HOW ARE SECOND LANGUAGE PHONEME CONTRASTS LEARNED?

by

Rachel L. Hayes

A Dissertation Submitted to the Faculty of the

DEPARTMENT OF LINGUISTICS

In Partial Fulfillment of the Requirements

For the Degree of

DOCTOR OF PHILOSOPHY

In the Graduate College

THE UNIVERSITY OF ARIZONA

2003

UMI Number: 3106997

ProQuest Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

THE UNIVERSITY OF ARIZONA ®
GRADUATE COLLEGE

As members of the Final Examination Committee, we certify that we have read the dissertation prepared by Rachel L. Hayes

entitled How are second language phoneme contrasts learned?

and recommend that it be accepted as fulfilling the dissertation requirement for the Degree of Doctor of Philosophy

Michael Hammond	5/19/03 Date
Natasha Warner	5/19/03 Date
Janet Nicol	5/19/03 Date
	Date
	Date

Final approval and acceptance of this dissertation is contingent upon the candidate's submission of the final copy of the dissertation to the Graduate College.

I hereby certify that I have read this dissertation prepared under my direction and recommend that it be accepted as fulfilling the dissertation requirement.

Dissertation Director	Michael Hammond Natasha Warner	5/19/03 Date

STATEMENT BY AUTHOR

This dissertation has been submitted in partial fulfillment of requirements for an advanced degree at The University of Arizona and is deposited in the University Library to be made available to borrowers under rules of the Library.

Brief quotations from this dissertation are allowable without special permission, provided that accurate acknowledgment of source is made. Requests for permission for extended quotation from or reproduction of this manuscript in whole or in part may be granted by the copyright holder.

SIGNED: Rachel Ter Hayes

ACKNOWLEDGEMENTS

I would first like to thank my advisors Mike Hammond and Natasha Warner for their support and dedication to my education. From day one, Mike has been a source of advice, encouragement, and inspiration to me. Thank you, Mike for being a great mentor. Although Natasha and I overlapped only two years at Arizona, she has had an extraordinary impact on my training as a researcher and on my goals for the future. Natasha is always generous with her time and resources and this project could not have happened without her commitment. It was in Janet Nicol's bilingualism seminar that I first became interested in speech perception, and I am grateful to Janet for encouraging me to pursue this type of research. Janet has been a source of great support and encouragement. My entire committee—Mike, Natasha and Janet—has been amazing.

The consistent and caring support of the entire Linguistics faculty and staff has been overwhelming. You are truly an exceptional bunch! I am especially grateful to Professors Diane Ohala, Adam Ussishkin, Andy Barss, Diana Archangeli, Dick Demers, LouAnn Gerken, Merrill Garrett, Sheila Dooley Collberg, Andrew Carnie, Simin Karimi, and Terry Langendoen, and to Rosemary, Gwen, Nova and Yolanda for always being so helpful.

I would like to thank Jessica Maye, whose dissertation work inspired some of the experiments presented here. I appreciate Jessica's generosity in sharing her experimental stimuli and making suggestions during the early stages of my dissertation work. Tami Kaplan, one of my first college professors, is the person who first encouraged me to go to graduate school. These years later, I remain grateful for her friendship and her faith in me.

Graduate school wouldn't have been nearly as much fun without my fellow students Cathy Hicks, Lika Chtareva, Bob Kennedy, Sonya Bird, Tania Granadillo, Luis Barragan, Meg O'Donnell, Erin O'Bryan, Peter Norquest, Kyoko Masuda, Jason Haugen, Heidi Orcutt, William Alexander, Alina Twist and many other fabulous people. I am especially grateful for the friendship of Cathy Hicks, Robin Nabi, Elizabeth Stein, Lika Chtareva and Sonya Bird.

I thank my mother, a model of incomprehensible strength and love, my father, who has always challenged me to think for myself, and my brother Christopher, who inspires me by following his dreams. Finally, I thank my darling husband Basil, who believes in me. He also reminds me that there is more to life than ANOVAs and phonemes!

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

Listeners are sensitive to phonetic differences that correspond to phonemic contrasts in their native language, and they often exhibit difficulty discriminating sounds that are not contrastive in their native language. Although a large literature shows that learners can improve their perception of novel contrasts with exposure to a second language, there is still little understanding of *how* learners accomplish this.

There are at least two possible sources of evidence that learners might use to acquire sensitivity to novel sound contrasts. First, learners might use their knowledge of minimal pairs in the language to determine which sounds are contrastive. For example, knowing the minimal pair *toe-doe* may provide learners with evidence that /t/ and /d/ are contrastive in English (Lexical-Contrast-Based Evidence). Second, learners might compute the statistical distributions of the acoustic-phonetic properties of their second language input. The logic is that two contrastive speech sounds will be represented by distinct distributions along a of a number of acoustic-phonetic dimensions (Distribution-Based Evidence). Although both are possible sources of the evidence learners use to acquire novel second language sound contrasts, the relative influence of these two types of information is not yet known.

Experiments 1 and 2 of this dissertation employ a variety of training techniques to determine the relative influence of Lexical-Contrast-Based and Distribution-Based evidence on participants' sensitivity to a novel contrast. Results indicate that while both

kinds of evidence affect sensitivity, Lexical-Contrast-Based evidence had a stronger influence on discrimination performance.

While Experiments 1 and 2 tested learners' sensitivity to novel contrasts, it is not yet clear that improved discrimination ability is of benefit in subsequent second language learning. Experiment 3 examined the linguistic relevance of participants' improved discrimination ability by testing learners' lexical representations for new words that differed minimally with respect to the trained contrast. Regardless of training condition, participants did not record the new contrast distinctly in their lexical representations. That participants exhibited sensitivity to the novel contrast but were nonetheless unable to record the contrast lexically suggests a dissociation between learners' acoustic-phonetic knowledge of their second language and their ability to represent that knowledge contrastively in their lexicon.

CHAPTER 1

INTRODUCTION

1.1. Introduction

Adults exhibit particular sensitivity to phonetic differences that correspond to phonemic contrasts in their native language (Werker *et al.* 1981, Werker & Tees 1984, Best 1994), but they often exhibit difficulty distinguishing sounds that are not phonemically contrastive in their native language (Goto 1971, Best & Strange 1992, Werker *et al.* 1981, MacKain *et al.* 1981), indicating that linguistic experience shapes sensitivity to phonetic distinctions. However, the question of how linguistic experience affects this sensitivity and its application to subsequent linguistic development—the development of phoneme inventories and contrastive lexical representations—remains. This dissertation addresses this question by comparing the effects of various types of linguistic training on participants' ability to discriminate two speech sounds that are not phonemically contrastive in their native language, and whether or not they can represent the speech sounds contrastively in their developing second language lexicon.

It has often been assumed that minimal pairs tell learners which sounds are contrastive in the language they are learning (e.g. MacKain 1982). For example, knowing that the words *toe* and *doe* have different meanings may provide the learner of English with the information necessary to determine that their initial sounds, /t/ and /d/, are contrastive in English. However, recent research provides evidence that listeners can learn to discriminate second language word pairs with more or less accuracy from simply listening to words in the language without any reference to what the words mean (Maye 2000 for adults; Maye *et al.* 2002 for infants). These studies by Maye and her colleagues, to be described in detail later in this chapter, indicate that listeners can calculate distributional patterns in their second language input, and that these patterns alone can influence listeners' discrimination of second language phoneme contrasts without the benefit of word meanings or minimal pairs.

To date, however, no studies have directly compared the effects of these distributional patterns (Distribution-Based evidence) and the availability of minimal pairs (Lexical-Contrast-Based evidence) on a learner's ability to learn second language phonemes. This dissertation compares these two types of evidence directly (Experiment 1), and investigates the interaction of the two types of evidence (Experiment 2).

Furthermore, although studies have shown that adults can discriminate new phonetic distinctions after being trained with Lexical-Contrast-Based or Distribution-Based evidence (see the review of this literature provided in section 1.3), none to date have shown that either kind of evidence is used to learn that the phonetic distinction must be encoded in the lexicon as phonemically contrastive. This dissertation also investigates whether Lexical-Contrast-Based or Distribution-Based learning contribute to the development of a target-language-like lexicon, and addresses this issue by testing participants' ability to establish contrastive lexical representations for new second language words (Experiment 3).

1.1.1 Summary of research questions

The broadest question driving this dissertation is: What is the evidence that second language learners use to learn novel phonemic contrasts? I address this question by considering the following sub-questions:

- How do Lexical-Contrast-Based and Distribution-Based evidence compare in their effects on learners' perception of novel phonemic contrasts? (Experiment 1)
- How do these types of evidence interact? (Experiment 2)
- Can short-term training using Lexical-Contrast-Based or Distribution-Based evidence be exploited in subsequent word learning? If so, under what training conditions can they be exploited? (Experiment 3)

In the present experiments, native speakers of English underwent various kinds of training in an unfamiliar language, where the training provided evidence for either the presence or absence of a phonemic contrast in the second language between two sounds that are not contrastive in English, and their performance on discrimination and word learning tests was evaluated.

1.1.2 Organization of the dissertation

In this first chapter, I will discuss the acquisition of first and second language phonemes, focusing on the kinds of evidence that are available to and used by language learners in acquiring a first or second language phoneme inventory. In chapters 2 and 3 I present Experiments 1 and 2, which compare the effects of the availability of the various kinds of evidence on language learners' ability to discriminate a novel phoneme contrast. Chapter 4 presents an experiment which tests whether participants who were provided with the various manipulations of evidence in Experiments 1 and 2 can use their knowledge of the phoneme contrast to create lexical representations that reflect the contrast for new second language words. Finally, Chapter 5 summarizes the experimental findings with a discussion of the evidence that listeners can use to learn novel second language phoneme contrasts, and presents implications of the findings for the notions of phonetic distinction, phonemic contrast, and lexical representation. This chapter also includes a discussion of the applications of this research to second language pedagogy and its relevance to current models of linguistic theory.

1.2 Native language and cross-language speech perception

Humans are born with the ability to discriminate both native and non-native sound contrasts (e.g. Aslin *et al.* 1998) and over the course of the first year of life exhibit a reduction in their ability to discriminate the contrasts that are not relevant to their native language. Native language learning involves retaining sensitivity to some speech sound contrasts while losing sensitivity to the rest, and infants begin to exhibit language-specific loss of discrimination ability during the second half of the first year of life (e.g. Werker & Tees 1984). Early exposure to the native language, then, provides infants with the evidence necessary to determine what phonetic variation in their linguistic input corresponds to native language phonemic contrasts and what variation can be ignored as linguistically irrelevant, and babies exhibit adult-like discrimination of native language sound contrasts already by the end of their first year (Kuhl *et al.* 1992). The question of how this learning takes place is addressed in section 1.3; the present discussion will focus on how monolingual adults perceive native and non-native speech sounds.

1.2.1 Native language experience shapes sensitivity to phonetic variation

Native language speech perception is characterized by sensitivity to phonetic differences that correspond to phonemic contrasts and by a lack of sensitivity to non-contrastive phonetic variation. In English, for example, the phoneme /d/ is pronounced word-initially as varying between voiceless unaspirated [t] and prevoiced [d]. Despite the phonetic differences among them, native speakers of English reliably perceive all variants along the [t]-[d] continuum as the phoneme /d/ in word-initial position, and discriminate these variants of /d/ from word-initial variants of the phoneme /t/, which is produced as aspirated [t^h] (with a variety of VOT durations). In word-initial position, then, native English listeners ignore the phonetic distinction between [t] and [d] but not the distinction between [t^h] and either [t] or [d].[1]

However, listeners whose native language is Spanish break up the phonetic space differently, as the phonemes /d/ and /t/ are manifested differently in Spanish than in English. In Spanish, word-initial /d/ is pronounced as [d], while word-initial /t/ is

[1] The phonetic variants presented here are simplified for the purpose of illustration. In actuality, a range of prevoicing values for [d] and a range of VOT values for [t^h] are included in the normal phonetic variation. A number of other phonetic characteristics also vary for these sounds. For the purpose of this illustration, only the variants [d], [t] and [t^h] will be discussed.

produced as some variant on a range between [t] and [t^{h}]. Note that the sounds [d] and [t] are contrastive in Spanish, as they belong to the phoneme categories /d/ and /t/, respectively. Also note that the sounds [t] and [t^{h}] both belong to the Spanish category /t/. Figure 1-1 shows how Spanish and English assign the sounds [d], [t] and [t^{h}] to the phonemes /d/ and /t/.

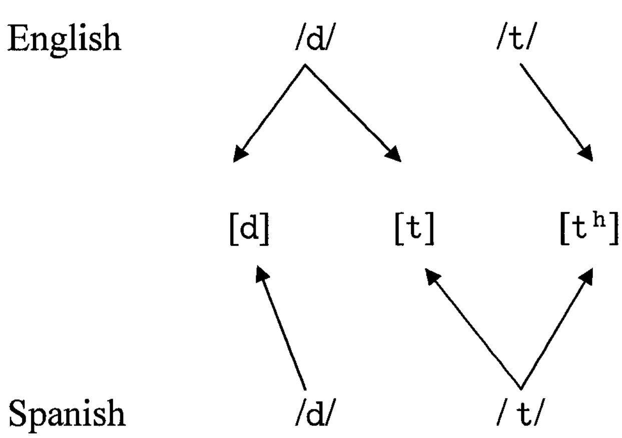


Figure 1-1. English and Spanish phonemes / t/ and /d/ and their phonetic variants

Native speakers of English perceive the distinction between [d] / [t] and [t^{h}] as contrastive, and they ignore the phonetic variability between [d] and [t], while native speakers of Spanish perceive the difference between [d] and [t] / [t^{h}] as contrastive and ignore the difference between [t] and [t^{h}]. Thus native language experience determines which phonetic variation listeners treat as contrastive and which variation they ignore as non-contrastive. However, as will be discussed in the following section, listeners do not always lose the ability to discriminate sub-phonemic phonetic variation.

1.2.2 Listeners are sensitive to some sub-phonemic phonetic variation

Recall that native speakers of English ignore variation along the continuum between [d] and [t] in word-initial position, since these variants belong to the English phoneme /d/. On the other hand, the sounds [d] and [t] belong to the Spanish phoneme categories /d/ and /t/, respectively. Therefore, native speakers of English must re-focus their attention to sub-phonemic phonetic variability that they are accustomed to ignoring in English (between [d] and [t]) in order to distinguish, for example, the Spanish words *dios* 'day' and *tios* 'aunts and uncles'. There is some evidence that listeners are sensitive to this kind of sub-phonemic phonetic variation under certain task conditions (Werker 1994, Pegg & Werker 1997, Werker & Logan 1985, Hayes 2002a, Polka 1992) and that this sensitivity may be exploited in second language learning.

Pegg and Werker (1997) demonstrated that even though [t] and [d] do not correspond to a phonemic contrast word-initially in English (because they are both variants of the same phoneme /d/), adult native speakers of English can discriminate the two sounds with better than chance accuracy in a discrimination task (though their discrimination of these sounds was less accurate than their discrimination of native phonemic contrasts). Thus the observed lack of sensitivity to non-native contrasts in adults does not necessarily constitute an absolute loss of ability to discriminate the sounds (Polka 1992, Pegg & Werker 1997).

A second type of evidence for listeners' sensitivity to sub-phonemic phonetic differences is related to the perceptual magnet effect (Kuhl 1991, Grieser & Kuhl 1989). Central to interpreting the perceptual magnet effect is the notion of category prototype, which is the "ideal" member of a category. Speech sound categories, like other categories such as the semantic category *fish*, have members that are more representative of the category than others (e.g. a trout is a prototypical fish while an eel is not). There is evidence that speech sound category prototypes act as "magnets", such that native listeners are worse at discriminating sound differences near a phonemic category prototype than differences of the same magnitude that are further away from the category prototype. That listeners' ability to discriminate phonetic differences varies within phonemic categories provides additional evidence that adult listeners are sensitive to some sub-phonemic phonetic differences.

Yet another type of evidence for listeners' ability to discriminate non-phonemic phonetic differences has to do with listeners' ability to assign an internal structure to their native language sound categories (see Miller 1994 for a review of this literature; and see Mondini *et al.* 2002 for internal category structure in highly-proficient bilinguals[2]). "Goodness" judgments by native speakers indicate that some members of a phoneme category are judged as better exemplars of the category than others (and that this structure

[2] Mondini *et al.* (2002) found that not only do native speakers of Dutch who have learned English show different category boundaries on the VOT continuum for English stop consonants, but the internal structure of their categories is also shifted.

shifts relative to contextual influences such as speaking rate; Miller *et al.* 1997, Wayland *et al.* 1994). These findings provide evidence that listeners can make use of both categorical and fine-grained acoustic-phonetic information when listening to speech; the primary contribution of this literature to the present research is that it further supports the observation that listeners can be sensitive to sub-phonemic phonetic differences.

The last type of evidence for listeners' sensitivity to sub-phonemic phonetic differences comes from a cross-language sound identification task. When played stimuli that vary along an acoustic dimension that corresponds to a native phonemic distinction, adult listeners typically perceive a sharp, not gradual, change from one speech sound to another (Repp 1984 and Lieberman & Blumstein 1988 review these findings). This is called categorical perception, and it indicates that listeners treat acoustic continua that are contrastive in their native language categorically.

On the other hand, adults do not perceive non-native phonemic contrasts in the same way as native speakers do (Goto 1971, Best & Strange 1992, Werker *et al.* 1981) and they do not perceive acoustic continua that do not contribute to native phoneme distinctions categorically (MacKain *et al.* 1981, Hayes 2002a). In fact, Hayes (2002a) found that monolingual English speakers perceive Japanese consonant duration continuously. In Japanese, most stop consonants can be phonemically short or long (e.g. /t/ versus /tt/). However, in English, variation in stop consonant duration is sub-phonemic, as stop consonant duration is not a contrastive parameter word-internally in English. Even though the acoustic dimension of consonant duration is not contrastive word-internally in their native English, monolingual English speakers in the Hayes

(2002a) study were able to detect some consonant length differences in Japanese, indicating that they do have some sensitivity to phonetic distinctions along the continuum,[3] and providing further evidence for listeners' ability to discriminate phonetic differences that do not correspond to phonemic contrasts in their native language.

These findings together indicate that listeners have at least some sensitivity to phonetic variability that is not phonemically contrastive in their native language, a fact that may prove useful when learning novel second language phonemes. In fact, I will argue in this dissertation that learners in the initial stages of second language acquisition can exploit the ability to discriminate some non-phonemic phonetic variability as a prerequisite to acquiring novel second language phonemic contrasts.

The discussion so far has demonstrated listeners' ability to discriminate sub-phonemic phonetic differences; however, there are some non-native contrasts that are not readily discriminable by listeners even under ideal experimental conditions. MacKain *et al.* (1981), for example, found that native speakers of Japanese with no English training discriminated English /r/ and /l/ with near-chance accuracy. Several other studies of Japanese listeners' perception of English /r/ and /l/ have produced similar results (Goto 1971, Logan *et al.* 1991, Miyawaki *et al.* 1975, Mochizuki 1981, Sheldon & Strange 1982, Yamada & Tohkura 1992). Other difficult cross-language contrasts include the

[3] However, learners' perceptual treatment of acoustic continua that are relevant to their second language becomes more target-like with more exposure to the second language. The "categoricalness" of perception along a non-native continuum increases with exposure to the second language. Participants in the Hayes (2002a) study who had one or more years of Japanese language experience exhibited more categorical treatment of consonant duration in Japanese than did the monolingual English speakers.

English /d/-/ð/ contrast for native speakers of French (Sundara *et al.* 2002), the French rounded/unrounded contrast for native speakers of English (Levy & Strange 2002), and Thai tones for native speakers of English (Wayland & Guion 2001). In addition to these cross-language studies, there are studies of second language learners that indicate that even listeners with significant exposure to their second language cannot discriminate novel second language contrasts. This leads to the question of why some non-native contrasts are easier for listeners to discriminate than others.

