Title	Perceptual distance between vowels: the effects of prototypicality and extremity
Author	Nakai, Satsuki
Qualification	PhD
Year	1999

The perceptual distance between vowels:
The effects of prototypicality and extremity

Satsuki Nakai

A dissertation presented for a degree of PhD
The University of Edinburgh
March, 1999

Declaration

I declare that this thesis is my own work.

Date: March 24, 1999

Name: Satsuki Nakai

Acknowledgement

A number of people helped me in various ways while I was producing this thesis. Some people gave me academic advice, while others gave me morale support. Some provided me with a place to stay at difficult times. Others, unknowingly, made me see something important during a casual conversation. Because it seems an endless task to describe how all my friends, the staff at the University of Edinburgh, my colleagues at the University of Wales, Bangor, and other researchers in the field helped me, and because all their help was equally valuable, I would like to simply list the names of all the people whose help I would like to acknowledge. I would like to thank:

Dora Alexopoulou, Kazuaki Ando, Georgios Arsenos, Alexander Astaras, Ellen Bard, Ocke-Schwen Bohn, Mathew Bull, Emilios Cambouropoulos, Rob Clark, Kostas Dabos, Norman Dryden, Eddie Dubourg, Nick Ellis, Demetrios Fragoyiannis, Elaina Frieda, Masato Fukao, Frank Guenter, Masahito Hasegawa, Yasushi Hayashi, Olaf Husby, Kazunori Imai, Mikiko Inoue, Steve Isard, Tomonori Ishikawa, Paul Iverson, Kino Iwazumi, Keith Johnson, Shuji Joko, Anastassios Karpodinis, Heather King, Yoshiki Kinoshita, Ariyuki Kondo, Yuko Kondo, Stasinos Konstantopoulos, Nicos Labaras, Bob Ladd, Dina Lew, Susana Lorenzo, Max Louwerse, Irene MacLeod, Cedric MacMartin, Tatsuyuki Marumo, Toshibumi Matsubara, Annyce Meiners, Glen Meiners, Emmanuel Melissaris, Ineke Mennen, Valsamis Mitsilegos, Chikayoshi Mochizuki, Ruta Noreika, Takashi Onozuka, Takeshi Osawa, Kristos Panayiotis, Kostas Papathanasiou, Deborah Pegg, David Pisoni, John Power, Bruno Repp, Martha Robinson, Nicholas Rossis, Narutoshi Sakano, Campbell Sandilands, Nagayasu Sato, Yoshiro Shibasaki, Richard Shillcock, Mike Smith, Stewart Smith, Antonella Sorace, Spyridakis Stefanos, Konstantinos Tourlas, Alice Turk, Marilyn Vihman, Rob van der Wijngaart, Keiji Yamane, Ikko Yoshida, and Jun Yoshida.

Finally, I am most grateful to my parents for their support and to my best friend Saleh Al Riyami for his unlimited patience and encouragement.

Bangor, July 30, 1999

Abstract

The present dissertation is concerned with a question of how the perceptual distance between vowels in foreign languages (L2) and those in one's first language (L1) is determined. According to current models of L2 acquisition, the degree of difficulty that the learner faces in perceptually learning L2 sounds is determined by their perceived similarity to the L1 sounds. For instance, Flege's (1995) Speech Learning Model predicts that the learner would fail to establish a separate category for those L2 sounds that are similar to L1 sounds.

Although empirical studies support a correlation between perceptual similarity between L1 and L2 sounds and the difficulty of perceptually acquiring L2 sounds, a measure for such similarity is yet to be established.

In the first part of the dissertation I investigate possible locations of vowel prototypes, which current models of L2 phonetics see as reference points in determining the perceptual distance between L1 and L2 vowels. As far as vowels are concerned, evidence suggests that unlike visual prototypes vowel prototypes may be at a more extreme location than the centre of the geometrically represented category distribution (Bradlow, 1993; Johnson et al., 1993; Iverson & Kuhl, 1995; Lotto et al., 1996; Frieda, 1997; Lively & Pisoni, 1997). If vowel prototypes are more extreme than the typical production value for the category, how the language specificity observed in their phonetic realisations is reflected in the locations of prototypes becomes a crucial issue for the current models of L2 phonetics, which assume phonetic prototypes to be language specific.

In the second part of the dissertation I re-examine Kuhl's Native Language Magnet Theory which holds that phonetic prototypes established in the course of L1 acquisition attract nearby members of the category, causing the perceptual space to shrink towards these prototypes (Kuhl 1991, 1992, 1993). If L1 prototypes have such an assimilation effect, this should be taken into account when measuring the perceptual distance between L1 and L2 sounds. However, the validity of Native Language Magnet Theory has been questioned by a number of researchers who replicated Kuhl's (1991) study, on which her Native Language Magnet Theory is based. An inspection of these replications reveals that the assimilation effect was always found around the most extreme stimuli used in the experiment, suggesting that it may be vowel extremity and not prototypicality that may be the cause of the observed assimilation effect.

In order to answer the above questions, the locations of native Japanese and Greek speakers' vowel prototypes and their discrimination sensitivity for the category /u/ are studied. Japanese and Modern Greek have phonologically comparable vowel systems, but Japanese /u/ is produced without lip rounding and different from cardinal, Greek /u/. Thus, language specificity is expected to be observed in the two language groups' prototypes for the category /u/. Furthermore, the effect of vowel extremity and prototypicality is expected to be distinguished in the Japanese listeners' abilities to distinguish between members of the category /u/, since Japanese /u/ is not cardinal and therefore not extreme.

If extreme vowels have an assimilation effect, the perceptual distance between L1 and L2 vowels may be inversely correlated to the extremity of L1 vowels, and as a consequence, the ease of perceptual learning of L2 vowels may be negatively correlated to the extremity of L1 vowels.

Table of Contents

Chapter 1 Introduction

This dissertation is concerned with how the perceptual distance between vowels in foreign languages (L2s)[1] and vowels in one's native language (L1) is determined on the assumption that the learner relates L2 sounds to those in their L1s. According to the two most influential current models of L2 phonetics, i.e., Best's Perceptual Assimilation Model and Flege's Speech Learning Model, the degree of difficulty that the learner faces in perceptually learning L2 sounds is determined by the degree of perceived similarity, or the perceptual distance, between L1 and L2 sounds. Best's Perceptual Assimilation Model assumes that the adult language learner perceives non-native phones in terms of their articulatory-phonetic (gestural) similarities to the L1 phonetic categories, and predicts that those L2 contrasts which are identified with a single L1 category would constitute difficulty for the learner, whereas those contrasts which are identified with separate L1 categories would not. Flege's Speech Learning Model assumes that the adult language learner relates L2 sounds to their L1 sounds based on their acoustic-phonetic (auditory) similarities, and maintains that the learner fails to establish a separate category for those L2 sounds which are similar to those in L1.

Although both of the above models have been generally supported by empirical studies (Bohn, 1995), a gauge for measuring the perceptual distance between L1 and L2 sounds, the key concept of the two models, is yet to be established (Strange, 1995). Although both Best and Flege seemingly assume that L2 phones are compared with some kind of mental representations, or prototypes, of L1 phones, what constitutes such prototypes is not defined by either model. In the prototype theory of speech perception, it has been assumed that phonetic prototypes are those category members that are perceived as 'best' and at the same time situated at the centre of the category in the geometrically represented stimulus space. However, as far as vowels are concerned, evidence suggests that those vowels that are more extreme than the production average are perceived as 'best' (cf. Johnson et al., 1993; Iverson & Kuhl, 1995; Lotto, et al., 1996; Frieda, 1997; Lively & Pisoni, 1997), even when they are outside of the language group's

---

[1] In this dissertation I use the term 'second language' (abbreviated as L2) in the sense used by Eckman (1996: 195), i.e., 'a language acquired either by an adult, or by a child after one language is already resident.' Although some researchers argue for the need of separate terms for language that is picked up naturally within the language community (second language) and language learned formally outside the language community (foreign language) (e.g., Richards, 1978: 4-7; Stern, 1983: 9-18), in reality most foreign language learners who have been subject to studies had had experience of both kinds of learning to different degrees, and neither of the above terms serves well in describing the languages they possess in addition to their L1s. For the same reason, 'learning' and 'acquisition' are interchangeably used to refer to both 'learning' (conscious language development in school-like settings) and 'acquisition' (natural development of language that is analogous to L1 acquisition) in Krashen's (1978, 1981) terminology.

production range (Bradlow, 1993). The question of possible locations of vowel prototypes is addressed in the first part of the present dissertation.

In the second part of the dissertation, I look at Kuhl's (1991, 1992, 1993) Native Language Magnet Theory. The Native Language Magnet Theory holds that phonetic prototypes established in the course of L1 acquisition attract nearby members of the category, causing the perceptual space to shrink towards these prototypes. According to Kuhl (1993), the assimilation by L1 prototypes interferes with adults' ability to perceive certain L2 sounds. If L1 prototypes have such an assimilation effect, it is an important factor that should be taken into account when measuring the perceptual distance between L1 and L2 sounds. However, the validity of Kuhl's Native Language Magnet Theory has been questioned by a number of researchers who replicated her 1991 study, on which the Native Language Magnet Theory is empirically based. In Kuhl's 1991 study, the production average of /i/ obtained from speakers of American English reported in Peterson & Barney (1952) was given the best rating in a goodness-rating task and was shown to assimilate other category members. In Iverson & Kuhl (1995), however, American listeners rated more extreme /i/ stimuli as best, and these extreme stimuli exhibited the assimilation effect. On the other hand, in Sussman & Lauckner-Morano (1995), Peterson & Barney's (1952) production average was found to have the assimilation effect despite the fact that it was not given the best rating by the listeners. An inspection of these studies reveals that the assimilation effect was always found around the most extreme stimuli used in the experiment, suggesting that it may be vowel extremity and not prototypicality[2] that is the cause of the observed assimilation effect. If vowel extremity is correlated with the assimilation effect, there may be directional asymmetry in the ease of perceptual learning of L2 vowels depending on their extremity.

Thus, in the present dissertation I investigate (i) possible locations of vowel prototypes, and (ii) whether the perceptual distance in the vowel space shrinks around language-specific prototypes or towards its periphery (extreme vowels). Experiments 1a-b (production and perception tasks) were conducted to re-examine the relationship between perceptually preferred vowels and the production values, and Experiments 2a-b (identification and discrimination tasks) were conducted in order to find out whether it is vowel prototypicality or extremity that is correlated with poor discrimination sensitivity, which has been taken as evidence of assimilation. Two languages that have not been used to test the Native Language Magnet Theory, Japanese and Modern Greek, were studied. These languages have vowel systems that are

---

[2] In this dissertation, I use 'prototypicality' to refer to the quality that is specific to a hypothetical phonetic prototype that plays a special role in speech perception, which may or may not match the production average, and distinguish it from 'typicality', which is associated with category members that are observed most frequently.

phonologically comparable in terms of vowel quality but differ in the phonetic realisations of the /u/ category. Specifically, Japanese /u/ is produced without lip rounding and not as extreme as Greek /u/. Thus, language specificity is expected to be observed in the locations of the two language groups' prototypes for the /u/ category. Furthermore, the effect of vowel extremity and prototypicality is expected to be distinguished in the Japanese listeners' discrimination sensitivity in the high back area of the vowel space, as their prototype for the /u/ category is expected not to be extreme.

The present dissertation is organised as follows:

In Chapter 2, I review the studies in the field of L2 phonetics during the past three decades, and describe the process of the development of Best's Perceptual Assimilation Model and Flege's Speech Learning Model. I then introduce the prototype theory of speech perception and Kuhl's Native Language Magnet Theory, and discuss the results of recent studies that are problematic to these theories.

In Chapter 3, I address specific research questions that the present dissertation is concerned with, and briefly explain the four experiments (Experiments 1a-b, 2a-b) designed to answer these questions. The vowel systems of Japanese and Modern Greek are also described here.

In Chapters 4 and 5, Experiments 1a-b (production and perception tasks) and Experiments 2a-2b (identification and discrimination tasks) are described in more detail and the results are discussed. As mentioned earlier, Experiments 1a-b were designed to re-examine the relationships between vowel prototypes and the production ranges. Experiments 2a-b were designed to test whether it is vowel prototypicality or extremity that correlates with poor discrimination sensitivity.

Finally, in Chapter 6, the results from the four experiments are summarised and their implications for the acquisition of L2 phonetics are discussed.

Chapter 2 Theoretical Background

2.1 An Overview of Second language (L2) Phonology Research

The emergence of accounts for L2 learners' sound systems that are independent from theories in other subfields of Second Language Acquisition (SLA)[3] is relatively recent. Until the mid-80's the field of L2 phonology/phonetics shared theories with other subfields in SLA (e.g., the Contrastive Analysis Hypothesis, the Interlanguage Theory; see Stern, 1983; Strange, 1995) and there was no model that specifically addressed the question of how L2 learners learn the new sound system.

In this chapter I first describe the findings in the field of L2 phonology/phonetics during the past three decades, and the process of the development of the most influential of current models that address such questions, Best's Perceptual Assimilation Model and Flege's Speech Learning Model in 2.1.1. I then describe the above two models in 2.1.2, and discuss the significance of these models from a historical perspective in 2.1.3.

In parts of this overview I do not make overt distinctions between phonetic and phonological levels of analysis, for the above models, which are often referred to as L2 phonetics models, stem from what was commonly referred to as L2- or interlanguage- phonology, and it is not easy to distinguish between the two. The L2 learners who served as subjects in various studies described in this chapter are adults who grew up in a monolingual environment, whose age ranges from 18 to 32, unless otherwise stated.

2.1.1 The Development of L2 Phonetics and Findings to date (from the late 50's to the 80's)

The rudiments of L2 phonetics research can be found in the application of the Contrastive Analysis Hypothesis, a theory that was influential in the US from the late 50's to the early 70's in the field of SLA, to the comparative phonological description of various languages and English. The Contrastive Analysis Hypothesis holds that L2 learners tend to transfer their L1 systems to L2, which leads to the use of L1 features in substitution for those of L2. Therefore, features common in L1 and L2 do not constitute problems, but L2 features not found in L1 are problematic to the learner. In other words, the key to ease or difficulty in L2 learning was regarded to lie in a comparison between L1 and L2. Thus, comparisons were made in order to predict language-specific difficulties in the learning of L2 and, in turn, to determine the areas of focus in various L2 classrooms. The Contrastive Analysis Hypothesis did not originate to deal exclusively with

---

[3] The term 'the field of SLA' refers to all disciplines dealing with different aspects of L2 acquisition, including phonetics, syntax, semantics, pragmatics, etc.

phonological aspects of L2 learning and was thought to be applicable to all aspects of L2 learning, from phonology to syntax (see Lado, 1957; Stern, 1983).

As for the phonological aspect of L2 learning, the Contrastive Analysis Hypothesis predicts that L2 phonemes that do not have similar counterparts in the learner's L1 phonology constitute problems for the learner, whereas L2 phonemes that have similar L1 counterparts do not. For example, in the first systematic statement of contrastive linguistics, *Linguistics Across Cultures,* Lado (1957) writes that the English phonemes commonly mispronounced by the Spanish learners, i.e., /v/, /ð/, /z/, /ʒ/, /ʃ/ and /dʒ/, along with some vowels, are those phonemes that are absent from the learners' L1 (Spanish) phonology. Thus, it was typical to assume that substitution of L1 features took place at an abstract, phoneme level in the application of the Contrastive Analysis Hypothesis to L2 learners' sound system (see Moulton, 1962; Stockwell & Bowen, 1965; Tarone, 1978:17).[4]

Furthermore, Lado thought that the learner's interpretation of L2 phonemes in terms of their L1 systems in perception led to incorrect L2 production. In other words, there was a straightforward causal relationship between the learner's perception and production of L2 phones (Lado 1957: 11, 21). Thus, Lado's observations are basically based on the learner's production, which was apparently regarded to be sufficient in discussing L2 learners' phonology. For example, he writes:

> We tend to transfer to that [foreign] language our phonemes and their variants, our stress and rhythm patterns, our transitions, our intonation patterns and their interaction with other phonemes... Much less known, and often not even suspected, may be the fact mentioned above that the speaker of one language listening to another does not actually hear the foreign language sounds units - phonemes. He hears his own. (Lado, 1957: 11).

The idea that L2 phones are perceived according to the learner's L1 phonology and that the imposition of L1 phonology on L2 sounds in perception leads to difficulties in L2 production can be found as early as in the 30's (Polivanov, 1931; Trubetzkoy, 1939) and was not new, but it was not until the late 50's that the principle was systematically applied in a dual description of phonology of L1 and L2 (cf. Stern 1983: 59-60).

Although Lado was aware that the learner's substitution of similar L1 sounds for L2 sounds led to production of target phones that are different from the target-language norms in phonetic detail, he regarded this sort of difference unproblematic, on the grounds that the substituted sound would not be misidentified by the target language group. For example, on the substitution of Spanish

---

[4] Analyses at an abstract phoneme level were a common practice, although Lado (1937: 16-19) was aware that the syllabic positions of phonemes and the phonetic contexts in which they are placed influence the production and perception of the phonemes, as is described shortly. Furthermore, Lado's analyses at a phoneme level received more attention, although his analysis extended to a suprasegmental level (Lado, 1957: 27-50). The issue of the importance of the syllabic positions and phonetic contexts of phonemes in L2 acquisition is discussed in more detail later in this section.

dental /d/ for English alveolar /d/, he comments, ' ... this difference in point of articulation is heard as a matter of "accent," Spanish accent, that does not change any word in the language' (Lado, 1957: 16). Lado's above view is in contrast with later researchers such as Flege, who claims that similar L2 sounds are more problematic for the learner, because foreign accents resulting from such substitution persist even in advanced learners (see 2.1.2).

Lado was also aware that the language-specific knowledge of allophonic variations depending on the position in syllable and phonetic context affects perception and production of phonemes, which recent instrumental work in L2 phonetics attests 30 years later. He writes:

> ... even when the native language has a similar phoneme and the variants are similar, if it does not occur in the same position as in the native language, the student will have trouble producing and hearing it in the position in which it does occur in the foreign language (Lado, 1957: 17).

Apparently, however, Lado's above insight was not extensively adopted in the application of the contrastive analysis and remained obscured.

By the early 80's, support for the Contrastive Analysis Hypothesis had started to erode, both on theoretical and empirical grounds. Theoretically, the behaviourist view of learning that viewed language learning as habit formation, on which the Contrastive Analysis Hypothesis was implicitly based, was heavily criticised by cognitivists who regarded language learning as understanding of a system of rules and its creative use (see, e.g., Sciarone, 1970). Empirically, several studies reported counterevidence against the prediction of the Contrastive Analysis Hypothesis. For instance, Dušková (1969) analysed syntactic and lexical errors made by Czech learners of English and showed that there are various error types, e.g., confusing similar forms in L2, and not all the errors could be explained as an imposition of the L1 system on L2.

Instead, the Interlanguage Theory proposed by Corder (1967, 1978, 1981) and Selinker (1972) became influential in the field of SLA. Interlanguage[5] is a term referring to the L2 learner's language from a viewpoint that it is a recognisable language variety, or a language system with its own characteristics and rules, just as child language is seen as a recognisable variety of a certain language, a variety that is more dynamic, or variable, than conventional language varieties, such as regional dialects. Within this framework, L2 learners' grammar was viewed as an autonomous system that was independent from L1, and emphasis was placed on similarities between L1 and L2 acquisition and universal sequence of development that all L2 learners follow irrespective of their

---

[5] The term interlanguage was first used in Reinecke's 1935 study of Hawaiian pidgin, but the first use of the term in the present meaning is found in Selinker (1972). At the beginning of the 70's, L2 learners' grammar was referred to using various terms, such as 'transitional dialect' (Corder, 1971) and 'approximative system' (Nemser, 1971). [In Richards, 1978: 2; Beebe, 1984].

L1.[6] (See Richards, 1978: 1; Stern, 1983: 469-471; Eckman, 1996: 196-197; Leather & James, 1996: 290).

The concept of interlanguage was adopted in the study of L2 phonology as well as other subfields in SLA, although not as much research was conducted in phonology as in syntax (cf. Richards, 1978: 2; Tarone, 1978: 15). As in other subfields of SLA, attempts were made to demonstrate that the learner's deviant production of L2 was not a result of the imposition of their L1 systems on L2, as the Contrastive Analysis Theory maintains. For instance, in her study of Japanese learners' production of English /z/, Dickerson (1975) argues that L2 learners' phonology is independent of their L1 phonology by showing that some of these phonemes that the Japanese learners produced for English /z/ were not in their L1 (Japanese) phoneme inventory (e.g., /dz/); in other words, they were not substitutions of L1 phonemes for L2 phonemes. In a similar vein, in her study of the pronunciation of English /r/, /l/, /i/, /ɪ/, /s/ and /θ/ by 5 groups of Asian learners of English (Japanese, Mandarin, Korean, Thai and Indonesian), Beebe (1984) reports that the learners produced some of the above English phonemes inaccurately as the Contrastive Analysis Hypothesis would predict, but they seldom substituted L1 phonemes for these English sounds as the Contrastive Analysis Hypothesis holds.

On the other hand, Johansson (1973) carried out a segmental-level analysis of repetitions of tape-recorded Swedish sentences collected from 180 speakers of 9 different L1 backgrounds (American English, Czech, Danish, Finnish, Greek, Hungarian, Polish, Portuguese and Serbo-Croatian), and concludes that a large number of substitutions made by the speakers could be predicted by the Contrastive Analysis Hypothesis, although there were some common tendencies for substitution followed by all the language groups. For example, she observed a general tendency for the subjects, irrespective of their L1 backgrounds, to substitute vowels that are less extreme in height for Swedish high vowels. Similarly, Wode (1976) longitudinally studied German children learning English as L2 and English children learning German as L2, and concludes that most phonological elements were strongly characterised by L1 transfer, although some were learned in a similar way to L1 acquisition. Specifically, German children learning English learned /r/ and /w/ in a similar way to English children acquiring L1, but vowels and other consonants were subject to L1 transfer.

As mentioned earlier, interlanguage theorists' focus was on finding in the learner's L2 production evidence for a system that was independent of the learner's L1 grammar. As for L2 phonology, their attempt was partially successful in that they did find some aspects that did not appear to result from L1 transfer. However, when all the results are put in perspective, L1 transfer seems more robust than a language-independent, universal sequence of development, especially at the beginning stage (Leather & James, 1996: 291). The failure to find evidence for predominance of

---

[6] Descriptively, however, interlanguage was always compared with the learner's L1 and the target language (L2) (Leather & James, 1996: 288).

a universal developmental sequence over L1 interference in the acquisition of L2 phonology has led some proponents of the Interlanguage Theory to claim that L1 transfer has its strongest effect in phonology (cf. Corder, 1978: 77; Tarone, 1978: 27; Hatch, 1978; Ritchie & Bhatia, 1996; Leather & James, 1996: 285). Their claim appears compatible with a multiple critical periods hypothesis (Seliger, 1978; Long, 1990) which holds that there is more than one maturationally constrained, sensitive period for different linguistic abilities, with the closure of that for phonology being the earliest (age 6) and syntax the latest (age 15). To summarise, although the change in focus of the field facilitated our understanding of L2 learners' phonology, the contention of interlanguage theorists that the L2 learner's language development is independent of the learner's L1 background seems to have proven only marginally true in phonology.[7]

In an attempt to reconcile the findings in support of the Contrastive Analysis Theory and the Interlanguage Theory, Eckman (1977) proposes a Markedness Differential Hypothesis, which holds that the areas of difficulty for the learner can be predicted on the basis of both a systematic comparison between L1 and L2, and the markedness relations stated in universal grammar.[8] Specifically, the Markedness Differential Hypothesis states that the learner has difficulty where there are differences between L1 and L2, as the Contrastive Analysis Theory holds, but the degrees of difficulty the learner faces in acquiring these L2 features depend on their markedness determined by universal grammar. That is, those L2 features that are more marked than the L1 features are difficult for the learner. For instance, if the learner's L1 only has a single obstruent in the word-final position (less marked), learning word-final consonant clusters (more marked) constitutes a problem for the learner. Although appealing, the Markedness Differential Hypothesis has been subject to a number of critical evaluations (Leather & James, 1996: 290). For instance, Hyltenstam (1984) found that Spanish-speaking subjects made the same errors as Finnish-speaking subjects with Swedish pronominal reflexes in relative clauses, despite the fact that Spanish has pronominal reflexes in relative clauses and Finnish does not, which was interpreted as counterevidence against the Markedness Differential Hypothesis. As a response to the counterevidence, Eckman (1984, 1996) proposes a Structural Conformity Hypothesis, which basically eliminates the element of contrastive analyses from the original Markedness Differential Hypothesis and simply states that the degrees of difficulty in acquiring L2 phones depend on their markedness. Eckman's Structural Conformity Hypothesis is, in a sense, a regression, since it ignores the fact that contrastive analyses

---

[7] Although most studies on L2 phonology/phonetics to date have been at the segmental level, some aspects of L2 phonology at the suprasegmental level have been also shown to be predicted systematically from the learner's L1 system and to be present even at an advanced stage of learning (cf. Leather & James, 1996: 274-275, 278, 291).

[8] Markedness is a dimension along which one can compare structures sharing features with respect to universal grammar. The criterion used to determine the degree of markedness differs depending on the definition, but the most frequently employed criterion is referred to as typological markedness, where A is more marked than B if the presence of A in a language implies the presence of B, but not vice versa (see Ioup & Weinberger, 1987: 420).

explain the L2 learner's difficulty to a great extent, at least in the realm of phonology. One of the possible problems with the Markedness Differential Hypothesis (and the Structural Conformity Hypothesis) is that it attempts to explain all the aspects of L2 acquisition without taking account of the evidence that the relationship between L1 transfer and the L1-independent developmental sequence may be different for different aspects of L2 acquisition.

In parallel with the Interlanguage Theory, Abramson & Lisker's (1967, 1973) studies on cross-linguistic differences in the perception of voice-onset time (VOT)[9] also influenced the field of L2 phonology and directed many researchers' interest, which had been mainly on speech production, to the perception of L2 contrasts, and to the L1 influence in the perception of L2 contrasts. Abramson & Lisker found that English, Spanish and Thai speakers' locations of category boundaries for voiced and voiceless distinction for initial stop consonants along a VOT continuum were different, reflecting cross-linguistic differences in their production of voiced-voiceless distinctions. Specifically, the English perceptual crossovers had higher VOT values than the Spanish ones for all the places of articulation, reflecting the difference between English aspirated voiceless consonants and Spanish unaspirated voiceless consonants. Thai speakers, on the other hand, produced two category boundaries along the VOT continuum for bilabial stops, reflecting the three categories of bilabial stops along the VOT dimension. Following Abramson & Lisker's study, difficulties in categorically perceiving /ra/-/la/ continua by Japanese and Korean listeners, whose L1s do not have these contrasts, were demonstrated (Goto, 1971; Miyawaki et al., 1975; Gillette, 1980; Sheldon & Strange, 1982). It was this group of researchers who first successfully demonstrated that the learner's correct categorisation of L2 contrasts in perception does not necessarily precede accurate production of the contrasts,[10] contrary to the common sense belief that difficulties in the production of L2 sounds arise from the influence of L1 phonology on the perception of L2 sounds (Polivanov, 1931; Trubetzkoy, 1939; Lado, 1957; For more details on the issue of L2 perception and production, see p 13).

Subsequently, through the 80's to date a substantial number of studies on the perception of L2 contrasts were carried out on a variety of language groups and contrasts, the results of which increased our understanding of L2 phonology, as shown below. Research conducted during this period is mostly on the learner's perception of L2, apart from studies on the relationships between L2 perception and production (cf. Strange, 1995). Furthermore, the research was increasingly concerned with the phonetics of L2 rather than phonology, reflecting the shift from impressionistic observation

---

[9] Voice-onset time is a term referring to the point in time at which vocal-cord vibration starts, after the release of a closure. In a fully voiced plosive (e.g., some instances of [b]), the vocal cords vibrate throughout. In a voiceless unaspirated plosive (e.g., [p]), there is a delay (or lag) before the voicing starts. In a voiceless aspirated plosive (e.g., [pʰ] in English *pin*), the delay is much longer, depending on the amount of aspiration (cf. Crystal, 1997: 24, 329).

[10] Although not as widely known, prior to this, Brèire (1968) reports cases where some subjects were able to produce L2 sounds before becoming able to discriminate them perceptually.

in the 60's and phonetic transcription in the 70's to more instrumental research in the laboratory that enabled researchers to manipulate and measure subtle differences in acoustic information. It must be also noted that a substantial number of studies during this period looked at advanced learners as well as learners at other levels, which was made more meaningful by the introduction of the instrumental approach. Summarised below are some of the major findings from the 80's to date:

(1) **Adults are capable of learning to distinguish L2 contrasts perceptually.**

Support for adult learners' abilities to learn to distinguish L2 contrasts can be found in studies comparing performance of L2 learners at different levels. Generally, advanced learners have been found to be more 'native-like' in their performance than beginners. For instance, MacKain et al. (1981) report that Japanese learners of English with extensive English experience discriminated and identified the English /r/ -/l/ contrast much better than learners with limited experience. Other support can be found in studies on the effects of some training techniques in adult learners' perception of L2 contrasts. For instance, Jamieson (1995) reports that short-term intensive training employing appropriate stimuli and tasks can improve adult learners' perception of L2 consonant contrasts: Jamieson and Morosan (1986) successfully trained native speakers of Canadian French to differentiate the English /θ/ -/ð/ contrast perceptually, using the fading technique where listeners are first trained with stimuli containing an exaggerated amount of the voiced/voiceless target frication and later with stimuli with a natural amount of frication. Jamieson (1995) also reports other successful techniques (multiple natural tokens, categorical discrimination tasks, etc.) with different contrasts and language groups, which include Korean and Japanese speakers trained on the English /r/ - /l/ contrast, English speakers trained on Hindi dental vs. retroflex consonant contrasts, and Mandarin speakers trained on the English /t/ - /d/ contrast.

(2) **Adults' inferior performance to children in perceptually discriminating L2 contrasts is a reflection of selective attention to acoustic cues that is presumably learned in the course of L1 acquisition, rather than an atrophy of sensory abilities.**

In the same study showing adult Japanese speakers' limited ability to differentiate stimuli along a synthesised /ra/-/la/ continuum, Miyawaki et al. (1975) also demonstrated that the Japanese subjects could detect the differences in isolated F3 components taken from the /ra/ -/la/ stimuli (perceived as nonspeech 'chirps'), which was the only acoustic cue that was available to American subjects who categorically perceived the /ra/ -/la/ continuum. Werker & Logan (1985) also report that adults could detect the differences in acoustic cues defining non-native contrasts when they attended to the stimuli as non-speech stimuli. The results have been interpreted to indicate that adults' inability to discriminate between non-native contrasts is not due to a loss of general auditory ability but their selective attention to cues that are phonologically contrasted in their L1 (cf. Pisoni & Tash, 1974; Jusczyk, 1993: 9; Bohn, 1995b: 87; Flege, 1995: 266; Strange, 1995: 79). Similarly, in reporting the success of short-term training for the perception of L2 contrasts, Jamieson (1995: 105) suggests that the adult learners' learning speed implies that training induces redirection of attention rather than fundamental auditory system changes. (For training that successfully redirected listeners' attention, also see Samuel, 1977; Carney et al., 1977; Pisoni et al., 1982).

(3) **Although adult L2 learners' perception of some L2 contrasts may improve, their perception will not match that of native speakers of the target language even after several years of L2 experience.**

As illustrated above, adult L2 learners are reported to be capable of learning to redirect attention to relevant cues in L2 input. However, this does not mean that after several years of L2 experience their perception becomes completely native-like. For example, experienced German learners of English (more than 5 years of residence in the US) in Bohn & Flege (1990b) did not differ from inexperienced German learners (less than one year of residence in the US) in the way they differentiated the English /i/ - /ɪ/ contrast. Both groups relied on both spectral and temporal cues, which is how German L1 speakers discriminate between the German /i/ - /ɪ/ contrast, while L1 English speakers predominantly used spectral cues in discriminating between the contrast. In the same study by Bohn & Flege, however, the experienced German learners distinguished the English /ɛ/-/æ/ contrast in a more native-like manner than the inexperienced German learners, i.e., the experienced learners used more spectral cues than the inexperienced ones in discriminating the contrast. Flege and his associates explain that the English /ɛ/-/æ/ contrast might have been learned better than the English /i/ - /ɪ/ contrast because German does not have /æ/ in its phonemic inventory, and hypothesise that L2 phones that are similar to L1 phones persist in perceptual difficulties. (See 2.1.2 for more about Flege's view).

(4) **Not all the dissimilar L2 contrasts are equally difficult to differentiate perceptually.**

Contradicting the prediction of the Contrastive Analysis Hypothesis that L2 sounds that do not have similar L1 counterparts constitute problems for the learner,[11] Best et al. (1988) report that both voicing and place contrasts among Zulu clicks, which are unlike any English phonemes, were well

---

[11] As mentioned earlier, Lado thought there was a straightforward causal relationship between incorrect perception and production of L2 phones and did not distinguish the two processes.

discriminated perceptually by native English speakers. (For cases where non-native contrasts are well discriminated, also see Brèire, 1966; Werker & Tees, 1984a; Polka, 1987; Best et al., 1982). Gottfried (1984) also reports a contradictory finding to the prediction of the Contrastive Analysis Hypothesis that experienced English learners of French had difficulty distinguishing the French /e/-/ɛ/ contrast perceptually, which constitutes a phonemic contrast also in the learners' L1 (English) but differs in phonetic detail. On the other hand, Werker & Tees (1983, 1984b) report that English listeners had persistent difficulty differentiating Hindi place contrasts perceptually, which do not have counterparts in English phonology, as the Contrastive Analysis Theory would predict. These results are interpreted by Best and others to indicate that the degrees of difficulty in perceptually discriminating L2 contrasts depend not simply on whether or not the L1 system has counterparts of individual L2 sounds, but on how they are mapped onto the L1 system (e.g., whether the contrasts are assimilated by a single L1 category or two; for more details on Best's view, see 2.1.2).

(5) Language-specific knowledge of allophonic variations as a function of position in syllable and phonetic context influences the learnability of L2 contrasts.

The influence of language-specific knowledge of allophonic variations on L2 perception was already noted by Lado as early as in the 50's, but the empirical attestation to the importance of allophonic variations in the learning of L2 phonetics took place much later. For example, Mochizuki (1981) and Pisoni & Lively (1995) report that the position of /r/ and /l/ in the syllable influenced the Japanese listeners' ability to discriminate between English /r/ and /l/. Specifically, their perceptual differentiation of the contrast was much more accurate in the postvocalic position than in the prevocalic position. Strange et al. (1993) also report that the categorisation of German /Y/ into English vowel categories and their goodness of fit (i.e., the extent to which German /Y/ sounded like the English vowels it was identified with) depended on the consonantal context. Furthermore, according to Morosan & Jamieson (1989) and Rochet (1995: 395-406), the knowledge required to discriminate a contrast in a certain phonetic environment is not applicable to the discrimination of the contrast in other phonetic environments. They report that learning to perceptually differentiate a non-native contrast in a certain phonetic environment does not transfer to the ability to differentiate the contrast in other environments. Thus, it appears that the identification of L2 phonemes is affected by the learner's knowledge of the allophonic variations of the phoneme as a function of syllabic position and phonetic context. This implies that the degrees of difficulty of perceptually discriminating L2 contrasts cannot be predicted simply on the basis of a comparison of L1 and L2 phoneme inventories, viz., a comparison of phoneme at an abstract level, as was a common practice of supporters of the Contrastive Analysis Hypothesis.

(6) **Perception does not always precede production in L2 learning.**

Contradicting the common belief that inaccurate production of L2 sounds arise from the imposition of L1 phonology on the perception of L2 sounds (Polivanov, 1931; Trubetzkoy, 1939; Lado, 1957) and that there is a straightforward causal relationship between perception and production, Goto (1971) and Sheldon & Strange (1982) showed that Japanese learners of English living in the US distinguished the English /r/ - /l/ contrast more accurately in production than perception. (Also see Gass, 1984; Flege & Eefting, 1987; Bohn & Flege, 1990a). However, it is not always the case that the learner distinguishes L2 contrasts better in production than perception. For instance, Flege (1993) examined the duration of vowels preceding /t/ and /d/ produced by advanced Chinese learners of English and found that the learners performed better in perception than production.

It was even shown that in some cases there may be no direct link between L2 learners' perceptual and productive patterns of L2 sounds. For example, Nemser (1971) reports cases where Hungarian learners of English perceived English interdentals (/θ, ð/) as labial fricatives (/f, v/) and produced them as stops (/t, d/). On the other hand, Flege & Schmidt's (1995) study on speaking-rate effects on the perception and production of English long-lag /p/ by Spanish learners suggests a close link between perception and production in L2. Rochet's study (1995) on the relationship between perceptual assimilation patterns of L2 vowel (French /y/) to L1 vowels (/i/ or /u/) by English and Portuguese speakers and their production of the target vowel also suggests that the perception of L2 may have direct influence on the production. Evidence for a link between perception and production in L2 can be also found in studies reporting that improvement in perception transfers to production and vice versa (Pimsleur, 1963; Mueller & Niedzielski, 1968; Weiss, 1992; Yamada et al., 1995; however, see Lane & Schneider, 1963; Jamieson, 1995).

To sum up, the once-prevalent belief that inaccurate production of L2 sounds derives from incorrect perception of these sounds was shown not to be true in all cases. Some attribute the greater success in production found in some studies to explicit articulatory training and social pressure to improve production (e.g., Sheldon & Strange, 1982; Sheldon, 1985; Flege, 1991). Others suggest that production/perception relationships may differ among different classes/aspects of sounds (e.g., Gordon & Myer, 1984; Bohn & Flege, 1990a; Strange, 1995). Yet others point out the methodological questions of comparing results from different kinds of tests, i.e., perception and production tests, may be responsible for the diverse results (e.g., Leather, 1988; Mack, 1989; Bohn & Flege, 1990a). In any case, it seems that there is a general consensus that the relationships between production and perception of L2 sounds are not as straightforward as it was once thought.

2.1.2 Current L2 Phonetics Models

Based on the empirical findings described above, L2 phonetics models that predict language-specific difficulties faced by language learners of various L1 backgrounds started to emerge in their own right in the mid-80's, after two decades of sharing theories with other SLA subdisciplines. Among the most influential of such models are Best's Perceptual Assimilation Model and Flege's Speech Learning Model (cf. Strange, 1995; Bohn, 1995b).

Best's Perceptual Assimilation Model is based on the assumption that the adult language learner perceives non-native phones in terms of their articulatory-phonetic (gestural) similarities to the L1 phonetic categories. Best's notion of articulatory-phonetic similarity is based on Browman and Goldstein's articulatory phonology, in which articulations refer to 'temporal and spatial properties (i.e., degree and location of constrictions) of the dynamic movements of vocal tract articulators such as lips, jaw, tongue body, glottis, etc.' (Best & Strange 1992: 306).[12] The Perceptual Assimilation Model holds that the degree of perceptual difficulty in differentiating L2 contrasts is predictable from whether and how they are assimilated to L1 phoneme categories.

According to the Perceptual Assimilation Model, four assimilation patterns are possible: (1) If the contrasting phones are both assimilated to a single L1 category equally well (or poorly), perceptual differentiation is difficult. However, (2) perceptual differentiation will be easier, if the contrasting phones assimilated to a single L1 category differ in the category goodness in their fit to the L1 category. (3) If the two phones are assimilated to two different L1 categories, they will be differentiated easily. (4) If the non-native phones are very discrepant from any L1 phonetic gestures, they will be perceptually differentiated on the basis of their psychoacoustic distinctiveness. For example, for native speakers of Japanese the ease of perceptually discriminating between American English approximant contrasts /w-j/, /w-r/ and /r-l/ decreases in this order, and this can be accounted for by the ways the above English approximants are assimilated to the Japanese phoneme categories. Specifically, discrimination of English /w/ from /j/ is the easiest, for they are assimilated to separate Japanese phonemes, i.e., /w/ and /j/, respectively. Discrimination of English /w/ from /r/ is more difficult, for both /w/ and /r/ are perceived as exemplars of a single Japanese category /w/. However, their goodness of fit to the L1 category differs, i.e., English /w/ is perceived as a better exemplar of Japanese /w/ than English /r/ is, and therefore discrimination is not as difficult as in the case of /r/ and /l/, which are both perceived as poor exemplars of Japanese /w/ (Best & Strange, 1992). On the other hand, discrimination of Zulu click contrasts should be easy for Japanese listeners, as they are not likely to be assimilated to any of the L1 phoneme categories, as was shown for English-speaking adults in Best et al. (1988).

---

[12] Best and her colleagues acknowledge, however, that the distinction between articulatory- and acoustic-phonetic similarities is difficult to make (Best et al., 1992: 306).

The Perceptual Assimilation Model is based on Werker et al.'s allophonic experience account and Burnham et al.'s psychoacoustic account for the perception of L2 contrasts. Werker and her colleagues (Werker et al., 1981; Tees & Werker, 1984) hold that allophonic variants of L1 contrasts provide the listener with experience that maintains some discrimination of phonetically similar non-native contrasts. For instance, Best et al. (1988: 346) state that the Hindi [dʰ] - [tʰ] occur as allophonic variants of the English [d]-[t] contrasts and therefore are discriminable for English-speaking adults, as observed in Werker & Tees (1984b).[13] Burnham (1986: 209-212), on the other hand, distinguishes between psychoacoustically 'robust' contrasts, which are psychoacoustically salient and typologically common, and 'fragile' contrasts, which are not psychoacoustically salient and typologically rare. For instance, the prevoiced/voiced contrast is less salient psychoacoustically than the voiced/voiceless contrast, since the latter has at least one extra perceptual cue in addition to voicing, i.e., the degree of transition in the first formant after the onset of voicing. Burnham holds that the robust contrasts remain distinguishable until relatively later stages of life (4-8 years) without exposure to those contrasts, and the ability to discriminate the contrasts is relatively easy to recover even if the L1 lacks those contrasts. On the other hand, fragile contrasts are lost in infancy (during the second half year of the infant's postnatal life).

Flege's Speech Learning Model also assumes that adult learners tend to interpret L2 sounds in terms of L1 phonetic categories, although his Speech Learning Model is different from Best's Perceptual Assimilation Model in that it is concerned with the learner's ability to distinguish perceptually between individual L1 and L2 sounds, while the Perceptual Assimilation Model is concerned with how L2 contrasts are differentiated perceptually. Thus, for example, the Speech Learning Model is concerned with the question of whether the French learner of English can discern the difference between French /u/ (L1 sound) and English /u/ (L2 sound) (see Flege, 1986), while the Perceptual Assimilation Model is concerned with the question of whether the learner can tell English /u/ and /ʊ/ (L2 sounds) apart but *not* with a question of whether the learner can distinguish French /u/ (L1 sound) from English /u/ (L2 sound). Furthermore, unlike Best, Flege apparently focuses more on acoustic-phonetic (auditory) than articulatory-phonetic similarity, although he does not commit himself on the issue of which similarity learners use as the basis for relating L2 sounds to L1 sounds. He notes, 'Such interlingual identification [i.e., equating sounds in L2 to those in L1; in Flege's later term 'equivalence classification'] appears to depend on the auditory, and perhaps articulatory, similarity of L1 and L2 phones' (Flege, 1984: 708, my emphasis; also see Flege, 1986: 35; Flege, 1995: 264).

Flege's Speech Learning Model differs from Best's Perceptual Assimilation Model also in that it takes account of longitudinal effects of L2 exposure in the learner's phonetics. It holds that

---

[13] Although [dʰ] is not commonly regarded to occur in English, Werker et al. (1981: 354) argue that some linguists (e.g., Ladefoged, 1975) describe the consonant as more like a voiced unaspirated stop, which does occur in English, rather than a breathy voiced stop.

L2 learners initially identify L2 sounds in terms of similar L1 sounds (equivalence classification) at a position-sensitive allophonic level, even if there are differences between them in phonetic detail. As learners receive more L2 input, however, they may gradually become able to discern phonetic differences between certain L2 sounds and their L1 counterparts. According to the Speech Learning Model, the persistence of equivalence classification is, on the one hand, correlated with the age when L2 learning starts, and inversely correlated with the amount of L2 experience, on the other. That is, the younger age at which L2 learning starts and the longer the L2 experience, the more likely it is that the learner will learn to discern the differences between similar L1 and L2 sounds.

More importantly, the Speech Learning Model holds that the likelihood of learning to discern an L2 sound from a similar L1 sound inversely correlates with the degree of perceived similarity between the two sounds (for a similar view, see Valdman, 1976). For each successfully discerned L2 sound, a new L2 perceptual category will be established over time. On the other hand, for those L2 sounds which learners fail to discern from their L1 counterparts, separate L2 category formation may continue to be blocked, not only leading to the production of the L2 sounds with a persistent foreign accent but also to shifts in production of the L1 counterparts away from the monolingual norm. The above hypotheses are supported by Flege's (1986) study in which advanced English learners produced French /y/ (a dissimilar L2 sound) more authentically than French /u/ (a similar L2 sound). Flege postulates that, since French /y/ does not correspond to any English vowel categories (hence it is dissimilar from any English sound), the English learners of French had established a new perceptual category for this sound, which led to the authentic production of the sound. French /u/, on the other hand, was identified with English /u/ and because the degree of similarity between the two is great, new category formation for French /u/ was blocked, which resulted in the inauthentic production of the sound. Furthermore, the experienced female French speakers of English produced French /u/ (L1 sound) with higher F2 values (more English-like value) than what is reported for monolingual French female speakers in Debrock and Forrez (1976), suggesting a merger of French and English /u/ categories in those advanced learners.[14]

Flege's idea that a single phonetic category subsumes similar L1 and L2 sounds in the L2 learner, which results in the formation of a category that differs from that of monolingual speakers of both languages derives from studies on bilingual adults' phoneme boundaries for VOT that are a compromise between the norms of the two languages they speak. Specifically, Flege (1987b) found that highly experienced French learners of English produced English /p, t, k/ with shorter (more French-like) VOT values than the English norm and French /p, t, k/ with longer (more English-like) VOT values than the French norm (Flege, 1995: 258-259, 264) (also see Caramazza et al., 1973; Williams, 1977; Flege, 1987a; for contradictory evidence, see Elman et al., 1977). The effects of L2 acquisition on L1 categories are also reported in Holmes (1995), where the German /e/-/a/

---

[14] However, L2 influence on production of L1 was not observed in English /u/ produced by experienced English learners of French.

category boundary of English/German bilinguals was found to be different from that of German monolinguals presumably because of their linguistic experience with the English /æ/ category (also see Nathan, 1987, 1990; Major, 1990).

2.1.3 Evaluation of Current L2 Phonetics Models from a Historical Perspective

As described above, both Best's Perceptual Assimilation Model and Flege's Speech Learning Model share with the Contrastive Analysis Hypothesis their major premise that adult learners tend to interpret L2 sounds in terms of their L1 systems. However, the current two models are different from the Contrastive Analysis Hypothesis in that the theoretical focus of the former is on the learner's perceptual difficulty[15] while that of the latter is on production.

Furthermore, the Contrastive Analysis Hypothesis derives from impressionistic observation in the language classroom, whereas the current models derive from empirical studies conducted in the laboratory, where manipulation and measurements of subtle differences in phonetic detail were made possible. Consequently, the former predicts L2 learners' difficulty in all-or-non fashion (Lado, 1957:1), whereas the latter acknowledges different degrees of difficulty arising from different degrees of perceived similarity between L1 and L2 phones resulting from various degrees of differences in phonetic detail (Flege, 1995: 239; Best, 1995: 195). In addition, Lado was not only a researcher but also an EFL teacher whose interest lied in helping the learner to produce 'comprehensible' speech and understand L2. This is in contrast with Best and Flege, who are interested in underlying phonological system L2 learners possess. The difference between Lado and Flege's interests is evident in the ways they discuss L2 learners' use of duration in discriminating English vowels:

> ... German speakers will identify the two sounds [English /i/ and /ɪ/] readily ... In most situations English /i/ is longer than /ɪ/, although this difference in length can be proved not to be phonemic feature ... The German speaker cannot hear clearly this difference in quality, but he hears a difference in length, because vowel length is phonemic in German ... The contrast between /i/ and /ɪ/ need not be considered a perception problem (Lado, 1957:22).
>
> The NS [native Spanish] subjects managed to partition both continua [/i - ɪ/ and /ɛ - æ/] into two response categories, but in neither instance did the data provide compelling evidence for the establishment of categories for vowels not found in Spanish [i.e., /æ/ and /ɪ/] ... many NS subjects identified members of the *beat-bit* continuum based primarily on vowel duration rather than on spectral quality, as was the case for NE [native English] subjects (Flege, 1995: 244).

Because of their different stances, Lado's Contrastive Analysis Hypothesis and Flege's Speech Learning Model make very different predictions in terms of which sounds would constitute

---

[15] Many of earlier Flege's studies look at production, but apparently in an attempt to find evidence for his hypotheses regarding L2 learners' perception.

problems for the learner, although they both compare individual L1 and L2 sounds on the basis of the same premise that learners interpret L2 sounds in terms of their L1 system. For example, the Contrastive Analysis Hypothesis would predict that English /d/ would not constitute a problem for the Spanish-speaking learner because English alveolar /d/ and Spanish dental /d/ are similar enough[16] and therefore English /d/ does not constitute a problem for Spanish learners of English, as substituting Spanish /d/ for English /d/ would not result in misunderstanding. On the other hand, the Speech Learning Model predicts that English /d/ would constitute a problem precisely because Spanish /d/ is similar to English /d/ (but differs in phonetic detail) and therefore it is not easy for the learner to establish a separate category for English /d/.

Finally, as mentioned earlier, although Lado was aware of the effects of knowledge of allophonic variation as a function of syllable position and phonetic context in interlingual identification, the principle does not seem to have been extensively incorporated in the application of his model to predict problem areas for the learner. On the other hand, the principle has been, either explicitly or implicitly, incorporated in both Best and Flege's models. Flege (1995: 233, 238-9) states that L1 and L2 are related perceptually to one another at a 'position sensitive allophonic level', and regards context-dependent phonetic segments as the appropriate level of analysis. Although less explicitly, Best's Perceptual Assimilation Model seems to take the same stance, for the level of analysis employed by Best and her associates has been at the context-dependent phonetic segment level.

Thus, the current models are better founded and more sophisticated in their predictions than the Contrastive Analysis Hypothesis even though they share the major premise that the basic cause of L2 learners' deviant phonetics is L1 interference. Their emphasis on L1 interference is largely justified by the aforementioned perceptual studies attesting L1 influence on the perception of L2 sounds and the failure of interlanguage theorists to convincingly demonstrate that an L1-independent, universal developmental sequence surpasses L1 interference in the learning of L2 phonetics.

However, it must be noted that there are findings suggesting that universal factors also play a role in the acquisition of L2 phonetics. For instance, as mentioned earlier, Johansson (1973) observed that all the 9 language groups she studied tended to replace Swedish high vowels with vowels that were less extreme in height in production. Bohn (1995a), on the other hand, observes that adult learners tend to rely on durational rather than spectral cues in discriminating L2 vowels even if their L1 vowels are not contrasted in length. Specifically, he found that both Spanish and Mandarin listeners relied more on durational cues in distinguishing the English /i/-/ɪ/ contrast, although they do not use such cues in distinguishing their L1 vowels. Based on the above

---

[16] As the perceptual distance between phones differs cross-linguistically and not enough is known for us to quantify the perceptual distance between particular phones for a particular language group, I have taken Lado's observation as true for English and Spanish speakers.

observation, Bohn hypothesises that the use of durational cues to differentiate a new vowel contrast is an L1-independent, universal strategy.

Hecht and Mulford (1982) studied the progress made by a 6-year-old Icelandic boy in his production of English phonemes and observed that the importance of the role transfer processes and developmental processes[17] play in the learning of L2 phonetics differed among different classes of sounds: L1 transfer was more predominant in the production of vowels, while developmental processes were more predominant in the production of affricates and fricatives. Hecht and Mulford's findings are compatible with Scholes (1967, 1968a) and Schouten's (1975) documentation that the listeners tended to map synthetic vowels and nasals into their L1 categories, and the perceptual reference to L1 categories was observed to a lesser extent for liquids, semivowels and fricatives. (For more evidence for developmental processes observed in L2 acquisition, see Flege & Davidian, 1985; Major, 1986). Leather & James (1996: 292-293) write, '[it] appears ... that the differential effects of L1 influence over time on L2 speech acquisition must be evaluated in the light of the competing forces of the typological (markedness) value of L2 phones as well as learner-perceived similarity between the phones of L2 and those of L1'. Given that there are L1-independent, universal factors that shape the learner's phonological progress, their relative importance in the development of L2 phonetics and the ways in which they interact with L1 interference should be incorporated in future models.

2.2 Gauge for Measuring Similarities

As illustrated in the previous section, Best's Perceptual Assimilation Model predicts the degree of perceptual differentiation of L2 contrasts on the basis of their assimilation patterns to L1 phonetic categories and their goodness of fit to these L1 categories. On the other hand, Flege's Speech Learning Model predicts the degree of difficulty of discerning an L2 sound from a similar L1 sound on the basis of the degree of perceived similarity between the two sounds. Thus, degrees of similarity, or the perceptual distance, between L1 and L2 phones is the key concept of both models. However, a gauge for measuring such distance is yet to be established (See Strange, 1995: 81; Bohn, 1995b: 90; Rochet, 1995: 387-390; Leather & James, 1996: 276).

In Best's view, the goodness of fit of an L2 phone to an L1 category is determined on the basis on how similar gestural constellations of the L1 and L2 categories are to each other. Giving an example, she writes, '[f]or a native listener of a language that has no dental stop but does have bilabial, alveolar, and velar stops, the tongue tip constriction of the dental stop is straightforwardly closer in native phonological space to the alveolar place than to others, because the articulation

---

[17] According to Ioup and Weinberger (1987: 419, 421), transfer processes refer to the imposition of the native language (L1) grammar on the structure of the target language (L2), and developmental

involved is the same and the place of constriction is more similar than those of bilabial or velar stops' (Best, 1995: 193-4). However, she does not provide specific rules that are generally applicable to measuring the perceptual distance between two gestural constellations. Moreover, it would be difficult to explain, at the above level of analysis, such cases where Japanese and Spanish speakers substitute /s/ (alveolar fricative) for English /θ/ (dental fricative), while Thai and Tagalog speakers substitute /t/ (alveolar stop) for the same phoneme,[18] when all the four language groups have both /s/ and /t/ in their L1 phoneme inventories, should the substitution derive from similarity between L1 and L2 gestural constellations.[19]

In comparison to Best's gauge for similarity, the gauge suggested in Flege & Hillenbrand (1984: 708) is more generally applicable. They tentatively assume that 'equivalence classification' of phones takes place between L1 and L2 phones transcribed by the same IPA (International Phonetic Alphabet) symbol because of their overall phonetic similarity (also see Flege, 1986:10). That is, L1 and L2 phones are regarded similar by the learner when they are transcribed by the same IPA symbol. For example, '... instances of /t/ occurring in French and English words are likely to be regarded by the learner as being different realisations of the same category' (Flege & Hillenbrand, 1984: 708). However, the above gauge would grade L2 sounds either categorically similar to, or dissimilar from, L1 sounds regardless of the variability in phonetic detail of the sounds transcribed by the same IPA symbol. If L1 and L2 sounds are transcribed by the same IPA symbol, they will be regarded similar and therefore differentiation between the two will be difficult; if not, they will be regarded different and discrimination will be easy. Thus, it does not, for example, predict different degrees of difficulty for English and Hungarian speakers in discerning Spanish /t/ from /t/ in their respective L1s, despite the fact English VOT (voice-onset time) for /t/ is much more different from the Spanish one in comparison to Hungarian VOT: Typical Spanish, English and Hungarian VOTs for syllable-initial /t/ are reported to be around 9 ms, 70 ms and 16 ms, respectively (average values taken from Lisker & Abramson, 1964). As already mentioned, in his more recent work, Flege (1995) emphasises that interlingual identification takes place at a position sensitive allophonic level rather than at a phoneme level; However, he does not provide a gauge that can be used to predict degrees of similarity between L1 and L2 phones in a non-circular fashion.

In order to predict different degrees of difficulty in perceptually learning L2 phones/contrasts based on the current models, a more sophisticated gauge for measuring degrees of similarity, or the perceptual distance, between L1 and L2 phones needs to be established. Such a

---

processes refer to a natural simplification of a difficult target phonological contrast utilised in child L1 acquisition, which often derives from universal properties of language.

[18] Such cases are reported in Lado (1957: 24) and Flege (1995:267). Rochet's (1995) study described earlier provides another example of two language groups substituting different L1 phones for a given L2 sound even though both groups posses both phones that are differentially substituted.

[19] Weinberger (1990) explains such differential substitutions from a different perspective; in the framework of Underspecification Theory (Archangeli, 1984) the least specified segments of the learner's phonological system is substituted for the L2 forms.

gauge should dictate: (i) between what exactly the measuring should take place (supposing that individual L2 phones the learner actually hears constitute one end of the distance, what constitutes the other end?); and (ii) how to quantify the distance between L1 and L2 sounds. As is explained below in more detail, the acoustic (and even psychoacoustic) distance between the components of speech sounds cannot be transformed into perceptual distance between the speech sounds in a straightforward manner. Apparently, the perceptual distance between speech sounds is not necessarily proportional to the acoustic (or psychoacoustic) distance between their components. How, then, is the acoustic/psychoacoustic distance between the components of speech sounds mapped onto perceptual distance? In the following, I discuss the above two points with reference to recent findings in speech perception.

2.2.1 The Locations of Prototypes (Reference Points)

In order to measure the perceptual distance between L1 and L2 phones, one needs to know between what exactly the measuring should take place. Supposing that the individual L2 phones the learner actually hears constitutes one end of the distance, what constitutes the other end? Seemingly, Best's Perceptual Assimilation Model and Flege's Speech Learning Model both assume it to be mental representations of L1 phones, with which L2 phones are compared. In Best's term, it is the 'native "ideal"', from which degrees of deviation of L2 phones are calculated (Best, 1995: 195). In Flege's term, it is 'central phonetic representations' (Flege, 1984: 692) or the 'perceptual target' (Flege, 1986: 31-2) with which L2 sounds are either identified or not identified. Thus, both Best and Flege seemingly assume that some kind of abstract, mental representations of L1 phonetic categories (commonly referred to as phonetic prototypes in the field of speech perception) serve as reference points in determining the perceptual distance between L1 and L2 phones. In other words, both models appear to be based on the prototype theory of speech perception which holds that listeners compare the incoming signal to a representation, or a template, stored in long-term memory (Massaro & Cohen, 1977; Oden & Massaro, 1978; Massaro & Oden, 1980a, 1980b; Nosofsky, 1986; also see Lively & Pisoni, 1997: 1665). Both Best and Flege take consonantal contexts into account in studying interlingual identification, and therefore the level of abstraction of Best and Flege's representations is different from that of the original prototype theory of speech perception, which typically discusses prototypes at the phoneme level. However, in my view, their models can be seen as an extension of the prototype theory in that they are based on the concept of the mental representation of a category rather than category boundaries or indefinite number of category exemplars.

The prototype theory of speech perception is often contrasted with the exemplar theory of speech perception which holds that the individual exemplars we actually encounter in experience (e.g., words produced by different talkers on different occasions) are stored in memory and that there is no abstract representation that stands for a given category (Hintzman, 1986; Medin & Barsalou, 1987; Nosofsky, 1988; Jusczyk, 1993: 7, 19). According to the exemplar theory, newly-encountered items act as retrieval cues to access the exemplars stored in memory, among which the ones that are most similar to the newly-encountered items are activated most strongly. Though seemingly discrepant, the prototype and exemplar theories are not necessarily incompatible with each other. For example, some proponents of exemplar-based model of speech deem that individual exemplars in each category are organised in accordance with degrees of similarity to one another, in which case, the exemplar that shares most in common with the rest of the category members can be compared to the prototype (Nosofsky, 1986: 56, Nosofsky, 1988; Nosofsky et al., 1989). According to their view, although a 'prototypelike' effect may be observed when several exemplars sharing characteristics are activated simultaneously in the process of cognition, the prototype is not stored in memory as an abstract representation.

Returning to the issue of measuring the perceptual distance between L1 and L2 sounds, the prototype theory presupposes a representation of a given L1 category, in other words a reference point, to be fixed, while in the framework of the exemplar theory it is not clear whether there are such fixed points from which the perceptual distance is calculated. In the present dissertation, following Best and Flege, I assume that there are mental representations of L1 sounds (hereafter called phonetic prototypes) at some level that serve as reference points in determining the perceptual distance between L1 and L2 sounds. Supposing that phonetic prototypes constitute the reference points in measuring the perceptual distance between L1 and L2 phones, how would we locate such prototypes?

The concept 'prototype' was introduced to the field of speech perception from Rosch and her colleagues' prototype theory (1975, 1981), which is based on studies on visual objects. Studies on visual objects (e.g., colour, geometric shape categories) indicate that category members differ from one another in a gradient fashion in terms of their representativeness as category members. Representative members, or prototypes, of visual categories are found to have privileged status in the process of cognition. For example, they are easily learned and remembered, and quickly processed (Mervis & Rosch, 1981). In the field of speech perception, it has been also shown that listeners perceive goodness of various members of a certain phonetic category to be different, which has been interpreted as an indication that the members of phonetic categories also differ from one another in their degrees of representativeness as category members (Kuhl, 1986; Miller & Volaitis, 1989; Kuhl, 1991). Furthermore, it has been claimed that representative members of phonetic categories, i.e., phonetic prototypes, also have a privileged status in certain perceptual tasks. For instance, it has been reported that prototypes are found to be better competitors in dichotic competition experiments and more effective adaptors in selective adaptation experiments (Miller, 1977; Repp, 1977; Samuel, 1982; Miller et al., 1983).

As briefly mentioned, in Rosch's framework, representativeness, which defines prototypes, is operationally defined by means of subjects' ratings of how good an example an item is of the category (Rosch, 1975). Mervis & Rosch (1981) write, 'Consistency in such ratings has been obtained. Individual subjects agree that some exemplars of a category are more representative than others, and different subjects consistently choose the same exemplars as most representative of the category' (Mervis & Rosch, 1981: 96). At the same time, prototypes have been assumed to be located at the heart of the distribution of the category members they represent. Thus, prototypes are regarded members possessing either average feature values (average prototype) or feature values observed most frequently (frequency prototype) (See, for instance, Goldman & Homa's 1977 study on recognition of schematic faces).

Apparently, the assumption that prototypes are those category members that are perceived as 'best' and at the same time situated at the centre of the category in the geometrically represented stimulus space was also adopted in the field of speech perception. For example, in describing phonetic prototypes Samuel (1982) writes, ' ... CV syllables whose consonant was a "good" (i.e., central) category exemplar ...' (Samuel 1982: 307, my emphasis). Similarly, Kuhl (1993) describes phonetic prototypes as instances that are 'good' and 'typical' (Kuhl 1993: 126). Thus, the production average, which is by definition located at the centre of the distribution of category members in the acoustic space, and goodness rating tasks, which elicit category ideals,[20] have been alternatively used in determining the locations of phonetic prototypes (e.g., Samuel, 1982; Repp & Crowder, 1990). Indeed, Kuhl (1991) reports that the first two formant-frequency values (F1 and F2)[21] of American listeners' best-rated /i/ stimulus coincided with the average values of /i/ produced by male speakers of American English reported in Peterson and Barney (1952). Moreover, as can be seen in Fig. 2:1, the goodness-ratings of variants of /i/ systematically decreased, as the distance between the stimuli and the best-rated stimulus increased, implying symmetrical structure of the category with the production average in its centre.

