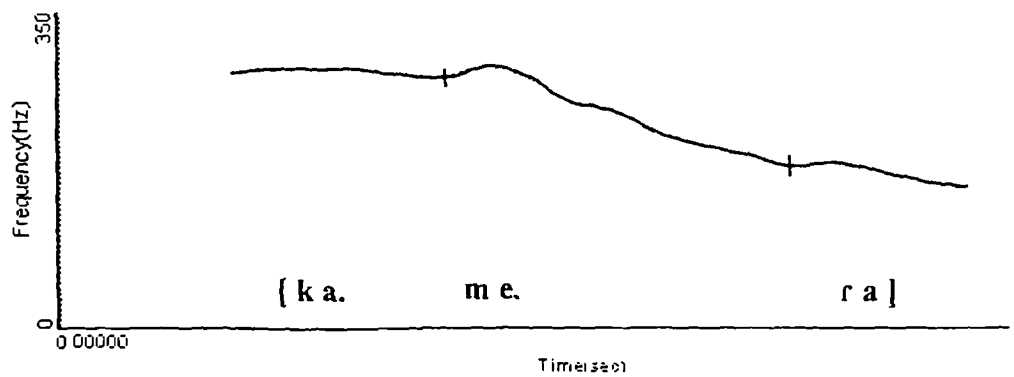

Figure 4.16 Japanese native speaker producing a Japanese word *camera*

After 5-7 months of exposure to native input, *NI* & *MO*'s timing of the falling pitch became target-like for English. This tendency was found for all of the test words. See for example, Figure 4.17 *NI* (month 6) *calibre*, and Figure 4.18 *MO* (Month 12) *calibre*: they started to get the timing correct. On the other hand, *SK* continued to show transfer from her Japanese timing for pitch. This tendency was repeated for the rest of the test words. The following is a chart indicating when learners started to produce test items in a target manner.

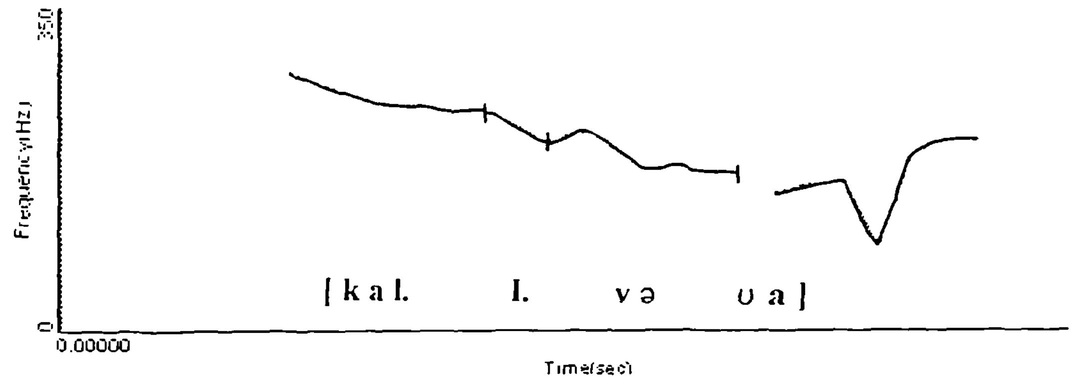

Figure 4.17 *NT*'s production of calibre at month 6

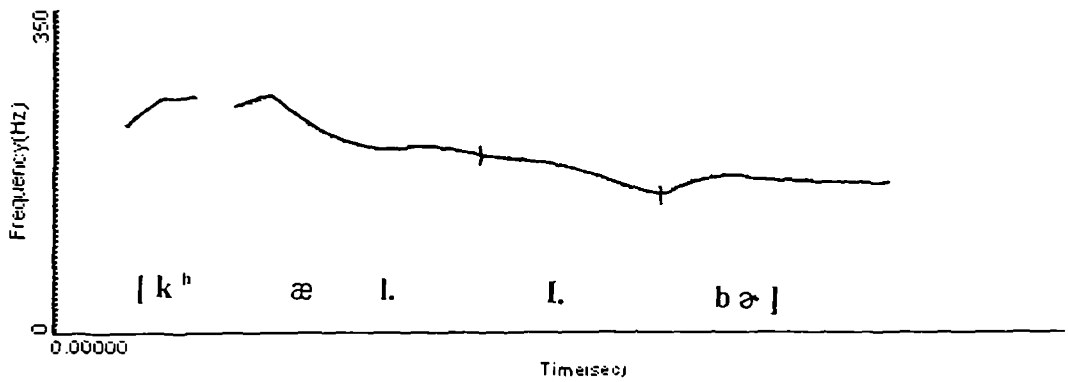

Figure 4.18 *MO*'s production of calibre at month 12

4.5 Discussion

4.5.1 The perception study

Data from the perception study suggest that adult L2 learners who have received years of primarily foreign-accented input can improve their L2 phonology once they are in the target language setting and exposed to abundant native-speaker-accented input. It is possible to reach a higher level of proficiency with exposure to native-speaker-accented input. In fact, of the three learners, the one with the lowest proficiency level upon her arrival (*NI*) demonstrated the most improvement in all three sub-components of phonology (i.e. she eventually caught up with the other higher proficiency learners for all three areas of phonology).

The results were consistent with Leather & James' (1991) study claiming that at the segmental level, transfer is prominent only during the early stages of acquisition, while transfer persists well into advanced stages of acquisition with respect to prosodic structure: *NI* showed improvement in segments and syllables from a very early stage of the data collection, while it was not until month 9 that she started to show development in prosody. This tendency was repeated in *MO*'s data with respect to segments and prosody: she showed improvement in segments from an early stage, but for prosody, she had a mysterious backslide. I will not discuss *SK* here, because she showed improvements only at the syllable level (which I attempt to account for later in this sub-section).

Although exposure to native-speaker-accented-input seems to induce improvement, it is possible that there is a fossilisation ceiling. For syllables, *MO* started off with the highest score among the three learners. Possibly because of this, she showed only a little development. *SK* did show some improvement for syllables at the beginning, but later plateaued at $\bar{x}$ = 84.7%. If we assume that there is a fossilisation ceiling (FC), perhaps around 85%, then these messy results start to make sense. For segments, *NI* and *MO* started off from below the FC, and continued to improve their perceptual ability in English. Likewise, for prosody, *NI* started from below the FC and continued to show improvement (here, *MO* shows a very different performance from the other two learners as stated below). For syllable structure, *MO* had probably hit the ceiling point before she arrived in the TL setting, therefore her phonology showed little change during the data collection. *SK* started below the FC on arrival and showed improvement until she hit the FC, when she plateaued. *NI*, on the other hand, started far below the FC and her acquisition showed a sharp upward curvilinear trend.

However, there are two inconsistencies which the FC is unable to account for, i.e. that *SK* plateaued at the segmental level and that *MO* exhibited backsliding in prosody. The reason *SK* plateaued at the segmental level and its consequence for other sub-components of phonology in acquisition will be discussed in 5.1.3 in connection with the production data. As for *MO*, she exhibited a clear backslide in prosody for ungrammatical items: she first scored 62.9%, more or less the same rate as *SK*, but scored increasingly worse as the months went by and ended with a 20% accuracy rate for ungrammatical forms. She seems to have resorted to the strategy of answering "correct" for either grammatical items *or* ungrammatical items. This may be one of the drawbacks of giving subjects the same test over and over. Nonetheless, it seems clear that *MO* had not acquired English accentuation by the end of study. The reasons behind this could be: 1) she hasn't acquired the prosodic rules of English accentuation at all; 2) we could interpret this confusion as a first step for her to reconstruct her interlanguage English prosody system.

Whether learners can further develop their L2 perception ability above the observed ceiling is an open question, which can not be answered with a one-year study. In their production, the learners did not show any trace of a fossilisation ceiling. Further research is awaited.

4.5.2 The production studies

Can L2 learners with years of non-native-accented input improve their phonology when they are exposed to native-speaker-accented input? If we look at study I, the production study, results show that although the learners' prosodic features continued to be distinctly non-native throughout the data collection, they did exhibit changes over time. Two of the subjects (*NI* & *MO*) started to produce formant frequencies appropriate for a target schwa and one of the subjects, *NI*, seemed to have acquired the targetlessness of F1.

Results from the second production study show that although at the beginning of the data collection all three learners' timing of falling pitch in English were late relative to English native speakers, in the later stage of the data collection, *NI* and *MO*'s timing of falling pitch became target-like for English. The falling pitch started from the *first* syllable, rather then the *second* as in their native language Japanese.

How, then, did the acquisition take place? What triggered acquisition? Recall that *NI* & *MO* succeeded in acquiring vowel reduction, and also got the timing correct for the target language. I would like to suggest that these seemingly separate processes are connected and suggest that the acquisition of vowel reduction is linked to correct timing, revealing the interaction of a segmental and a prosodic process. It is therefore not surprising that *SK* failed to reduce vowels and also exhibited very little change over time in prosody. In fact, the data show that acquisition of vowel reduction precedes acquisition of timing and this tendency was consistent with the rest of the test words that had a HLL accentuation pattern, as seen in Table 4.6 to Table 4.9.