1.2.3 Second language speech perception

A large literature addresses the issue of which non-native contrasts are more difficult for second language learners to master, and suggests that several factors contribute to the relative discriminability of non-native contrasts (e.g. Polka 1992, Bohn 1995, Iverson *et al.* 2003). The degree of conflict between the phonemic inventories of the native and second languages is generally agreed to be a main determiner of relative perceptual difficulty for second language learners; this observation is formalized in Best's Perceptual Assimilation Model (Best 1994), Flege's Speech Learning Model (Flege 1995b), and Kuhl's Native Language Magnet Model (Kuhl 1991, 1993). Other potential contributing factors are also addressed in this literature on second language speech perception (e.g. Flege 1993, Best *et al.* 2001, Pisoni *et al.* 1982, Polka 1992, Werker & Logan 1985, Werker & Tees 1984). Much of this literature focuses on pinpointing the sources of difficulty for monolingual listeners in cross-language perception tasks—the perception of non-native sounds by monolinguals not only provides insights into adults' native language perceptual systems, but also provides a window to the initial state for second language speech perception. However, as I am primarily interested in this dissertation in how perceptual learning proceeds from the initial state, the remainder of this discussion will focus on the development of second language speech perception beyond the initial state. The development of second language speech perception can be studied in two ways: cross-sectional studies of learners at different stages in their acquisition of their second language (Flege *et al.* 1997, Wayland & Guion 2001, Levy & Strange 2002, Kato & Tajima 2002, Hayes 2002a, Hayes & Masuda in prep); and training studies where monolingual listeners are trained on some aspect of a "second language" in a laboratory setting over the course of an experiment session. Since the questions of interest in this dissertation focus on the evidence that second language learners extract from their linguistic input to learn novel second language phonemes, the remainder of this discussion will address the latter type of study, where participants learn to discriminate new sounds during experimental training sessions. These training studies make it possible to manipulate the evidence that learners receive about their "second language" in order to understand how different types of evidence in the linguistic input impact the perception of second language sounds.

1.3 Training studies

Training studies allow researchers to investigate the types of information that learners can make use of in their linguistic input to learn new second language sounds. This section provides an overview of various types of training that have been explored.

1.3.1 Explicit training studies

Training studies vary in the explicitness of information provided to learners (see Landahl & Ziolkowski 1995, Marsh & Mineo 1977, Pruitt 1996, Jamieson & Marosan 1986, Flege 1995a, Wang *et al.* 1999, Guenther *et al.* 1999, Catford & Pisoni 1970 for examples of various levels of explicitness during training). For example, in a very explicit training study, Wang *et al.* (1999) trained native speakers of English to perceive Mandarin tones by presenting stimuli containing each of four tones and asking participants to identify which tone they heard (tones were labeled one through four). Immediately following the participants' responses, they were told which tone they had heard; that is, they were given explicit labels for the tones that they heard and were trained to identify the tones with these labels. Wang *et al.* (1999) reported that following this explicit training, participants exhibited improved discrimination of the tones. Similarly explicit training studies are reported in Kingston & Moreton (1998) and Kingston *et al.* (1996).

As I am primarily interested in the types of information that second language learners are able to infer from their linguistic input under more "natural" circumstances, such explicit training methods will not be used in the experiments reported in this dissertation. The remainder of this review focuses on a division among studies that present two main kinds of implicit evidence for a novel distinction—Lexical-Contrast-Based and Distribution-Based.

1.3.2 Lexical-Contrast-Based training studies

In the majority of perception training studies, participants are presented pairs of words that differ only in the new contrast; they are instructed to indicate whether the pairs are the "same" word or two "different" words, and they are given immediate feedback as to whether they correctly discriminated the words. This is called "contrast training", and it exploits the contrastive property of phonemes—that phonemes are the minimal units of speech that distinguish words in the language.

For example, Strange & Dittman (1984) trained native speakers of Japanese to perceive the /r/-/l/ distinction in English. Training consisted of a same-different discrimination task where participants were given immediate feedback as to whether or not they had correctly discriminated the words *rock* and *lock*. The majority of participants in this study showed improvement between pre- and post-training identification of the /r/ and /l/ sounds. This kind of feedback training, where participants are taught a novel contrast with evidence that the contrast distinguishes words in the target language, provides participants with Lexical-Contrast-Based evidence for a novel contrast (see Tajima *et al.* 2002, Flege 1989, Flege 1995a for more examples of this type of training).

An alternative and even less explicit method for providing Lexical-Contrast-Based evidence for a phonemic contrast is to teach participants that sound strings differing in the novel contrast have different meanings. In this kind of training, it is assumed that a learner can determine that two strings of speech are distinct because they have different meanings. If a learner knows that the string [rɔk] refers to the "hard gray object in the yard", while the string [lɔk] refers to the "complex metal object that prevents the door from being opened", she might be able to infer the importance of the /r/-/l/ contrast in the language she is learning even without being explicitly told that *rock* and *lock* are different words. In the experiments reported in Chapters 2 and 3, participants who received Lexical-Contrast-Based evidence for a distinction were provided this kind of evidence for a novel second language contrast. They were presented sets of minimal pairs along with their meanings such that the only evidence these learners had for the contrast came from their having identified that the phonetic difference coincided with a difference in word meaning. To date, no other studies have provided this particular kind of Lexical-Contrast-Based training.

In the experiments, performance by participants who were trained with Lexical-Contrast-Based evidence was compared to that by participants who received another type of evidence: Distribution-Based evidence, which depends on learners' ability to calculate the distributional properties of the acoustic information in their linguistic input. This type of training is considered in the next section.

1.3.3 Distribution-Based training studies

In Distribution-Based training, participants receive only acoustic information about the target language, and no information (explicit or implicit) about what distinguishes words in the language (Maye 2000, Maye *et al.* 2002, Smits *et al.* submitted). This kind of training exploits the distributional tendencies in speech with respect to phonemic distinctions—separate phonemes have distinct distributions on some acoustic dimensions.

Using Distribution-Based evidence, Maye (2000) trained native speakers of English either to discriminate or not to discriminate voiceless unaspirated stops from prevoiced stops (e.g. [g] versus [k]). These differences are sub-phonemic in English, yet they are discriminable by native speakers of English under certain task conditions (as are [d] and [t] discussed above; Pegg & Werker 1997). Maye created a series of eight tokens on continua between prevoiced [g] and voiceless unaspirated [k] (in three different vowel environments ([ga]-[ka], [gæ]-[kæ] and [gɚ]-[kɚ])—these served as the experimental tokens in the experiment. One set of participants heard a Bimodal distribution of these experimental tokens and one set heard a Monomodal distribution of the tokens during training (see Figure 1-2). Tokens 1 through 8 are the most [g]-like (token 1) to the most [k]-like (token 8).

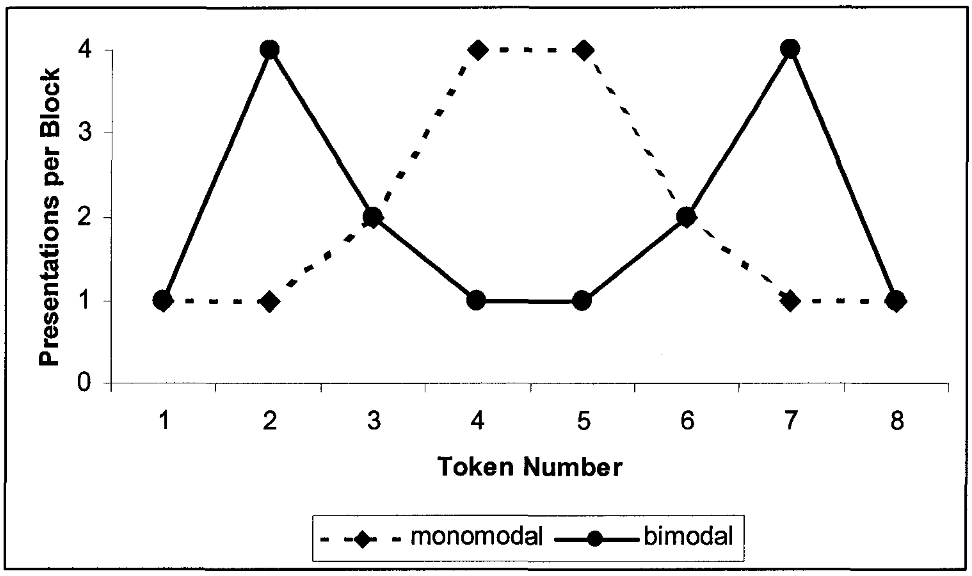


Figure 1-2. Representation of distributions of experimental training tokens presented in Maye (2000)

The Bimodal group heard tokens 2 and 7 four times as often as the tokens in the middle of the continuum, while the Monomodal group heard tokens 4 and 5 four times as often as the tokens at the endpoints of the continuum. Maye assumed that the distributions of tokens (phonetic exemplars) along a continuum reflect the phonemic status of the continuum, such that Bimodal training provides listeners with two distinct clusters of tokens (phonemes) along the continuum, while Monomodal training provided them with only one cluster (phoneme) along the continuum. All participants also heard a set of filler tokens. The fillers were included to distract participants from focusing too much attention on the trained distinction, and also to provide more of the phonetic variability that is present in natural linguistic input. After a training session where participants listened to one of the two distributions of the various [g]-[k] tokens plus the variety of filler tokens, participants took a discrimination test where they were presented pairs of tokens and were asked to decide whether the two were the same word in the target language or two different words. By asking them whether they were the same word, not whether they were "the same utterance", it was expected that they would respond based on their knowledge of the phonemic structure of the language, not based on sensitivity to small sub-phonemic acoustic differences.

Included in the test were pairs of the same experimental token (Experimental-Same; token 1 & token 1; token 8 & token 8), pairs of different experimental tokens (Experimental-Different; token 1 and token 8; token 8 and token 1), pairs of identical filler tokens (Filler-Identical; e.g. the same production of the filler [ma] presented twice), pairs of same filler tokens (Filler-Same; e.g. two separately-produced tokens of the filler [ma]), and pairs of different filler tokens (e.g. [ma] & [la]). The filler pairs were included in the discrimination test in order to direct interpretation of the experimental items. The Experimental-Same pairs were included as controls for the Experimental-Different pairs—if they understood the task, participants were expected to correctly respond "same" to the Experimental-Same pairs. The Filler-Different pairs were included so that at least some pairs would be "different" for all participants regardless of training condition. All participants were expected to respond "same" to the Filler-Identical items, since there were no differences at all between them. The Filler-Same tokens, on the other hand, were not phonemically different despite the small sub-phonemic acoustic differences between them. If participants responded "different" to the Filler-Same tokens, this would indicate that they were attending to small sub-phonemic differences in the task, not to phonemic differences, and might indicate that "different" responses to the Experimental-Different pairs reflected only their sensitivity to the small sub-phonemic differences and not knowledge that the differences corresponded to a phonemic contrast. On the other hand, if they responded "same" to the Filler-Identical tokens, this would indicate that they were ignoring sub-phonemic differences in the task and that their "different" responses to Filler-Different tokens could be interpreted as indicating knowledge of the phonemic contrast. It was found that participants trained with the Bimodal distribution of tokens were better able to discriminate tokens 1 and 8 than were participants trained on the Monomodal distribution of the same stimuli, while performance on all other pairs was near-ceiling. Maye interpreted these results as indicating that participants in the Bimodal group were able to learn the phonemic status of the difference between [g] versus [k] based only on the Distribution-Based evidence they received during training, and that minimal pairs (or Lexical-Contrast-Based evidence) are not necessary for learning phonemes. This work has paved the way for several other research questions, including the relative influence of Lexical-Contrast-Based and Distribution-Based evidence in novel second language phoneme learning.

1.4 Do both Distribution-Based and Lexical-Contrast-Based evidence contribute to the acquisition of phonemic contrasts?

I approach the question of whether both Distribution-Based and Lexical-Contrast-Based evidence contribute to the acquisition of phonemic contrasts by providing participants with various kinds of Distribution-Based and Lexical-Contrast-Based evidence about the presence or absence of a novel distinction in a made-up language. By training participants on a made-up language it is possible to control for prior experience with the language and ensure that participants all have the same initial state. The training that participants received was manipulated such that some received Distribution-Based evidence in support of or against a novel distinction, some received Lexical-Contrast-Based evidence in support of or against a novel distinction, and some received a combination of the two types of evidence during training. Following a training session, participants were tested on their ability to discriminate the novel contrast in a sound discrimination test. These experiments are detailed in Chapters 2 and 3.

1.5 From phonetic differences to phonemic distinctions

The results of the discrimination test can provide information about participants' knowledge of phonetic distinction, but cannot indicate whether or not they can exploit the phonetic distinction for future linguistic development. Knowledge of a phonetic distinction is useful linguistically only if it is exploited to distinguish words in the language. In order to test whether participants who are trained on a novel phonetic difference can learn new words that contain the new sounds, participants in this study participated in a word learning experiment. The word learning experiment can provide two kinds of information about the acquisition of novel second language contrasts: first, it provides information about how the different types of evidence (Lexical-Contrast-Based and Distribution-Based) differ in their effects on learners' ability to represent the novel contrast in their new second language "lexicons"; second, it provides information about the time-course for learning to distinguish a new contrast and learning to represent the contrast lexically. If participants' discrimination performance on the novel contrast precedes their ability to represent the contrast lexically, this provides evidence for an intermediate stage between having knowledge of a phonemic contrast (as evidenced by discrimination performance in Experiments 1 and 2) and being able to exploit the contrast lexically (as evidenced by word learning performance in Experiment 3).

1.6 A note on phonetic variability and training studies

It has been observed that phonetic training studies that employ synthetic speech, as do the experiments presented here, may provide "misleading" or "incomplete" information about the phonemes that participants are expected to learn (Pisoni *et al.* 1985, Logan *et al.* 1991). The issue is that synthetic continua often provide very little of the phonetic variability that listeners encounter in more natural linguistic settings. Several studies have addressed this issue by including training stimuli produced by a variety of talkers and presenting the target segments in a variety of phonetic contexts in order to increase phonetic variability in the training stimuli (Pisoni & Lively 1995, Lively *et al.* 1993, Logan *et al.* 1991). It is claimed that variability in talkers and phonetic contexts helps adults learn more stable and robust phonetic categories and that training that includes these kinds of phonetic variability is important to learners' ability to generalize from their learning to stimuli produced by different talkers and to novel phonetic contexts (Logan *et al.* 1991). The issue of phonetic variability is addressed in Experiments 1 and 2.

CHAPTER 2

EXPERIMENT 1: DISTRIBUTION-BASED VERSUS LEXICAL-CONTRAST-BASED EVIDENCE IN SECOND LANGUAGE SPEECH PERCEPTION

2.1 Introduction

This experiment tests how the availability of two different types of evidence during a training session—Distribution-Based and Lexical-Contrast-Based evidence—affects adult monolingual English speakers' ability to discriminate a novel phonemic contrast in a subsequent perception task. As discussed earlier, there are a variety of factors that determine relative perceptibility of second language phoneme contrasts. Some novel contrasts are not easily discriminable by non-native listeners (e.g. the English /r/-/l/ distinction is very difficult for native speakers of Japanese to learn), while some are more easily discriminable (e.g. monolingual English speakers can discriminate tokens along a continuum between [d] and [t] even though the differences are sub-phonemic in English). Because I am interested in whether or not participants can learn to perceive a phonetic difference as contrastive during a relatively short training session, it was necessary to use a contrast where participants were already sensitive to the related phonetic differences prior to training. These experiments, then, test whether or not various types of training can turn already-existing sensitivity to sub-phonemic differences into the ability to discriminate the sounds as phonemically contrastive. As discussed in Chapter 1, native English speakers are able to detect differences among prevoiced stops and voiceless unaspirated stops even though the difference is not phonemic in English (Pegg & Werker 1997, Maye 2000). Therefore, the prevoiced [g] voiceless versus voiceless unaspirated [k] distinction is used in these experiments (as in Maye 2000, Experiment 2).

The first type of evidence is called Distribution-Based evidence, and is exactly the same training provided participants in Maye (2000; these two training conditions are a partial replication of Experiment 2 from Maye 2000). Participants in training conditions that received Distribution-Based evidence were presented stimuli in two different distributions along an acoustic continuum between [g] and [k]. The Monomodal training group heard the stimuli with a single distributional peak along the continuum, consistent with there being no phonemic contrast between [g] and [k], and similar to the distribution of these sounds in English. The Bimodal training group heard the same stimuli, only with two distributional peaks along the continuum, consistent with there being a phonemic contrast between [g] and [k]. It was hypothesized that, after the training session, the Bimodal group would be more sensitive to the acoustic difference between the endpoints on the continuum than the Monomodal group would be. The question addressed with these training groups is whether participants can learn to discriminate or learn not to discriminate a novel sound contrast based on this kind of distributional evidence alone.

The second type of evidence is Lexical-Contrast-Based evidence, where participants were taught the meanings of the stimuli that they heard via the simultaneous presentation of pictures and auditory stimuli. Participants heard only stimuli near the two endpoints of the continuum between [g] and [k], and the pictures associated with these near-endpoint stimuli indicated either that the sounds were contrastive or not—stimuli beginning with [g] and [k] were either paired with the same pictures or with different pictures. Participants in the No-Contrast training group were presented tokens near both endpoints of the continuum accompanied by the *same* pictures, indicating that the difference between [g] and [k] does not contrast lexical items in the new language. As with Monomodal training, the training for the No-Contrast group was consistent with there being no phonemic contrast between [g] and [k] in the new language. Participants in the Contrast training group were presented the near-endpoint tokens accompanied by *different* pictures, indicating that the continuum between [g] and [k] is lexically contrastive. As with Bimodal training, the training for the Contrast groups was consistent with there being a phonemic contrast between [g] and [k]. The question addressed with these training groups is whether participants learn to discriminate or not to discriminate a novel sound contrast based on lexical evidence alone.

	Gives evidence that there is a contrast between [g] and [k]	**Gives evidence that there is NOT a contrast between [g] and [k]**
Distribution-Based Evidence	BIMODAL	MONOMODAL
Lexical-Contrast-Based Evidence	DISTINCTION	NO-DISTINCTION

Table 2-1. Summary of training conditions in Experiment 1

Table 2-1 provides a summary of all four training conditions detailed above. A fifth training condition was included as a control for the inherent "distributional" evidence provided to participants in the distinction and no-distinction groups. Presenting these participants with only the near-endpoint stimuli entails providing them with a type of bimodal distribution, which confounds Distribution-Based and Lexical-Contrast-Based evidence for the Contrast and No-Contrast training groups. This issue and the training condition included to address it are discussed in section 2.3 below. There was a sixth training condition in which participants received no training and served as a control group. The construction of training stimuli for each training condition will be discussed in further detail below.

2.2 Participants

Participants were 132 native English speakers enrolled in undergraduate courses at the University of Arizona who received either course credit or $5.00 payment for their voluntary participation. All participants were monolingual English speakers whose parents and/or primary childhood caretakers were also monolingual English speakers. Participants had between one and thirteen years of foreign language instruction (mean = 4.3 years). All spoke English exclusively on a daily basis, with their families, and at home.

Participants were randomly assigned to one of the five training conditions plus the sixth condition in which participants received no training, with a total of 22 participants per condition. The training conditions are described in further detail in the next section.

2.3 Training stimuli and methods

All training stimuli were the same as those created for and used in Maye (2000; Experiment 2). The experimental training stimuli were three continua between [g] and [k]: [ga] to [ka], [gæ] to [kæ] and [gɚ] to [kɚ]. First, a native speaker of American English produced each of the syllables, and from each a series of four digitally-synthesized items were created to make a continuum of eight tokens, such that tokens 1 through 4 were created from an original [g]-syllable production, and tokens 5 through 8 were created from an original [k]-syllable production.[4] The acoustic dimensions that were manipulated in the synthesis were duration of prevoicing and formant onset frequencies. The training stimuli ranged from nine milliseconds of prevoicing (the most [g]-like) to 0 milliseconds of prevoicing (the most [k]-like).[5] Exact details of the formant onset frequencies and prevoicing durations are provided in the Appendix.