---

[20] I use the term 'category ideal' to refer to perceptually preferred members of categories to distinguish them from the statistical average of category members based on production data. As is shown below, these two may not be the same as far as vowels are concerned.

[21] F1 and F2 are known to convey the most crucial acoustic information regarding vowel quality, F1 correlating inversely with vowel height, and F2, with vowel backness (cf., Pols et al., 1969).

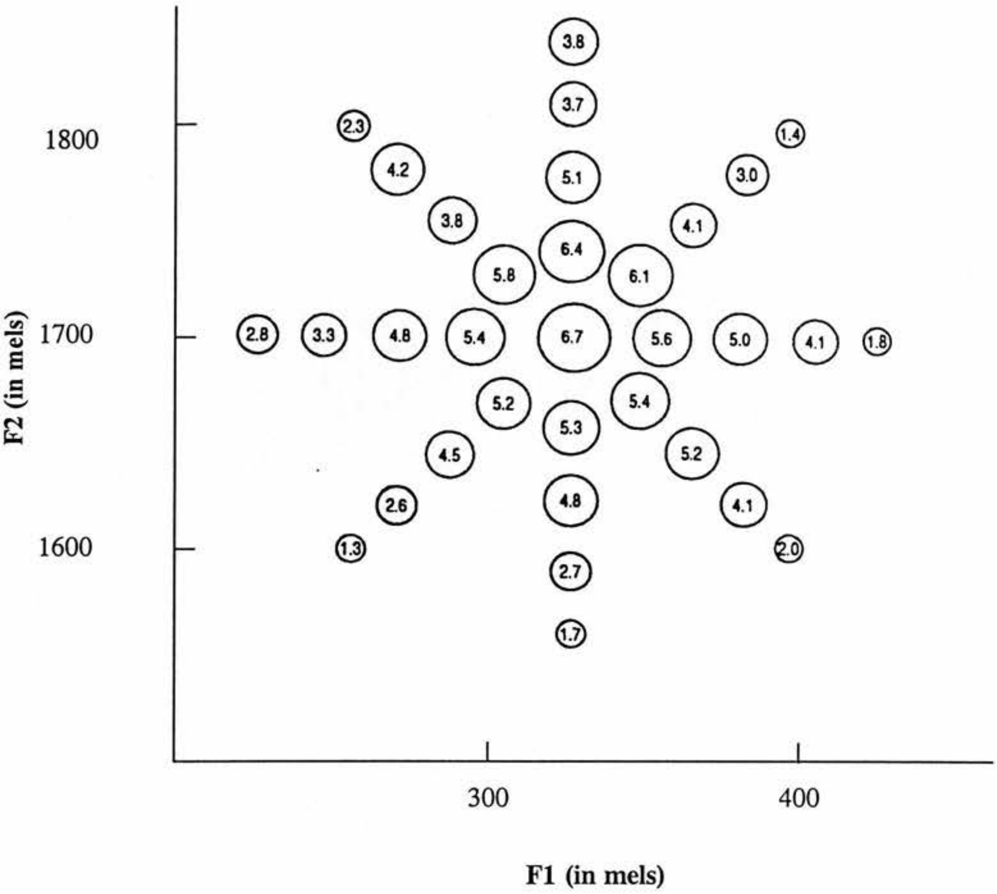


Figure 2:1. Category goodness (typicality) ratings for the prototypical /i/ and variants surrounding the prototype. The goodness was judged using a scale from 1 (a poor exemplar) to 7 (a good exemplar). (From Kuhl, 1991: 94; the x- and y- axes were added by the author).

However, evidence from other studies on phonetic prototypes suggests that the production average and the best-rated exemplar of the phonetic category may not always match. For example, in a study using an identical stimulus set to Kuhl's (1991), Lively (1993) reports that American listeners gave higher ratings to more extreme (in this case, higher and more front) /i/-stimuli than Kuhl's (1991) best-rated stimulus that matched Peterson & Barney's (1952) production average. He also reports that listeners' choices were more variable than Kuhl (1991) indicates. In later replications of Kuhl (1991), listeners also preferred stimuli that were more extreme than Peterson & Barney's (1952) production average (cf. Iverson & Kuhl, 1995; Lotto et al., 1996; Frieda, 1997; Lively & Pisoni, 1997). The listener's preference for extreme vowels is shown more extensively in Johnson et al.'s (1993) study where 11 English vowels /i, ɪ, eɪ, ɛ, æ, ʌ, ɑ, ɔ, ou, ʊ, u/ were studied; subjects systematically chose, as the best vowel sounds, stimuli that were more extreme than their own production of these vowels. Furthermore, Bradlow (1993) reports that both English and Spanish listeners exhibited preference for vowels at extremes and gave high ratings to vowel stimuli at the outer edge of an F1-F2 vowel space,[22] even to the stimuli that fell outside the production range of their respective L1 vowels (Bradlow, 1993: 78-83). Thus, as can be seen in Fig. 2:2 taken from Bradlow (1993), the above studies did not yield a picture of vowel categories that were symmetrically structured around the production averages as indicated in Kuhl (1991).

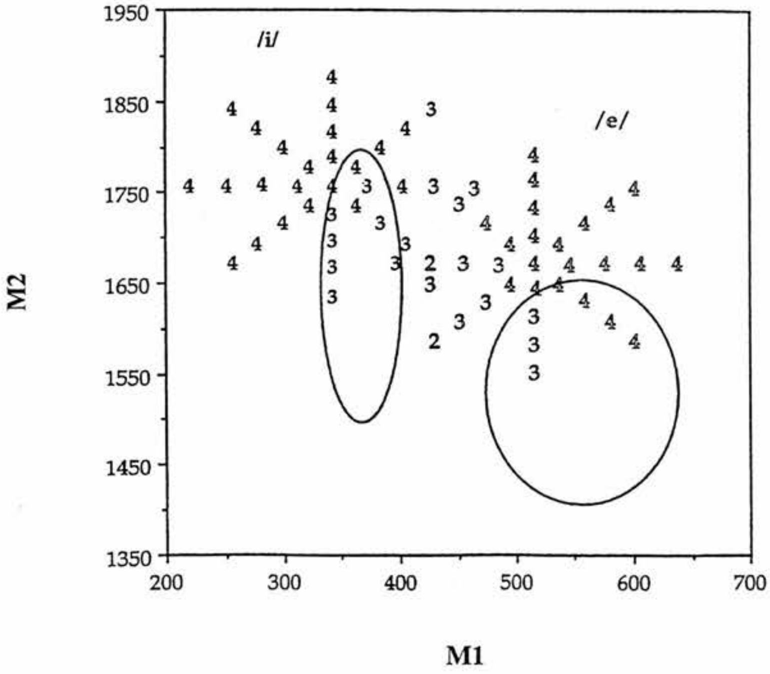


Figure 2:2. Spanish listeners' ratings of English /i/-/e/ stimuli. M1 and M2 stand for F1 and F2 in mels, respectively. The numbers in bold are the goodness ratings of the category /i/, and the outlined numbers are those of /e/. A rating of '1' indicates a 'very bad and unclear' exemplar, and a rating of '5' indicates a 'very good and clear' exemplar. The ellipses are drawn around the Spanish /i/ and /e/ categories from production data. (Adapted from Bradlow, 1993: 82).

Thus, evidence seems to be in favour of the view that vowel category ideals are more extreme than production averages. However, the above studies should be interpreted with caution, as they all used synthesised, steady-state vowel stimuli for perception tasks, while vowels in the production data were in consonantal contexts, which could have led to the separate locations of the production averages and category ideals. In Peterson & Barney (1952) and Johnson et al. (1993), vowels were produced in an 'hVd' context. In Bradlow (1993), English vowels were produced in a 'bVd' context, and Spanish vowels, in a 'bVda' context. Thus, the vowels in these production tasks may have been less extreme than their isolated counterparts in the perception tasks due to the coarticulatory effects, i.e., the mutual influence of neighbouring sounds [see Lindblom (1963) and

---

[22] Conventionally, the acoustic vowel space has F1 and F2 as co-ordinates, on which the estimated centre frequencies of the F1 and F2 values of the steady-state portion of the vowel production (or the prominence in the spectrum) are plotted (see, for instance, Peterson & Barney, 1952).

Stevens & House (1963) for the 'undershoot' of vowels due to such effects]. In other words, the materials in the production tasks and stimuli in the perception tasks in the above studies were not strictly comparable. Although the 'hVd' context employed in Peterson & Barney (1952) and Johnson et al. (1993) has been referred to as the 'null context' and regarded to yield F1 and F2 values that are not different from those of isolated vowels, /d/ following a vowel can lower F1 of the preceding vowel and centralise F2 of front and back vowels towards its locus, i.e., around 1800 Hz, when the speaking rate is fast. In fact, data presented in Stevens & House (1963: 115-116), where three speakers of American English produced vowels both in an 'hVd' context and in isolation, suggest that F1 and F2 values can differ in the two contexts, with the difference in formant values varying from 0 ~ 100 Hz for F1 and from 0 ~ 240 Hz for F2 for various vowels produced by different speakers, although in this particular report the differences between the two conditions were not significant when they were averaged across speakers. Thus, it is conceivable that the two conditions yield different formant values, if the degree of coarticulation is great in the 'null context', either due to a fast speaking rate or individuals' speaking styles (cf. Stevens & House, 1963: 124).

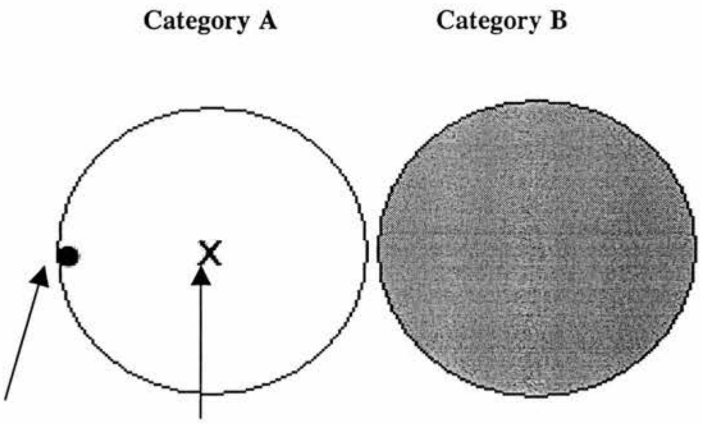
Furthermore, it is possible that the listeners in the above studies gave higher ratings to extreme vowels because they interpreted the term 'good' as 'distinct', which departs from the original definition of prototypes, i.e., representative. Johnson et al. (1993: 517) had listeners choose and rate vowel stimuli that were the 'best' vowel sounds in one condition and those 'that matched their own production' in another, and found that the listeners gave higher ratings to the same stimuli in the second condition. Thus, the production averages may be elicited from listeners as category ideals, given different instructions, e.g., 'choose typical vowel sounds' rather than 'best vowel sounds'. (For a similar view, see Aaltonen et al., 1997: 1096). For instance, Elsendoorn's (1984) Dutch subjects were instructed to adjust the duration of synthesised vowels to a 'correct' value and produced synthesised vowels whose duration matched production values.[23] In this case, although the match between perception and production could have been due to the feature of vowels under investigation, i.e., vowel duration, or the compatible materials used for the two tasks (the same CVC words were used for both tasks), it is also possible that the match was due to the instructions given to the subjects.

At the same time, as for vowel categories, it is conceivable that category ideals and production averages do not match. According to Mervis & Rosch (1981: 98), prototypes 'have most attributes in common with other category members', but they also have 'fewest attributes in common with related contrast categories'. In the case of most vowels (except for vowels like schwa which are surrounded by other vowel categories), the production average satisfies the first description, while extreme vowels fit the second, considering the distributional structure of the F1-

---

[23] This was the case, however, only with their L1 (Dutch) vowels. The subjects chose durations that did not match their production of L2 (English) vowels.

F2 vowel space.[24] That is, most vowels are located along the periphery of the acoustic vowel space and tend not to have a neighbouring category on both sides along F1 and F2. As can be seen in Fig.2:3, when a category has only one neighbouring category along a given dimension,Mervis & Rosch's (1981) description of prototypes designate two different locations of prototypes. Thus, there are two possible prototype locations for vowel categories, i.e., the category ideal and the production average.[25] If the locations of vowel category ideals and production averages do not match, a further question needs to be addressed: Assuming that a prototype is used as a reference point in determining the perceptual distance between L1 and L2 vowels, is it the category ideal or the production average that is used as a reference point? I come back to this issue in the next section.



A category member that has fewest attributes in common with contrast category B.

A category member that has most attributes in common with other category members.

Figure 2:3. Two possible prototype locations when a category does not have contrast category on both sides. The two large circles represent hypothetical contrast categories, Category A (white circle) and Category B (grey circle). The filled circle and the cross each represents the locations of two possible prototypes of Category A in accordance with Mervis & Rosch's (1981) description of category prototypes.

Supposing that vowel category ideals are more extreme than production averages, another question arises when one attempts to apply the prototype theory to L2 phonetics models: Are the

---

[24] I limit the discussion to F1 and F2 of vowels, for I have not found studies suggesting the possibility that the production average and the category ideal may not match for other aspects of vowels or classes of sounds. In the case of vowel duration and VOT of stop consonants, for example, it is reported that perceptually preferred values match typical values in production as far as L1 is concerned (Elsendoorn, 1984; Flege, 1995: 260; Flege & Schmidt, 1995).

[25] Lotto et al. (1996: 9-11) report that European Starlings preferred, among a vowel stimulus set with which they were positively reinforced, those which were most dissimilar from the opposing

locations of category ideals language specific? Conventionally, phonetic prototypes have been assumed to derive from the language input one receives. They are thought to '[develop] as an individual hears many tokens of a sound' (Flege, 1986: 31) and be 'based on experience' (Johnson et al., 1993: 516). Some further assume that phonetic prototypes are representations of the ambient language input that reflect distributional properties of the input (e.g., Kuhl, 1993: 130), suggesting that prototype locations are within the speakers' production range of the categories. In this view, vowel prototypes may be more extreme than the production averages but should be within the production range of given vowel categories. In fact, Frieda's (1997) study suggests that the F2 of the category ideal of English /i/, although it was more extreme than that of the normal production, did not differ from that of the hyperarticulated (or exaggerated) production,[26] suggesting that the category ideals may be located at the extremes of the production range. On the other hand, others think that prototypes are abstracted by the perceiver rather than realised in the physical environment and that representations are composed of ideal features or dimensional values for objects in the category (e.g., Oden & Massaro, 1978), implying that category ideals may be located outside of the speakers' production range of the category it represents. Indeed, Bradlow's (1993) study showing listeners' preference for vowels that are more extreme than the production range of their language suggests that this may be the case. If this is the case, however, whether or not the cross-linguistic differences observed in the phonetic realisations of a category are reflected in the locations of these category ideals is questionable. In fact, Hoffman (1973) reports that Japanese and English listeners chose the same exemplar as the best or 'focal' instance of the category /i/ despite possible cross-linguistic differences in the phonetic realisation of the category.[27] Thus, vowel prototypes may be universal as colour prototypes have been said to be and may not be language-specific (cf. Mervis & Rosch, 1981: 95). If vowel prototypes are not language-specific, Best and Flege's L2 phonetics models, which draw on language-specificity of the L1 prototypes, should modify one of their major assumptions.

Alternatively, it may be the case that subjects' goodness ratings are sensitive to the stimulus range used in the task and therefore do not yield absolute locations of category ideals. For instance, Lively & Pisoni (1997: 1669, 1672) report that individual American subjects' best-rated /i/ stimuli

---

stimulus set, which they see as analogous to human listeners' preference for extreme vowels and the structure of phonetic categories.

[26] It has been reported that in hyperarticulated, or exaggerated speech, vowels are realised with F1 and F2 values that are more extreme than those in normal speech. Hyperarticulated speech can be elicited from speakers by instructing them to speak 'clearly', 'as when communicating with a non-native listener', etc. (cf., Picheny et al., 1986; Moon & Lindblom, 1989; Johnson et al., 1993).

[27] According to Tsujimura (1996: 18), Japanese /i/ is not accompanied with lip spreading like English /i/, which is known to affect the quality of vowels (see Jones, 1950). In fact, a comparison of the F1 and F2 values of isolated /i/ obtained from speakers of American English reported in Stevens & House (1963: 115) and those obtained from Japanese speakers in the present study suggests that /i/ may be realised with more extreme F2 values in American English than in Japanese, with the mean value of the former being 2340 Hz and the latter, 2226 Hz.

varied depending on the stimulus range. The effect of stimulus range on goodness-ratings of vowel stimuli is also reported by Galama & Nearey (1995).[28] These findings are not so surprising, considering the context sensitive nature of vowel identification (cf. Ladefoged et al., 1957; Fry et al., 1962; Repp et al., 1979; Repp & Liberman, 1987: 92; Nearey, 1989).

2.2.2 Factors that Determine Perceptual Distance Between L1 and L2 Vowels

Supposing that we know the locations of the L1 vowel prototypes in an F1-F2 vowel space that are used as reference points in determining the perceptual distance between L1 and L2 vowels, how would we measure the distance between the L1 prototype and an L2 vowel? The psychoacoustic distance between two sounds has been shown to be different from the acoustic distance measured using an arithmetic frequency ratio even in simple nonspeech stimuli such as sinusoidal signals, or pure tones (e.g., Beranek, 1949; Fant, 1973; Nearey 1976, 1978; Gulick et al., 1989). For instance, it is known that the pitch of sound (a sensation that can be described along a high-low dimension) varies as a function of frequency, but the perceived pitch of a sound is not always linearly related to its frequency. The relationship is almost linear at low frequencies (below 1000 Hz) but logarithmic at higher frequencies (above 1000 Hz). Thus, two kinds of pitch scales, i.e., the units of mel and bark, were established as psychoacoustic scales that reflect the ear's sensitivity to differences in pitch. The mel scale is determined on the basis of the listeners' judgements of the distance between pure tones of given frequencies; For instance, a pure tone of 100 mels is heard as half a tone of a tone of 200 mels. On the other hand, the bark scale is determined on the basis of the masking effects, i.e., the width of a band of noise within which a pure tone of a given frequency is no longer audible (cf. Fant, 1973: 47; Gulick et al., 1989: 246-247; Ladefoged, 1996: 80).

The issue of determining the perceptual distance between speech sounds is even more complicated, as shown in categorical perception experiments (e.g., Liberman et al., 1957, 1967; Fujisaki & Kawashima, 1969, 1970). Categorical perception experiments have shown that in the speech mode[29] stimuli that are acoustically evenly spaced are better discriminated around the category boundary, whereas the stimuli are not as readily discriminated within categories. In other words, the perceptual distance is greater across categories than within categories. According to

---

[28] The effect of contexts on goodness ratings reported in Galama & Nearey (1995) appears to be of a smaller magnitude than that reported in Lively & Pisoni (1997). This may be due to the fact that Galama & Nearey used stimuli that differed orthogonally in F1 and F2. Considering Lively & Pisoni's report that the context effect on goodness ratings was greatest when F2 alone was varied, the fact that F1 and F2 varied together in Galama & Nearey (1995) may have limited the context effect on goodness ratings.

[29] The speech mode of perception is thought by some to be different from the auditory mode of perception. For example, as described earlier, the Japanese listeners, who could not detect the difference in F3 when it was presented as a component of a /ra-la/ stimulus continuum (the speech mode), were nevertheless able to detect the difference in F3 when it was presented in isolation as non-speech stimuli (the auditory mode) (see Miyawaki et al., 1975).

experiments on cross-linguistic categorical perception, these boundary locations are language-specific (Abramson & Lisker, 1970; Goto, 1971; Miyawaki et al., 1975; Gillette, 1980; Sheldon & Strange, 1982). That is, the perceptual distance between speech sounds stretches at language-specific category boundaries, which means that the perceptual distance between speech sounds that straddle category boundaries cannot be accounted for simply in terms of the psychoacoustic distance between the components of the stimuli. Considering the language specificity in the locations of perceptual category boundaries which apparently warp the perceptual space, cross-linguistic differential substitution (some language groups' substituting one L1 phoneme and other groups' substituting other L1 phonemes for the same L2 phoneme) may be a reflection of cross-linguistic differences in the ways the perceptual space is warped by language-specific category boundaries. In fact, Rochet's (1995) study on English and Portuguese speakers' perception and production of French /y/ shows that the different locations of category boundaries between the /i/ and /u/ categories in English and Portuguese can explain the different patterns of interlingual identification observed in the two language groups, resulting in different patterns of L1 substitution for the L2 sound. Specifically, English speakers' category boundary between /i/ and /u/ is higher in F2 (around 1900 Hz) than that of Portuguese speakers (around 1600 Hz), and as a consequence, French /y/ whose F2 is approximately from 1200 Hz to 2100 Hz is equated to /u/ in perception and produced as such by English speakers, while it is equated to /i/ in perception and produced as such by Portuguese speakers.

According to Kuhl's Native Language Magnet Theory (Kuhl, 1991, 1993; Kuhl et al., 1992; Iverson & Kuhl, 1995), it is not only around the category boundaries that the perceptual space for speech sound is warped. The Native Language Magnet Theory holds that language-specific mental representations of phonetic categories, or phonetic prototypes established in the course of L1 acquisition, attract nearby members of the category, causing the perceptual space to shrink around these prototypes. Kuhl further holds that the assimilation by L1 prototypes interferes with adults' ability to perceive certain L2 sounds. L1 prototypes assimilate L2 sounds in their vicinity of the perceptual space, and, as a result, L2 sounds become indistinguishable from these L1 prototypes. Kuhl writes, '... the nearer a new (L2) sound is to a native-language magnet [i.e., the prototype] the more it will be assimilated by it, making the new sound indistinguishable from the native-language sound' (Kuhl, 1993: 131).

Empirically, the Native Language Magnet Theory is based on a series of experiments on phonetic prototypes by Kuhl and her colleagues. As mentioned earlier, Kuhl et al. conducted a series of experiments to explore the internal structure of phonetic categories and found that different instances of the vowel category /i/ (synthesised /i/ stimuli varying in F1 and F2) varied in the degrees of goodness perceived by American listeners (Kuhl, 1986; Grieser & Kuhl, 1989; Miller & Volaitis, 1989). In a subsequent experiment (Kuhl, 1991), American listeners exhibited difficulty in perceiving differences between the best-rated stimulus, which was designated as the prototype, and other members of the category. On the other hand, the listeners' discrimination sensitivity away from the prototype was found to be good. On the basis of the above results, Kuhl (1991) hypothesises that phonetic prototypes assimilate neighbouring, other members of the category, pulling them towards these prototypes [this effect was termed the 'perceptual magnet effect'; see Samuel (1982: 307) for a similar view]. In order to show that the above effect derives from language-specific prototypes, Kuhl et al. (1992) further tested 6-month-old Swedish and American infants with stimuli centred around American-prototype /i/[30] and Swedish-prototype /y/, and obtained results that infants from both countries showed a significantly stronger magnet effect only for their native-language prototype.

Should language-specific vowel prototypes cause the perceptual vowel space to shrink towards themselves, as the Native Language Magnet Theory holds, one would need a scale that takes account of the location of the L1 prototype and its assimilation effect in order to measure the perceptual distance, or decide the degree of similarity, between certain L1 and L2 vowels. Conversely, the issue of which of the two possible prototype locations, i.e., the category ideal or the production average (if they are different), serves as a reference point in determining the perceptual distance between L1 and L2 vowels may be solved by probing which of them exhibits the assimilation effect. However, whether the poorer discrimination around the prototypical stimulus in comparison to a non-prototypical stimulus observed in Kuhl (1991) [hereafter Kuhl's (1991) P and Kuhl's (1991) NP] is due to its prototypicality is open to question.

First of all, as mentioned earlier, Kuhl's (1991) P that matched Peterson & Barney's (1952) production average has not been rated as the best stimuli by American listeners in the replications of Kuhl (1991). Lively (1993), Iverson & Kuhl (1995), Lotto et al. (1996), Frieda (1997) and Lively & Pisoni (1997) all report that American listeners gave higher ratings to more extreme /i/-stimuli than Kuhl's (1991) P. Thus, whether Kuhl's (1991) P was in fact prototypical to American listeners is questionable, should it be category ideals that play a special role in speech perception.

Furthermore, a number of studies replicating Kuhl (1991) indicate that Kuhl's (1991) NP was not perceived as an instance of /i/ by American listeners as reported in Kuhl (1991). In Sussman & Lauckner-Morano (1995), for instance, Kuhl's (1991) NP was transcribed as [i] only 8% of the time by American subjects in a phonetic transcription task and judged to be [i] only 20% of the time in a forced-choice [i]/not-[i] task. Lively (1993), Iverson & Kuhl (1995), Lotto et al. (1996) and Lively & Pisoni (1997) all report similar results. These results imply that the better discrimination sensitivity observed around Kuhl's (1991) NP perhaps is due to cross-category discrimination that took place around Kuhl's (1991) NP whereas discrimination around Kuhl's (1991) P was poorer since it was within-category. In fact, in their replication of Kuhl (1991), Lotto

---

[30] Although Swedish also has /i/ in its vowel inventory, Swedish listeners gave the American-prototype /i/ an average rating of 1.8 as Swedish /i/ and 2.6 as Swedish /e/ in a goodness rating task

et al. (1996) report that when the contrast effects on stimuli presented in pairs were taken into account, viz., when the stimuli were identified in pairs as they were presented in the discrimination task, the better discrimination sensitivity around Kuhl's (1991) NP could be fully accounted for as arising from cross-category discrimination, i.e., it was due to the listener's labelling of the paired stimuli into different phonemic categories.

The questionable identity of Kuhl's (1991) P and NP motivated researchers to replicate Kuhl (1991) with more appropriate P/NP, the results of which lead us to further questions. Sussman & Lauckner-Morano (1995) gave listeners a 'change/no-change' discrimination task, using Kuhl's (1991) P (i.e., the production average) and stimuli that fell between Kuhl's (1991) P and NP, which they considered better candidates for non-prototypical members of the /i/ category than Kuhl's (1991) NP. They report that discrimination sensitivity was significantly poorer around Kuhl's (1991) P in comparison to the other stimuli when a bias-free measure of d' was used.[31] Sussman & Lauckner-Morano (1995: 550) write, '[the] results for this investigation were partially in agreement with Kuhl's suggestion that discrimination is poorer around a prototypical stimulus'. However, drawing on previous studies showing improvement in auditory discrimination sensitivity from /i/ to /ɛ/ (e.g., Macmillan et al., 1988), a direction along which their stimulus continuum stretched, Sussman & Lauckner-Morano suggest that the poorer discrimination observed around Kuhl's (1991) P may be due to a general decline in auditory sensitivity towards extreme /i/. Schouten & van Hessen's (1992) study using a stimulus continuum stretching from /i/ to /a/ also suggests a decline in auditory sensitivity towards the extremity of the vowel space.[32] Indeed, in Iverson & Kuhl (1995) the discrimination sensitivity curve declines towards extreme /i/, beyond Kuhl's (1991) P, which may be an indication of a correlation between poor discrimination sensitivity and vowel extremity, and not prototypicality (cf. Fig. 2:5). However, Iverson & Kuhl's (1995) subjects gave best-ratings to those most extreme stimuli towards which their discrimination sensitivities declined, based on which Iverson & Kuhl (1995: 556) argue that category goodness (and hence prototypicality) influences discrimination within the category. That is, discrimination sensitivity does not decline towards the production average, but towards the category ideal, which is another possible interpretation of their results. On the other hand, Aaltonen et al. (1997) report that Finnish subjects' discrimination sensitivity for the /i/ category was poorer at a relatively low F2 location irrespective

---

using a scale from 1 (poor) to 7 (good). The American-prototype /i/ was given an average rating of 5.4 as an English /i/ by American listeners (Kuhl, 1992: 608).

[31] In Sussman & Lauckner-Morano (1995) the difference was not significant when miss rates were used as a measure of discrimination sensitivity. Different measures of discrimination sensitivity used in the prototype experiments are discussed in more detail later.

[32] Sussman & Lauckner-Morano (1995) also cite Schouten & van Hessen (1992), but as showing an improvement in auditory sensitivity from /i/ to /a/. However, I could not find such evidence in Schouten & van Hessen's study. Rather, their discrimination results seem to indicate a decline towards the extremity of the vowel space. Seemingly, Sussman & Lauckner-Morano misunderstood Schouten & van Hessen's (1992: 1848) graphic showing cumulative discrimination results starting from the /i/ end of the continuum to the /a/ end.

of the locations of the individuals' category ideals. Aaltonen et al.'s results may be, however, attributable to F3 of the stimuli that was fixed at 3010 Hz, which was very close to the F2 values at the extreme range of the stimulus continuum (the highest F2 value in their study was 2966 Hz). As Chistovich and her colleagues (1979) have shown, the listener perceives two formants as one when the formants are within a distance of 3.5 Bark. Apparently, these closely spaced, perceptually integrated formants form a more salient cue than widely separated formants; according to Traunmüller's (1981) study, more variability in formant separation is tolerated by the listener for widely separated formants than closely spaced formants. Therefore, the better discrimination sensitivity around an extreme stimulus demonstrated by Aaltonen et al.'s (1997) subjects may be due to the stimulus continuum whose F2 and F3 were close together at its extreme end. Altogether, it is unclear from studies to date whether it is vowel prototypicality or extremity that is correlated with poor discrimination sensitivity.

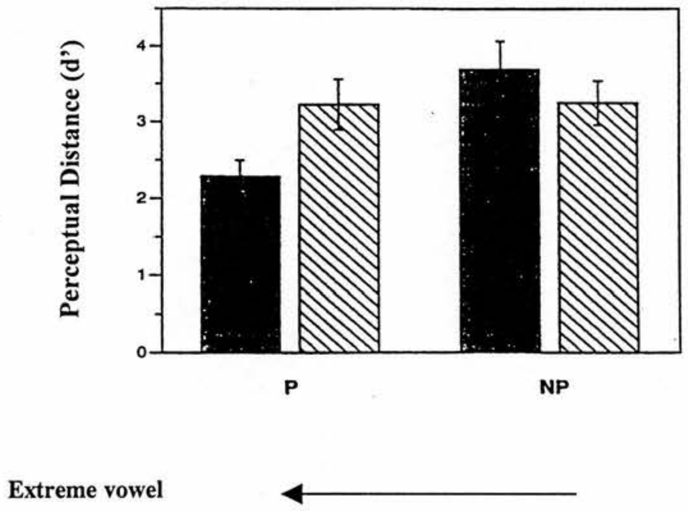


Figure 2:4. Average d' scores showing American listeners' discrimination sensitivity around Kuhl's (1991) P. P stands for Kuhl's (1991) P, and NP, for Kuhl's (1991) NP. (Adapted from Iverson & Kuhl 1995: 557).

Furthermore, whether the reported 'perceptual magnet effect' can be observed at the level of phonetic coding[33] needs to be examined, as all the replications of Kuhl (1991) including Sussman & Lauckner-Morano (1995) and Iverson & Kuhl (1995) used long vowel stimuli (500 ms), which are known to elicit the auditory rather than the phonetic mode of coding. Auditory coding is the mode that is also employed when processing nonspeech sounds (cf. Repp et al., 1979:143) and has been regarded universal, as shown by Miyawaki et al. (1975) where the Japanese subjects, who

---

[33] By 'phonetic coding' I mean the processing of stimuli through mediation of phonetic labels.

could not reliably discriminate between /r/ and /l/ stimuli like the American subjects, nevertheless discriminated between isolated F3 tokens, on the basis of which American subjects distinguished between /r/ and /l/ (also see Best et al., 1981). Phonetic coding, on the other hand, has been known to produce language-specific patterns of speech perception, as shown in a number of categorical perception experiments (e.g., Abramson & Lisker's, 1967, 1973; Goto, 1971; Miyawaki et al., 1975; Gillette, 1980; Sheldon & Strange, 1982).

The concept of phonetic versus auditory modes of coding derives from the dual-coding model, which was originally formulated by Fujisaki & Kawashima (1969, 1970) to explain results obtained from ABX discrimination tests employing vowel stimuli differing in length, where short stimuli elicited more categorical-like perception whereas long stimuli elicited continuous perception. In Fujisaki & Kawashima (1969: 71) these two modes are referred to as 'recognition of discrete category and continuous quality', both of which are utilised in processing speech sounds. Pisoni (1971, 1973), who extended the concepts to apply to the AX paradigm, refer to them as the 'phonetic memory code' and the 'auditory memory code'.

Given the findings that non-speech stimuli can be categorically perceived, and that categorical perception can be observed also in non-humans (cf., Cutting & Rosner, 1974; Miller et al., 1976; Pisoni, 1977), the distinction between phonetic and auditory codings seems more blurred than the terminology suggests. However, it is the former mode of coding where language specificity in speech perception has been observed. From the viewpoint of L2 acquisition where language-specific patterns of perception of L2 categories are observed depending on the L1 background, the perceptual magnet effect should manifest itself at the phonetic level of coding, should it be the cause of the observed language-specific patterns of L2 perception, as the Native Language Magnet Theory maintains. However, in neither Kuhl (1991) nor its replications were the experimental designs optimal for phonetic coding. Thus, whether Kuhl's (1991) findings and its replications have relevance to the phonetic level of speech perception needs to be examined.[34]

In addition, two measures of discrimination sensitivity that have been used in the studies of the perceptual magnet effect, i.e., miss rates and d', are compared in the present study. Miss rates are the rate at which the listener responds 'same' when the members of a stimulus pair is actually different. When miss rates are used as a measure of discrimination sensitivity, no distinction is made between basic sensitivity and response bias, i.e., the listener's willingness to select a certain response in a given task, in this case, respond 'same' or 'different' (cf. Macmillan & Creelman, 1991: 10-33). On the other hand, d' separates basic sensitivity from response bias, which is regarded a 'response strategy' (cf., Macmillan et al. 1988: 1265; Sussman & Lauckner-Morano 1995: 550). As just mentioned, Sussman & Lauckner-Morano (1995: 541) used both miss rates and

---

[34] Although improved discrimination sensitivity from /i/ to /ɛ/ is also found in Repp et al. (1979) using shorter stimuli (240 ms), the stimulus length seems still a lot longer than those produced in

d', and obtained results that discrimination was significantly poorer around Kuhl's (1991) P when the sensitivity was measured in d' but not when it was measured in miss rates. They report, '... d-prime provides a better measure of the perceptual magnet effect than do miss scores' (Sussman & Lauckner-Morano, 1995: 550).

The adequacy of the two metrics as measures of the perceptual magnet effect is assessed in terms of assumptions underlying the two metrics: The use of miss rates as a measure of discrimination sensitivity presupposes that no hits ['different' response to pairs of stimuli that are different (henceforth referred to as D pairs)] are correct by chance, given the equation: (Hit rate) = 1 - (miss rate). If no hits are correct by chance, the false-alarm rate ['different' response to pairs of stimuli that are the same (henceforth referred to as S pairs)] is 0, assuming that half of guessed 'different' responses would lead to hits and the other half, to false-alarms. If the false-alarm rate is 0, the correct-rejection rate ('same' response to S pairs) is 100%, given the equation: (Correct-rejection rate) = 1 - (false-alarm rate). On the other hand, d' was formulated to normalise bias between different observers and compare results from different experimental designs (Luce, 1963: 148; Macmillan, 1993: 24, 42) and presupposes no change in response bias in a single observer within the same experiment. Macmillan (1993: 41-42) states that d' is based on the claim that '(a) [changes] in response bias do not affect sensitivity, and (b) changes in sensitivity do not affect bias'. However, 'the second claim has not been well supported', and that '[subjects] do not appear to agree on which meaning is intended when an experimenter instructs them to maintain a constant response strategy in the face of varying sensitivity'. Thus, although d' seems a more accurate measure of discrimination sensitivity in that it takes account of response bias, which miss rates do not, whether it is in fact a better measure of the perceptual magnet effect is questionable, if its underlying assumptions are not satisfied.

To sum up, the results of recent studies are inconclusive as to the locations of vowel prototypes, which I assume to serve as reference points in determining the perceptual distance between L1 and L2 vowels. As we have seen, this is in part due to the dual description of prototypes, i.e., the 'best' and 'most typical' exemplar ('category ideal' and 'production average'), which may not necessarily match for vowel categories, as discussed above. Furthermore, it is not clear from studies to date which of these two possible prototypes correlates with poor discrimination sensitivity, in other words, which of these two assimilates other category members, supposing that Kuhl's Native Language Magnet Theory is correct. In addition, whether or not the observed assimilation effect manifests itself at the phonetic level of coding needs testing, which is of an interest from the viewpoint of L2 acquisition. In Chap. 3, I present specific questions regarding vowel prototypes on the basis of discussion in this chapter.

---

natural speech or consonants that are shown to be perceived more categorically (i.e., in the phonetic mode) in similar experimental paradigms.

Chapter 3 Research Questions and Experimental Design

As described in Chap. 2, both of the two influential current models of L2 phonetics, i.e., Best's Perceptual Assimilation Model and Flege's Speech Learning Model predict degrees of learners' difficulty in perceptually learning L2 sounds (contrasts) on the basis of degrees of perceived similarity, or the perceptual distance, between L1 and L2 sounds. However, a gauge for measuring such similarity is yet to be established. As explained earlier, both Best's Perceptual Assimilation Model and Flege's Speech Learning Model seemingly assume the existence of mental representations of L1 phones, or L1 prototypes, against which L2 phones are compared and from which the perceptual distance between L1 and L2 sounds is calculated. Then, such a gauge should dictate (i) the reference points representing L1 sounds, from which the perceived distance between L1 and L2 sounds is measured, and (ii) how to quantify the perceptual distance between the two sounds. Based on the discussions in Chap. 2, in the rest of this dissertation I look at (i) two possible locations of such prototypes for vowel categories and (ii) whether perceptual distance in the vowel space shrinks around language-specific prototypes at the level of phonetic coding, as Kuhl's Native Language Magnet Theory holds, or towards its periphery (extreme vowels), as Macmillan (1988) and Schouten & van Hessen's (1992) studies suggest. In 3.1.1, I spell out specific questions concerning the above two issues that are dealt with in this dissertation, and in 3.1.2 I touch on the implications of the answers to these questions for the perceptual learning of L2 vowels. In 3.2.1, I describe Japanese and Greek vowels, which are studied in the present dissertation, and in 3.2.2, I roughly describe experiments designed to answer the research questions posed in 3.1.1.

3.1.1 Research Questions

Questions 1a and 1b are concerned with the locations of vowel category ideals in relation to language groups' production range of the vowels, while Questions 2a and 2b are concerned with the effects of vowel prototypicality and extremity on the perceptual distance between vowels.

Question 1a: Do vowel category ideals match the production averages of the language group?

On the one hand, Kuhl (1991) obtained a result that the production average of English /i/ produced by male American speakers of English matched their best-rated /i/ stimulus (the category ideal), confirming the general assumption that the vowel category ideal, like category ideals in other domains, is situated at the heart of the geometrically represented category distribution, in this case the language group's production range of the category. On the other hand, others (Lively, 1993; Iverson & Kuhl, 1995; Lotto et al., 1996; Frieda, 1997; Lively & Pisoni, 1997) report that listeners gave higher ratings to /i/ stimuli that were more extreme than the production average, contradicting Kuhl (1991). Although the majority of recent findings seem to suggest that the vowel category ideal is more extreme than the language group's production average, its location needs to be re-examined, for these studies compare vowels produced in consonantal contexts with those chosen in goodness-rating tasks using steady-state isolated vowel stimuli. Therefore, the discrepancy between the locations of the category ideal and the production average of a vowel category observed in those studies could have been due to the incompatible materials used in the perception and production tasks. In addition, Johnson et al.'s (1993) study suggests that the listeners may interpret the term 'good (best)' as 'distinct', which departs from the original definition of the prototype, i.e., 'representative'.

Question 1b: If the category ideals and the production averages of vowel categories do not match, do the locations of category ideals match the most extreme realisations of the vowel categories?

If the location of the vowel category ideal is in fact more extreme than the speakers' production average, do the locations of vowel category ideals reflect the language-specificity observed in the phonetic realisations of the categories? On the one hand, it is conceivable for a category ideal to be more extreme than the production average and yet be within the speakers' production range of the category; in other words, the vowel prototype may be located at an extreme within the production range of the language group, as Frieda's study (1997) suggests. On the other hand, it is also conceivable that vowel category ideals do not have language-specific locations in the acoustic vowel space, as can be seen in Bradlow's (1993: 78-83) study, where American and Spanish listeners all gave high ratings to extreme vowels, demonstrating no apparent connection between the locations of the category ideals and cross-linguistic differences observed in the phonetic realisations of the categories [also see Hoffman (1973) for a similar finding with Japanese and American listeners' category ideals for /i/].

Question 2a: If the category ideals and the production averages of vowels do not match, does either of these correlate with poor discrimination sensitivity at the phonetic level of coding? If so, which one?

If the locations of category ideals and production averages of vowels do not match, which of these are more likely candidates for the reference point in speech perception? As discussed in Chap. 2, in the light of Mervis & Rosch's (1981) definition of prototypes as category members that have 'fewest attributes in common with related contrast categories', it is plausible that most vowel prototypes are not in the centre of the category distribution, i.e., where the production average lies, considering the distributional structure of the vowel space. It is also plausible, however, that vowel prototypes are in the centre of the category distribution, given Mervis & Rosch's other definition of prototypes that they 'have most attributes in common with other category members'. Assuming that prototypicality correlates with poor discrimination sensitivity, as Kuhl's Native Language Magnet Theory holds, comparisons between discrimination sensitivity around the category ideal and a vowel's production average should indicate which of the two is the prototype that assumes a special role in speech perception. However, the results of the studies on vowel prototypes to date are inconclusive in this respect also. For instance, Kuhl (1991) suggests that discrimination sensitivity may decline towards the production average, while Iverson & Kuhl (1995) suggest that sensitivity may decline towards the category ideal and not the production average. (Also see Sussman & Lauckner-Morano, 1995). Furthermore, the above studies all employ an experimental design known to elicit auditory coding (long stimuli and/or short inter-stimulus intervals), which is thought by some to differ from the phonetic level of coding, in which language-specific patterns of perception of sounds are observed. Thus, whether the reported assimilation effect, whether around the production average or category ideal, can be observed at the phonetic level of coding needs testing.

Question 2b: Does vowel extremity correlate with poor discrimination sensitivity at the level of phonetic coding?

In addition to Question 2a, the question of whether it is vowel prototypicality or extremity that correlates with poor discrimination sensitivity needs to be answered. As pointed out earlier, considering evidence suggesting that vowel extremity may correlate with poor discrimination sensitivity at the level of auditory coding (Macmillan, 1988; Schouten & van Hessen, 1992), it is not clear from studies to date whether it is vowel extremity or prototypicality that correlates with poor discrimination sensitivity, since discrimination sensitivity around the 'prototype' has been typically compared with discrimination sensitivity around a less extreme stimulus (see Kuhl, 1991; Sussman & Lauckner-Morano, 1995; Iverson & Kuhl, 1995). In order to answer the above question, discrimination around the two possible prototype locations needs to be compared with discrimination sensitivity around a more extreme, non-prototypical stimulus. Furthermore, whether the reported effect of vowel extremity on discrimination sensitivity manifests itself at the phonetic level of coding, which is the concern of the present study, is yet to be examined.

3.1.2 Implications for the Perceptual Learning of L2 Phonetics

As discussed earlier, both Best's Perceptual Assimilation Model and Flege's Speech Learning Model hold that the perceptual distance between L1 and L2 phones determines the degree of ease of perceptual learning of L2 phones. Furthermore, as I discussed in Chap. 1, both models seemingly share in common with the prototype theory of speech perception an assumption that mental representations of L1 categories serve as reference points in determining the perceptual distance between L1 and L2 phones.

However, as summarised above, as far as vowel categories are concerned, where such mental representations, or prototypes, are located in relation to the phonetic realisations of the categories is open to question. They may be the production averages, i.e., located at the centre of category distribution, at an extreme of the language group's production range, or outside the production range and may not reflect the language-specificity observed in the realisations of the categories they represent. If vowel prototypes do not have language-specific locations, it would be difficult to explain the cross-linguistic differences observed in the degrees of difficulty in acquiring L2 vowels simply in terms of the locations of prototypes.

Furthermore, although Kuhl's Native Language Magnet Theory holds that the L1 prototype pulls the perceptual space towards itself, making it difficult for the learner to distinguish L2 sounds from their L1 prototypes, it is not clear from studies to date whether it is vowel prototypicality or extremity that correlates with poor discrimination sensitivity. If vowel extremity correlates with poor discrimination sensitivity, it follows that the perceptual distance between two vowels becomes increasingly shorter towards the periphery of the vowel space. In the light of Flege's Speech Learning Model, this implies that L2 vowels that are more extreme than one's L1 vowels may be more difficult to learn perceptually in comparison to L2 vowels that are less extreme, given the same psychoacoustic distance, for the perceptual distance would be shorter in the former case than the latter. Furthermore, having an extreme L1 vowel may be a disadvantage for the learner, as it may be more difficult to discriminate L2 vowels in its vicinity than it is if the L1 vowel is not extreme.

3.2 Experimental Design and Choice of Languages

Two sets of experiments were designed in order to answer the four questions laid out in 3.1.1:

Question 1a:	Do vowel category ideals match the production averages of the language group?
Question 1b:	If vowel category ideals and production averages do not match, do the category ideals match the most extreme realisations of the categories?
Question 2a:	If vowel category ideals and production averages do not match, does either of these correlate with poor discrimination sensitivity at the level of phonetic coding? If so, which one?
Question 2b:	Does vowel extremity correlate with poor discrimination sensitivity at the level of phonetic coding?

To answer Questions 1a and 1b, the locations of vowel category ideals in relation to the production average and the production range of the language group were examined in Experiments 1a (a production experiment) and 1b (a perception experiment), using native speakers of Japanese and Modern Greek. The locations of the category ideals and production averages were also used in subsequent experiments (Experiments 2a and 2b) in order to answer Questions 2a and 2b presented above. The present experiments differed from the previous experiments reporting listeners' preference for extreme vowels (Bradlow, 1993; Johnson et al., 1993; Iverson & Kuhl, 1995; Lotto et al., 1996; Frieda, 1997; Lively & Pisoni, 1997) in that comparable materials were used for production and perception experiments. That is, vowels were produced in isolation to match the synthesised steady-state vowels used in the perception task. Furthermore, the listeners were instructed to choose vowel sounds that were 'closest' to their L1 vowels rather than the 'best' ones, for it was thought that listeners' preferences for extreme vowels found in previous experiments could have been partially due to the subjects' interpretation of the term 'best vowel' as 'distinct vowel'.

Japanese and Modern Greek, two languages that had not been previously tested for the perceptual magnet effect, were chosen, since they have phonologically comparable vowel systems in terms of vowel qualities (five vowels: /i, e, a, o, u/). Of the five vowels, production averages and category ideals of three vowels /i, a, u/ were obtained from the respective language groups. These vowels, unlike English lax vowels, can occur in isolation, and therefore can be regarded as vowels in a null context. Because the Japanese high back vowel /u/ is not as extreme as Greek /u/ and realised with substantially higher F2 values, cross-linguistic differences were expected to be observed between Japanese and Modern Greek in the F2 values of the production averages and category ideals of /u/.

Experiments 2a (an identification task) and 2b (a discrimination task) were designed to answer Questions 2a and 2b, in other words, to investigate whether listeners' discrimination sensitivity decreased towards the category ideals, the production averages, or the periphery of the acoustic vowel space at the level of phonetic coding. Within-category discrimination sensitivity curves for the vowel category /u/ were obtained from the same Japanese and Greek subjects, reusing the synthesised vowel stimuli corresponding to the high back region of the vowel space from those used in Experiment 1b (the perception task), in which the category ideals of three vowels /i, a, u/ were located. Experiment 2a was an identification task designed to find out which stimulus pairs used in the discrimination task (Experiment 2b) were both identified as /u/ by each subject. As the present study was concerned with whether there were differences in degrees of discriminability between members of a single vowel category (within-category discrimination), it was crucial that discrimination between stimuli identified as belonging to different categories was eliminated from the analyses. Experiment 2b consisted of a discrimination task that yielded the listeners' discrimination sensitivity curves in the high back region of the acoustic vowel space. In order to elicit phonetic coding in the discrimination task, vowel stimuli of a short duration (85 ms) and a long inter-stimulus interval (1 s) were employed. As the phonetic realisations of Japanese /u/ and Greek /u/ are substantially different along F2, cross-linguistic differences were expected to be observed in the discrimination curves of the two language groups in this region of the vowel space, if discrimination sensitivity decreased towards language-specific vowel prototypes, as Kuhl's Native Language Magnet Theory maintains. Furthermore, it was expected that the effects of vowel prototypicality and extremity would be distinguished in the Japanese listeners' discrimination sensitivity, whose /u/ (commonly transcribed as [ɯ]) is not extreme.

In the following, I briefly describe the Japanese and Modern Greek vowel systems (3.2.1), and elaborate on the design of the above four experiments (3.2.2).

3.2.1 Choice of Languages: Japanese and Modern Greek Vowel Systems

3.2.1.1 Japanese Vowel System

Japanese has five phonemically distinct short vowels,[35] i.e., high front, mid front, low, mid back, and high back vowels, which are phonemically transcribed as /i/, /e/, /a/, /o/ and /u/ respectively (Bloch, 1950:

---

[35] According to Vance (1987: 13), each of the five short vowel qualities also occurs as long vowels. There are no significant quality differences between Japanese long and short vowels as there are for German long and short vowels. (Also see Shibatani, 1990: 161-162). In Modern Greek, vowel length is not a distinctive feature (cf. Mackridge, 1985: 18).

115; Vance, 1987: 9-11; Tsujimura, 1996: 18-19). According to Vance's (1987: 10-11) summary of Sakuma (1973: 32-34) and Kawakami's (1977: 21-23) descriptions of Japanese vowels,[36] Japanese /i/ is close to cardinal [i]. Japanese /e/ has a quality between [e] and [ɛ]; and /o/ has a quality between [o] and [ɔ]. Japanese /a/ is described by Sakuma (1973: 33-34) as having a quality between [a] and [ɑ], while Kawakami (1977: 22) states that the range of individual variations for the vowel is wide and includes [a] and [ɑ]. Japanese /u/, which is commonly transcribed as [ɯ], is known to lack lip rounding. Vance (1987: 11) observes that lip compression[37] is observed in /u/ carefully pronounced by Tokyo speakers, but the compression is substantially weaker or absent in normal speech. Sakuma (1973: 34) claims that, in addition to the lack of lip rounding, the point of articulation of Japanese /u/ is relatively front. Acoustically, Japanese /u/ is realised with substantially higher F2 than cardinal [u]. (See Fig. 3:1 showing wide-band spectrograms of Japanese and Greek /u/s).

According to Shibatani (1990: 160-161, 187), dialectal variations in the articulation of the vowels are observed. For instance, although the dialects of the Western and the Eastern part of Japan, the two major Japanese dialects, both have vowel systems consisting of the five vowels listed above, /u/ is slightly rounded in the Western dialect, while it is not in the Eastern dialect. Furthermore, a confusion of /i/ and /e/ takes place in some dialects, e.g., the *shitamachi* ('downtown') variation of the Tokyo dialect. However, due to the standardisation of Japanese which started in the mid 19th century, some of these dialectal differences have disappeared in younger generations (cf. Imaishi et al., 1982).

---

[36] Vance (1987) mainly looks at the normative Standard Japanese *(hyoojungo)*, which is based on a variety of the Tokyo dialect [the Yamanote ('uptown') dialect] (Vance 1987: 1). (For a detailed description of *hyoojungo* and the Yamanote variety, see Nakamura & Kindaichi, 1955).

[37] Ladefoged (1971: 62-63) defines 'lip rounding' as forming 'a small lip aperture by bringing the corners of the mouth forward and protruding the lips', and 'lip compression' as 'closing the jaw and bringing the lips together vertically, so that the side portions are in contact, but there is a gap in the centre'. According to Kelly (1966), the distinction between vowels of rounded vs. compressed lips occurs in Unhobo, one of the West African languages. Ladefoged's 'lip rounding' vs. 'lip compression' corresponds to Heffner's (1950) 'horizontal lip rounding' vs. 'vertical lip rounding' and Sweet's (1890) 'inner rounding' vs. 'outer rounding'.

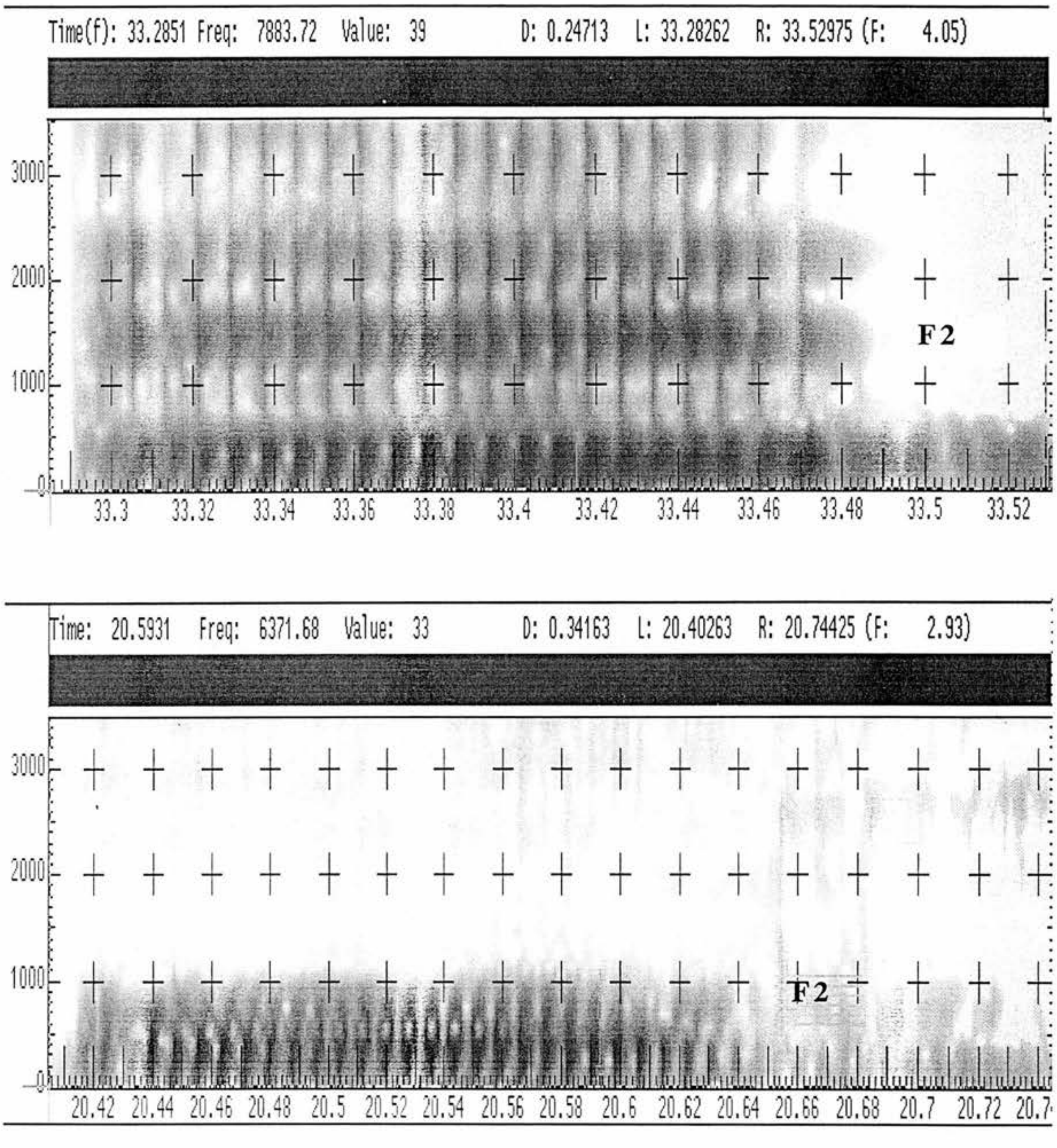


Figure 3:1. Wide-band spectrograms showing Japanese /u/ (the top panel) and Greek/u/ (the bottom panel). The spectrograms were produced from the data obtained in Experiment 1a, in which isolated /u/ was produced by Japanese and Greek male speakers.

3.2.1.2 Modern Greek Vowel System

Modern Greek has a vowel system consisting of five phonemically distinct vowels that are comparable to those of Japanese. Modern Greek also consists of high front, mid front, low, mid back and mid high vowels, each of which are transcribed as /i/, /e/, /a/, /o/ and /u/ (Mackridge, 1985: 18; Joseph & Philippaki-Warburton, 1987: 236). According to Jongman et al. (1989: 239), Greek point vowels are less extreme than those in German or American English, but well separated in the acoustic vowel space (Jongman et al., 1995: 239). Furthermore, unlike Japanese /u/, Greek /u/ is accompanied by lip rounding (Joseph & Philippaki-Warburton, 1987: 236) and therefore should be closer to cardinal [u] (cf. Jones, 1950: 18-22). A diagram in Bradlow (1993) comparing common vowels in Spanish, English and Modern Greek suggests that Greek /u/ has lower F2, i.e., farther back and/or more rounded, than Spanish and English /u/ (see Fig. 3:2).

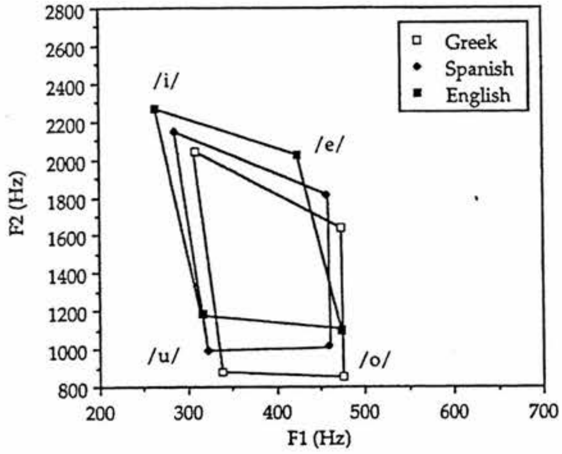


Figure 3:2. Greek, Spanish, and English /i, e, o, u/ in an F1-F2 space. The ordinate and abscissa each represents F2 and F1 Hz values of the four vowels. Unfilled and filled squares represent Greek and English vowels, respectively. Spanish vowels are represented by filled diamonds. Greek data was taken from Jongman et al. (1989) where four male native speakers of Modern Greek produced four repetitions of the five Greek vowels placed between a bilabial and an alveolar consonant. Spanish and English vowels, which were placed in a comparable consonantal context, average five repetitions produced by four male native speakers of General American English and Madrid Spanish. (The figure is adapted from Bradlow, 1993: 34).

According to Mackridge (1985: 5-6), the pronunciation of moderately educated people from all parts of Greece tends to be hardly distinguishable from that of an Athenian, and although there are regional differences between Thessalonikan and Athenian Greek, these are mainly morphological and not phonetic (cf. Joseph & Philippaki-Warburton, 1987).

3.2.2 Experimental Designs

3.2.2.1 Experiments 1a (a production task) and 1b (a perception task)

Experiments 1a and 1b were designed to examine the locations of Japanese and Greek subjects' category ideals for /i/, /u/ and /a/ in relation to the production range of these vowels. The locations of the category ideals in an F1-F2 space and the production averages (F1 and F2) of Japanese and Greek /u/ were also used in the subsequent experiments set up to answer the questions regarding the effects of vowel prototypicality and extremity on discrimination sensitivity. As noted earlier, isolated vowels were used both for the production and perception tasks in order to ensure that the two tasks were comparable. Furthermore, the use of the term 'good' in the perception task was avoided, as the listeners might interpret the term 'good' as meaning 'distinct', which departs from the original definition of prototypes, i.e., 'representative'.

In Experiment 1a, Japanese and Modern Greek /i, a, u/ were produced in isolation by the respective language groups in two different conditions: (i) a normal speech condition, in which the subjects were instructed to produce the above vowels in a natural way; and (ii) a hyperarticulated speech condition, in which the subjects were instructed to produce the vowels in an exaggerated way. The data from the normal speech condition were intended for obtaining the production averages of the above vowels of each language group. The data from the hyperarticulated speech condition were intended for obtaining the extreme end of the production range of the vowels produced by each language group. As mentioned earlier, it has been shown that in hyperarticulated speech vowels are realised with F1 and F2 values that are more extreme than in normal speech (cf. Picheny et al., 1986; Moon & Lindblom, 1989; Johnson et al., 1993). Furthermore, Frieda's (1997) study suggests that the extreme, hyperarticulated vowels, if not vowels produced in a normal speech condition, may match the category ideals.

In Experiment 1b, the same subjects located the category ideals for the above three vowels in a perception task using Johnson et al.'s (1993) method of adjustment (MOA) technique (for other studies using the MOA technique, see Scholes, 1967, 1968b; Nooteboom, 1973; Ganong & Zatorre, 1980; Samuel, 1982; Elsendoorn, 1984; Johnson, 1989). In the present experiment, subjects selected category ideals for the above three vowels from presynthesised vowel stimuli that were assigned to squares arranged in a grid-shape on the computer screen in such a way that F1 and F2 of the vowels systematically varied along each side of the grid. An MOA task was chosen rather than a goodness-rating task, since this method made it possible to present vowel stimuli covering the whole range of vowels that can be produced by adult male speakers within a reasonable time frame, and thus provided a possible way of getting around the problem of the stimulus range effect on the listener's selections of category ideals reported by Lively & Pisoni (1997). The stimuli were steady-state, comprising compatible materials with those obtained in Experiment 1a, in which isolated vowels were produced. Furthermore, the subjects were instructed to choose from these vowel stimuli those that were 'closest' to /i/, /a/ and /u/ of their respective L1, instead of 'best' sounds.

3.2.2.2 Experiments 2a (an identification task) and 2b (a discrimination task)

Experiments 2a (an identification task) and 2b (a discrimination task) were designed to study the change in Japanese and Greek speakers' within-category discrimination sensitivity for category /u/ at the level of phonetic coding. More specifically, the major purposes of Experiments 2a and 2b were (i) to investigate whether discrimination sensitivity decreased at either of the two possible prototype locations, i.e., the production average or the category ideal, and (ii) to distinguish between the effects of vowel prototypicality and extremity on discrimination sensitivity, when the discrimination task was designed to elicit phonetic coding of the stimuli.

Experiment 2a (the identification task) was conducted primarily to decide which portion of the stimulus continuum used in Experiment 2b was regarded as /u/ by each subject. As the present experiment was concerned with the change in within-category discrimination sensitivity, it was crucial not to include in the analyses portions of the discrimination sensitivity curves where stimuli were perceived as belonging to different vowel categories. In the identification task, stimuli were presented with adjacent stimuli in pairs, replicating D pairs (pairs consisting of different stimuli) prepared for the discrimination task (Experiment 2b), so that contrast effects on identification of vowels could be taken into account (for literature reporting contrast effects on identification of vowels, see Liberman et al., 1957; Repp et al., 1979). Lotto et al. (1996) argue that identification tasks presenting stimuli singly do not accurately indicate whether discrimination between a certain pair of sounds is within- or cross-category discrimination, since two sounds presented in a pair are more likely to be assigned to different phonemic categories than stimuli presented singly, in which case it can be no longer regarded within-category discrimination.