Table 4.6 Acquisition of vowel reduction and timing (production of *calibre*)

	Vowel reduction	Acquisition of timing of falling pitch
NI	Month 5	Month 6
MO	Month 7	Month 10
SK	—	—

Table 4.7 Acquisition of vowel reduction and timing (production of *parapet*)

	Vowel reduction	Acquisition of timing of falling pitch
NI	Month 5	Month 7
MO	Month 5	Month 6
SK	—	—

Table 4.8 Acquisition of vowel reduction and timing (production of *cantaloup*)

	Vowel reduction	Acquisition of timing of falling pitch
NI	Month 7	Month 8
MO	Month 6	Month 6
SK	—	—

Table 4.9 Acquisition of vowel reduction and timing (production of *matador*)

	Vowel reduction	Acquisition of timing of falling pitch
NI	Month 7	Month 9
MO	Month 6	Month 8
SK	—	—

With respect to the idea that the acquisition of vowel reduction triggers acquisition of timing, I speculate that L2 learners perceive it as unnatural to place the falling pitch on a syllable which does not have full quality & length, which in turn means that they will either place the falling pitch earlier (i.e. on the first syllable) or delay it until the last syllable. Placing the accent (falling pitch) on the last syllable is unlikely because: 1) the L1, Japanese, does not prefer that structure 2) typologically speaking, word-final accent is a marked option (Prince & Smolensky 1993). This of course does not exclude the existence of such languages. A well known example is French. In fact, when part of this thesis was presented, I was informed that 12% of natural languages have word final stress. However, the fact that word final stress is a marked option in world languages remains. Therefore, learners are likely to choose to put the accent (falling pitch) on the first syllable when they acquire vowel reduction, which is the target form in English.

It should be added that acquisition of reduced vowels seems to play an important role in L1 acquisition as well. According to Allen & Hawkins (1980), one of the first rhythmically important skills the child must learn in order to produce fluent English phrases is that of producing adult-like weak syllables. Young children's ability to reduce weak syllables adequately develops much later than their production of stressed syllables. By age three, heavy syllable nuclei are acquired, but the acquisition of light nuclei is delayed until the age of four to five. Two year olds tend to use far fewer reduced syllables than do adults, so that their speech rhythm has fewer syllables per foot, or more beats per utterance: in short it sounds more syllable-timed.

In summary, at least for Japanese EFL learners studied in this study, it was found that two of the subjects who managed to improve segmental features (i.e. vowel reduction), succeeded in improving one aspect of English prosodic features (i.e. timing of falling pitch). The third learner *SK*, who did not acquire vowel reduction did not succeed in to acquiring the timing of falling pitch. It might be the case that acquisition of vowel reduction is a necessary condition for acquiring falling pitch, or at least we might say that the acquisition of the former triggers the latter.

4.5.3 Perception vs. production

Differences seem to exist in the way individual learners perceptually process the L2 input and the differences seem to result in improvement or lack of improvement in production.

NI exhibited dramatic improvement in segments at an early stage of the data collection and showed improvement in all three sub-components of phonology. *SK*'s perception data reveals that she plateaued at the segmental level and also at the prosodic level.[13] The results indicate that perception and production is related and it is not the case that they are two separate processes which are acquired completely independently.

[13] For future studies, I will investigate how syllable level processes interact with segmental and prosodic processes. To investigate this point, I would look at test words which contain consonant clusters and have a HLL accentuation pattern. In this case, learners will have to get the (1) syllable structure correct; (2) vowel quality & length correct; and (3) timing of the falling pitch correct. In what order do learners get them correct? Furthermore, to obtain a full picture of phonological L2A, I will investigate the correlation of production and perception ability at the phonological level. This includes whether learners' perception precedes production at various phonological levels.

Table 4.10 Summary of the results of the perception data and the production data

		perception		production
NI	segments	overall upward curvilinear trend	⇗	⇗
	syllables	a dramatic improvement	⤴	—
	prosody	improvement	⇗	⇗
MO	segments	mild upward curvilinear trend	⇗	⇗
	syllables	highest score first, later plateaued	⇗ ⇨	—
	prosody	some improvement, later, no data	—	⇗
SK	segments	no improvement	⇨	⇨
	syllables	some improvement, later plateaued	⇗ ⇨	—
	prosody	very little improvement	⇨	⇨

In summary, adult Japanese L2 learners were studied as they continued to develop their L2 in a target language setting. The nature of phonological acquisition and the effect of such input on learners over time was examined. Three main conclusions were reached. It was concluded that it is possible for L2 learners who have mainly been exposed to non-native-speaker-accented input during their initial foreign language experience to improve their L2 phonology when they are in a target language setting exposed to native-speaker-accented input. It is suggested that the acquisition of vowel reduction is the key to the L2 acquisition of the timing of falling pitch in the production of English, revealing the interaction of a segmental process and a prosodic process. Finally, it is argued that perception and production are related and it is not the case that they are two separate processes which are acquired completely independently.

Chapter 5

Summary and Conclusion

A longitudinal study was carried out to examine the effect of primary linguistic data over time on learners who had received a limited amount of input, often foreign-accented.

In Chapter 1, factors which were suggested to be responsible for the incomplete attainment of L2 phonology were examined. I first looked at studies which point to age as the factor responsible for incomplete attainment in L2 phonology, and then reviewed the literature suggesting there exist neuro-biological constraints on adult L2A. I then looked at studies which point to a profound effect for socio-psychological factors. Finally, I discussed to the importance of abundant exposure to native-accented-input for successful L2A.

In Chapter 2, how biological maturation affects L2 phonology was further discussed, focusing on perception. First, I examined studies which show that perception and production are related, and that perception precedes production, which led to the idea that the difficulty L2 learners experience with L2 speech may have its roots in their difficulty in perceiving the L2 sounds. I then discussed the development of perceptual abilities in first language acquisition and reviewed literature which suggests *mis-perception* of target language (TL) input is a possible cause for L2 learners' failure to achieve native proficiency. I then discussed implications for pronunciation teaching. Prosodic characteristics may be one of the most important features in the perception of foreign accent and perceived foreign accent is in turn closely tied to the reduced intelligibility of accented speech for native speakers.

In Chapter 3, the phonologies of English and Japanese were compared. For segmental features, it was discussed that English schwa is underspecified in the second formant trajectories, in other words, targetless in F2. I also reviewed studies which discuss variability in vowels in English and in Japanese. This led to implications of English targetless schwa for Japanese learners. I next looked at suprasegmental features, namely syllable structure, rhythmic style and pitch-accent in English and Japanese, and then considered the implications of English suprasegmental features for Japanese learners.

In Chapter 4, a longitudinal study which examined the effect of native-accented-input on learners who had received a limited amount of input, often foreign-accented was introduced. Data were collected from three Japanese EFL learners starting right after they arrived in the U.K. and continuing on a monthly basis for over a year. Subjects regularly performed several tasks testing their perception and production.

Data from the perception study suggested that adult L2 learners who have received years of primarily foreign-accented input can improve their L2 phonology once they are in the target language setting and exposed to native-speaker-accented input, i.e. it is possible for adults to reach a higher level of proficiency with exposure to native-speaker-accented input. In fact, of the three learners, the one with the lowest proficiency level upon her arrival (*NI*) demonstrated the most improvement in all three sub-components of phonology; she eventually caught up with the other higher-proficiency learners for all three areas of phonology.

Although exposure to native-speaker-accented-input seems to induce improvement, it is still possible that there is a 'fossilisation ceiling'. For syllables, *MO* started off with the highest score among the three learners, and possibly because of this, she showed only a little development. *SK* did show some improvement for syllables at the beginning, but later plateaued at $\bar{x}$= 84.7%. It was suggested that there might exist a fossilisation ceiling (FC), perhaps around 85%. For example, for segments, *NI* and *MO* started off from below the FC, and continued to improve their perceptual ability in English. Likewise, for prosody, *NI* started from below the FC and continued to show improvement (here, *MO* shows very different performance from the other two learners, as stated below). For syllable structure, *MO* had probably hit the ceiling before she arrived in the TL setting, therefore her phonology showed little change during the data collection. *SK* started below the FC on arrival and showed improvement until she hit the FC, when she plateaued. *NI*, on the other hand, started far below the FC and her acquisition showed a sharp upward curvilinear trend. Whether learners can further develop from their L2 perception ability above the observed ceiling is an open question which can not be answered with data from only a one-year study.

Results from the two production studies, one looking at the learners' acquisition of vowel reduction and the other at their acquisition of the timing of falling pitch, also show that primary linguistic data from native speakers has a positive effect on L2 learners. If we look at production study I, the results show that, although the learners' production continued to be distinctly non-native throughout the data collection, they did exhibit changes over time. Two of the subjects (*NI* & *MO*) started to produce formant frequencies appropriate for English schwa and one of the subjects, *NI*, seemed to have acquired the targetlessness of F2. Results from the second production study also show that native-speaker input has a positive effect on L2 learners. At the beginning of the data collection, all three learners' timing of falling pitch in English was late relative to English native speakers'. However, it was found that in the later stage of the data collection, *NI* and *MO*'s timing of falling pitch became target-like for English. The falling pitch started from the *first* syllable, rather then the *second* as in their native language Japanese.

The results from the production study I also confirm the prediction that the production of schwa by Japanese speakers of English requires the acquisition of a new coarticulatory strategy: Japanese speakers of English need to learn to produce targetless schwa and thus contrast targeted and targetless vowels. The results from the study further support Kondo's finding that non-fluent speakers show strong transfer from their L1 to the L2 production of schwa and do not show the targetlessness in F2, while the production of schwa by fluent L2 speakers was characterised as being very similar to the English native speakers' production, which in turn was characterised systematically and largely as a function of context, i.e. targetlessness in F2. The study reported on in this thesis confirms that the two different patterns observed in Kondo's study for fluent and non-fluent L2 speakers of English represent different developmental stages.