The six training conditions in this experiment differed in both the distribution of tokens in the training blocks and in whether or not the auditory tokens were accompanied by pictures. In each training block, participants heard sixteen experimental training items

---

[4] As discussed in section 1.6, phonetic variability (e.g. a variety of talkers and tokens) would provide more naturalistic stimuli for the present experiments. The training stimuli used in the experiments reported here were all produced by the same speaker, and each stimulus of interest is synthesized from a single production. For this reason, the training that participants receive in these experiments do not include the kinds of phonetic variability just discussed. However, the training stimuli do include synthesized variability, where participants (depending on training condition) hear tokens with various values along the acoustic dimension of interest.

[5] Because it is unclear whether a range from zero to nine milliseconds of prevoicing constitutes a perceptible difference between these stimuli, listeners probably extracted most information about the differences among items from the formant transitions. However, these items were constructed such that they elicited the intended training effect nonetheless.

for each of the three vowel environments, totaling 48 training stimuli per block. The training blocks for each training condition are described in detail below.

Additionally, participants heard filler stimuli during training. All heard the same filler training stimuli and the same number of each. Filler training items began with /m/ and /l/ in the same vowel environments as the experimental training items: /ma/, /la/, /mæ/, /læ/, /mɚ/ and /lɚ/. The filler items were presented in exactly the same way as the experimental stimuli. Four different tokens of each filler item were presented twice each (4 tokens x 6 filler words x 2 presentations each) for a total of 48 filler items per block. The total for experimental plus filler training stimuli was 96 per block, and the entire training session consisted of four repetitions of the block (this equals 384 auditory stimuli total during training). Training stimuli were presented with a one-second inter-stimulus interval, leading to a total of approximately nine minutes of training, during which participants listened to the "words in the new language" without making any response.

2.3.1 Distribution-Based training conditions

Participants in the Monomodal training condition heard experimental training stimuli in a monomodal distribution between [g] and [k]: tokens from the center of each continuum (tokens 4 and 5) were presented four times as often as tokens at the endpoints of the continuum (tokens 1 and 8). Participants in the Bimodal training condition heard experimental training stimuli in a bimodal distribution between [g] and [k]: tokens near the two endpoints of each continuum (tokens 2 and 7) were presented four times as often as tokens in the center of the continuum. Participants in both of these training conditions heard the exact same set of experimental training stimuli; the only difference between the Bimodal and Monomodal training conditions is the number of times they heard each token. The distributions of experimental training tokens for the two distribution-based training conditions are demonstrated in Table 2-2, and Figure 2-1 provides a graphical representation of the distributions.

token number	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**total**
Monomodal training	1	1	2	4	4	2	1	1	16
Bimodal training	1	4	2	1	1	2	4	1	16

Table 2-2. Distributions of experimental training tokens (number of times each token occurs per training block)

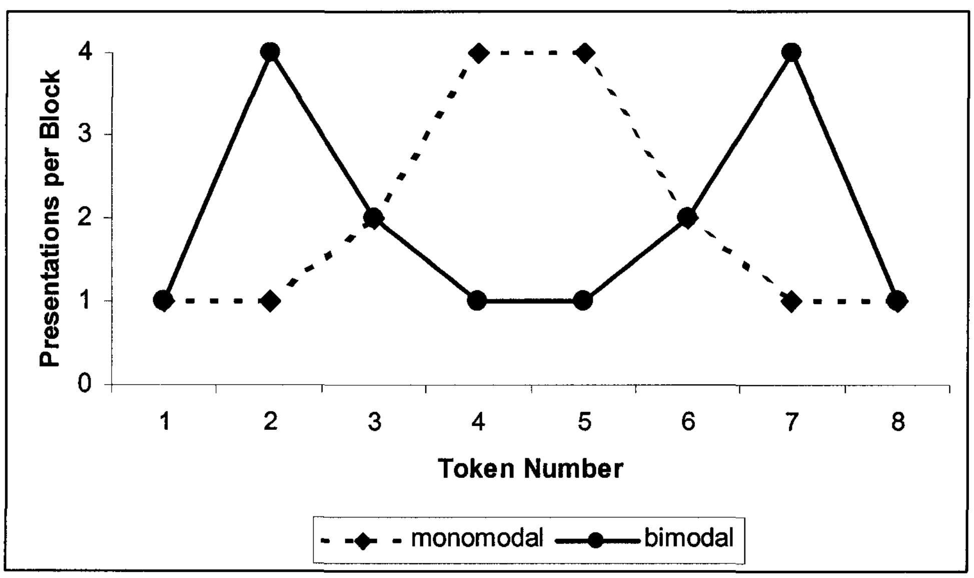


Figure 2-1. Representation of the distributions of experimental training tokens

It is important to keep in mind that participants in these training conditions were not provided meanings for any of the training stimuli; they simply listened to the stimuli. In order to ensure that participants paid attention to the task, they were given a sheet of paper with 384 boxes, and they were instructed to check a box for each word they heard (this was also done by Maye 2000). The task of checking a box after every word was intended to give participants incentive to listen to each word, but they were instructed not to allow the task of checking boxes interfere with listening to the words.

To summarize, the Bimodal and Monomodal groups were trained by presenting auditory stimuli in particular distributions, without participants receiving any indication of what the words meant; that is, without Lexical-Contrast-Based evidence for a distinction between [g] and [k]. Therefore participants in these training conditions were not able to use word meaning (or thus minimal pairs) to infer the importance of the acoustic difference between the sounds [g] and [k]—they received only distributional evidence for or against a distinction between the two sounds. In the Lexical-Contrast-Based training conditions, however, participants were shown pictures indicating word meanings along with the auditory presentation of the training stimuli. Participants who received this Lexical-Contrast-Based training, therefore, received lexical evidence for or against a distinction between [g] and [k].

2.3.2 Lexical-Contrast-Based training conditions

In the Lexical-Contrast-Based Training conditions, participants were shown pictures on a computer screen simultaneous with the presentation of the auditory stimuli, and were told that the pictures indicated the meanings of the words they heard. Participants in these training conditions heard only tokens 2 and 7 from each auditory continuum described above (see Table 2-2). Tokens 2 and 7 were chosen because they were the most frequently-presented tokens for the Bimodal training condition—the peaks in the bimodal distribution (see Figure 2-1). Therefore, the acoustic distance between the peaks in the bimodal distribution and the tokens heard by participants in the Lexical-Contrast-Based training conditions was identical (this was done in an attempt to make all training conditions maximally comparable).

There were two kinds of Contrast-Based Training: Contrast and No-Contrast. In the Contrast training condition, participants were presented pictures paired with auditory stimuli consistent with a contrast between words beginning with [g] and [k]. For example, token 2 from the [ga] to [ka] continuum was paired with a picture of a bunch of grapes while token 7 was paired with a picture of an eagle. Because these two auditory stimuli are minimally contrastive with respect to one segment and are assigned different meanings, these two stimuli make up a minimal pair, and were intended to indicate to the participant that the [g]-[k] sounds in tokens 2 and 7 belong to two separate phonemes because the difference between them creates a Lexical-Contrast. Table 2-3 provides an exhaustive list of the experimental training stimuli and their associated pictures/meanings.

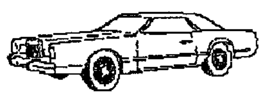
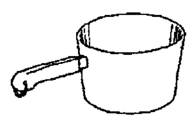
Participant heard...	Participant saw...
[ga]-[ka] token 2	
[ga]-[ka] token 7	
[gæ]-[kæ] token 2	
[gæ]-[kæ] token 7	
[gɚ]-[kɚ] token 2	
[gɚ]-[kɚ] token 7	

Table 2-3. Experimental training stimuli for the Contrast training condition

In the No-Contrast training condition, participants were presented pictures paired with auditory stimuli consistent with there not being a Lexical-Contrast between words beginning with [g] and [k]. For example, both token 2 and token 7 from the [ga]-[ka] continuum were paired with a picture of a bunch of grapes. Despite the acoustic difference between the two tokens, they were presented as having the same meaning, therefore suggesting to the listener that the [g]-[k] sounds in tokens 2 and 7 belong to the same phoneme. Table 2-4 provides an exhaustive list of the experimental training stimuli for the no distinction training condition along with their associated pictures/meanings.

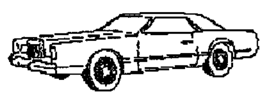
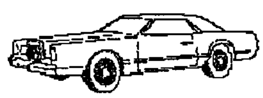
Participant heard...	**Participant saw...**
[ga]-[ka] token 2	
[ga]-[ka] token 7	
[gæ]-[kæ] token 2	
[gæ]-[kæ] token 7	
[gɚ]-[kɚ] token 2	
[gɚ]-[kɚ] token 7	

Table 2-4. Experimental training stimuli for the No-Contrast training condition

Because the words were presented with pictures indicating their meanings, it was assumed that participants would notice which of the words they heard had different meanings, which might provide the evidence they need to learn that [g] and [k] are or are not contrastive in the new language.

Filler training stimuli for the Lexical-Contrast-Based training groups were also presented with accompanying pictures. All four tokens of each filler item were accompanied by the same picture, giving participants some phoneme-internal phonetic variability in their training. The fillers are presented in Table 2-5.

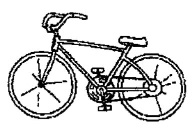
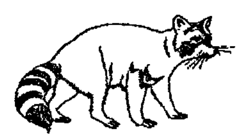
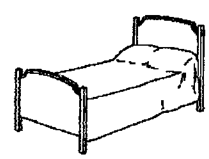
ma$_1$ ma$_2$ ma$_3$ ma$_4$		la$_1$ la$_2$ la$_3$ la$_4$	
mæ$_1$ mæ$_2$ mæ$_3$ mæ$_4$		læ$_1$ læ$_2$ læ$_3$ læ$_4$	
mɚ$_1$ mɚ$_2$ mɚ$_3$ mɚ$_4$		lɚ$_1$ lɚ$_2$ lɚ$_3$ lɚ$_4$	

Table 2-5. Exhaustive list of filler items, with their accompanying pictures

A fifth training condition, presented next, addresses a potential confound in comparing performance by participants in the Bimodal/Monomodal (Distribution-Based) training conditions to performance by participants in the Contrast/No-Contrast (Lexical-Contrast-Based) training conditions.

2.3.3 Distribution-Based training without phonetic variability

The Distribution-Based training conditions were intended to provide distributional information without lexical information for a distinction (or lack thereof) between [g] and [k], while the Lexical-Contrast-Based training conditions were intended to provide lexical information without distributional information for a distinction (or lack thereof) between [g] and [k]. While it was possible to accomplish the former—by not including pictures to indicate what the words meant in the Distribution-Based training conditions, it was not possible to accomplish the latter. In fact, the Lexical-Contrast-Based training conditions do provide distributional evidence for a contrast between [g] and [k]. Essentially, the distribution of tokens in the Contrast and No-Contrast training conditions is bimodal, as participants heard only tokens 2 and 7. Compare the distribution for the Bimodal and Contrast/No-Contrast training conditions in Figure 2-2.

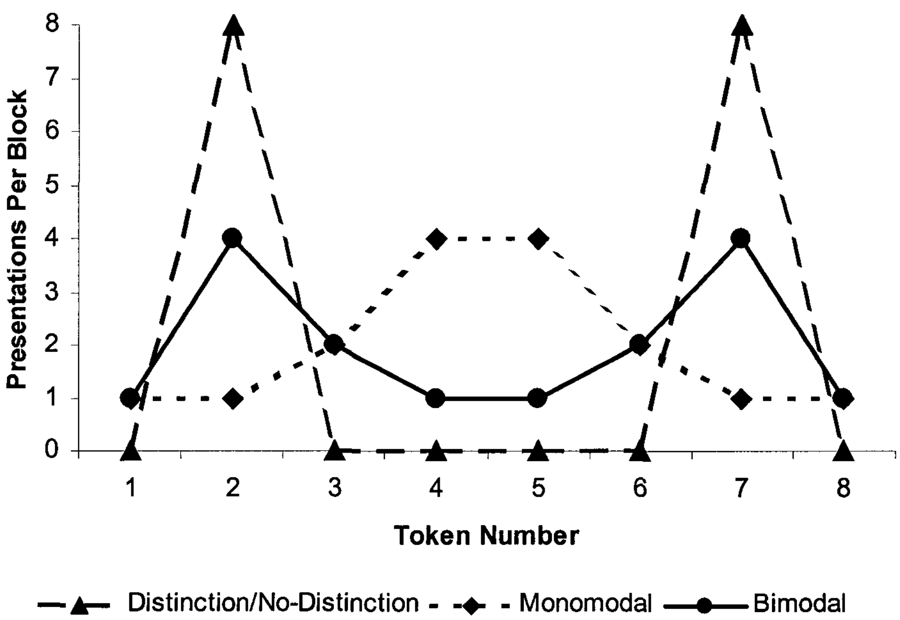


Figure 2-2. Graphical representation of distributions of experimental training tokens

Participants in the Contrast and No-Contrast training conditions heard a distinctly bimodal distribution of the tokens. This can create two potential problems: first, it appears that the Contrast and No-Contrast training groups receive both lexical and distributional information about the language during training. Thus performance by participants in both of these groups was potentially influenced by both types of information, possibly making it difficult to interpret their discrimination test performance. Second, participants in the No-Contrast group actually received lexical information that there is not a phonemic contrast, but, to the extent that they were influenced by the bimodal distribution of the tokens in their auditory input, they additionally received distributional evidence that there is a phonemic contrast. To summarize, because of the confound between lexical and distributional evidence in the Contrast and No-Contrast training conditions, it is not possible to know whether discrimination success by participants in either of these two groups results from the distributional evidence or the lexical evidence they received during training.

To address this problem, a fifth training condition, called Near-Endpoint training, was added. Participants in the Near-Endpoint training condition heard only tokens 2 and 7 (identical to the distribution heard by the Contrast and No-Contrast training conditions), but they were not provided with pictures to indicate word meaning. If participants in the Contrast training condition exhibit discrimination of [g] and [k] that is superior to participants' discrimination in the Near-Endpoint training condition, then the addition of lexical information is beneficial to learning to discriminate the new contrast. And if performance by participants in the No-Contrast training group is worse than that by participants in the Near-Endpoint group, then we also learn that the impact of lexical information overrules the effect of the distributional evidence that is inherent to the Lexical-Contrast-Based training conditions.

The Near-Endpoint condition further allows us to consider the influence of a type of phonetic variability on learning to discriminate new sound contrasts. The issue of phonetic variability and the acquisition of sound contrasts will be considered by comparing their performance at test to that of the Bimodal group.

Participants in the Near-Endpoint condition heard only tokens 2 and 7, without intermediate tokens, while participants in the Bimodal group heard tokens 2 and 7 most frequently, but not exclusively. Instead, participants in the Bimodal training condition heard more phonetic variability in their input. Therefore, it is also possible to test the hypothesis that phonetic variability in the input is critical to the development of long-term, robust phonetic learning that was discussed in section 1.6. However, it is important to note that the type of phonetic variability referred to in the discussion in section 1.6 included the phonetic variability introduced by a variety of speakers, speaking rates, etc. The type of phonetic variability in the present experiments is variability along a synthetically-manipulated acoustic contintuum. If discrimination by the Bimodal group is superior to that of the Near-Endpoint group, this serves as further support for this hypothesis.

2.3.4 The No-Training condition

In order to understand the effect of the different training conditions on participants' test performance, it was necessary to include a group of participants who received no training at all. They provide a baseline for interpreting the results of the other training conditions on the subsequent discrimination test.

2.3.5 Summary of the training conditions

Figure 2-3 provides a summary of the training conditions in this experiment.

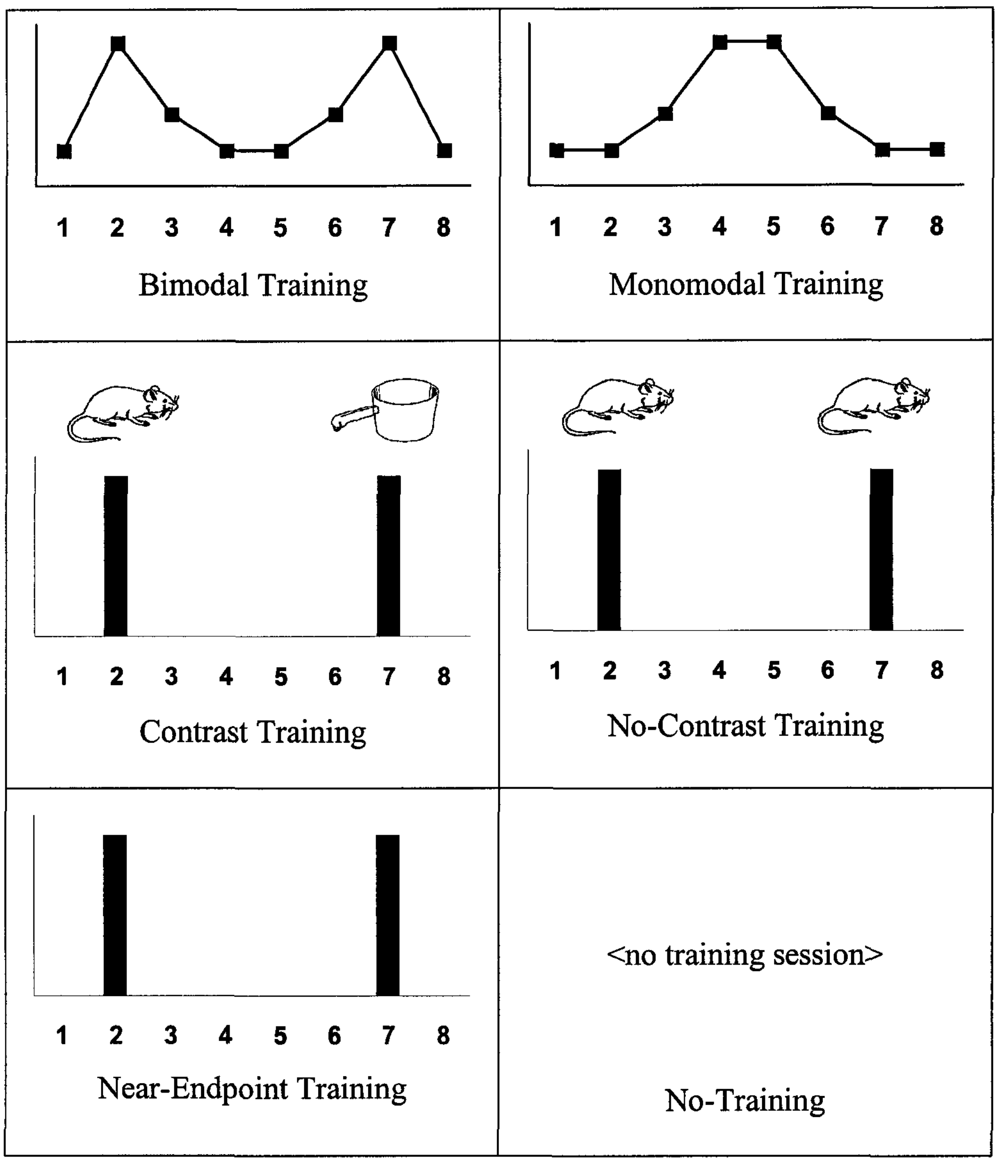


Figure 2-3. Summary of Experiment 1 training conditions

There were six training conditions in Experiment 1. Participants in all training conditions took a discrimination test immediately following the training session. Details of the test phase are provided next.

2.4 Testing stimuli and methods

Immediately following the training session, all participants took the same discrimination test. They were instructed that they would be hearing pairs of stimuli in the new language they had just learned and that their task was to determine whether the two stimuli in each pair were the same word repeated twice (by pressing a button labeled "same") or two different words in the new language (by pressing a button labeled "different"). They were told that they should use what they had learned about the language during the training session to make these decisions. The purpose of this test was to determine whether and how discrimination of [g] and [k] differed according to training condition.