In Experiment 2b, Greek and Japanese subjects' discrimination sensitivities along F2 were tested in the high back region of the vowel space, where Japanese /u/ and Greek /u/ were located, reusing the stimuli in the corresponding area of the vowel space among those used in the MOA task (Experiment 1b). Both the category ideal (if it reflects cross-linguistic differences in phonetic realisations of the category) and the production average of Japanese /u/ were expected to have less extreme F2 than Greek /u/, as Japanese /u/ is realised without lip rounding. Therefore, it was expected that cross-linguistic differences would be observed in the change in Japanese and Greek discrimination sensitivities along F2 in this area of the vowel space, if discrimination sensitivity decreased towards language-specific vowel prototypes (i.e., either the category ideal or the production average). Specifically, the Japanese subjects would produce discrimination sensitivity curves that decline towards a point that is more towards the interior of the vowel space than the Greek subjects, if their discrimination sensitivity decreases either towards the language group's production average or a language-specific category ideal. On the other hand, if it is vowel extremity that correlates with low discrimination sensitivity, Japanese and Greek subjects' sensitivity curves within category /u/ would both decrease towards extreme vowels.

Experiment 2b differed from previous studies on the perceptual magnet effect (e.g., Kuhl 1991) in that it yielded a discrimination curve that stretched beyond the possible prototypes (the production average and the category ideal) towards the periphery of the vowel space for the Japanese subjects. Thus, the effect of vowel prototypicality and extremity on discrimination sensitivity was expected to be distinguished in the Japanese subjects' discrimination sensitivities. Furthermore, a short stimulus duration (85 ms) and a long inter-stimulus interval (1 s) were employed in order to facilitate phonetic coding of the stimuli. The listener's discrimination sensitivity was measured using miss rates and d' in order to examine whether Sussman & Lauckner-Morano's (1995) report that d' is a better measure of a perceptual magnet effect was applicable to the results of the present experiment.

Chapter 4 Relationships between Vowel Category Ideals (Perceptually Preferred Vowels) and Production

As discussed in Chap. 2, the two most influential current models of L2 phonetics (Best's Perceptual Assimilation Model and Flege's Speech Learning Model) apparently draw on the concept of phonetic prototypes, from which the perceptual distance between L1 and L2 sounds is calculated. However, it is not clear from studies to date where such prototypes are located for vowel categories. Experiments 1a (a production experiment) and 1b (a perception experiment) were conducted to examine the relationships between the locations of category ideals (perceptually preferred category members) in an F1-F2 vowel space and the production values of /i, a, u/ in Japanese and Modern Greek. In this chapter, Experiments 1a and 1b are described in more detail in 4.1 and 4.2, respectively, and the results are presented in 4.3.

4.1 Experiment 1a: Production

In Experiment 1a, Greek and Japanese subjects produced /i/, /a/ and /u/ in isolation in hyperarticulated and normal speech conditions, the results of which are compared with the locations of category ideals obtained in Experiment 1b.

Subjects

Ten Japanese and ten Greek phonetically naïve, adult male native speakers participated in the experiment as volunteers. They all reported normal hearing. The age of the subjects ranged from 23 to 42 years, with the average age of 29.6. They all resided in Edinburgh for their studies at the time of the experiment, and their lengths of stay in UK ranged from 3 months to 4 years and 4 months, with an average length of stay being 2 years. They all rated their English proficiency as either intermediate or advanced.

All the Japanese subjects spent most of their childhood in Japan and spoke Japanese as L1. Similarly, all the Greek speakers spent most of their childhood in Greece and spoke Modern Greek as L1. As for the Japanese subjects, except for J2 who lived in Germany from the age of 2 to 6, all the subjects lived in Japan at least up to the age of 20. According to his self-report, J2 does not speak German. As his performance in the present study was not noticeably different from the rest of the language group, he is included in the analyses. As for Greek subjects, all the subjects lived in Greece at least up to the age of 18.

Out of ten Japanese subjects nine of them were from the Eastern part of Japan and one subject (Subject J 10) was from the Western part in terms of Shibatani's (1990: 160-161, 187) classification of major dialectal regions. Out of ten Greek subjects, eight subjects reported to speak either Athenian or standard Greek, one of them spoke Chiotika, and another, Thessalonikan. Ideally, subjects' regional backgrounds should be controlled, but it was not possible due to the fact that the study took place in Edinburgh, Scotland. As for the Greek speakers, as mentioned in Chap.3, according to Mackridge (1985: 5-6), the pronunciation of moderately educated people from all parts of Greece tends to be hardly distinguishable from that of an Athenian. Therefore, the Greek speakers were considered as a homogeneous group. As for the Japanese speakers, two subjects (Subjects J3 and J6) reported their regional accent to be Northern, and one subject (Subject J10) reported it to be Western, while the rest spoke standard Japanese or the Tokyo dialect, of which the standard Japanese is a variety (cf. Chap. 3). The diversity in the subjects' regional background is taken into consideration when analysing the data where between-subject differences seem to have arisen from regional background. However, as Imaishi et al. (1982) report, some dialectal differences are disappearing in the younger generation due to the standardisation of pronunciation. Thus, it was expected that the speakers would not show such great variability in the production of the three vowels under investigation.

Many Greek subjects also spoke other foreign languages fluently in addition to English (self-report), which again could not be controlled in the present study. (For detailed information on each subject, see Appendix A). The results of the present experiment, however, did not show any systematic differences between those who reported to speak another language and those who did not.

Materials

Three vowels /i, a, u/ were transcribed in Japanese and Greek orthography. Each vowel was transcribed on 5 separate cards, yielding 15 cards altogether for each language group for each condition (3 vowels × 5 repetitions).

Procedure

The 15 cards were randomised and presented to the subjects, who produced the isolated vowels in hyperarticulated and normal speech conditions in two separate blocks. In the first block, the subjects were instructed to read the vowel on each card 'as clearly as possible' and 'as if they were talking to a person with hearing difficulties' (a hyperarticulated speech condition). These instructions were adopted from Moon & Lindblom (1989). When the subject seems to have understood the instructions as meaning 'speak loudly', it was clarified that the purpose of the experiment was to obtain 'clear' speech samples and not 'loud' ones. In a subsequent block, they were instructed to produce the vowels 'in a natural manner' (a normal speech condition). Five repetitions of each of the three vowels were produced by each subject in each condition, yielding 600 tokens in total for measurements (3 vowels × 5 repetitions × 2 conditions × 10 subjects × 2 groups = 600 tokens).

Recordings were conducted in a sound-treated studio at the Department of Linguistics, University of Edinburgh. The original recordings were made on a Sony PCM2700A DAT recorder using an AKGC567 Lapel microphone, which were then copied onto a Maxwell CD-R74H CD. The recordings were digitised at a sampling rate of 16000 Hz and low-pass filtered at 7800 Hz with a 16-bit amplitude resolution.

Analyses

F1 and F2 of the steady-state portion of each isolated vowel were measured using Entropic's X-Waves speech analysis software package on a SUN workstation. Both LPC (Linear Predictive Coding) spectra and wide-band spectrograms were used in measuring the formants (cf. Figs. 4:1 ~ 2). The LPC spectra were calculated using a covariance LPC analysis type (STRCOV1) with 18 LPC coefficients and a pre-emphasis factor of 100%, within a 25 ms rectangular window placed in the middle of the steady portion of the vowel which was visually determined on the spectrograms. The number of LPC coefficients was decreased to 17 or 16 when the software package picked up too many peaks. The F1 and F2 values generated by the STROCOV1 analyses were checked against readings from the spectrogram. When the values generated by the STROCOV1 analyses did not match the reading from the spectrogram, the power spectrum generated using a DFT (Discrete Fourier Transform) analysis was used to determine the approximate F1 and F2 values (cf. Fig. 4:3). All the vowel tokens were measured twice. Some (approximately 10%) of them were measured for the third time, as the first two measurements were discrepant. The two measurements were regarded discrepant when the difference was more than 15 Hz for F1 and 50 Hz for F2. When the first and second measurements were close, values from the second measurements were recorded, since the second measurement was felt to be more reliable.

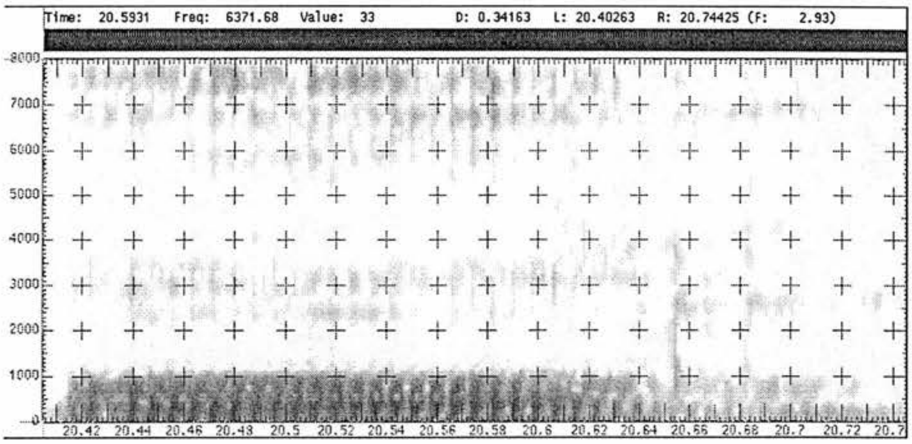


Figure 4:1. A wide-band spectrogram of an isolated vowel token produced in Experiment 1a.

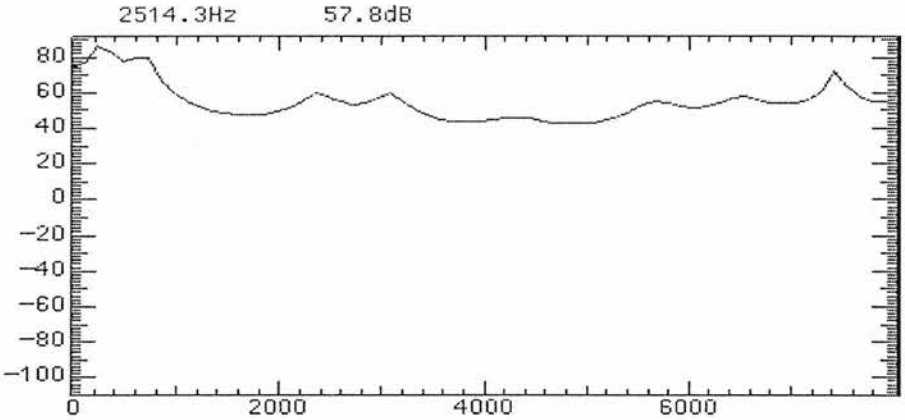


Figure 4:2. An LPC spectrum of the isolated vowel token in Figure4: 1.

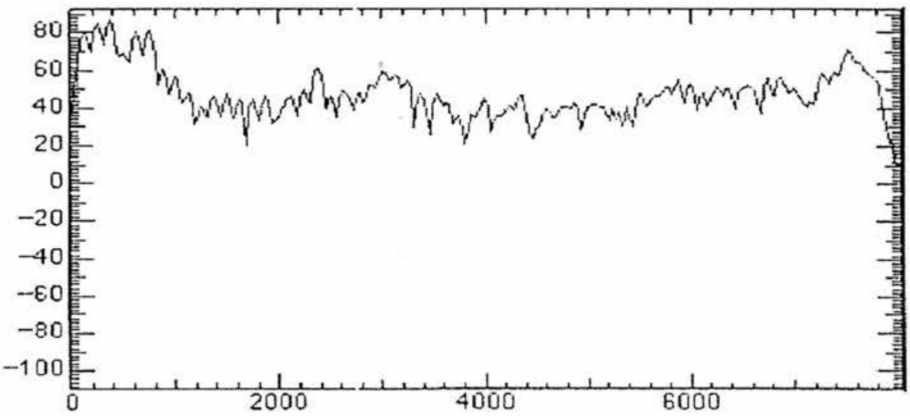


Figure 4:3. A DFT power spectrum of the isolated vowel token in Figure 4:1.

4.2 Experiment 1b: MOA

In Experiment 1b, Johnson et al.'s (1993) MOA (method of adjustment) procedure was adopted to locate Japanese and Greek category ideals for three vowels /i, a, u/, which were compared with the production data obtained in Experiment 1a. The same subjects chose vowel sounds judged to be the most representative of each of the three vowel categories in their respective L1 from 368 presynthesised vowel stimuli varying systematically in F1 and F2, which were presented as squares of a grid on a computer screen.

Subjects

The same subjects as in Experiment 1a participated in this experiment.

Materials

Three hundred sixty-eight steady-state vowels of a large range of F1 (215 ~ 957 Hz) and F2 (610 ~ 2511 Hz) were synthesised using Sensimetrics SenSyn cascade/parallel formant synthesiser developed by Klatt (cf. Klatt, 1980; Klatt, 1990). The stimuli were produced on a SUN workstation using the cascade portion of the synthesiser, a method normally used for laryngeal sound sources (cf. Klatt, 1990: Chap. 3).

All the stimuli consisted of 4 formants and were 85 ms long, with 55 ms of fixed amplitude that gradually attenuated in the last 30 ms that weakened the click at the end of the stimuli. The length of the stimuli was made short, so that they could be reused in a subsequent discrimination task (Experiment 2b) where the phonetic coding was the object of study.[38] The F0 of the stimuli was set at 120 Hz initially and fell linearly to 105 Hz, making the stimuli sound like those produced by a male speaker. (A synthesis specification file is given in Appendix B).

The ranges of F1 and F2 were determined to be from 215 Hz to 957 Hz and from 610 Hz to 2511 Hz, respectively, so that they would cover the production values of the three Japanese and Greek vowels obtained in Experiment 1a. Consequently, the F1 and

---

[38] According to David Pisoni (personal communication), the optimal length of vowel stimuli for eliciting phonetic coding is 50 ~ 100 ms. The author gratefully acknowledges the assistance of Dr. Pisoni in designing the experiment.

F2 ranges differed from those in Johnson et al. (1993).[39] The step-size for F1 was 43 mels and that for F2 was 45 mels in the present study, which again differed from Johnson et al.'s (1993) study that used the Bark scale. The mel scale was employed in order to make the following discrimination task (Experiment 2b) reusing the stimuli from the MOA task comparable to Kuhl (1991) whose Native Language Magnet Theory was tested. As mentioned in Chap. 2, the mel is a psychological unit for pitch, which corresponds to '... equal increments of the pitch as sensed by naive listeners' (Fant, 1973: 47), and has been widely used in calculating a perceptually equidistant step-size for formants. Following Kuhl (1991), Fant's formula was used to convert formant frequencies from Hertz to mels (Fant 1973: 48):

$$y = (1000/\log 2) \log (f/1000 + 1),$$

where f is the frequency in Hz and y is the mel value. The F1 and F2 values of the stimuli and the exact step-sizes between adjacent stimuli in mels are given in Table 4:1. Altogether, there were 17 possible F1 values and 26 possible F2 values, giving 442 possible F1-F2 combinations, of which 74 combinations were eliminated, where F1 would be either too close to, or have a higher value than, F2. The cut-off point was where the difference between F1 and F2 was smaller than 250 Hz. In total, 368 vowels were produced.

---

[39] In Johnson et al. (1993), F1 varied from 250 Hz to 900 Hz, and F2 varied from 800 Hz to 2800 Hz.

<F1>

Hz	Mels	Difference (mels)
215	280.96	
		43.3
252	324.23	
		43.1
290	367.37	
		43.0
329	410.34	
		42.8
369	453.12	
		42.6
410	495.70	
		43.3
453	539.03	
		43.0
497	582.07	
		42.7
542	624.80	
		43.3
589	668.12	
		42.9
637	711.05	
		42.6
686	753.60	
		43.0
737	796.60	
		43.4
790	839.96	
		42.9
844	882.84	
		43.2
900	926.00	
		42.6
957	968.64	

<F2>

Hz	Mels	Difference (mels)
610	687.06	
		45.0
661	732.05	
		45.3
714	777.37	
		44.8
768	822.12	
		45.0
824	867.11	
		45.2
882	912.27	
		44.5
941	956.80	
		45.4
1003	1002.16	
		44.7
1066	1046.84	
		45.4
1132	1092.21	
		44.6
1199	1136.85	
		45.2
1269	1182.06	
		45.1
1341	1227.12	
		44.9
1415	1272.02	
		44.7
1491	1316.73	
		45.0
1570	1361.77	
		45.3
1652	1407.08	
		45.0
1736	1452.07	
		45.2
1823	1497.23	
		44.8
1912	1542.01	
		44.9
2004	1586.88	
		44.9
2099	1631.80	
		45.4
2198	1677.17	
		44.9
2299	1722.03	
		44.8
2403	1766.81	
		45.1
2511	1811.88	

Table 4:1. F1 and F2 values and the step-sizes of MOA stimuli. F1 and F2 values are given in both Hz and mels. The step-sizes are given in mels.

Following Johnson et al. (1993), F3 was estimated using Nearey's (1989: 2095) regression formulas for front vowels ($F2 > 1500$ Hz) and back vowels ($F2 \leq 1500$ Hz). Nearey obtained those formulas by applying Broad and Wakita's (1977) coefficients to the American English data in Peterson & Barney (1952) and the Swedish data in Fant (1973):

$F3 \text{ (front vowels)} = .522F1 + 1.197F2 + 57$,
$F3 \text{ (back vowels)} = .7866F1 -.365F2 + 2341$.

F4 was set at least 300 Hz higher than F3 and not lower than 3500 Hz.

The bandwidths (the width of the formant resonance) of F1, F2 and F3 (B1, B2, and B3) were calculated using Johnson et al.'s (1993: 510) regression formulas given below. According to Johnson et al., the formulas provide a rough fit to the bandwidth values suggested by Klatt (1980), although extreme formant values result in unnatural bandwidths.

$B1 = 29.27 + .061 \times F1 - .027 \times F2 + .02 \times F3$,
$B2 = -120.22 - .116 \times F1 + .107 \times F3$,
$B3 = -432.1 + .053 \times F1 + .142 \times F2 + .151 \times F3$.

The bandwidth of F4 was kept at a typical value of 200 Hz.

As the perceived loudness of the stimuli varied depending of the formant values and the spacing between them, the overall amplitude of the stimuli was adjusted after the synthesis, so that they would be perceptually comparable in loudness. Stimuli whose F1 and F2 were close together needed to be up-scaled because of the principle of loudness summation within the critical band, and stimuli consisting of high-frequency formants needed to be down-scaled due to the intrinsic loudness of high-frequency formants (cf. Gulick, 1989: 246-282). The adjustment was done manually by listening to each stimulus and changing the overall gain scale factor for the amplitude of voicing. The overall gain scale factor for the amplitude of voicing for each stimulus is given in Table 4:2.

F1 (Hz)

F2 (Hz)

	215	252	290	329	369	410	453	497	542	589	637	686	737	790	844	900	957
610	60	58	57	55													
661	60	58	56	55	54	53											
714	60	57	56	55	54	53	51										
768	59	57	56	55	54	52	51	50									
824	59	57	56	55	53	52	51	50	49								
882	59	57	56	54	53	52	51	50	49	49							
941	59	57	55	54	53	52	51	50	49	49	48						
1003	59	56	55	54	53	52	51	50	49	49	48	48	47				
1066	58	56	55	54	53	51	51	50	49	49	48	48	47	47			
1132	58	56	55	54	52	51	51	50	49	49	48	48	47	47	46		
1199	58	56	55	53	52	51	51	50	49	49	48	48	47	47	46	46	
1269	58	56	54	53	52	51	51	50	49	49	48	48	47	47	46	46	45
1341	58	55	54	53	52	51	51	50	49	49	48	48	47	47	46	46	45
1415	57	55	54	53	52	51	51	50	49	49	48	48	47	47	46	46	45
1491	57	55	54	53	52	51	51	50	49	49	48	48	47	47	46	46	45
1570	57	55	54	53	52	51	51	50	49	49	48	48	47	47	46	46	45
1652	57	55	53	53	52	51	51	50	49	49	48	48	47	47	46	46	45
1736	57	54	53	53	52	51	50	50	49	49	48	48	47	47	46	46	45
1823	56	54	53	53	52	50	50	50	49	49	48	48	47	47	46	46	45
1912	56	54	53	53	51	50	50	50	49	49	48	48	47	47	46	46	45
2004	56	54	53	52	51	50	50	50	49	49	48	48	47	47	46	46	45
2099	56	54	52	52	51	50	50	50	49	49	48	48	47	47	46	46	45
2198	56	53	52	52	51	50	50	50	49	49	48	48	47	47	46	46	45
2299	55	53	52	52	51	50	50	50	49	49	48	48	47	47	46	46	45
2403	55	53	52	52	51	50	50	50	49	49	48	48	47	47	46	46	45
2511	55	53	52	52	51	50	50	50	49	49	48	48	47	47	46	46	45

Table 4:2. Overall gain scale factors of the MOA stimuli for the amplitude of voicing (in dB)

A 17 x 26 grid was created on a computer screen using a programme written by Norman Dryden, the Department of Linguistics, University of Edinburgh. To each square of the grid each of the 368 presynthesised vowel stimuli was assigned in such a way that F1 of the stimuli systematically varied along the vertical side of the grid while F2 varied along the horizontal side (cf. Fig. 4:4). The grid served as a subject-stimuli interface that allowed the subjects to listen to each stimulus by clicking on the square using the mouse. A corner of the grid was cut off, where F1 would be either too close to, or higher than, F2. The cut-off portion of the grid was visually displayed with an 'x' on the square, indicating that the square was disabled and no sound could be produced by clicking on the square.

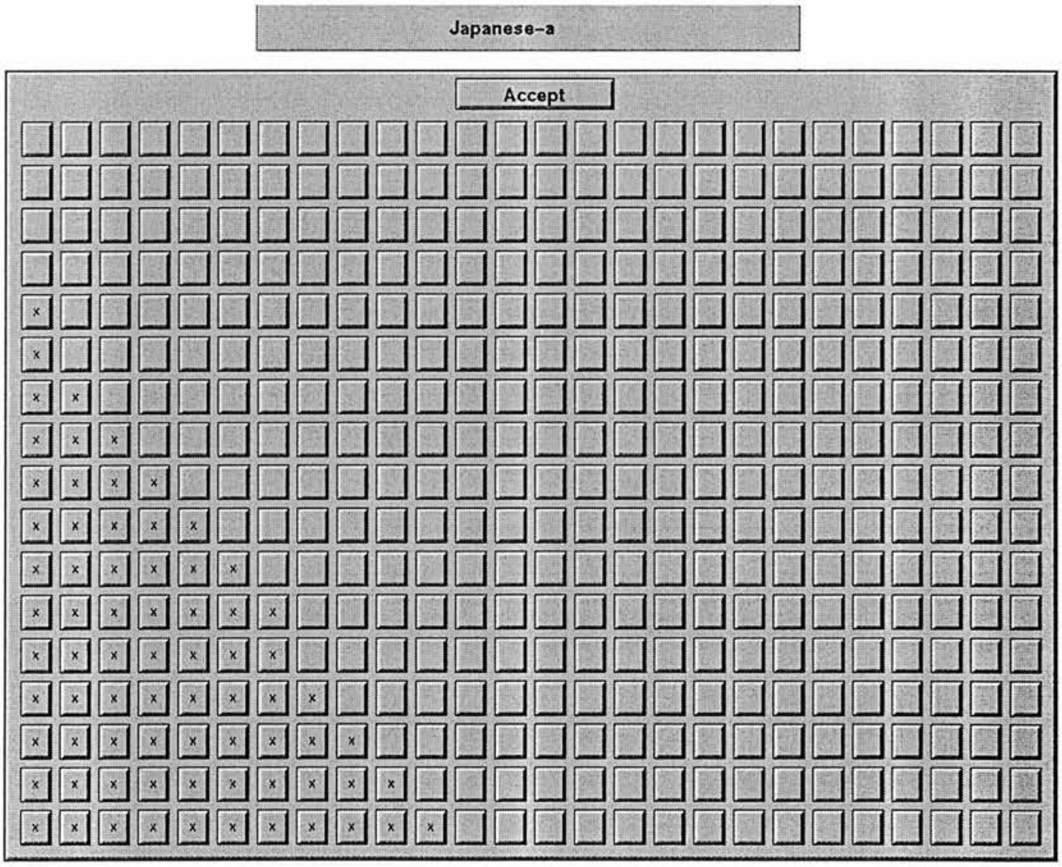


Figure 4:4. The MOA interface.

Procedure

The experiment was conducted in the experiment room at the Department of Linguistics, University of Edinburgh. The subjects were instructed to choose from the 368 vowel stimuli those judged to be 'the closest to /i, a, u/ in their respective L1 (Japanese and Modern Greek)' by moving the mouse freely on the above grid and listening to the stimuli, which were played binaurally through SONY MDR CD 550 headphones at a comfortable-listening level. The subjects were given oral instructions to find vowels that sounded 'closest' to their L1 vowels. When the subjects were not certain about what the experimenter meant by 'closest', the instructions were rephrased using the term 'most representative'.

The subjects chose each of the three vowels five times in five separate blocks in the order specified by the cue displayed above the grid that presented the three vowels in a randomised order, thus each yielding 15 responses (3 vowels × 5 repetitions). As neither the Japanese nor Greek script was available on the computer terminal, the vowels were specified in the Roman alphabet embedded in English phrases, such as 'Japanese-a' and 'Greek-a' (see Fig. 4:4 above). These phrases were read to the subjects when giving instructions, to ensure that they knew which vowels the phrases referred to.

In order to prevent the subjects from remembering the locations of the stimuli they chose previously, the orientation of the grid was changed randomly after every block. That is, at times F1 increased towards the bottom of the screen and at other times towards the top. Similarly, F2 sometimes increased towards the left and at other times towards the right. Furthermore, the subjects were told that the purpose of the experiment was not to test their consistency and therefore they needed not worry about what they chose in the previous block.

The task was self-paced. On average, each subject spent approximately 20 minutes to complete the task.

4.3 Results and Analyses

In the way described in 4.1 ~ 4.2, Experiment 1a (the production experiment) yielded F1 and F2 values of three Japanese and Greek vowels /i, a, u/ produced in two kinds of speech conditions, i.e., normal and hyperarticulated (exaggerated) speech conditions, while Experiment 1b (the MOA task) yielded F1 and F2 values of the category ideals of the above three vowels obtained from the two language groups.

In this section the results of the two experiments are discussed:

In 4.3.1, the production values of F1 and F2 of /i, a, u/ obtained from Japanese and Greek speakers in the normal speech condition are presented, and the cross-linguistic differences in the phonetic realisations of these vowel categories are studied;

In 4.3.2, F1 and F2 values of the three vowels obtained from the two language groups in the hyperarticulated speech condition are presented and compared with those obtained in the normal speech condition;

In 4.3.3, the relationships between F1 and F2 values of category ideals of the three vowels chosen by the Japanese and Greek speakers and the production data obtained from the respective language groups are examined;

Finally, in 4.3.4, the results are summarised and their implications for the nature of vowel category ideals are discussed.

4.3.1 A comparison between Japanese and Greek Vowels /i, a, u/: A normal speech condition

F1 and F2 values in Hz of the steady-state portions of /i, a, u/ produced by ten Japanese and Greek male speakers were measured and plotted in an F1-F2 vowel space (Fig. 4:5), where each ellipse covers two standard deviations from the category mean of each vowel category in each language (5 repetitions × 10 speakers). The mean formant values and standard deviations are given in Table 4:3 for each vowel category and language group. The mean F1 and F2 values obtained from individual speakers are given in Appendix C.

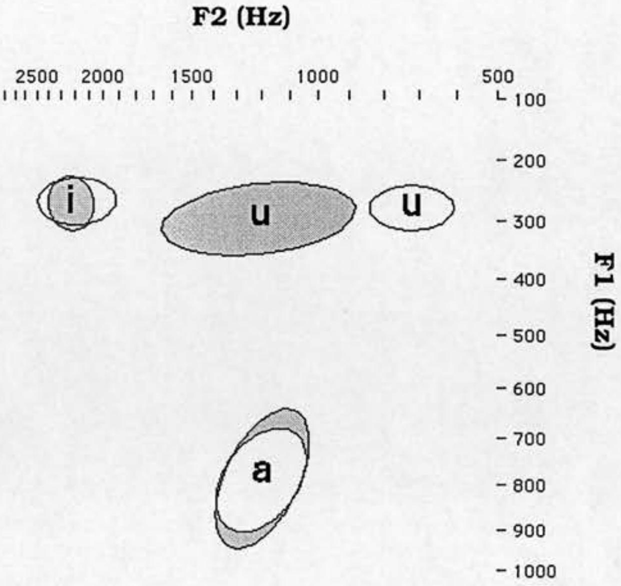


Figure 4:5. F1 and F2 values of Japanese and Greek /i, a, u/ produced by 10 male speakers in the normal speech condition. Grey ellipses represent Japanese vowels, and white ellipses represent Greek vowels. The x-axis and y-axis represent F2 and F1 values in Hz, respectively.

Vowel	Formant	Japanese formant value (Hz)	Greek formant value (Hz)
/i/	F1 (sd)	270.6 (22.7)	267.0 (19.6)
	F2 (sd)	2226.5 (87.7)	2186.2 (139.0)
/a/	F1 (sd)	787.6 (76.7)	790.0 (56.0)
	F2 (sd)	1203.1 (90.3)	1207.4 (88.4)
/u/	F1 (sd)	297.0 (31.0)	278.7 (18.8)
	F2 (sd)	1229.0 (187.3)	723.0 (57.0)

Table 4:3. Mean F1 and F2 Hz values of Japanese and Greek /i, a, u/ and standard deviations. (10 males in each group, 5 repetitions).

Overall, the results were in agreement with the reports in the literature introduced in Chap. 3. All of the three Greek vowels were cardinal, occupying the three corners of the acoustic vowel space (cf. Jongman et al., 1989: 239). Japanese /i/ and /a/ also occupied two corners of the vowel space, greatly overlapping with their Greek counterparts, whereas Japanese /u/ was realised with considerably higher F2 values in comparison to Greek /u/, which is consistent with the report in the literature (cf.

Sakuma, 1973: 34; Kawakami, 1977: 23; Vance 1987; 10-11). The only discrepancy between the results of the present study and the reports in the literature is that the variation in F2 of Japanese /a/ was not as great as Kawakami (1977: 22) describes.

In order to statistically compare the differences in the locations of the Japanese and Greek vowels in an F1-F2 space, two-tailed, unpaired t-tests were performed on the F1 and F2 Hz values of the 50 individual tokens of each vowel category in each language (5 repetitions × 10 speakers), with an exception of Greek /u/ which constituted 49 tokens.[40] The results are summarised in Table 4:4.

As expected, the difference between F2 of Japanese and Greek /u/ was statistically significant [t (58.2)[41] = 18.26, p < .0001]. The results also indicated that Japanese /u/ was realised with higher F1 than Greek /u/ [t (81.0) = 3.55, p = .001]. This implies that Japanese /u/ was more open than its Greek counterpart, which may be attributed to the absence of lip rounding in the production of Japanese /u/. Furthermore, the difference between F2 of Japanese and Greek /i/ was marginally significant [t (98) = 1.73, p = .086], with the mean F2 of Japanese /i/ being higher than that of Greek /i/. In the light of Liljencrants & Lindblom's (1972) Dispersion Theory which holds that a language's vowel system is structured in such a way that vowels are sufficiently apart from one another, this may be due to the fact that Japanese /u/ is realised with F2 that is higher than that of Greek /u/. As for /a/, and F1 of /i/, there were no statistically significant differences between Japanese and Modern Greek.

Vowel	Formant	Japanese mean (Hz)	Greek mean (Hz)	Mean difference	t-value (df)	p-value
/i/	F1	270.6	267.0	3.6	.84 (98)	n.s.
	F2	2226.5	2186.2	40.2	1.73 (98)	p = .086
/a/	F1	787.6	790.0	-2.3	-.17 (98)	n.s.
	F2	1203.1	1207.4	-4.4	-.24 (98)	n.s.
/u/	F1	297.0	278.7	18.3	3.55 (81)	p = .001
	F2	1229.0	723.0	506.0	18.26 (58)	p < .0001

Table 4:4. Results of two-tailed, unpaired t-test on F1 and F2 values of three Japanese and Greek vowels produced in a normal speech condition.

---

[40] One token of /u/ produced by a Greek speaker (Subject G2) had F1 and F2 too close to each other for a measurement and was thus excluded from the analysis.

[41] Degrees of freedom for the /u/ category derive from t-tests based on separate variance estimates, which were used instead of the equal variances line of values for the t-test, since the Levene Test for homogeneity of variance revealed that the homogeneity of variance assumption was violated for the /u/ category (cf. Kinnear & Gray, 1997: 138).

The differences observed between the formant values of the Japanese and Greek vowels cannot be only due to a difference in the over-all vocal tract length between the two language groups, which could have also contributed to the differences in the formant values (e.g., the Japanese speakers had shorter vocal tracts and produced vowels with higher formants), considering the fact that the differences were not observed for all the formants or vowels. According to Stevens & House (1963: 117-8), systematic differences should be obtained for the formant values of vowels produced by talkers whose vocal-tract lengths are different, although one cannot just use a simple scaling factor to convert formant frequencies of a particular vocal tract length to another.[42]

Another noticeable difference between Japanese and Greek vowels from Fig. 4:5 is the category size of each vowel in the two languages. In general, Greek vowels did not show large variations in their phonetic quality, as can be seen from the small size of the ellipses, which is consistent with Hawks and Fourakis' (1995: 248) observation. Japanese /i/, however, seemed to vary even less in F2 than Greek /i/ and cluster more tightly in the periphery of the vowel space. Again, the tight clustering of Japanese /i/ may be due to the high F2 values of Japanese /u/. On the other hand, Japanese /u/ seemed to be realised with a more variability in F2 than Greek /u/, with the standard deviation of the former being more than three times as much as the latter in Hz (a standard deviation of 187.3 Hz vs. 57.0 Hz; see Table 4:3 above). The greater variability in F2 values of Japanese /u/ may be due to large between-speaker differences, which could possibly be regional. As mentioned in Chap. 3, Shibatani (1990) observes that Japanese /u/ is slightly rounded in the Western dialect, while it is not in the Eastern dialect, which may have led to a wide range of F2 values of /u/ produced by the Japanese speakers. Alternatively, it may be the case that Greek /u/ is more 'quantal' than Japanese /u/. According to Stevens' quantal theory (1989: 3-4), '... there are certain ranges of the articulatory parameter within which the acoustic parameter is quite sensitive to changes in the articulation'. In the light of the quantal theory, it is conceivable that Greek /u/ was produced with an articulatory configuration within which the acoustic parameter was relatively unaffected by small changes in articulation, while the acoustic parameter of the configuration for Japanese /u/ was more susceptible to small changes.

A comparison of individual speakers' standard deviations in the two groups suggest that the greater variability in F2 values of Japanese /u/ is mainly due to

---

[42] An attempt was made to obtain the formant values of each speaker's production of schwa, from which his vocal-tract length could have been estimated, which, in turn, would have predicted the amount of shifts in formant values due to the difference in vocal-tract length. However, half of the speakers produced vowels like either /a/ or /e/, and therefore the data were not used.

between-speaker differences, rather than within-speaker differences. As can bee seen in Table 4:5 (next page), individual speakers' standard deviations in F2 of /u/ do not seem to differ significantly between the two groups, with Japanese group's mean standard deviation being 38.5 Hz and Greek group's being 32.5 Hz. Thus, the large standard deviation in F2 of Japanese /u/ may be due to greater differences in the locations of F2 targets for Japanese /u/ than Greek /u/. Indeed, an examination of the results of the perception task (Experiment 1b) suggested that this might be the case, which is shown later in 4.3.3.

In order to examine whether the greater variability in the F2 values of Japanese /u/ can be attributed to the various regional backgrounds of the Japanese subjects, the subjects' regional backgrounds and their F2 values of /u/ were compared. Subject J10, the only Western Japanese dialect speaker in Shibatani's (1990) definition, in fact produced /u/ with the lowest F2 (858 Hz) in the group (the average of the remaining subjects is 1270 Hz). The low F2 of /u/ produced by Subject J10 may have been resulted from the slight lip rounding reported in the production of /u/ by the speakers of Western Japanese dialects (cf. Shibatani, 1990). However, even when Subject J10 is excluded from the analysis, the standard deviation of Japanese-/u/ F2 values amounts to 147 Hz, which is more than twice as big as the Greek standard deviation. This suggests that F2 of the Japanese /u/ may be variable, and that Japanese listeners may tolerate a relatively wide range of F2 values for the /u/ category.

Japanese speaker	sd in F2 (Hz)	Greek speaker	sd in F2 (Hz)
J1	32.1425	G1	49.6976
J2	28.8984	G2	23.8529
J3	21.1136	G3	10.8586
J4	27.4166	G4	42.956
J5	58.8567	G5	25.7383
J6	52.4986	G6	34.1119
J7	28.6496	G7	25.2496
J8	38.3002	G8	36.4329
J9	66.1584	G9	59.3584
J10	31.2833	G10	16.4237
mean	38.53179	mean	32.46799

Table 4:5. Japanese and Greek speakers' standard deviations in F2 of /u/ (a normal speech condition).

4.3.2 A Comparison between Normal and Hyperarticulated Vowels

In Experiment 1a isolated vowels were produced in a hyperarticulated speech condition as well as in a normal speech condition, with the purpose of obtaining the most extreme F1 and F2 values of Japanese and Greek /i, a, u/. This was to examine whether the locations of the category ideals of the three Japanese and Greek vowels match their most extreme realisations, if category ideals were more extreme than the production averages. Here, extremity is defined along both F1 and F2 for /i/ and /u/: extremity is defined as having low F1 and high F2 for /i/, and as having low F1 and F2 for /u/. Extreme /a/ is defined in terms of F1 alone: the higher the F1, the more extreme /a/ is.

F1 and F2 values (Hz) of the three vowel categories obtained in normal and hyperarticulated speech conditions are plotted for each language group in F1-F2 vowel spaces (Figs. 4:6, 4:7), where ellipses cover two standard deviations from the category means of F1 and F2 of each vowel category. The mean formant values and standard deviations obtained in the two conditions are compared for each language group in Tables 4:6 ~4:7, which are presented just below Figs. 4:6 ~ 4:7 where these values are plotted.

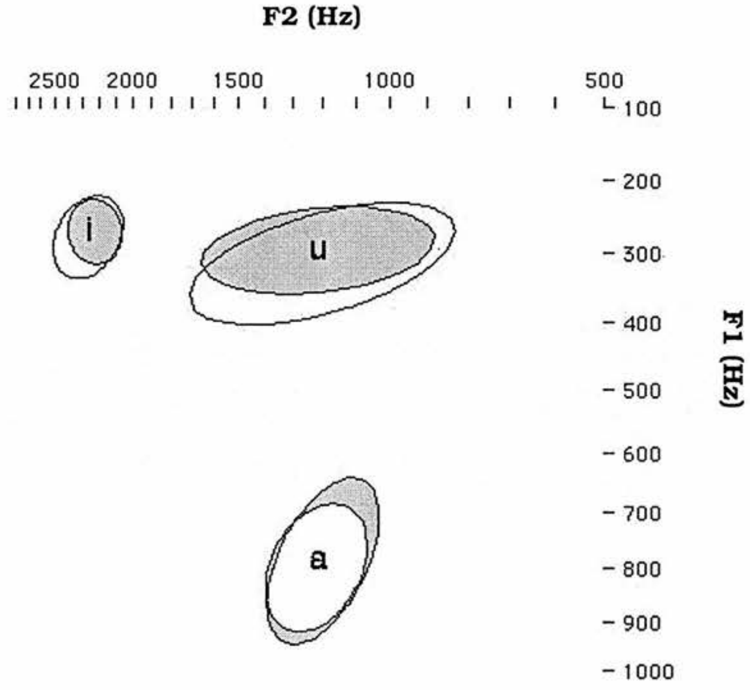


Figure 4:6. /i, a, u/ produced by 10 Japanese speakers in normal (grey ellipses) and hyperarticulated (white ellipses) conditions.

Vowel	Formant	Formant value (Hz)		Mean difference (Hz)
		Normal speech condition	Hyperarticulated speech condition	
/i/	F1 (sd)	270.6 (22.7)	278.3 (29.4)	-7.7
	F2 (sd)	2226.5 (87.7)	2271.7 (119.4)	-45.2
/a/	F1 (sd)	787.6 (76.7)	799.2 (58.7)	-11.6
	F2 (sd)	1203.1 (90.3)	1225.0 (83.3)	-21.9
/u/	F1 (sd)	297.0 (31.1)	315.7 (44.0)	-18.7
	F2 (sd)	1229.0 (187.3)	1219.6 (211.9)	9.4

Table 4:6. A comparison between mean F1 and F2 values of Japanese vowels /i, a, u/ produced in normal and hyperarticulated speech conditions.

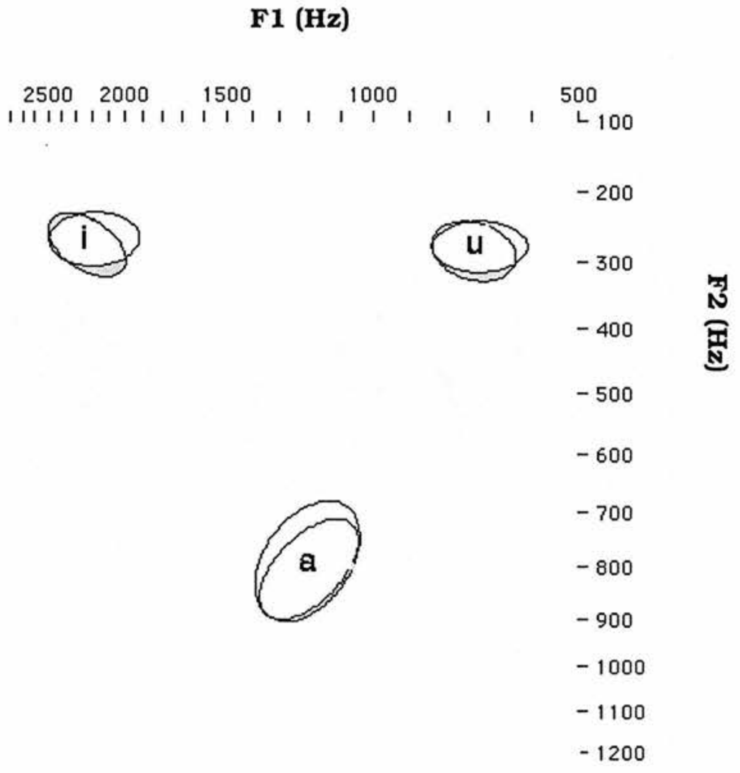


Figure 4:7. /i, a, u/ produced by 10 Greek speakers in normal (grey ellipses) and hyperarticulated (white ellipses) conditions.

Vowel	Formant	Formant value (Hz)		Mean difference (Hz)
		Normal speech condition	Hyperarticulated speech condition	
/i/	F1 (sd)	267.0 (19.6)	275.4 (23.2)	-8.4
	F2 (sd)	2186.2 (139.0)	2232.8 (125.9)	-46.6
/a/	F1 (sd)	790.0 (56.0)	804.5 (48.0)	-14.6
	F2 (sd)	1207.4 (88.4)	1202.7 (83.3)	4.7
/u/	F1 (sd)	278.7 (18.8)	285.2 (22.0)	-6.4
	F2 (sd)	723.0 (57.1)	736.9 (51.4)	-13.9

Table 4:7. A comparison between mean F1 and F2 values of Greek vowels /i, a, u/ produced in normal and hyperarticulated speech conditions.

As can be seen in Figs. 4:6 ~ 4:7, isolated vowels produced in hyperarticulated and normal speech conditions did not differ significantly for either language group. The small difference may be interpreted as indicating that those vowels were hyperarticulated to a certain degree in the normal speech condition, considering that the recordings were not made of natural speech. Alternatively, given that the vowels were produced in isolation in the present experiment whereas previous studies reporting the expansion of hyperarticulated vowel space used vowels in consonantal contexts (Picheny et al., 1986: Moon & Lindblom, 1989; Johnson et al., 1993), the more extreme vowels in the hyperarticulated speech condition obtained in those studies might be attributable to smaller degrees of 'undershoot' due to coarticulation (influence of articulatory configurations of preceding and following segments) as Moon et al. (1989) maintain. For example, Johnson et al. (1993: 519-520) used Moon & Lindblom's (1989) 'null' context, i.e., a /hVd/ frame, for both normal and hyperarticulated speech, and observed expanded hyperarticulated vowel space, which is mainly along F2 and towards higher F1. This is possibly due to /d/ in the /hVd/ frame which may lower F1 of the preceding vowel and make F2 of both front and back vowels less extreme. As the F2 locus of /d/ is around 1800 Hz, the F2 of the preceding /i/ might be lowered and that of /u/ increased, for instance, in comparison to their isolated versions, when the vowel length is not longer than the formant transition which takes around 45 ms.[43]

Furthermore, although the difference is small, there seems a slight tendency for hyperarticulated vowels in both languages to be produced with higher F1 and F2 than those produced in the normal speech condition, which contradicts the prediction that hyperarticulated vowels would have more extreme F1 and F2 values. If hyperarticulated vowels were produced with more extreme F1 and F2 values, /i/ would have lower F1 and higher F2 (less open and further front), /a/ would have higher F1 (more open), and /u/ would have lower F1 and F2 (less open and further back), compared to the normal speech condition. In order to see whether the observed tendency is statistically significant, simple factorial analyses of variance were performed on the F1 and F2 values of the 200 individual vowel tokens[44] produced in the two speech conditions (5 repetitions × 10 speakers × 2 language groups × 2 conditions) separately for each vowel category, as a function of condition (normal and hyperarticulated), language (Japanese and Modern Greek), and subjects nested under language as a random factor, from which the simple effects of condition and the condition-language interaction were computed using

---

[43] Although this can be easily tested, the test was not conducted, as it was beyond the scope of the present study.

[44] The number of tokens for the category /u/ was 199, as one token from the normal speech condition was excluded, for its F1 and F2 were too close together for a measurement.

formulae given in Winer (1971: 464-8). The results summarised in Table 4:8 indicate that the simple effect of condition was significant only for F1 of /u/ [F (1,18) = 6.11, p < .025] and F2 of /i/ [F (1, 18) = 12.49, p < .01], with higher values for F1 of hyperarticulated /u/ and F2 of hyperarticulated /i/ compared to those in the normal speech condition. The condition-language interaction was not significant with any of the vowels or formants, suggesting that the two language groups were not behaving in different ways from one another across the two speech conditions.

Vowel	Formant	Simple effect of condition		Condition x Language interaction	
		F-ratio	p-value	F-ratio	p-value
/i/	F1	F (1, 18) = 2.67	n.s.	F (9, 9) < 1	n.s.
	F2	F (1, 18) = 12.49	p < .01	F (9, 9) < 1	n.s.
/a/	F1	F (1, 18) = 2.22	n.s.	F (9, 9) < 1	n.s.
	F2	F (1, 18) = 2.34	n.s.	F (9, 9) = 1.56	n.s.
/u/	F1	F (1, 18) = 6.11	p < .025	F (9, 9) = 1.40	n.s.
	F2	F (1, 18) < 1	n.s.	F (9, 9) < 1	n.s.

Table 4:8. Results of ANOVA on F1 and F2 Hz values of Japanese and Greek /i, a, u/ produced in normal and hyperarticulated speech conditions.

As briefly mentioned earlier, a popular view of hyperarticulated vowels is that they are more extreme than those produced in normal speech (Picheny et al., 1986: Moon & Lindblom, 1989; Johnson et al., 1993). However, when the data in these studies are closely looked at, the above view seems a rather simplified picture of hyperarticulated speech. For instance, Picheny et al.'s (1986: 442) data comparing vowels in various consonantal contexts produced in the two conditions suggest that in a hyperarticulated speech condition F2 of front vowels is reliably higher (more extreme) for all the speakers, whereas F2 of back vowels does not seem to differ as much between the two conditions. Furthermore, F1 of hyperarticulated high vowels is lower (less open, more extreme) for some speakers but higher (more open, less extreme) for others, suggesting no general directions for the difference between the two conditions. Frieda, et al. (in press) also reports that F2 of hyperarticulated /i/ produced by American subjects was consistently higher than that of normally produced /i/, but that F1 differed significantly between the two conditions only for some of the speakers, who produced higher F1 (less extreme, more open) in the hyperarticulated speech condition. This suggests that there may be two competing goals in hyperarticulated speech, namely, a goal to open the mouth more widely which results in higher F1, and a goal to hit the production target, i.e., the values achieved by isolated vowels.

To summarise, F1 and F2 of hyperarticulated vowels produced in the present experiment did not differ significantly from those of the vowels produced in the normal speech condition except for F1 of /u/ and F2 of /i/. Although this may be partially due to a certain degree of hyperarticulation in the 'normal' condition, it can be also inferred that the more extreme F1 and F2 values observed in hyperarticulated vowels in previous studies may have mainly derived from reduced degrees of coarticulation, which isolated vowels are not subject to (cf. Moon & Lindblom, 1989; Johnson et al., 1993). Considering possible 'undershoot' effects on a vowel preceding /d/, expansions of hyperarticulated vowel spaces reported in studies using vowels embedded in /hVd/ frames may have been due to a reduced degree of coarticulation and 'undershoot' of the target values, although conventionally /hVd/ frames have been used as 'null' contexts. Furthermore, although the tendency was not statistically significant except for F1 of /u/, hyperarticulated vowels tended to be realised with higher F1 that indicates a more open mouth, suggesting two goals of hyperarticulated speech: achieving the target values and facilitating the transmission of speech.

Recall that the hyperarticulated speech condition was included in Experiment 1a so as to obtain extreme F1 and F2 values and to compare them with category ideals obtained using the MOA technique in Experiment 1b, if the MOA choices are more extreme than the production averages of vowels produced in the normal speech condition. Since vowels in the hyperarticulated speech condition did not differ significantly from those produced in the normal speech condition except for the F1 of /u/ and F2 of /i/, formant values produced in both conditions are collapsed when compared with the MOA choices. As for the F1 of /u/ and F2 of /i/ which were significantly different between the two speech conditions, the formant values obtained in the two conditions are considered separately where appropriate.

4.3.3 A Comparison between MOA Choices and Production Data

In this section, I present the results of the MOA task (Experiment 1b) in which the same Japanese and Greek subjects chose from synthesised vowel stimuli those judged to be most representative of /i, a, u/ in their respective L1s, which I have called category ideals to distinguish them from prototypes that draw on production averages. The locations of category ideals are discussed in relation to the production data obtained in Experiment 1a. This section is further divided into 3 parts:

In 4.3.3.1, the MOA choices and the production values of Japanese and Greek /i, a, u/ are collectively compared;

In 4.3.3.2, within- and between- subject variability in MOA choices are discussed;

In 4.3.3.3, the relationships between individual subjects' production values and their MOA choices are examined.

4.3.3.1 An Overall Comparison between MOA Choices and Production Data

In this section, I examine the relationships between the F1 and F2 values of Japanese and Greek /i, a, u/ obtained from the production tasks (Experiment 1a) and those of the category ideals chosen by Japanese and Greek subjects in the MOA task as most representative of the three vowels in their respective L1s (Experiment 1b). As explained earlier, a major purpose of Experiments 1a and 1b was to re-examine the relationships between vowel category ideals and the production averages using comparable materials for production and perception tasks and without using the term 'good/best' in eliciting the locations of category ideals, thereby eliminating possible factors that might have led to the listeners' bias towards extreme vowels in previous studies. As noted earlier, extremity is defined as having low F1 and high F2 for /i/, and as having low F1 and F2 for /u/. Extreme /a/ is defined as having high F1 for /a/. Three kinds of relationships between the production values and MOA choices are conceivable: (i) the locations of MOA choices (category ideals) match the production averages of the vowels obtained from each language group; (ii) MOA choices lie at the extreme end of the language group's production ranges; (iii) MOA choices are more extreme than the language group's production ranges.

The results of the MOA task and the production range of each of the vowel categories /i, a, u/ are summarised in Tables 4:9 ~ 4:10 for each language group. As noted earlier, the production data presented here combine those produced in normal and hyperarticulated speech conditions except for F2 of /i/ and F1 of /u/, which differed significantly between the two conditions. For F2 of /i/ and F1 of /u/, values obtained in the normal speech condition alone are used.

Vowel	Formant	MOA means (Hz)	Production means (Hz)
/i/	F1 (sd)	269.4 (49.6)	274.5 (26.4)
	F2 (sd)	2382.3 (169.2)	2249.1 (106.7)
/a/	F1 (sd)	806.4 (75.5)	793.4 (68.2)
	F2 (sd)	1190.5 (96.9)	1214.0 (87.1)
/u/	F1 (sd)	294.2 (34.1)	306.3 (39.0)
	F2 (sd)	1048.0 (231.6)	1224.3 (199.0)

Table 4:9. Average F1 and F2 of Japanese /i, a, u/ in Hz values obtained from the MOA and production tasks. Standard deviations are given in parentheses.

Vowel	Formant	MOA means (Hz)	Production means (Hz)
/i/	F1 (sd)	251.6 (32.6)	271.2 (21.8)
	F2 (sd)	2340.5 (125.3)	2209.5 (134.0)
/a/	F1 (sd)	860.9 (73.6)	797.2 (52.4)
	F2 (sd)	1438.9 (407.6)	1205.1 (85.5)
/u/	F1 (sd)	262.2 (35.7)	282.0 (20.7)
	F2 (sd)	732.5 (98.1)	730.0 (54.5)

Table 4:10. Average F1 and F2 of Greek /i, a, u/ in Hz values obtained from the MOA and production tasks. Standard deviations are given in parentheses.

As can be seen in Figs. 4:9 ~ 4:10, where the number of times each MOA stimulus was chosen is tallied, the MOA choices generally spread beyond the production range of the corresponding vowel category, but they are concentrated on the most extreme end of the production range.

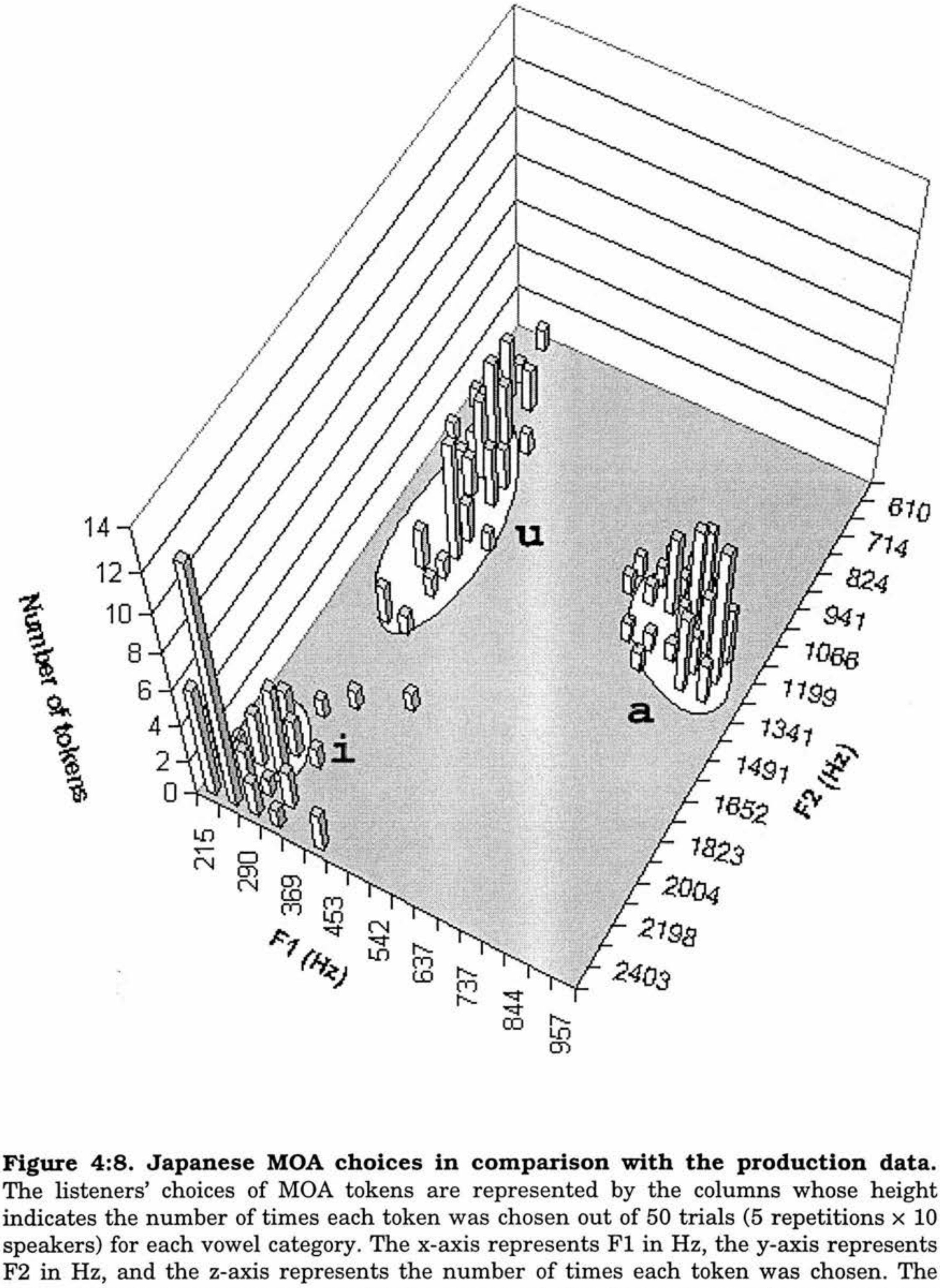


Figure 4:8. Japanese MOA choices in comparison with the production data. The listeners' choices of MOA tokens are represented by the columns whose height indicates the number of times each token was chosen out of 50 trials (5 repetitions × 10 speakers) for each vowel category. The x-axis represents F1 in Hz, the y-axis represents F2 in Hz, and the z-axis represents the number of times each token was chosen. The white ellipses represent the production range of the Japanese vowels covering two standard deviations from the production means.

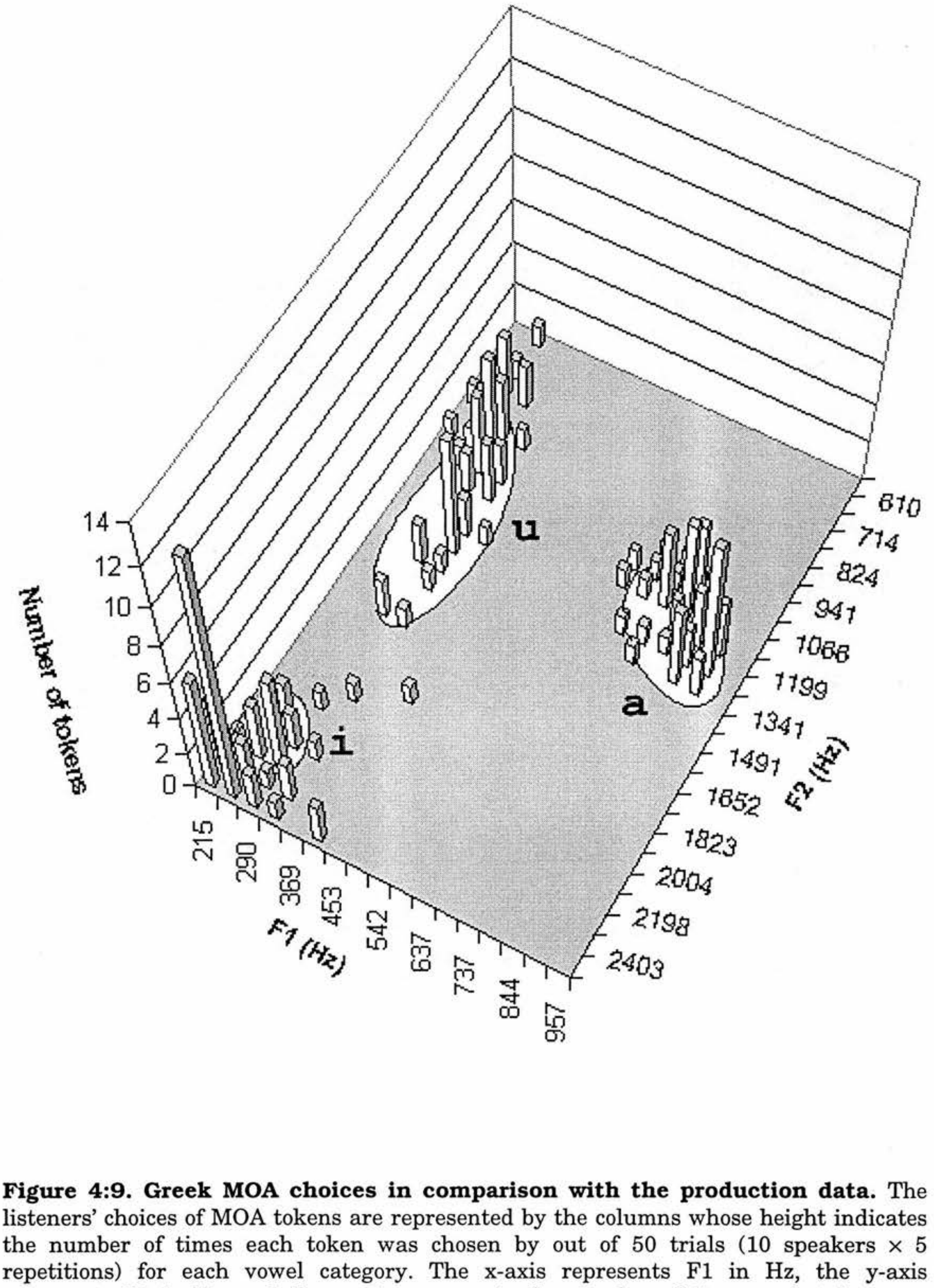


Figure 4:9. Greek MOA choices in comparison with the production data. The listeners' choices of MOA tokens are represented by the columns whose height indicates the number of times each token was chosen by out of 50 trials (10 speakers × 5 repetitions) for each vowel category. The x-axis represents F1 in Hz, the y-axis represents F2 in Hz, and the z-axis represents the number of times each token was chosen. The white ellipses represent the production range of the Greek vowels covering two standard deviations from the production means.

In order to verify the observation that the listeners preferred stimuli that were relatively extreme in comparison to the production values, two tailed, unpaired t-tests[45] were performed separately for each vowel category and language group on the F1 and F2 values of the individual 50 tokens from the MOA task and 100 tokens,[46] from the collapsed production data (normal and hyperarticulated speech conditions). As mentioned earlier, since F1 of hyperarticulated /u/ and F2 of hyperarticulated /i/ differed significantly from those produced in the normal speech condition, they were not collapsed with those obtained in the normal speech condition. The results of the t-tests are given in Tables 4:11 ~ 4:12.

Vowel	Formant	Mean formant value (Hz)		Mean difference (Hz)	t-value (df)	p-value
/i/	F1	(MOA)	269.4	-5.1	-.68 (63.3)[47]	n.s.
		(production)	274.5			
	F2	(MOA)	2382.3	133.1	5.08 (69.1)	p < .0001
		(production)	2249.1			
/a/	F1	(MOA)	806.4	13.0	1.07 (148)	n.s.
		(production)	793.4			
	F2	(MOA)	1190.5	-23.6	-1.50 (148)	n.s.
		(production)	1214.0			
/u/	F1	(MOA)	294.2	-12.1	-1.87 (148)	p = .064
		(production)	306.3			
	F2	(MOA)	1048.0	-176.3	-4.84 (148)	p = .0001
		(production)	1224.3			

Table 4:11. A comparison between production averages of F1 and F2 of Japanese /i, a, u/ and MOA choices for the three vowel categories.