I have therefore suggested that the seemingly separate processes, namely the acquisition of vowel reduction and the acquisition of correct timing are connected. It might be the case that the acquisition of vowel reduction is a necessary condition for acquiring falling pitch, or at least one might say that the acquisition of the former triggers the latter.

Differences seem to exist in the way individual learners perceptually process the L2 input, and these differences seem to result in improvement or lack of improvement in production. The results from the perception and production studies indicate that perception and production are related; it is not the case that they are two separate processes which operate independently.

Furthermore, the results from the perception study provided support for the claim that L1 transfer persists well into advanced stages of acquisition with respect to suprasegmental/prosodic features, and is most resistent to change, as was claimed in Leather & James (1991).

The results further highlight the importance of prosody in L2 phonology and the need for explicit instruction in this area. Taken together (1) prosodic characteristics are some of the most important features in the perception of foreign accent and perceived foreign accent results in reduced intelligibility; (2) learners' overall pronunciation of their second language and their perception ability are related, which suggests that (3) perception training in prosody may be effective.

As to how prosody can be best taught, further research is needed. We need to investigate the nature of phonological problems in specific L1- L2 contact situations and suggest what should be included in specific teaching programmes, how prosody should be presented and to empirically test the methods used to develop the most effective pronunciation teaching programmes for post-puberty L2 learners.

References

Abercrombie, D. (1964). A phonetician's view of verse structure. *Linguistics* 6, 5-13.

Abercrombie, D. (1967). *Elements of general phonetics.* Chicago: Aldine Publishing Company.

Adams, C. (1979). *English speech rhythm and foreign learners.* The Haugue: Mouton.

Allen, G. D. (1975). Speech rhythm: its relation to performance universals and articulatory timing. *Journal of Phonetics* 3, 75-86.

Anderson-Hsieh, J., Johnson , R. & Koehler, K. (1992). The effects of foreign accent and speaking rate on native judgements of non-native pronunciation and deviance in segmentals, prosody and syllable structure. *Language learning* 42, 529-555.

Archangeli, D. & Pullyblank, D. (1994). *The content and structure of phonological representations.* Cambridge, Mass.: MIT Press.

Archangeli, D. (1984). *Underspecification in Yawelmani Phonology and Morphology.* Ph.D. thesis. MIT. Published 1988, New York: Garland.

Archangeli, D. (1988). Aspects of underspecification theory. *Phonology* 5, 183-207.

Archibald, J. (1992). Transfer of L2 parameter settings: some empirical evidence from Polish metrics. *Canadian Journal of Linguistics* 37, 301-339.

Archibald, J. (1993). *Language learnability and L2 phonology: the acquisition of metrical parameters.* Dordrecht: Kluwer.

Archibald, J. (1994). A formal model of learning L2 prosodic phonology. *Second*

Language Research 10, 211-243.

Archibald, J. (1995a). *Phonological acquisition and phonological theory.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Archibald, J. (1995b). A longitudinal study of the acquisition of English stress. *Calgary Working Papers in Linguistics* 17, 1-10. The University of Calgary, Canada.

Asher, J. J. & Garcia, R. (1969). The optical age to learn a foreign language. *Modern Language Journal* 53, 334-342.

Aslin, R. N., Pisoni, D. B., Hennessy, B. L & Perey, A. J. (1981). Discrimination of voice onset time by human infants: new findings and implications for the effect of early experience. *Child Development* 52, 1135-1145.

Atal, B. & Hanauer, S. (1971). Speech analysis and synthesis by linear prediction of the speech wave. *Journal of the Acoustical Society of America* 50, 637-655.

Au, S. Y. (1988). A critical appraisal of Gardner's social-psychological theory of second-language (L2) learning. *Language Learning* 38 (1), 75-100.

Barry, W. (1992) Comments on chapter 2. In Cocherty, G. J. & Ladd, D. R. (eds.) *Papers in Laboratory Phonology II: Gesture, Segments, Prosody.* Cambridge: Cambridge University Press, 65-67.

Bates, S. A. R. (1995). *Towards a definition of schwa: an acoustic investigation of vowel reduction in English.* Ph.D. thesis. University of Edinburgh.

Beckman, E. M. (1986). *Stress and non-stress accent.* Dordrecht, Holland: Fortis Publications.

Beckman, M. & Pierrehumbert, J. (1986). Intonational structure in Japanese and English. *Phonology Yearbook* 3, 255-309.

Beckman, M. (1982). Segment duration and the 'mora' in Japanese. *Phonetica*

39, 113-135.

Beckman, M. E. (1986). *Stress and non-stress accents.* Dordrecht: Fortis. Netherlands Phonetic Archives 7.

Beebe, L. M. (1985). Input: choosing the right stuff. In Gass, S. M. & Madden, C. G. (eds.) *Input in second language acquisition.* Cambridge: Newbury House, 404-414.

Bergem, D. R. van. (1993). Acoustic vowel reduction as a function of sentence accent, word stress and word class. *Speech communication* 12, 1-23.

Bergem, D. R. van. (1994). A model of coarticulatory effects on schwa. *Speech Communication* 14, 143-162.

Bertinetto, P. M. (1988). Reflections on the dichotomy 'stress' vs. 'syllable-timing'. In Vayra, M. (ed.) *Quaderni del Laboratorio de Linguistica* 2. Scuola Normale Superiore Pisa.

Berwick, R. (1985). *The acquisition of syntactic knowledge.* Cambridge, Mass.: MIT Press.

Best, C. T. (1994). The emergence of native-language phonological influences in infants: a perceptual assimilation model. In Goodman, J. C. & Nusbaum, H. C. (eds.) The development of speech perception: the transition for speech sounds to spoken words. Cambridge, Mass.: MIT Press, 167-224.

Best, C. T., McRoberts, G.W. & Sithole, N. N. (1988). The phonological basis of perceptual loss for nonnative contrasts: maintenance of discrimination among Zulu clicks by English-speaking adults and infants. *Journal of Experimental Psychology: Human Perception and Performance* 14, 345-360.

Bever, T. (1981). Normal acquisition processes explain the critical period for language learning. In Diller, K. (ed.) *Individual differences and universals in* *language learning aptitude.* Rowley: Newbury House, 176-198.

Birdsong, D. (1999). *Second language acquisition and the critical period hypothesis.* London: Lawrence Erlbaum Associates.

Bley-Vroman, R., Felix, S. & Ioup, G. (1988). The accessibility of universal grammar in adult language learning. *Second Language Research* 4 (1), 1-32.

Bloch, B. (1942). Studies in colloquial Japanese IV: phonemics. *Language* 26, 86-125.

Bolinger, D. (1975). *Aspects of Language.* New York: Harcourt Brace Jovanovch.

Bolinger, D. (1958). A theory of pitch accent in English. *Word* 14, 110-149.

Bolinger, D. (1985). Two views of accent. *Journal of Linguistics* 21, 79-123.

Bongaerts, T. (1999). Ultimate attainment in L2 pronunciation: the case of very advanced late L2 learners. In Birdsong, D. (ed.) *Second language acquisition and the critical period hypothesis.* London: Lawrence Erlbaum Associates, 133-159.

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

Borzone de Manrique, A. M., Signorini, A. & Massone, M. I. (1982). Rasgos prosódicos: el acento. *Fonoloaudiológica* 28, 19-36.

Bradlow, A., Posoni, D., Yamada, R. & Tohkura, Y. (1996). Training Japanese listeners to identify English /r/ and /l/: some effects of perceptual learning on speech production. *Journal of the Acoustical Society of America* 101, 2299-2310.

Broselow, E. & Park, H-B. (1995). Mora conservation in second language prosody. In Archibald, J. (ed.) *Phonological Acquisition and phonological theory.*

Hove: Erlbaum, 151-168.

Browman, C. P. & Goldstein, L. (1992a). Articulatory phonology. An overview. *Phonetica* 49, 166-180.

Browman, C. P. & Goldstein, L. (1990). Tiers in articulator phonology, with some implications for casual speech. In Kingston, J. & Beckman, M. E. (eds.) *Papers in Laboratory Phonology I: Between the Grammar and Physics of Speech.* Cambridge: Cambridge University Press.

Browman, C. P. & Goldstein, L. (1992b). Targetless schwa: an articulatory analysis. In Cocherty, G. J. & Ladd, D. R. (eds.) *Papers in Laboratory Phonology II: Gesture, Segments, Prosody.* Cambridge: Cambridge University Press, 26-56.

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

Brown, C. & Matthews, J. (1993). The acquisition of segmental structure. McGill Working Papers in Linguistics 9, 46-76.

Brown, C. & Matthews, J. (1997). The role of feature geometry in the development of phonemic contrasts. In Hannahs, S. J. & Young-Scholten, M. (eds.) Generative studies in the acquisition of phonology. Amsterdam: John Benjamins, 67-112.

Catford, J. C. (1977). *Fundamental problems in phonetics.* Edinburgh: Edinburgh University Press.

Chamot, A. U. (1978). Grammatical problems in learning English as a third language. In Hatch, E. (ed.) *Second language acquisition: a book of readings.* Rowley, Mass: Newbury House, 176-189.

Champagne-Muzar, C., Schneiderman, E. I. & Bourdages, J. S. (1988). Second language accent: the role of pedagogical environment. *International Review of Applied Linguistics in Language Teaching* 31(2), 143-160.