2.4.1 Experimental test pairs

There were five types of pairs presented in the discrimination test. All stimuli in all pairs were taken from the token sets presented during training. For each of the continua of experimental training stimuli ([ga] to [ka], [gæ] to [kæ], and [gɚ] to [kɚ]), tokens 1 and 8 were paired to create Experimental-Different pairs. In half of these pairs, token 1 was presented first, and in half, token 8 was presented first (token 1-token 8; token 8-token 1). These are the critical test pairs, as participants' accuracy in responding "different" to these pairs indicates their ability to discriminate [g] and [k] in the new language.

For each continuum, two Experimental-Same pairs were also created (token 1-token 1; token 8-token 8). These were included to serve as the first type of control items for interpreting performance on the Experimental-Different test pairs. All participants, regardless of training condition, were expected to respond "same" to these pairs. "same" responses to these pairs would indicate that participants understood the task, so that "different" responses to the Experimental-Different pairs can be interpretable. These two types make up the experimental pairs and are listed in Table 2-6.

	Experimental-Different	**Experimental-Same**
[ga] to [ka]	token 1-token 8 ([ga]-[ka]) token 8-token 1 ([ka]-[ga])	token 1-token 1 ([ga]-[ga]) token 8-token 8 ([ka]-[ka])
[gæ] to [kæ]	token 1-token 8 ([gæ]-[kæ]) token 8-token 1 ([kæ]-[gæ])	token 1-token 1 ([gæ]-[gæ]) token 8-token 8 ([kæ]-[kæ])
[gɚ] to [kɚ]	token 1-token 8 ([gɚ]-[kɚ]) token 8-token 1 ([kɚ]-[gɚ])	token 1-token 1 ([gɚ]-[gɚ]) token 8-token 8 ([kɚ]-[kɚ])

Table 2-6. Exhaustive list of experimental pairs in the discrimination test

Each of the six Experimental-Different and the six Experimental-Same pairs was presented twice, for a total of 24 experimental pairs in the discrimination test.

2.4.2 *Filler test pairs*

There were also 24 filler pairs, half of which require the response "same" and half "different". There were two types of filler pairs that require "same" responses. First, recall that there were four tokens each of the filler stimuli in the training session, each a separately recorded production. The Filler-Identical pairs consisted of identical tokens (e.g. the same token of [ma] presented twice). The Filler-Same pairs consisted of non-identical tokens of the same word (e.g. two different tokens of [ma]). The Filler-Different pairs consist of filler words paired with different filler words (e.g. [ma]-[la] or [lɚ]-[mɚ]; the vowels were always the same). Decisions about which tokens of each filler word to use in the test were made randomly.

	Filler-Different	Filler-Identical	Filler-Same
[ma] & [la]	[ma]$_1$-[la]$_1$ [ma]$_2$-[la]$_2$ [la]$_3$-[ma]$_3$ [la]$_4$-[ma]$_4$	[ma]$_1$-[ma]$_1$ [la]$_2$-[la]$_2$	[ma]$_1$-[ma]$_2$ [la]$_3$-[la]$_4$
[mæ] & [læ]	[mæ]$_1$-[læ]$_1$ [mæ]$_2$-[læ]$_2$ [mæ]$_3$-[læ]$_3$ [læ]$_4$-[mæ]$_4$	[mæ]$_1$-[mæ]$_1$ [læ]$_2$-[læ]$_2$	[mæ]$_1$-[mæ]$_2$ [læ]$_3$-[læ]$_4$
[mɚ] & [lɚ]	[mɚ]$_1$-[lɚ]$_1$ [mɚ]$_2$-[lɚ]$_2$ [lɚ]$_3$-[mɚ]$_3$ [lɚ]$_4$-[mɚ]$_4$	[mɚ]$_3$-[mɚ]$_3$ [lɚ]$_4$-[lɚ]$_4$	[lɚ]$_2$-[lɚ]$_3$ [lɚ]$_1$-[lɚ]$_4$

Table 2-7. Exhaustive list of filler pairs in the discrimination test

Each of the filler pairs presented in Table 2-7 was presented once in the discrimination test (12 Filler-Different + 6 Filler-Identical + 6 Filler-Same = 24 filler pairs), plus the 24 experimental pairs, making a total of 48 pairs in the discrimination test.

2.4.3 Practice discrimination test

At the very beginning of the experiment, before the training session, participants took a practice discrimination test, where they were instructed to listen to pairs of English words and respond "same" if the pairs were of the same English word and "different" if the pairs were different English words. Participants were told that they would be doing this test again in a new language after a training session in the language. The purpose of the practice test was to get participants accustomed to the task and to pressing the "same" and "different" response buttons. The practice test was also intended to show participants that they should pay attention to phonemic differences between the words, and not to non-phonemic acoustic differences in making their "same" or "different" decisions. One half of the ten practice pairs were different words (e.g. *need* and *lead*) and one half were two separately-produced tokens of the same word (e.g. $need_1$ and $need_2$). All practice test pairs were unmodified productions by a male native speaker of American English, and were presented once each. Table 2-8 lists the practice test pairs.

Practice-Same	Practice-Different
$need_1$-$need_2$	*need-lead*
$loot_1$-$loot_2$	*root-loot*
$noon_1$-$noon_2$	*noon-moon*
$heal_1$-$heal_2$	*hear-heal*
$vast_1$-$vast_2$	*vast-fast*

Table 2-8. Exhaustive list of practice test pairs

Results of the practice test were not included in the analysis. After the practice test, participants continued directly to the training phase.

2.4.4 Organization of the experiment session

Each experiment session began with a short oral questionnaire in which participants were asked to detail their language background. They were then told that they would first be taking a short practice test using English words followed by a training phase where they would be listening to words in a language that they had never heard before, and that their task during the training phase was to listen carefully to the words in order to learn about the sounds of the new language. They were told that following the training phase, they would take the test again, only this time the test would use words from the *new* language. Participants were given a chance to ask questions before beginning the experiment. All parts of the experiment took place in a soundproof booth. Visual and auditory stimuli were presented using the DMDX experiment presentation software. All auditory stimuli were presented over headphones, and visual stimuli were presented on a computer screen that participants viewed through a window in the soundproof booth. Participants responded to test items by pressing labeled buttons on a keyboard that was placed on a table in front of them.

2.5 Results

Participants' responses in the discrimination test were scored as either correct ("same" responses to Experimental-Same, Filler-Same, and Filler-Identical pairs, and "different" responses to Experimental-Different and Filler-Different pairs) or incorrect ("different" responses to Experimental-Same, Filler-Same, and Filler-Identical pairs, and "same" responses to Experimental-Different and Filler-Different pairs). Overall, participants' performance on the Experimental-Same, Filler-Same, and Filler-Different test pairs was near perfect, while discrimination accuracy for the Experimental-Different pairs was under 50% for participants in all training groups.

A 2-factor ANOVA with training condition (6 levels: Bimodal, Monomodal, Contrast, No-Contrast, Near-Endpoint, and No-Training) and item condition (5 levels: Experimental-Same, Experimental-Different, Filler-Same, Filler-Different, Filler-Identical) was performed on the percent correct data, with training condition as a between-subjects variable and item condition as a within-subjects variable. There was a significant effect of item condition ($F(4,504) = 828.1$, $p < .001$) and a significant effect of training condition ($F(5,126) = 2.882$, $p < .05$), and a significant interaction of the two ($F(20,504) = 4.321$, $p < .001$). See Figure 2-4 for all Experiment 1 results.

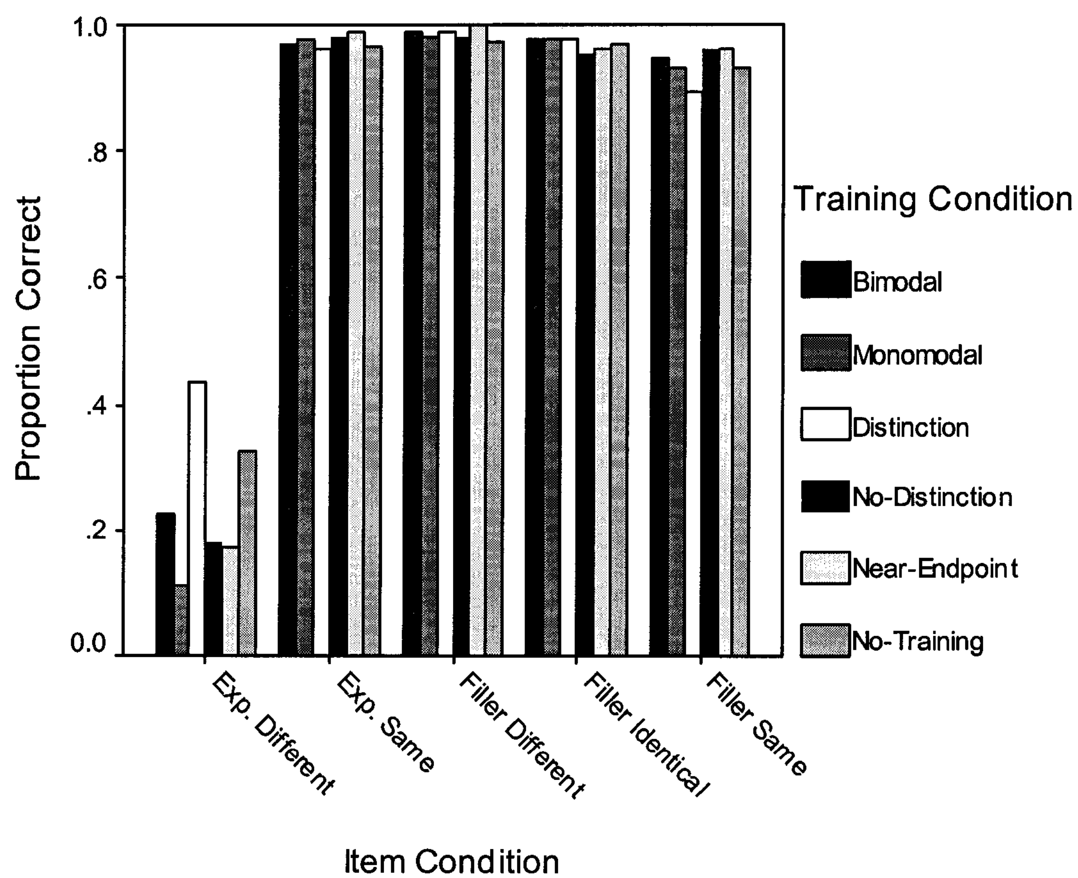


Figure 2-4. All Experiment 1 data, by item condition and training condition

The item condition of primary interest is Experimental-Different; the other four serve as controls to determine whether the Experimental-Different results are interpretable. Recall that performance on every item condition other than Experimental-Different was expected to have very low error rates, as native speakers of English were expected to respond "different" to Filler-Different pairs (e.g. [ma]-[la]) and "same" to Filler-Identical (e.g. $[ma]_1$-$[ma]_1$), Filler-Same (e.g. $[ma]_1$-$[ma]_2$), and Experimental-Same pairs (e.g. [ga]/[ka] token 1 - [ga]/[ka] token 1). Additionally, no training condition differences were predicted for this analysis. A 6 training condition x 4 item condition ANOVA was performed on the discrimination accuracy data for the four control item conditions, with training condition as a between-subjects variable and item condition as a within-subjects variable. There was a significant effect of item condition ($F(3,378) = 12.563$, $p < .001$) but no significant effect of training condition ($F(5,126) < 1$), and the interaction of the two was not significant ($F(15,378) = 1.133$, $p > .1$). That there was no main effect of training condition indicates that participants' performance on the Experimental-Different items can be interpreted as reflecting whether or not they perceive a phonological distinction between [g] and [k].

The significant effect of item condition in this analysis suggests that not all of the control conditions were equally difficult. To determine where the differences were among these control item conditions, a series of *post hoc* comparisons were done. See Figure 2-5 for the control condition means averaged over all training conditions.

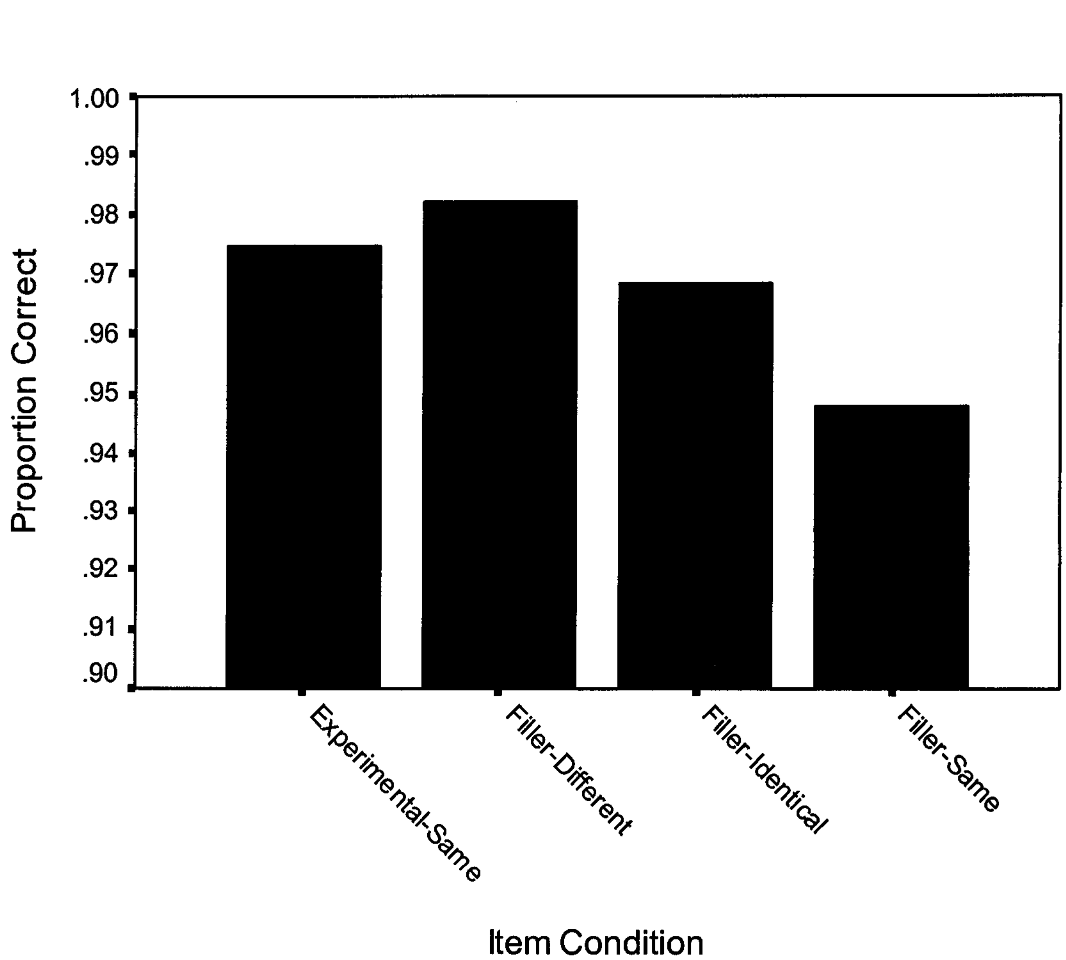


Figure 2-5. Means for the control item conditions, averaged across training condition

The Filler-Different items should have been very easy for the participants, since the difference between [ma] and [la] should be easily detectable by native speakers of English. On the other hand, the Filler-Same items were most likely the most difficult items, since they were two different productions of the same word—participants might have confused them as being two different words because of their acoustic differences. Participants exhibited greater accuracy on the Filler-Different items than the Filler-Same items ($F(1,438) = 18.369$, $p < .001$). Similarly, the Filler-Same items were more difficult than the Filler-Identical items ($F(1,438) = 5.884$, $p < .05$). The most important finding with respect to these control conditions, however, is that there was no main effect of training condition and no interaction between training condition and item condition.

There are several planned comparisons in this experiment. All of the planned comparisons consider data from the Experimental-Different pairs only. Next is a discussion of each comparison individually, followed by a discussion of the larger picture findings of Experiment 1. First, Figure 2-6 presents Experimental-Different data for all training conditions.

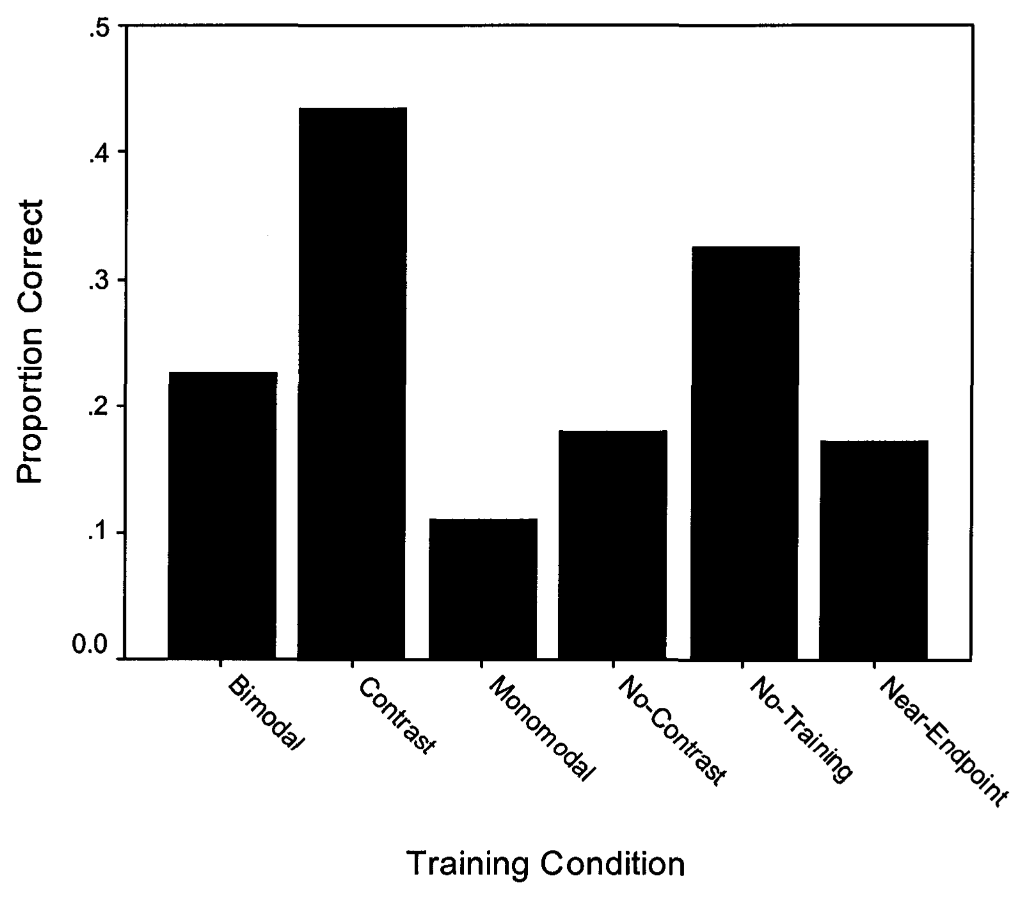


Figure 2-6. Experimental-Different discrimination accuracy for all training conditions (data repeated from Figure 2-4)

The first comparison of interest is between the Bimodal and Monomodal groups. Did Distribution-Based training in support of a distinction between [g] and [k] (Bimodal training) result in more accurate discrimination of the sounds than Distribution-Based training against the distinction (Monomodal training)? The Bimodal group and Monomodal group data (see the first and third bars from the left in Figure 2-6) was submitted to a between-subjects ANOVA with training condition as the independent variable (two levels: Bimodal and Monomodal) and percent correct as the dependent variable. The Bimodal training group, with 22.7% correct, performed significantly better than the Monomodal training group, with 11% correct, ($F(1,42) = 4.516$, $p < .05$). This replicates the results first reported in Maye (2000) that Distribution-Based training can affect discrimination of a novel phonemic contrast.

Next, did Lexical-Contrast-Based training in support of a distinction between [g] and [k] (Contrast training) result in more accurate discrimination of the sounds than Lexical-Contrast-Based training against the contrast (No-Contrast training)? The second and fourth bars in Figure 2-6 are the data for these two groups. The Contrast training group, with 43.6% correct, performed significantly better than the No-Contrast training group, with 18.2% correct, ($F(1,42) = 4.516$, $p < .05$). This supports the hypothesis that the availability of lexical information can affect discrimination of a novel phonemic contrast.