---

[45] Unpaired t-tests were performed in order to collectively compare each language group's MOA choices with the production values. Individual subjects' MOA choices and production values are compared in 4.3.3.3.

[46] Again, Greek /u/ tokens in the production data amounted only to 99 due to one missing value.

[47] Where degrees of freedom are not (number of tokens - 1), they derive from t-tests based on separate variance estimates, which were used instead of the equal variances line of values for the t-test (cf. Kinnear & Gray, 1997: 138).

Vowel	Formant	Mean formant value (Hz)		Mean difference (Hz)	t-value (df)	p-value
/i/	F1	(MOA)	251.6	-19.6	-3.84 (71.5)	p < .0001
		(production)	271.2			
	F2	(MOA)	2340.5	131.0	5.76 (148)	p < .0001
		(production)	2209.5			
/a/	F1	(MOA)	860.9	63.7	5.47 (74.6)	p < .0001
		(production)	797.2			
	F2	(MOA)	1438.9	233.8	4.01 (51.2)	p < .0001
		(production)	1205.1			
/u/	F1	(MOA)	262.2	-19.7	-3.62 (66.1)	p = .001
		(production)	282.0			
	F2	(MOA)	732.5	2.5	.17 (64.7)	n.s.
		(production)	730.0			

Table 4:12. A comparison between production averages of F1 and F2 of Greek /i, a, u/ and MOA choices for the three vowel categories. F1 and F2 values are in Hz.

As can be seen in Table 4:11, the results of the t-tests indicated that Japanese MOA choices for /i/ were significantly more extreme than the production average along F2 [t (69.1) = 5.08, p < .0001], but not along F1. Japanese MOA choices for /u/ were significantly more extreme than the production average along F2 [t (148) = -4.84, p < .0001], and F1, to a lesser extent [t (148) = -1.87, p = .064], with the p-value being just short of the significant level of .05. As for Japanese /a/, MOA choices were not significantly more extreme than the production values, although the mean difference was in the direction of MOA choices being more extreme than the production average. As can be seen in Fig. 4:8, although the Japanese listeners seemingly preferred /a/ with relatively extreme F1 values, since the production values of Japanese /a/ themselves stretch almost up to the most extreme F1 value, the production values and MOA choices did not differ significantly. In addition, the average F2 of the Japanese subjects' choice of /a/ is 1191 Hz, where the maximum F1 is smaller (900 Hz) than what is available for stimuli whose F2 is higher (957 Hz) due to the cut-off of the corner of the MOA grid. Thus, it is possible that the Japanese listeners' choices of F1 for /a/ was restricted because of their choices of F2 for this vowel. As shown below, F1 of Greek MOA choices for /a/ were significantly more extreme than the production values. Assuming that both

Japanese and Greek subjects preferred /a/ stimuli with extreme F1, the fact that only Greek MOA choices were significantly different from the production values may be partially explained by the fact that the Greek subjects chose a range of F2 where a more extreme F1 value (957 Hz) was available.

As for the Greek vowels, as can be seen Table 4:12, MOA choices were significantly more extreme than the production values except for F2 of the /u/ category. As for F1 and F2 of /i/ and F1 of /a/, t (71.5) = -3.84, t (148) = 5.76, and t (74.6) = 5.47 in this order, for all of which p < .0001; as for F1 of /u/, t (66.1) = -3.62, p = .001. As in the case of Japanese /a/, the production values of F2 of Greek /u/ were in the extreme corner of the vowel space, and therefore MOA choices did not differ from the production values, although the Greek listeners did prefer tokens with most extreme F2 values, as apparent in Fig. 4:9.

Thus, the results of t-tests suggest that both Japanese and Greek subjects perceptually prefer stimuli that were more extreme than the production average except for those vowels whose production values were most extreme. Furthermore, listeners' preference for extremity seems to be found both along F1 and F2 for the two high vowels examined here, i.e., /i/ and /u/. However, when the mean differences between the production values and MOA choices were compared with the step-size used in the MOA task, i.e., approximately 43 mels for F1 and 45 mels for F2, the differences observed between the production and MOA values were smaller than the step-sizes for F1 of the two high vowels. Specifically, the differences in F1 between two adjacent MOA stimuli in the high vowel area ranged from 37 Hz to 40 Hz, which are larger than the mean differences between production values and MOA choices observed for the high vowels (5.1 Hz difference for Japanese /i/, 12.1 Hz for Japanese /u/, 19.6 Hz for Greek /i/, and 19.7 Hz for Greek /u/). Thus, although the possibility remains that the listeners preferred extreme F1 for the above high vowels, it is also possible that the listeners chose MOA stimuli whose F1 was closest to the production average, which happened to be more extreme than the production average. In other words, listeners' preference for extremity may be restricted along F2 for the high vowels. Therefore, the listeners' preference for high vowels with extreme F1 values should not be interpreted as conclusive from the results of the present study.

On the other hand, the mean difference observed between Greek production values and their MOA choices for F1 of /a/ (63.7 Hz) is larger than the MOA step-size (54-56 Hz) in the corresponding area. Therefore, the Greek listeners' preference for /a/ stimuli with extreme F1 seems a solid result. Similarly, the mean differences between production values and MOA choices observed along F2 of Japanese /i, u/ and Greek /i/ were larger than the MOA step-size in the corresponding areas. The F2 step-size in the

MOA task was approximately 45 mels, which correspond to 51 Hz ~ 108 Hz, with 51 Hz being the difference between the stimuli with the smallest F2 values used for the current set of stimuli (in the extreme /u/ region), and 108 Hz being the difference between the stimuli with the largest F2 values (in the extreme /i/ region). As can be seen in Tables 4:11 ~ 4:12, the mean differences between the production values and MOA choices were 133.1 Hz for Japanese /i/, 176.3 Hz for Japanese /u/, and 131.0 Hz for Greek /i/, all of which were larger than the largest F2 step-size employed in the MOA task. Therefore, the listeners' preference for extreme F2 for the above vowels in comparison to the production values also seems robust. It should be noted, however, that the listeners did not seem to simply choose extreme F2 values but those which roughly correspond to the most extreme values in production. As can be seen in Fig. 4:8, the /u/ stimuli chosen by the Japanese listeners seem to cluster around the extreme end of the production range, and were scarce in the region where F2 was most extreme in the MOA grid. This implies that the listeners may have preferred F2 values that were around the most extreme of the production values.

The listeners' preference for higher F2 for /i/ in the perception task was statistically significant even in comparison with hyperarticulated production values, although hyperarticulated /i/ was found to have significantly more extreme F2 than /i/ produced in the normal speech condition, as shown in 4.3.2. The results of unpaired, two-tailed, t-tests implemented on F2 of the hyperarticulated production of /i/ and MOA choices were: t (88.1) = 3.77, p < .0001 for Japanese /i/; and t (98) = 4.29, p < .0001 for Greek /i/. This is in contradiction with the results obtained by Frieda (1997) using speakers of American English where hyperarticulated production of /i/ did not differ from their MOA choices in F2. The discrepancy may be due to the fact that American English vowels are unlikely to be sustained monophthongs. In other words, dynamic characteristics of Japanese and Greek /i/s are different from those of American English /i/, which may have led to discrepant results.

To summarise, in line with previous studies, vowel category ideals seemed to be extreme in comparison to the speakers' production average of the vowels, even when care was taken to eliminate possible factors that may lead to the listener's bias towards extreme vowels, except for F1 of Japanese /a/ and F2 of Greek /u/ whose production values were around the most extreme values available in the present MOA task. However, when the MOA step-size was taken into consideration, the listeners' preference for extreme vowels was convincingly shown only along F2 for the two high vowels /i, u/. Furthermore, the listeners' MOA choices tended to cluster around the extreme end of the production range, suggesting that vowel category ideals may be related to the production range of the vowels, possibly situated at the most extreme end of the speakers' production range.

4.3.3.2 Variability in MOA Choices

Another noticeable difference between the MOA choices and production values from Tables 4:9-4:10 above is the difference in the size of the vowel categories obtained from the two tasks. In general, standard deviations of F1 and F2 of the MOA choices were larger than those of the production values; in other words, the listeners' MOA choices were more variable than the production data. The MOA- and production-based vowel categories are plotted for each language group in Figs. 4:10-11. The finding that the vowel categories obtained from the MOA task were generally larger than the corresponding production-based categories is a surprising result, given the assumption that the category ideal is a representative member of the category, in other words, a selected member among all the category members. Although MOA choices are more variable than the production range for both languages and all the vowels examined here, the amount of variability seems to differ along F1 and F2 dimensions for different vowels and languages. For instance, the MOA categories are the smallest for /i/ across the two languages. At the same time, Japanese MOA choices for /i/ seem more variable along F1 than Greek /i/. Japanese choices for /u/ are also much more variable along F2 than the Greek counterpart. On the other hand, Greek choices are much more variable for /a/ along F2 than Japanese choices.

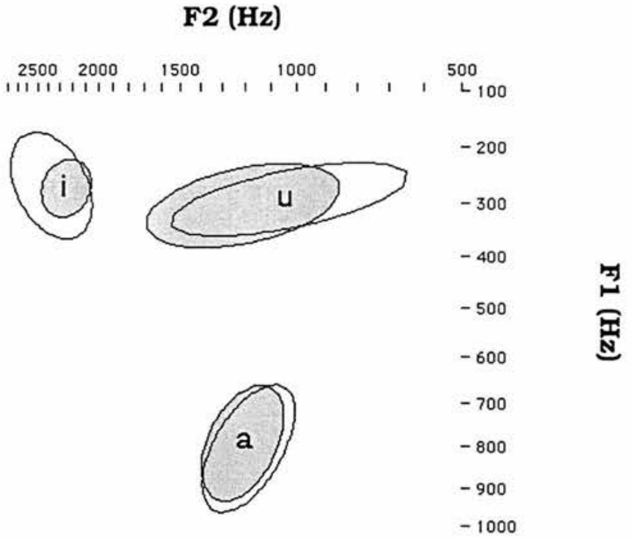


Figure 4:10. Japanese production values and MOA choices of category ideals for /i, a, u/. Grey ellipses represent two standard deviations from the production means of each vowel category. The production data collapse hyperarticulated and normal speech conditions. White ellipses represent two standard deviations from the means of MOA choices. The x-axis represents F2 in Hz, and the y-axis represents F1 in Hz.

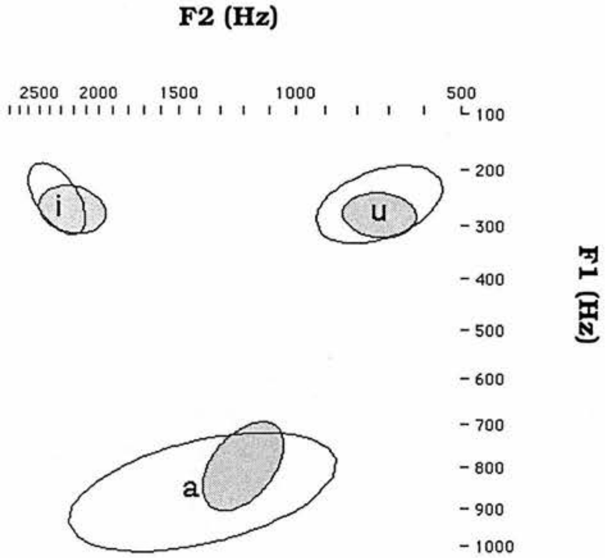


Figure 4:11. Greek production values and MOA choices of category ideals for /i, a, u/. Grey ellipses represent two standard deviations from the production means of each vowel category. The production data collapse hyperarticulated and normal speech conditions. White ellipses represent two standard deviations from the means of MOA choices. The x-axis represents F2 in Hz, and the y-axis represents F1 in Hz.

In order to examine whether the larger variability in the MOA choices in comparison to the production values was due to inconsistency within each subject or disagreement between subjects from the same language group, the individual subjects' range of MOA choices are plotted in F1-F2 vowel spaces (Figs. 4:12 ~ 13) separately for the two language groups. Data from subjects whose standard deviations were 0 both in F1 and F2 are excluded from these figures; Subjects J5 and G2's choices of /i/ are not plotted in Figs. 4:12-13. The means of each subject's MOA choices and standard deviations are summarised in Hz separately for each vowel and each language group in Tables 4:13-4:17. Those subjects whose data are not included in Figs. 4:12-4:13 due to standard deviations of 0 both in F1 and F2 are shaded in Tables 4:13-4:17.

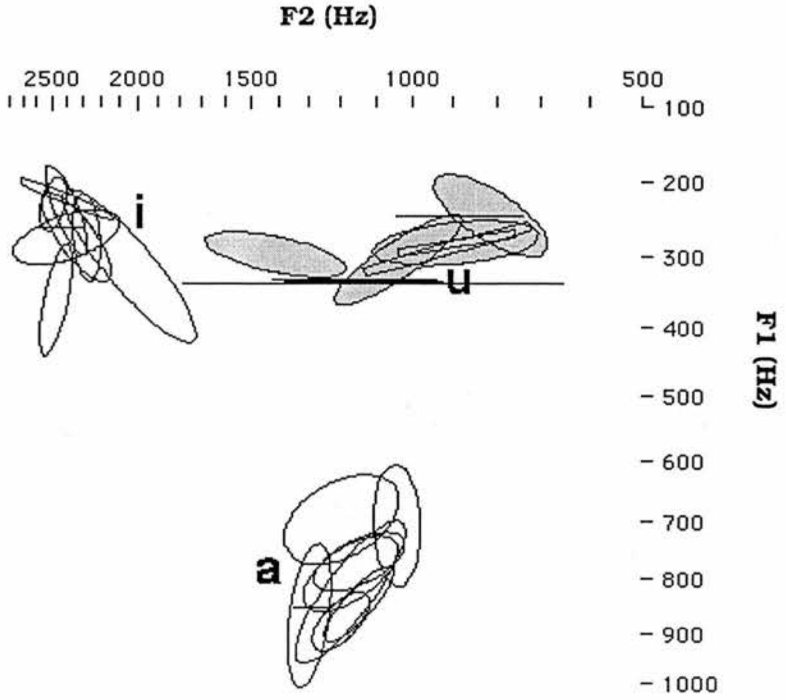


Figure 4:12. Individual Japanese listeners' MOA choices of category ideals for /i, a, u/. Each ellipse represents individual subjects' choices of MOA stimuli. Ellipses representing the /u/ category are painted in grey. The x-axis represents F2 in Hz, and the y-axis represents F1 in Hz.

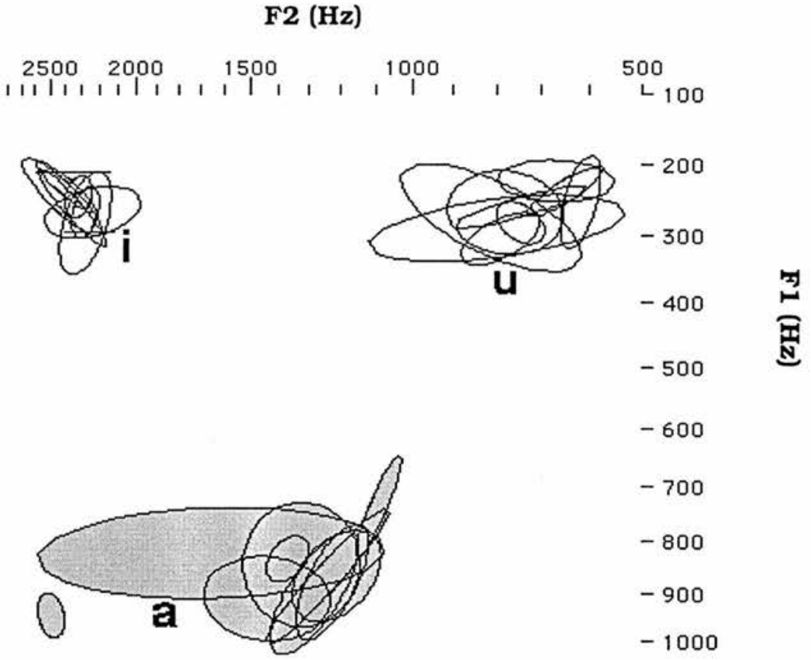


Figure 4:13. Individual Greek listeners' MOA choices of category ideals for /i, a, u/. Each ellipse represents individual subjects' choices of MOA stimuli. Ellipses

representing the /a/ category are painted in grey. The x-axis represents F2 in Hz, and the y-axis represents F1 in Hz.

Vowel	Formant	Subject	MOA means (Hz)	Standard deviation (Hz)
/i/	F1	J1	229.8	20.2657
		J2	274.8	20.8135
		J3	237.4	33.4858
		J4	275	34.3802
		J5	252	0
		J6	222.4	16.5469
		J7	314.2	60.0766
		J8	361.4	44.3655
		J9	275	33.541
		J10	252	0
/i/	F2	J1	2489.4	48.2991
		J2	2407	178.4152
		J3	2403.8	74.9613
		J4	2342	119.0756
		J5	2511	0
		J6	2407	178.4152
		J7	2047.2	197.835
		J8	2467.8	59.154
		J9	2278.8	45.1686
		J10	2468.6	94.8093

Table 4:13. Mean F1 and F2 and standard deviations of individual Japanese subjects' MOA choices for the /i/ category.

Vowel	Formant	Subject	MOA means (Hz)	Standard deviation (Hz)
/i/	F1	G1	259.8	31.4356
		G2	215	0
		G3	267.2	20.8135
		G4	267.2	20.8135
		G5	290	0
		G6	229.8	20.2657
		G7	215	0
		G8	244.6	16.5469
		G9	282.8	41.4934
		G10	244.8	31.3161
/i/	F2	G1	2362.8	118.7695
		G2	2511	0
		G3	2278.8	45.1686
		G4	2240.6	150.8221
		G5	2218.6	83.7872
		G6	2404.6	106.0038
		G7	2362.8	118.7695
		G8	2319.8	46.5102
		G9	2279.4	85.5821
		G10	2426.8	136.1844

Table 4:14. Mean F1 and F2 and standard deviations of individual Greek subjects' MOA choices for the /i/ category.

Vowel	Formant	Subject	MOA means (Hz)	Standard deviation (Hz)
/a/	F1	J1	844	0
		J2	877.6	30.6725
		J3	707.2	58.2469
		J4	768.8	29.0293
		J5	844.4	38.8947
		J6	823.4	73.197
		J7	696.6	41.9559
		J8	867.4	72.8958
		J9	790.2	37.8312
		J10	844.8	55.0018
/a/	F2	J1	1326.6	32.1994
		J2	1199.6	48.4438
		J3	1040.8	34.5065
		J4	1146.2	74.3586
		J5	1145.6	55.6983
		J6	1173.6	91.1608
		J7	1201.2	98.1157
		J8	1297.8	39.436
		J9	1173	76.9708
		J10	1200.2	68.5033

Table 4:15. Mean F1 and F2 and standard deviations of individual Japanese subjects' MOA choices for the /a/ category.

Vowel	Formant	Subject	MOA Means (Hz)	Standard Deviation (Hz)
/a/	F1	G1	867.4	72.8958
		G2	945.6	25.4912
		G3	737.8	52.0019
		G4	822.8	49.0836
		G5	833.2	24.1495
		G6	866.8	49.3072
		G7	878.2	63.4602
		G8	911.6	47.3318
		G9	845	68.1836
		G10	900.8	68.1814
/a/	F2	G1	1242.8	103.6349
		G2	2489.4	48.2991
		G3	1092.4	36.1497
		G4	1691.4	384.0557
		G5	1370.6	40.5315
		G6	1214	76.6942
		G7	1227.6	61.4394
		G8	1448	129.8576
		G9	1328.2	108.4306
		G10	1284.2	77.7766

Table 3:17. Mean F1 and F2 and standard deviations of individual Greek subjects' MOA choices for the /a/ category.

Vowel	Formant	Subject	MOA means (Hz)	Standard deviation (Hz)
/u/	F1	J1	297.8	17.4413
		J2	282.4	16.9941
		J3	305.8	35.3299
		J4	290.2	27.225
		J5	252	0
		J6	329	0
		J7	329	0
		J8	329	0
		J9	244.8	31.3161
		J10	282.4	16.9941
/u/	F2	J1	1418.2	141.657
		J2	896.2	114.7702
		J3	1042.4	98.0653
		J4	923.6	175.7421
		J5	895.2	86.958
		J6	1188.2	122.3303
		J7	1227.8	284.5851
		J8	1175.2	130.358
		J9	813.8	75.1279
		J10	899.6	164.8326

Table 4:16. Mean F1 and F2 and standard deviations of individual Japanese subjects' MOA choices for the /u/ category.

Vowel	Formant	Subject	MOA Means (Hz)	Standard Deviation (Hz)
/u/	F1	G1	259.6	16.9941
		G2	252.4	37.5007
		G3	305.6	21.3612
		G4	290.2	27.225
		G5	282.4	16.9941
		G6	267.4	33.5976
		G7	222.4	16.5469
		G8	229.8	20.2657
		G9	275.2	43.2863
		G10	237.2	20.2657
/u/	F2	G1	747.8	84.019
		G2	620.2	22.8079
		G3	790.8	50.9823
		G4	807.2	162.2889
		G5	746.4	29.577
		G6	780.4	72.7929
		G7	672.6	68.5332
		G8	692.8	29.0293
		G9	815.8	118.221
		G10	651.2	43.3901

Table 4:17. Mean F1 and F2 and standard deviations of individual Greek subjects' MOA choices for the /u/ category.

As shown in Figs. 4:12-4:13, both Japanese and Greek individuals' MOA choices only partially overlap one another, suggesting that there are certain degrees of disagreement among subjects as to the locations of the category ideals for the three vowels. Disagreement among subjects in the same language group regarding the locations of category ideals is also reported in previous studies. For instance, Samuel (1982) reports that there was a great inter-subject difference in the locations of category ideals, even among speakers of the same dialect. (Also see Aaltonen et al., 1997: 1093; Lively & Pisoni, 1997: 1668 Frieda, et al., in press; see, however, Johnson et al., 1993: 513; Iverson & Kuhl, 1995: 555).

Furthermore, Figs. 4:12 ~ 4:13 suggest that differences in the size of MOA categories observed between Japanese and Greek /i/ and /a/ categories may be largely due to a few anomalous subjects.[48] Specifically, the large variability observed in the choice of F2 for Greek /a/ can be largely accounted for by Subject G2, whose choices of F2 for the /a/ category (2489 Hz on the average) is more than 1000 Hz higher than the Greek average (1439 Hz), and Subject G4, whose standard deviation in F2 of his MOA choices for /a/ (384 Hz) is nearly as large as that of the entire Greek group (408 Hz) (see Fig. 4:13 and Tables 4:10 and 4:17). Similarly, a larger variability in the choice of MOA for Japanese /i/ observed along F1 seems to be greatly accounted for by Subject J7, whose standard deviation in F1 of the MOA choices (65.0 Hz) is larger than that of the entire Japanese group (55.1 Hz), and Subject J8, whose choices of F1 for /i/ (444.5 Hz) is 100 Hz higher than the Japanese average (343.1 Hz) (see Fig. 4:12 and Tables 4:9 and 4:13). At the same time, a certain degree of inconsistency is observed for all the subjects (except for Subject J5 and G2's choices of /i/) in their choices of category ideals. Considering the fact that there was no control over what stimuli in the MOA grid the subject would listen to before he made a decision, the inconsistency in the individual's choices of category ideals may be due to the context-sensitive nature of the listener's goodness judgements, as reported in Lively & Pisoni (1997: 1675).

Although it seems a general tendency for the MOA categories to be larger than the production value categories, the amount of variability in individuals' MOA choices seem to be different along F1 and F2, as can be seen in Figs. 12-13, which suggests that the listener may be insensitive to differences in vowel quality along different dimensions to different degrees. To compare the differences in the degrees of variability in individuals' MOA choices along F1 and F2 for each vowel category, the standard

---

[48] On the other hand, a greater variability in Japanese choices of F2 for the category /u/ does not seem to be due to a few anomalous subjects. Rather, it seems better accounted

deviations of each subject's MOA choices in F1 and F2 were computed using a mel scale, a psychoacoustically equidistant unit for pitch, whose means were compared separately for each vowel category and language group. Average F1 and F2 standard deviations for each vowel category is tabulated in mels separately for each language group in Table 4:18. (For individual subjects' standard deviations in mels, see Appendix D).

Vowel	Language	Mean F1 sd in mels	Mean F2 sd in mels	Mean difference
/i/	Japanese	29.42 (26.35 Hz)	43.73 (99.61 Hz)	-14.31 (-73.27 Hz)
	Greek	20.99 (18.27 Hz)	38.94 (89.16 Hz)	-17.95 (-70.9 Hz)
/a/	Japanese	35.18 (43.77 Hz)	40.83 (61.94 Hz)	-5.64 (-18.17 Hz)
	Greek	40.61 (52.01 Hz)	63.09 (106.69 Hz)	-22.47 (-54.68 Hz)
/u/	Japanese	16.28 (14.53 Hz)	98.08 (137.13 Hz)	-81.81 (-122.6 Hz)
	Greek	29.03 (25.40 Hz)	55.49 (68.16 Hz)	-26.45 (-42.76 Hz)

Table 4:18. A comparison of mean F1 and F2 standard deviations of individual Japanese and Greek subjects' MOA choices for /i, a, u/ in mels.

As can be seen Table 4:18, generally speaking, across the two language groups, the average standard deviations were larger in F2 than F1 for all the three vowel categories. A greater degree of instability in the listener's choices of F2 is also reported in Johnson et al. (1993: 512-513) and Frieda, et al. (in press) for an MOA task, and in Lively & Pisoni (1997) for a goodness-rating task. This is consistent with Disner's (1983:5) view that F1 is perceptually more salient than F2 due to its high amplitude. Furthermore, the difference between the mean standard deviations in F1 and F2 is especially noticeable for the Japanese subjects' choices of /u/ whose mean standard deviation in F1 is 16.3 mels whereas that in F2 is 98.1 mels. The great inconsistency in the Japanese listeners' choices of F2 for the /u/ category may also be a reflection of the large variability in the production values of this category (see Table 4:3).

4.3.3.3. A Comparison between Individual Subjects' MOA Choices and Production Values

As seen in 4.3.3.2, in Experiment 1b a fair amount of between-subject differences were found in both Japanese and Greek listeners' MOA choices, and there did not seem to be a single member that represents any of the vowel categories examined here for

---

for as a reflection of the large variability in the production data. This issue is discussed shortly.

either languages. Rather, MOA categories were generally larger than categories based on production values. Considering a certain degree of within-subject inconsistency shown in the previous section, it is conceivable that the between-subject differences observed in their MOA choices may have derived from subjects' inaccuracy in their choices of MOA stimuli. Alternatively, it may be the case that individual subjects interpreted the task differently, which may have resulted in a large between-subject differences in the MOA choices within the same language group. However, it is also possible that the differences reflected true differences in the locations of the category ideals among individuals. Supposing that individual speakers' mental representations, of which category ideals are a possible form, serve as targets in production (Flege, 1986: 31-2; Johnson et al., 1993), and that production values differ among different speakers of a given language, there may be a correlation between individuals' production values and MOA choices. In fact, such a correlation is reported in Fox (1982) for the point vowels /i, a, u/ [also see Frieda, et al. (in press) for /i/], and in Newman (1996) for VOT in /pa/ and /ba/ syllables (however, see Bailey & Haggard's (1973) study on the link between perception and production for VOT using a /kill-gill/ continuum). In this section, I examine whether the individual differences in MOA choices reflect individual differences in the production values of these vowels.

Shown in Figs. 4:14-4:15 are individual speakers' production of the three vowels under investigation, i.e., /i, a, u/, and their MOA choices plotted together for comparison. (Mean production values and MOA choices for each subject are in Appendices C & D). The ellipses in Figs. 4:16-4:17 cover two standard deviations from each subject's production means of each vowel, while the symbols represent their MOA choices for each vowel category. As can be seen in these figures, the individuals appeared to have chosen MOA stimuli whose F1 and F2 were in the vicinity of their own production values, although the degrees of approximation of individuals' MOA choices to their production values varied.

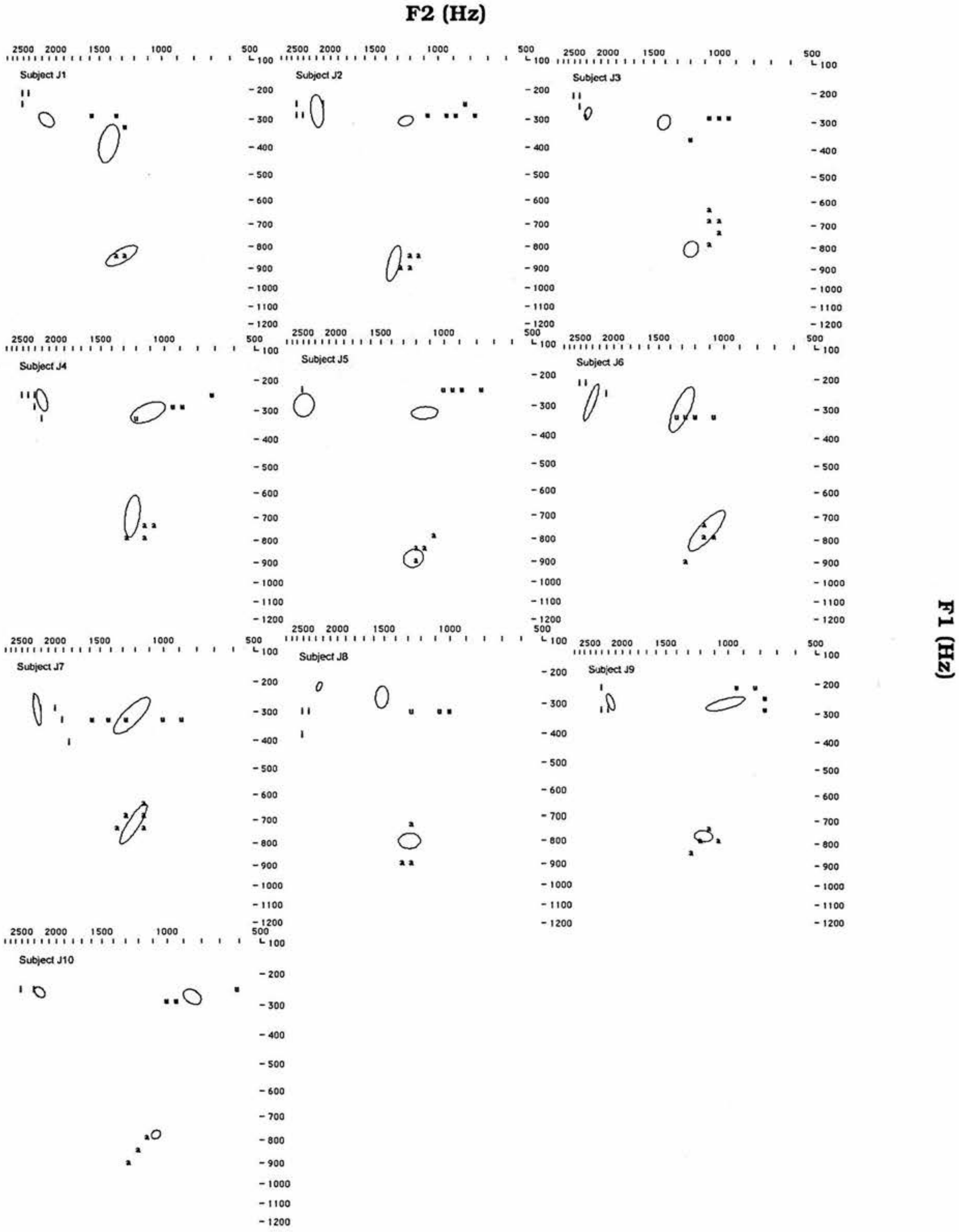


Figure 4:14. Individual Japanese subjects' production values and MOA choices. The ellipses cover two standard deviations from the production means of each subject. Their MOA choices are overlaid in the form of symbols representing each vowel category.

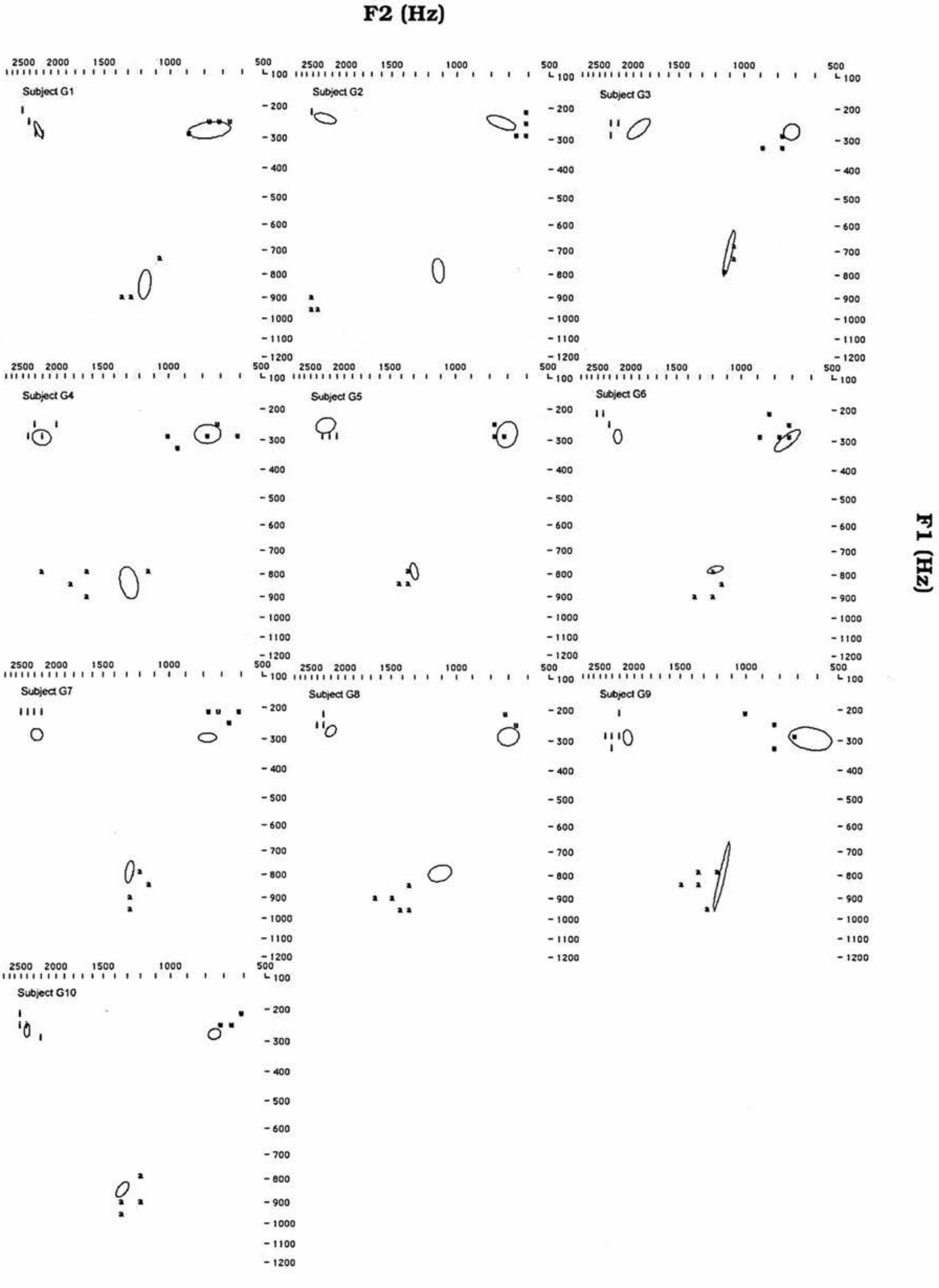


Figure 4:15. Individual Greek subjects' production values and MOA choices. The ellipses cover two standard deviations from the production means of each subject. Their MOA choices are overlaid in the form of symbols representing each vowel category.

The magnitude of correlation between individual subjects' MOA choices and their production values was examined using two-tailed Pearson tests, which were conducted on F1 and F2 of each vowel for each language group.

As for F1, Japanese production values and MOA choices for the category /a/ approached the significant level of .05 (r = .21, p = .072), but no significant correlation was found for other Japanese vowels or any of the Greek vowels. This may be due to the fact that the standard deviations in F1 of the production values of these vowels of the two language groups (cf. Tables 4:9 ~ 4:10) were smaller than, or nearly just as small as, the step-size in the MOA task. In other words, the step-size in the MOA task may have been too large to reflect between-subject differences in production values. Specifically, the F1 standard deviations of the production values of Japanese and Greek /i/ are 26.4 Hz and 21.8 Hz, respectively, and those of Japanese and Greek /u/ are 39 Hz and 21 Hz, while the MOA step-size in the high vowel region is from 37 Hz to 40 Hz (cf. Table 4:1). Similarly, the F1 standard deviation of Greek productions of /a/ is 52 Hz, while the MOA step-size in the low vowel region is from 51 Hz to 57 Hz. On the other hand, the F1 standard deviation of Japanese /a/ is 68 Hz, slightly bigger than the step-size, which might have improved the correlation.

As for F2, the correlation between Japanese production of /u/ and MOA choices was significant (r = .48, p = .026), but no significant correlation was observed otherwise. Again, the large MOA step-sizes in comparison to each language group's standard deviations of the production values may be the reason why a significant correlation was not found for the remaining vowels. As can be seen Tables 4:9 ~ 4:10, the F2 standard deviations of the production values of /i/ are 107 Hz for Japanese and 134 Hz for Greek, while the MOA step-size of the /i/ region is from 92 Hz to 108 Hz. The F2 standard deviations of the production values of /a/ are 87 Hz for Japanese and 85 Hz for Greek, while the MOA step-size in the central vowel area is around 70 Hz. Furthermore, the F2 standard deviation of the production of Greek /u/ is 55 Hz, whereas the MOA step-size in the back vowel region is form 50 Hz to 56 Hz. On the other hand, the F2 standard deviation of the production values of Japanese /u/ is 199 Hz, while the MOA step-size in the corresponding area is from 63 to 66 Hz, which might have led to a significant correlation between individuals' production values and MOA choices for Japanese /u/.

Altogether, from Figs. 4:14-4:15, it seems that in the MOA task individual subjects chose vowel sounds with F1 and F2 values that were in the vicinity of their own production values, suggesting that individual subjects' category ideals may serve as production targets, in which case category ideals of a language group could be as variable as the language group's phonetic realisations of the category. However, in the present experiment the correlation between individual subjects' MOA choices and production values was shown to be significant only for F2 of Japanese /u/. As explained, this may be partially due to the MOA step-size that was too large to reflect between-speaker differences in the production values of the remaining vowels. At the same time, it is conceivable that a smaller MOA step-size may not have improved the correlation, if the step-size was too small for the listeners to detect the differences in vowel quality between adjacent stimuli. Thus, what can be drawn from the present experiment is that the MOA choices may reflect their own production values, where between-speaker differences in the production values are large in magnitude.

4.3.4 Summary and Discussion

The major goal of Experiments 1a and 1b was to re-examine the location of vowel category ideals for Japanese and Greek /i, a, u/ placed in an F1-F2 space in relation to the production values of these vowels using comparable materials for production and perception tasks and without drawing on listeners' goodness judgements in locating category ideals, thereby eliminating possible factors that might have led to the listener's bias towards extreme vowels found in previous studies. As shown in 4.3.3.1, in line with the literature, Japanese and Greek vowel category ideals were found to be generally more extreme in comparison to the production averages, with exceptions of Greek /u/ and Japanese /a/, whose production values corresponded to the most extreme values that were available in the MOA task. The result is consistent with the view that the listener's preference for extreme vowels may be a universal tendency. To date, the listeners' preferences for extreme vowels have been reported for speakers of English and Spanish. (See Johnson et al., 1993; Bradlow, 1993; Iverson and Kuhl, 1995; Frieda, 1997; Lively & Pisoni, 1997).[49] However, in the present study listeners' preference for extreme vowels were convincingly shown only for the F2 of high vowels (with an exception of Greek /u/) and F1 of Greek /a/. Listeners' preference for front and back vowels with extreme F2 values is also found in Bradlow (1993) and Lively & Pisoni (1997). Furthermore, although the listeners preferred stimuli that were extreme in comparison to the production values, their choices were clustered around the extreme end of the production range of each vowel category, suggesting a link between the locations of category ideals and the production range.

The results also suggested that the subjects' MOA choices, or the locations of individuals' category ideals, differed considerably within each language group,

---

[49] In Aaltonen et al.'s (1997) study half of their Finnish subjects also exhibited such preferences. The other half preferred stimuli that are near the category boundary

suggesting that there may be no single location for the category ideal for a language group. It was thought that the inter-subject differences in the locations of category ideals might reflect the differences in the locations of the production targets, if category ideals serve as individuals' production targets. However, in the present experiment, the correlation between individuals' production values and their choices of MOA stimuli was convincingly demonstrated only for F2 of Japanese /u/ whose production values varied most between the subjects.

Finally, a fair degree of inconsistency within subjects was also observed in their MOA choices, although some subjects were more consistent than others. In line with the literature (Johnson et al., 1993; Lively & Pisoni, 1997; Frieda et al., in press), the listeners were generally more inconsistent in their choices of F2 than F1, suggesting that listeners may be more insensitive to the changes in vowel quality in F2 than F1.

---

instead, but this may be, as the authors speculate, because the fixed F3 of the stimuli that was atypical for Finnish /i/.

Chapter 5 The Effects of Vowel Prototypicality and Extremity on Discrimination Sensitivity

As reported in Chap. 4, the results of Experiments 1a and 1b were in confirmation of previous reports that vowel category ideals are more extreme than the speakers' production averages. Furthermore, the results suggested that vowel category ideals might cluster around the most extreme end of the production values, possibly reflecting the speaker's production targets. Thus, there are two possible locations of a vowel prototype towards which listeners' discrimination sensitivity might decline, assuming that prototypicality correlates with poor discrimination sensitivity, as Kuhl's Native Language Magnet Theory holds. However, as pointed out in Chap. 2, it is not clear from studies to date whether it is vowel prototypicality or extremity that correlates with poor discrimination sensitivity. Moreover, it is open to question whether the reported assimilation effect can be observed at the level of phonetic coding, since the perceptual magnet effect has been studied typically using long stimuli and/or short inter-stimulus intervals, which are found not to be optimal for phonetic coding (cf. Fujisaki & Kawashima, 1969; Pisoni, 1973, 1975). If the perceptual magnet effect is caused by language-specific prototypes, the effect should be observed at the level of phonetic coding, given the universal nature of auditory coding (cf. Repp et al., 1979; Miyawaki et al., 1975; Best et al., 1981). Change in discrimination sensitivity at the level of phonetic coding is of a more direct relevance to the present study, as its ultimate goal is to find out the factors that may cause language specific patterns of speech perception.

Experiments 2a (an identification task) and 2b (a discrimination task) were designed to investigate whether listeners' discrimination sensitivity within a vowel category decreased towards their category ideals (MOA choices), the production averages, or extreme vowels, i.e., the periphery of the acoustic vowel space, when the experimental conditions were optimal for phonetic coding. Specifically, in Experiments 2a and 2b, discrimination sensitivity curves for the vowel category /u/ were obtained from the same Japanese and Greek subjects, reusing the synthesised vowel stimuli in the high back region of the vowel space in Experiment 1b (the MOA task). In addition, both miss rates and d' were used as indices of the listener's discrimination sensitivity. This is to examine whether d' is in fact a better measure of the perceptual magnet effect, namely, whether d' better captures poorer discrimination sensitivity around the prototype, as Sussman & Lauckner-Morano (1995) report.

Experiment 2a was an identification task designed to investigate which stimulus pairs were both identified as /u/ by each subject and to determine, in turn, which portion of the discrimination sensitivity curve obtained in Experiment 2b could be regarded as reflecting changes in discrimination sensitivity within the /u/ category. As the present study is concerned with whether there are differences in degrees of discriminability between members of a single vowel category, it is crucial to find out how the stimuli used in Experiment 2b were labelled by the listeners, so that stimulus pairs identified as members of vowel categories other than /u/ are eliminated from the analyses.

Experiment 2b consisted of a discrimination task that examined the Japanese and Greek listeners' discrimination sensitivity in the high back area of the acoustic vowel space. In order to observe the effects of vowel prototypicality and extremity on discrimination sensitivity, a same-different discrimination task was conducted, which produced a continuous curve showing the change in discrimination sensitivity along F2 across the /u/ category. As both the production average and individuals' category ideals of Japanese /u/ obtained in Experiments 1a and 1b had substantially less extreme F2 values than those of Greek /u/, cross-linguistic differences were expected to be observed in the two language groups' discrimination curves along F2, if discrimination sensitivity decreased towards either of the two possible prototype locations. Furthermore, as the Japanese MOA choices, or category ideals, for the /u/ category had substantially more extreme values than the Japanese production average, it was expected that the effects of these two possible prototypes on Japanese listeners' discrimination sensitivity would be distinguished. On the other hand, if vowel extremity alone is correlated with poor discrimination sensitivity, both Japanese and Greek listeners' discrimination sensitivity would decrease towards stimuli with extreme F2 values.

In what follows, Experiment 2a (the identification task) and Experiment 2b (the discrimination task) are described in more detail in 5.1 and 5.2, respectively, followed by a discussion of the results in 5.3, which are summarised in 5.4.

5.1 Experiment 2a: Identification[50]

As just explained, Experiment 2a was an identification task designed to determine which stimulus pairs prepared for Experiment 2b (the discrimination task) were both identified as /u/ by each subject, and therefore which portion of the discrimination sensitivity curve obtained in Experiment 2b could be regarded as reflecting each subject's discrimination sensitivity within the /u/ category. The subjects identified each sound in the stimulus pairs consisting of two adjacent stimuli of the stimulus continuum, which comprised D pairs (stimulus pairs consisting of different stimuli) in the discrimination task, so that contrast effects on the identification of the stimuli would be taken into account in interpreting the discrimination sensitivity curves obtained in Experiment 2b. As mentioned in Chap. 2, according to Lotto et al. (1996), identification tasks that present stimuli singly do not accurately indicate whether discrimination between pairs of sounds are within- or cross- category discrimination, since two sounds presented in a pair are more likely to be assigned to different phonemic categories due to contrast effects (cf. Stevens et al., 1969:9), in which case discrimination between the pair can be no longer regarded as within-category. In order to make the experiment time minimal, the S pairs (stimulus pairs consisting of the same stimulus), which are not generally regarded to be subject to contrast effects, were not included in the identification task.

Subjects

Subjects were the same as those who participated in Experiments 1a (the production task) and 1b (the MOA task).

Materials

From the high back region (low F1 and F2) of the MOA grid used in Experiment 1b, 30 D pairs of a fixed F1 and varying F2 were created for each subject. The stimuli were 85-

---

[50] Following David Pisoni's advice (personal communication), the identification task was conducted before the discrimination task in order to elicit phonetic coding from the subjects. An identification task forces subjects to assign phonetic interpretations to the synthesised stimuli, which is found to be carried over to the next task using the same stimuli. The difficulty of leaving the phonetic mode after assigning phonetic interpretations to sounds is also demonstrated in Bailey et al. (1977) (in Repp et al. 1979: 143).

ms-long synthesised steady-state vowels modelled on a male voice. The F0 of the stimuli was set at 120 Hz initially and fell linearly to 105 Hz. F1 was fixed for each subject at approximately halfway between the average F1's obtained from his production values (the normal speech condition) and his MOA choices. Thus, F1 of the stimulus continuum was set differently for each subject, ranging from 252 Hz to 329 Hz (One Japanese and five Greek subjects were given a stimulus set whose F1 was 252 Hz, eight Japanese and five Greek subjects were given a set whose F1 was 290 Hz, and one Japanese subject was given a set whose F1 was 329 Hz). Ideally, F1 should have been set at halfway between individuals' MOA choices and the language group's production average, as these two were considered possible prototype locations in the present study. However, this was not possible, since some subjects were leaving before the time all the subjects would have completed Experiment 1a (the production task), and therefore the production average of the entire language group was not available when these subjects took part in Experiments 2a and 2b. The F1's of the stimulus sets given to the individual subjects are given in Table 5:1, together with the average F1's of their production values and MOA choices. F2 of the stimuli varied from 610 Hz to 1570 Hz in a step-size of 45 mels, comprising 16 stimuli, from which 30 one-step D pairs [15 combinations of adjacent stimuli × 2 temporal orders (AB and BA)] were created. This range of F2 roughly covered the production range of Japanese and Greek /u/ and the stimuli chosen by the two language groups in the MOA task. The stimulus set given to the subjects is schematised in Table 5:2 with its sixteen F2 values in Hz and mels. The stimulus pairs were identical to the D pairs given to each subject in the discrimination task (Experiment 2b).

Subject	F1 of discrimination stimuli (Hz)	Average F1 of production values (Hz)	Average F1 of MOA choices (Hz)	$\frac{\text{(Production value + MOA)}}{2}$ (Hz)
J1	329	363	298	330.5
J2	290	303	282	292.5
J3	290	291	306	298.5
J4	290	313	290	301.5
J5	290	329	252	290.5
J6	290	268	329	298.5
J7	290	285	329	307
J8	290	271	329	300
J9	252	271	245	258
J10	290	275	282	278.5
G1	252	269	260	264.5
G2	252	256	252	254
G3	290	264	306	285
G4	290	283	290	286.5
G5	290	268	282	275
G6	290	287	267	277
G7	252	297	222	259.5
G8	252	281	230	255.5
G9	290	301	275	295.5
G10	252	275	237	256

Table 5:1. F1 of the stimulus set given to each subject, and the average F1 values obtained from each subject in the production study (the normal speech condition) and MOA task. (J1 ~ J10: Japanese subjects; G1~ G10: Greek subjects).

F1: fixed for each subject (see Table 5:1).

F2	610 Hz	(687 mels)
	661	(732)
	714	(777)
	768	(822)
	824	(867)
	882	(912)
	941	(957)
	1003	(1002)
	1066	(1047)
	1132	(1092)
	1199	(1137)
	1269	(1182)
	1341	(1227)
	1415	(1272)
	1491	(1317)
	1570	(1362)

Table 5:2. The stimulus set given to each subject. F1 was fixed for each subject.

Procedure

The experiment was conducted in the experiment room at the Department of Linguistics, the University of Edinburgh on a separate day from Experiments 1a-b.[51] Each of the Japanese and Greek subjects identified each vowel sound constituting the 30 D pairs as /i/, /u/, or /o/ (a forced three-choice identification task). Categories /i/ and /o/ were given as alternative choices, as they are adjacent to the /u/ category in both language groups' vowel

---

[51] As Experiment 2b (a discrimination task) was rather long (45 – 60 minutes per subject) and tiresome, it was difficult to conduct all the four experiments on the same day. As Experiments 1a-b were concerned with the locations of prototypes, which are thought to be stable, this should not have affected the interpretation of Experiments 2a-b, should the preceding two experiments in fact captured the locations of such prototypes.

systems. The same inter-stimulus interval (1s) and inter-trial interval (2s) as in the discrimination task (Experiment 2b) were used in order to make the two tasks comparable. The subjects identified each vowel in the 30 D pairs presented binaurally through headphones, by clicking on one of the three buttons presented on the computer screen, each of which was labelled /i/, /u/ and /o/. A programme written by Norman Dryden, the Department of Linguistics, the University of Edinburgh, was used to implement the task. The 30 pairs were randomised and presented five times in five separate blocks, yielding 150 responses from each subject (30 pairs × 5 repetitions = 150 responses). Before the start of the task, the subjects were familiarised with the procedure using stimuli taken from different areas of the MOA grid used in Experiment 1b. The task was self-paced and lasted for about 15 –20 minutes for each subject.

Analyses

Based on the 150 identification responses from each subject, the number of times each stimulus was identified as /i/, /u/ and /o/ in the context of its adjacent stimuli was tallied and converted to percentiles. The results were used to determine which portion of each subject's discrimination curve obtained in Experiment 2b could be regarded as reflecting his discrimination sensitivity within the /u/ category. The results were also used to calculate predicted miss rates and d', two measures of discrimination sensitivity, which were compared with the actual miss rates and d' obtained in Experiment 2b. If improvement observed in the obtained sensitivity, if any, is limited to the stimuli and the degree predicted by the results of the identification task, it follows that the changes in discrimination sensitivity may not be due to the perceptual magnet effect but the identification of the paired stimuli as different phonemes that resulted in cross-category discrimination, as Lotto et al. (1996) argue.

Each subject's predicted miss rates and predicted d' scores were calculated on the basis of predicted hit rates derived from his identification performance. Predicted hit rates were computed in an analogous way to Liberman et al. (1957: 363), where it is assumed that a stimulus pair consisting of different phonemes would result in a hit ('different' responses to D pairs). In other words, it is assumed that the subject is able to tell two members of a stimulus pair apart if he assigned different labels to the pair of stimuli, and that subjects are not biased and make no mistakes in the identification task. As demonstrated later in 5.3.1, in Experiment 2a paired stimuli around the category boundary were not always given contrasting labels within a single trial but were, at times, given different labels from one trial to another, indicating the ambiguous status of these stimuli. It was felt that there was a greater chance for those ambiguous pairs to be contrasted within a trial in the discrimination task than those which are identified consistently as the same vowel across trials, given a larger number of trials (10 times) in the discrimination task. Therefore, predicted hit rates were calculated using two formulas: One formula only took account of the assignment of different labels to a pair of stimuli within a trial, and thus served as a conservative measure of predicted hit rates, while the other formula took account of different labels given to a certain stimulus pair across trials, which served as a less conservative measure. The first formula is expressed as:

P hit = 1 - (pii + puu + poo), (1)

where 'P hit' is the predicted hit rate, 'pxx' is the probability at which both stimuli in a pair would be identified as /x/ in the discrimination task, which equals the rate at which the first and second stimuli in the pair were both identified as /x/ within the same trial in the identification task. Thus, 'pii' is the rate at which both stimuli in a given pair were identified as /i/ within the same trial, 'puu' is the rate at which both stimuli in the pair were identified as /u/, and so on. For instance, if Stimuli A and B are identified as in Table 5:3 below, the predicted hit rate for Stimuli A-B pair is .1, as:

P hit = 1 - (white portion in Table 5:3)
= 1 - (.4 + .5 + 0) = .1.

Thus, Formula (1) only takes account of the assignment of different labels to the two sounds in a pair within a trial. On the other hand, Formula (2) computes predicted hit rates on the basis of the identification of a given stimulus pair across trials:

P hit = {[p(A)i × p(B)í] + [p(A)u × p(B)ú] + [p(A)o × p(B)ó] + [p(B)i × p(A)í] + [p(B)u × p(A)ú] + [p(B)o × p(A)ó]}/2, (2)

where 'p(A)i' in 'p(A)i × p(B)í' is the rate at which a given stimulus in the pair, in this case Stimulus A, is identified as /i/ when presented before Stimulus B across trials, and 'p(B)í' is the rate at which Stimulus B is not identified as /i/ when presented after Stimulus A across trials. Similarly, 'p(B)i' in 'p(B)i × p(A)í' is the rate at which Stimulus B is identified as /i/

when presented before Stimulus A across trials, and ‘p(A)i‘ is the rate at which Stimulus A is not identified as /i/ when presented after Stimulus B. Therefore, the example presented above in Table 5:3 would yield a predicted hit rate of .5, as:

$$P\ \text{hit} = [(.4 \times .4) + (.6 \times .6) + (0 \times 1) + (.4 \times .6) + (.6 \times .4) + (0 \times 1)]\ /2 = .5.$$

Trial	Stimulus A	Stimulus B	Presentation Order
1	/i/	/i/	AB
2	/u/	/i/	AB
3	/u/	/u/	AB
4	/u/	/u/	AB
5	/i/	/i/	AB
6	/u/	/u/	BA
7	/i/	/i/	BA
8	/i/	/i/	BA
9	/u/	/u/	BA
10	/u/	/u/	BA

Table 5:3. Hypothetical identification results for a stimulus pair A-B. The pair is presented 5 times each in two temporal orders. The trials in which context effects are observed are shadowed.

Formula (1) and Formula (2) were taken to indicate the minimum and maximum possibilities of occurrences of cross-category discrimination, respectively. Therefore, to the extent that the actual hit rate fell between the two predictions, it was regarded as due to discriminability of the stimuli arising from assignment of different labels to the paired stimuli. The probability that the paired vowel sounds would be identified as different phonemes was thus converted to a predicted hit rate, which was then transformed into predicted miss rates (‘same’ responses to D pairs) and d’ scores. Predicted miss rates were computed as:

$$P\ \text{miss} = 1 - P\ \text{hit}, \qquad (3)$$

where P hit is the predicted hit rate, and P miss is the predicted miss rate. Predicted d' was computed using Macmillan & Creelman's (1991) formula:

$$d' = v\,[z(H) - z(F)], \qquad (4)$$

where v is a constant, H and F are hit and false-alarm rates ('different' responses to S pairs), and z is the z-transformation, which converts hit and false-alarm rates to a z-score, i.e., standard deviation units. As can be seen in Formula (4), in principle, d' scores are based on the difference between hit and false-alarm rates, and the intention is to discount hits resulting from the subject's guessing and are therefore correct by chance. As the S pairs were not labelled by the subjects in the present experiment on the assumption that they would be identified as the same vowel, the false-alarm rate was assumed to be .025 for all the pairs. F = .025 was used to match the value with that used for the results of Experiment 2b, a value recommended by Macmillan & Creelman (1991) when the actual false-alarm rate is 0 when each S pair is presented 10 times.[52]

5.2 Experiment 2b: Discrimination

In Experiment 2b, Greek and Japanese subjects' discrimination sensitivities along F2 were obtained, reusing the stimuli in the high back region of the vowel space in the MOA task (Experiment 1b), in order to test whether, at the level of phonetic coding, the listener's discrimination sensitivity decreased towards the language group's production average, individual listeners' category ideals, or the periphery of the vowel space.

Subjects

The same Japanese and Greek subjects as in Experiments 1a (the production task), 1b (the MOA task) and 2a (the identification task) participated in the discrimination task.

Materials

The stimuli in the high back region of the MOA grid used in Experiment 1b were reused to create stimulus continua for the discrimination task. As explained in Chap. 4, all

---

[52] $F = .025$ derives from $1/2\times2N = 1/2\times20 = .025$, where N is the number of repetitions.

the stimuli used in the MOA task were 85-ms-long synthesised steady-state vowels modelled on a male voice (for full details, see 4.2). The stimuli were made short, so that they could be used in the discrimination task that was designed to elicit phonetic coding. As Explained in 5.1, a stimulus continuum of a fixed F1 and varying F2 was prepared for each subject. The F1's of the stimulus sets given to the individual subjects ranged from 252 Hz to 329 Hz (cf. Table 5:1). F2 of the stimuli varied in a step-size of 45 mels, from 610 Hz to 1570 Hz for all the subjects (16 stimuli), a range of F2 that roughly covered the production range of Japanese /u/ and Greek /u/, and the stimuli chosen in the MOA task by the two language groups (cf. Table 5:2). As Japanese /u/ is not extreme, all the Japanese subjects were tested on stimuli including those whose F2's were more extreme than both their production values and MOA choices. The stimulus continua yielded for each subject 16 S pairs and 30 D pairs consisting of 15 combinations of adjacent stimuli arranged in two temporal orders (AB and BA, 15 combinations × 2 temporal orders = 30 pairs).

Procedure

The experiment was conducted in the experiment room at the Department of Linguistics, the University of Edinburgh. The 16 S and 30 D pairs were mixed in a randomised order and played binaurally through headphones to the subjects at a comfortable listening level. The subjects were instructed to answer whether the two sounds were 'same' or 'different' by clicking on the buttons displayed on the computer screen (a roving same-different discrimination task). A roving discrimination task that presents paired stimuli in a random order was preferred to a fixed discrimination task, in which one of the stimulus in the pair is fixed across trials, so that learning effects would be minimal. A programme written by Norman Dryden, the Department of Linguistics, the University of Edinburgh, was used to implement the paradigm. Each pair was presented ten times to each subject in five separate blocks, yielding 460 responses [(16 S pairs + 30 D pairs) × 10 repetitions = 460 responses]. The inter-stimulus interval was 1s. A relatively long inter-stimulus interval was employed to facilitate phonetic coding. A 2s inter-trial interval was programmed to start when the subject pressed the response button. The subjects were first familiarised with the procedure using different stimuli from those in the test trials. As the experiment was long, the subjects had a 10-15 minute break after the completion of the first three blocks and continued with the task after the break. The experiment lasted for 45 – 60 minutes per subject including the break.

Analyses

Changes in each subject's discrimination sensitivity along varying F2 (610 Hz ~ 1570 Hz) were calculated using two metrics, i.e., miss rates and d'. As explained earlier, miss rates are the rate at which the subject responds 'same' when the pair of stimuli are actually different. D' is a sensitivity measure that normalises the subjects' response biases (the willingness to say 'same' or 'different'), by, in principle, subtracting false-alarm rates ('different' responses to S pairs) from hit rates ('different' responses to D pairs), and thereby correcting for 'different' responses that are correct by chance [see Formula (4) given in 5.1]. The sensitivity curves were checked against the identification results obtained in Experiment 2a in order to determine which portion of the sensitivity curve could be regarded as reflecting changes in discrimination sensitivity for members of the /u/ category. Furthermore, the obtained discrimination sensitivity was compared with predictions made on the basis of each subject's identification performance as described in 5.1. If improvement observed in obtained sensitivity is limited to the stimuli and degree predicted by the identification results, it implies that the changes in discrimination sensitivity may be simply a reflection of differences in sensitivity for within- and cross-category discrimination (cf., Lotto et al., 1996).

Miss rates and d' were compared, as Sussman & Lauckner-Morano (1995) used both measures in their replication of Kuhl (1991) and report that d' is more sensitive to the perceptual magnet effect than miss rates. Thus, the present study compared the two measures to see whether d' would serve as a better measure of the perceptual magnet effect. The two metrics were evaluated in the light of the assumption underlying each metric: As explained in Chap. 2, the use of miss rates as a measure of discrimination sensitivity presupposes 100% correct-rejection rates ('same' responses to S pairs) along the stimulus continuum, while the use of d' as a measure of perceptual distance assumes a constant response bias across the stimulus continuum. Response bias along the stimulus continuum was computed for each subject using the basic bias measure c for Signal Detection Theory (Macmillan & Creelman, 1991: 10-33). C is defined as:

$$c = -.5\ [z\,(H) + z\,(F)].$$

As in the previous formula, H stands for the hit rate, F is the false-alarm rate, and z is the z-transformation. Negative c values arise when the false-alarm rate exceeds the miss rate, and indicate that the subject is more inclined to respond 'different'. Positive c values arise when the false-alarm rate is lower than the miss rate, and indicate that the subject is more inclined to respond 'same'.

5.3 Results and Discussion

In this section the results of Experiment 2a (the identification task) and 2b (the discrimination task) are discussed. In 5.3.1, the results of the identification task are presented, which is followed by 5.3.2, where the results of the discrimination task are analysed.