Chao, Y. R. (1930). A system of tone letters. *Le Maître Phonétique, troisème série* 30, 24-7. Reprinted in Fangyan (1980).

Chihara, T. & Oller, J. W. (1978). Attitudes and attained proficiency in EFL: a sociolinguistic study of adult Japanese speakers. *Language Learning* 28, 55-68.

Choi, J. D. (1995). An acoustic-phonetic underspecification account of Marshallese vowel allophony. *Journal of Phonetics* 23, 323-347.

Chun, D. (1988). The neglected role of intonation in communicative competence and proficiency. *Modern Language Journal* 72, 295-303.

Clahsen, H. & Muysken, P. (1986). The availability of universal grammar to adult and child learners: a study of the acquisition of German word order. *Second Language Research* 2 (2), 93-119.

Classé, A. (1939). *The rhythm of English prose.* Oxford: Basil Blackwell.

Clément, R., Gardner, R. C. & Smythe, P. C. (1977). Attitudes and attained proficiency in EFL: a socio-linguistic study of adult Japanese speakers. *Language Learning* 28, 55-68.

Clements, G. N. (1985). The geometry of phonological features. *Phonology Yearbook* 2, 223-250.

Clements, G. N. (1987). Toward a substantive theory of feature specification. Paper presented at the UCLA Symposium on Segment Structure.

Cohn, A. C. (1993). Nasalisation in English: phonology or phonetics. *Phonology* 19, 43-81.

Corder, S. P. (1973). *Introducing applied linguistics.* Harmondsworth, Middlesex:

Penguin.

Crystal, D. (1985). *A Dictionary of Linguistics and Phonetics.* Oxford: Blackwell.

Cutler, A. & Norris, D. (1988). The role of strong syllables in segmentation for lexical access. *Journal of Experimental Psychology: Human Perception and Performance* 14, 110-121.

Cutler, A. (1976) Phoneme-monitoring reaction time as a function of preceding intonation contour. *Perception and Psychophysics* 20, 55-60.

Dalby, R. M. (1984). *Phonetic Structure of Fast Speech Processes in American English.* Ph.D. thesis. Indiana University.

Dauer, R. M. (1983). Stress-timing and syllable-timing reanalyzed. *Journal of Phonetics* 11, 51-62.

De Bot, K. & Mailfert, K. (1982). The teaching of intonation: fundamental research and classroom applications. *TESOL Quarterly* 16, 71-77.

Delattre, P. (1969). An acoustic and articulatory study of vowel reduction in four languages. *International Review of Applied Linguistics* VII /4, 295-325.

Dennis, J. & Scott, J. (1975). *Creole formation and reorganization.* Paper presented at the International Congress on Pidgins and Creoles, Honolulu, Hawaii.

Dresher, B. & Larkins, J. (1980). Non-semantic auditory discrimination: foundation for second language learning. *Modern Language Journal* 56, 227-230.

Dresher, B. E & Kaye, J (1990). A computational learning model for metrical phonology. *Cognition* 34, 137-195.

Dulay, H. & Burt, M. (1974). Natural sequences in child second language acquisition. *Language Learning* 24, 37-53.

Durand, J. (1990). *Generative and Non-linear Phonology.* London: Longman.

Dušková, L. (1969). On sources of errors in foreign language learning. *International Review of Applied Linguistics* 7, 11-36.

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

Eilers, R. E., Gavin, W. J. & Oller, D. K. (1982). Cross-linguistic perception in infancy: early effects of linguistic experience. *Journal of Child Language* 9, 289-302.

Elias-Olivares, L (1976). *Ways of speaking in a Chicago community: a sociolinguistic approach.* Unpublished Ph.D. thesis. University of Texas at Austin.

Elliott, A. (1995). Field independence/dependence, hemispheric specialization, and attitude in relation to pronunciation accuracy in Spanish as a foreign language. *Modern Language Journal* 79, 356-371.

Engstrand, O. (1988). Articulatory correlates of stress and speaking rate in Swedish vcv utterances. *Journal of the Acoustical Society of America* 83, 1863-1875.

Felix, S. (1980). Interference, interlanguage and related issues. In Felix, S. (ed.) *Second language development.* Tubingen: Gunter Narr, 93-107.

Flege, J. E. & Hillenbrand, J. (1984). Limits on phonetic accuracy in foreign language speech production. *Journal of the Acoustical Society of America* 76 (3), 708-721.

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

Flege, J. E. (1981). The phonological basis of foreign accent. *TESOL Quarterly* 15, 443-455.

Flege, J. E. (1987). A critical period for learning to produce a foreign language?

Applied Linguistics 8, 162-177.

Flege, J. E. (1991). Age of learning affects the authenticity of Voice-Onset Time (VOT) in stop consonants produced in a second language. *Journal of the Acoustical Society of America*, 395-411.

Flege, J. E. (1992). Speech learning in a second language. In Ferguson, L. M. & Stoel-Gammon, C. (eds.) *Phonological Development: models, research and implications.* Timonium, MD: York Press, 565-604.

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

Flege, J. E. (1995). Second language speech learning: theory, findings, and problems. In Strange, W. (ed.) *Speech perception and linguistic experience: issues in cross-language science* 171: 303-306.

Flege, J. E. (1999). Age of learning and second language speech. In Birdsong, D. (ed.) *Second language acquisition and the critical period hypothesis.* London: Lawrence Erlbaum Associates, 101-131

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., Munro, M. & Fox, R. (1992). The interlingual identification of Spanish and English vowels. *Journal of the Acoustical Society of America,* 128-143.

Fowler , C. A. (1981). Production and perception of coarticulation among stressed and unstressed vowels. *Journal of Speech and Hearing Research* 24, 127-139.

Fowler, C. A. & Best, C. T. & McRoberts, G. W. (1990). Young infants' perception of liquid coarticulatory influences on following stop consonants. *Perception and Psychophysics* 48, 559-570.

Fry, D. B. (1958). Experiments in the perception of stress. *Language and Speech* I, 126-152.

Gardner, R. C. & Lambert, W. E. (1972). *Attitudes and motivation in second-language learning.* Rowley, MA.: Newbury House.

Gardner, R. C., Clément, R., Smythe, P. C. & Smythe, C. L. (1979). Attitudes and motivation test battery – revised manual. *Research Bulletin* No. 15, Language Research Group, Department of Psychology, University of Western Ontario.

Gardner, R. C. (1980). On the validity of affective variables in second language acquisition: conceptual, contextual, and statistical considerations. *Language Learning* 30 (2), 255-270.

Gay, T. (1978). A cinefluographic study of vowel production. *Journal of Phonetics* 2, 255-266.

Gimson, A. C. (1980). *An Introduction to the Pronunciation of English.* London: Edward Arnold.

Gottfried, T. & Beddor, P. (1988). Perception of temporal and spectral information in French vowels. *Language and Speech* 31, 57-75.

Guiora, A. Z. (1972). Construct validity and transpositional research: toward an empirical study of psychoanalytic concepts. *Comprehensive Psychiatry* 13, 139-150.

Guiora, A. Z., Beit-Hallahmi, B., Brannon, R. C. L., Dull, C. Y. & Scoval, T. (1972). The effects of experimentally induced changes in ego states on pronunciation ability in a second language: an exploratory study. *Comprehensive Psychiatry* 13, 421-428.

Halle, M & Vergnaud, J. R (1987) *An essay on stress.* Cambridge, MA: MIT Press.

Halle, M. & Vergnaud, J. R. (1987). Stress and the cycle. Paper presented at the colloquium *Phonologie Pluri-linéaire*, Lyon.

Hamayan, E., Genesee, F. & Tucker, G. R. (1977). Affective factors and language experience in second language learning. *Language Learning: a Journal of Applied Linguistics* 27, 225-241.

Han, M. S. (1962). The feature of duration in Japanese. *Study Sounds* 10, 65-80.

Han, M. S. (1994). Acoustic manifestations of mora timing in Japanese. *Journal of the Acoustical Society of America* 96 (1), 73-82.

Hancin-Bhatt, B. (1994a). *Phonological transfer in second language perception and production.* Ph.D. thesis. University of Illinois at Urbana-Champaign.

Hancin-Bhatt, B. (1994b). Segment transfer: a consequence of a dynamic systems. *Second Language Research* 10 (3), 241-269.

Hanke, W. L. (1966). *Dynamic Articulatory Model of Speech Production Using Computer Simulation.* Ph.D. dissertation. MIT.

Haraguchi, S. (1994). *Oninron: gendai no eigogaku shiriizu* 3. Tokyo: Kaitakusha.

Haraguchi, S. (1977). *The tone pattern of Japanese: an autosegmental theory of tonology.* Tokyo: Kaitakusha.

Hatch, E. (1978). *Second language acquisition: a book of readings.* Rowley, Mass: Newbury House.

Hill, J. (1970). Foreign accents, language acquisition and cerebral dominance revisited. *Language Learning* 20, 237-248.

Hirayama, T. (1985). *Zen Nihon no hatsuon to akusento.* Tokyo: The Japan Broadcasting Corporation (NHK).

Hockett, C. (1955). A manual of phonology. *Memoir 11* (Indiana University Publications in Anthropology and Linguistics, Bloomington; Bloomington 1955; University of Chicago 1974).

Hockett, C. F. (1958). *A course in modern linguistics.* New York: Macmillan.

Hoequist, Jr. C. (1983). Syllable durations in stress-, syllable, and mora-timed languages. *Phonetica* 40, 203-237.