Now that we know that there is an effect of both Lexical-Contrast-Based *and* Distribution-Based evidence on performance in this discrimination task, the next question is which type of evidence has a greater influence on learning the contrast. The comparison between the Bimodal training condition and the Contrast training condition can indicate whether the lexical information provided in the Contrast condition serves to enhance participants' discrimination ability more than the distributional information provided in the Bimodal condition. Both groups received training that indicated that there is a distinction between [g] and [k], but one received Distribution-Based evidence and the other Lexical-Contrast-Based evidence. Thus this comparison can help to show which type of evidence has more impact on learning to discriminate a novel phonemic contrast. The Contrast training group, with 43.6% discrimination accuracy, performed significantly better than the Bimodal training group, with 22.7% accuracy, ($F(1,42) = 7.039$, $p < .05$). This provides preliminary evidence that the effect of Lexical-Contrast-Based evidence has a greater influence than that of Distribution-Based evidence on learning to discriminate a novel phoneme contrast. However, there was a confound in the Contrast and No-Contrast training conditions between Lexical-Contrast-Based evidence and Distribution-Based evidence, as participants in the Contrast and No-Contrast groups were actually presented auditory tokens in a bimodal distribution on the continuum because they heard only tokens 2 and 7. The Near-Endpoint training condition, discussed next, was introduced to address this confound.

The comparison between the Contrast training condition and the Near-Endpoint training condition can indicate whether the lexical information provided in the Contrast condition serves to enhance participants' discrimination ability, or whether their discrimination performance is simply an effect of the bimodal distribution of tokens in

Contrast training (tokens 2 and 7 only). The Contrast training group, with 43.6% discrimination accuracy, performed significantly better than the Near-Endpoint training group, with 17.4% accuracy, ($F(1,42) = 10.528$, $p < .01$). It can therefore be concluded that the addition of lexical information in the Contrast training condition serves to enhance discrimination performance by these participants over the Near-Endpoint training participants who heard training tokens in the same distribution but without lexical information in support of the contrast. This finding, coupled with the finding for the Contrast versus Bimodal conditions, suggests an important influence of lexical information in learning novel phoneme contrasts. So far, it has been shown that discrimination by participants who received lexical information plus auditory tokens in a bimodal distribution without phonetic variability (Contrast training group) was better than that by participants who received no lexical information with the same distribution of auditory tokens (Near-Endpoint training group).

Further confirmation of this result might be found by comparing discrimination performance by the No-Contrast group to that by the Near-Endpoint group. Participants in both Lexical-Contrast-Based training conditions heard tokens in the same distribution (tokens 2 and 7 only—a bimodal distribution), but the No-Contrast group received lexical evidence against a distinction between [g] and [k]. Therefore, this comparison can help to further indicate the relative influence of lexical and distributional information. The data of interest in this comparison can be found in the fourth and sixth bars from the left in Figure 2-6. However, the No-Contrast training group, with 18.2% discrimination accuracy, did not perform differently from the Near-Endpoint training group, with 17.4%

accuracy, (F(1,42) = .013, p = .911). Thus the addition of lexical evidence against a distinction does not appear to influence discrimination, at least when combined with distributional information (items 2 and 7 only) that supports a distinction.

The next comparison tests the impact of phonetic variability during training on discrimination performance. Recall that participants in the Bimodal training group heard every token on the continuum, with distributional peaks at tokens 2 and 7, and that the Near-Endpoint training group heard only tokens 2 and 7. See the first and sixth bars in Figure 2-6 for data from these two groups. There was no significant difference between performance by participants in the Bimodal training group, with 22.7%, and performance by participants in the Near-Endpoint training group, with 17.4% accuracy (F(1,42) > 1, p > .1). The result for these training groups, then, does not provide support for the hypothesis that phonetic variability in training enhances sensitivity to a novel phoneme contrast. This does not mean that phonetic variability does not lead to more robust phonetic learning, as this is a null result, and the kind of variability explored here is only one of several possible types of phonetic variability that might be explored.

Next we can consider the more general issue of the impact of training of any kind on discrimination performance, by comparing the Bimodal, Contrast and Near-Endpoint conditions to the No-Training condition. The purpose of these comparisons is to test the hypothesis that participants with training that supports a contrast between [g] and [k] would exhibit better discrimination accuracy than participants without any training in the new language. The No-Training condition was intended to provide a baseline against which the effects of training could be compared.

There was no significant difference between performance by participants in the No-Training group, with 32.6%, and performance by participants in the Bimodal training group ($F(1,42) = 1.453$, $p > .1$), the Contrast training group ($F(1,42) = 1.361$, $p > .1$), or the Near-Endpoint training group ($F(1,42) = 3.28$, $p = .077$). This null result indicates at least that the No-Training participants discriminated the sounds quite well, and that the impact of training that indicated the phonemic status of the distinction between [g] and [k] did not serve to significantly enhance the discrimination accuracy that participants had prior to the experiment.

Given that there were no differences for the Bimodal, Contrast and Near-Endpoint conditions versus the No-Training condition, it might be inferred that participants prior to training were relatively good at discriminating [g] and [k], which is not unexpected given the studies presented in Chapter 1 that suggest that native speakers of English do have some sensitivity to the difference between [g] and [k] in word-initial position. Another way of considering this issue, then, is to determine whether the Monomodal and No-Contrast training conditions had the effect of *suppressing* a distinction that the participants were in fact able to discriminate prior to training.

There was a significant difference between performance by participants in the No-Training group and performance by participants in the Monomodal training group ($F(1,42) = 8.106$, $p < .01$), but there was not a significant difference between discrimination by the No-Training group and the No-Contrast training group ($F(1,42) = 2.989$, $p = .091$). Thus it appears that Distribution-Based evidence against the contrast between [g] and [k] *suppresses* an ability that participants had to discriminate the sounds before training. It might be possible to interpret the No-Training versus No-Contrast difference as a non-significant trend. A more liberal interpretation of these comparisons, then, is that participants start with a baseline ability to discriminate [g] and [k] that is suppressed by training of either kind against the distinction.

The next issue to consider is whether Distribution-Based or Lexical-Contrast-Based evidence against the distinction results in poorer discrimination accuracy. Since we know that the only effect of training appears to be to suppress discrimination of [g] and [k], we also want to know which of the two training conditions against the contrast had a greater suppressive effect. The difference between the Monomodal training and No-Contrast training groups' discrimination accuracy, at 11% and 18.2% respectively, is not significant ($F(1,42) = 1.566$, $p > .1$). It cannot be concluded, therefore, that the Distribution-Based and Lexical-Contrast-Based training against a distinction affect discrimination accuracy differently, even though the Contrast training group exhibited more accurate discrimination than the Bimodal group.

2.6 Discussion of Experiment 1 results

2.6.1 Summary of results

Table 2-9 lists the planned comparisons and the results for each.

Significant Comparisons (p < .05)
Bimodal > Monomodal
Contrast > No-Contrast
Contrast > Bimodal
Contrast > Near-Endpoint
No-Training > Monomodal

Non-Significant Comparisons (p > .1)
Bimodal, Near-Endpoint
Monomodal, No-Contrast
Bimodal, No-Training
No-Contrast, Near-Endpoint
Contrast, No-Training

Non-Significant Trends (.05 < p < .1)
Near-Endpoint (>) No-Training
No-Contrast (>) No-Training

Table 2-9. Summary of Experiment 1 comparisons

2.6.2 Why did participants in the No-Training condition discriminate [g] and [k] so well?

Recall that Pegg & Werker (1997) demonstrated that native speakers of English without any training are able to discriminate sub-phonemic phonetic differences between prevoiced [d] and voiceless unaspirated [t] under certain task conditions. Similarly, they can discriminate prevoiced [g] and [k] to some extent, as reported in Maye 2000. In fact, this ability was exploited in the experiments presented here because it was desirable to use a distinction that was not too difficult for native speakers of English to detect— it is thus not surprising that participants in the No-Training condition discriminated [g] and [k] with the accuracy that they did—what is surprising is that the accuracy with which participants in the No-Training condition discriminated [g] and [k] was coupled with near-perfect discrimination of the Filler-Same and the Filler-Identical test pairs.

The Filler-Identical pairs, as discussed above, were included in the set of test items in order to give all participants regardless of training condition some pairs to which they would respond "same". These pairs were presentations of identical filler stimuli (e.g. two of the same exact production of the filler word [ma]), and all participants responded "same" to these pairs as expected. The Filler-Same pairs, on the other hand, were two different productions of the same filler word (e.g. two different productions of the filler word [ma]).

The Filler-Same tokens, then, were not *phonemically* different despite the small sub-phonemic acoustic differences between them. It is assumed that "different" responses to the Filler-Same tokens would have indicated that participants were attending to small sub-phonemic differences in the discrimination task, and not to phonemic differences, and might have indicated that "different" responses to the Experimental-Different pairs reflected only their sensitivity to the small sub-phonemic differences and *not* knowledge that the differences corresponded to a phonemic contrast. Therefore, the fact that participants in the No-Training group *ignored* sub-phonemic differences between filler items (indicated by "same" responses to Filler-Same items) should mean that "different" responses to the Experimental-Different pairs can be interpreted as indicating that they have knowledge of the phonemic contrast.

There is no reason to expect that these monolingual English speakers treat [g] and [k] as contrastive because they are not contrastive in English…or are they? It may in fact be argued that prevoiced [g] belongs to the English category /g/, while unaspirated [k] is an allophone of English /k/ when it follows [s] as in the word *ski*, and therefore these phones are indeed contrastive in English. However, it is important to recall that prevoiced [g] and unaspirated [k] both occur as allophones of /g/ in word-initial position, and that /k/ is realized as [k^h], *not* [k], in word-initial position. Therefore, these two phones are not contrastive word-initially in English. Since all [g] and [k] sounds in these experiments appeared word-initially, it is instead concluded that the participants in this study were trained on a novel distinction that is specific to a particular context.[6] Thus it is assumed that the monolingual English speakers do not treat [g] and [k] as contrastive word-initially prior to the experiment.

Another possible explanation for the apparently "phonemic" treatment of the [g]-[k] distinction by the No-Training group may be that in the absence of any other

---

[6] This may be a critical point, as it has been shown that listeners without training perform abysmally on the discrimination of entirely novel non-native contrasts (where the contrast does not exist in the native language even in different environments in the native and second languages). It may be found that the effect of Bimodal or Contrast training is to enhance discrimination of some contrasts but not others, or to suppress discrimination of some contrasts but not others. This issue is left for future research.

information about the new language (no training prior to the discrimination task), they deduced the phonemic status of the [g]-[k] distinction from hearing the sounds juxtaposed repeatedly during the discrimination task. The test therefore may have served as training in itself. It is possible to explore this possibility by checking the No-Training participants' data for order effects. If their discrimination accuracy for Experimental-Different items is higher in the second half of the task than in the first, for example, it may be that participants in the No-Training condition learned about the contrast during the discrimination test itself. Figure 2-7 below shows discrimination accuracy for the Experimental-Different items in the first half and the second half of the discrimination test for the No-Training condition.

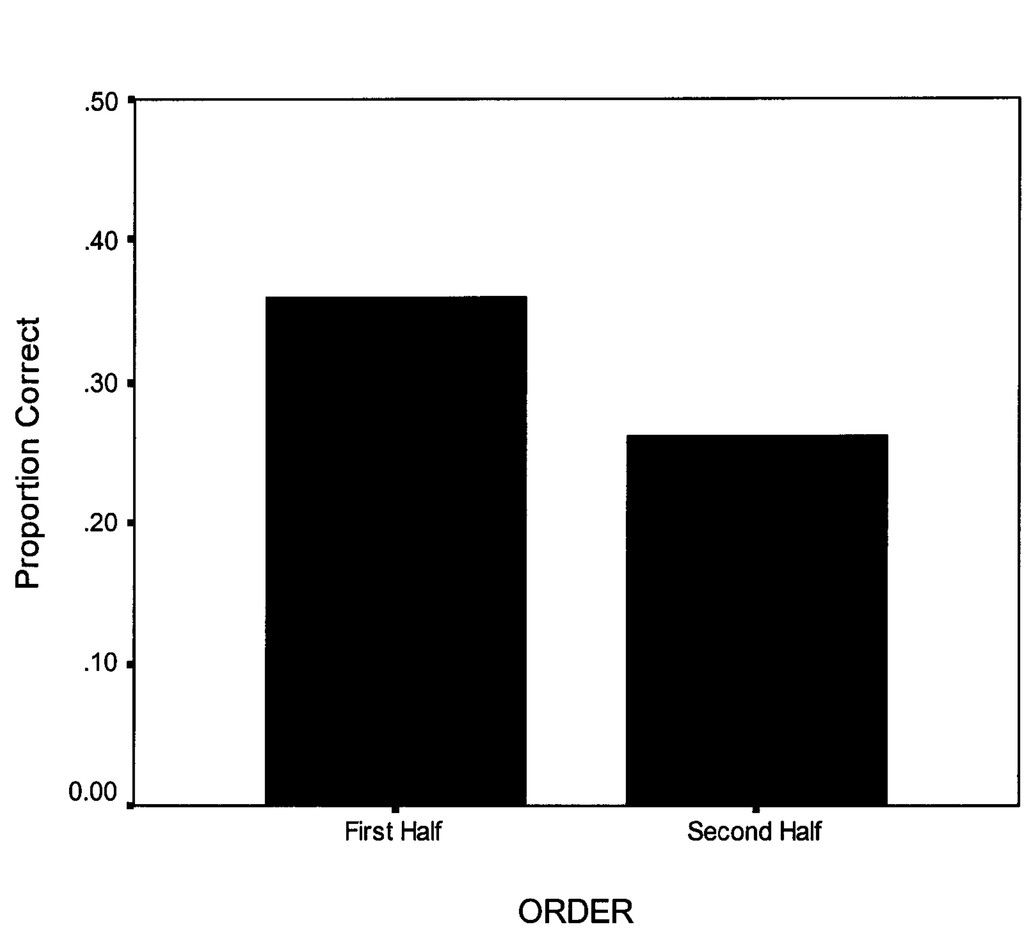


Figure 2-7. No-Training discrimination accuracy on Experimental-Different items for first-half and the second-half the discrimination test

There was no significant difference in discrimination accuracy for the first half (35.9%) and second half (26.2%) of the discrimination test ($F(1,62) < 1$). In addition to the lack of significant difference between the accuracy rates for the first and second halves of the discrimination task, the direction of the means for the first half and second half is the opposite of that expected if the discrimination test served as training for participants in the No-Training condition. Thus this comparison does not suggest that participants learned about the novel contrast over the course of the discrimination test.

The question of why participants in the No-Training condition were sensitive to the acoustic difference between [g] and [k] in the Experimental-Different stimuli such that they were able to discriminate them quite accurately, yet they appropriately ignored comparable acoustic differences in the Filler-Same pairs remains. It is possible that participants were faced with *two different kinds* of acoustic differences. In the Experimental-Different pairs, the difference between the two was systematic—between [g] and [k]. However, in the Filler-Same pairs, the differences between the two stimuli were any of a variety of random differences that occur between two independent productions of the same word. The systematicity of the [g]-[k] difference in the Experimental-Different pairs may have focused listeners' attention on these pairs and may account for their sensitivity to the differences while at the same time the randomness of the difference in the Filler-Same pairs may have been more readily ignored. This final explanation seems to be the most plausible; it is therefore assumed that participants in the No-Training condition *did not* treat the distinction between [g] and [k] as phonemically contrastive even though their "different" responses to these Experimental-Different pairs was coupled with "same" responses to the Filler-Same pairs.

2.6.3 Distribution-Based versus Lexical-Contrast-Based evidence and novel phonemic contrast learning

This experiment was designed to shed some light on the relationship between Distribution-Based and Lexical-Contrast-Based evidence in the acquisition of novel phoneme contrasts. The most accurate discrimination of the novel phonemic contrast was by participants who were presented Lexical-Contrast-Based evidence for a distinction between [g] and [k] (the Contrast group). However, as discussed above, these participants also received a kind of Distribution-Based evidence for the distinction, as they heard only tokens 2 and 7 during training. One potential solution for this issue might be a follow-up experiment that presents the auditory tokens on a flat distribution between 1 and 8, with tokens 1-4 associated with one picture and tokens 5-8 associated with another. However, this distribution may in fact provide participants with evidence against a distinction between [g] and [k], and thus does not eliminate the confound. Experiment 2 provides a different kind solution for this problem.

2.6.4 Conclusions drawn from Experiment 1

There are several conclusions that can be drawn from the results of Experiment 1. First, the results indicate that both Distribution-Based and Lexical-Contrast-Based evidence impact discrimination accuracy by learners. Participants in the Contrast condition exhibited the most accurate discrimination of all the conditions, and were significantly more accurate than the Bimodal group, who received only Distribution-Based evidence for the distinction. However, a potential confound between Distribution-Based and Lexical-Contrast-Based information in the training presented to participants in the Contrast group makes a comparison of the Bimodal and Contrast groups difficult to interpret. The Near-Endpoint condition was added to clarify the issue. It was found that the Contrast group, participants who received both Lexical-Contrast-Based and Distribution-Based (the near endpoints only) evidence for a distinction between [g] and

[k], also discriminated the sounds more accurately than the Near-Endpoint group. This indicates that the addition of lexical information in the Contrast training condition enhances the influence on discrimination. However, there were no differences between the Bimodal and Near-Endpoint groups, suggesting that phonetic variability during did not significantly enhance discrimination accuracy.

Finally, it was found that participants in the No-Training group did not discriminate [g] and [k] more poorly than participants in any of the conditions that provided evidence in support of a contrast, and that the No-Training group performed significantly better than the Monomodal group and nearly-significantly better than the No-Contrast group. These findings together suggest that participants started the experiment with some ability to discriminate [g] and [k], and that training had the effect of suppressing that ability. The main findings of Experiment 1 are summarized as follows:

- Distribution-Based and Lexical-Contrast-Based training both influenced monolingual English listeners' discrimination of [g] and [k].
- Both Distribution-Based and Lexical-Contrast-Based training *against* a phonemic contrast between [g] and [k] suppressed listener's perception of the distinction compared to an untrained group.
- Neither Distribution-Based nor Lexical-Contrast-Based training *for* a phonemic contrast between [g] and [k] enhanced listener's perception of the distinction compared to an untrained group.
- Phonetic variability in the experimental training stimuli did not enhance listener's perception of the distinction between [g] and [k].

Because it is a main objective of this project to uncover the relative influence of Distribution-Based and Lexical-Contrast-Based evidence in learning novel phoneme distinctions, a second experiment was performed to tease apart the effects of both types of evidence during training. The issue of the relative influence of Distribution-Based evidence and Lexical-Contrast-Based evidence is thus considered in further detail in Experiment 2.

CHAPTER 3

EXPERIMENT 2: INTERACTION OF LEXICAL-CONTRAST-BASED AND DISTRIBUTION-BASED LEARNING

3.1 Introduction

Experiment 1 provided information about the relative influence of Distribution-Based and Lexical-Contrast-Based evidence in the acquisition of novel phonemic contrasts. However, it was not possible to determine exactly how these two types of evidence interact in novel phoneme learning. This experiment aims to explicitly address the way these two types of evidence interact. In this experiment, participants were presented both types of evidence at the same time in combination. Thus participants received Contrast-Based evidence like that of the Monomodal or Bimodal training groups from Experiment 1 plus Lexical-Contrast-Based evidence like that of the Contrast or No-Contrast training groups at the same time. Table 3-1 demonstrates how the training conditions in Experiment 2 were constructed.