5.3.1 Results of the Identification Task (Experiment 2a)

In the way described in 5.1, the Japanese and Greek subjects labelled both sounds in the D pairs prepared for the discrimination task (Experiment 2b) as /i/, /u/, or /o/. The results were used to decide which D pairs in the discrimination task were both regarded as members of a single vowel category /u/, and which portions of the discrimination sensitivity curves obtained in Experiment 2b therefore reflected the change in discrimination sensitivity within the /u/ category.

The results of the identification task are summarised in Figs. 5:1 ~ 5:4, which present the percentage of identification of each stimulus as one of the above three vowels in two different contexts, i.e., the context of each of its two adjacent stimuli. Figs. 5:1a ~ 5:2c show identification results of individual Japanese and Greek subjects, while Figs. 5:3 ~ 5:4 show collective results from each language group. In all of these figures, the ordinate gives the percentile at which the given stimulus was identified as /o/, /u/ and /i/ accumulatively in this order, which is each represented by gray, white, and dark areas. The abscissa gives F2 values of the stimuli in Hz and, in parentheses, F2 of the adjacent stimuli with which given stimuli were paired. As there was no consistent effect of the order of presentation of the stimuli in a pair in the identification task, the plots collapse identification of a given stimulus paired with one of its two adjacent stimuli in two temporal orders. That is, if Stimulus A was identified as /i/ 50% of the time and as /u/ 50% of the time when it was presented <u>before</u> Stimulus B, and as /u/ all the time when it was presented <u>after</u> Stimulus B, the plot would read 25% /i/-identification [(50% + 0%) / 2 = 25%] and 75%-/u/ identification [(50%+100%) / 2 = 75%] for Stimulus A identified in the context of Stimulus B.

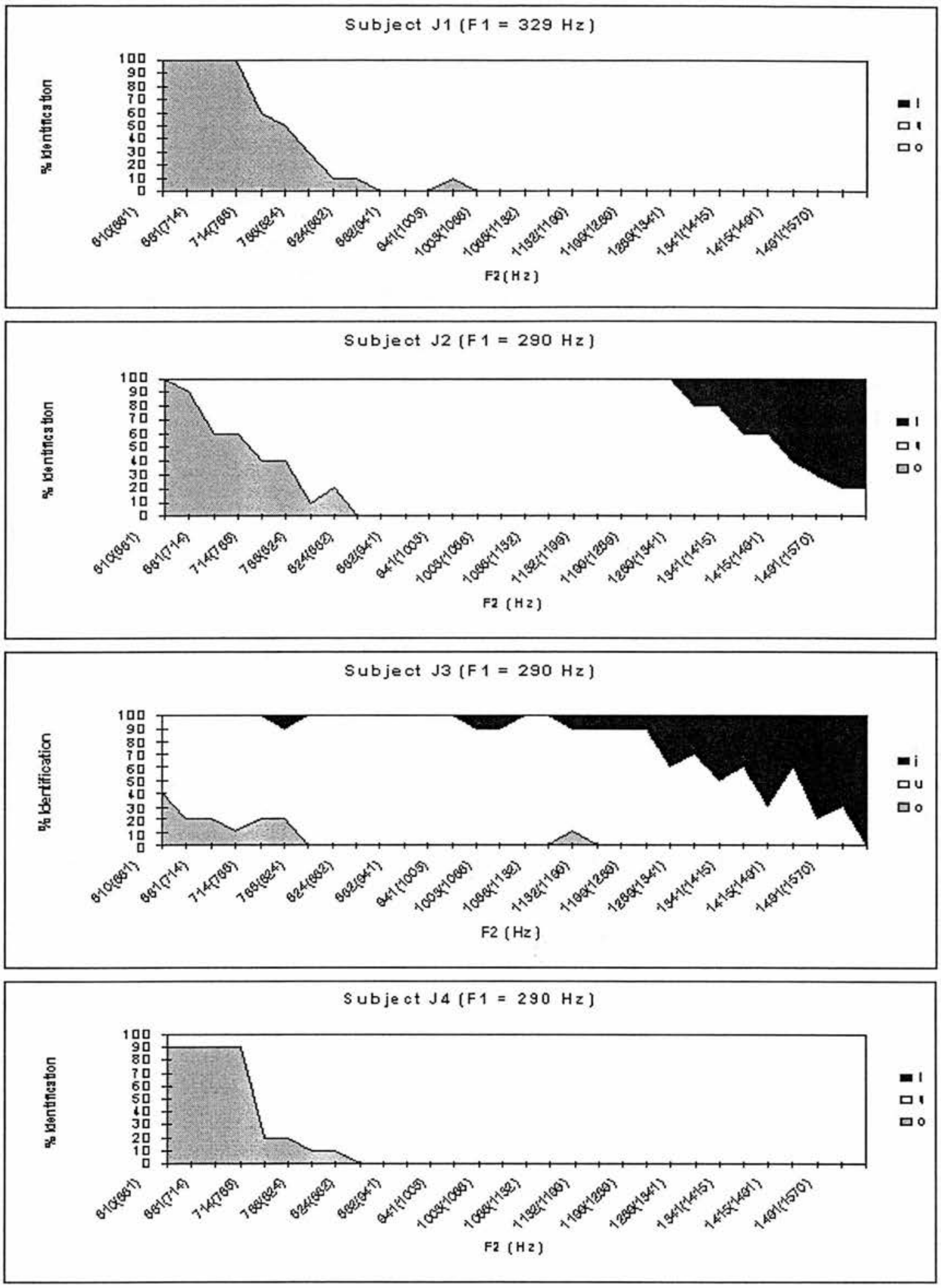


Figure 5:1a. Individual Japanese' identification results. The F1 values of the stimuli given to each subject is given at the top of each graphic.

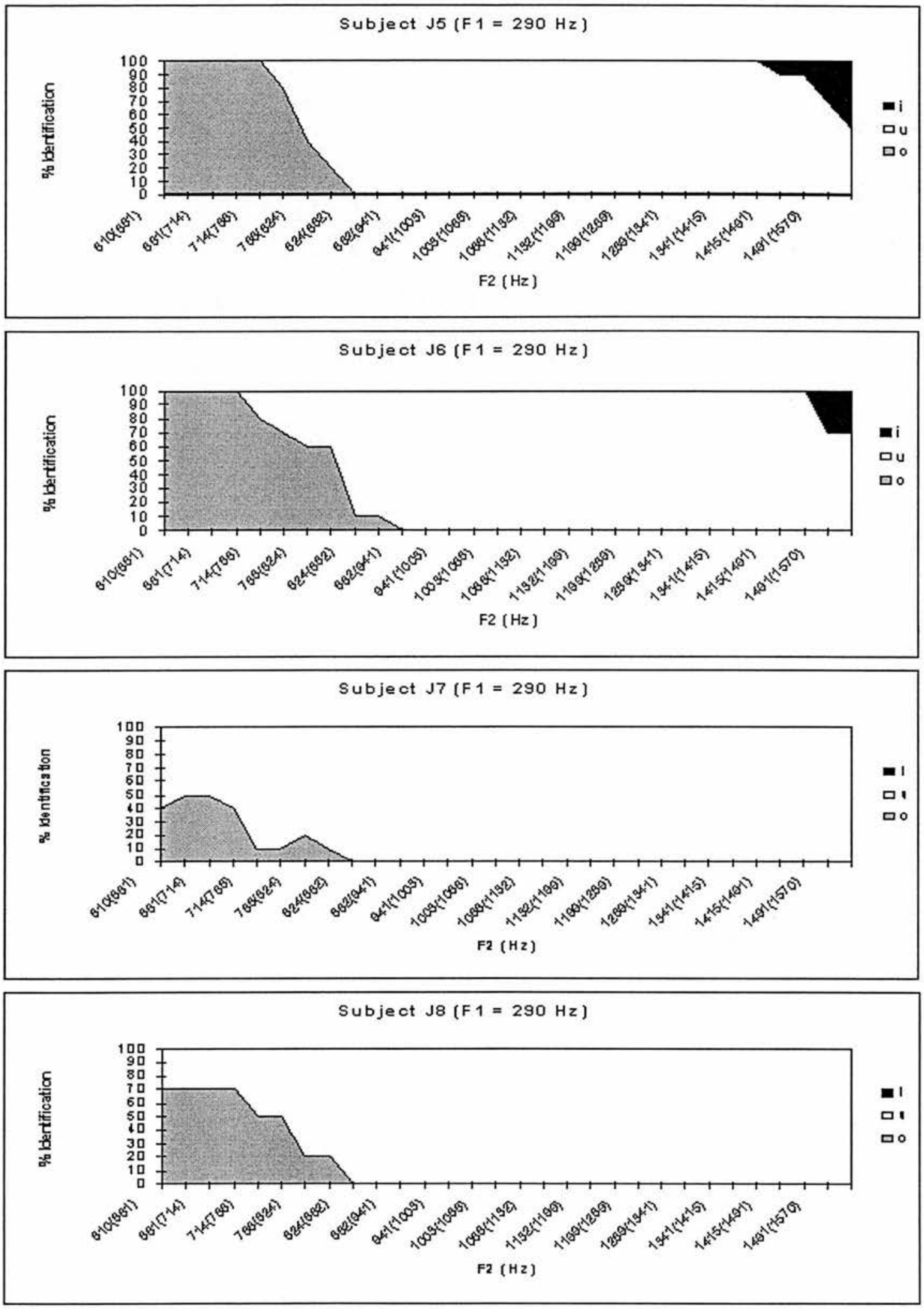


Figure 5:1b. Individual Japanese' identification results. The F1 values of the stimuli given to each subject is given at the top of each graphic.

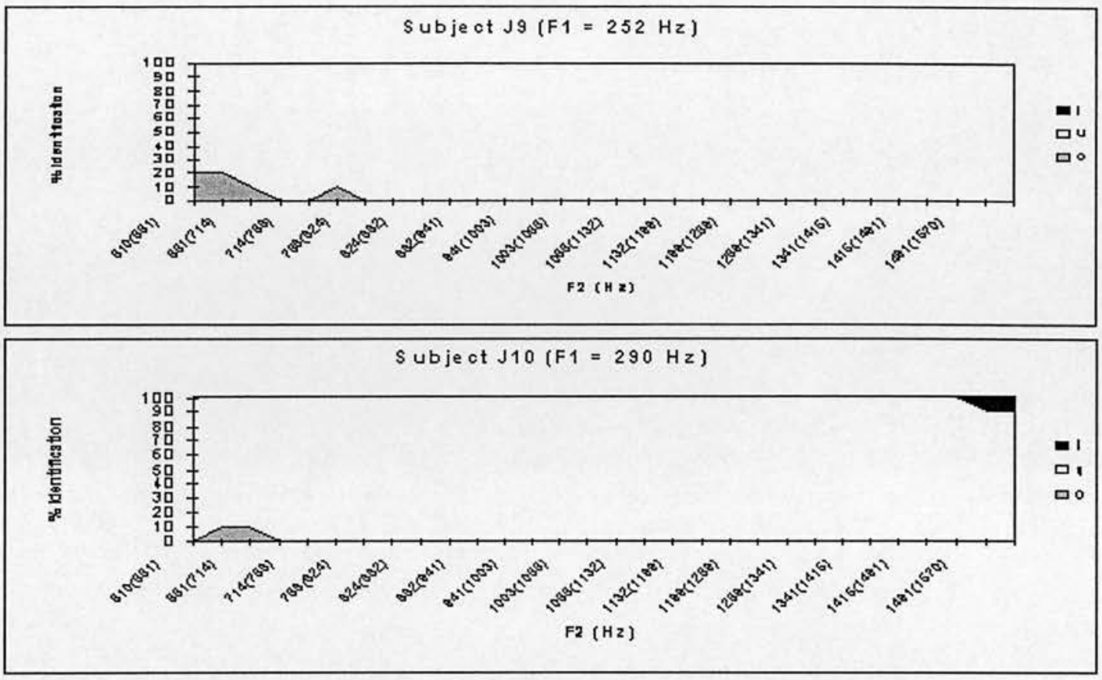


Figure 5:1c. Individual Japanese' identification results. The F1 values of the stimuli given to each subject is given at the top of each graphic.

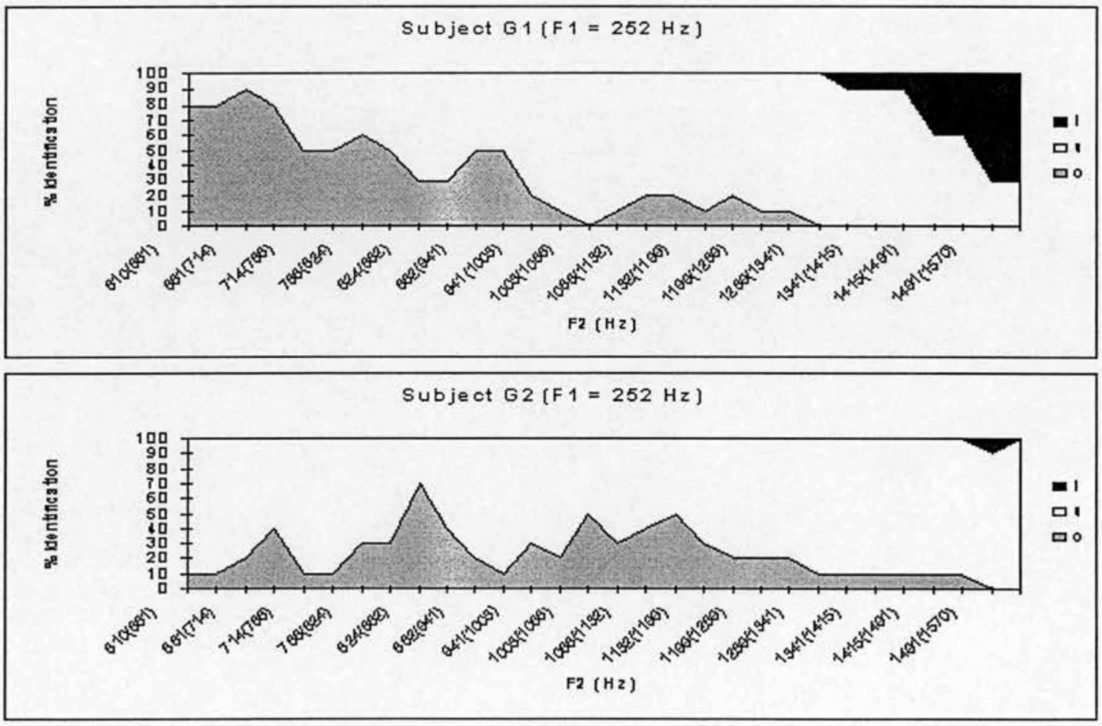


Figure 5:2a. Individual Greek' identification results. The F1 values of the stimuli given to each subject is given at the top of each graphic.

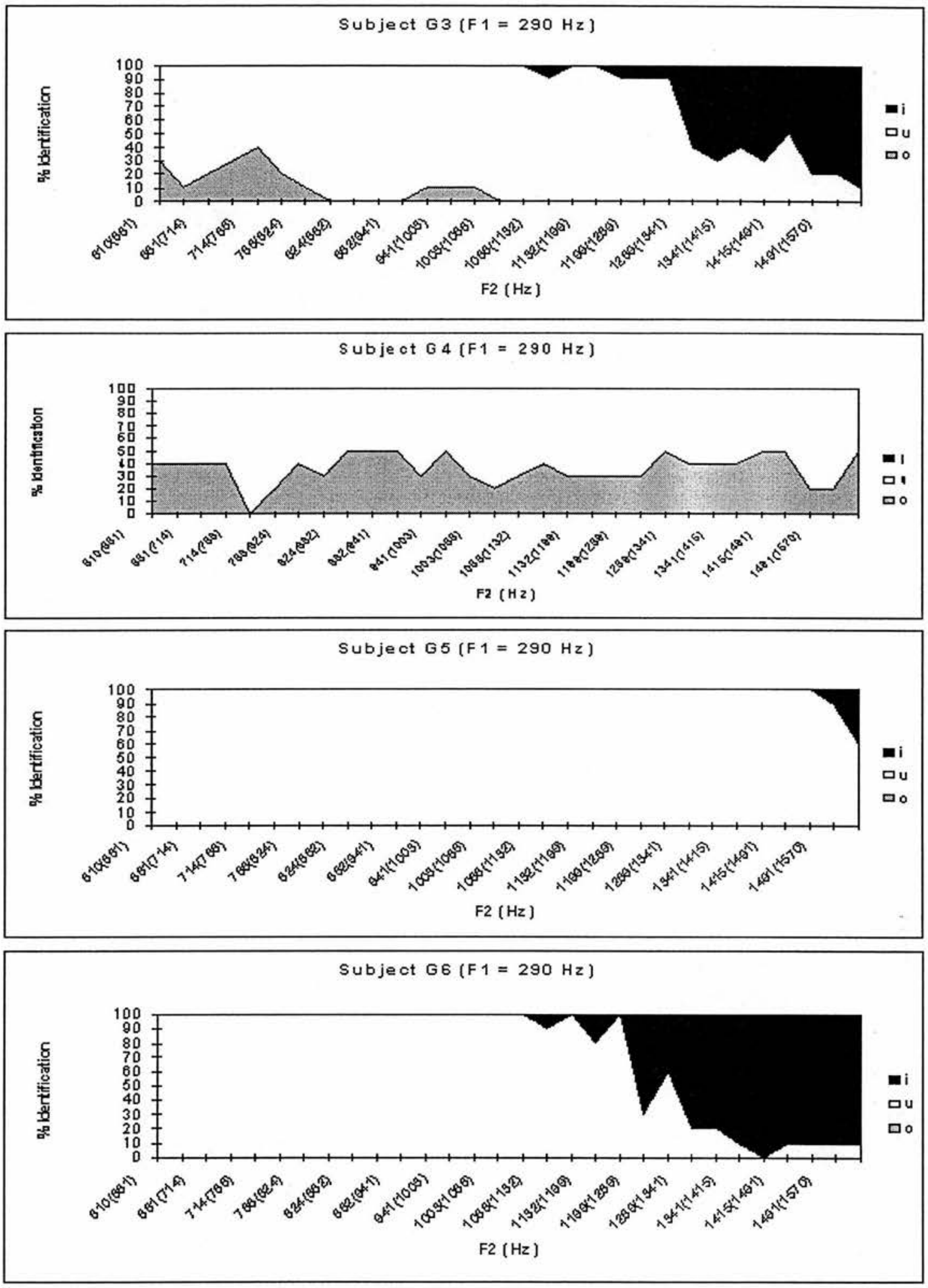


Figure 5:2b. Individual Greek' identification results. The F1 values of the stimuli given to each subject is given at the top of each graphic.

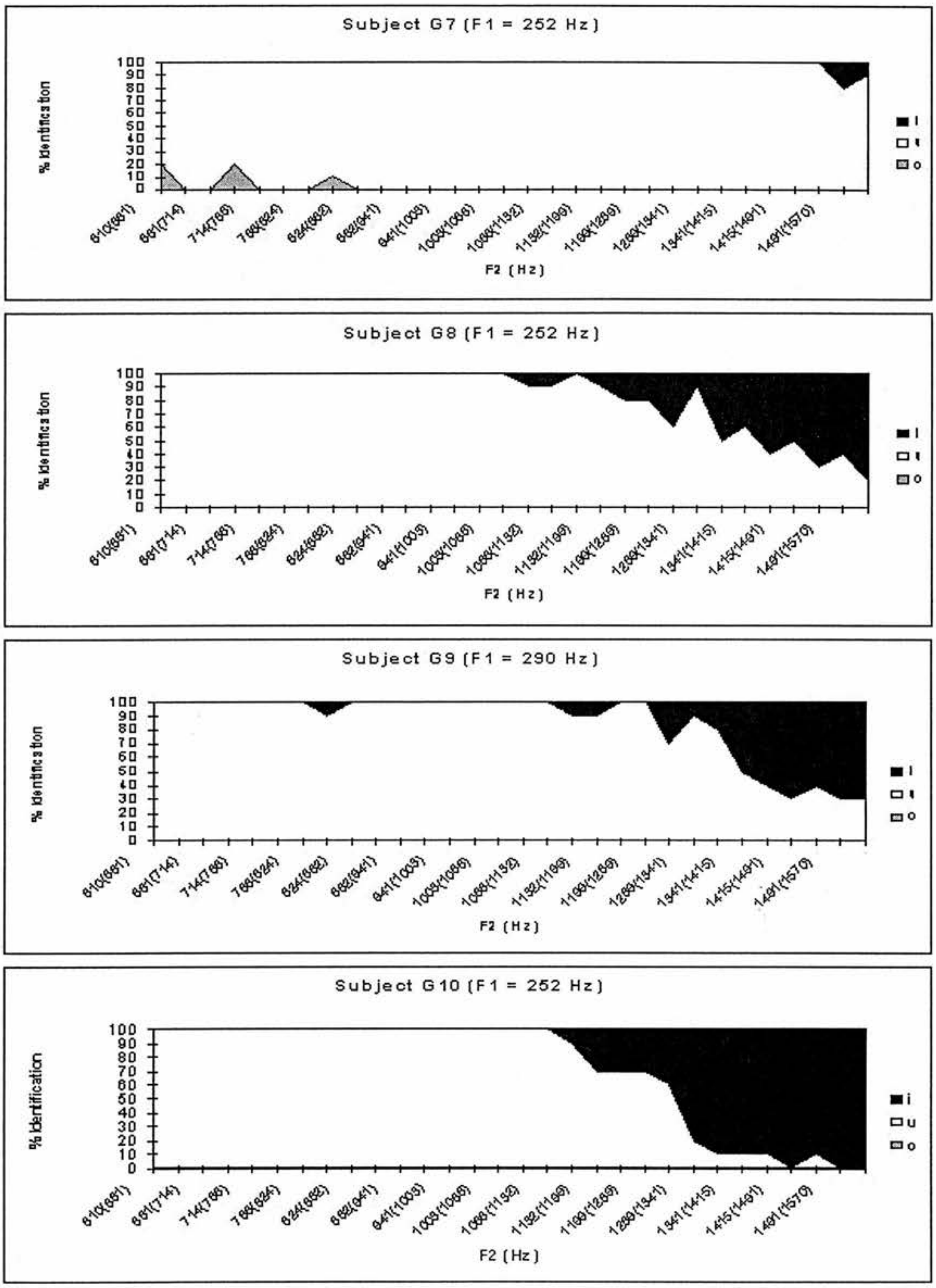


Figure 5:2c. Individual Greek' identification results. The F1 values of the stimuli given to each subject is given at the top of each graphic.

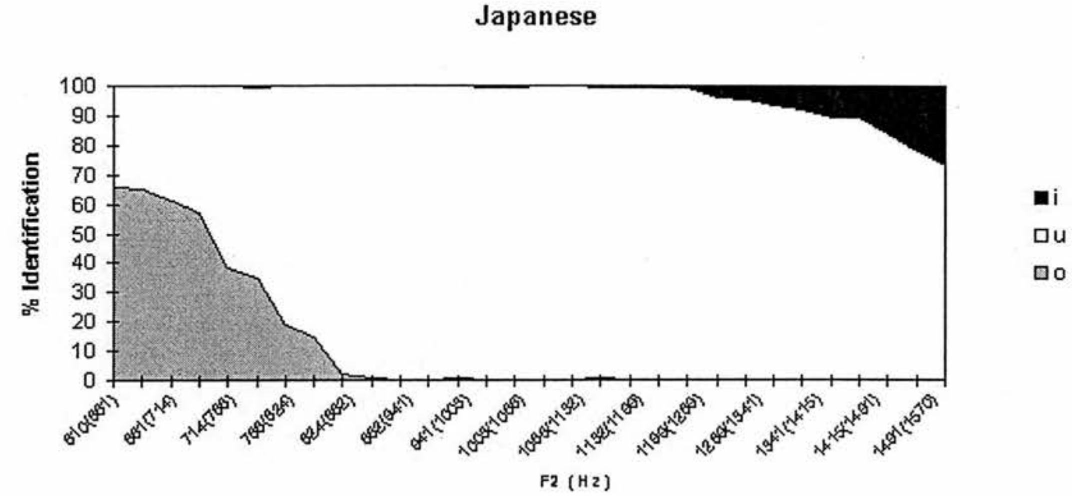


Figure 5:3. Japanese identification results (pooled).

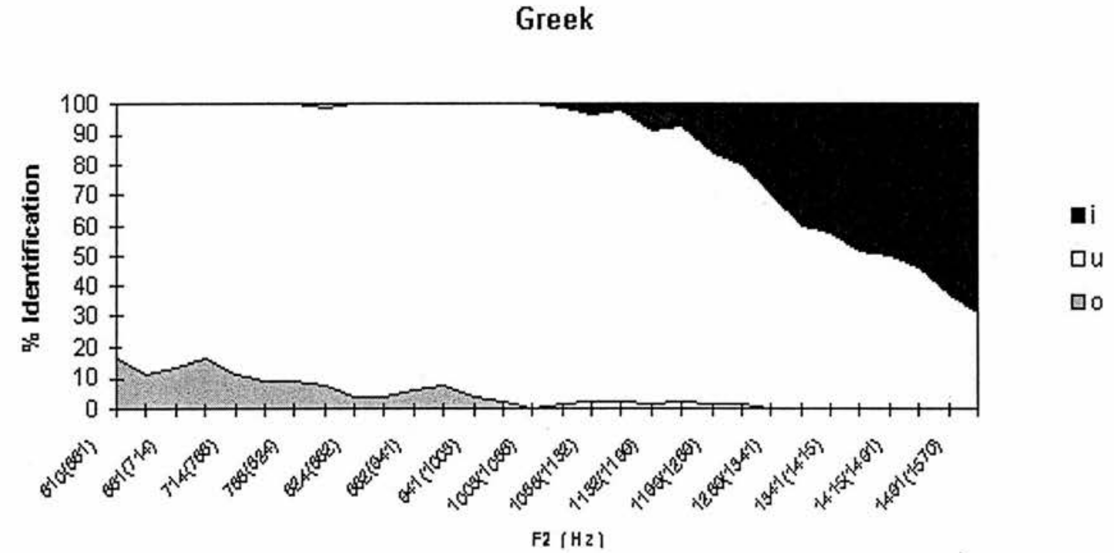


Figure 5:4. Greek identification results (pooled). Two anomalous subjects, Subjects G2 and G4 are excluded.

As can be seen, the subjects labelled most stimuli as /u/, some of the stimuli with the lowest F2 values as /o/, and those with the highest F2 values as /i/. In most cases, i.e., except for two clear exceptions, i.e., Subjects G2 and G4's identification performance, the percentages of /i/- and /o/-identification systematically increased towards high and low F2, respectively, i.e., to the right and left of the stimuli identified as /u/.

The observation that stimuli with higher F2 values were identified as /i/ is consistent with the finding that F2 correlates inversely with vowel backness. In addition, as can be seen in Figs. 5:3 ~ 5:4, as a whole, the percentage of /i/-identification was smaller for the Japanese subjects, suggesting that the perceptual category boundary between /i/ and /u/ may be at higher F2 values for the Japanese than for the Greek listeners, which is in keeping with the cross-linguistic difference observed between the Japanese and Greek in the production of /u/ (cf. 4.3).

It is less clear, however, why stimuli with low F2 values were identified as /o/, which is mainly observed for the Japanese subjects when the two anomalous subjects, i.e., Subjects G2 and G4, are discounted. As for the Japanese listeners, the average formant values of isolated Japanese /o/ reported in Imaishi et al. (1984: 92) suggest that Japanese /o/ is realised with lower F2 values than Japanese /u/ (see Table 5:4), which could have led to the identification of some stimuli with the lowest F2 values as /o/ by the Japanese subjects in spite of the relatively low F1 values of the stimulus sets.[53] With regard to Greek vowels, I have not found literature giving the formant values of isolated vowels, and therefore it is not possible to directly compare formant values of both Greek /u/ and /o/ with those of the stimulus sets in the present study. However, given that F2 of Greek /u/ obtained in Experiment 1a ranged from 520 Hz to 841 Hz (cf. Appendix C), it is not likely to be due to their F2 values that some Greek subjects perceived those stimuli with the lowest F2 values as /o/. Furthermore, F2 and F3 values of Greek /u/ in the word 'puse' and /o/ in 'pote' produced by male speakers in Jongman et al. (1989: 243-244) suggest that the two vowels differ mainly in F1 and do not differ significantly in F2 or F3 (see Table 5:5). Thus, there is no clear reason why those stimuli with low F2 should be perceived as /o/ by the Greek listeners in spite of their low F1 values. Perhaps, a comparison of individual vowel formants is not a satisfactory way of comparing perceptual and production data, as Lively & Pisoni (1997: 1672) and Frieda, et al. (in press) point out. According to Syrdal & Gopal (1986), for instance, vowel frontness and height are better captured in F3 - F2 and F1 - F0 bark differences, respectively, than independent F2 and F1 values. Another possibility is that the subjects felt that they should use all the labels given to them as options. In other words, it may be an artefact of experimental settings.

---

[53] Subject J10 identified the stimuli as /o/ the fewest times among the Japanese subjects, which may be due to the dialectal difference in the production of /u/ reported in (Shibatani, 1991: 160-161, 187). That is, unlike the Eastern dialect, /u/ is produced with a slight lip-rounding in the Western dialect, which Subject J10 speaks. However, his identification performance does not seem strikingly different form that of Subject J9 who is a speaker of standard Japanese, which derives from the Eastern dialect.

	/u/	/o/
F1	333 (297)	421
F2	1119 (1229)	708
F3	2352 (2223)	2527

Table 5:4. Formant values of isolated Japanese /u/ and /o/ produced by 10 male speakers of standard Japanese reported in Imaishi et al. (1984). The formant values of isolated /u/ produced by ten male speakers of Japanese in the presented study are also given in parentheses for comparison. All the values are in Hz.

	/u/	/o/
F1	339	476
F2	879	854
F3	2355	2366

Table 5:5. Formant values of Greek /u/ and /o/ produced by four male speakers in Jongman et al. (1989:244-245). All the values are in Hz.

As seen in Figs. 5:1a ~ 5:2c, for some subjects (Subjects J3, G1, G2, G3, G4), there were no or very few stimuli that were identified as /u/ at all times. As mentioned earlier, two subjects (Subjects G2 and G4) did not classify the stimuli on the basis of F2 in a systematic fashion as other subjects did. This may be due to a certain degree of unnaturalness that is inherent to synthesised steady-state vowel stimuli. The remaining subjects (Subjects J3, G1, G3) classified the stimuli into three groups depending on their F2 values in the same manner as others, i.e., they identified more stimuli with high F2 values as /i/ and more stimuli with low F2 values as /o/. However, they do not seem to have found most stimuli satisfactory tokens of /u/.

In an informal interview after the task, six out of the twenty subjects reported that some of the stimuli sounded like German /Y/ (Subjects J9, J10, G3, G5, G7 and G9), and one subject (Subject G10) reported that some stimuli sounded like French /y/, which are possibly the stimuli with F2 values that are too high for /u/ but not high enough for /i/. Six of them (Subjects J9, J10, G3, G5, G7 and G9) reported that they had labelled these foreign sounds /u/ most of the time, while one subject (Subject G10) reported that he labelled them /i/ half of the time and /u/ half of the time. In addition, Subject J2 reported that some stimuli had an /e/-like quality, which he labelled /i/, perhaps also referring to a similar range of stimuli as those heard as German /Y/ or French /y/ by some subjects. Altogether, almost half of the subjects reported to have heard sounds that were neither satisfactory /i/ or /u/ in the stimulus continuum. Considering that half of the subjects reported that they spoke French and/or German in addition to English and their L1, this is likely to be because of their knowledge of these languages. However, given Hawks et al.'s (1995: 243) report that Greek vowels are well separated in the perceptual vowel space, i.e., there is a gap between the two, it may be the case that the Japanese and Greek /i/ and /u/ categories may not be next to each other in the perceptual vowel space. Although there are no such studies on the Japanese perceptual vowel space to my knowledge, considering Stevens et al.'s (1969) study using speakers of American English suggesting that there may be a natural category between /i/ and /u/, it is conceivable that the Japanese /i/ and /u/ categories are also well separated in the perceptual vowel space.

No consistent contrast effect was found in the identification of the stimuli. Figs. 5:1a ~ 5:2c, where the identification of each stimulus in the context of its two neighbouring stimuli is shown separately for each context, indicate that the identification of stimuli near the category boundary was not necessarily affected by the context sound in the form of contrast. If the two stimuli near the boundary were contrasted, a stimulus would be labelled as a given phoneme when paired with a stimulus further away from the phoneme, and *not* as the given phoneme when paired with a stimulus closer to the phoneme. For instance, a stimulus near the /u/-/i/ category boundary would be identified as /u/ in the context of a more /i/-like sound, and as /i/ in the context of a more /u/-like sound, which would result in a saw-like category boundary, as seen in Subject G8's plot (Fig. 5:2c). Instead of contrast effects, assimilation effects (Kanamori et al., 1971) were observed at times, i.e., a stimulus near the category boundary was given the same label as the context stimulus which is further away from the boundary. For example, sometimes a stimulus near the /u/-/i/ category boundary was identified as /u/ in the context of a more /u/-like sound and as /i/ in the context of a more /i/-like sound, which resulted in step-like boundaries, as seen in Subject J8's plot (Fig. 5:1b). Repp et al. (1979: 139) report that contrast effects were pronounced in the identification task when the inter-stimulus interval was short (300 msec) but were negligible when the inter-stimulus interval was long (1,920 msec). However, Lotto et al. (1996: 8) report that there was a considerable amount of contrast effects in their identification task employing the same inter-stimulus interval as in the present study (1s) and a longer stimulus duration (500 ms) than the present study (85 ms). Putting all together, the amount of contrast effects may be correlated with the amount of auditory memory available in the identification task. In other words, the context effect observed in the present task was modest in comparison to

Lotto et al. (1996) perhaps because less auditory memory was available in the present task due to the short length of the stimuli (85ms).

As explained in 5.2, from the results of the identification task, a predicted hit rate for each D pair was calculated for each subject, based on which predicted miss rates and predicted d' scores were calculated. The predicted miss rates and d' are compared with the results of the discrimination task in the next section.

5.3.2 Results of the Discrimination Task (Experiment 2b)

As described in 5.2, the same Japanese and Greek subjects participated in a roving same/different discrimination task using the MOA stimuli corresponding to the high back area of the acoustic vowel space. The purpose of the experiment was to test whether the listener's discrimination sensitivity within the /u/ category decreased, at the level of phonetic coding, towards either of the two possible prototype locations, i.e., the language group's production average and individuals' category ideals (MOA choices), or extreme vowels. In order to facilitate the phonetic coding of the stimuli, short stimuli (85 ms) and long inter-stimulus intervals (1s) were employed.

As the focus of the present study was on the change in within-category discrimination sensitivity, the results were compared with those of the identification task (Experiment 2a) to determine which portion of the discrimination sensitivity curve reflected the subject's ability to discriminate between members of a single category /u/. In order to see to what extent change in discrimination sensitivity could be explained in terms of identification performance, the obtained sensitivity curves were also compared with the two kinds of predictions introduced earlier made on an assumption that discrimination sensitivity equals 0 when the two stimuli presented in a pair are given the same label. Each subject's discrimination sensitivity was computed using two metrics of discrimination sensitivity, i.e., miss rates and d', in order to assess Sussman & Lauckner-Morano's (1995) claim that d' is a better measure of the perceptual magnet effect. The two metrics were evaluated in terms of assumptions underlying each metric: 100% correct rejection rates underlying the use of miss rates as a measure of discrimination sensitivity, and a constant bias assumed by the use of d' as a measure of perceptual distance. Following Macmillan (1993), I use the term 'discrimination sensitivity' to refer to 'discrimination accuracy'. Although in Kuhl's Native Language Magnet Theory, it is assumed that perceptual distance is always reflected in discrimination sensitivity, the present data suggest that this may not be the case, as shown later in this chapter.

Over all, more than half of the subjects' discrimination sensitivities for the /u/ category were near chance level and little evidence for a single assimilation point within the /u/ category was found. Although some subjects' d' curves declined towards the extreme end of the /u/ category, this was true only for a minority of the subjects. At the same time, a close inspection of the data revealed that the assumptions underlying miss rates and d' were not met in the present data, leading to a methodological question regarding the use of the two measures as measures of perceptual distance. The only consistent effect found in the present task was a presentation order effect in the subjects' response bias, which suggests that there may be a directional asymmetry in the perceptual distance between vowels depending on their relative extremity (and possibly prototypicality) and the presentation order.

The remainder of this section is organised as follows: In 5.3.2.1 and 5.3.2.2, the results are analysed using the two metrics of discrimination sensitivity, miss rates and d', respectively, and checked against the assumptions underlying the two metrics. In 5.3.2.3, the results of the discrimination tasks are discussed, and the replications of Kuhl (1991) are evaluated in the light of the findings in the present study.

5.3.2.1 Miss rates and correct-rejection rates

In Figs. 5:5a ~ 5:6b, the miss rate obtained from each subject is plotted for each stimulus pair together with two kinds of predicted miss rates computed on the basis of the subject's identification performance obtained in Experiment 2a, using the formulas given in 5.1. Those subjects who labelled no or very few stimuli consistently as /u/ in the identification task (Subjects J3 and G1-G4) are eliminated from subsequent analyses. Subject G7 is also excluded from the analyses, as his 100%-/u/ identification area is dislocated from both his category ideal location (MOA choice) and the Greek production average. The ordinate in Figs. 5:5a ~ 5:6b represents the percentage of miss rates, while the abscissa represents the F2 values of the stimuli. Each data point is plotted halfway between the F2 values of the two stimuli to be discriminated. The solid line gives the actual miss rates obtained in the experiment, and the two broken lines give two kinds of predicted miss rates. As explained in 5.1, one kind of predicted miss rates gives more conservative estimates by taking account of the identification of paired stimuli as different phonemes within each trial only, whereas the other takes account of the identification of the stimuli across trials, allowing for the maximum possibility of discriminability arising from cross-category discrimination. The predicted miss rates are 100 % for pairs of stimuli both identified as the same vowel at all times, while they are lower than 100 % for pairs that were identified as different vowels at times, predicting lower miss rates (i.e., higher hit rates) for those pairs of stimuli. The average F2 values of each subject's MOA choices and the language group's productions are presented near the closest F2 values on the abscissa and are indicated by arrows.

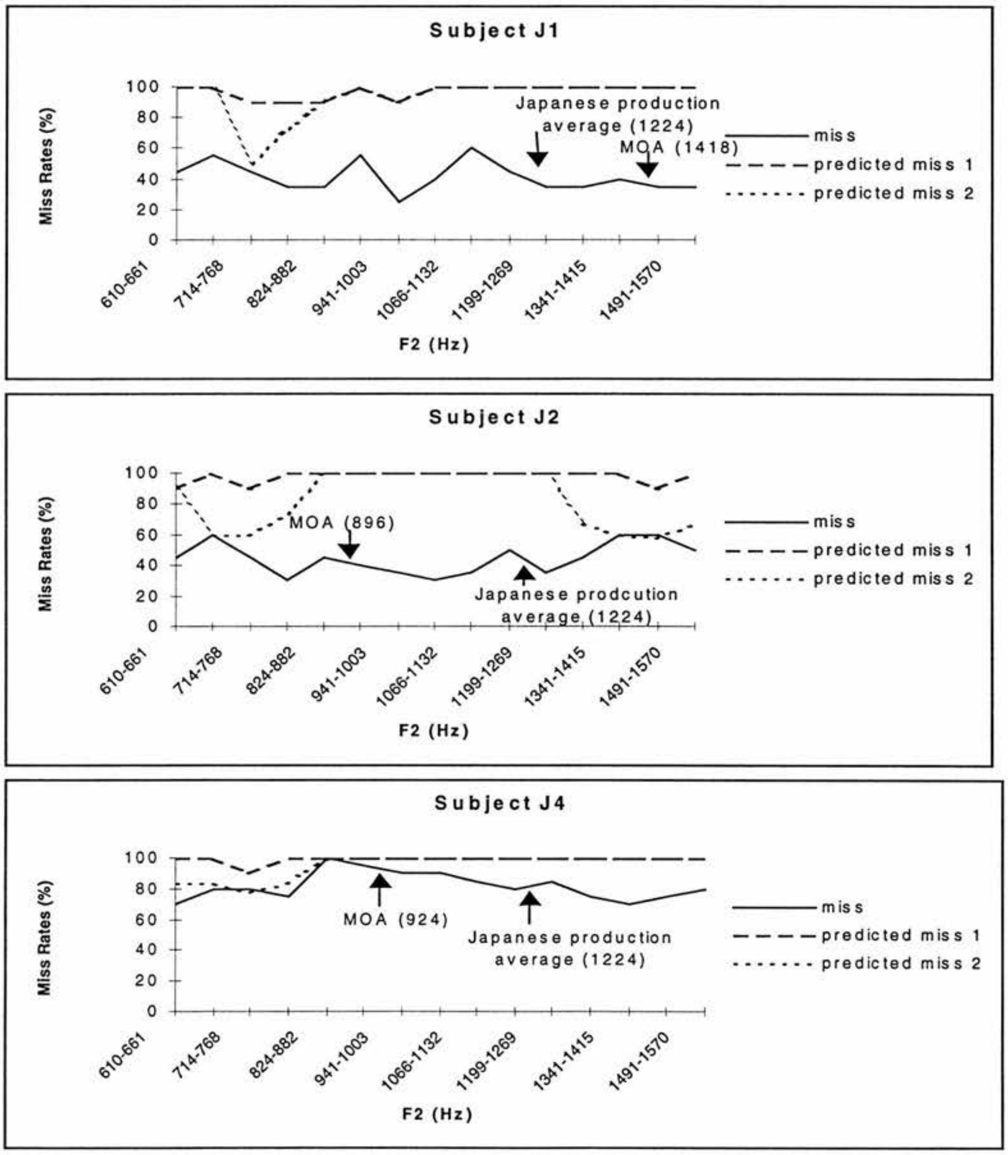


Figure 5:5a. Japanese subjects' obtained and predicted miss rates.

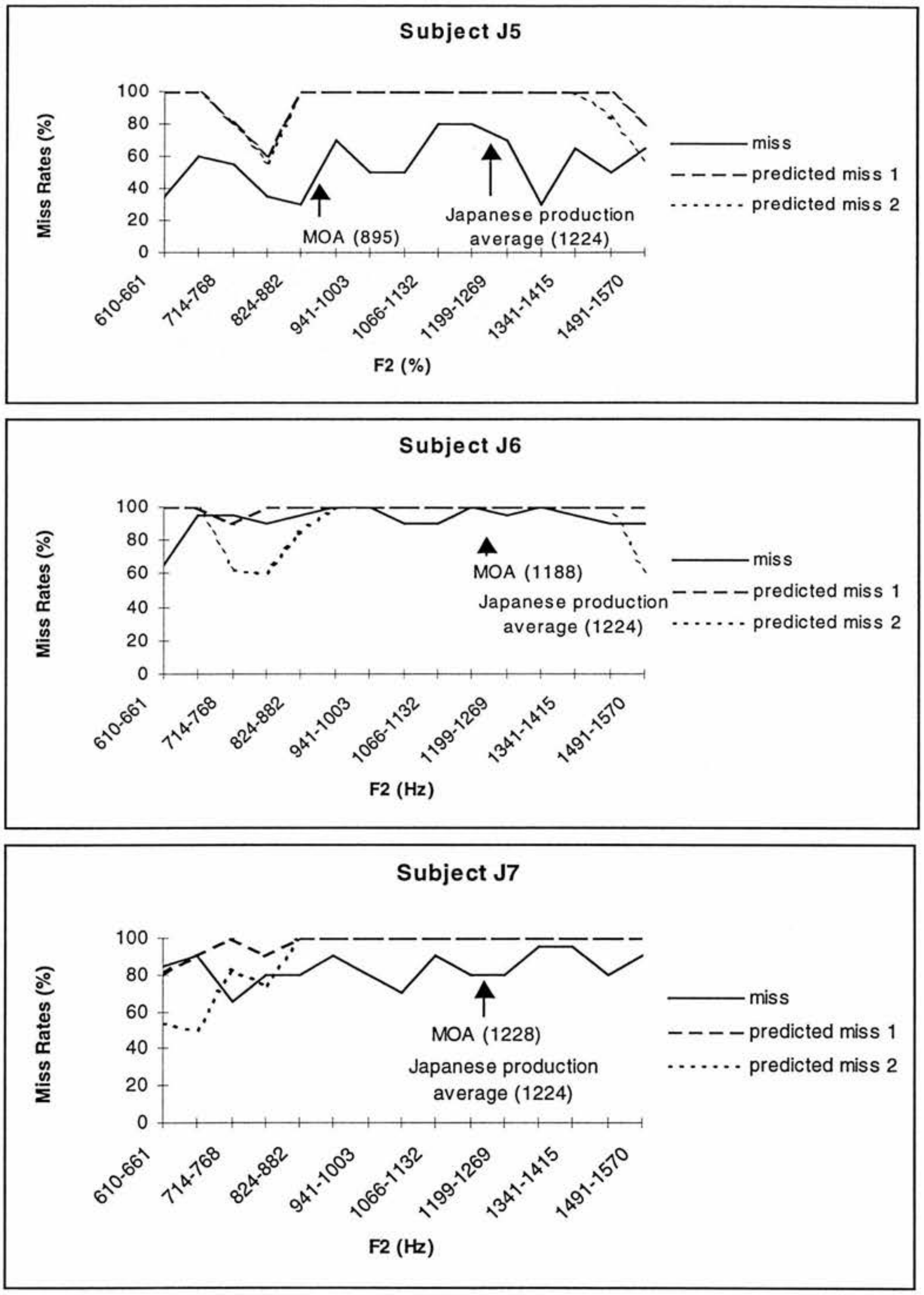
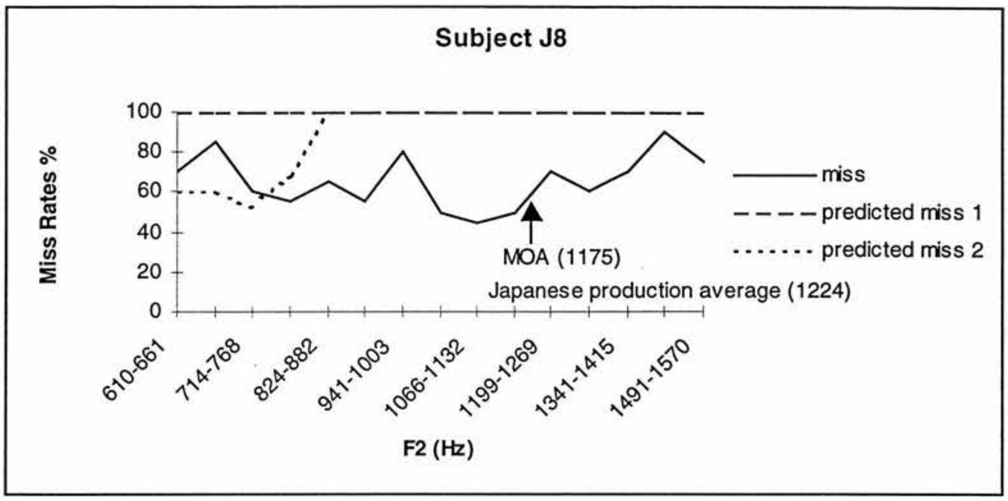
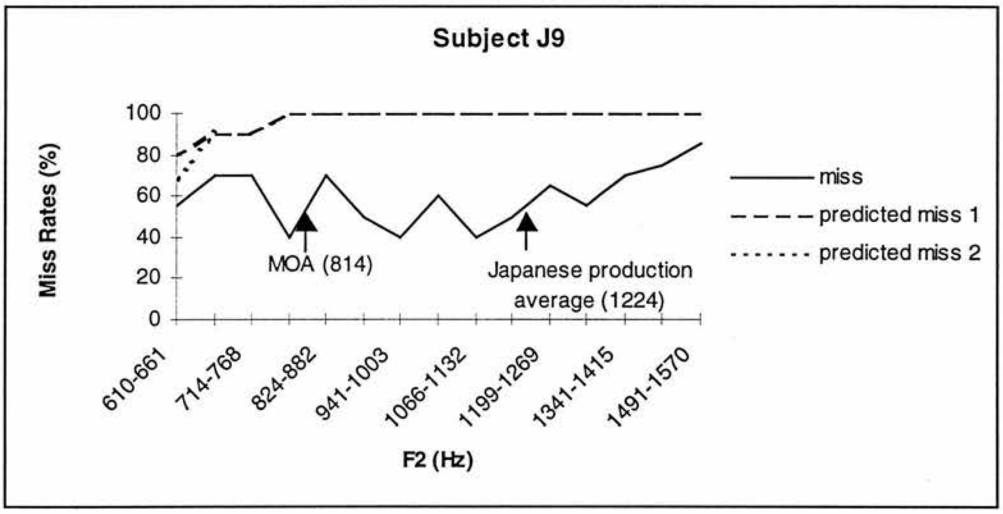


Figure 5:5b. Japanese subjects' obtained and predicted miss rates.





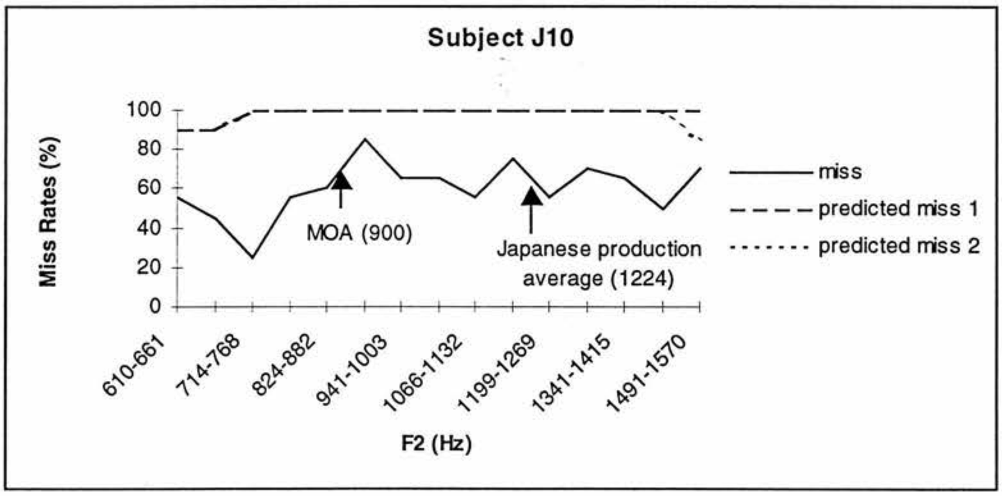
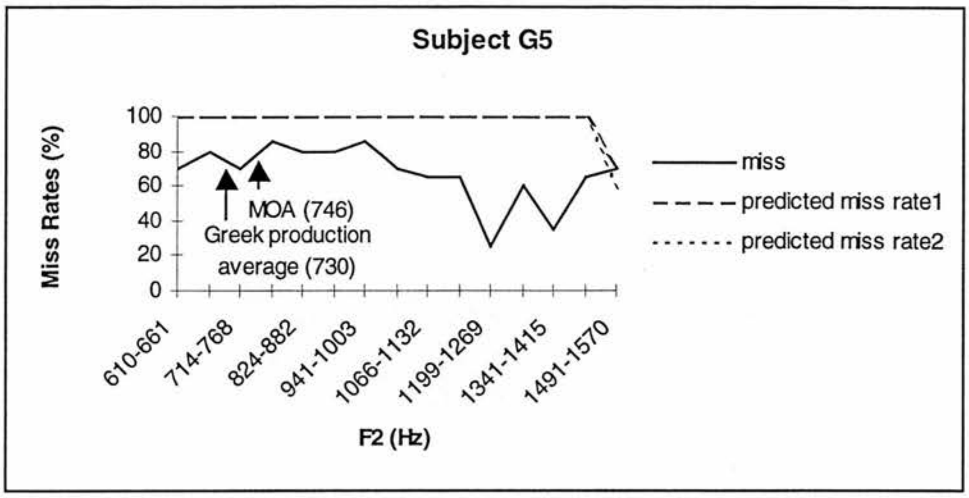
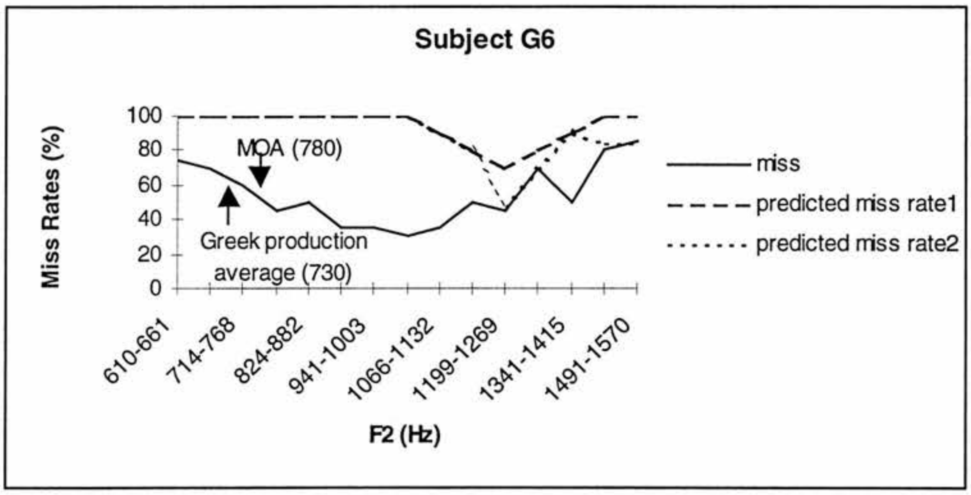


Figure 5:5c. Japanese subjects' obtained and predicted miss rates.





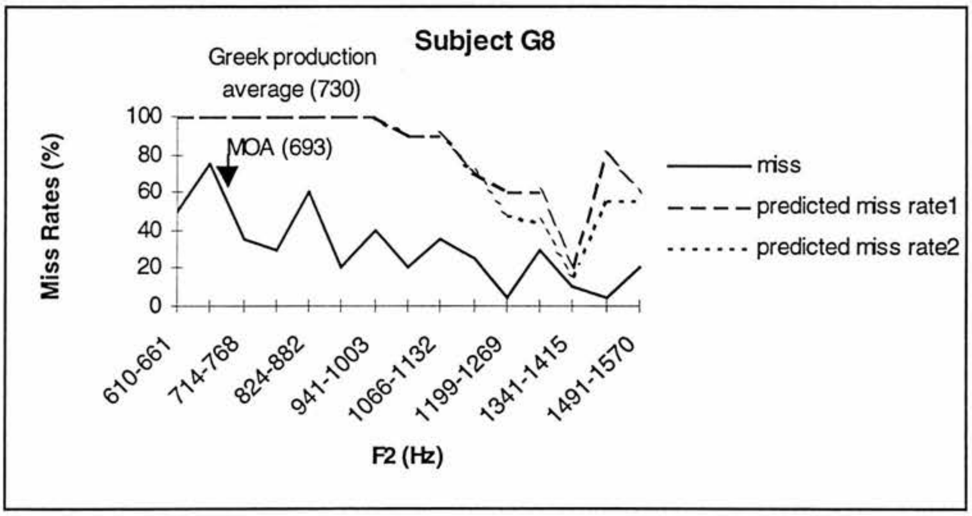
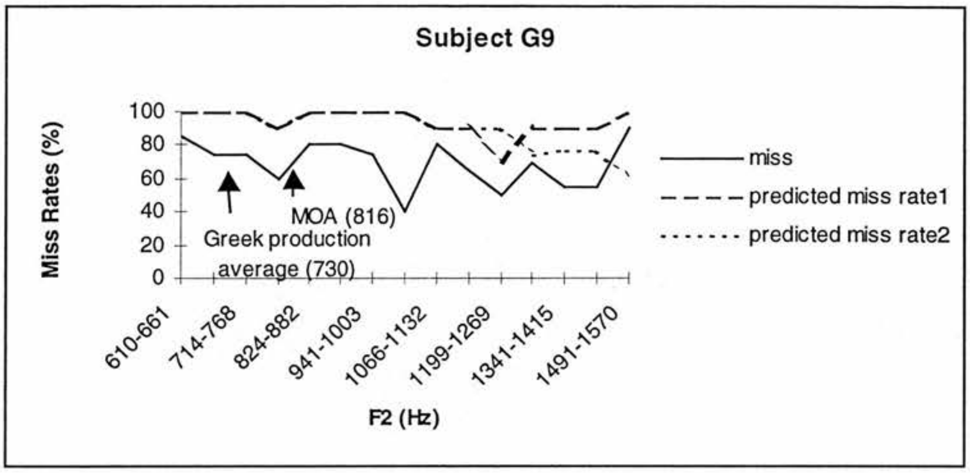


Figure 5:6a. Greek subjects' obtained and predicted miss rates.



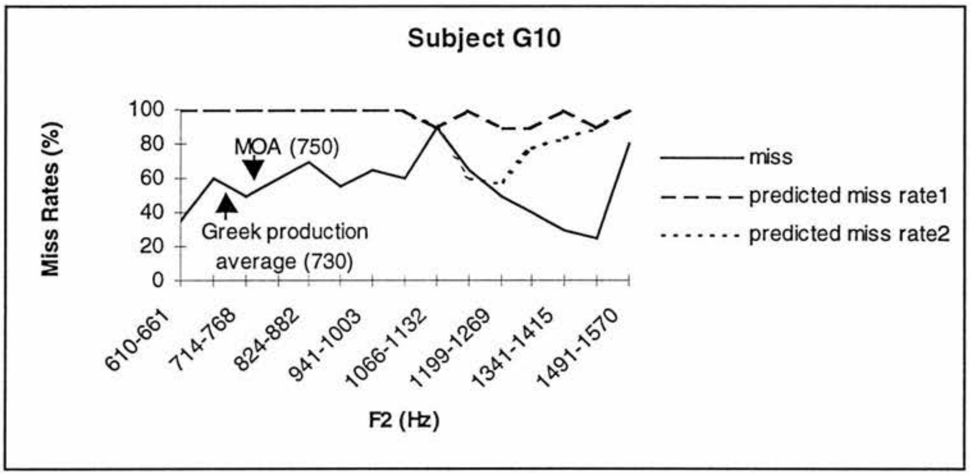


Figure 5:6b. Greek subjects' obtained and predicted miss rates.

As can be seen in Figs. 5:5a ~ 5:6b, the obtained miss rates are consistently lower than the predicted miss rates for most subjects; in other words, the subjects' discrimination sensitivities appear to be better than the predictions based on the identification of the stimuli, which is in keeping with the findings in the literature that vowel perception is more continuous than consonant perception (e.g., Stevens et al., 1969; Repp, 1979). However, as is shown below, the subject's correct-rejection rates ('same' responses to S pairs) are substantially lower than 100%, suggesting that some of the hits were correct by chance, which in turn implies that the obtained miss rates may be deflated in general.

As can be seen, no apparent systematic patterns can be found in the way the obtained miss rates fluctuate. Some of the dips in the miss rates, which indicate more hits, seem attributable to the subjects' identification of the stimuli in the pair as different vowels, given that they correspond fairly well to those in the predicted miss rates. However, it is rather difficult to explain those peaks and valleys found within the 100%-/u/ region, i.e., the portion of the solid line in Figs. 5:5a ~ 5:6b that corresponds to the portion in the predicted miss rates that reads 100%. As those peaks and valleys are observed in different locations for different subjects, they are not likely to have resulted from unequal spacing along the stimulus continuum.

For almost half of the subjects (Subjects J2, J5, J6, J9, J10 and G8), either of their MOA choices or the production averages, or both of them seem to be located near the local peaks in their miss rates, which suggest lower discrimination sensitivity, as Native Language Magnet Theory would predict. However, most of these peaks are small in size and/or not the only peak. Thus, the miss rates do not provide convincing evidence for the inner structure of the category where the prototype assimilates other category members. As for the effect of vowel extremity on discrimination sensitivity, some subjects' miss rates (Subjects J1, J4, J10, G6, G8) exhibit a general tendency for a decline from the most extreme end of the 100%-/u/ region (the portion of the solid line that corresponds to the portion in the predicted miss rates that reads 100%) towards less extreme stimuli, suggesting a correlation between poor discrimination sensitivity and vowel extremity, but this is true only for one third of the subjects.

Some of the valleys in the miss rates within the /u/ category may reflect a category boundary between /u/ and the foreign sound resembling German /Y/ or French /y/ some subjects reported to have heard after the identification task. As reported in 5.3.1, six out of seven subjects who reported to have heard a foreign sound also reported that they labelled the sound /u/ most of the time. However, those dips which may correspond to the possible extra boundaries cannot be identified from the present identification results, since only three vowels, i.e., /i/, /u/ and /o/, were given as choices in the identification task.

When each subject's correct-rejection rates are examined, however, the use of miss rates as a measure of discrimination sensitivity does not seem appropriate for the present data. In Figs. 5:7a ~ 5:8b each subject's miss and correct-rejection rates are compared. The ordinate stands for the percentages of miss rates (solid lines) and correct-rejection rates (broken lines), and the abscissa represents the F2 values of the stimuli to be discriminated. Each data point is plotted halfway between the F2 values of the two stimuli to be discriminated, and thus correct-rejection rates are the averages of those for two adjacent S pairs.