Homma, Y. (1981). Durational relationship between Japanese stops and vowels. *Journal of Phonetics* 9, 273-281.

Horwitz, E. K., Horwitz, M. B. & Cope, J. (1986). Foreign language classroom anxiety. *The Modern Language Journal* 70, 125-132.

Hulst, van dur H. & Weijer, van de, J. (1995). Vowel harmony. In Goldsmith, J. (ed.) *The handbook of phonological theory.* Oxford: Blackwell, 495-534.

Hutchinson, S. P. (1973). *An objective index of the English-Spanish pronunciation dimension.* Unpublished Masters thesis. University of Texas at Austin.

Hyltenstam, K. (1977). Implicational patterns in interlanguage syntax variation. *Language Learning* 27 (2), 383-411.

Imaishi, K. (1980). *Onseigakuteki hikaku.* Nichieigo hikaku kouza 1: onsei to keitai. Tokyo: Taishuukan, 18-67.

Ioup, G. (1987) Is there a structural foreign accent? A comparison of syntactic and phonological errors in second language acquisition. *Language Learning* 34 (2), 1-17.

Ioup, G., Boustagui, E., Tigi, M. E. & Moselle, M. (1994) Reexamining the critical period hypothesis. *Studies in Second Language Acquisition* 16, 99-112.

Ito, J. (1986). *Syllable theory in prosodic phonology.* Ph.D. thesis. University of Massachusetts at Amherst. Circulated by the Graduate Linguistic Student Association. Published by Garland, New York, 1988.

Johnson, J. S. & Newport, E. L. (1989). Critical period effects in second language learning: the influence of the maturation state on the acquisition of English as a second language. *Cognitive Psychology* 21, 60-99.

Jun, S-A. & Beckman, M. E. (1993). A gestural overlap analysis of vowel devoicing in Japanese and Korean. Paper presented at the 1993 Annual Meeting of The Linguistic Society of America, Los Angeles, California.

Jusczyk, P. W. (1985). On characterizing the development of speech perception. In Mehler, J. & Fox, R. (eds.) *Neonate cognition: beyond the blooming buzzing confusion.* Hillsdale, NJ: Erlbaum.

Jusczyk, P. W. (1992). Developing phonological categories from the speech signal. In Ferguson, L. M. & Stoel-Gammon, C. (eds.) *Phonological Development: models, research and implications.* Timonium, MD: York Press, 17-64.

Keating, P. A. & Huffman, M. K. (1984). Vowel variation in Japanese. *Phonetica* 59, 50-61.

Keating, P. A. (1988). Underspecification in phonetics. *Phonology*, 5 275-292.

Keating, P. A. (1990). The window model of coarticulation. In Kingston, J. and Beckman, M. E. (eds.) *Papers in Laboratory Phonology I: between the grammar and physics of speech.* Cambridge: Cambridge University Press, 451-470.

Kent, R. & Read, C. (1992). *The acoustic analysis of speech.* San Diego: Singular.

Kenworthy, J. (1987). *Teaching English pronunciation.* New York: Longman.

Kenyon, J. S. (1946). *American Pronunciation.* Michigan: Ann Arbor.

Kindaichi, H. (1954). *Haneru on, tsumaru on.* [References are to the version in Kindaichi 1967: 154-167].

Kindaichi, H. (1967). *Nihongo onin no kenkyuu.* Tokyo: Tokyodo.

Kingston, J. (1992) Comments on Chapter 2. In Cocherty, G. J. & Ladd, D. R. (eds.) *Papers in Laboratory Phonology II: Gesture, Segments, Prosody.*

Cambridge: Cambridge University Press, 60-65.

Kiparsky, P. (1985). Some consequences of lexical phonology. *Phonology* 2, 85-138.

Kondo, Y. (1995). *Production of schwa by Japanese speakers of English: a crosslinguistic study of coarticulatory strategies.* Ph.D. thesis. University of Edinburgh.

Koopmans-van Beinum, F. (1994). What's in the schwa? *Phonetica* 51, 68-79.

Koopmans-van Beinum, F. J. (1980). *Vowel contrast reduction: an acoustic and perceptual study of Dutch vowels in various speech conditions.* Ph.D. thesis. University of Amsterdam.

Krashen, S. D. (1973). Lateralization, language learning, and the critical period: some new evidence. *Language Learning* 23, 63-74.

Krull, D. (1989). Second formant locus patterns and consonant-vowel coarticulation in spontaneous speech. *Phonetic Experimental Research.* Institute of Linguistics, University of Stockholm (PERIUS) X, 87-108.

Kubota, R. (1987). *Analysis of the learning problems experienced by Japanese LEP students and suggestions for global bilingual assessment.* MAT Thesis, School for International Training.

Kubozono, H. (1989). The mora and syllable structure in Japanese: evidence from speech errors. *Language and Speech* 32(3), 249-278.

Kubozono, H. (1989). The mora and syllable structure in Japanese: evidence from speech errors. *Language and Speech* 32, 249-278.

Kuhl, P. K. (1979). Speech perception in early infancy: perceptual constancy for spectrally dissimilar vowel categories. *Journal of the Acoustical Society of America* 66, 1168-1179.

Kuhl, P. K. (1987). Perception of speech and sound in early infancy. In Salapatek, P. & Cohen, L (eds.) *Handbook of infant perception* 2. New York: Academic Press.

Kuhl, P. K. (1993). Early linguistic experience and phonetic perception: implications for theories of developmental speech perception. *Journal of Phonetics*, 125-139.

Kuhl, P. K., Williams, K. A. Lacerda, F., Stevens, K. N. & Lindblom, B. (1992). Linguistic experience alters phonetic perception in infants by 6 months of age. *Science* 255, 606-608.

Labov, W. (1972). *Sociolinguistic patterns.* Philadelphia: University of Pennsylvania Press.

Ladefoged, P. (1975). *Three areas of experimental phonetics.* London: Oxford University Press.

Ladefoged, P. (1982). *A course in Phonetics.* Orlando: Harcourt Brace Jovanovich.

Lado, R. (1957). *Linguistics across cultures.* Ann Arbor, Michigan.: University of Michigan Press.

Lamendella, J. T. (1977) General principles of neurofunctional organization and their manifestation in primary and nonprimary language acquisition. *Language Learning* 27, No. 1, 155-196.

Lea, W. A. (1977). Acoustic correlates of stress and juncture. In Hyman, L. M. (ed.) Studies in stress and accent. *Southern California Occasional Papers in Linguistics* 4, 83-119.

Leather, J. (1983). Second-language pronunciation learning and teaching. *Language Teaching Abstracts* 16, 198-219.

Leather, J. & James, A. (1991). The acquisition of second language speech.

Studies in Second Language Acquisition 13, 305-341.

Lehiste, I. (1970). *Suprasegmentals.* Cambridge, Mass.: MIT Press.

Lehiste, I. (1977). Isochrony reconsidered. *Journal of Phonetics* 5, 253-263.

Lehn, W. & Slager, W. (1959). A contrastive study of Egyptian Arabic and American English. The segmental phonemes. *Language Learning* 9, 25-33.

Lenneberg , E. (1967). *Biological foundations of language.* New York: Willey.

Levitt, A., Jusczyk, P. W., Murray, J. & Carden, G. (1988). Context effects in two-month-old infants' perception of labiodental/interdental fricative contrasts. *Journal of Experimental Psychology: Human Perception and Performance* 14, 361-368.

Lieberman, M. & Pierrehumbert, J. (1984). Intonational invariance under changes in pitch range and length. In Arnoff, M. & Oehrle, R. T. (eds.) *Language sound structure.* Cambridge, Mass.: MIT Press, 157-233.

Lieberman, O. (1960). Some acoustic correlates of word stress in American English. *JASA* 32, 451-454.

Lightbown, P. (1983). Exploring relationships between developmental and instructional sequences in L2 acquisition. In Seliger, H. & Long, M (eds.) *Classroom-oriented research in second language acquisition.* Rowley, Mass.: Newbury House, 217-243.

Lindblom, B. & Moon, S. J. (1988). Formant undershoot in clear and citation form speech. *Phonetic Experimental Research.* Institute of Linguistics, University of Stockholm (PERILUS) 8, 21-33.

Lindblom, B. (1963). Spectrographic study of vowel reduction. *Journal of the Acoustical Society of America* 35, 1773-1781.

Lindblom, B. (1988). Phonetic universals in the consonant system. In Hyman and

Li, C. (eds.) *Language, speech and mind.* Studies in honour of Victora A. Fromkin. London: Routledge and Kegan Paul.

Lindblom, B. (1990). Explaining phonetic variation: a sketch of the H and H theory. In Hardcastle, W. J. & Marchal, A. (eds.) *Speech Production and Speech Modelling* Netherlands: Kluwer Academic Publishers, 403-439.

Long, M. (1985). Input and second language acquisition theory. In Gass, S. & Madden, C. (eds.) *Input in second language acquisition.* Rowley, Mass.: Newbury House, 377-393.

Long, M. (1990). Maturational constraints on language development. *Studies in Second Language Acquisition* 12, 25-285.

Luanghongkum, T. (1977). *Rhythm in standard Thai.* Ph.D. thesis. University of Edinburgh.

Lukmani, Y. M (1972). Motivation to learn and language proficiency. *Language Learning* 22, 261-273.

Maassen, B. & Povel, D. J. (1985). The effect of segmental and suprasegmental corrections on the intelligibility of deaf speech. *Journal of the Acoustical Society of America* 78, 877-886.