		Lexical-Contrast-Based	
		Contrast	**No-Contrast**
Distribution-Based	**Bimodal**	BIMODAL+CONTRAST	BIMODAL+NO-CONTRAST
	Monomodal	MONOMODAL+CONTRAST	BIMODAL+NO-CONTRAST

Table 3-1. Combination training conditions used in Experiment 2; conditions where the Distribution-Based and Lexical-Contrast-Based evidence conflict are shaded

The training conditions in this experiment allow us to address the following questions:

- Do participants exhibit even more accurate discrimination of [g] and [k] when they are trained with both Distribution-Based and Lexical-Contrast-Based evidence in support of a distinction than when they are trained with only one or the other?[7]
- Do participants exhibit even poorer discrimination of [g] and [k] when they are trained with both Distribution-Based and Lexical-Contrast-Based evidence against a distinction than when they are trained with only one or the other?
- What happens when Distribution-Based and Lexical-Contrast-Based evidence conflict? Which type of evidence has a stronger influence on learning or suppressing a novel contrast?

In Experiment 1, no differences were found in discrimination accuracy between participants who received Bimodal and Contrast training (both in support of a contrast between [g] and [k]) or between participants who received Monomodal and No-Contrast training (both providing evidence against a contrast between [g] and [k]). It was thus not possible to determine the relative impact of Lexical-Contrast-Based and Distribution-Based training on novel phonemic contrast learning. By training participants in this experiment using the combinations of evidence described in Table 3-1 it might be possible to tease apart the relative influence of these types of evidence.

[7] Recall that the Contrast training group from Experiment 1 received Lexical-Contrast-Based evidence plus a kind of Distribution-Based evidence because they heard only the near-endpoint tokens during training. In this experiment, participants who receive both types of training hear the entire range of tokens in the phonetically variable Bimodal distribution—this more accurately mimics the kind of input that learners are likely to receive in a natural setting.

3.2 Participants

Participants were 88 native English speakers who met the same criteria as those in Experiment 1, but they did not participate in Experiment 1. These participants had between 1 and 14 years of foreign language instruction (mean = 4.2 years). All spoke English exclusively on a daily basis, with their families, and at home.

Participants were randomly assigned to one of the four training conditions, for a total of 22 participants per condition. The training conditions are described in further detail in the section 3.3 below.

3.3 Training stimuli and methods

As mentioned above, the training conditions in Experiment 2 are combinations of the properties of the training conditions in Experiment 1—combinations of Distribution-Based and Lexical-Contrast-Based learning. Each is described in detail below.

3.3.1 Bimodal+Contrast training

Participants in the Bimodal+Contrast training condition heard the experimental auditory tokens in a bimodal distribution between [g] and [k]; in the same way as the Bimodal training condition in Experiment 1. Additionally, they saw pictures associated with the tokens, as Lexical-Contrast-Based evidence. Tokens 1-4 were associated with one picture; tokens 5-8 with another.

This condition is very similar to the Contrast training condition in Experiment 1, where participants received Lexical-Contrast-Based evidence for the distinction and heard only tokens 2 and 7. However, in the Bimodal+Contrast training condition, the distribution of auditory stimuli includes phonetic variability and is exactly the same as that for the Bimodal training group, allowing for a direct comparison to the Bimodal participants' data. There was no significant difference between discrimination accuracy by the Bimodal training and the Near-Endpoint training groups in Experiment 1, where the former heard a bimodal distribution of the experimental training tokens that included phonetic variability and the latter heard a bimodal distribution without phonetic variability. Though there was no difference between these two groups in Experiment 1, it is possible that the addition of Lexical-Contrast evidence in a comparison between the Bimodal+Contrast condition from Experiment 2 and the Contrast condition from Experiment 1 will change the influence of phonetic variability. Comparisons between training conditions from Experiment 1 and Experiment 2 such as this will be addressed in section 3.6.

The Bimodal+Contrast condition is an agreement condition, where the Distribution-Based and the Lexical-Contrast-Based evidence are consistent. In this condition, the two types of evidence both support a distinction between [g] and [k]. Therefore participants in this training condition are expected to exhibit discrimination performance at test that is *at least* as accurate as that by the Bimodal and Contrast training groups in Experiment 1. Whether the combination of these two types of training *further* improves discrimination accuracy will be determined by comparing performance by the Bimodal+Contrast group to that of both the Bimodal and Contrast groups in Experiment 1.

3.3.2 Bimodal+No-Contrast training

The Bimodal+No-Contrast training condition is a combination of the Bimodal and No-Contrast conditions. Participants in this condition heard auditory tokens in a bimodal distribution but all tokens are associated with the same picture. In this way, participants in this training condition received conflicting information about the status of [ɡ] and [k] in the new language—the Distribution-Based evidence is consistent with there being a distinction, while the Lexical-Contrast-Based evidence is consistent with there being no distinction between [ɡ] and [k]. By providing participants in this condition with one type of evidence for the distinction and the other against the distinction we can begin to tease apart the relative influence of each type of evidence on novel phoneme contrast learning.

3.3.3 Monomodal+Contrast training

The Monomodal+Contrast training condition is a combination of the Monomodal and Contrast training conditions. In this condition, participants heard auditory tokens in a monomodal distribution but tokens 1 through 4 are associated with one picture and tokens 5 through 8 with another. Like the Bimodal+No-Contrast condition, participants in this training condition receive conflicting information about the status of [ɡ] and [k] — the Distribution-Based evidence is consistent with there being no distinction between [ɡ] and [k], while the Lexical-Contrast-Based evidence is consistent with there being a distinction between the two sounds.

The Bimodal+No-Contrast and Monomodal+No-Contrast training conditions make up a critical comparison for understanding the relative influence of the two types of evidence on the acquisition of novel phonemic contrasts. If the Bimodal+No-Contrast group exhibits superior discrimination accuracy to the Monomodal+Contrast group, this provides evidence that Distribution-Based information about phonemic status has a stronger influence on the language learner than Lexical-Contrast-Based information. However, if the Monomodal+Contrast group exhibits superior performance to the Bimodal+No-Contrast group, this provides evidence that Lexical-Contrast-Based information about phonemic status has a stronger influence on the language learner than Distribution-Based information.

3.3.4 Monomodal+No-Contrast training

The Monomodal+No-Contrast training condition is a combination of the Monomodal and No-Contrast training conditions. In this condition, participants hear auditory tokens in a monomodal distribution and all tokens are associated with the same picture. Like the Bimodal+Contrast condition, this is an agreement condition—both the Distribution-Based information and the Lexical-Contrast-Based evidence are consistent with there being no distinction between [g] and [k].

3.3.5 *Summary of the training conditions*

Figure 3-1 provides a summary of the training conditions in this experiment.

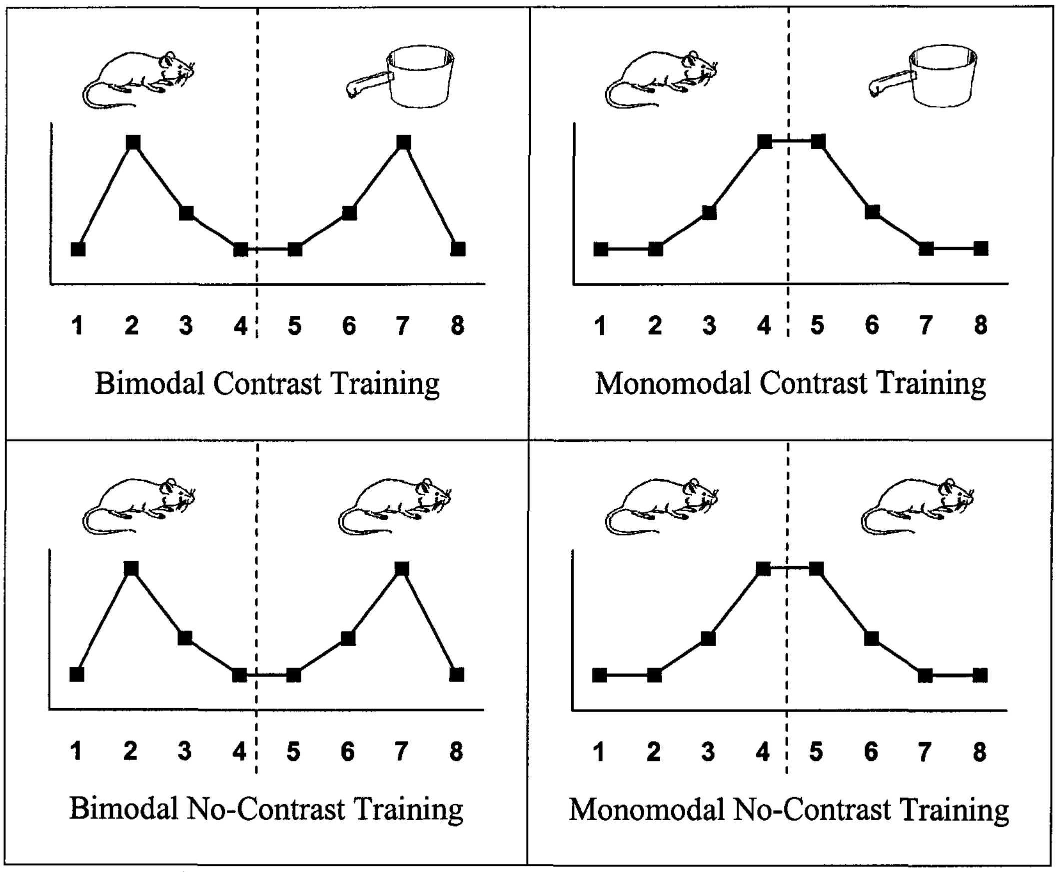


Figure 3-1. Summary of training conditions in Experiment 2

3.4 Testing stimuli and methods

The test phase is identical to that in Experiment 1, a discrimination test. All training methods, testing methods, and testing stimuli were identical.

3.5 Results

As in Experiment 1, participants' responses in the discrimination test were scored as either correct ("same" responses to Experimental-Same, Filler-Same, and Filler-

Identical pairs and "different" responses to Experimental-Different and Filler-Different pairs) or incorrect ("different" responses to Experimental-Same, Filler-Same, and Filler-Identical pairs and "same" responses to Experimental-Different and Filler-Different pairs).

A 2-factor ANOVA with training condition (6 levels: Bimodal, Monomodal, Contrast, No-Contrast, Near-Endpoint, and No-Training) and item condition (5 levels: Experimental-Same, Experimental-Different, Filler-Same, Filler-Different, Filler-Identical) was performed on the percent correct data, with training condition as a between-subjects variable and item condition as a within-subjects variable. There was a significant effect of item condition ($F(4,336) = 724.81$, $p < .001$) and a significant interaction of item condition and training condition ($F(12,336) = 2.817$, $p < .05$), but the main effect of training condition was not significant ($F(3,84) = 1.067$, $p > .1$). See Figure 3-2 for all Experiment 2 results.

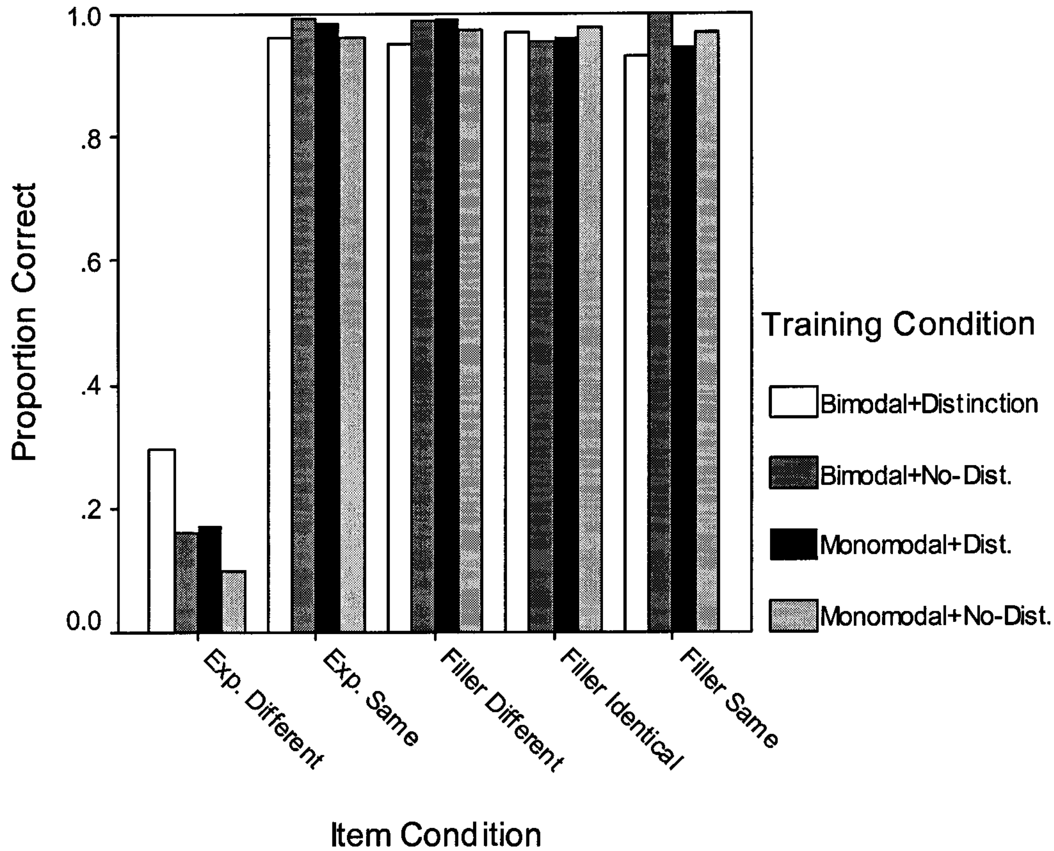


Figure 3-2. All Experiment 2 data, by item condition and training condition

As in Experiment 1, the item condition of primary interest is Experimental-Different; the others serve as controls to determine whether the Experimental-Different results are interpretable. A 4 training condition x 4 item condition ANOVA was performed on the percent correct data for the control conditions only, with training condition as a between-subjects variable and item condition as a within-subject variable. There was no significant effect of item condition ($F(1,84) = 2.708$, $p > .1$), no significant effect of training condition ($F(3,84) = 1.684$, $p > .1$), and the interaction of item condition and training condition was not significant ($F(3,84) = 1.492$, $p > .1$). These findings for the control item conditions indicate that participants' performance on the Experimental-Different items can be interpreted as reflecting whether or not they perceive a phonological distinction between [g] and [k].

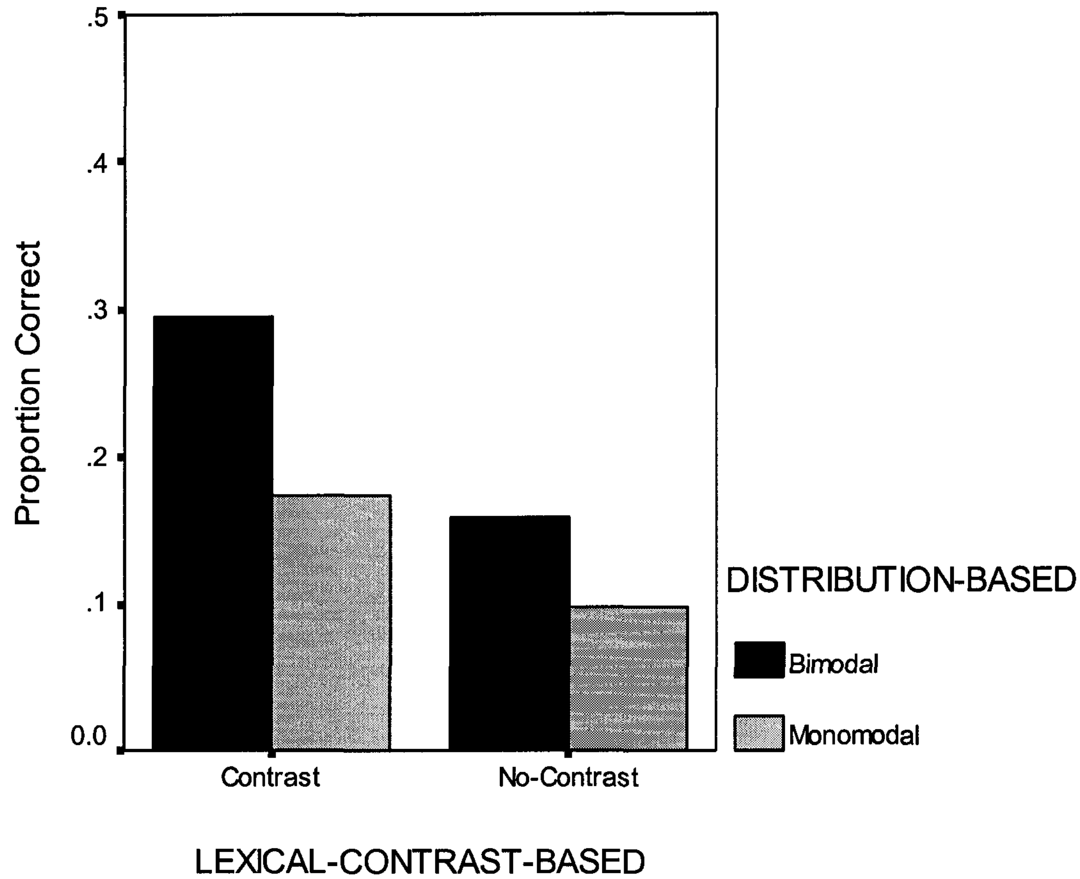


Figure 3-3. Discrimination accuracy for the Distribution-Based and Lexical-Contrast-Based conditions in Experiment 2

The data in Figure 3-3 is presented this way in order to show how the two types of evidence interact. A 2 Distribution-Based (Bimodal and Monomodal) x 2 Lexical-Contrast-Based (Contrast and No-Contrast) ANOVA was performed on this percent correct data for the Experimental-Different items only. There was a main effect of

Lexical-Contrast-Based evidence ($F(1,84) = 5.194$, $p < .05$) but the effect of Distribution-Based evidence was not significant ($F(1,84) = 3.816$, $p = .054$), and the interaction of the two was not significant ($F(1,84) < 1$).

There are several planned comparisons in this experiment. Next is a discussion of each comparison individually, followed by a discussion of the larger-picture findings of Experiment 2. To facilitate the interpretation of these individual comparisons, Figure 3-4 shows data for each group.

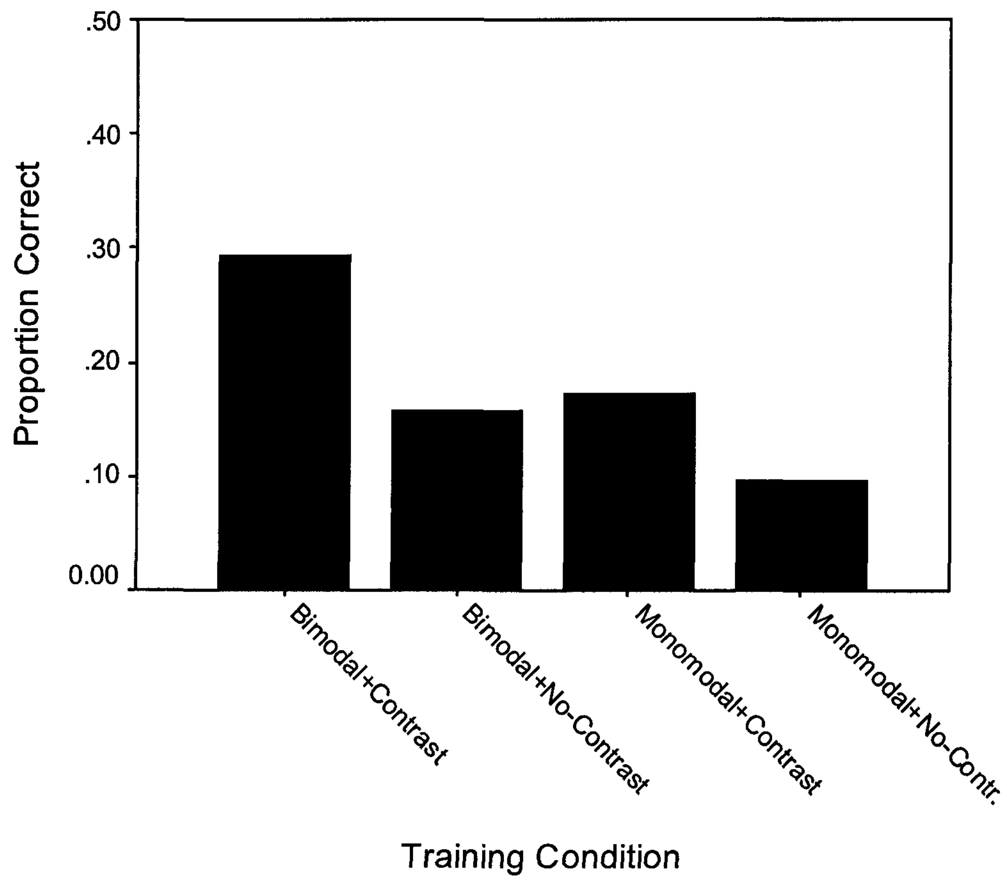


Figure 3-4. Discrimination accuracy for Experimental-Different pairs by all Experiment 2 training groups

The first comparison of interest concerns whether the participants who were provided only evidence in support of a contrast discriminated [g] and [k] more accurately than participants who received evidence against the contrast. Participants in the Bimodal+Contrast versus Monomodal+No-Contrast training conditions received agreeing evidence for or against a distinction between [g] and [k]. It was expected that there would be a significant difference in discrimination performance for these two groups, with the Bimodal+Contrast condition exhibiting superior discrimination to the Monomodal+No-Contrast condition. The Bimodal+Contrast training group, with 29.6% discrimination accuracy, performed significantly better than the Monomodal+No-Contrast training group, with 9.8% accuracy, ($F(1,42) = 6.674$, $p < .05$). This finding further confirms the findings from the comparisons of Contrast versus No-Contrast and Bimodal versus Monomodal in Experiment 1. It does not, however, contribute to an understanding of how the types of training interact. The following comparisons address this issue directly.