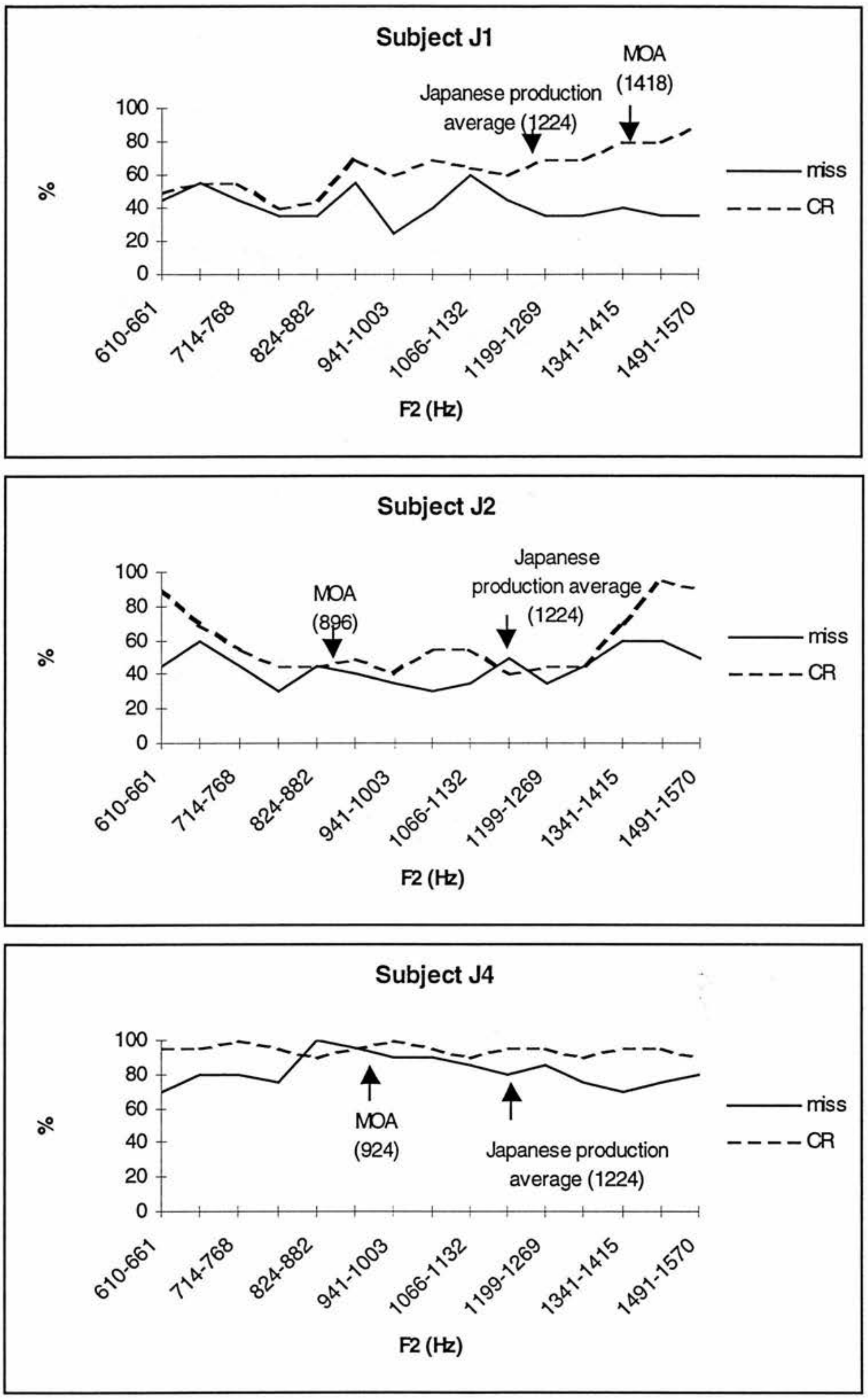
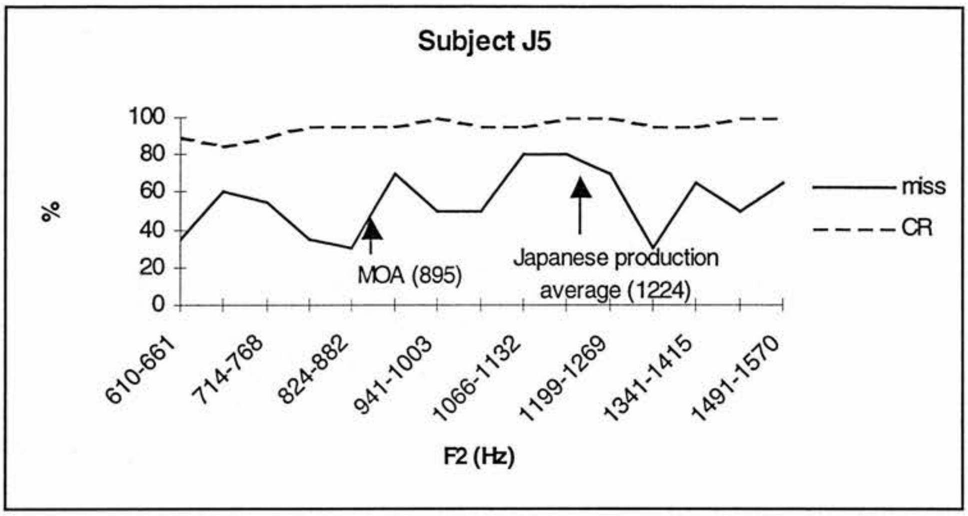
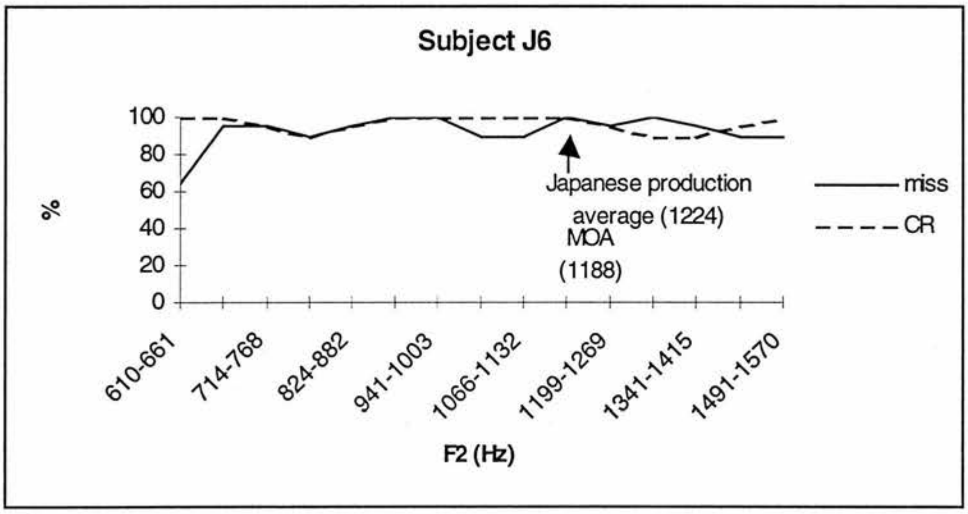


Figure 5:7a. Japanese subjects' miss and CR rates.





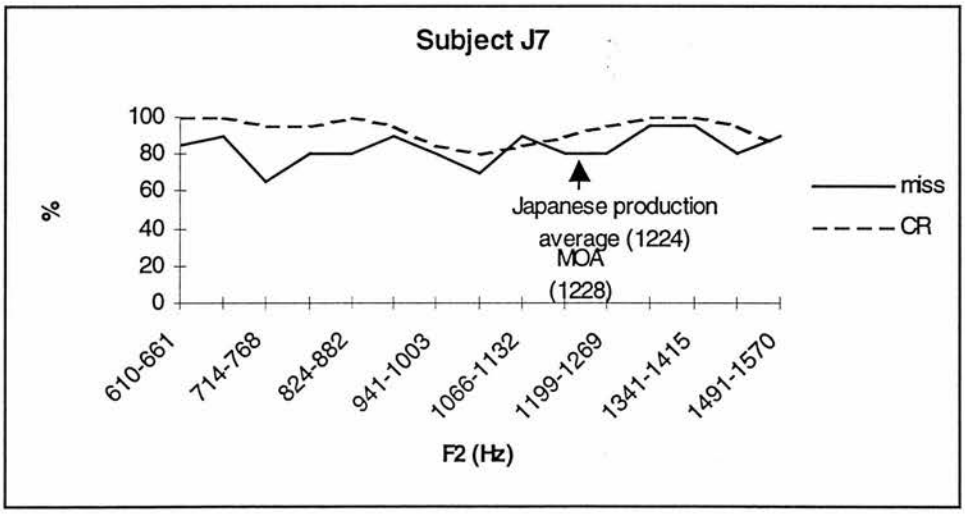
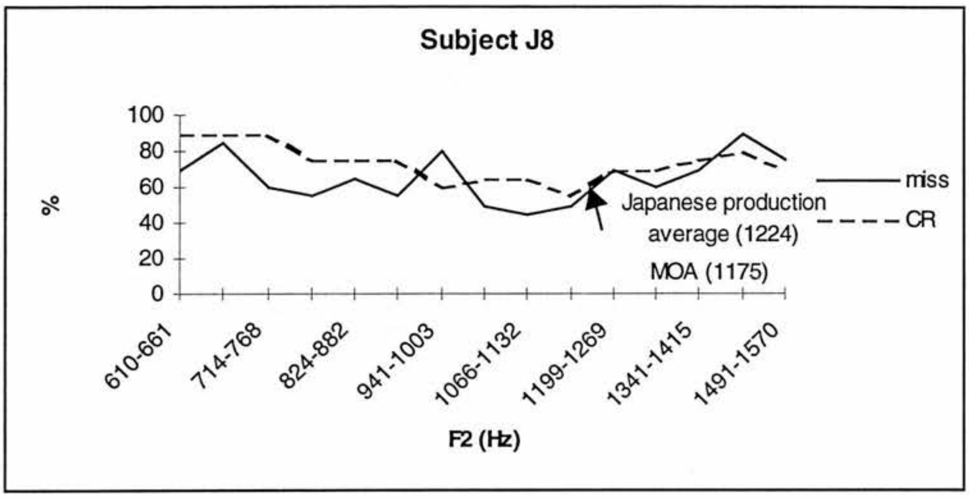
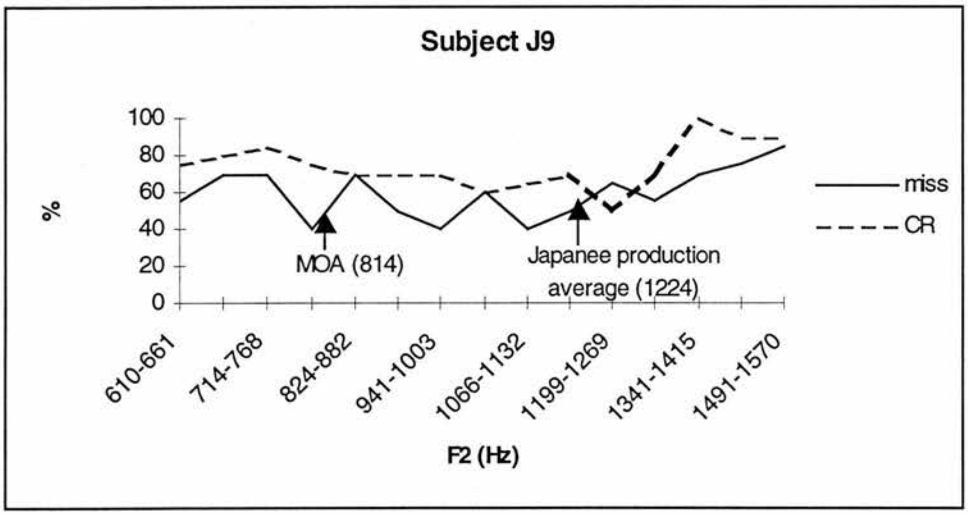


Figure 5:7b. Japanese subjects' miss and CR rates.





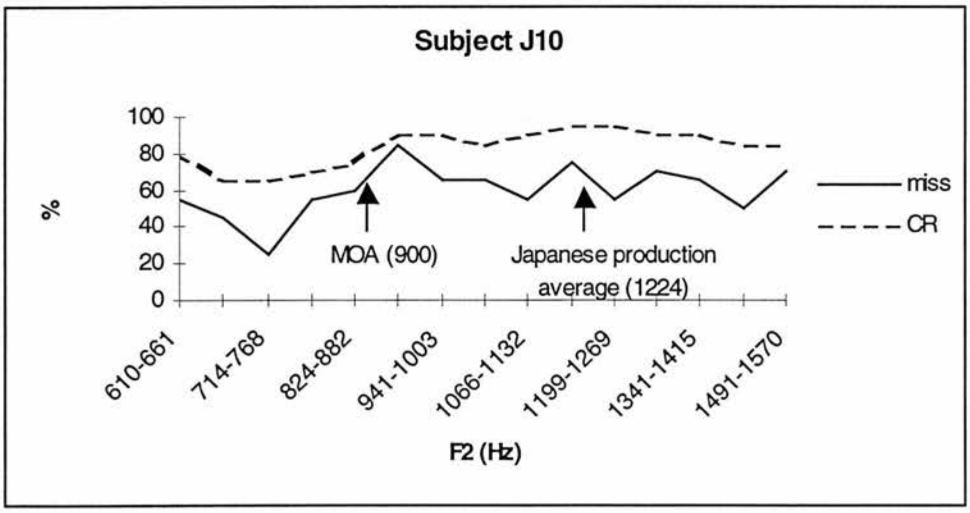


Figure 5:7c. Japanese subjects' miss and CR rates.

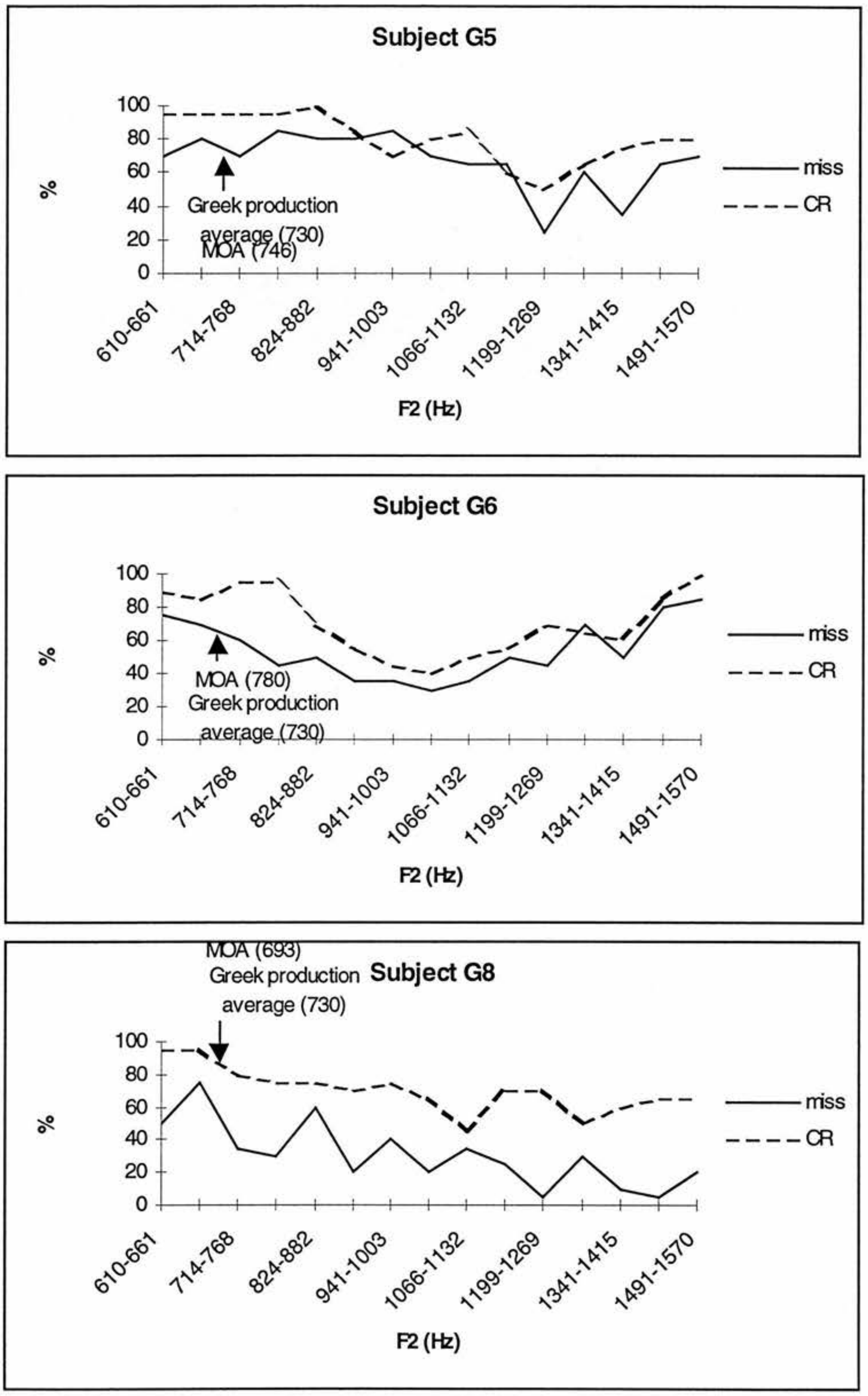
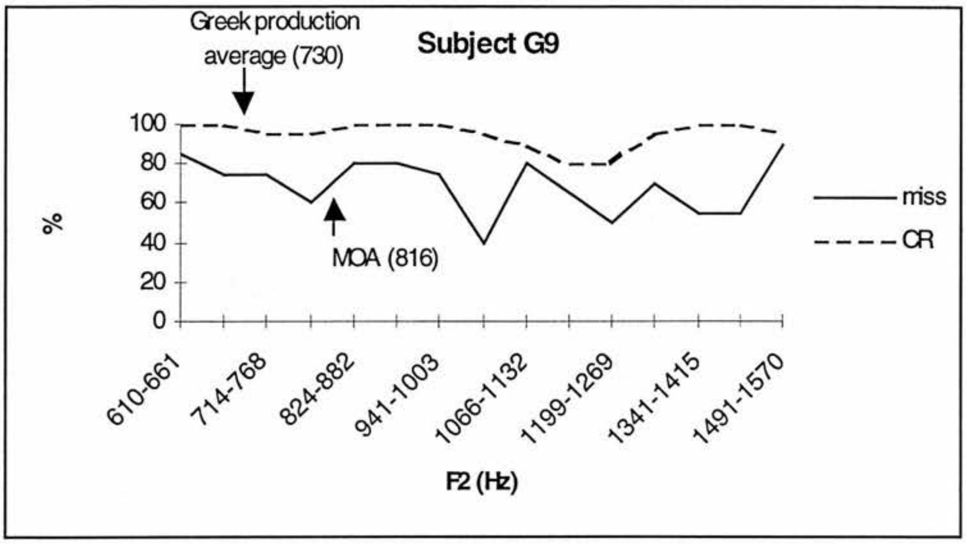


Figure 5:8a. Greek subjects' miss and CR rates.



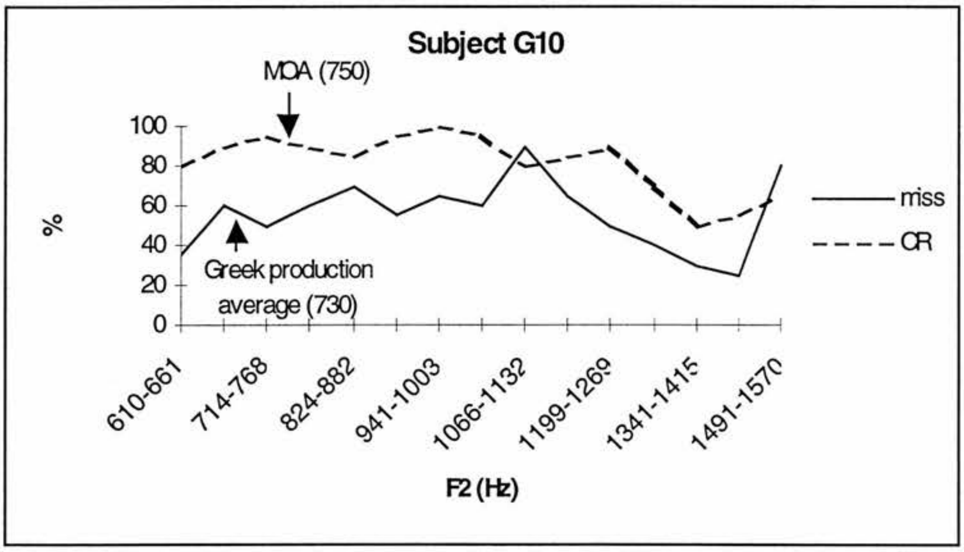
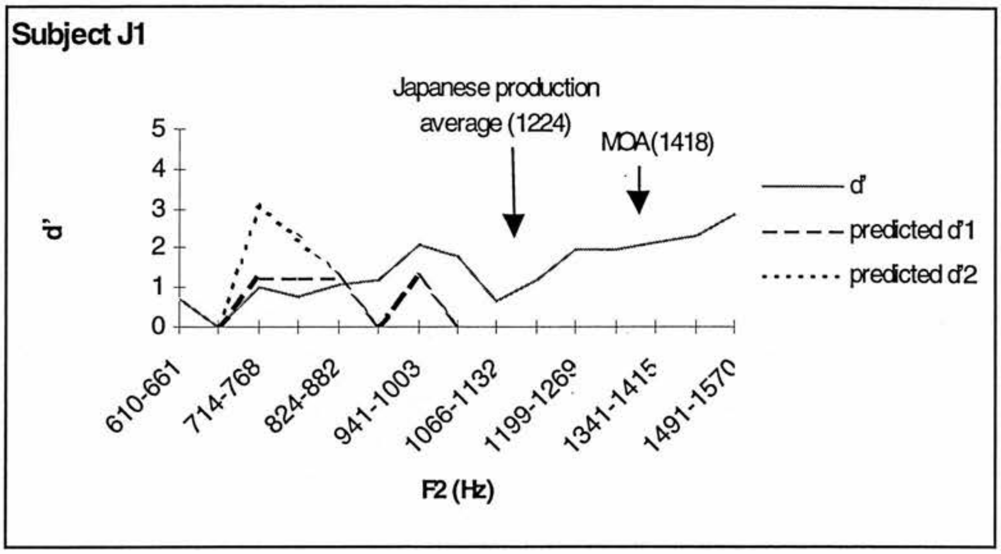


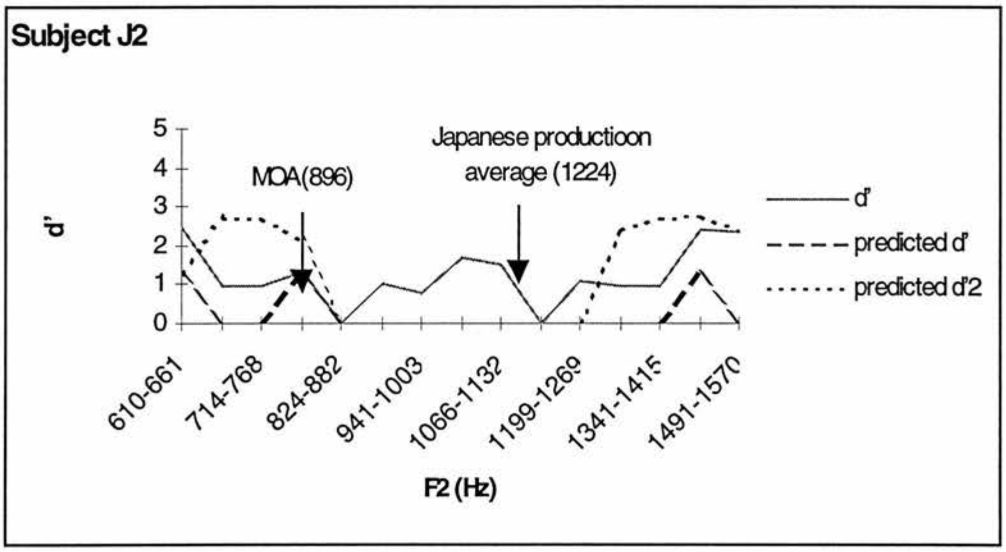
Figure 5:8b. Greek subjects' miss and CR rates.

As can be seen in Figs. 5:7a ~ 5:8b, the correct-rejection rates are almost always lower than 100% for all the subjects, which explains at least partially why the obtained miss rates are lower than predicted. If correct-rejection rates are lower than 100%, false-alarm rates ('different' responses to S pairs) are above 0%, provided the equation: (False-alarm rate) = 1 - (correct-rejection rate). The fact that false-alarm rates are above 0 % suggests that some hits are correct by chance, taking false alarms as evidence for guessing. As a consequence, the obtained miss rates are likely to be deflated. The more the subject is willing to guess, the smaller the miss rates will be, regardless of sensitivity. Moreover, correct-rejection rates vary considerably (ca. 30 ~ 50 %) along the stimulus continuum for nearly two thirds of the subjects (Subjects J1, J2, J8, J9, G5, G6, G8, G10), with some of the valleys in the miss rates (higher hit rates) corresponding to the valleys in the correct-rejection rates (higher false-alarm rates), which implies that the miss rates may have been deflated to different degrees at different points along the stimulus continuum. In other words, lower miss rates in the areas corresponding to lower correct-rejection rates are not necessarily indicative of better sensitivity, as the lower miss rates may be partially due to a greater tendency of the subject to guess, or a greater bias towards the 'different' response, which could have led to more hits. Thus, miss rates do not seem to serve a reliable measure of discrimination sensitivity for the present data.

5.3.2.2 d' and response bias

Changes in each subject's discrimination sensitivity are also examined using d', a bias free measure of discrimination sensitivity. Again, Subjects J3, G1-G4 and G7 are eliminated from the analyses. As explained in 5.1, d' separates subjects' discrimination sensitivities from their response biases (tendencies to select a certain answer over the other) by discounting the portion of hit rates regarded as correct by chance. Each subject's d' curve is plotted in Figs. 5:9a ~ 5:10b, where the ordinate stands for d' scores and the abscissa indicates the stimuli's F2. Again, each data point is plotted halfway between the locations of the two stimuli to be discriminated. Grey solid lines represent obtained d' scores and dark broken lines represent two kinds of predicted d' scores based on the subject's identification performance in Experiment 2a. As explained in 5.1, one kind of predicted d' gives a more conservative estimate, and the other gives a less conservative estimate (for full details, see 5.1). The predicted d' reads 0 where the paired stimuli were both given the same label at all times, while it has positive values where they were given different labels, predicting better discrimination for pairs labelled as different phonemes. As mentioned earlier, to the extent that the improvement in obtained d' scores falls between the two estimates, the better discrimination can be regarded to have arisen from cross-category discrimination. As in previous figures, the average F2 values of each subject's MOA choices and the language group's productions are presented near the closest F2 values and indicated by arrows.





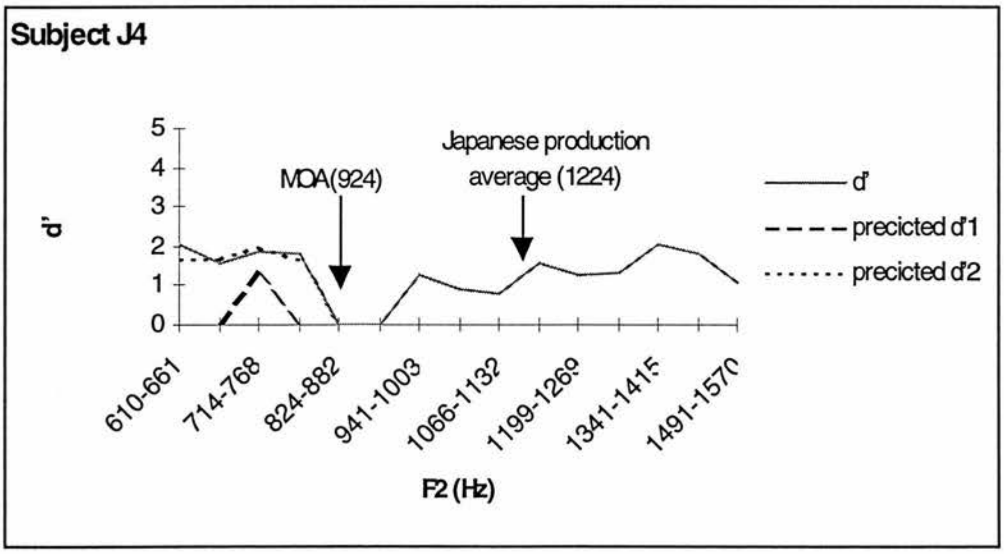
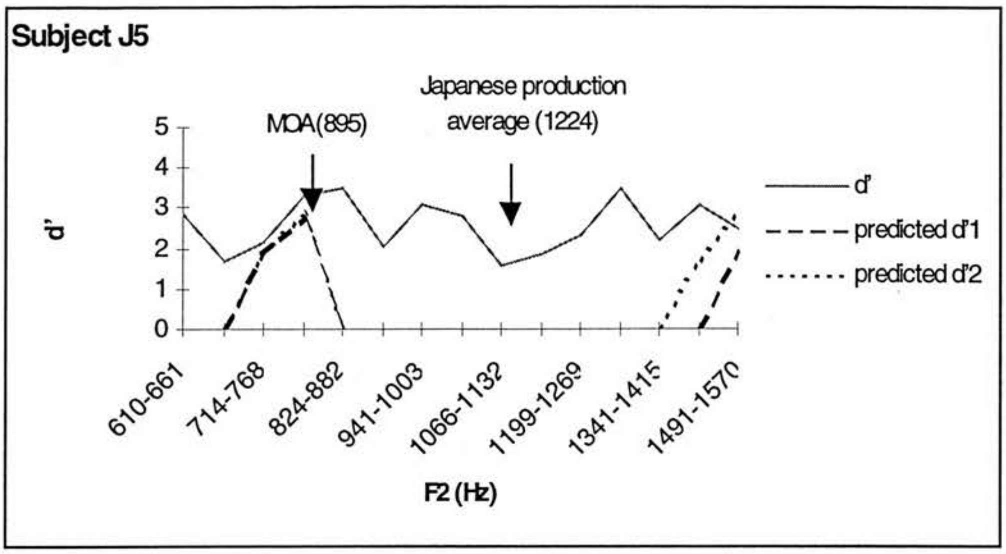
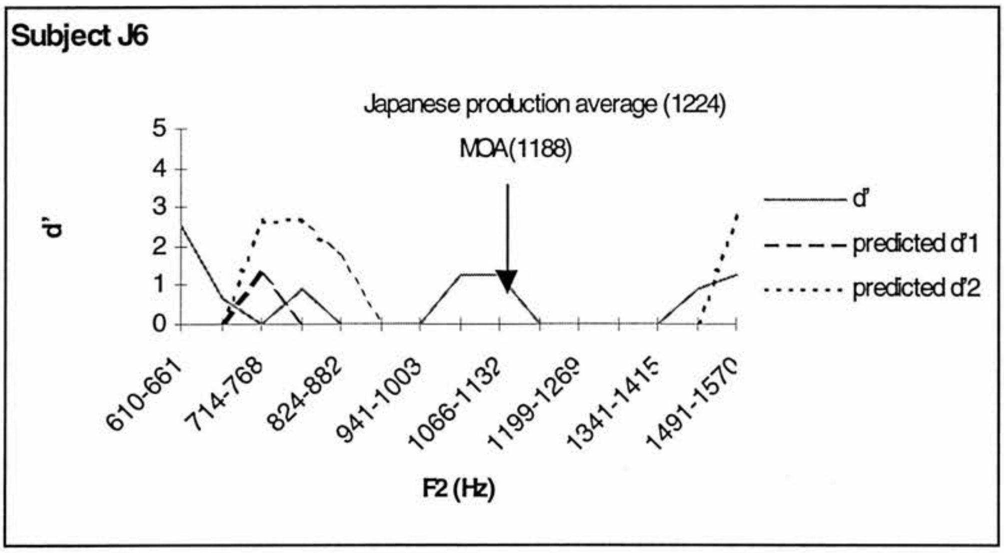


Figure 5:9a. Obtained and predicted d' (Japanese subjects).





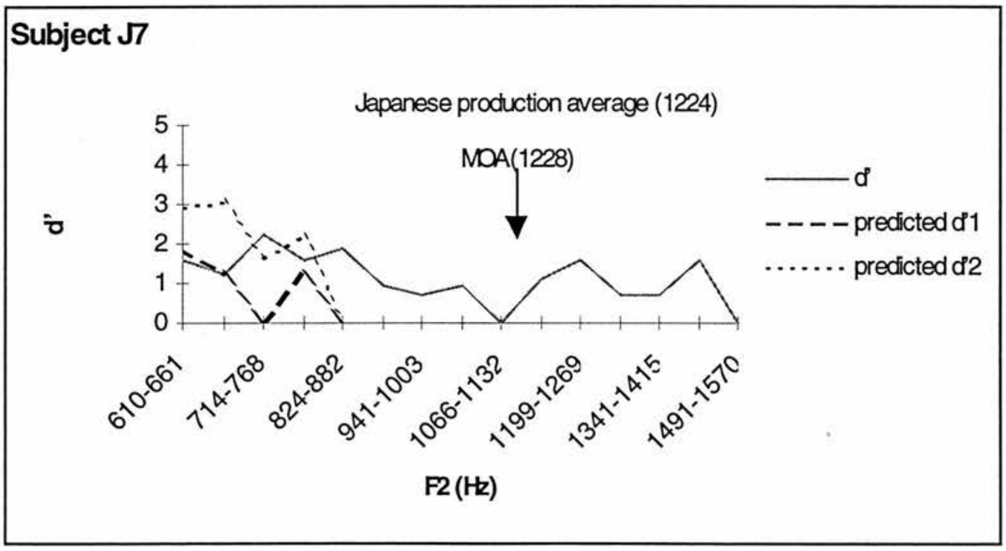


Figure 5:9b. Obtained and predicted d' (Japanese subjects).

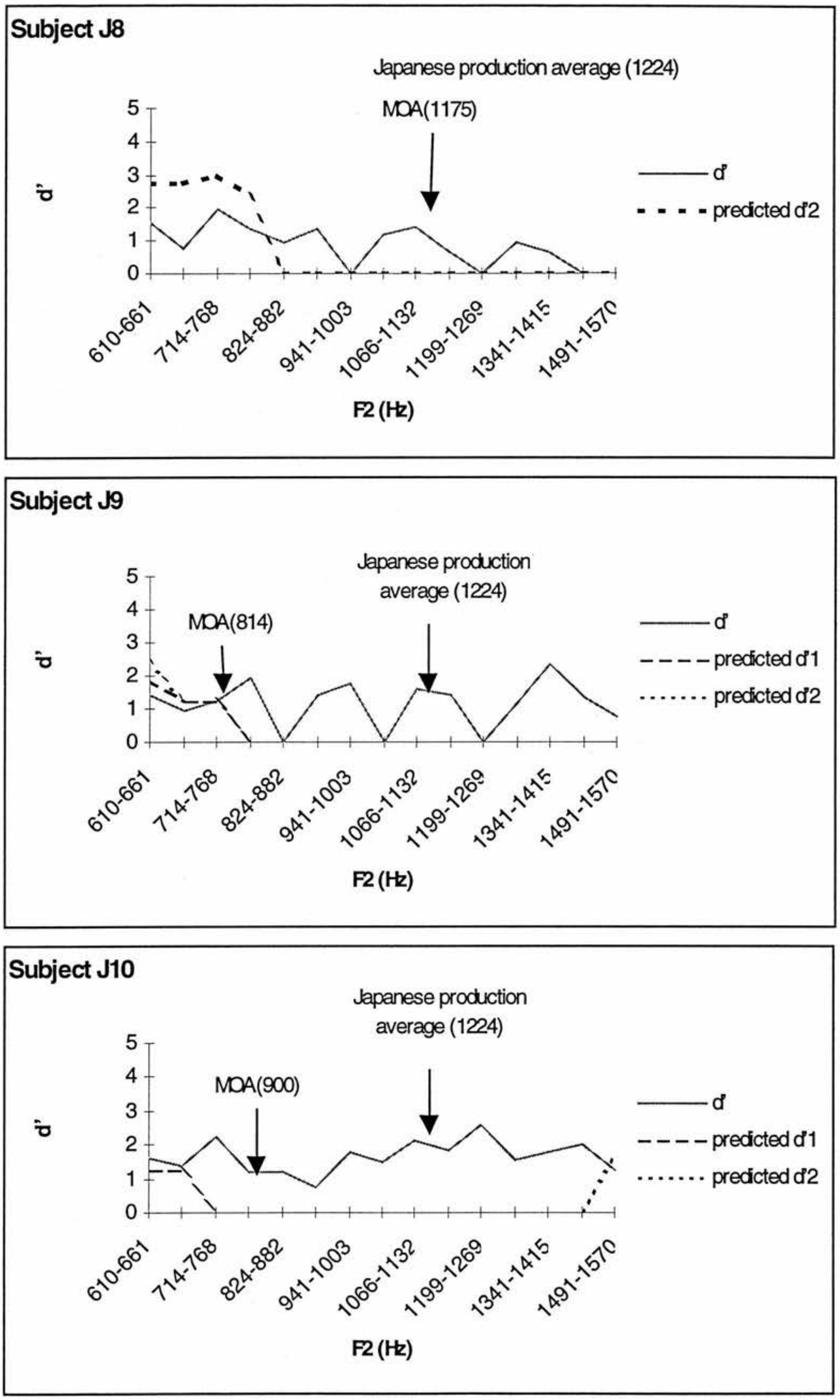


Figure 5:9c. Obtained and predicted d' (Japanese subjects).

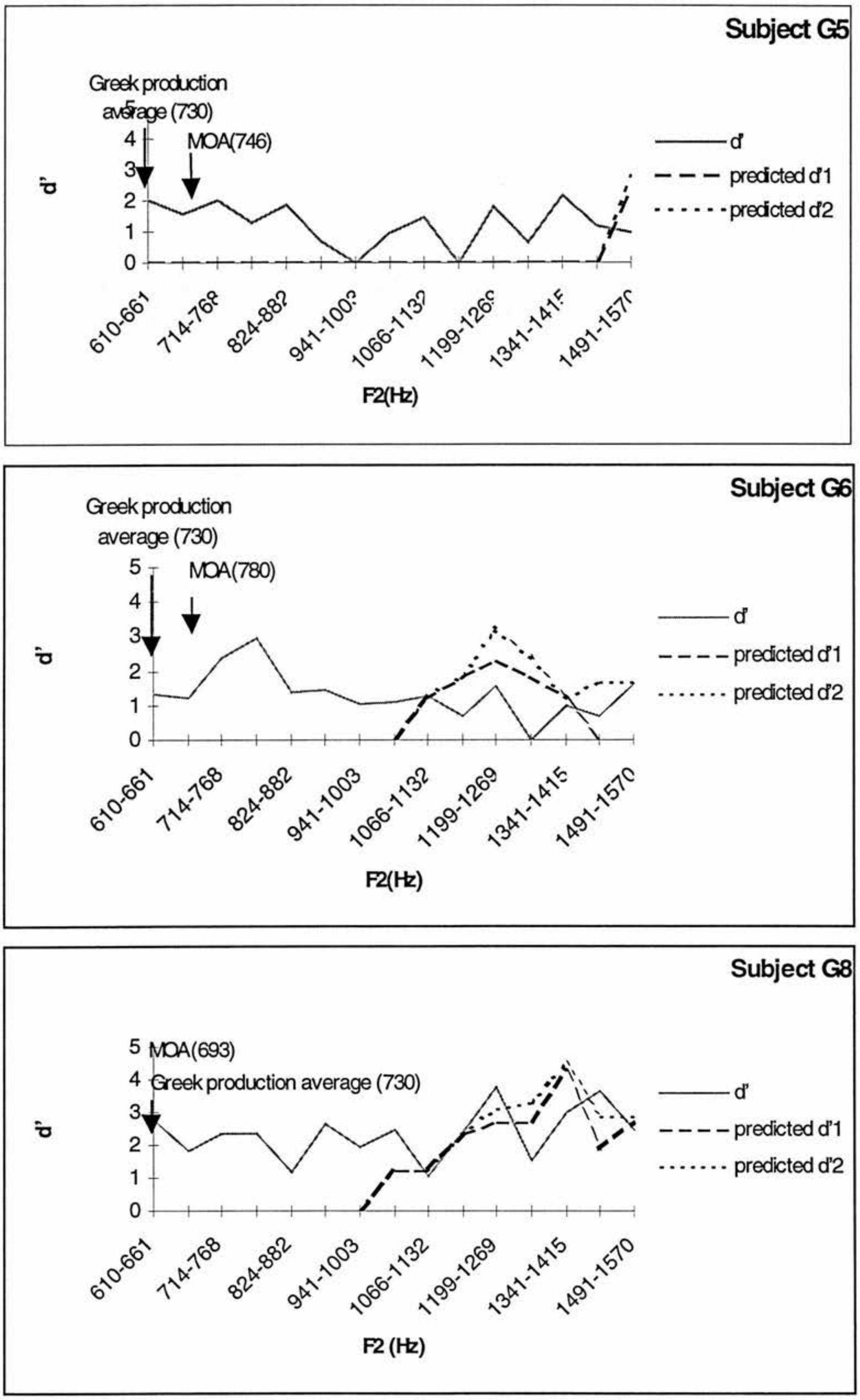
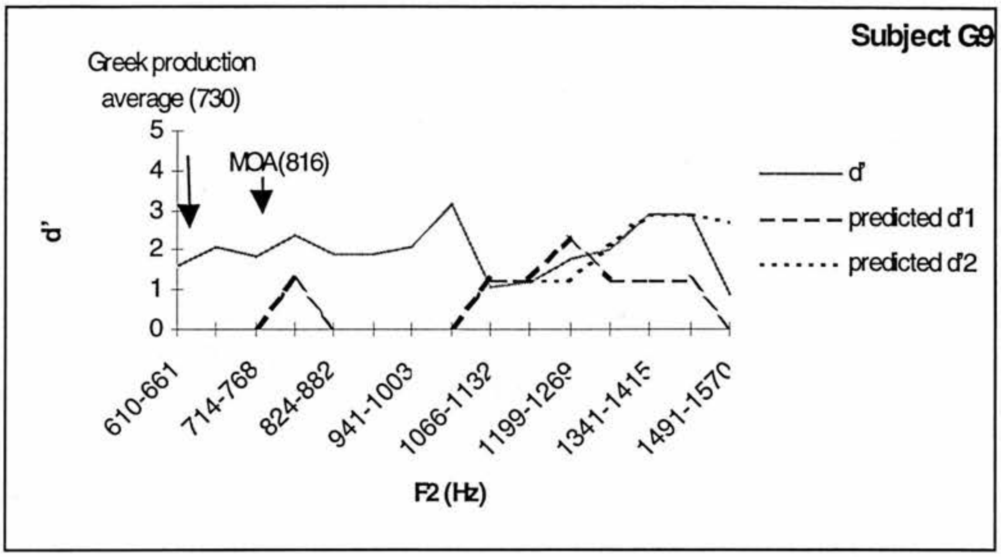


Figure 5:10a. Obtained and predicted d' (Greek subjects).



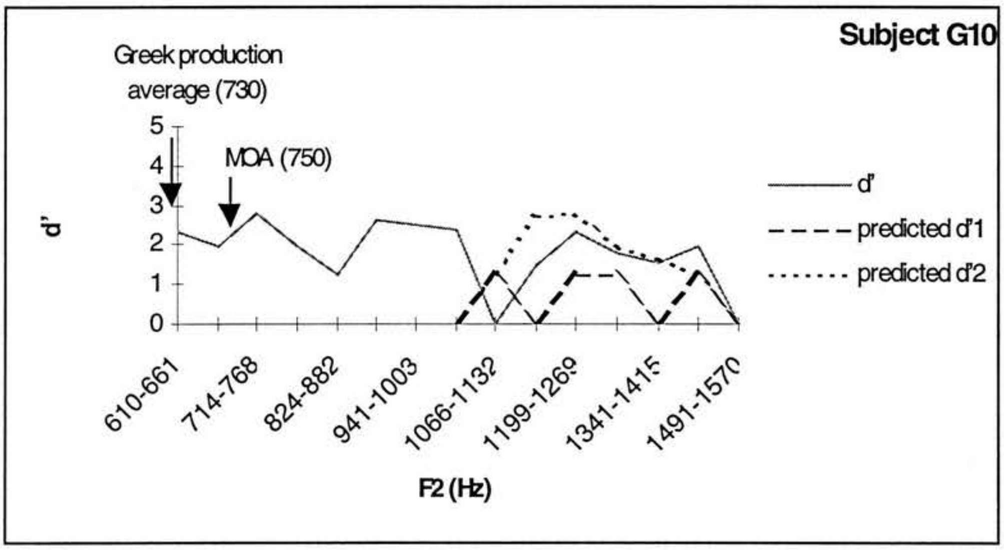


Figure 5:10b. Obtained and predicted d' (Greek subjects).

As can be seen in Figs. 5:9a ~ 5:10b, although the obtained d' has peaks that match those in the predicted d' fairly well for half of the subjects (Subjects J1, J4, J7, J9, G8, G9 and G10), it seems to fluctuate elsewhere in an even more random fashion than in the obtained miss rates. About two thirds of the subjects' d' curves have more than one dip indicating low discrimination sensitivity within the 100%-/u/ area, some of which do not match any of the possible locations towards which discrimination sensitivity was thought to decline (cf. Subjects J7, J9, G5, G6, G8 and G10's d'). Thus, here again, no convincing evidence for a single assimilation point for the category is found. But why the greater ruggedness?

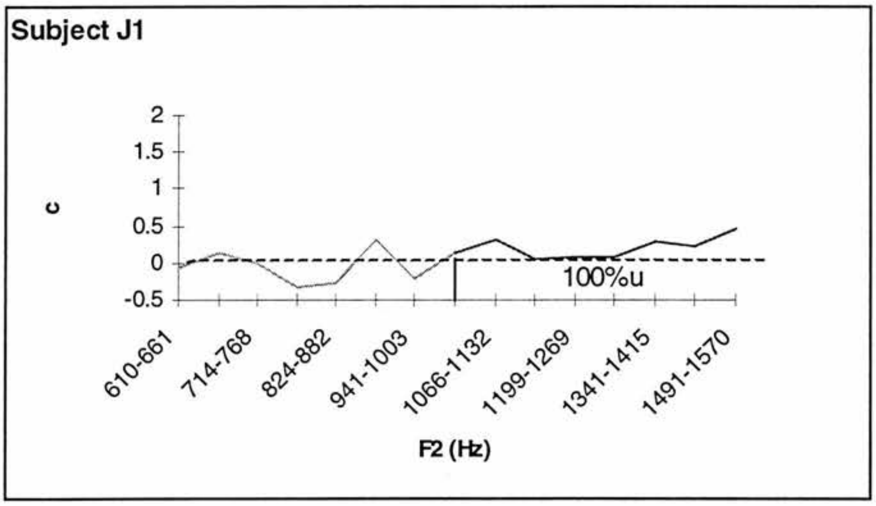
A comparison of Figs. 5:7a ~ 5:8b and Figs. 5:9a ~ 5:10b reveals that d' is basically upside-down projections of miss rates, although local peaks and valleys in miss rates seem somewhat amplified in d', while more global shifts in the miss rates tend to be lost in d', which makes the local peaks/valleys in the miss rates seem even more exaggerated in d'. For instance, the local peaks in Subject J2's miss rates (Fig. 5:7a) in the mid-range of the stimulus continuum that do not appear to be more than random fluctuations constitute substantial dips in d' (also compare Subjects J4 and J6's miss rates and d' for amplified local peaks/valleys). Furthermore, Subject G10's miss rates (Fig. 5:8b) gradually increase from the smallest F2 values towards the mid-range and then decrease, but no such pattern can be found in his d' (also compare Subjects G5, G6, and G8's miss rates and d' for lost global patterns).

According to Macmillan et al. (1988: 1269), the amplification of random variations in performance can take place when the performance is near chance level. They state that d' increases rapidly as performance edges above chance (H = F), so that even small random variation of observed proportions around values near chance can easily yield a substantial d'. Indeed, Figs. 5:7a ~ 5:8b indicate that for almost two thirds of the subjects (Subjects J2, J4, J6, J7, J8, J9, G5 and G6) the miss rates [1 - (hit rates)] and correct-rejection rates [1- (false alarm rates)] are rather close together, i.e., only about half of their 'same' responses are correct, in most parts of the stimulus continuum. That is, their performance was close to chance level along most of the stimulus continuum. Thus, the greater ruggedness observed in these subjects' d' in comparison to their miss rates appear to be largely due to their chance-level performance.

As for the rest of the subjects (Subjects J1, J5, J10, G8, G9 and G10) whose miss and correct-rejection rates are not as close as the rest, half of the subjects' d' (Subjects J1, J10 and G9) exhibits a decline towards the extreme end of the category. However, such a decline is not observed in the remaining subjects' d' (Subjects J5, G8 and G10). More subjects' d' may have declined towards the extremity of the vowel space, however, if a different formula had been used for converting mels into Hz when determining the spacing between the stimuli. As explained in Chap. 4, the present study used Fant's (1973) formula to create the spacing. Fant's (1973:48) figure showing the relationship between the mel scale and the mel approximation derived using the above formula suggests that the step-size used in the present experiment could have systematically increased, in psychoacoustic terms, towards the extreme end of the stimulus continuum.

The observation that global patterns found in miss rates are lost in some subjects' d' suggests that these subjects' miss rates and correct-rejection rates shifted in parallel, namely, hit and false-alarm rates shifted in the same direction, given the equation: $d' = v\,[z(H) - z(F)]$. This, in turn, suggests shifts in response bias, provided that response bias is expressed as c = -.5 [z(H) + z(F)]. As mentioned in Chap. 2, shifts in response bias along the stimulus continuum violate d's assumption of a constant bias. Thus, each subject's response bias along the stimulus continuum was examined.

The basic bias measure c for Signal Detection Theory was computed using the formula given in 5.2. Figs. 5:11a ~ 5:12b plot each subject's c, where the dark portion of the solid line represents response biases for stimulus pairs that were identified as /u/ at all times by each subject, and the grey portions represent response biases for stimuli that were not identified as /u/ at all times. C scores are given along the ordinate, and the F2 values of the stimulus pairs were given along the abscissa. To reiterate, negative c values arise when the false-alarm rate exceeds the miss rate, indicating that the subject was more inclined to respond 'different'. Positive c values arise when the false-alarm rate is lower than the miss rate, indicating that the subject was more inclined to respond 'same'.



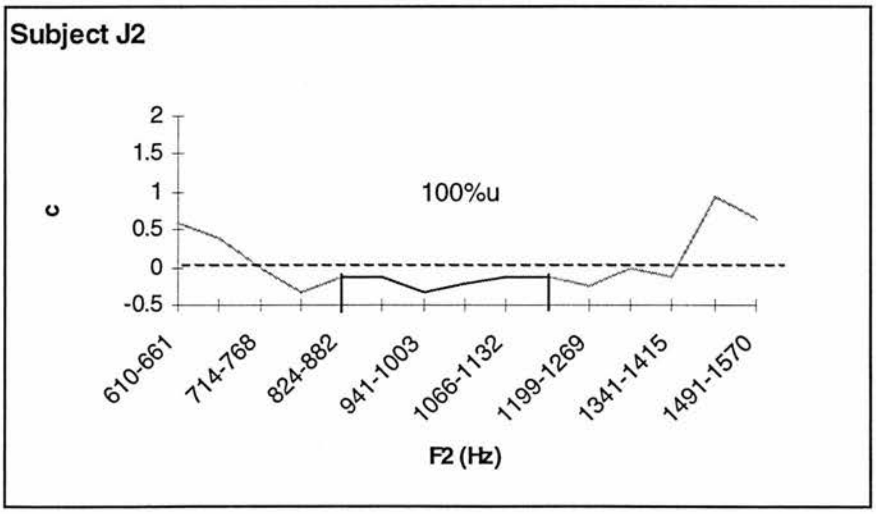
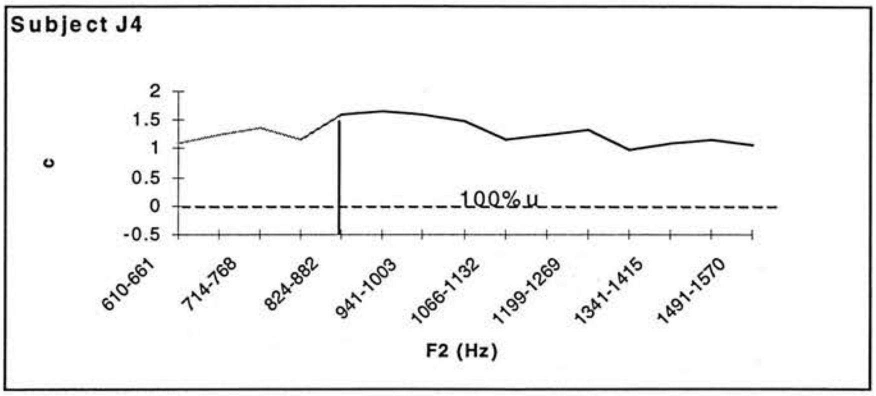
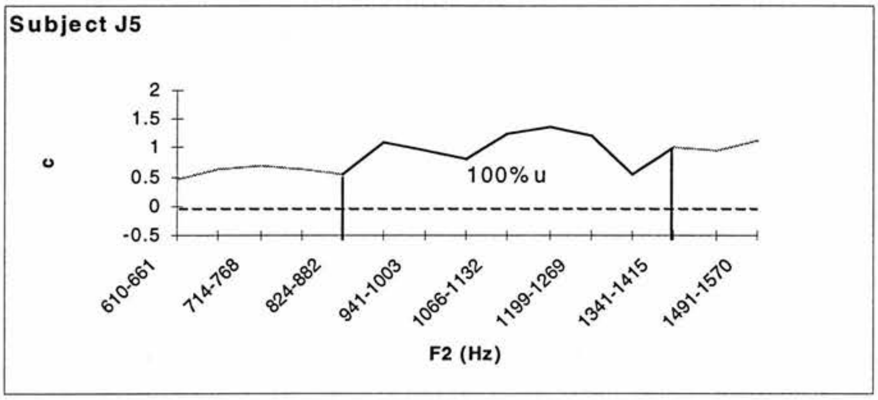
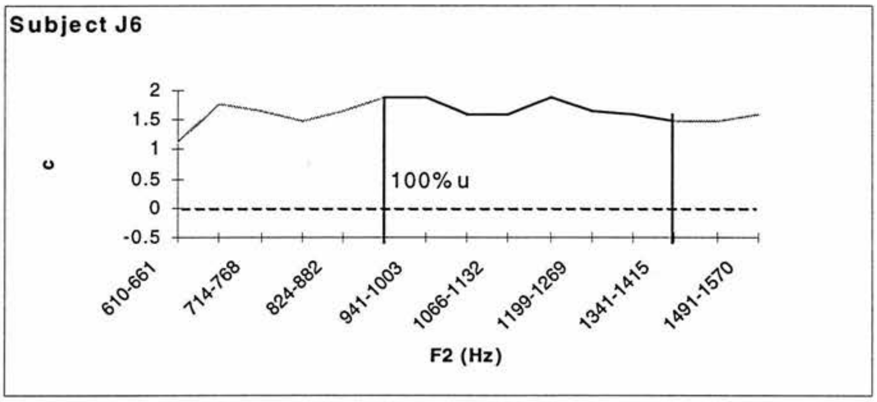


Figure 5:11a. Japanese subjects' C (response bias).







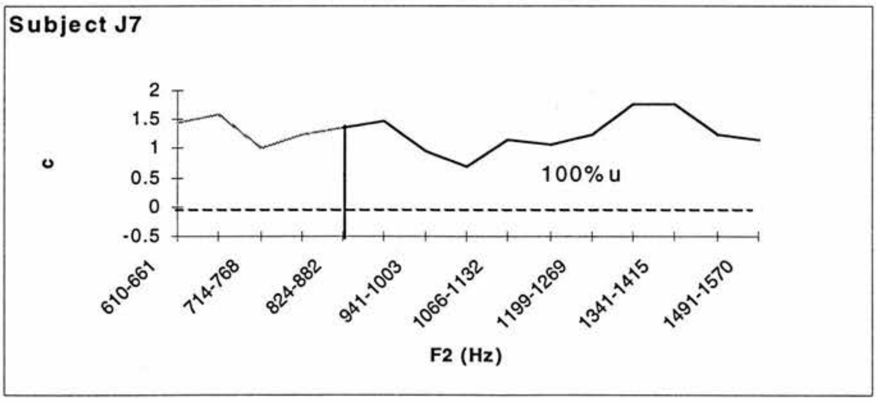
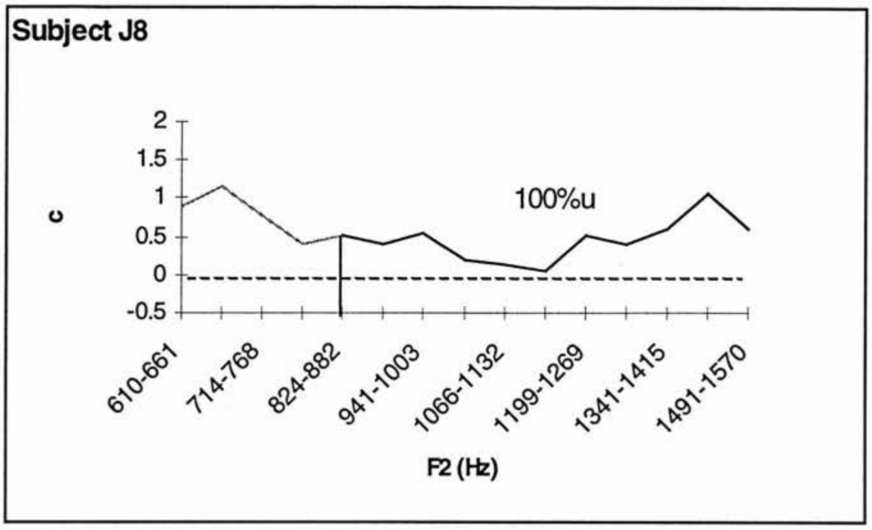
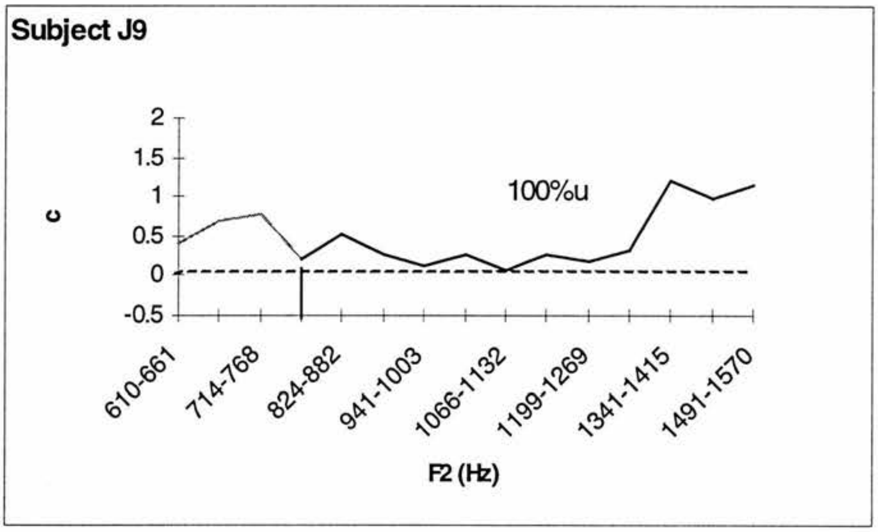


Figure 5:11b. Japanese subjects' C (response bias).





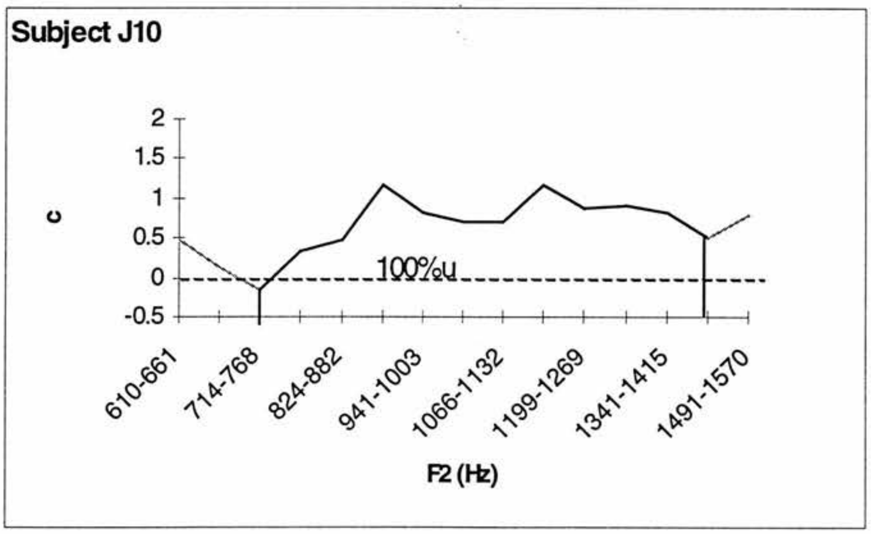


Figure 5:11c. Japanese subjects' C (response bias).

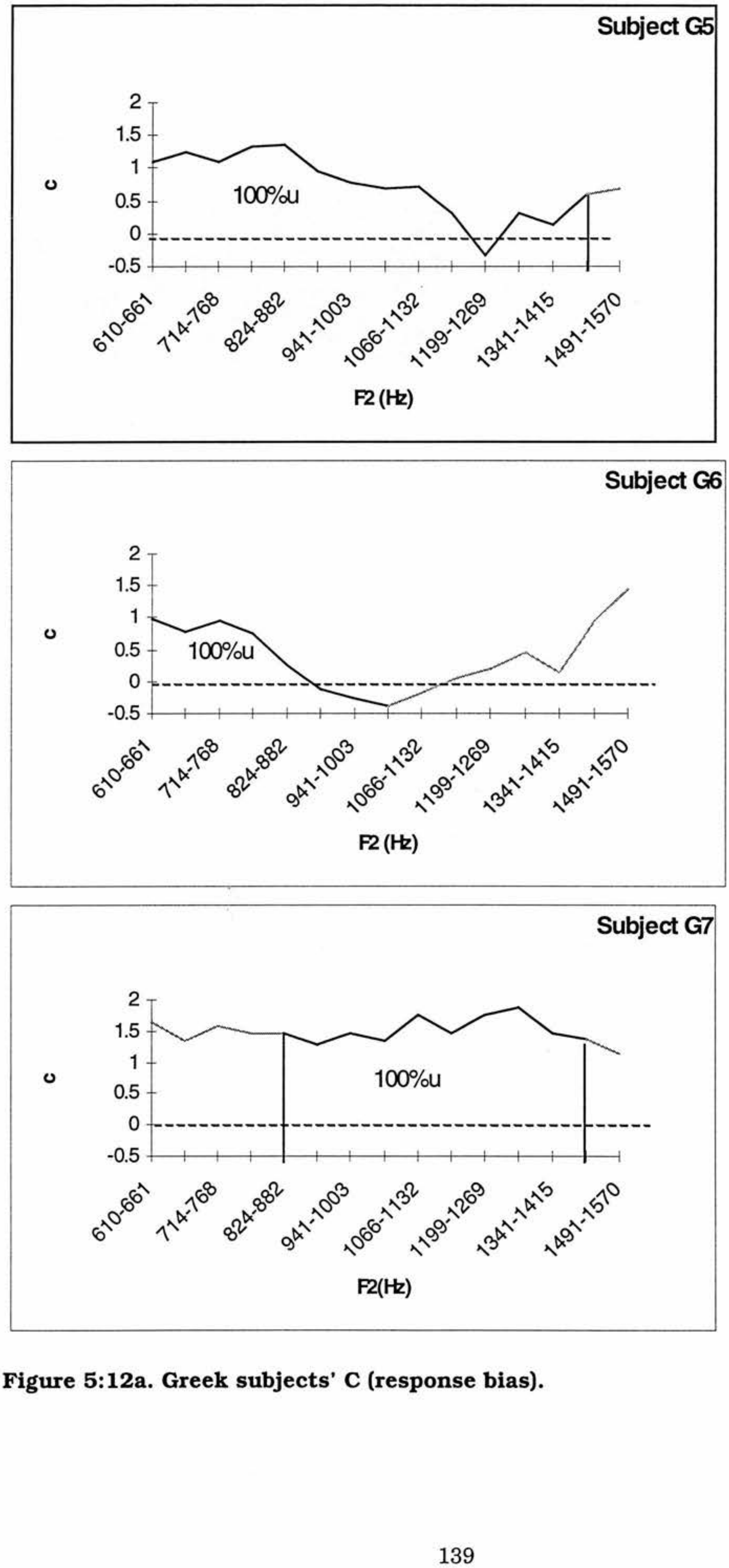


Figure 5:12a. Greek subjects' C (response bias).