MacIntyre, P. D. & Gardner, R. C. (1989). Anxiety and second-language learning. Toward a theoretical clarification. *Language Learning* 39, 251-275.

MacIntyre, P. D. & Gardner, R. C. (1994). The subtle effects of language anxety on cognitive processing in the second language. *Language Learning* 44 (2), 283-305.

Magan, H. (1984). Vowel-to-vowel coarticulation in English and Japanese. *Journal of the Acoustical Society of America* 75, Suppl.1:S11.

Magan, H. (1989). *An acoustic study of vowel-to-vowel coarticulation in English.*

Ph.D. thesis. Yale University.

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

Major, R. C. (1981). Stress-timing in Brazilian Portuguese. *Journal of Phonetics* 9, 343-351.

Major, R. C. (1987) Foreign accent: recent research and theory. *International Review of Applied Linguistics* XXV/3, 185-202.

Major, R. C. (1987). Phonological similarity, markedness, and rate of L1 acquisition. *Studies in Second Language Acquisition* 9, 63-82.

Manuel, S. & Krakow, R. A. (1984). Universal and language particular aspects of vowel-to-vowel coarticulation. *Haskins Laboratories Status Report on Speech Research* SR-77/78, 69-78.

McCawley (1987). What is a tone language? In Fromkin, V. A. (ed.) *Tone: a linguistic survey.* New York: Academic Press, 113-131.

McCawley, J. D. (1968). *The phonological component of a grammar of Japanese.* The Hague: Mouton.

Meador, D., Flege, J. E. & MacKay, I. R. A. (1997). Native Italian speakers' perception of English vowels. *Journal of the Acoustical Society of America* 101 (A), 2973-2987.

Mehler, J. (1985). Language related dispositions in early infancy. In Mehler, J. & Fox, R. (eds.) *Neonate cognition: beyond the blooming, buzzing confusion.* Hillside, NJ: Erlbaum.

Menzerath, P. & Lacerda, A. (1933). *Koartikulation, Steuerung und Lautabgrenzung: Eine experiemntelle Untersuchung.* Berlin and Bonn: Gerd. Dümmler.

Michaels, D. (1973). Sinhalese sound replacement and feature hierarchies. *Linguistics* 170, 12-22.

Miyawaki, K., Strange, W., Verbrugge, R., Liberman, 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.

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

Nakatani, H. H., O'Connor, K. D. & Aston, C. H. (1981). Prosodic aspects of American English speech rhythm. *Phonetica* 38, 84-106.

Nelson, C. (1982). Intelligibility and non-native varieties of English. *The other tongue: English across cultures* 15, 58-73.

Neufeld, G. (1978). On the acquisition of prosodic and articulatory features in adult language learning. *Canadian Modern Language Review* 34 (2), 163-174.

Newport, E. (1984). Constraints on learning: studies in the acquisition of American Sign Language. *Papers and Reports on Child Language Development* 23, 1-22.

Nosofsky, R. M. (1986). Attention, similarity, and the identification-categorization relationship. *Journal of Experimental Psychology: General* 115, 39-57.

Nosofsky, R. M. (1988). Exemplar-based accounts of relations between classification, recognition, and typicality. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 14, 700-708.

O'Connor, J. D. (1965). *The perception of time intervals. Progress Report: Phonetics Laboratory.* University College London 2: 11-15.

O'Connor, J. D. (1973). *Phonetics.* Harmondsworth, Middlesex: Penguin Books.

Öhman, S. E. G. (1966). Coarticulation in VCV utterances: spectrographic measurements. *Journal of the Acoustical Society of America* 39, 151-168.

Oller, J. W. & Perkins, K. (1978). Intelligence and language proficiency as sources of variance in self-reported affective variables. *Language Learning* 28, 85-97.

Oller, J., Hadson, A. J.& Liu, P. F. (1977). Attitudes and attained proficiency in ESL: a sociolinguistic study of native speakers of Chinese in the United States. *Language Learning* 27 (1), 1-27.

Osberger, M. J. & Levitt, H. (1979). The effect of timing errors on the intelligibility of deaf children's speech. *Journal of the Acoustical Society of America* 66, 1316-1324.

Oyama, S. (1973). *A sensitive period for the acquisition of a second language.* Ph.D. thesis. Harvard University.

Oyama, S. (1976). A sensitive period of the acquisition of a nonnative phonological system. *Journal of Psycholinguistic Research* 5 (3), 261-283.

Oyama, S. (1982a). A sensitive period for the acquisition of a nonnative phonological system. In Krashen, S., Scarcella, R. & Long, M. (eds.) *Child-adult differences in second language acquisition.* Rowley, Mass: Newbury House, 20-38.

Oyama, S. (1982b). The sensitive period for the comprehension of speech. In Krashen, S., Scarcella, R. & Long, M. (eds.) *Child-adult differences in second language acquisition.* Rowley, Mass: Newbury House, 39-51.

Parrondo-Rodriguez, A. E. (1999). *The L2 acquisition of syllable structure and stress in Spanish.* Ph.D. thesis. University of Durham.

Pater, J. (1993). Theory and methodology in the study of metrical parameter (re)setting. *McGill Working Papers in Linguistics* 9, 211-244.

Pater, J. (1997). Metrical parameter missetting in second language acquisition. In Hannahs, S. J. & Young-Scholten, M. (eds.). Focus on phonological Acquisition. Amsterdam: John Benjamins, 245-261.

Patkowski, M. S. (1980). The sensitive period for the acquisition of syntax in a second language. *Language Learning* 30 (2), 449-472.

Patkowski, M. S. (1990). Age and accent in a second language: a reply to James Emil Flege. *Applied Linguistics* 22, 73-89.

Paulston, C. B. (1975). Ethnic relations and bilingual education: accounting for contradictory data. *Working Papers in Bilingualism* 6, 1-44.

Pennington, M. (1989). Teaching pronunciation from the top down. *RELC Journal* 20, 207-225.

Pennington, M. C. (1998). The teachability of phonology in adulthood: a re-examination. *International Review of Applied Linguistics* 36, 323-341.

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

Pennington, M. C. & Ellis, N. C. (2000). Cantonese speakers' memory for English sentences with prosodic cues. *Modern Language Journal* 84 (3), 372-389.

Peterson, G. E. & Lehiste, I. (1960). Duration of syllable nuclei in English. *Journal of the Acoustical Society of America* 32, 693-703.

Pierrehumbert, J. & Beckman, M. E. (1988). *Japanese Tone Structure.* Cambridge, Mass.: The MIT Press.

Pike, K. L. (1946). *The intonation of American English.* Ann Arbor: University of Michigan Press.

Pimsleur, P. (1963). Discrimination training in the teaching of French pronunciation.

Modern Language Journal 47, 199-203.

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

Plann, R. (1977). Acquiring a second language in an immersion situation. In Brown, H., Yorio, C. & Crymes, R. (eds.) TESOL'77, TESOL, Washington. D. C., 213-123.

Port, R. F., Al-Ani, S. & Maeda, S. (1980). Temporal compensation and universal phonetics. *Phonetica* 37, 235-252.

Port, R. F., Dalby, J. & O'Dell, M. (1987). Evidence for mora timing in Japanese. *Journal of the Acoustical Society of America* 81, 1574-1585.

Port, R. F., Dalby, J. & O'Dell, M. (1987). Evidence for mora timing in Japanese. *Journal of the Acoustical Society of America* 81, 1574-1564.

Pulleyblank, D. (1988). Vocalic underspecification in Yoruba. *Linguistic Inquiry* 19, 233-270.

Purcell, E. T. & Suter, R. W. (1980). Predictors of pronunciation accuracy: a reexamination. *Language Learning* 30 (2), 271-287.

Rabiner, L. R., Rosenberg, A. E. & Levinson, S. E. (1978). Considerations in dynamic time warping algorithms for discrete word recognition. IEEE Transactions on Acoustics. *Speech and Signal Processing* 26, 575-582.

Recasens, D. (1984). V-to-V coarticulation in Catalan VCV sequences: an articulatory and acoustical study. *Journal of Phonetics* 12, 61-73.

Recasens, D. (1986). An acoustic analysis of V-to-C and V-to-V coarticulatory effects on Catalan and Spanish VCV sequences. *Status Report on Speech Research.* Sr-86/87. Haskins Laboratories.

Recasens, D. (1991). An electropalatographic and acoustic study of consonant-to-vowel coarticulation. *Journal of Phonetics* 19, 177-192.

Richie, W. (1968). On the explanation of phonic interference. *Language Learning* 18, 183-197.

Rochet, B. (1995). Perception and production of L2 speech sounds by adults. In Strange, W. (ed.) *Speech perception and linguistic experience: theoretical and methodological issues.* Timonium, MD: York Press, 379-410.

Rosansky, E. (1975). The critical period for the acquisition of language: some cognitive developmental considerations. *Working Papers on Bilingualism* 6, 92-102.

Sagey, E. (1986). *The representation of features and relations in non-linear phonology.* Ph.D. thesis. MIT.

Schein, B. & Steriade, D. (1986). On geminates. *Linguistic Inquiry* 17, 691-744.

Schermerhorn, R. A. (1970). *Comparative ethnic relations: a framework for theory and research.* New York: Random House.

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

Schmidt, A. M. & Flege, J. E. (1995). Effects of speaking rate changes on native and non-native production. *Phonetica* 52, 45-51.

Schumann, J. H. (1975). Affective factors and the problem of age in second language acquisition. *Language Learning* 25 (2), 209-235.