Participants in the Bimodal+Contrast versus Bimodal+No-Contrast training groups received Distribution-Based evidence in support of a contrast, but differed in whether the Lexical-Contrast-Based evidence was consistent with a contrast between [g] and [k]. It was hypothesized that since participants in the Bimodal+Contrast condition received both Distribution-based and Lexical-Contrast-Based evidence in support of a contrast, they would exhibit superior performance on the discrimination task than the Bimodal+No-Contrast group, who received Lexical-Contrast-Based evidence against the distinction. The Bimodal+Contrast training group, with 29.6% discrimination accuracy, did not perform significantly better than the Bimodal+No-Contrast training group, with 15.9% accuracy, ($F(1,42) = 3.114$, $p = .085$). However, this may be interpreted as a non-significant trend of interest,[8] and will be discussed further later.

Participants in the Distribution-Based training condition received evidence in support of a contrast, but differed in whether the Lexical-Contrast-Based evidence was consistent with a contrast between [g] and [k]. It was hypothesized, that since the Monomodal+No-Contrast group received both Distribution-based and Lexical-Contrast-Based evidence against a contrast, they would exhibit less accurate performance on the discrimination task than the Monomodal+Contrast group. The Monomodal+Contrast training group, with 17.4% discrimination accuracy, did not perform significantly better than the Monomodal+No-Contrast training group, with 9.8% accuracy, ($F(1,42) = 2.132$, $p > .1$).[9] This means that distributional Monomodal training neutralized the impact of the Lexical-Contrast-Based evidence presented to participants in these groups.

The comparison between Monomodal+Contrast versus Bimodal+No-Contrast is a critical comparison for understanding how Distribution-Based and Lexical-Contrast-Based evidence interact in novel phoneme contrast learning. Participants in both of these training groups received conflicting information about the presence or absence of a contrast between [g] and [k]. The Monomodal+Contrast training group, with 17.4%

[8] Actually, a one-tailed t-test is probably appropriate for this comparison, and results in a significant difference between the Monomodal+Contrast and Monomodal+No-Contrast groups.
[9] A one-tailed t-test might be appropriate here, but it does not produce a significant difference between these two groups.

discrimination accuracy, did not discriminate better than the Bimodal+No-Contrast training group, with 15.9% accuracy, ($F(1,42) = 2.287$, $p > .1$). Thus we do not have evidence that either Lexical-Contrast-Based or Distribution-Based evidence plays a stronger role in learning novel sound contrasts under this analysis.

3.6 Experiment 1-Experiment 2 Comparisons

It is assumed that data from Experiments 1 and 2 is comparable, as participants in both followed exactly the same test methods, and the training sessions were identical in length and the number of training stimuli. In fact, in terms of the amount of information presented during training (pictures plus auditory stimuli), Experiment 2 participants received the same amount of visual and auditory information as the Contrast and No-Contrast training group in Experiment 1. To the extent that the Contrast training group is comparable to the other conditions in Experiment 1, Experiment 1 and 2 training condition results can be compared.

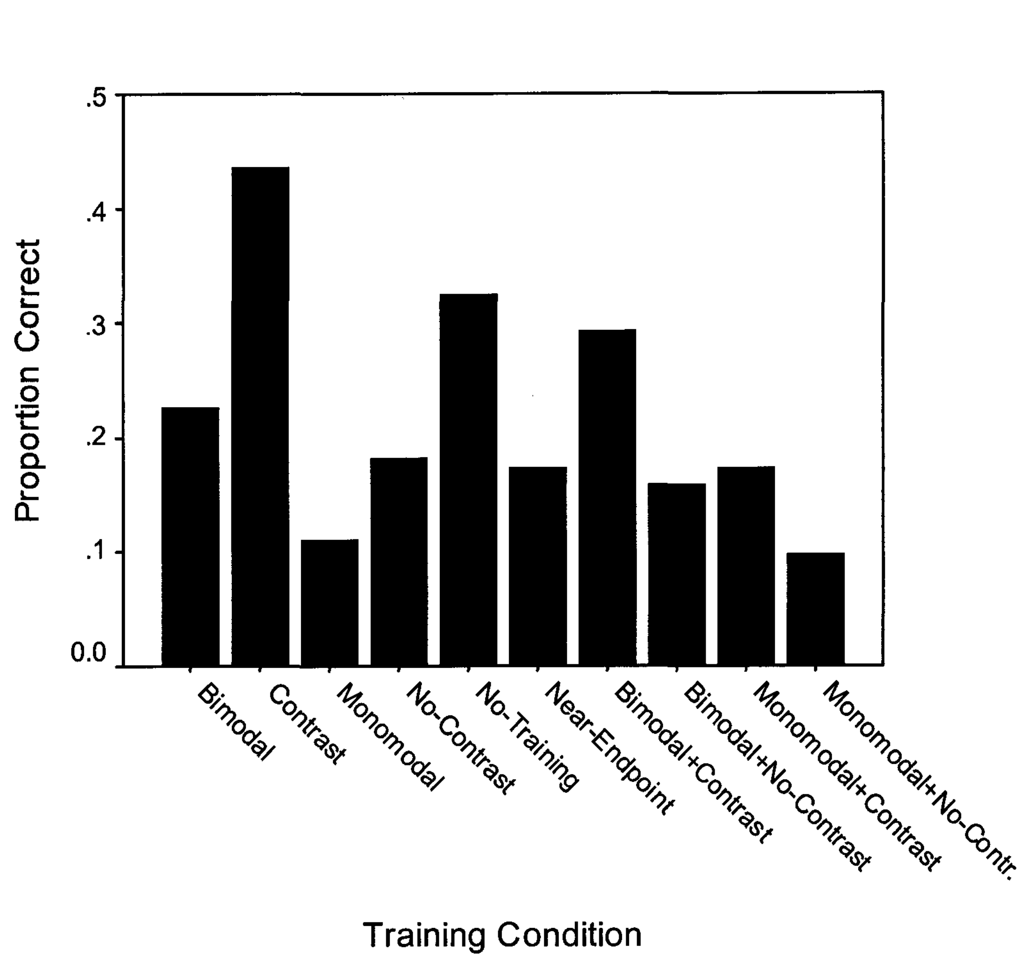


Figure 3-5. Proportion correct for Experimental-Different pairs by all Experiment 1 and 2 training groups

First, does the addition of Lexical-Contrast-Based evidence in the Bimodal+Contrast group lead to better discrimination accuracy over the Bimodal group, that received only Distribution-Based evidence? The discrimination accuracy rates for these groups are not significantly different ($F(1,42) < 1$), suggesting that Lexical-Contrast-Based evidence does not improve on what participants can learn from the Distribution-Based evidence alone.

As discussed above, the Contrast group received both Lexical-Contrast-Based and Distribution-Based evidence for a distinction between [g] and [k]. However, the Distribution-Based evidence that they received did not include phonetic variability. In the comparison between the Bimodal+Contrast and Contrast training conditions, we are able to compare the impact of phonetic variability on discrimination success when auditory training stimuli included phonetic variability (Bimodal+Contrast) or did not (Contrast). The discrimination accuracy rates for these groups are not significantly different ($F(1,42) = 2.287$, $p > .1$). This might serve as evidence that phonetic variability in the input did not serve to reinforce sensitivity to the distinction. This is consistent with the comparison of the Bimodal and Near-Endpoint training groups in Experiment 1, where there was no significant difference in discrimination accuracy for groups that received phonetic variability versus no phonetic variability in the absence of Lexical-Contrast-Based evidence.

In Experiment 1 it was found that although neither the Contrast nor the Bimodal training groups' discrimination accuracy was significantly better than that of the group that was not trained, the group that received Monomodal training in fact exhibited poorer discrimination accuracy than the untrained participants (while the difference between the No-Contrast and No-Training groups was not significant). This result was interpreted as suggesting that participants came to the experiment with some sensitivity to the distinction between [g] and [k], such that Monomodal training had the effect of suppressing a distinction that they were already able to detect. The question now is whether a combination of Distribution-Based and Lexical-Contrast-Based evidence against a contrast (Monomodal+No-Contrast) has this effect of suppressing sensitivity to the contrast as well. The No-Training group exhibited significantly more accurate discrimination of the Experimental-Different pairs than the Monomodal+No-Contrast group ($F(1,42) = 8.483$, $p < .01$). This finding is consistent with the results presented above for the Monomodal versus No-Training conditions in Experiment 1. These findings together suggest that training in these experiments served largely to suppress a contrast to which participants were already sensitive. This issue is discussed further in Chapter 5.

There was no significant difference in Experiment 1 between the Bimodal or Contrast groups and the No-Training group. Of interest in the present comparison is whether the Bimodal+Contrast group, that received both Distribution-Based and Lexical-Contrast-Based evidence, discriminated [g] and [k] more accurately than the participants who received no training (the No-Training group). There was no difference between the two groups ($F(1,42) < 1$). Thus it appears that the effect that training had on participants' discrimination performance was primarily to suppress sensitivity to the contrast (in the Monomodal and Monomodal+No-Contrast groups), and not to enhance sensitivity to the contrast. This is an interesting result in light of the results of Experiment 3, a word learning experiment, presented in Chapter 4.

3.7 Discussion of Experiment 2 results

3.7.1 Summary of results

Table 3-2 lists the planned comparisons just discussed and the results for each.

Significant Comparisons (p < .05)
Bimodal+Contrast > Monomodal+No-Contrast
No-Training > Monomodal+No-Contrast

Non-Significant Comparisons (p > .05)
Monomodal+Contrast, Monomodal+No-Contrast
Monomodal+Contrast, Bimodal+No-Contrast
Bimodal, Bimodal+Contrast
Contrast, Bimodal+Contrast
Bimodal+Contrast, No-Training

Non-Significant Trends (.05 < p < .1)
Bimodal+Contrast (>) Bimodal+No-Contrast

Table 3-2. Summary of Experiment 2 results

That the Bimodal+Contrast versus Monomodal+No-Contrast was significant indicating that Distribution-Based plus Lexical-Contrast-Based evidence in support of a novel sound contrast together result in greater discrimination accuracy than both kinds of evidence against the contrast. That the No-Training versus Monomodal+No-Contrast comparison was significant, however, provides evidence that participants were quite good at discriminating the contrast of interest, and that training against the contrast served to actually suppress listener's sensitivity to the contrast.

3.7.2 Conclusions drawn from Experiment 2

In this experiment, the Bimodal+Contrast group exhibited more accurate discrimination performance than the Mononomodal+No-Contrast group, indicating that a combination of Lexical-Contrast-Based evidence and Distribution-Based evidence in support of a contrast makes participants more able to discriminate the sounds than both kinds of evidence against the contrast.

However, most of the other planned comparisons in this experiment were not significant. The comparison that was included to indicate how the two types of evidence interact when they provide conflicting evidence, Bimodal+No-Contrast and Monomodal+Contrast, was not significantly different. Based on the results of Experiment 1, it might have been predicted that the Bimodal+Contrast would have more accurate discrimination than the Bimodal+No-Contrast group (similar to the Contrast and No-Contrast groups), but this comparison was only a non-significant trend.

The lack of a significant difference between both the Bimodal and the Contrast groups and the Bimodal+Contrast group suggests that the combination of the two types of evidence in support of a contrast does not significantly improve discrimination accuracy. Finally, paralleling the results in Experiment 1, the No-Training group performed significantly better than the Monomodal+No-Contrast group, but did not discriminate worse than the Bimodal+Contrast group. Again, it appears that the effect of training was to suppress an ability to discriminate sounds that participants had prior to the experiment session.

The main findings of Experiment 2 are summarized as follows:

- Lexical-Contrast-Based evidence does not improve on what participants can learn from the Distribution-Based evidence alone;
- Phonetic variability in the input does not serve to reinforce sensitivity to a phonemic distinction, consistent with the comparison of the Bimodal and Near-Endpoint training groups in Experiment 1; and
- Training in these experiments served largely to suppress a contrast to which participants were already sensitive.

The experiment presented in Chapter 4 tests whether participants in all of the above training conditions were able to extend their sensitivity to phonetic differences that they exhibited in Experiments 1 and 2 to establishing contrastive lexical representations of words that contain these sounds.

CHAPTER 4

EXPERIMENT THREE: LEXICAL REPRESENTATION OF NOVEL CONTRASTS

4.1 Introduction

While the results of the discrimination test in Experiments 1 and 2 can tell us whether or not the different training conditions affected how well participants were able to perceive the difference between the sounds [g] and [k], we still do not know whether or not participants are able to exploit this knowledge in subsequent linguistic development in the new language. Experiment 3 was performed in order to determine whether participants were able to establish lexical representations of new words in the new language that included representation of the [g]-[k] distinction. In other words, Experiment 3 tests whether participants' demonstrated sensitivity to language-specific sound contrasts (from Experiments 1 and 2) can be extended to learning new second language words—that is, the experiment tests the linguistic relevance of the different types of training. As discussed in Chapter 1, the ability to represent a phonetic difference in the lexicon is necessary before learners can be said to have learned a phonemic contrast.

There are two parts to Experiment 3. In the first part, participants were taught words in the new language associated with pictures to show their meanings. This part is called the "word learning phase". In the second part, called the "word testing phase", participants were given a matching test to evaluate their lexical representations for the new words.

4.2 Participants

Participants in this experiment are a subset of those who participated in Experiments 1 and 2. For many of the participants in each training condition, the discrimination test that they performed as part of Experiment 1 or 2 was immediately followed by this word learning experiment. Thus participants in Experiment 3 have identical characteristics to those in Experiments 1 and 2.

Those who participated in Experiment 3 were selected on the basis of scheduling constraints, and not on the basis of their performance in Experiment 1 or 2. They are therefore assumed to be a random sample of Experiment 1 and 2 participants, with eight from each training condition.

4.3 Training conditions

Immediately prior to participation in Experiment 3, participants had received training and taken a discrimination test as part of Experiment 1 or 2. There was no break between the experiments, so it is assumed that participants still had access to what they had learned during the training phase. The training conditions, then, are the ones already discussed in detail earlier, repeated for reference in Table 4-1.

From Experiment 1:	From Experiment 2:
Bimodal	
Monomodal	Bimodal+Contrast
Contrast	Bimodal+No-Contrast
No-Contrast	Monomodal+Contrast
Near-Endpoint	Monomodal+No-Contrast
No-Training	

Table 4-1. Training conditions from Experiments 1 and 2

4.4 Word Learning methods and stimuli

In the Word Learning Phase, participants were presented "words" in the new language along with pictures to indicate the meanings of the words. They were instructed to memorize these words and their meanings, and were later tested on their memory of the words in a matching test.

It was expected that participants would establish memory representations of the words that reflected their phonological knowledge of the new language. Thus it was predicted that if they had learned during training in Experiment 1 or Experiment 2 that the [g] and [k] sounds were distinct (as evidenced by their discrimination test performance), they might be able to exploit that knowledge further in learning words in the new language that contain these sounds. To test this prediction, participants were taught 18 words in the new language, 12 of which contained a [g] or a [k].

Experimental words were constructed using tokens 1 and 8 from the experimental continua from Experiments 1 and 2, because it was important that the "words" that participants learned in this experiment had identical acoustic properties to the stimuli that they were trained on earlier. Tokens 1 and 8 of the experimental stimuli continua were used because they were the same tokens used in the discrimination test from Experiments 1 and 2. In this way, the initial consonants in the tokens that participants discriminated in Experiments 1 and 2 were acoustically identical to the initial consonants in the words they learned this experiment.

The words needed to be different from each other without adding extra syllables that would burden memory unnecessarily, so they were constructed by simply adding codas to the training stimuli, which were all single open syllables. The procedure for doing this is explained below. First, Table 4-2 lists the training stimuli that were modified to use in this experiment, plus the codas that were added to them to form "words". For each continuum, four words were created—two from token 1 and two from token 8.

Training Stimuli from Experiments 1 and 2		Coda Added	"Word" Created
[ga] to [ka]	token 1	-ʃ	gaʃ
	token 1	-nt	gant
	token 8	-pt	kapt
	token 8	-v	kav
[gæ] to [kæ]	token 1	-vz	gævz
	token 1	-t	gæt
	token 8	-ŋ	kæŋ
	token 8	-sp	kæsp[10]
[gɚ] to [kɚ]	token 1	-k	gɚk
	token 1	-st	gɚst
	token 8	-ʒ	kɚʒ
	token 8	-m	kɚm

Table 4-2. The 12 experimental words taught during the word learning phase

The filler words were created from one token each of the filler stimuli from Experiments 1 and 2. They are listed in Table 4-3.

[10] The item [kæsp] may be problematic because it is likely to be perceived as identical to the English word *gasp*.

Training Stimuli from Experiments 1 and 2	Coda Added	"Word" Created
ma la	-mp -n	mamp lan
mæ læ	-ŋk -z	mæŋk læz
mɚ lɚ	-fs -p	mɚfs lɚp

Table 4-3. The six filler words taught during the word learning phase

To create these words, each of the training stimuli was prepared for the addition of a coda by deleting the final 50 milliseconds of the vowel. Figure 4-4 shows the training syllable [kɚ] before and after this manipulation.[11]

[11] All manipulations were made from the waveform representations of the words; cutoffs were made at 0-crossings to avoid "pops" and "clicks" in the sound files.

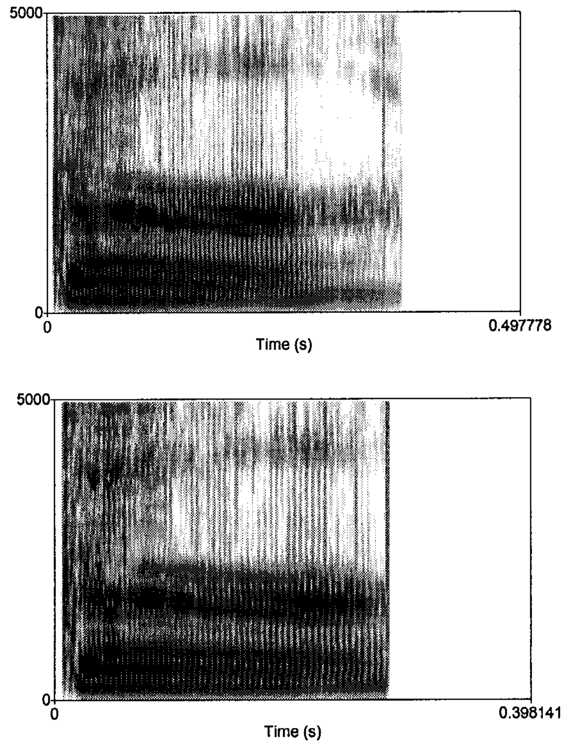


Figure 4-1. The training stimulus [kɚ], token 8, before (top) and after (bottom) the final 50 milliseconds of the vowel was removed

Next, a female native speaker of English produced several tokens of the "words" listed in Table 4-3. Ideally, the codas that were added to the ends of the training stimuli would have been produced by the same speaker as the training stimuli, but this was not possible for logistical reasons. From these recordings, the final 50 milliseconds of the vowel, plus the coda consonant, were copied from the whole word and pasted to the end of the shortened training stimuli. Figure 4-2 shows a spectrogram of the stimulus [kɚst].