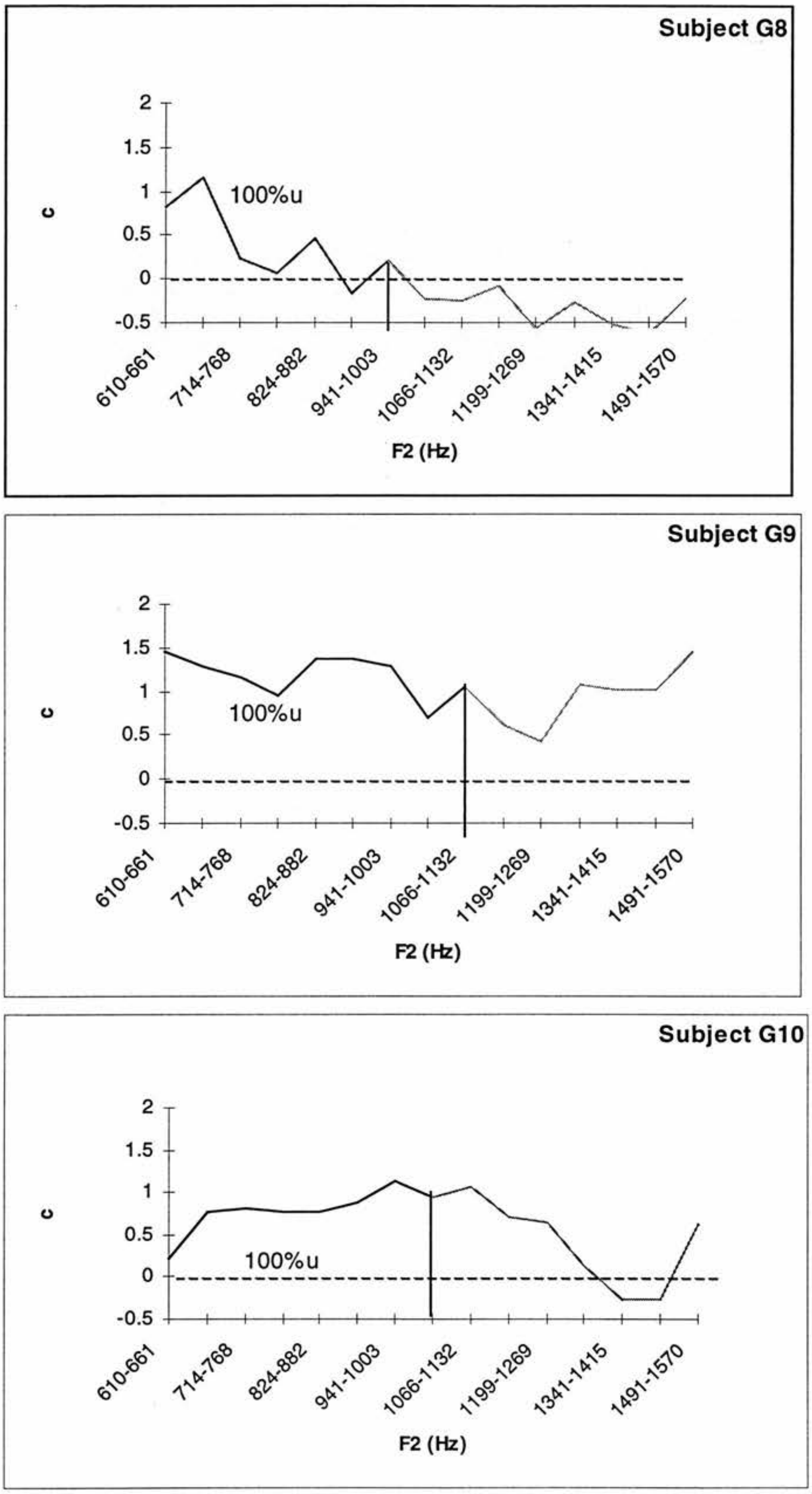


Figure 5:12b. Greek subjects' C (response bias).

As can be seen in Figs. 5:11a ~ 5:12b, the obtained c curves reveal that response biases shifted along the stimulus continuum considerably for most subjects, violating Signal Detection Theory's assumption of a constant bias (Macmillan, 1993: 41-42). Figs. 5:11a ~ 5:12b demonstrate that most Greek subjects' (Subjects G5, G6, G8 and G10) c is generally high for the stimuli with relatively low F2 and declines towards stimuli whose F2's are from mid- to high range, indicating that these subjects were more biased towards the 'same' response for the stimuli that were most consistently labelled as /u/ by these subjects. Similarly, the overall shape of some Japanese subjects' c (Subjects J4, J5 and J10) is convex in the mid-range where the stimuli were most consistently labelled as /u/ by the Japanese subjects. The fact that these subjects were more biased towards the 'same' response for stimuli away from the category boundary suggests that the perceptual distance is reflected in response bias, assuming that the perceptual distance between equally-spaced stimuli is larger at the category boundary. It is plausible that very small differences in the perceptual distance between stimuli may not always result in differences in discrimination sensitivity, considering studies such as Pisoni & Tash (1974), in which it was shown that within-category differences influence reaction times more than discrimination accuracy.

In fact, for some subjects, c appears to reflect the perceptual distance between the stimuli better than d'. For instance, Subject G6's d' (cf. Fig. 5:10a) has the lowest dip in a location where the predicted d' is the highest which marks his category boundary, contradicting the assumption that the perceptual distance is larger at the category boundary. An inspection of his miss and correct-rejection rates in Fig. 5:8a indicate that both his miss and correct-rejection rates are relatively low in the boundary region; in other words, both his hit and false-alarm rates are high in this region. That is, the low d' in the category boundary region is not due to high miss rates but low correct-rejection rates (or high-false alarm rates), which may be due to the unstable identity of the stimuli around the category boundary (cf. Repp & Liberman, 1987). On the other hand, Subject G6's c (cf. Fig. 5:12a) indicates that he was more biased towards the 'different' response in the category boundary region, which is compatible with the view that the perceptual distance between stimuli is larger around the category boundary. Thus, the present data indicate that larger perceptual distance may not be always accompanied by higher discrimination sensitivity, or better discrimination accuracy, as Macmillan puts it (Macmillan, 1993: 23). Albeit to a lesser extent, the deflation of d' in the category boundary region due to high false-alarm rates can be also found in Subjects J1, J10, G5 and G 10's d'. As a result, their d' curves are not the highest in the boundary region, which should be the case, if cross-category discrimination is easier than within-category discrimination. Thus, the implicit assumption that discrimination sensitivity and perceptual distance go hand in hand does not seem applicable to the present data. In some cases, the perceptual distance between the stimuli was better reflected in changes in response bias than discrimination sensitivity. Should this be the case, d', which corrects for response bias, fails to incorporate crucial information regarding the perceived distance between stimuli, if not sensitivity.

However, response bias c does not seem to reflect the assumed perceptual distance faithfully in all cases. For example, the overall shape of Subjects J2, J8 and J9's c curves are concave (see Figs 5:11a, 5:11c), indicating that they were more biased towards the 'different' response for stimuli that were identified as /u/ most consistently by the Japanese subjects, while they were more biased towards the 'same' response at the two ends of the stimulus continuum including category boundary regions, contradicting the expectation that a smaller perceptual distance would lead to more 'same' responses. According to Luce (1963: 147-154), response biases are thought to be controlled by the subjects to optimise something, for instance, money return when it is given as an incentive. Assuming that the subjects shifted their biases on the basis of the relative perceptual distance between the stimuli to optimise performance, it may be the case that these subjects estimated more of the pairs from the /u/ category to be 'different' than other subjects. Overall, two thirds of the subjects in the present study shifted their response biases along the stimulus continuum, which may have reflected perceptual distance between the stimuli, but this was not done so in a unanimous fashion. Should it be the case that changes in response bias reflect the perceptual distance between the stimuli, how they do so may be subject to strategies individuals decide to adopt to optimise performance, and therefore they do not seem to provide straightforward information regarding the perceptual distance between the stimuli.

The only consistent effect found in the present data was the presentation order effect in the subjects' response bias. In Figs. 5:13a ~ 5:14b two c curves were presented for each subject depending on whether the stimulus with a lower F2 was presented first or second. The solid line represents the c scores when the stimulus with a lower F2 was presented first, and the broken line represents those when the stimulus with a lower F2 was presented second. The dark portions of the lines represent cases where the stimuli were both identified as /u/ at all times. As can be seen in the figures, generally speaking, both Japanese and Greek subjects were more inclined to respond 'different' when the stimulus with a higher F2 was presented first. Although F3 of the stimuli also varied along the stimulus continuum, the effect is more likely to be due to the direction of change in F2 of the stimuli, given that no sudden shift in the order effect was observed for the pair of stimuli with the highest F2 values (1491 Hz and 1570 Hz) where the direction of the shift in F3 changed: F3 systematically decreased as F2 increased up to 1491 Hz and then started to increase, as defined by Nearey's (1989) formula (cf. 4.2).

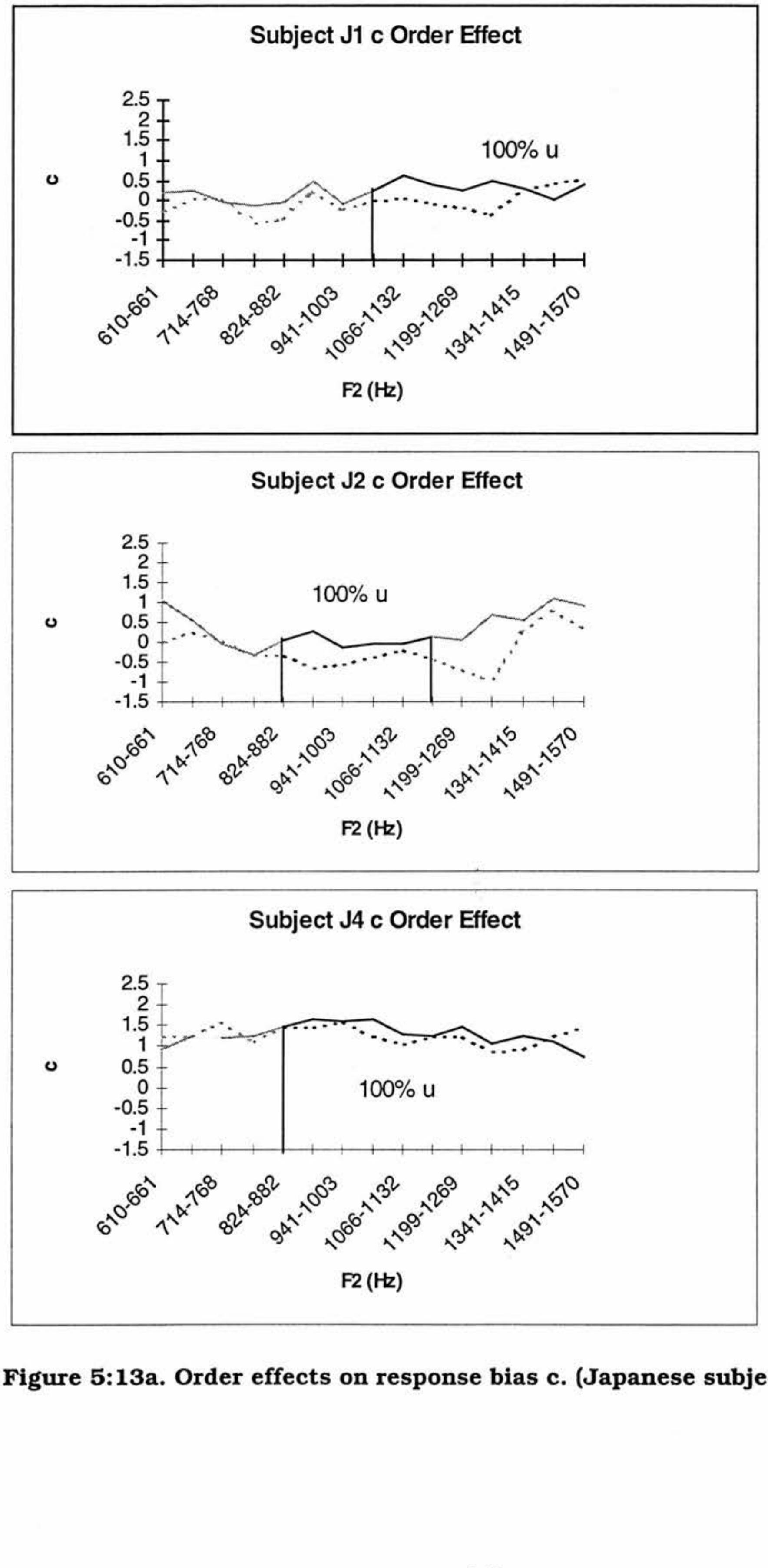
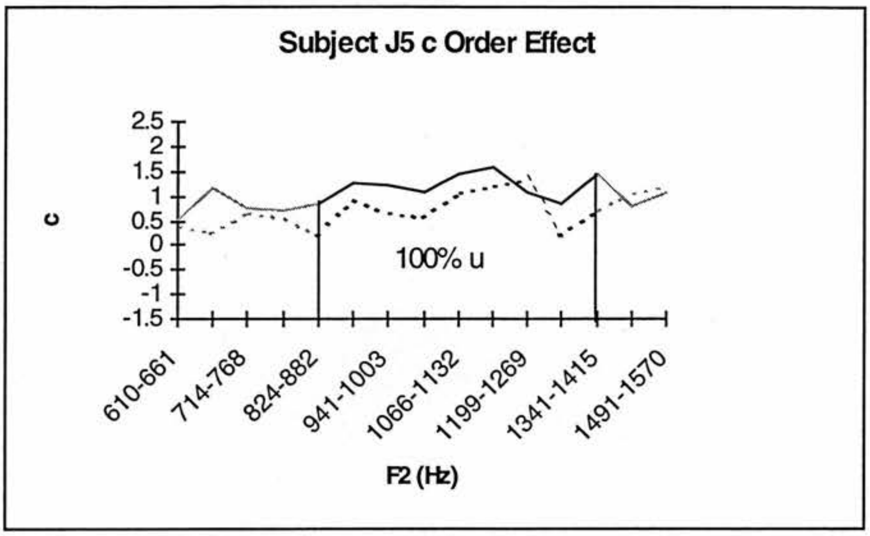
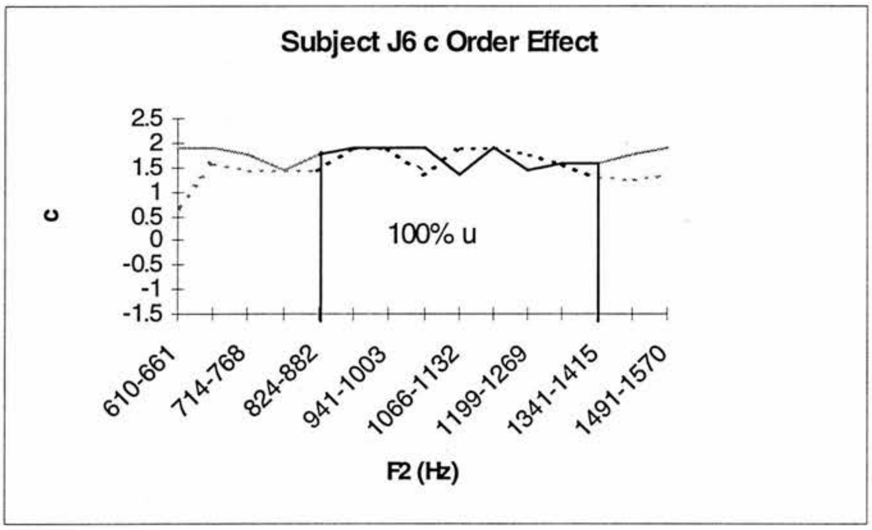


Figure 5:13a. Order effects on response bias c. (Japanese subjects).





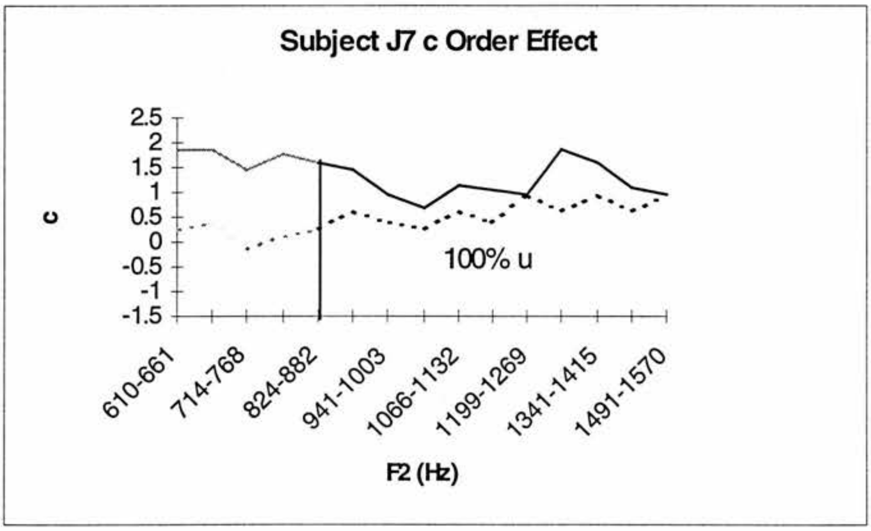
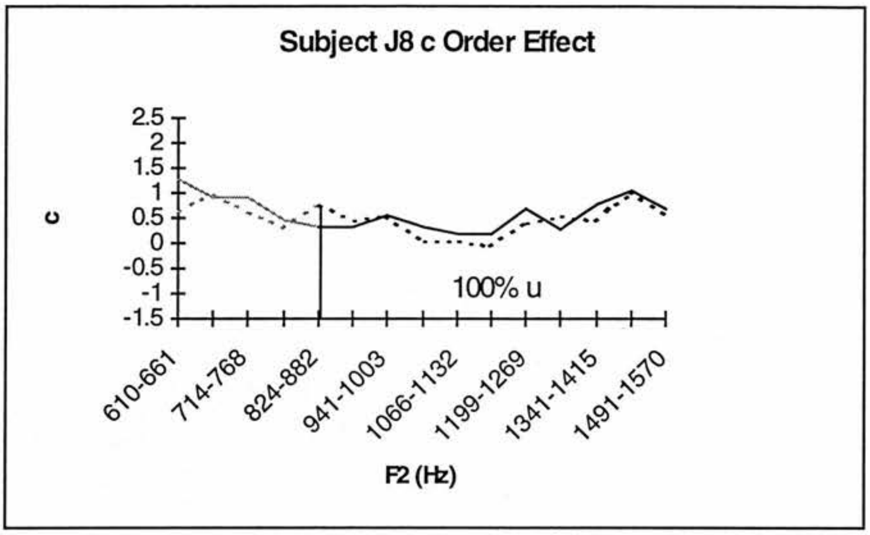
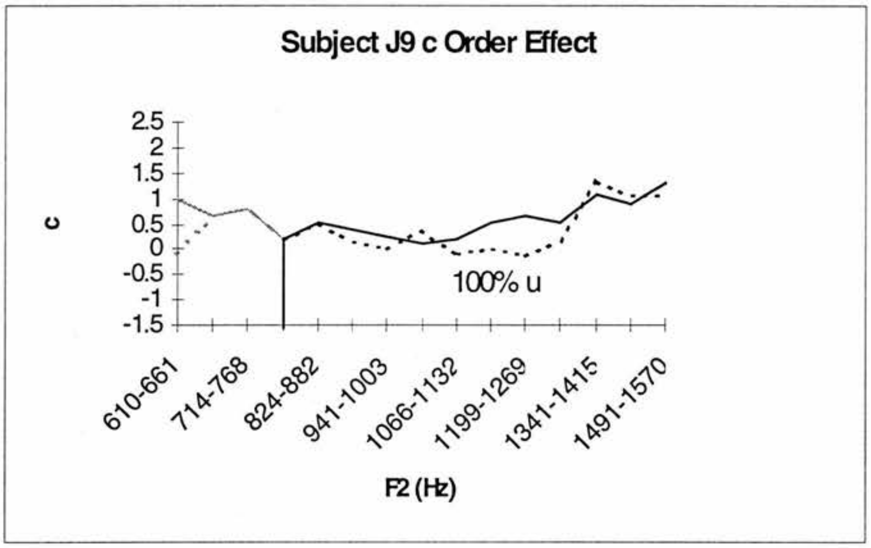


Figure 5:13b. Order effects on response bias c. (Japanese subjects).





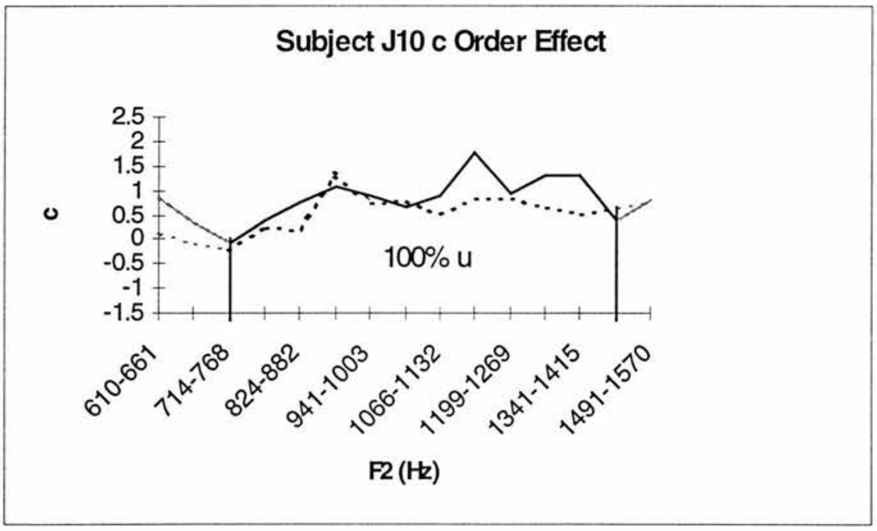
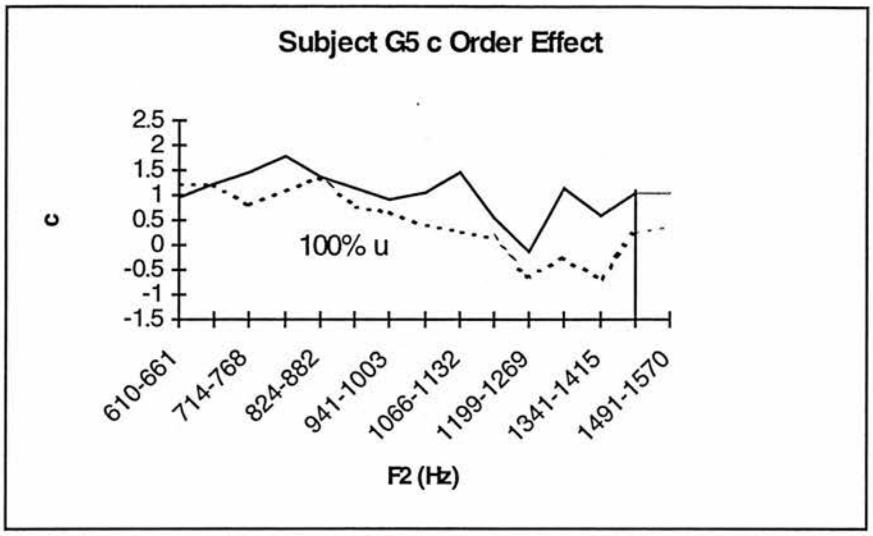
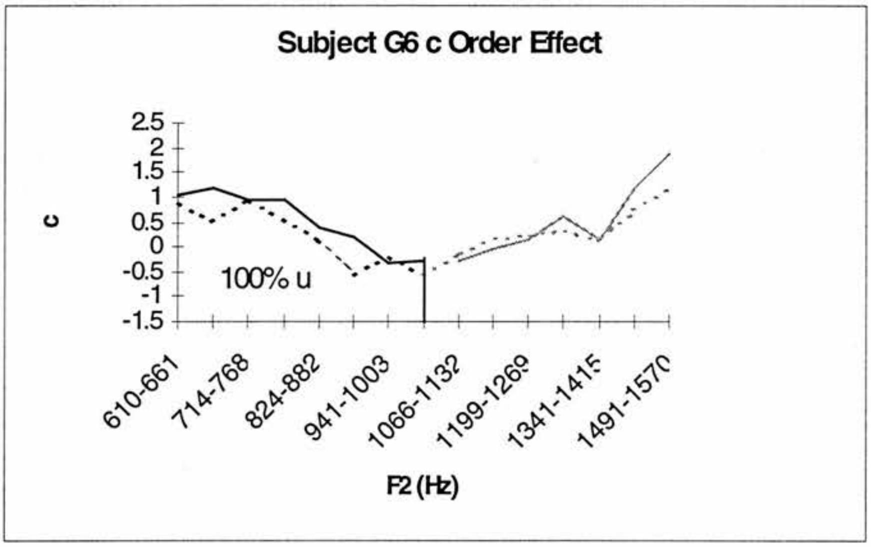


Figure 5:13c. Order effects on response bias c. (Japanese subjects).





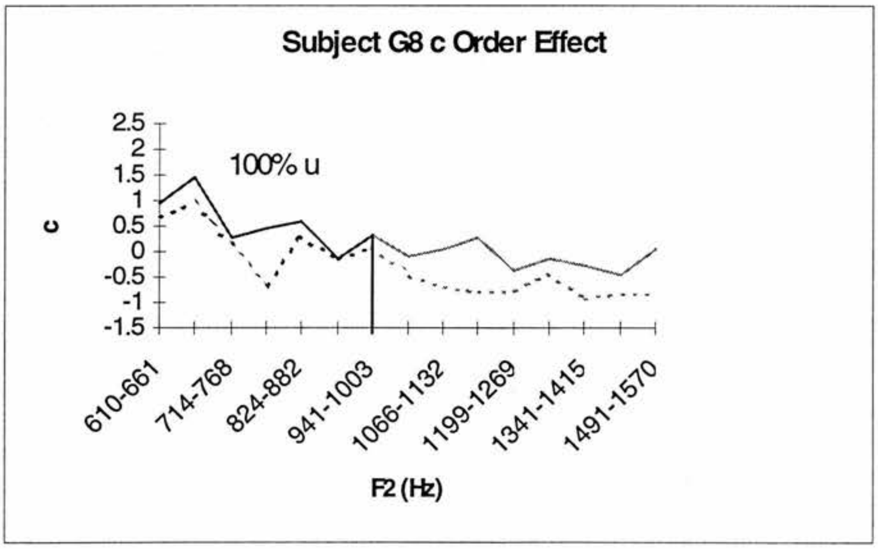
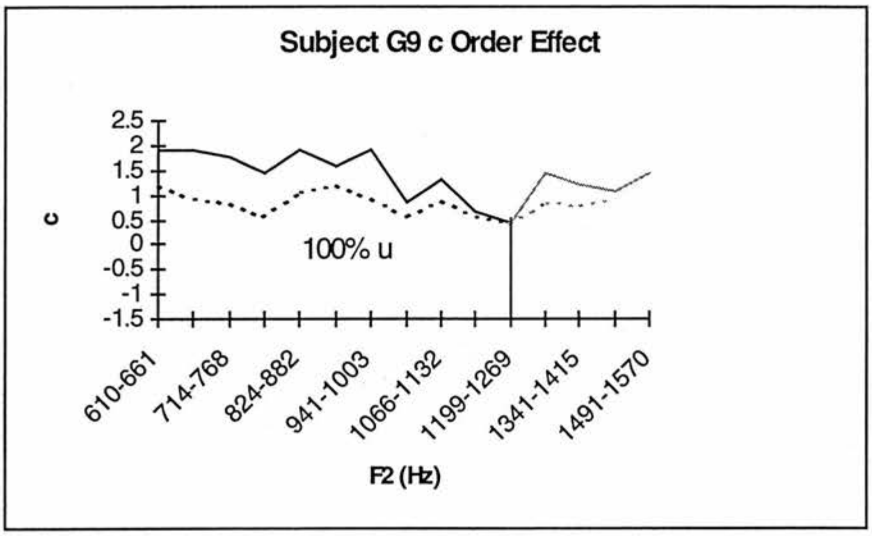


Figure 5:14a. Order effects on response bias c. (Greek subjects).



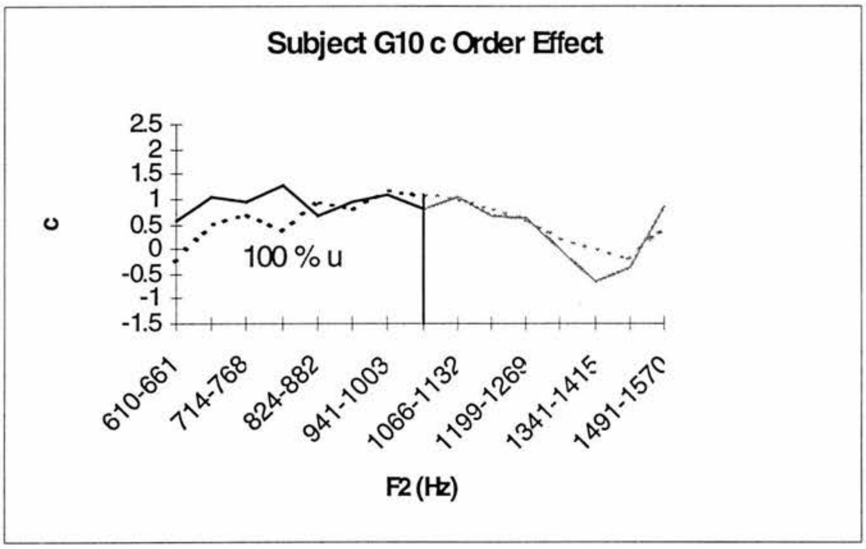


Figure 5:14b. Order effects on response bias c. (Greek subjects).

Assuming that response bias reflects the perceptual distance between the stimuli, the above results can be interpreted as indicating that the listeners perceived the distance between the stimuli to be shorter when they heard a stimulus of a relatively small F2 value (or more extreme F2, given the F2 range) before a stimulus with a less extreme F2. The observed order effect is another finding that is incompatible with Macmillan's (1993) idea of perceptual distance between stimuli, if the effect were to be taken as evidence for directional asymmetry in perceptual distance. According to Macmillan (1993: 42) the perceptual distance between a pair of stimuli is symmetric: d' (x, y) = d' (y, x).

The order effect found in response bias is consistent with Repp & Crowder's (1990) finding that American listeners responded 'different' more frequently in a same-different discrimination task when a less extreme variant of a vowel category was presented before a more extreme variant; in the case of /u/, when the stimuli with higher F1 and F2 were presented first, the listeners were more inclined to answer 'different'. Flanagan (1955) reports a similar finding as directional asymmetry in just-noticeable differences between vowel formants. Polka & Bohn (1996) also report a similar finding with English and German infants (age 6-8 and 10-12 months) for the German /dut/-/dyt/ contrast and the English /dɛt/-/dæt/ contrast; the infants' discrimination performance in both language groups was significantly poorer when syllables carrying the more extreme vowel, i.e., /dut/ and /dæt/, were presented first (also see Polka & Werker, 1994; for evidence for presentation effect in F0, see Verhoeven, 1992; Ladd & Morton, 1997). The presentation order effect is further discussed in 5.3.2.3.

5.3.2.3 Discussion

As described above, most subjects' discrimination sensitivities in the present study were near chance level, and neither miss rates nor d' provided a convincing evidence for a single assimilation point within the category /u/, towards which the listener's discrimination sensitivity declines. No effects of the prototypicality on discrimination sensitivity were observed. Although some subjects' d' exhibited a decline in discrimination sensitivity towards the extreme end of the category, this was not the case for the majority of the subjects. The only consistent outcome was the presentation order effect on the subject's response bias, which may be interpreted as directional asymmetry in the perceptual distance between the stimuli depending on their relative extremity. Specifically, the subjects were more inclined to respond 'same' when a relatively more extreme /u/ was presented first. In this section, I attempt to explain the results of the two replications of Kuhl (1991) that motivated the present study, i.e., Sussman & Lauckner-Morano (1995) and Iverson & Kuhl (1995), in the light of the present findings, and discuss the implications of the order effect for perceptual distance between vowels.

As discussed in Chap. 2, considering that both in Sussman & Lauckner-Morano (1995) and Iverson & Kuhl (1995) a decline in discrimination sensitivity was observed towards the most extreme stimuli used in the task, it is conceivable that the listener's discrimination sensitivity decreased towards the periphery of the vowel space. However, in the present study such a decline in sensitivity was observed only for a few subjects.

Given the order effect observed in the present study, the significantly poorer discrimination sensitivity around Kuhl's (1991) P (prototype stimulus) found in Sussman & Lauckner-Morano (1995) may be due to the presentation order effect of stimuli differing in their relative extremity. In their study, Kuhl's (1991) P (the production average) served the standard in a 'change/no-change' task in the prototype condition, and less extreme stimuli followed the standard in 'change' trials. In the non-prototype condition, a non-prototypical standard was followed by more extreme stimuli in 'change' trials. That is, in the prototype condition a more extreme stimulus was always presented before a less extreme stimulus, while in the non-prototype condition the presentation order was reversed. Thus, it is conceivable that the listeners were less willing to respond 'change' in the prototype condition due to the presentation order effect, which could have contributed to poorer discrimination sensitivity. However, the order effect would not explain the results of Iverson & Kuhl (1995) where the presentation order was counterbalanced and discrimination sensitivity was yet shown to decline towards the periphery of the vowel space, where the subjects' category ideals were located.

Alternatively, the discrimination task in the present study could have been too difficult to reflect the change in sensitivity. Although the step-size employed in the present experiment (45 mels) is larger than the smallest step-size in Iverson & Kuhl (1995) and Sussman & Lauckner-Morano (1995) (30 mels) where a systematic effect of the perceptual magnet was reported, it is conceivable that the short stimulus duration (85 ms) employed in the present study to facilitate the phonetic coding led to the observed floor effect in most subjects' performance. Iverson & Kuhl (1995) and Sussman & Lauckner-Morano (1995), on the other hand, employed long stimuli (500 ms), which is likely to have led to auditory coding. If the discrepant results arose from the difference between phonetic and auditory coding, the results are in apparent contradiction with my argument that the perceptual magnet effect should be more evident at the level of phonetic coding, assuming that the effect is of a language-specific nature, as maintained by the Native Language Magnet Theory. Alternatively, considering Lively & Pisoni's (1997) report that a greater generalisation (a failure to detect differences) in vowel quality was observed for stimuli varying in F2 alone in comparison to those varying in both F1 and F2, the fact that the stimulus varied only along F2 in the present study could be responsible for the observed floor effect. In Iverson & Kuhl (1995) and Sussman & Lauckner-Morano (1995), F1 and F2 of the stimuli were simultaneously varied. It is plausible that the listener is insensitive to differences in F2 alone, given that F2 is lower in amplitude than F1 (cf. Disner, 1983:5).

It is also conceivable that the formula used in converting Hz into mels are responsible for the discrepancy between the present study and the above two replications of Kuhl (1991). As already mentioned, the present study used a mel scale and Fant's (1973) formula to create psychoacoustically equidistant step-sizes. As brought up earlier, the step sizes created using Fant's (1973) formula may have resulted in step-sizes that have systematically increased, in psychoacoustic terms, towards the extreme end of the stimulus continuum for the Hz range (610 ~ 1570 Hz) used in the present discrimination task. Thus, the step-size employed in the present study may have biased the subjects away from exhibiting a decline towards the periphery of the vowel space. Should this be the case, the small number of subjects who exhibited a decline in discrimination sensitivity towards the extreme vowels may have been due to the spacing that increased gradually towards the periphery of the vowel space. Conversely, the step size used in Sussman & Lauckner-Morano (1995) and Iverson & Kuhl (1995) could have been systematically decreased towards the periphery in the high front corner of the vowel space where their test stimuli were located. As the formula is the 'best fit' to the psychoacoustically equidistant Hz values obtained in experiments and only give approximate Hz values that correspond to a psychoacoustic unit, the spacing between stimuli are more skewed in some ranges of Hz values than others.

At the same time, a close inspection of miss rates and d' suggested that perceptual distance between stimuli from a single category may not be necessarily reflected in discrimination accuracy. The presentation order effect on response bias found in the present discrimination task suggests an effect of relative extremity on perceptual distance between vowels, and possibly prototypicality. What may be seen a close analogy to the observed order effect is reported by Rosch (1975) using non-speech stimuli (e.g., colour), in which the psychological distance between a category prototype and a non-prototypical category member was expressed to be shorter in a physical space when the prototype was presented first and used as a reference point, compared to the distance when a non-prototypical member was presented first and used as a reference point. Specifically, Rosch compared the psychological distance between prototypical and non-prototypical stimuli (e.g., focal and non-focal colours) reflected in subjects' placement of the two kinds of stimuli on a semicircular space in two conditions: In one condition a prototypical stimulus was fixed at the origin of the semicircle, i.e., at the reference point, and the subjects placed a non-prototypical stimulus on the semicircle according to the perceived distance between the two stimuli. In the other condition, a non-prototypical stimulus was fixed at the reference point, and the subjects placed a prototypical stimulus. When these two conditions were compared, the distance between the two stimuli was significantly shorter when prototypical stimuli were fixed at the reference point than the reverse case.

If Rosch's study is relevant to the presentation order effect observed in the present study, the listener may have perceived the distance between the stimuli to be shorter when a relatively more prototypical stimulus was presented first. Provided the listeners' preference for extreme vowel stimuli found in the MOA task (Experiment 1b) and elsewhere (Bradlow, 1993; Johnson et al., 1993; Frieda, 1997), and the instability of the listener's choices of category ideals as evidenced in the MOA task, it is conceivable that the listeners perceived the more extreme stimulus of a given pair relatively more prototypical and found the perceptual distance to be shorter when the more extreme stimulus was presented first. In other words, the directional asymmetry found in the perceptual distance between the stimuli may be related to the relative prototypicality of the stimuli, which may be closely tied to their extremity. In any case, the presentation order effect suggests that there may be directional asymmetry in the perceived distance of L2 vowels depending on their relative extremity in relation to the L1 prototype. The implication of the order effect on the acquisition of L2 vowels is discussed in the next chapter.

5.4 Summary

Experiments 2a (the identification task) and 2b (the discrimination task) were conducted in order to test Kuhl's (1991, 1992, 1993) Native Magnet Language Theory which holds that language-specific phonetic prototypes assimilate other category members and cause the perceptual space to shrink around these prototypes, leading to low discrimination sensitivity. Studies on vowel prototypes to date are unclear concerning the locations of the prototypes and have yielded inconclusive results as to whether the observed decline in discrimination sensitivity is due to the prototypicality or extremity of the vowel stimulus designated as prototype. Thus, Experiments 2a and 2b were designed to measure Japanese and Greek listeners' discrimination sensitivities for members of /u/ differing in F2, where the effects of prototypicality and extremity on discrimination sensitivity were thought to be distinguished. Experiment 2a consisted of an identification task designed to determine which portion of the discrimination sensitivity curves obtained in Experiment 2b can be regarded within-category. Experiment 2b was a same/different discrimination task designed to examine changes in the subjects' discrimination sensitivities across the category /u/ along the F2 dimension. Two measures of discrimination sensitivity, miss rates and d', were used to test whether d' was in fact a better measure of the perceptual magnet effect, as Sussman & Lauckner-Morano (1995) report.

As shown in 5.3.1, cross-linguistic differences were observed in the identification task in a parallel fashion to the differences observed in the production task (Experiment 1a). Fewer Japanese subjects and more Greek subjects labelled stimuli with high F2 as /i/, implying that Japanese category boundary for /u/ has a higher F2 value than that of Greek /u/, which is compatible with the cross-linguistic differences observed in the production data (Experiment 1a). However, over one third of the Japanese and Greek subjects reported that they had heard some foreign sounds between the /u/- and /i/-categories, which suggests that these vowels may be well separated in their perceptual vowel space and that the category boundary between /i/ and /u/ depicted by the identification results may not be in a precise location, as the identification task was a forced-choice task where no choice was given for another category between /i/ and /u/. Furthermore, both Japanese and Greek subjects identified those stimuli with the lowest F2 values as /o/. The Japanese subjects' identification of those stimuli as /o/ can be explained by the differences in F2 between Japanese /u/ and /o/ found in production data (Imaishi et al, 1984). However, Greek subjects' identification of those stimuli as /o/ cannot be explained simply by comparing the F1 and F2 values of these two vowels found in the production data. Perhaps, a comparison of individual vowel formants is not a satisfactory way of comparing perceptual and production data, as Lively & Pisoni (1997: 1672) and Frieda, et al. (in press) point out.

As demonstrated in 5.3.2, almost two thirds of the subjects' discrimination performance was near chance level in Experiment 2b, and the results did not provide convincing evidence for a single assimilation point towards which the listener's discrimination sensitivity decreases. There was some evidence for the decline in discrimination sensitivity towards the extreme vowels within the 100% /u/ identification region, which could not be explained in terms of contrast effects, as was the case in Lotto, et al. (1996), given that both stimuli in the pair were judged to belong to a single category by the subjects. The decline in discrimination sensitivity towards extreme vowels could have been present in more subjects' discrimination sensitivity, had the step-size been determined using a different formula from that defined by Fant's (1973).[54]

At the same time, the data obtained in the discrimination task revealed that there was a considerable degree of shift in each subject's response bias along the stimulus continuum, which implies that perceptual distance between stimuli from a single category may not be accurately reflected in discrimination sensitivity. Pisoni & Tash (1974) also report that within-category differences are not necessarily reflected in discrimination accuracy; in their study, they found that within-category differences were better reflected in reaction times. In the present study, the most consistent effect was the presentation order effect on response bias depending on the relative extremity of the vowel stimuli. Specifically, the subjects were more inclined to respond 'same' when a relatively more extreme /u/ was presented first, implying that there may be directional asymmetry in the perceived distance between vowels depending on their relative

---

[54] For instance, Makhoul & Cosell's (1976) formula would have created a stimulus continuum whose step size is systematically larger towards the extreme end of the continuum in comparison to that employed by the present study.

extremity. In the light of Rosch's (1975) study, the order effect may be related to the relative prototypicality of the vowel stimuli as well as their extremity. The close relationship between vowel extremity and prototypicality is also suggested by the results of the MOA task (Experiment 1b) and elsewhere (Johnson et al. 1993; Bradlow, 1993; Frieda, 1997), where listeners' preference for extreme vowels is demonstrated. The implication of the presentation order effect on the ease of acquisition of L2 vowels is discussed in the next chapter.

Chapter 6 Conclusion

As described in Chaps. 2-3, the present dissertation is motivated by the question of how the degrees of perceived similarity between L1 and L2 vowels are determined. As illustrated in Chap. 2, the two current influential models of L2 phonetics, i.e., Best's Perceptual Assimilation Model and Flege's Speech Learning Model, predict the degree of difficulty of acquiring L2 sounds and contrasts on the basis of perceived similarity between L1 and L2 vowels. Best's Perceptual Assimilation Model maintains that the ease of perceptually learning L2 contrasts is determined by how these contrasts are assimilated to L1 categories on the basis of articulatory-phonetic (gestural) similarities between L1 and L2 sounds. On the other hand, Flege's Speech Learning Model maintains that the ease of the formation of a new perceptual L2 category is correlated inversely to the degree of acoustic-phonetic similarity between L1 and L2 phones. Therefore, according to these models, the degree of perceived similarity (whether it is articulatory or acoustic) between L1 and L2 phones is a key to predicting the degrees of difficulty of the perceptual learning of L2 phones and contrasts. However, how the degrees of similarity between L1 and L2 phones are determined is yet to be investigated.

In this connection, possible locations of L1 vowel prototypes were investigated. Although both models explicitly or implicitly assume some kind of mental representation of a phonetic category, or a phonetic prototype, which serves as a reference point in determining the perceived similarity, or the perceptual distance between L1 and L2 sounds, neither model specifies what such prototypes constitute. With regard to vowels, which are the focus of the present study, evidence suggests that category ideals may be more extreme than the language group's production averages of the categories (cf. Bradlow, 1993; Johnson et al., 1993; Iverson & Kuhl, 1995; Lotto et al., 1996; Lively & Pisoni, 1997; Frieda, 1997). This, in turn, suggests that the assumption originated in the visual domain (cf. Rosch, 1975; Goldman & Homa, 1977; Mervis & Rosch, 1981) that prototypes are the most preferred category member and at the same time situated at the centre of the category distribution may not be applicable to vowel categories. If vowel category ideals do not match the production averages, it follows that there are two possible prototype locations for vowel categories that may serve as reference points in determining the perceptual distance between L1 and L2 vowels, i.e., the category ideal and the production average. Traditionally, phonetic prototypes have been assumed to reflect the language specificity observed in the phonetic realisations of the categories (cf. Flege, 1986: 31; Johnson et al., 1993: 516; Kuhl, 1993: 130). Thus, where, if not at the centre of the category distribution, those category ideals are located in terms of production values was investigated.

The present study is also concerned with the contention of Kuhl's (1991, 1992, 1993, 1995) Native Language Magnet Theory that phonetic prototypes assimilate other category members (a perceptual magnet effect). If the L1 prototype assimilates other category members, causing the perceptual distance to shrink towards itself, such an effect should be taken into account when considering the perceptual distance between L1 and L2 sounds. However, the interpretation of the results of Kuhl (1991), on which her Native Language Magnet Theory is originally based, has been questioned by other researchers who replicated her study. Sussman & Lauckner-Morano (1995), Lotto et al. (1996) and Lively & Pisoni (1997) all report that Kuhl's (1991) NP (a non-prototypical /i/ for American listeners) was not identified as /i/ by their listeners, which suggests that the poorer discrimination found around Kuhl's (1991) P (a prototypical /i/) in comparison to Kuhl's (1991) NP may be simply due to the difference between within- and cross- category discrimination that took place around the respective stimuli, and may not be due to a perceptual magnet effect. Furthermore, in Kuhl (1991) the production average of /i/ reported in Peterson & Barney (1952) was given the best ratings by American listeners and therefore designated as the prototype. However, in its replications more extreme stimuli than Kuhl's (1991) P were given the best ratings, which suggests that Kuhl's (1991) P may not have been actually a prototypical /i/ for American listeners (cf. Iverson & Kuhl, 1995; Lotto et al., 1996; Frieda, 1997; Lively & Pisoni, 1997). In fact, in their replication of Kuhl (1991) Iverson & Kuhl (1995) report that discrimination sensitivity declined not towards Kuhl's (1991) P (the production average) but towards the best-rated stimuli which were more extreme than Kuhl's (1991) P. On the other hand, Sussman & Lauckner-Morano (1995) report that discrimination around Kuhl's (1991) P was poorer than that around a non-prototypical category member that was more reliably identified as /i/ than Kuhl's (1991) NP. Thus, the results of the studies on the perceptual magnet effect are not consistent in terms of where the perceptual magnet was found. In Iverson & Kuhl (1995) best-rated stimuli (category ideals) are reported to be the magnet, while in Sussman & Lauckner-Morano (1995) the production average seems to be the magnet. Considering that discrimination sensitivity around the 'prototype' was always compared with that around a less extreme, non-prototypical stimulus, it may not be the prototypicality but the extremity of the vowel that correlates with poor discrimination sensitivity. Furthermore, considering that a decline in discrimination sensitivity towards the periphery of the vowel space is reported to be found at the level of auditory coding (cf. Macmillan et al., 1988; Schouten & van Hessen, 1992) and that both Iverson & Kuhl (1995) and Sussman & Lauckner-Morano (1995) employed long stimuli, which is not regarded optimal for eliciting phonetic coding (cf. Fujisaki & Kawashima, 1969; Pisoni, 1973, 1975), what seems a decline in discrimination sensitivity towards extremity observed in the above studies may be specific to auditory coding. Given the universal nature of auditory coding (cf. Miyawaki et al., 1975; Repp et al., 1979:143; Best et al., 1981), the perceptual magnet effect, which is a language-specific phenomenon according to Kuhl, is expected to be observed at the level of phonetic coding. Thus, the questions that motivated the present study are:

Question 1a: Do vowel category ideals match the language group's production averages?

Question 1b: If the category ideals and the production averages of vowels do not match, do the locations of category ideals match the most extreme realisations of the vowel categories?

Question 2a: If the category ideals and the production averages of vowels do not match, does either of these correlate with poor discrimination sensitivity at the phonetic level of coding? If so, which one?

Question 2b: Does vowel extremity correlate with poor discrimination sensitivity at the level of phonetic coding?

In order to answer the above questions, four experiments, i.e., a production task (Experiment 1a), an MOA task (Experiment 1b), an identification task (Experiment 2a), and a discrimination task (Experiment 2b) were conducted using native speakers of Japanese and Modern Greek.

In this chapter, I summarise the results of the four experiments and discuss the implications of the findings for the L2 phonetics. The findings from Experiments 1a and 1b (the production and MOA tasks) are discussed in 6.1, and those from Experiments 2a and 2b (the identification and discrimination tasks) are discussed in 6.2. In 6.3 I discuss the implications of the findings for L2 phonetics.

6.1 Findings from Experiments 1a and 1b (the production and MOA tasks)

Experiments 1a (the production task) and 1b (the MOA task) were conducted in order to investigate the locations of vowel category ideals in relation to the production values of the language group (Questions 1a and 1b). Japanese and Modern Greek were chosen, since the two languages have phonologically comparable vowel systems in terms of inventory size and vowel qualities (five vowels: /i, e, a, o, u/). F1 and F2 values of the two language groups' category ideals for three vowels /i, a, u/ were compared to those of their production values. As the two languages differ in the phonetic realisations of the /u/ category, the locations of the category ideals for the two language groups were expected to be different, if they reflected the cross-linguistic differences in production. Specifically, it was expected that the Japanese category ideal for /u/ would have higher F2 than the Greek one.

In order to make the two tasks comparable, the above three vowels were produced in isolation in Experiment 1a, while steady-state synthesised vowel stimuli were used in Experiment 1b, where the subjects chose their category ideals from stimuli varying in F1 and F2, using the MOA technique adopted from Johnson et al. (1993). All the subjects were male, so that possible F0 differences in production would be minimised, and the production values could be directly compared with the synthesised vowels modelled on a male voice. The present experiments differed from the previous studies reporting listeners' preference for extreme vowels in that isolated vowels were obtained in the production experiment in order to allow a direct comparison between the production experiment and the perception experiment that used steady-state vowel stimuli. In previous studies, vowels were always produced in consonantal contexts while the perception experiment used steady-state vowel stimuli, which could have led to the results that the stimuli chosen in the perception experiment were more extreme than the production averages. In addition, the subjects were instructed to choose sounds that were the 'closest to their L1 vowels' rather than the 'best vowels' in eliciting the category ideals, as it was thought possible that the term 'best' might be interpreted as meaning 'distinct', which departs from the original definition of the prototype, i.e., 'representative'.

Even when care was taken to eliminate possible factors that might have led to the listener's preference for extreme vowels found in previous studies, both Japanese and Greek listeners chose vowels that were generally more extreme than the production averages except for Greek /u/, whose production values corresponded to the most extreme stimuli employed in the MOA task. However, the listeners' preference for extremity in comparison to the production values were reliably demonstrated only for the F2 of the two high vowels /i/ and /u/ (with an exception of Greek /u/) and F1 of Greek /a/. At the same time, the listeners' choices were clustered around the extreme end of the production range of each vowel category. Thus, the results suggest that vowel category ideals may correspond to the language group's most extreme realisations of the category. In other words, although they do not correspond to the production average of the language group, vowel category ideals seemingly reflect cross-linguistic differences in the phonetic realisations of the category. The observation that the Japanese subjects did not choose those stimuli with the most extreme F2 for the /u/ category suggests that this may be the case. Furthermore, considering studies showing English and Spanish listeners' preference for extreme vowels (Bradlow, 1993; Johnson et al., 1993; Iverson & Kuhl, 1995; Lotto et al., 1996; Lively & Pisoni, 1997; Frieda, 1997), the listener's preference for extreme vowels may be a universal tendency.

In addition to the above findings, there was a fair degree of between-subject differences in the choices of category ideals within each language group, which could not be explained in terms of dialectal differences. It was thought possible that these differences might be reflected in individuals' own production values, if category ideals served as individuals' production targets. Therefore, the correlation between F1 and F2 values of individuals' production and those of their category ideals was examined. A correlation between individuals' production values and their category ideals (MOA choices) was demonstrated for F2 of Japanese /u/ whose production values varied most between the subjects, but not for other vowels or formants, possibly due to the large spacing of the stimuli in the MOA task in comparison to the variability of the production values.

Furthermore, there was a fair amount of inconsistency in each subject's choices of category ideals across trials, suggesting that the locations of category ideals obtained in an experimental condition are not absolute. In general, the inconsistency was greater in the subjects' choices of F2 than those of F1, which is in line with the findings in the literature (cf. Johnson et al., 1993: 512-513; Lively & Pisoni, 1997; Frieda, in press).

Thus, Experiments 1a-b left us with two possible locations of language-specific vowel prototypes, i.e., the production average and the category ideal. According to the Native Language Magnet Theory, one of them is correlated with poor discrimination sensitivity. This question was hoped to be answered in Experiments 2a and 2b (the identification and discrimination tasks).

6.2 Findings from Experiments 2a and 2b (the identification and discrimination tasks)

Experiments 2a (a forced-choice identification task) and 2b (a same-different discrimination task) were conducted to answer Questions 2a and 2b, that is, to investigate whether listeners' discrimination sensitivity decreases towards the category ideal, the production average, or the periphery of the vowel space at the level of phonetic coding. If the perceptual magnet effect results in language-specificity in speech perception, as the Native Language Magnet Theory maintains, the effect should be evident at the phonetic level of coding, which is known to be language specific. In order to facilitate phonetic coding, vowel stimuli of a short duration (85 ms) and a long inter-stimulus interval (1 s) were employed.

The same Japanese and Greek subjects participated in the identification and discrimination tasks reusing the MOA stimuli from the high back region of the vowel space, where Japanese /u/ and Greek /u/ are located. This region of the vowel space was chosen because the effects of vowel extremity and prototypicality were thought to be distinguished in the Japanese subjects' discrimination sensitivity, as neither their production average nor category ideals were in the extreme corner of the vowel space. Thus, if vowel prototypicality correlated with poor discrimination, the Japanese subjects' discrimination sensitivity would decline towards either of the two possible prototype locations, i.e., the production average or the category ideals, and not towards the periphery of the vowel space. If vowel extremity correlated with poor discrimination, their discrimination sensitivity would decline towards the periphery of the vowel space.

Furthermore, as Japanese production average and category ideals were sufficiently apart, it was thought possible to discern which of them correlated with poor discrimination sensitivity. On the other hand, Greek subjects' discrimination sensitivity was expected to decrease towards the periphery of the vowel space, where both their production average and category ideals were situated.

The identification task was conducted in order to find out which stimulus pairs used in the discrimination task were both identified as /u/ by the subjects. As the present study was concerned with changes in within-category discrimination sensitivity, it was essential to find out which stimulus pairs used in the discrimination task were both labelled as /u/.

In addition, miss rates and a bias-free measure of d' were used to examine the change in each listener's discrimination sensitivity, in order to test whether d' is in fact a better measure of the perceptual magnet effect, as Sussman & Lauckner-Morano (1995) report. The two metrics were checked against their underlying assumptions: 100 % correct-rejection rates underlying the use of miss rates as a measure of discrimination sensitivity, and a constant response bias assumed by the use of d' as a measure of perceptual distance between stimuli.

The results of the identification task indicated that most stimulus pairs used in the discrimination task were perceived as /u/ by both Japanese and Greek subjects, as expected. Furthermore, cross-linguistic differences were observed in the locations of the Japanese and Greek listeners' category boundaries, in such a way as to reflect the differences observed in the production task (Experiment 1a). That is, fewer Japanese and more Greek subjects labelled stimuli with the highest F2 values as /i/, which implies that the Japanese category boundary between /i/ and /u/ may have a higher F2 value than the Greek one. Furthermore, the Japanese subjects labelled stimuli with the lowest F2 as /o/ in spite of the low F1 of the stimuli, possibly because /o/ is realised with lower F2 than /u/ in Japanese (cf. Imaishi et al., 1984). A few Greek listeners also labelled some stimuli with the lowest F2 as /o/, which could not be explained in terms of a simple comparison between the formants of those stimuli and Greek production values. To some extent, this may be an experimental artefact; some subjects might have felt that they should use all the three options (/i, u, o/) available to them. However, given that certain stimuli (in this case those with the lowest F2 values) were more likely to be identified as /o/, the acoustic characteristics of these stimuli must be relevant to the observed identification performance. Perhaps, a comparison of individual formant values is not a satisfactory way of studying vowel perception, as Lively & Pisoni (1997) and Frieda (in press) point out. Additionally, over one third of the Japanese and Greek subjects reported that they had heard some foreign sounds between /u/- and /i/- categories, suggesting that both Japanese and Greek /i/ and /u/ categories may be well separated in the perceptual vowel space, and that the boundary locations obtained in the present identification task may not have demarcated their perceptual categories accurately, as the identification task did not offer a choice between /i/ and /u/.

The results of the discrimination task, on the other hand, did not turn out to be quite as expected. Almost two thirds of the subjects' discrimination performance was near chance level despite the step-size that was larger than the smallest step-sizes employed in previous experiments reporting a perceptual magnet effect (Iverson & Kuhl, 1995; Sussman & Lauckner-Morano, 1995), and the results did not provide convincing evidence for a single assimilation point towards which the listener's discrimination sensitivity declines.

The near-chance level performance observed in the present study may have been due to the short stimulus duration (85 ms) and the long inter-stimulus interval (1s), which were employed to elicit phonetic coding. Provided that within-category discrimination is more difficult at the level of phonetic coding, this certainly is a possibility. However, as I argued in Chap.2, if phonetic coding is primarily responsible for the floor effect, it is difficult to see how the perceptual magnet effect would result in language-specific patterns of speech perception, as Kuhl's Native Magnet Theory maintains.

Or, the discrepancy may be due to the formula used in creating a 'psychoacoustically equidistant' step-size. As pointed out earlier, Fant's (1973) formula may have generated a step-size that systematically increased, in psychoacoustic terms, towards the extreme F2 values for the Hz range used in the present study. As those formulas are the 'best fit' to the psychoacoustically equidistant Hz values and only give approximate Hz values that correspond to a psychoacoustic unit, the spacing between stimuli are more skewed in some ranges of Hz values than others. For instance, Makhoul & Cosell's (1976) formula for converting mels into Hz would have generated stimulus spacing that grew smaller towards the extreme F2 values in comparison to that employed in the present experiment.

Alternatively, the floor effect may have been due to the dimension along which the stimulus continua stretched. In the present discrimination task the stimuli varied only along F2. Given the results of Experiment 1b where the MOA choices varied more in F2 than F1 especially for high vowels, the listener may be less sensitive to differences in vowel quality differing in F2 than F1. Thus, the step-size employed in the current study could have been large enough for the perceptual magnet effect to manifest itself, if F1 and F2 of the stimuli had been simultaneously varied, as was the case in Iverson & Kuhl (1995) and Sussman & Lauckner-Morano (1995). This, however, leads to a question of what constitutes 'psychologically equidistant step-size'. The fact that a step-size employed in the present study was larger than the smallest step-size used in Kuhl (1991) but produced floor effects for most subjects implies that the independent manipulation of formants using the mel, a scale for the pitch of a pure tone, would not yield psychologically equidistant step-size for speech stimuli whose formants are supposedly integrated by the listener. Lively & Pisoni (1997) also found that the subjects' discrimination sensitivities were poorer when F2 of the stimuli alone was varied. They state, 'The pattern of results observed here indicates that the underlying psychological distance between variants may not be equated by independently scaling formant values in mels. Rather, more sophisticated scaling techniques may be required to ensure equal stimulus spacing with complex, multidimensional stimuli such as these synthetic vowels' (Lively & Pisoni, 1997: 1672).

At the same time, an inspection of the data revealed that neither of the assumptions underlying the two metrics of discrimination sensitivity used in the present study, i.e., miss rates and d', was met by the present data. Specifically, each subject's correct-rejection rates varied along the stimulus continuum, suggesting that miss rates were not an accurate measure of discrimination sensitivity for the present study. Similarly, each subject's response bias varied along the stimulus continuum, suggesting that d' may not have accurately reflected the perceptual distance between the stimuli. As discussed in Chap. 5, the subject's response bias was seemingly affected by both his basic sensitivity and expectation regarding how many pairs consist of different stimuli. This implies that the step-size and/or the proportion of S and D pairs (pairs consisting of the same stimulus and different stimuli) need to be carefully adjusted for each subject in order for discrimination accuracy to serve as a measure of perceptual distance between stimuli, which presents problems to the experimenter. First, this would potentially mean a considerable amount of work before obtaining data that can be used as a measure of perceptual distance. In addition, if subjects were tested more than once, a practice effect on their performance would be inevitable. Finally, when would we know that either the floor or ceiling effect is indeed due to no change in discrimination sensitivity along the stimulus continuum?

The only consistent effect found in the results of the discrimination task was a presentation order effect in the subjects' response biases. Specifically, the subjects were more inclined to choose the 'same' response when a more extreme instance of /u/ was presented first than in the reverse case, which implies that relative extremity of vowel stimuli may have an effect on the perceived distance between vowels, depending on the order in which they are presented. Considering the listeners' preference for extreme vowels observed in Experiment 1b (the MOA task) and elsewhere (Johnson et al., 1993; Bradlow, 1993; Frieda, 1995), it is possible that the more extreme stimulus is regarded more prototypical by the listener, and that the order effect is related to vowel prototypicality as well as extremity. A close analogy can be found in Rosch's (1975) work where the perceptual distance was perceived to be shorter when the prototype was presented first than in the reverse case. If the above interpretation is valid, it may be the category ideal, not the production average, that plays a special role in speech perception.

Furthermore, the presentation order effect observed in the discrimination task suggests that there may be directional asymmetry in the perceptual distance between L1 and L2 vowels depending on their relative extremity (For a similar view, see Polka & Bohn, 1996).

6.3 Conclusion: Implications for L2 phonetics

The present dissertation attempted to investigate how the degrees of similarity between L1 and L2 vowels may be determined in the framework of Best and Flege's L2 phonetics models, which, explicitly or implicitly, assume that there is some kind of mental representation of L1 phonetic categories, which serves as reference points in determining perceptual distance between L1 and L2 phones. As explained earlier, such mental representations, or prototypes, have been conventionally described as 'typical' and at the same time the 'best' exemplar of the category by the proponents of prototype theory of speech perception, and the production average and the best-rated instance of a phonetic category (a category ideal) have been interchangeably referred to as the 'prototype'. However, whether the category ideal corresponds to the production average, as has been often assumed, is open to question.

The results of Experiments 1a-b (the production and MOA tasks) suggested that Japanese and Greek vowel category ideals are more extreme than each language group's production averages, as suggested for English and Spanish (cf. Johnson et al., 1993; Bradlow, 1993; Frieda, 1995). At the same time, the listeners' choices were clustered around the extreme end of each language group's production range, suggesting a link between the production range of the language group and the location of the category ideal, which left us with two possible language-specific prototypes for vowel categories. In the light of Kuhl's Native Language Magnet Theory, it was hoped that the question of which of these plays a special role in speech perception might be answered in Experiments 2a-b (the identification and discrimination tasks), but the question remained unanswered. Should there be mental representations of L1 vowel categories, against which L2 vowels are compared, what constitutes such mental representations needs to be clarified in order for us to further investigate what determines the degrees of perceived similarity between L1 and L2 vowels.

If it is the category ideal that plays such a role, the between-subject disagreement observed in the locations of the category ideals suggests that there may not be a single location for the category ideal for a language group. The fact that the Japanese listeners' choices of F2 for the category /u/ were correlated with their own production values suggests that individuals' category ideals may be as variable as their production values. Thus, assuming that there are mental representations of L1 phonetic categories, the concept of a language group's mental representation needs to be replaced by individual speakers' mental representations, which may cluster in a certain area for a language group. The concept of multiple representations for a language group implies that the perceived distance between L1 and L2 vowels may vary among individuals even within the same language group, and that the individual differences in the ease of acquiring a certain L2 vowel may, at least partially, arise from differences in the locations of the category ideal in relation to the L2 vowel, if the ease of perceptual learning of L2 vowels is determined by the perceived distance between L1 and L2 vowels, as Best and Flege maintain.