Schumann, J. H. (1976). Social distance as a factor in second language acquisition. *Language Learning* 26 (1), 135-143.

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

Scovel, T. (1981). The recognition of foreign accents in English and its implications for psycholinguistic theories of language acquisition. In Savard, J-G & Laforge, L. (eds.) *Proceedings of the 5th Congress of AILA.* Laval: University of Laval Press, 389-401.

Seliger, H. (1978). Implications of the multiple critical periods hypothesis for second language learning. In Ritchie, W. (ed.) *Second language acquisition research.* New York : Academic Press, 11-19.

Selkirk, E. (1980). The role of prosodic categories in English word stress. *Linguistic Inquiry* 11, 563-605.

Shearme, J. N. & Holmes, J. N. (1962). An experimental study of the classification of sounds in continuous speech according to their distribution in the formant-1, formant-2 plane. *Proceedings of the 4th International Congress of Phonetic Sciences.* The Hague: Mouton, 233-240.

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, 243-61.

Shen, Y. & Peterson, G. G. (1962). Isochronism in English. *Occasional Papers* 9. University of Buffalo, Studies in Linguistics, 1-36.

Shibatani, M.(1990). *The Languages of Japan.* Cambridge: Cambridge University Press.

Shuy, R., Wolfram, W. & Riley, W. (1967). *Linguistic correlates of social stratification in Detroit speech.* East Lansing: Michigan State University.

Steriade, D. (1987). Redundant values. In Bosch, A., Need, B. & Schiller, E.

(eds.) *Papers from the parasession on autosegmental and metrical phonology.* Chicago: Chicago Linguistic Society 23, 339-362.

Stockwell, R. & Bowen, J. (1965). *The sounds of English Spanish.* Chicago: University of Chicago Press.

Stockwell, R., Bowen, J. & Martin, J. (1965). *The grammatical structures of English and Spanish.* Chicago: University of Chicago Press.

Suenobu, M., Kanzaki, K. & Yamane, S. (1992). An experimental study of the intelligibility of Japanese English. *International Review of Applied Linguistics in Language Teaching* 30, 146-156.

Sugito, M. (1969). Doutaisokutei ni yoru nihongo no akusento no kenkyuu. *Gengo kenkyuu* 55.

Sugito, M. (1980). *Akusento, intoneeshon no hikaku II, Nichiei hikaku kouza 1: onsei to keitai.* Tokyo: Taishuukan. Reprinted in Sugito, M. (1996).

Sugito, M. (1996). *Nihonjin no eigo. Nihongo onsei no kenkyuu* 2. Tokyo: Izumi Shoin, 111-138.

Suter, R. W. (1976). Predictors of pronunciation accuracy in second language learning. *Language Learning* 26 (2), 233-253.

Swain, M. (1985). Communicative competence: some roles of comprehensible input and comprehensible in its development. In Gass, S. M. & Madden, C. G. (eds.) *Input in second language acquisition.* Cambridge: Newbury House Publishers.

Tahta, S. (1980). An investigation into second language learning. M.Phil thesis. City University, London.

Tahta, S., Wood, M. & Lowenthal, K. (1981). Foreign accents: factors relating to transfer of accent from the first language to a second language. *Language and*

Speech 24 (3), 265-272.

Tajima, K. Port, R. & Dalby, J. (1997). Effects of temporal correction on the intelligibility of foreign-accented English. *Journal of Phonetics*, 25 1-24.

Takefuta, Y & Black, J. W. (1977). Some acoustic correlates of foreign accent in the word accentuation of Japanese English and American Japanese. *Language Laboratory* 14. The Language Laboratory Association of Japan.

Taylor, B. (1974) Toward a theory of language acquisition. *Language Learning* 24, 23-45.

Terbeek, D. A. (1977). A cross-language multi-dimensional scaling study of vowel perception. *Working Papers in Phonetics* 37. University of California at Los Angeles.

Tiffany, W. R. (1959). Nonrandom sources of variation in vowel quality. *Journal of Speech and Hearing Research* 2, 305-317.

Tiffen, B. (1992). A study of intelligibility of Nigerian English. In van Essen, A. & Burkart, E. I. (eds.) *Homage to W. R. Lee: essays in English as a foreign or second language.* Berlin: Fortis, 255-259.

Tuller, B., Harris, K. S. & Scott-Kelso, J. A. (1982). Stress and rate: differential transformation of articulation. *Journal of the Acoustical Society of America* 71 (6), 1534-1543.

Vance, J. V. (1987). *An introduction to Japanese phonology.* Albany: State University of New York Press.

Verbrugge, R. R. & Shankweiler, D. P. (1977). Prosodic information for vowel identity. *Journal of the Acoustical Society of America* 61, Suppl. No.1 S 39.

Vihman, M. (1992). Early syllables and the construction of phonology. In Ferguson, L. M. & Stoel-Gammon, C. (eds.) *Phonological Development:*

models, research and implications. Timonium, MD: York Press, 393-422.

Wardhaugh, R. (1970). The contrastive analysis hypothesis. *TESOL Quarterly* 4, 123-130.

Wechsler, D. (1958). *The measurement and appraisal of adult intelligence.* Baltimore: Williams & Wilkins.

Weinreich, U. (1953). *Language in contact.* Linguistic Circle of New York.

Werker, J. F. & Lalonde, C. E. (1988). Cross-language speech perception: initial capabilities and developmental change. *Developmental Psychology* 24, 672-683.

Werker, J. F. & Pegg, J. E. (1992). Infant speech perception and phonological acquisition. In Ferguson, L. M. & Stoel-Gammon, C. (eds.) *Phonological Development: models, research and implications.* Timonium, MD: York Press, 393-422.

Werker, J. F. & Polka, L. (1993). Developmental changes in speech perception: new challenges and new directions. *Journal of Phonetics* 21, 83-101.

Werker, J. F. & Tees, R. C. (1984). Cross-language speech perception: evidence for perceptual reorganization during the first year of life. *Infant Behaviour and Development* 7, 49-63.

Werker, J. F. (1989). Becoming a native listener. *American Scientist* 77, 54-59.

Werker, J. F. (1993). The contribution of the relation between vocal production and perception to a developing phonological system. *Journal of Phonetics* 21, 177-180.

Whittaker, H., Bub, D. & Laventer, S. (1981). Eurolinguistic aspects of language acquisition and bilingualism. In Winitz, H. (ed.) Native language and foreign language acquisition. *Annals of the New York Academic of Sciences* 379, 59-

74.

Wieden, W. & Nemser, W. (1991) The pronunciation of English in Austria. Tübingen: Gunter Narr.

Young-Scholten, M. (1995). The negative effects of 'positive' evidence on L2 phonology. In Eubank, L., Selinker, L. & Smith, M. S. (eds.) *The current state of interlanguage.* Amsterdam: Benjamins, 107-121.

Appendix A

Table 1. Segments: test scores of the three subjects

	M0	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12
NI	70.00%	83.30%	83.30%	90.00%	86.70%	90.00%	80.00%	80.00%	83.30%	93.30%	83.30%	90%	86.60%
MO	73.30%	86.70%	90.00%	86.70%	93.30%	90.00%	96.70%	96.70%	90.00%	90.00%	86.60%	86.60%	90%
SK	70.00%	63.30%	63.30%	60.00%	66.70%	66.70%	66.70%	66.70%	60.00%	70.00%	70%	63.30%	70%

Table 2. Syllable structure: test scores of the three subjects

	M0	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12
NI	66.30%	66.30%	73.80%	81.30%	86.30%	83.80%	83.80%	88.80%	90.00%	88.80%	86.30%	87.50%	81.30%
MO	85.00%	88.80%	86.30%	88.80%	86.30%	91.30%	88.80%	90.00%	86.30%	90%	90%	92.50%	88.80%
SK	76.30%	78.80%	83.80%	86.30%	85.00%	86.30%	87.50%	88.80%	87.50%	88.80%	82.50%	85%	85%

Table 3. Prosody (ungrammatical items): test scores of the three subjects

	M0	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12
NI	61.90%	59.00%	52.40%	65.70%	70.50%	71.40%	62.90%	76.20%	70.50%	80%	82.90%	81%	81%
MO	62.90%	48.60%	54.30%	41.00%	40.00%	36.20%	26.70%	21.90%	20.00%	21%	20%	21%	21%
SK	52.40%	54.30%	55.20%	58.10%	62.90%	61.90%	62.90%	64.80%	49.50%	61.00%	63.80%	63.80%	63.80%