At the same time, the fair amount of inconsistency observed in individual listeners' choices in the MOA task and the lack of correlation between individual subjects' production values and their MOA choices for other vowel categories suggests a difficulty in eliciting such category ideals with accuracy. As seen in Chap. 4, in the present study, the subjects were generally much more consistent in their production than their MOA choices. According to Lively & Pisoni (1997: 1676), '... implicit in a phonetic prototype theory, the same prototype must be applied in each case because it is the idealized representation for the category as a whole'. If this assumption is correct, neither the MOA task nor the goodness rating task seems to serve as an ideal tool for locating such prototypes. Without the precise location of the prototype, however, one cannot test the prototype theory of speech perception or speech learning models that draw on the concept of prototypes. Nor can the degrees of perceived similarity between L1 and L2 phones be predicted from the results of such perception tasks, assuming that the locations of category ideals are the key to measuring perceptual distance between L1 and L2 phones. In this connection, Lively & Pisoni (1997: 1676-7) further note, '... adapting prototype theory to deal with contextual shifts in goodness ratings may involve the postulation of some sort of perceptual normalization mechanism... the assumption of idealized, highly abstract, prototypical representations for phonetic categories may have to be abandoned in favour of multiple context-sensitive representations'. If this is the case, it follows that elusiveness is something inherent to the prototype, which is not simply a question of the adequacy of the experimental paradigm.

As described earlier, the results of Experiments 2a-b (the identification and discrimination tasks) did not demonstrate convincing evidence for a decline in discrimination sensitivity towards the prototype, as Kuhl's Native Language Magnet predicts, or towards extreme vowels, as I speculated, which I argued would lead to asymmetry in perceptual distance between vowels depending on their extremity. The presentation order effect on response bias found in the discrimination task, however, suggests that there may be directional asymmetry in the perceptual distance between vowels depending on their relative extremity. Given that the order effect has been found in quite a few studies, using different kinds of stimuli and paradigms (Flanagan, 1955; Repp & Crowder, 1990; Verhoeven, 1992; Polka & Bohn, 1996; Ladd & Morton, 1997), the order effect may not be just an experimental artefact, but what is the underlying mechanism? Repp & Crowder (1990) try to explain the order effect as the quality of the vowel presented first decaying in memory and changing its quality towards the interior of the vowel space, but dismiss the idea and write, '[it] seems ... that the category boundary somehow "attracts" vowels in memory ... such a process cannot be reconciled with the idea of covert phonetic categorisation', as '[phonetic] categorisation amounts to an assimilation to the prototype ...' (Repp & Crowder, 1990: 2088). As I already brought up, considering Rosch's (1975) study showing that the distance between a category prototype and a non-prototypical category member was perceived shorter when the prototype was presented first and used as a reference point than in the reverse case, it may be the case that the more extreme vowel stimulus was regarded more prototypical by the listener, and the distance between the pair of vowels was perceived to be shorter when the more prototypical stimulus was presented first. Given the listener's preference for extreme vowels and the instability of the locations of those category ideals observed in the MOA task, this seems a possible explanation.

However, assuming that the above interpretation of the order effect is valid, we run into another problem when we attempt to make predictions for the acquisition of L2 vowels, as we do not know whether the prototype compares to the first or the second stimulus in real-life speech perception. Assuming that the prototype theory of speech perception is valid, do listeners hear a sound and then compare it with their internal representations? Or, do they have the internal representations 'evoked' before the comparison takes place, with which the incoming signal is compared? If the former is the case, the second stimulus compares to the prototype; if the latter is the case, the first stimulus compares to the prototype. If it is the first stimulus that plays an analogous role to the prototype, it follows that it is more difficult to distinguish a less extreme L2 vowel from a similar L1 vowel perceptually, since the perceptual distance is supposedly shorter when a more extreme stimulus is presented first. If it is the second stimulus that compares to the prototype, the prediction would be the exact opposite. Or, are the stimuli in the discrimination task each compared with the inner representation before the comparison between the two takes place, based on which the distance between the two stimuli is calculated? If the discrimination of the stimuli is mediated through the prototype, it no longer makes sense to discuss which of the two stimuli in the discrimination task plays an analogous role to that of the prototype. Nor does the presentation order effect have direct relevance to the ease of acquisition of L2 vowels, as our interest is in the distance between the L1 prototype and L2 sounds, and not the distance between two physical sounds which is determined through the mediation of the prototype. In any case, if we were to assume that the observed presentation order effect is relevant to speech perception, we need an explanation of what mechanism underlies such an effect.

Apart from the interpretation of the order effect, the above question leads to a further question of how much resemblance there is between comparing two physical sounds for the purpose of discrimination and real-life speech perception, which may be described, in the framework of the prototype theory of speech perception, as comparing a physical sound with an abstract mental representation (i.e., the prototype) for the purpose of identification. Moreover, there is a question of whether or not a physical sound can play the same role as the prototype that is thought to be an abstract mental representation. In my view, supposing that the prototype is an abstract representation, the best-rated stimulus may be a physical sound that is perceived as the closest to the prototype in a given context but nothing more, and whether these best-rated stimuli can be regarded as an equivalent to the abstract representation is questionable.

From the above viewpoint, goodness-rating tasks may be a more direct and useful measure of the perceptual distance between the prototype and physical sounds, if the goodness of the stimuli can be taken to reflect the perceived distance between the abstract mental representation and the physical stimuli. Should the results of the goodness-rating task be taken to indicate the perceptual distance between the prototype and the stimulus, the fact that extreme vowels tend to be given higher ratings implies that extreme vowels are perceived to be closer to the prototype. For instance, in Bradlow (1993) both American and Spanish subjects gave higher ratings to stimuli that were more extreme than the production range of respective language groups (cf. Fig. 2:2). Considering the results of Experiments 1a and 1b indicating that the prototype may correspond to the most extreme realisations of the speakers' production range, it may be the case that those stimuli that are more extreme than the prototype are perceived to be closer to the prototype than those that are less extreme. That is, L2 vowels that are more extreme than the L1 prototype may be more difficult to perceptually distinguish from the prototype.

Should this be the case, a conflict between vowel extremity and perceptual distance may be observed in the acquisition of L2 vowels. According to Flege's Speech Learning Model, the ease of acquiring an L2 phone is inversely correlated with the perceptual distance between L1 and L2 phones. Assuming that the listener's preference for extreme vowels is a universal tendency, however, it may be difficult to establish a new representation for less extreme, L2 vowels, even if the perceptual distance between L1 and L2 vowels is relatively large. On the other hand, extreme L2 vowels may be relatively difficult to distinguish from the L1 prototype, but they may overwrite the L1 prototype as a result of exposure to these L2 vowels. Hecht & Mulford's (1982) study seems to indicate difficulty in establishing a new category for less extreme, L2 vowels, while Johansson's (1973) study seemingly presents a case where extreme L2 vowels are not distinguished from the learner's less extreme L1 vowels (cf. 2.1.3). On the other hand, in Flege (1986) experienced French speakers of English seem to differentiate their L2 (less extreme) /u/ from L1 (more extreme) /u/ in production better than experienced American speakers of French do; this could be taken to suggest that perceived similarity between L1 and L2 vowels may be a greater hindrance to the acquisition of L2 vowels than perceptual preference for extreme vowels. Or, given that we do not know for certain in which direction the asymmetry in perceived distance may be observed, it may be the case that it is more difficult to distinguish less extreme vowels from a more extreme, vowel prototype. The possible directional asymmetry in perceptual distance between L1 and L2 vowels would be an interesting issue to be addressed in future studies of L2 phonetics.

Appendix A: Subjects' background

*All the information is based on the subject's self-report.

[Japanese subjects]

Subject	L1 (regional accent)	Date of birth	Length of residence in UK	English proficiency	Knowledge of other foreign languages	Experience of living abroad (age)
J1	Japanese (standard)	8/6/1970	1 year 4 months	intermediate		
J2	Japanese (standard)	4/2/1971	2 years 2 months	advanced		Germany (age 2-6) US (age 21-22)
J3	Japanese (Northern)	9/27/1973	1 year 2 months	intermediate		
J4	Japanese (standard)	7/20/1956	8 months	intermediate		
J5	Japanese (standard)	6/19/1960	8 months	intermediate		
J6	Japanese (Northern)	11/4/1960	9 months	intermediate		
J7	Japanese (standard)	5/25/1971	9 months	intermediate		
J8	Japanese (standard)	10/23/1965	1 year 6 months	intermediate		
J9	Japanese (standard)	8/8/1970	4 years 4 months	advanced	French	
J10	Japanese (Western)	2/28/1955	8 months	intermediate		

[Greek subjects]

Subject	L1 (regional accent)	Date of birth	Length of residence in UK	English proficiency	Knowledge of other foreign languages	Experience of living abroad (age)
G1	Greek (Athenian)	5/10/1972	4 years 1 month	advanced		
G2	Greek (Chiotika)	12/1/1965	2 years	advanced	French	Belgium (age 18-19)
G3	Greek (Athenian)	2/9/1973	3 months	advanced	German	
G4	Greek (standard)	11/14/1973	2 years 5 months	advanced	German	Germany (age 19-20)
G5	Greek (standard)	10/2/1971	3 years 5 months	advanced	French, Spanish	
G6	Greek (Athenian)	11/24/1961	3 years 3 months	advanced		
G7	Greek (Athenian)	1/24/1975	2 years 6 months	advanced	German	
G8	Greek (Athenian)	11/23/1964	5 years 3 months	advanced		
G9	Greek (Athenian)	12/16/1972	2 years 5 months	advanced	German, French	
G10	Greek (Thessalonikan)	2/25/1974	4 months	advanced	French	US (age 18-22)

Appendix B: A synthesis specification file

```
                         Synthesis specification for file

SenSyn      Version 1.1      Sensimetrics Corporation

Max output signal (overload if greater than 0.0 dB) is  -19.2 dB
Total number of waveform samples = 850


CURRENT CONFIGURATION:
60 parameters

SYM V/C  MIN   VAL   MAX  DESCRIPTION
---------------------------------------------
DU   C    30    85  5000  Duration of the utterance, in msec
UI   C     1     5    20  Update interval for parameter reset, in msec
SR   C 5000 10000 20000  Output sampling rate, in samples/sec
NF   C     1     4     6  Number of formants in cascade branch
SS   C     1     2     3  Source switch (1=impulse, 2=natural, 3=LF model)
RS   C     1     8  8191  Random seed (initial value of random # generator)
SB   C     0     1     1  Same noise burst, reset RS if AF=AH=0, 0=no,1=yes
CP   C     0     0     1  0=Cascade, 1=Parallel tract excitation by AV
OS   C     0     0    20  Output selector (0=normal,1=voicing source,..)
GV   C     0    50    80  Overall gain scale factor for AV, in dB
GH   C     0    60    80  Overall gain scale factor for AH, in dB
GF   C     0    60    80  Overall gain scale factor for AF, in dB
F0   V     0  1200  5000  Fundamental frequency, in tenths of a Hz
AV   V     0    60    80  Amplitude of voicing, in dB
OQ   v    10    50    99  Open quotient (voicing open-time/period), in %
SQ   v   100   200   500  Speed quotient (rise/fall time, LF model), in %
TL   v     0     0    41  Extra tilt of voicing spectrum, dB down @ 3 kHz
FL   v     0     0   100  Flutter (random fluct in f0), in % of maximum
DI   v     0     0   100  Diplophonia (alt periods closer), in % of max
AH   v     0     0    80  Amplitude of aspiration, in dB
AF   v     0     0    80  Amplitude of frication, in dB
F1   v   180   497  1300  Frequency of 1st formant, in Hz
B1   v    30    63  1000  Bandwidth of 1st formant, in Hz
DF1  v     0     0   100  Change in F1 during open portion of period, in Hz
DB1  v     0     0   400  Change in B1 during open portion of period, in Hz
F2   v   550  1491  3000  Frequency of 2nd formant, in Hz
B2   v    40    56  1000  Bandwidth of 2nd formant, in Hz
F3   v  1200  2187  4800  Frequency of 3rd formant, in Hz
B3   v    60   136  1000  Bandwidth of 3rd formant, in Hz
F4   v  2400  2487  4990  Frequency of 4th formant, in Hz
B4   v   100   200  1000  Bandwidth of 4th formant, in Hz
F5   v  3000  3700  4990  Frequency of 5th formant, in Hz
B5   v   100   200  1500  Bandwidth of 5th formant, in Hz
F6   v  3000  4990  4990  Frequency of 6th formant, in Hz (applies if NF=6)
B6   v   100   500  4000  Bandwidth of 6th formant, in Hz (applies if NF=6)
FNP  v   180   500  2000  Frequency of nasal pole, in Hz
BNP  v    40    90  1000  Bandwidth of nasal pole, in Hz
FNZ  v   180   500  2000  Frequency of nasal zero, in Hz
BNZ  v    40    90  1000  Bandwidth of nasal zero, in Hz
FTP  v   300  2150  3000  Frequency of tracheal pole, in Hz
BTP  v    40   180  1000  Bandwidth of tracheal pole, in Hz
```

```
FTZ  v  300  2150  3000  Frequency of tracheal zero, in Hz
BTZ  v   40   180  2000  Bandwidth of tracheal zero, in Hz
A2F  v    0     0    80  Amp of fric-excited parallel 2nd formant, in dB
A3F  v    0     0    80  Amp of fric-excited parallel 3rd formant, in dB
A4F  v    0     0    80  Amp of fric-excited parallel 4th formant, in dB
A5F  v    0     0    80  Amp of fric-excited parallel 5th formant, in dB
A6F  v    0     0    80  Amp of fric-excited parallel 6th formant, in dB
AB   v    0     0    80  Amp of fric-excited parallel bypass path, in dB
B2F  v   40   250  1000  Bw of fric-excited parallel 2nd formant, in Hz
B3F  v   60   300  1000  Bw of fric-excited parallel 3rd formant, in Hz
B4F  v  100   320  1000  Bw of fric-excited parallel 4th formant, in Hz
B5F  v  100   360  1500  Bw of fric-excited parallel 5th formant, in Hz
B6F  v  100  1500  4000  Bw of fric-excited parallel 6th formant, in Hz
ANV  v    0     0    80  Amp of voice-excited parallel nasal form., in dB
A1V  v    0    60    80  Amp of voice-excited parallel 1st formant, in dB
A2V  v    0    60    80  Amp of voice-excited parallel 2nd formant, in dB
A3V  v    0    60    80  Amp of voice-excited parallel 3rd formant, in dB
A4V  v    0    60    80  Amp of voice-excited parallel 4th formant, in dB
ATV  v    0     0    80  Amp of voice-excited par tracheal formant, in dB
```

```
Varied Parameters:
time   F0   AV
   0 1200   60
   5 1190   60
  10 1180   60
  15 1170   60
  20 1160   60
  25 1150   60
  30 1140   60
  35 1130   60
  40 1120   60
  45 1110   60
  50 1100   60
  55 1090   60
  60 1080   48
  65 1070   36
  70 1060   24
  75 1050   12
  80 1050    0
```

Appendix C: Individual speakers' F1 and F2 values of the three vowels /i, a, u/ produced in Experiment 1a

/i/ (normal speech condition)

Japanese subjects	Mean F1 (Hz)	sd
J1	296.00	15.35
J2	284.17	32.50
J3	263.59	3.35
J4	277.67	8.67
J5	303.53	5.12
J6	245.87	4.01
J7	267.09	4.67
J8	243.23	9.70
J9	267.90	16.65
J10	257.26	7.66

Greek subjects	Mean F1 (Hz)	sd
G1	265.45	12.20
G2	237.26	10.75
G3	252.73	3.62
G4	288.70	5.81
G5	248.45	13.05
G6	278.67	8.76
G7	284.33	9.01
G8	259.60	6.73
G9	288.59	15.69
G10	266.70	14.24

Japanese subjects	Mean F2 (Hz)	sd
J1	2113.30	66.76
J2	2153.06	32.19
J3	2263.56	17.82
J4	2162.22	24.85
J5	2407.94	27.12
J6	2265.56	29.86
J7	2271.98	13.77
J8	2241.79	26.56
J9	2159.73	42.26
J10	2225.74	47.47

Greek subjects	Mean F2 (Hz)	sd
G1	2291.29	24.7
G2	2254.93	84.32
G3	1864.62	38.68
G4	2206.03	35.12
G5	2193.13	30.88
G6	2178.98	24.52
G7	2233.84	31.16
G8	2170.90	39.71
G9	2077.42	39.80
G10	2391.33	20.58

/a/ (normal speech condition)

Japanese subjects	Mean F1 (Hz)	sd
J1	820.16	14.62
J2	900.82	49.30
J3	819.17	4.36
J4	686.00	40.97
J5	893.44	10.73
J6	726.08	22.16
J7	681.31	27.20
J8	805.41	21.88
J9	766.46	17.34
J10	777.08	11.06

Greek subjects	Mean F1 (Hz)	sd
G1	832.18	16.27
G2	788.86	17.30
G3	674.03	41.52
G4	811.80	36.08
G5	774.36	15.63
G6	772.59	4.49
G7	767.95	14.79
G8	794.12	20.83
G9	845.54	78.51
G10	837.77	11.06

/a/ (normal speech condition)

Japanese subjects	Mean F2 (Hz)	sd
J1	1264.34	68.19
J2	1346.29	30.84
J3	1204.07	25.21
J4	1236.10	31.32
J5	1242.49	46.48
J6	1065.53	29.03
J7	1177.17	48.91
J8	1258.14	17.54
J9	1163.32	42.75
J10	1073.19	13.16

Greek subjects	Mean F2 (Hz)	sd
G1	1167.32	27.17
G2	1138.61	19.26
G3	1090.22	24.01
G4	1285.20	48.93
G5	1306.04	16.82
G6	1177.04	26.42
G7	1271.14	12.56
G8	1115.10	61.32
G9	1180.37	24.64
G10	1343.16	24.36

/u/ (normal speech condition)

Japanese subjects	Mean F1 (Hz)	sd
J1	363.15	9.5
J2	302.82	10.27
J3	291.27	6.00
J4	313.15	21.06
J5	329.01	7.8
J6	268.31	10.92
J7	285.08	4.82
J8	270.82	6.35
J9	271.11	11.33
J10	275.11	11.32

Greek subjects	Mean F1 (Hz)	sd
G1	268.88	14.43
G2	256.48	9.68
G3	264.41	6.36
G4	283.16	16.66
G5	268.04	21.99
G6	286.70	12.80
G7	297.25	10.80
G8	281.26	14.87
G9	301.26	21.11
G10	275.26	5.40

Japanese subjects	Mean F2 (Hz)	sd
J1	1428.95	32.14
J2	1233.8	28.90
J3	1407.756	21.11
J4	1162.70	27.42
J5	1170.40	58.86
J6	1274.30	52.50
J7	1173.25	28.65
J8	1514.32	38.30
J9	1066.80	66.16
J10	858.18	31.28

Greek subjects	Mean F2 (Hz)	sd
G1	737.73	49.70
G2	707.01	23.85
G3	727.60	10.86
G4	789.75	42.96
G5	709.15	25.74
G6	704.44	34.11
G7	784.64	25.25
G8	697.64	36.43
G9	619.10	59.36
G10	749.65	16.42

/i/ (hyperarticulated speech condition)

Japanese subjects	Mean F1 (Hz)	sd
J1	306.73	9.9
J2	262.71	26.20
J3	280.90	9.08
J4	251.22	18.07
J5	295.85	29.08
J6	308.15	10.35
J7	317.48	14.30
J8	249.70	2.93
J9	253.09	8.07
J10	257.40	11.13

Greek subjects	Mean F1 (Hz)	sd
G1	284.91	8.85
G2	228.46	4.17
G3	277.21	17.67
G4	294.75	18.87
G5	255.95	14.27
G6	294.55	9.83
G7	285.65	12.47
G8	275.37	3.3
G9	295.25	11.91
G10	262.31	9.40

Japanese subjects	Mean F2 (Hz)	sd
J1	2149.83	47.73
J2	2215.23	44.63
J3	2302.68	17.90
J4	2232.25	25.93
J5	2561.24	60.10
J6	2375.14	26.47
J7	2236.22	37.06
J8	2257.51	13.60
J9	2170.84	16.93
J10	2216.42	33.03

Greek subjects	Mean F2 (Hz)	sd
G1	2240.02	18.31
G2	2349.67	64.05
G3	1992.96	16.14
G4	2206.98	97.93
G5	2327.48	24.76
G6	2188.27	34.83
G7	2303.26	19.40
G8	2221.99	19.96
G9	2087.01	24.14
G10	2410.85	21.88

/a/ (hyperarticulated speech condition)

Japanese subjects	Mean F1 (Hz)	sd
J1	863.96	7.67
J2	857.48	30.99
J3	797.53	22.82
J4	701.12	54.01
J5	882.81	29.93
J6	791.49	40.48
J7	755.12	8.86
J8	801.26	12.76
J9	766.87	7.97
J10	774.24	9.39

Greek subjects	Mean F1 (Hz)	sd
G1	857.40	44.34
G2	774.97	34.92
G3	748.26	20.60
G4	863.28	12.88
G5	802.12	9.98
G6	783.01	8.38
G7	806.75	13.25
G8	785.07	17.57
G9	772.70	73.59
G10	851.71	20.33

/a/ (hyperarticulated speech condition)

Japanese subjects	Mean F2 (Hz)	sd
J1	1335.07	47.85
J2	1312.26	23.90
J3	1185.21	29.80
J4	1221.68	34.13
J5	1196.27	19.78
J6	1170.39	54.69
J7	1249.10	33.36
J8	1314.92	47.67
J9	1191.39	25.45
J10	1073.83	21.38

Greek subjects	Mean F2 (Hz)	sd
G1	1175.98	21.36
G2	1113.80	18.03
G3	1114.86	11.79
G4	1272.38	27.62
G5	1286.62	7.77
G6	1189.65	35.56
G7	1285.28	20.20
G8	1107.28	15.72
G9	1150.96	34.83
G10	1330.36	30.32

/u/ (hyperarticulated speech condition)

Japanese subjects	Mean F1 (Hz)	sd
J1	409.00	39.93
J2	311.35	6.17
J3	313.77	9.37
J4	300.65	16.46
J5	321.51	12.82
J6	333.62	30.08
J7	342.30	18.70
J8	291.04	22.85
J9	261.78	13.59
J10	271.96	15.92

Greek subjects	Mean F1 (Hz)	sd
G1	284.90	9.28
G2	237.55	6.57
G3	283.63	12.24
G4	280.27	17.93
G5	297.19	11.55
G6	312.07	16.00
G7	295.77	4.37
G8	291.11	16.58
G9	288.14	18.51
G10	280.90	11.46

Japanese subjects	Mean F2 (Hz)	sd
J1	1396.85	58.17
J2	1234.50	34.76
J3	1418.01	36.97
J4	1069.29	56.28
J5	1108.35	27.64
J6	1336.68	28.33
J7	1283.26	66.89
J8	1533.36	23.25
J9	975.41	43.74
J10	840.29	20.44

Greek subjects	Mean F2 (Hz)	sd
G1	807.10	51.47
G2	769.71	28.97
G3	714.04	30.12
G4	751.72	23.93
G5	699.55	32.97
G6	749.54	23.37
G7	767.24	25.97
G8	707.50	22.67
G9	653.63	52.76
G10	748.53	19.94

Appendix D: Individual subjects' MOA choices in mels

/i/ (MOA choices)

Japanese subjects	Mean F1 (mels)	sd
J1	298.27	23.70
J2	350.11	23.63
J3	306.90	38.65
J4	350.08	38.65
J5	324.23	0
J6	289.61	19.35
J7	393.00	64.98
J8	444.48	46.75
J9	350.09	38.64
J10	324.23	0

Greek subjects	Mean F1 (mels)	sd
G1	332.83	36.14
G2	280.96	0
G3	341.49	23.63
G4	341.49	23.63
G5	367.37	0
G6	298.27	23.70
G7	280.96	0
G8	315.58	19.35
G9	358.68	47.26
G10	315.55	36.16

Japanese subjects	Mean F2 (mels)	sd
J1	1802.87	20.16
J2	1766.85	77.98
J3	1766.87	31.77
J4	1739.98	51.18
J5	1811.88	0
J6	1766.85	77.98
J7	1605.06	93.29
J8	1793.85	24.69
J9	1713.06	20.06
J10	1793.91	40.18

Greek subjects	Mean F2 (mels)	sd
G1	2240.02	18.31
G2	2349.67	64.05
G3	1992.96	16.14
G4	2206.98	97.93
G5	2327.48	24.76
G6	2188.27	34.83
G7	2303.26	19.40
G8	2221.99	19.96
G9	2087.01	24.14
G10	2410.85	21.88

/a/ (MOA choices)

Japanese subjects	Mean F1 (mels)	sd
J1	882.84	0
J2	908.74	23.64
J3	770.96	49.02
J4	822.62	23.75
J5	822.90	30.42
J6	865.70	57.82
J7	762.29	35.82
J8	900.12	57.87
J9	839.86	30.49
J10	882.95	43.02

Greek subjects	Mean F1 (mels)	sd
G1	900.12	57.87
G2	960.11	52.53
G3	796.74	19.07
G4	865.75	43.18
G5	874.26	38.46
G6	900.16	19.18
G7	908.69	38.49
G8	934.42	48.95
G9	882.85	35.86
G10	925.85	52.53

/a/ (MOA choices)

Japanese subjects	Mean F2 (mels)	sd	Greek subjects	Mean F2 (mels)	sd
J1	1218.11	20.15	G1	1164.03	68.35
J2	1136.96	31.77	G2	1802.87	20.16
J3	1028.97	24.47	G3	1064.99	24.85
J4	1101.11	49.34	G4	1416.15	212.02
J5	1100.99	37.61	G5	1245.08	24.59
J6	1119.07	60.41	G6	1145.98	49.31
J7	1137.16	63.58	G7	1155.05	40.21
J8	1200.08	24.68	G8	1290.01	75.26
J9	1118.96	51.33	G9	1217.98	66.69
J10	1137.08	44.93	G10	1191.01	49.44

/u/ (MOA choices)

Japanese subjects	Mean F1 (mels)	sd	Greek subjects	Mean F1 (mels)	sd
J1	375.96	19.22	G1	332.85	19.29
J2	358.74	19.29	G2	324.18	43.20
J3	384.52	38.35	G3	384.56	23.54.
J4	367.34	30.44	G4	367/34	30.44
J5	324.23	0	G5	358.74	19.29
J6	410.34	0	G6	341.46	38.64
J7	410.34	0	G7	289.61	19.35
J8	410.34	0	G8	298.27	23.70
J9	315.55	36.16	G9	350.05	49.18
J10	358.74	19.29	G10	306.92	23.70

Japanese subjects	Mean F2 (mels)	sd	Greek subjects	Mean F2 (mels)	sd
J1	1271.97	84.02	G1	804.24	68.27
J2	921.03	86.42	G2	696.06	20.12
J3	1028.96	68.20	G3	840.15	40.32
J4	939.11	129.44	G4	849.10	129.57
J5	921.12	66.70	G5	804.22	24.51
J6	1127.94	80.62	G6	831.25	58.61
J7	1146.06	186.45	G7	741.13	58.74
J8	1119.04	87.73	G8	759.24	24.82
J9	858.05	58.54	G9	858.23	92.13
J10	921.00	132.73	G10	723.12	37.76

Bibliography

Aaltonen, O., Eerola, O., Hellström, Å., Uusipaikka, E., & Lang, A.H. (1997). Perceptual magnet effect in the light of behavioral and psychological data. *Journal of the Acoustical Society of America*, **101(2)**, 1090-1105.

Abramson, A.S. & Lisker, L. (1967). Discriminability along the voicing continuum: cross-language tests. *Proceedings of the 6th International Congress of Phonetic Science, Prague*, 569-573.

Abramson, A.S. & Lisker, L. (1973). Voice timing perception in Spanish word-initial stops. *Journal of Phonetics*, **1**, 1-8.

Archangeli, D. (1984). *Underspecification in Yawelmani Phonology and Morphology,* Unpublished doctoral dissertation.

Bailey, P.J. & Haggard, M.P. (1973). Perception and production: Some correlations on voicing of an initial stop. *Language and Speech*, 189-195.

Bailey, P.J., Summerfield, Q., & Dorman, M.F. (1977). On the identification of sine-wave analogues of certain speech sounds. In *Status Report on Speech Research,* **SR-51/52**, 1-25. New Haven, Conn.: Haskins Laboratories.

Beebe, L.M. (1984). Myths about interlanguage phonology. *Studies in Descriptive Linguistics*, **13**, 51-61.

Beranek, L.L. (1949). *Acoustics*. NY: McGraw-Hill.

Best, C.T., Morrongielllo, B., & Robson, R. (1981). Perceptual equivalence of acoustic cues in speech and nonspeech perception. *Perception & Psychophysics*, **29** **(3)**, 191-211.

Best, C.T. (1995). A direct realist view of cross-language speech perception. In W. Strange (Ed.), *Speech Perception and Linguistic Experience: Issues in Cross-language Speech Research* (pp. 171-201). Timonium MD: York Press, Inc.

Best, C.T., MacKain, K.S., & Strange, W. (1982). Across-language study of categorical perception for semivowel and liquid contrasts, *Paper presented at the 103rd meeting of the Acoustical Society of America.*

Best, C.T., McRoberts, G.W., & Sithole, N.N. (1988). Examination of perceptual reorganization for nonnative speech contrasts: Zulu click discrimination by English-speaking adults and infants. *Journal of Experimental Psychology: Human Perception & Performance*, **14**, 345-360.

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

Bloch, B. (1950). Studies in colloquial Japanese IV: Phonemics. *Language*, **26**, 86-125.

Bohn, O.-S. (1995a). Cross-language speech perception in adults: First language transfer doesn't tell it all. In W. Strange (Ed.), *Speech Perception and Linguistic Experience: Issues in Cross-Language Speech Research* (pp. 279-304). Timonium MD: York Press.

Bohn, O.-S. (1995b). What determines the perceptual difficulty encountered in the acquisition of nonnative contrasts?. *Proceedings of the 13th International Congress of Phonetic Sciences, Stockholm*, **4**, 84-91.

Bohn, O.-S. & Flege, J.E. (1990a). Perception and production of a new vowel category by adult second language learners. *New Sounds 90: Proceedings of the 1990 Amsterdam Symposium on the Acquisition of Second-Language Speech*, 37-56.

Bohn, O.-S. & Flege, J.E. (1990b). Interlingual identification and the role of foreign language experience in L2 vowel perception. *Applied Psycholinguistics*, **11**, 303-328.

Bradlow, A.R. (1993). *Language-Specific and Universal Aspects of Vowel Production and Perception: A Cross-Linguistic Study of Vowel Inventories*. Ithaca: DMLL Publications.

Bréire, E. (1966). An investigation of phonological interference. *Language*, **42**, 768-796.

Bréire, E. (1968). *A Psycholinguistic Study of Phonological Interference*. The Hague: Mouton.

Broad, D. & Wakita, H. (1977). Piecewise planar representation of vowel formant frequencies. *Journal of the Acoustical Society of America*, **62**, 1467-1473.

Burnham, D.K. (1986). Developmental loss of speech perception: Exposure to and experience with a first language. *Applied Psycholinguistics*, **7**, 207-240.

Caramazza, A. & Yeni-Komshian, G. (1974). Voice onset time in two French dialects. *Journal of Phonetics*, **2**, 239-245.

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

Carney, A.E., Widin, G.P., & Veimeister, N.F. (1977). Noncategorical perception of stop consonants differing in VOT. *Journal of the Acoustical Society of America*, **62**, 961-970.

Chistovich, I.A., Sheikin, R.L., & Lublinskaya, V.V. (1979). The "center of gravity" effect in vowel spectra and critical distance between the formants: Psychoacoustical study of the perception of vowel like stimuli. *Hearing Research*, **1**, 143-157.

Corder, S.P. (1967). The significance of learners' errors. *IRAL*, **5**, 161-170.

Corder, S.P. (1978). Language-learner language. In J.C. Richards (Ed.), *Understanding Second & Foreign Language Learning: Issues & Approaches* (pp. 71-93). Rowley, Massachusetts: Newbury House Publishers, Inc.

Corder, S.P. (1981). *Error Analysis and Interlanguage*. Oxford: Oxford University Press.

Cowan, N. & Morse, P.A. (1986). The use of auditory and phonetic memory in vowel discrimination. *Journal of the Acoustical Society of America*, **79**, 500-507.

Crothers, J. (1978). Typology and universals of vowel systems. In C.A. Greenberg, C.A. Ferguson, & E.A. Moravcsik (Eds.), *Universals of Human Language, Vol. 2: Phonology* (pp. 93-152). Stanford: Stanford University Press.

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

Cutting, J.E. & Rosner, B.S. (1974) Discrimination functions predicted from categories in speech and music. *Perception & Psychophysics,* **20**, 87-88.

Debrock, M. & Forrez, G. (1976). Analyse methématique des voyelles orales du néerlandais et du français, Méthode et resultats. *Revue de Phonétique Appliquée*, **37**, 27-73.

Dickerson, L.J. (1975). The learner's interlanguage as a system of variable rules. *TESOL Quarterly*, **9**, 401-407.

Disner, S. (1983). Vowel quality: The relation between universal and language-specific factors. *UCLA Working Papers in Phonetics*, **58.**

Dooley, G.J. & Moore, B.C.J. (1988). Detection of linear frequency glides as a function of frequency and duration. *Journal of the Acoustical Society of America*, **84**, 2045-2057.

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

Eckman, F.R. (1977). Markedness and the contrastive analysis hypothesis. *Language Learning*, **24**, 37-53.

Eckman, F.R. (1984). Universals, typologies, and interlanguage. In W. Rutherford (Ed.), *Language Universals and Second Language Acquisition* (pp. 79-105). Amsterdam: John Benjamins.

Eckman, F.R. (1996). A functional-typological approach to second language acquisition theory. In W.C. Ritchie & T.K. Bhatia (Eds.), *Handbook of Second Language Acquisition* (pp. 195-211). San Diego: Academic Press.

Eimas, P.D. (1963). The relation between identification and discrimination along speech and non-speech continua. *Language and Speech*, **6**, 206-217.

Elman, J.L., Diehl, R.L., & Buchwald, S.E. (1977). Perceptual switching in bilinguals. *Journal of the Acoustical Society of America*, **62**, 971-974.

Elsendoorn, B.A.G. (1984). Production and perception of English vowel duration by Dutch speakers of English. *Proceedings of the 10th International Congress of Phonetic Sciences*, 673-676.

Fant, G. (1973). *Speech Sounds and Features*. Cambridge, Massachusetts: The MIT Press.

Flanagan, J. (1955). A difference limen for vowel formant frequency. *Journal of the Acoustical Society of America*, **27(3)**, 288-291.

Flege, J.E. (1984). The detection of French accent by American listeners. *Journal of the Acoustical Society of America*, **76(3)**, 692-707.

Flege, J.E. (1986). Effects of equivalence classification on the production of foreign language speech sounds. In A. James & J. Leather (Eds.), *Sound Patterns in Second Language Acquisition* (pp. 9-39). Dordrecht: Foris Publications.

Flege, J.E. (1987a). A critical period for learning to pronounce foreign languages? *Applied Linguistics*, **8**, 162-77.

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

Flege, J.E. (1991). Perception and production: The relevance of phonetic input to L2 phonological learning. In T. Hueber & C. Ferguson (Eds.), *Crosscurrents in Second Language Acquisition and Linguistic Theories* (pp. 249-289). Amsterdam: John Benjamins.

Flege, J.E. (1993). Production and perception of a novel, second-language phonetic contrast. *Journal of the Acoustical Society of America*, **93,3**, 1589-1608.

Flege, J.E. (1995). Second language speech learning theory, findings, and problems. In W. Strange (Ed.), *Speech Perception and Linguistic Experience: Issues in Cross-language Speech Research* (pp. 233-277). Timonium, MD: York Press, Inc.

Flege, J.E. & Davidian, R. (1985). Transfer and developmental processes in adult foreign language speech production. *Journal of Applied Psycholinguistics Research*, **5**, 323-347.

Flege, J.E. & Eefting, W. (1987). Cross-language switching in stop consonant perception and production by Dutch speakers of English. *Speech Communication*, **6,3**, 185-202.

Flege, J.E. & Hammond, R. (1982). Mimicry of non-distinctive phonetic differences between language varieties. *Studies in Second Language Acquisition*, **5**, 1-18.

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

Fox, R.A. (1982). Individual variation in the perception of vowels: Implications for a perception-production link. *Phonetica*, **39**, 1-22.

Frieda, E.M. (1997). Comparison of native speakers' perception and production of the English vowel /i/. *Journal of the Acoustical Society of America*, 3110.

Frieda, E. M., Walley, A.C., Flege J. E. & Sloane, M. E. (in press). Adults' perception and production of the English vowel /i/. *Journal of Speech and Hearing Research.*

Fry, D.B., Abramson, A.S., Eimas, P.D. & Liberman, A.M. (1962). The identification and discrimination of synthetic vowels. *Language and Speech*, **5**, 171-189.

Fujisaki, H. & Kawashima, T. (1969). On the modes and mechanisms of speech perception. *Annual Report of the Engineering Research Institute, Faculty of Engineering, University of Tokyo*, **28**, 67-73.

Fujisaki, H. & Kawashima, T. (1970). Some experiments on speech perception and a model for the perceptual mechanism. *Annual Report of the Engineering Research Institute, Faculty of Engineering, University of Tokyo*, **29**, 207-214.

Galama, D.A. & Nearey, T.M. (1995). Competition and categorization: The effect of choice set on vowel category structure. *Paper presented at 1995 ASA meeting.*

Ganong, W.F. & Zatorre, R.J. (1980). Measuring phoneme boundaries four ways. *Journal of the Acoustical Society of America*, **68**, 431-439.

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

Gillette, S. (1980). Contextual variation in the perception of L and R by Japanese and Korean speakers. *Minnesota Papers in Linguistics and the Philosophy of Language*, **6**, 59-72.

Goldman, D. & Homa, D. (1977). Integrative and metric properties of abstracted information as a function of category discriminability, instance variability, and experience. *Journal of Experimental Psychology: Human Learning & Memory*, **3(4)**, 375-385.

Gordon, P.C. & Myer, D.E. (1984). Perceptual-motor processing of phonetic features in speech. *Journal of Experimental Psychology: Human Perception & Performance*, **10**, 153-178.

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

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

Grieser, D. & Kuhl, P.K. (1989). Categorization of speech by infants: Support for speech-sound prototypes. *Developmental Psychology*, **25(4)**, 577-588.

Gulick, W.L., Gescheider, G.A., & Frisina, R.D. (1989). Pitch and loudness. In *Hearing: Physiological Accounts, Neural Coding, and Psychoacoustics* (pp. 246-282). Oxford: Oxford University Press.

Hatch, E. (1978). Acquisition of syntax in a second language. In J.C. Richards (Ed.), *Understanding Second and Foreign Language Learning: Issues and Approaches* (pp. 34-70). Rowley, Massachusetts: Newbury House.

Hawks, J.W. & Fourakis, M.S. (1995). The perceptual vowel spaces of American English and Modern Greek: A comparison. *Language and Speech*, **38(3)**, 237-252.

Hecht, B.F. & Mulford, R. (1982). The acquisition of a second language phonology: Interaction of transfer and developmental factors. *Applied Psycholinguistics*, **3**, 313-328.

Heffner, R.-M.S. (1950). *General Phonetics*. Madison: University of Wisconsin Press.

Hintzman, D.L. (1988). Judgements of frequency and recognition memory in a multiple-trace memory model. *Psychological Review*, **95**, 528-551.

Hoffman, M.A.B. (1973). Focal vowels. *Paper presented at the 48th meeting of the Linguistic Society of America, San Diego, CA.*

Holmes, F. (1995). Perception of vowel quality by German-English bilinguals. *Proceedings of the 13th International Congress of Phonetic Sciences, Stockholm*, **3**, 624-627.

Hyltenstam, K. (1984). The use of typological markedness conditions as predictors in second language acquisition: The case of pronominal copies in relative clauses. In R. Andersen (Ed.), *Second Languages: A Cross-Linguistic Perspective* Rowley, MA: Newbury House.

Imaishi, M., Sato, K., Miwa, J., Yoshida, N., Ohashi, K., & Kato, M. (1984). Nihongo hoogen onsei no supekutoru bunseki shiryoushuu. In Ministry of Education (Ed.), *Jyouhouka shakai ni okeru gengo no hyoujunka* (pp. 1-109). Ministry of Education.

Ioup, G. & Weinberger, S. (1987). *Interlanguage Phonology: The Acquisition of a Second Language Sound System.* Cambridge, Massachusetts: Newbury House.

Iverson, P. & Kuhl, P.K. (1995). Mapping the perceptual magnet effect for speech using signal detection theory and multidimensional scaling. *Journal of the Acoustical Society of America*, **97(1)**, 553-562.

James, A.R. & Leather, J. (1987). *Sound Patterns in Second Language Acquisition.* Dordrecht: Foris Publications.

Jamieson, D.G. (1995). Techniques for training difficult non-native speech contrasts. *Proceedings of the 13th International Congress of Phonetic Sciences, Stockholm*, **4**, 100-107.

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

Johansson, F.A. (1973). *Immigrant Swedish Phonology: A Study in Multiple Contact Analysis.* Lund, Sweden: CWK Gleerup.

Johnson, K. (1989). On the perceptual representation of vowel categories. *Research on Speech Perception Progress Report*, **15**, 343-358.

Johnson, K., Flemming, E., & Wright, R. (1993). The hyperspace effect: phonetic targets are hyperarticulated. *Language*, **69(3)**, 505-528.

Jones, D. (1950). *The Pronunciation of English.* Cambridge: Cambridge University Press.

Jongman, A., Fourakis, M., & Sereno, J.A. (1989). The acoustic vowel space of Modern Greek and German. *Language and Speech*, **32(3)**, 221-248.

Joseph, B.D. & Philippaki-Warburton, I. (1987). *Modern Greek.* Kent: Croom Helm Ltd.

Jusczyk, P.W. (1993). From general to language-specific capacities: the WRAPSA model of how speech perception develops. *Journal of Phonetics*, **21**, 3-28.

Kanamori, Y., Kasuya, H., Arai, S., & Kido, K. (1971). Effect of context on vowel perception. *Seventh International Congress on Acoustics, Budapest*, **20C5**, 37-40.

Kawakami, S. (1977). *Nihon Onsei Gaisetsu.* Tokyo: Oofuusha.

Keating, P.A. (1988). The window model of coarticulation: Articulatory evidence. *UCLA Working Papers in Phonetics*, **69**, 3-29.

Kelly, J. (1966). A note on lip-rounding. *Le Maître Phonétique*, **125**, 8-9.

Kinnear, P.R. & Gray, C.D. (1997). *SPSS for Windows Made Simple*. East Sussex: Psychology Press.

Klatt, D.H. (1980). Software for a cascade/parallel formant synthesizer. *Journal of the Acoustical Society of America*, **67**, 971-995.

Klatt, D.H. (1990). Description of the cascade/parallel formant synthesizer. In *Klatttalk: The Conversion of English Text to Speech* .

Krashen, S.D. (1978). The monitor model for second language acquisition. In R.C. Gingras (Ed.), *Second-Language Acquisition and Foreign Language Teaching* (pp. 1-26). Arlington, VA: Center for Applied Linguistics.

Krashen, S.D. (1981). *Second Language Acquisition and Second Language Learning*. Oxford: Pergamon.

Kuhl, P.K. (1986). Reflections on infants' perception and representation of speech. In J.S. Perkell & D.H. Klatt (Eds.), *Invariance and Variability in Speech Processes* (pp. 19-30). Hillsdale, NJ: Erlbaum.

Kuhl, P.K. (1991). Human adults and human infants show a "perceptual magnet effect" for the prototypes of speech categories, monkeys do not. *Perception & Psychophysics*, **50(2)**, 93-107.

Kuhl, P.K. (1993). Early linguistic experience and phonetic perception: implications for theories of developmental speech perception. *Journal of Phonetics*, **21**, 125-139.

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

Lacerda, F. (1993). Sonority contrasts dominate young infants' vowel perception. *Journal of the Acoustical Society of America,* **93,** 2372.

Ladd, D.R. & Morton, R. (1997). The perception of intonational emphasis: continuous or categorical? *Journal of Phonetics,* **25,** 313-342.

Ladefoged, P. (1971). *Preliminaries to Linguistic Phonetics*. Chicago: University of Chicago Press.

Ladefoged, P. (1975). *A Course in Phonetics*. NY: Harcourt Brace Jovanovich.

Ladefoged, P. (1996). *Elements of Acoustic Phonetics*. Chicago: The University of Chicago Press.

Ladefoged, P. & Broadbent, D.E. (1957). Information conveyed by vowels. *Journal of the Acoustical Society of America*, **29**, 98-104.

Lado, R. (1957). *Linguistics Across Cultures: Applied Linguistics for Language Teachers*. Ann Arbor, Michigan: The University of Michigan Press.

Leather, J. & James, A. (1996). Second language speech. In W.C. Ritchie & T.K. Bhatia (Eds.), *Handbook of Second Language Acquisition* (pp. 269-316). San Diego: Academic Press.

Leather, J.H. (1988). *Speech Pattern Elements in Second Language Acquisition,* Unpublished doctoral dissertation, University College, London.

Liberman, A.M., Cooper, F.S., Shankweilier, D.P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review*, **74(6)**, 431-461.

Liberman, A.M., Harris, K.S., Hoffman, H.S., & Griffith, B.C. (1957). The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology*, **54**, 358-368.

Liljencrants, J. & Lindblom, B. (1972) Numerical simulation of vowel quality systems: the role of perceptual contrast. *Language,* **48(4)**, 839-862.

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

Lively, S.E. (1993). An examination of the "perceptual magnet" effect. *Journal of the Acoustical Society of America*, 2423.

Lively, S.E. & Pisoni, D.B. (1997). On prototypes and phonetic categories: a critical assessment of the perceptual magnet effect in speech perception. *Journal of Experimental Psychology: Human Perception & Performance*, **23(6)**, 1665-1679.

Long, M.H. (1990). Maturational constraints on language development. *Studies in Second Language Acquisition*, **12,3**, 251-285.

Lotto, A.J., Kluender, K.R., & Holt, L.L. (1996). Effects of language experience on perceptual organization of vowel sounds. *Paper presented at the 5th Conference on Laboratory Phonology.*

Luce, R.D. (1963). Detection and recognition. In R.D. Luce, R.R. Bush, & E. Galanter (Eds.), *Handbook of Mathematical Psychology, Vol. 1* (pp. 103-189). NY: John Wiley and Sons, Inc.

Mack, M. (1989). Consonant and vowel perception and production: early English-French bilinguals and English monolinguals. *Perception & Psychophysics*, **46.2**, 187-200.

MacKain, K., Best, C. & Strange, W. (1981). Categorical perception of English /r/ and /l/ by Japanese bilinguals. *Applied Psycholinguistics,* **2,** 369-390.

Mackridge, P. (1985). *The Modern Greek Language: A Descriptive Analysis of Standard Modern Greek.* Oxford: Oxford University Press.

Macmillan, N.A. (1993). Signal detection theory as data analysis method and psychological decision model. In G. Keren & C. Lewis (Eds.), *A Handbook for Data Analysis in the Behavioral Sciences: Methodological Issues* (pp. 21-58). Hillsdale, New Jersey: Lawrence Erlbaum Associates, Inc., Publishers.

Macmillan, N.A. & Creelman, C.D. (1991). *Detection Theory: A User's Guide.* NY: Cambridge University Press.

Macmillan, N.A., Goldberg, R.F., & Braida, L.D. (1988). Resolution for speech sounds: basic sensitivity and context memory on vowel and consonant continua. *Journal of the Acoustical Society of America*, **84**, 1262-1280.

Maddieson, I. (1984). Vowels. In *Patterns of Sounds* Cambridge: Cambridge University Press.

Major, R. (1986). Paragoge and degree of foreign accent in Brazilian English. *Second Language Research*, **2**, 53-69.

Major, R. (1990). L2 acquisition, L1 loss, and the critical period hypothesis. *New Sounds 90: Proceedings of the 1990 Amsterdam Symposium on the Acquisition of Second-Language Speech*, 207-224.

Makhoul, J. & Cosell, L. (1976). LPCW: An LPC vocoder with linear predictive spectral warping. *Proceedings of ICASSP*, 466-469.

Massaro, D.W. & Cohen, M.M. (1977). The contribution of voice-onset time and fundamental frequency as cues to the /si/-/zi/ distinction. *Journal of the Acoustical Society of America*, **22**, 373-382.

Massaro, D.W. & Oden, G.C. (1980a). Speech perception: a framework for research and theory. In N. Lass (Ed.), *Speech and Language: Advances in basic research and practice, Vol. 3* (pp.129-165). New York: Academic Press.

Massaro, D.W. & Oden, G.C. (1980b). Evaluation and integration of acoustic features in speech perception. *Journal of the Acoustical Society of America*, **67**, 996-1013.

Medin, D.L. & Barsalou, L.W. (1987). Categorization processes and categorical perception. In S. Harnad (Ed.), *Categorical Perception: The Groundwork of Cognition* (pp. 455-490). New York: Cambridge University Press.

Mervis, C.B. & Rosch, E. (1981). Categorization of natural objects. *Annual Review of Psychology*, **32**, 89-115.

Miller, J.L. (1977). Properties of feature detectors for VOT: the voiceless channel of analysis. *Journal of the Acoustical Society of America*, **62**, 641-648.

Miller, J.L., Connine, C.M., Schermer, T.M., & Kluender, K.R. (1983). A possible auditory basis for internal structure of phonetic categories. *Journal of the Acoustical Society of America*, **73(6)**, 2124-2133.

Miller, J.L. & Volaitis, L.E. (1989). Effect of speaking rate on the perceptual structure of a phonetic category. *Perception & Psychophysics*, **46**, 505-512.

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 & Psychophysics*, **18**, 331-340.

Mochizuki, M. (1981). The identification of /r/ and /l/ in natural and synthesized speech. *Journal of Phonetics*, **9**, 283-303.

Moon, S.-J. & Lindblom, B.E.F. (1989). Formant Undershoot in clear and citation-form speech: a second progress report. *Transmission Laboratory-Quarterly Progress Status Report*, **1**, 121-123.

Morosan, D.E. & Jamieson, D.G. (1989). Evaluation of a technique for training new speech contrasts: generalization across voices, but not word-position or task. *Journal of Speech and Hearing Research*, **32**, 501-511.

Moulton, W. (1962). *The Sounds of English and German*. Chicago: University of Chicago Press.

Mueller, T.H. & Niedzielski, H. (1968). The influence of discrimination training on pronunciation. *Modern Language Journal*, **52**, 410-416.

Nakamura, M. & Kindaichi, H. (1955). Tokyogo. *Kokugogakkai*, 680-697.

Nathan, G. (1987). On second-language acquisition of voiced stops. *Journal of Phonetics*, **15**, 313-322.

Nathan, G. (1990). On the non-acquisition of an English sound system. *New Sounds 90: Proceedings of the 1990 Amsterdam Symposium on the Acquisition of Second-Language Speech*, 294-299.

Nearey, T.M. (1976). *Features for vowels*. Unpublished Ph.D. dissertation, University of Connecticut, Storrs.

Nearey, T.M. (1978). *Phonetic Feature for Vowels*. Indiana: Indiana University Linguistics Club.

Nearey, T.M. (1989). Static, dynamic, and relational properties in vowel perception. *Journal of the Acoustical Society of America*, **85 (5)**, 2088-2113.

Nemser, W. (1971). *An Experimental Study of Phonological Interference in the English of Hungarians*. Bloomington: Indiana University Publications.

Newman, R.S. (1996). Individual differences and the link between speech perception and speech production: A first report. *Speech Research Status Report, 5, State University of New York at Buffalo*, **5**.

Nooteboom, S.G. (1973). The perceptual reality of some prosodic durations. *Journal of Phonetics*, **1**, 25-46.

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.

Nosofsky, R.M., Clark, S.E., & Shin, H.-J. (1989). Rules and exemplars in categorization, identification, and recognition. *Journal of Experimental Psychology: Learning, Memory and Cognition*, **15**, 282-304.

Oden, G.C. & Massaro, D.W. (1978). Integration of featural information in speech perception. *Psychological Review*, **85**, 172-191.

Peterson, G. & Barney, H. (1952). Control methods used in a study of the vowels. *Journal of the Acoustical Society of America*, **24(2)**, 175-184.

Picheny, M.A., Durlach, N.I., & Braida, L.D. (1986). Speaking clearly for the hard of hearing II: Acoustic characteristics of clear and conversational speech. *Journal of Speech and Hearing Research*, **29**, 434-446.

Pimsleur, P. (1963). Discrimination training in the teaching of French pronunciation. *Modern Language Journal*, **47**, 190-203.

Pisoni, D.B. (1971). *On the Nature of Categorical Perception of Speech Sounds.* Doctoral thesis. University of Michigan.

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

Pisoni, D.B. (1975). Auditory short-term memory and vowel perception. *Memory and Cognition*, **3 (1)**, 7-18.

Pisoni, D.B. (1977). Identification and discrimination of the relative onset of two component tones: implications for the perception of voicing in stops. *Journal of the Acoustical Society of America,* **61**, 1352-1361.

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

Pisoni, D.B. & Lively, S.E. (1995). Methodological issues in training listeners to perceive nonnative contrasts. In W. Strange (Ed.), *Speech Perception and Linguistic Experience: Issues in Cross-Language Research* (pp. 433-459). Timonium, MD: York Press, Inc.

Pisoni, D.B., Logan, J.S., & Lively, S.E. (1990). Perceptual learning of nonnative speech contrasts: Implications for theories of speech perception. In H.C. Nusbaum & J. Goodman (Eds.), *Development of speech perception: The transition from recognizing speech sounds to spoken words.* Cambridge, MA: MIT Press.

Pisoni, D.B. & Luce, P. (1987). Acoustic-phonetic representations in word recognition. *Cognition,* **25**, 21-52.

Pisoni, D.B. & Tash, J. (1974). Reaction times to comparisons within and across phonetic categories. *Perception & Psychophysics*, **15**, 285-290.

Polivanov, E.D. (1974). A.A. Leontév (Ed.), *E.D.Polivanov: Selected works* (pp. 223-237). The Hague: Mouton.

Polka, L. (1989). Cross-language speech perception in adults: phonemic, phonetic, and acoustic contributions. *Journal of the Acoustical Society of America,* **89(6),** 2961-77.

Polka, L. & Bohn, O.-S. (1996). A cross-language comparison of vowel perception in English-learning and German-learning infants. *Journal of the Acoustical Society of America*, **100(1)**, 577-592.

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

Pols, L.C.W., van der Kamp, L.J.T., & Plomp, R. (1969). Perceptual and physical space of vowel sounds. *Journal of the Acoustical Society of America*, **46**, 458-467.

Repp, B.H. (1977). Dichotic competition of speech sounds: The role of acoustic stimulus structure. *Journal of Experimental Psychology: Human Perception & Performance*, **3**, 37-50.

Repp, B.H. (1984). Categorical perception: issues, methods, findings. In *Speech & Language: Advances in Basic Research and Practice, Vol. 10* (pp. 243-335). New York: Academic Press, Inc.

Repp, B.H. & Crowder, R.G. (1990). Stimulus order effects in vowel discrimination. *Journal of the Acoustical Society of America*, **88(5)**, 2080-2090.

Repp, B.H., Healy, A.F., & Crowder, R.G. (1979). Categories and context in the perception of isolated steady-state vowels. *Journal of Experimental Psychology: Human Perception & Performance*, **5(1)**, 129-145.

Repp, B.H. & Liberman, A.M. (1987). Phonetic category boundaries are flexible. In S. Harnad (Ed.), *Categorical Perception: The Ground work of Cognition* (pp. 89-112). Cambridge: Cambridge University Press.

Richards, J.C. (1978). Introduction: understanding second and foreign language learning. In J.C. Richards (Ed.), *Understanding Second & Foreign Language Learning: Issues & Approaches* (pp. 1-14). Rowley, Massachusetts: Newbury House Publishers, Inc.

Ritchie, W.C. & Bhatia, T.K. (1996). Second language acquisition: introduction, foundations and overview. In W.C. Ritchie & T.K. Bhatia (Eds.), *Handbook of Second Language Acquisition* (pp. 1-46). San Diego: Academic Press.

Rochet, B.L. (1995). Perception and production of L2 speech sounds by adults. In W. Strange (Ed.), *Speech Perception and Linguistic Experience: Issues in Cross-Language Speech Research* (pp. 379-410). Timonium, MD: York Press, Inc.

Rosch, E. (1975). Cognitive reference points. *Cognitive Psychology*, **7**, 532-547.

Sakuma, K. (1973). *Hyoojun Nihongo no Hatsuon, Akusento,* expanded ed. Tokyo: Kooseisha Kooseikaku.

Samuel, A.G. (1977). The effect of discrimination training on speech perception: noncategorical perception. *Perception & Psychophysics*, **22**, 321-330.

Samuel, A.G. (1982). Phonetic prototypes. *Perception & Psychophysics*, **31**, 307-314.

Scholes, R.J. (1967). Phoneme categorization of synthetic vocalic stimuli by speakers of Japanese, Spanish Persian, and American English. *Language and Speech*, **10**, 46-68.

Scholes, R.J. (1968a). Phonemic interference as a perceptual phenomenon. *Language and Speech*, **11**, 86-103.

Scholes, R.J. (1968b). Perceptual categorization of synthetic vowels as a tool in dialectology and typology. *Language and Speech*, **11**, 194-207.

Schouten, M.E.H. (1975). *Native Language Interference In the Perception of Second Language Vowels,* Doctoral dissertation, University of Utrecht.

Schouten, M.E.H. & van Hessen, A.J. (1992). Modeling phoneme perception I: categorical perception. *Journal of the Acoustical Society of America*, **92(4)**, 539-552.

Sciarone, A.G. (1970). Contrastive analysis: possibilities and limitations. *International Review of Applied Linguistics*, **8**, 115-131.

Seliger, H.W. (1978). Implications of a multiple critical periods hypothesis for second language learning. In W.C. Ritchie (Ed.), *Second Language Acquisition Research* (pp. 11-19). NY: Academic Press.

Selinker, L. (1972). Interlanguage. *IRAL*, **1974**, 31-54.

Sheldon, A. (1985). The relationship between production and perception of the /r/-/l/ contrast in Korean adults learning English. A reply to Borden, Gerber and Milsark. *Language Learning*, **35,1**, 107-113.

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

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

Smith, E.E. & Medin, D.L. (1981). *Concepts and categories*. Cambridge, MA: Harvard University Press.

Stern, H.H. (1983). *Fundamental Concepts of Language Teaching*. Oxford: Oxford University Press.

Stevens, K.N. (1989). On the quantal nature of speech. *Journal of Phonetics*, **17**, 3-45.

Stevens, K.N. & House, A.S. (1963). Perturbation of vowel articulations by consonantal context: an acoustical study. *Journal of Speech and Hearing Research*, **6**, 111-128.

Stevens, K.N., Liberman, A.M., Studdert-Kennedy, M., & Ohman, S.E.G. (1969). Cross-language study of vowel perception. *Language and Speech*, **12**, 1-23.

Stockwell, R. & Bowen, R.D. (1965). *The Sounds of English and Spanish*. Chicago: University of Chicago Press.

Strange, W. (1995). Phonetics of second-language acquisition: past, present, future. *Proceedings of the 13th International Congress of Phonetic Sciences, Stockholm*, **4**, 76-83.

Strange, W. Trent, S.A., Stack, J.W., X. Ling & Rodriguez, A.I. (1993). Consonant context affects perceived similarity of North German and American English vowels. *Journal of the Acoustical Society of America*, **94**, 1866.

Sussman, J.E. & Lauckner-Morano, V.J. (1995). Further test of the "perceptual magnet effect" in the perception of [i]: identification and change/no-change discrimination. *Journal of the Acoustical Society of America*, **97(1)**, 539-552.

Sweet, H. (1890). *A Premier of Phonetics*. Oxford: Oxford University Press.

Syrdal, A. & Gopal, H.S. (1986). A perceptual model of vowel recognition based on auditory representations of American English vowels. *Journal of the Acoustical Society of America*, **79**, 1086-1100.

Tarone, E.E. (1978). The phonology of interlanguage. In J.C. Richards (Ed.), *Understanding Second & Foreign Language Learning: Issues & Approaches* (pp. 15-33). Rowley, Massachusetts: Newbury House Publishers, Inc.

Tees, R.C. & Werker, J.F. (1984). Perceptual flexibility: Maintenance or recovery of the ability to discriminate nn-native speech sounds. *Canadian Journal of Psychology*, **38**, 579-590.

Traunmüller, H. (1981). Perceptual dimension of openess in vowels. *Journal of the Acoustical Society of America*, **69**, 1465-1475.

Trubetzkoy, N.S. (1939). *Principles of Phonology.* Translated by C.A. Baltaxe, Berkeley, CA: University of California Press.

Tsujimura, N. (1996). *An Introduction to Japanese Linguistics*. Cambridge, MA: Blackwell Publishers Inc.

Valdman, A. (1976). *Introduction to French Phonology and Morphology*. Rowley, MA: Newbury House.

Vance, T. (1987). *An Introduction to Japanese Phonology.* Albany: State University of New York Press.

Verhoeven, J.W.M. (1992). *Perceptual Aspects of Dutch Intonation.* Doctoral thesis. University of Edinburgh.

Weinberger, S. (1990). Minimal segments in L2 phonology. *New Sounds 90: Proceedings of the 1990 Amsterdam Symposium on the Acquisition of Second-Language Speech*, 137-179.

Weiss, W. (1992). Perception and production in accent training. *Llisterri 95*, **102**, 69-81.

Werker, J.F., Gilbert, J.V.H., Humphrey, K., & Tees, R.C. (1981). Developmental aspects of cross-language speech perception. *Child Development*, **52**, 349-355.

Werker, J.F. & Logan, J.S. (1985). Cross-language evidence for three factors in speech perception. *Perception & Psychophysics*, **37**, 35-44.

Werker, J.F. & Tees, R.C. (1983). Developmental changes across childhood in the perception of non-native speech sounds. *Canadian Journal of Psychology*, **37**, 278-286.

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

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

Williams, L. (1977). The perception of stop consonant voicing by Spanish-English bilinguals. *Perception & Psychophysics*, **21**, 289-297.

Winer, B.J. (1971). *Statistical Principles in experimental Design.* New York: McGraw-Hill Book Company.

Wode, H. (1976). Developmental sequences in naturalistic second language acquisition. *Working papers in bilingualism*, **11**, 1-31.

Yamada, R.A., Tohkura, Y., Bradlow, A.R., & Pisoni, D.B. (1995). Effects of perceptual learning on the production of a non-native contrast. *Journal of the Acoustical Society of America*, **97(5)**, 3418.

Zwicker, E. & Terhardt, E. (1980). Analytical expressions for critical-band rate and critical bandwidth as a function of frequency. *Journal of the Acoustical Society of America*, **68**, 1523-1525.