Appendix B

NI data		F1	F2
Month 0	parapet 1	740	1260
	parapet 2	750	1300
	calibre 1	730	1280
	calibre 2	710	1280
	cantaloup 1	700	1300
	cantaloup 2	700	1250
	matador 1	730	1300
	matador 2	720	1280
Month 1	parapet 1	700	1300
	parapet 2	730	1280
	calibre 1	710	1290
	calibre 2	730	1300
	cantaloup 1	750	1330
	cantaloup 2	750	1300
	matador 1	760	1270
	matador 2	720	1300
Month 2	parapet 1	700	1280
	parapet 2	730	1310
	calibre 1	730	1270
	calibre 2	740	1300
	cantaloup 1	730	1280
	cantaloup 2	730	1300
	matador 1	720	1280
	matador 2	700	1300
Month 3	parapet 1	710	1290
	parapet 2	700	1300
	calibre 1	740	1260
	calibre 2	710	1250
	cantaloup 1	730	1310
	cantaloup 2	700	1300
	matador 1	710	1300
	matador 2	750	1320
Month 4	parapet 1	750	1270
	parapet 2	710	1290
	calibre 1	700	1290
	calibre 2	750	1270
	cantaloup 1	730	1320
	cantaloup 2	730	1330
	matador 1	710	1300
	matador 2	700	1330
Month 5	parapet 1	710	1280
	parapet 2	730	1290
	calibre 1	760	1270
	calibre 2	710	1300
	cantaloup 1	740	1390

	cantaloup 2	720	1400
	matador 1	720	1410
	matador 2	740	1460
Month 6	parapet 1	630	1350
	parapet 2	650	1380
	calibre 1	650	1390
	calibre 2	700	1400
	cantaloup 1	680	1490
	cantaloup 2	660	1480
	matador 1	700	1430
	matador 2	670	1500
Month 7	parapet 1	660	1510
	parapet 2	660	1590
	calibre 1	680	1580
	calibre 2	690	1480
	cantaloup 1	700	1530
	cantaloup 2	690	1600
	matador 1	700	1610
	matador 2	650	1620
Month 8	parapet 1	660	1560
	parapet 2	660	1600
	calibre 1	700	1630
	calibre 2	680	1620
	cantaloup 1	630	1520
	cantaloup 2	610	1530
	matador 1	630	1390
	matador 2	650	1500
Month 9	parapet 1	590	1550
	parapet 2	630	1540
	calibre 1	650	1590
	calibre 2	670	1390
	cantaloup 1	630	1450
	cantaloup 2	650	1500
	matador 1	590	1590
	matador 2	610	1650
Month 10	parapet 1	640	1430
	parapet 2	630	1390
	calibre 1	640	1440
	calibre 2	630	1570
	cantaloup 1	590	1640
	cantaloup 2	610	1700
	matador 1	550	1490
	matador 2	600	1550
Month 11	parapet 1	590	1710
	parapet 2	600	1400
	calibre 1	610	1550

	calibre 2	650	1610
	cantaloup 1	630	1670
	cantaloup 2	610	1490
	matador 1	630	1440
	matador 2	580	1590
Month 12	parapet 1	640	1540
	parapet 2	660	1730
	calibre 1	590	1550
	calibre 2	590	1450
	cantaloup 1	600	1730
	cantaloup 2	610	1680
	matador 1	590	1650
	matador 2	570	1740

MO data		F1	F2
Month 0	parapet 1	790	1370
	parapet 2	750	1400
	calibre 1	810	1330
	calibre 2	790	1280
	cantaloup 1	830	1300
	cantaloup 2	780	1340
	matador 1	800	1400
	matador 2	840	1430
Month 1	parapet 1	800	1400
	parapet 2	760	1370
	calibre 1	760	1320
	calibre 2	840	1300
	cantaloup 1	800	1410
	cantaloup 2	750	1440
	matador 1	860	1390
	matador 2	810	1350
Month 2	parapet 1	800	1350
	parapet 2	730	1400
	calibre 1	800	1370
	calibre 2	840	1390
	cantaloup 1	840	1450
	cantaloup 2	780	1420
	matador 1	880	1460
	matador 2	800	1430
Month 3	parapet 1	780	1400
	parapet 2	800	1430
	calibre 1	780	1370
	calibre 2	810	1370
	cantaloup 1	830	1340
	cantaloup 2	800	1400
	matador 1	810	1450
	matador 2	790	1400
Month 4	parapet 1	750	1370
	parapet 2	810	1390
	calibre 1	790	1400
	calibre 2	770	1390
	cantaloup 1	750	1460
	cantaloup 2	830	1440
	matador 1	810	1390
	matador 2	800	1400
Month 5	parapet 1	770	1420
	parapet 2	750	1390
	calibre 1	730	1370
	calibre 2	800	1400
	cantaloup 1	740	1490

	cantaloup 2	720	1520
	matador 1	740	1450
	matador 2	790	1490
Month 6	parapet 1	770	1420
	parapet 2	800	1390
	calibre 1	770	1370
	calibre 2	750	1400
	cantaloup 1	740	1490
	cantaloup 2	730	1480
	matador 1	810	1570
	matador 2	720	1520
Month 7	parapet 1	730	1510
	parapet 2	760	1450
	calibre 1	710	1420
	calibre 2	690	1410
	cantaloup 1	750	1560
	cantaloup 2	710	1600
	matador 1	800	1590
	matador 2	750	1620
Month 8	parapet 1	710	1520
	parapet 2	660	1440
	calibre 1	730	1500
	calibre 2	610	1470
	cantaloup 1	630	1580
	cantaloup 2	610	1630
	matador 1	590	1600
	matador 2	630	1650
Month 9	parapet 1	650	1560
	parapet 2	710	1610
	calibre 1	630	1590
	calibre 2	670	1590
	cantaloup 1	710	1680
	cantaloup 2	650	1630
	matador 1	590	1630
	matador 2	620	1620
Month 10	parapet 1	660	1490
	parapet 2	630	1610
	calibre 1	620	1660
	calibre 2	640	1580
	cantaloup 1	590	1690
	cantaloup 2	610	1640
	matador 1	550	1730
	matador 2	590	1610
Month 11	parapet 1	590	1510
	parapet 2	690	1430
	calibre 1	510	1620

	calibre 2	530	1530
	cantaloup 1	560	1670
	cantaloup 2	610	1710
	matador 1	630	1700
	matador 2	580	1610
Month 12	parapet 1	650	1510
	parapet 2	700	1430
	calibre 1	550	1690
	calibre 2	530	1620
	cantaloup 1	600	1510
	cantaloup 2	610	1680
	matador 1	590	1670
	matador 2	550	1710
Month 13	parapet 1	530	1470
	parapet 2	600	1480
	calibre 1	610	1500
	calibre 2	580	1480
	cantaloup 1	570	1680
	cantaloup 2	550	1640
	matador 1	590	1690
	matador 2	550	1650

SK data		F1	F2
Month 0	parapet 1	830	1360
	parapet 2	800	1370
	calibre 1	850	1350
	calibre 2	810	1330
	cantaloup 1	890	1300
	cantaloup 2	820	1330
	matador 1	800	1370
	matador 2	840	1340
Month 1	parapet 1	830	1340
	parapet 2	850	1330
	calibre 1	880	1310
	calibre 2	790	1300
	cantaloup 1	840	1390
	cantaloup 2	800	1380
	matador 1	850	1330
	matador 2	810	1320
Month 2	parapet 1	870	1350
	parapet 2	840	1330
	calibre 1	800	1310
	calibre 2	810	1340
	cantaloup 1	890	1390
	cantaloup 2	860	1370
	matador 1	840	1400
	matador 2	810	1410
Month 3	parapet 1	850	1400
	parapet 2	880	1340
	calibre 1	820	1350
	calibre 2	830	1350
	cantaloup 1	800	1340
	cantaloup 2	880	1400
	matador 1	810	1370
	matador 2	830	1360
Month 4	parapet 1	830	1330
	parapet 2	810	1370
	calibre 1	790	1320
	calibre 2	830	1340
	cantaloup 1	850	1380
	cantaloup 2	830	1400
	matador 1	820	1390
	matador 2	800	1400
Month 5	parapet 1	870	1330
	parapet 2	850	1390
	calibre 1	820	1370
	calibre 2	800	1350
	cantaloup 1	830	1330

	cantaloup 2	880	1360
	matador 1	870	1360
	matador 2	820	1360
Month 6	parapet 1	870	1340
	parapet 2	810	1360
	calibre 1	810	1350
	calibre 2	830	1330
	cantaloup 1	840	1390
	cantaloup 2	890	1400
	matador 1	810	1380
	matador 2	820	1380
Month 7	parapet 1	870	1410
	parapet 2	870	1350
	calibre 1	810	1360
	calibre 2	790	1340
	cantaloup 1	860	1390
	cantaloup 2	890	1410
	matador 1	800	1390
	matador 2	850	1400
Month 8	parapet 1	810	1380
	parapet 2	860	1350
	calibre 1	790	1400
	calibre 2	810	1420
	cantaloup 1	890	1380
	cantaloup 2	890	1370
	matador 1	850	1400
	matador 2	830	1380
Month 9	parapet 1	890	1360
	parapet 2	810	1410
	calibre 1	790	1380
	calibre 2	780	1380
	cantaloup 1	810	1380
	cantaloup 2	860	1400
	matador 1	770	1390
	matador 2	810	1380
Month 10	parapet 1	790	1390
	parapet 2	830	1400
	calibre 1	820	1360
	calibre 2	840	1370
	cantaloup 1	880	1400
	cantaloup 2	850	1420
	matador 1	800	1330
	matador 2	820	1370
Month 11	parapet 1	840	1340
	parapet 2	850	1390
	calibre 1	790	1400

	calibre 2	830	1380
	cantaloup 1	860	1370
	cantaloup 2	890	1410
	matador 1	830	1390
	matador 2	780	1400
Month 12	parapet 1	790	1410
	parapet 2	830	1430
	calibre 1	850	1390
	calibre 2	830	1400
	cantaloup 1	790	1380
	cantaloup 2	840	1400
	matador 1	790	1380
	matador 2	800	1410
Month 13	parapet 1	790	1370
	parapet 2	800	1380
	calibre 1	810	1360
	calibre 2	780	1410
	cantaloup 1	830	1430
	cantaloup 2	850	1400
	matador 1	820	1390
	matador 2	790	1380