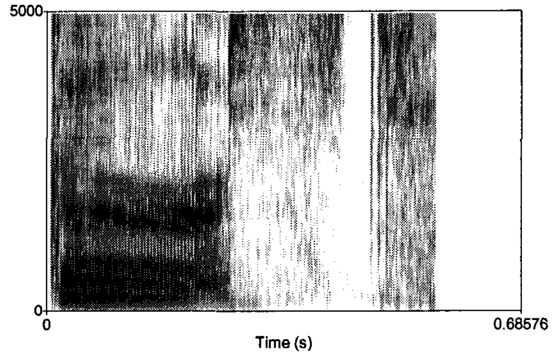
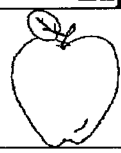
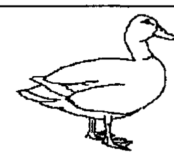
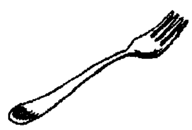
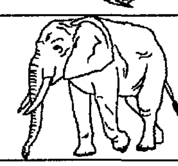
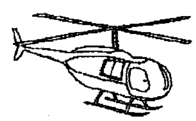
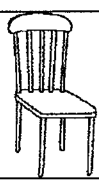


Figure 4-2. The "word" [kɚst]

Despite the small discontinuities visible in this spectrogram, these manipulated utterances sounded like fluently-produced words. Each "word" was presented with a picture to indicate its meaning. Table 4-4 lists all of the words and their meanings.

Experimental Words

gaʃ		kæŋ	
gant		kæsp	
kapt		gɚk	
kav		gɚst	
gævz		kɚʒ	
gæt		kɚm	

Filler Words

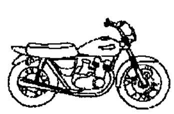
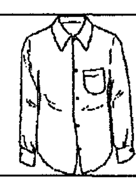
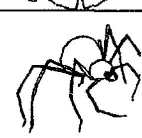
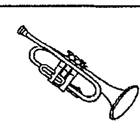
mamp		læz	
lan		mɚfs	
mæŋk		lɚp	

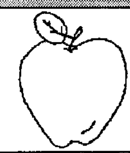
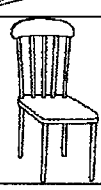
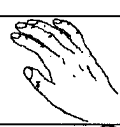
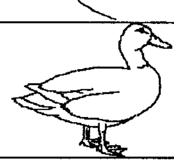
Table 4-4. Exhaustive list of experimental and filler words and their meanings

The set of words taught in Experiment 3 did not contain minimal pairs in order to prevent confounding word learning with Lexical-Contrast-Based training. If there had been minimal pairs in the word learning phase, participants may have learned from these pairs that the sounds [g] and [k] were contrastive, neutralizing the effects of the preceding training on participants' responses.

It was assumed that participants would store only contrastive information in their lexical representations of the words that they learned. Therefore, they were expected to be able to discriminate [g]-initial words from [k]-initial words in their representations if they treated the sounds as phonemically distinct, but not if they only had sensitivity to the phonetic difference between the sounds.

4.5 Practice test methods and stimuli

The next part of this experiment was a short practice test in which participants' memory of the words they had been taught was tested in order to make sure that they had reached a threshold of memory before going on to the test phase. It is important to note that their sensitivity to the distinction between [g] and [k] was not being tested here; instead, they were asked, for example, whether [kav] matched the picture of the hand (which they were taught during training was [kæsp]). Each of the pictures from training was presented twice: once accompanied by a correctly-matched word and once with an incorrectly-matched word. The participants' task was to determine whether or not the word they heard and the picture they saw were correctly matched. They responded by pressing a YES or a NO key. They were given three seconds to respond to each item, with a 1-second interval between items. All practice test items are presented in Table 4-5.

Picture	Correctly-Matched Word (Correct Response: Yes)	Incorrectly-Matched Word (Correct Response: No)
	gaʃ	kæŋ
	gant	kɚm
	kapt	gæt
	kav	gɚst
	gævz	gɚk
	gæt	kapt
	kæŋ	gant
	kæsp	kav
	gɚk	kæsp

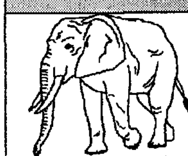
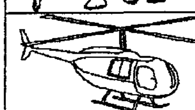
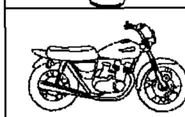
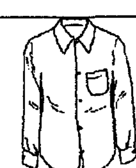
Picture	Correctly-Matched Word (Correct Response: Yes)	Incorrectly-Matched Word (Correct Response: No)
	gɚst	gaʃ
	kɚʒ	gævz
	kɚm	kɚʒ
	mamp	lɚp
	lan	læz
	mæŋk	mamp
	læz	mɚfs
	mɚfs	mæŋk
	lɚp	lan

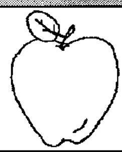
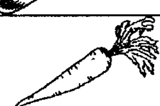
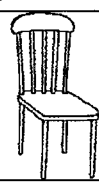
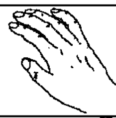
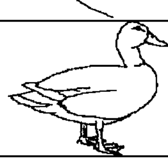
Table 4-5. Exhaustive list of practice test items

An accuracy threshold of 85% in this practice test was required before participants were permitted to continue to the final word learning test. If a participant scored below 85%, the program automatically returned to the beginning of the word learning phase. If, after two attempts at the practice test (twice through the loop of word learning and the practice test), a participant still did not reach the threshold, she did not continue with the experiment. Whether on the first or the second try, a participant who reached threshold on the practice test continued directly on to the final word learning test. 4.7% of participants did not reach the threshold and were excluded from the final word learning test.

4.6 Final word learning test methods and stimuli

In this last part of the experiment, participants were tested on whether or not they distinguished [g] and [k] in their lexical representations of the words learned during the experiment. The method for the final word learning test was identical to the practice test; the difference was in the incorrectly-matched test pairs. In this test, the Experimental Incorrectly-Matched test pairs were incorrect only in whether the word began with a [g] or a [k]. For this part of the experiment, a counterpart with the wrong initial consonant was made for each of the experimental words that were used in training. These "wrong" words were created by the same method as the original words from the word learning phase. Filler items were the six filler pictures, each presented once with the correctly-matched word and once with another filler word (Filler Incorrectly-Matched test pairs).

In the final word learning test, participants saw each picture two times: once with the correctly-matched word and once with an incorrectly-matched word; again, their task was to decide whether the word they heard and the picture they saw were correctly matched. Table 4-6 provides an exhaustive list of the final word learning test items.

Picture	Correctly-Matched Word (Correct Response: Yes)	Incorrectly-Matched Word (Correct Response: No)
	gaʃ	kaʃ
	gant	kant
	kapt	gapt
	kav	gav
	gævz	kævz
	gæt	kæt
	kæŋ	gæŋ
	kæsp	gæsp
	gɚk	kɚk

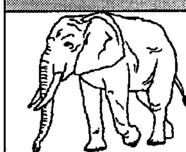
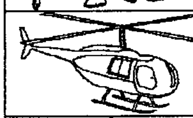
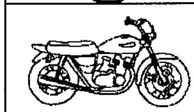
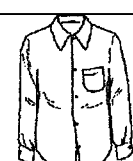
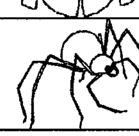
Picture	Correctly-Matched Word (Correct Response: Yes)	Incorrectly-Matched Word (Correct Response: No)
	gɚst	kɚst
	kɚʒ	gɚʒ
	kɚm	gɚm
	mamp	lamp
	lan	man
	mæŋk	læŋk
	læz	mæz
	mɚfs	lɚfs
	lɚp	mɚp

Table 4-6. Exhaustive list of final word learning test items

There were 24 experimental test items and 12 filler test items, for a total of 36 items in the final word learning test.

Experimental Item Conditions	Filler Item Conditions
Experimental Correctly-Matched Experimental Incorrectly-Matched	Filler Correctly-Matched Filler Incorrectly-Matched

Table 4-7. Item conditions in the final word learning test

If participants responded accurately to the Filler Correctly-Matched, the Filler Incorrectly-Matched, and the Experimental Correctly-Matched items, then this would mean that participants were able to remember words learned during the word learning phase. Only if this is true is it then possible to interpret the results of the Experimental Incorrectly-Matched items.

4.7 Results

Accurate responses to Experimental Incorrectly-Matched items would indicate that the participant noticed that the word presented with the picture began with the wrong consonant—[g] instead of [k] or [k] instead of [g]. In order to respond accurately to these items, participants would have to have created representations for the words in memory that encoded the [g]-[k] distinction. Then, when the word-picture item was presented, the participants would have to have noticed that the word they heard did not match their memory of the word.

Responses were scored as accurate ("yes" responses for the correctly-matched items and "no" responses for the incorrectly-matched items) or inaccurate, and response accuracy ("matching accuracy") was computed. The data from participants in all 10 training conditions from Experiments 1 and 2 were included in this analysis. A 2-factor ANOVA with training condition (10 levels: Bimodal, Monomodal, Contrast, No-Contrast, Near-Endpoint, No-Training, Bimodal+Contrast, Bimodal+No-Contrast, Monomodal+Contrast and Monomodal+No-Contrast) and item condition (4 levels: Experimental-Correct, Experimental-Incorrect, Filler-Correct and Filler-Incorrect) was performed on the percent correct data, with training condition as a between-subjects variable and item condition as a within-subjects variable. There was a significant effect of item condition ($F(3,210) = 1257.804$, $p < .001$) but no significant effect of training condition ($F(9,70) = 1.120$, $p > .1$), and no significant interaction of item condition and training condition ($F(27,210) < 1$). See Figure 4-3 for all Experiment 3 results.

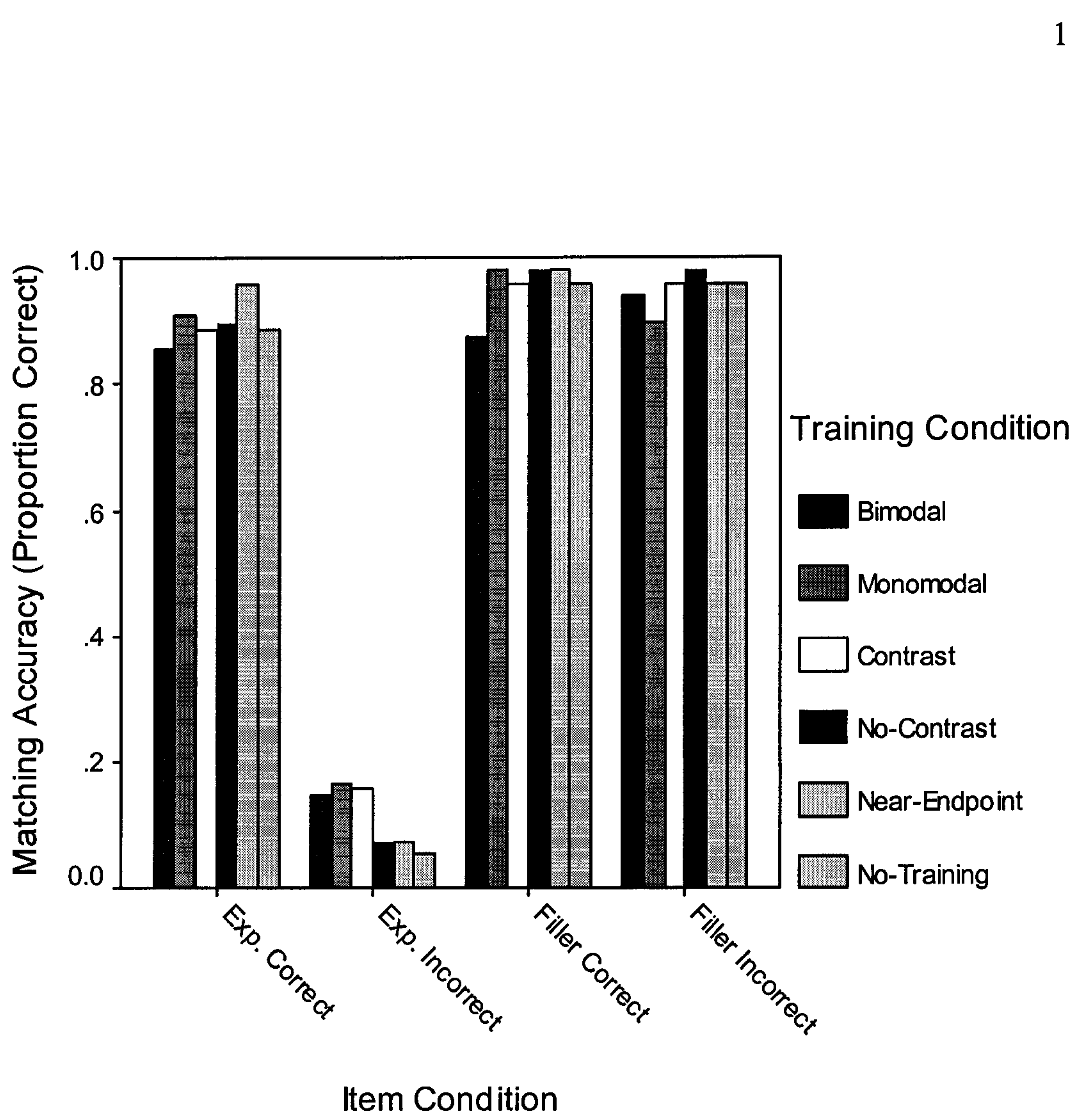


Figure 4-3. Final word learning test matching accuracy for all item conditions

For the Experimental-Incorrect test pairs only, there was no effect of training condition ($F(9,70) = 1.106$, $p > .1$). Additionally, accuracy rates in this condition, between 5.2% and 18.8% with a mean of 11.5%, were not much greater than inaccuracy rates in the filler conditions (Experimental-Correct: mean = 9.7%; Filler-Correct: mean = 4.2%; and Filler-Incorrect: mean = 4.6%). This suggests that regardless of training condition, participants were not able to extend their knowledge of the phonetic difference that they evidenced in Experiments 1 and 2 to create lexical representations that reflect a phonemic distinction between the sounds. While overall performance on the

Experimental-Incorrect test pairs was extremely low, it might be of interest to determine whether any participants in this study were able to represent the contrast lexically.

4.8 Did any participants extend the contrast to word learning?

Of the 80 participants whose data is included in this analysis, 31 responded correctly to no Experimental-Incorrect test pairs, 18 had an accuracy rate of 8%, 15 had an accuracy rate of 17%, seven had an accuracy rate of 25%, five had an accuracy rate of 33%, three had an accuracy rate of 42%, and one had an accuracy rate of 50%. To determine whether the participants who performed relatively well in the Experimental-Incorrect condition were better able to encode the [ɡ]-[k] contrast in their lexicons, or were simply guessing randomly while other participants were more strongly biased toward responding "same", a signal detection analysis was carried out for these participants. The three participants with accuracy rates of 42% were from the Bimodal, Monomodal+No-Contrast and Bimodal+Contrast training groups. This analysis compared their rates of "different" responses in the Experimental-Incorrect condition (where "different" is the correct response) and the Experimental-Correct condition (where "same" is the correct response), and their d' scores were 1.2, 0.24 and 0.47, respectively. These are all relatively small d' values, suggesting that perhaps only the participant from the Bimodal group indicated an ability to encode the [ɡ]-[k] distinction to some extent in her lexicon, and even this participant did not show a strong ability to do this. The participant with 50% accuracy (from the Bimodal+No-Contrast training group) can also be said to have represented the [ɡ]-[k] distinction to some extent in her developing lexicon because she responded "same" 100% of the time to Experimental-Correct items, yet responded "different" to 50% of the Experimental-Incorrect items. (Because the false positive rate is exactly zero, d' cannot be calculated for this participant.)

Because the four participants with accuracy rates above 40% belonged to the Bimodal, Bimodal+Contrast, Monomodal+No-Contrast, and the Bimodal+No-Contrast groups, training condition alone does not explain this high rate of accuracy, since two of these four participants were in training conditions that supported a contrast, one had conflicting evidence during training, and one was in a training condition that provided two kinds of evidence against the contrast. Table 4-8 presents matching accuracy rates for each of the item conditions for these four participants.

Participant	**A**	**B**	**C**	**D**
Training Condition	Bimod.	Bimodal+ Contrast	Monomod.+ No-Contrast	Bimod.+ No-Contrast
Experimental-Incorrect	42%	42%	42%	50%
Experimental-Correct (Control)	92%	75%	67%	100%
Filler-Incorrect (Control)	100%	100%	100%	83%
Filler-Correct (Control)	100%	100%	100%	100%

Table 4-8. Accuracy rates on all item conditions for the four participants who had >40% accuracy on Experimental-Incorrect test pairs in the final word learning test

If accuracy on the Experimental Incorrect pairs by these participants were the result of guessing, then they would have been as inaccurate on the three control conditions as they were accurate on the Experimental-Incorrect pairs. However, these four participants exhibited very accurate performance on the three control conditions, suggesting that their relatively accurate performance on the Experimental-Incorrect pairs reflects knowledge that [g] and [k] are phonemically distinctive. Further evidence for this can be found by examining the discrimination performance by these participants in the previous experiments.

Participant	**A**	**B**	**C**	**D**
Training Condition	Bimod.	Bimodal+ Contrast	Monomod.+ No-Contrast	Bimod.+ No-Contrast
Experimental-Incorrect Detection Accuracy	42%	42%	42%	50%
Experimental Different Discrimination Accuracy	42%	50%	58%	50%

Table 4-9. Discrimination accuracy on Experimental-Different items from Experiments 1 and 2 for the four participants who had >40% accuracy on Experimental-Incorrect test pairs

The average discrimination accuracy for Experiment-Incorrect items across all training groups in Experiments 1 and 2 was 21.8%. Thus the discrimination accuracy rates for the four participants in Table 4-9 are well above the average. This further suggests that these four participants were especially sensitive to the [g]-[k] distinction in their lexical representations, but the explanation for their performance on the word learning test is yet unclear.

4.9 Conclusions drawn from Experiment 3

Although most participants (regardless of training condition) did not encode the [g]-[k] distinction in their lexical representations of the new "second language" words, there were four participants who showed some evidence of having learned the phonemic distinction by encoding it lexically. However, these participants were from four different training conditions and therefore no generalizations about the type of training needed to learn to encode phonemic distinctions lexically can be made.

It is not concluded from this that the types of training that participants received are not sufficient to learn novel phoneme distinctions. Instead, future studies might manipulate the amount of training that participants receive in order to determine whether and with what types of evidence participants can develop distinctive lexical representations of a newly-trained contrast. Since the entire experiment session (training, discrimination test, word learning, and word test) was completed within 45 minutes, this may not be enough time for participants to have learned the contrast in a way that can be represented in their developing lexical representations for second language words. These results, however, indicate that it is possible for second language learners to have made perceptual gains in their treatment of a novel second language contrast (as evidenced in Experiments 1 and 2) without yet being able to use the contrast to distinguish lexical representations of new second language words.

The curious fact that the four participants (out of 80 participants total in this experiment) performed so well remains to be explained. It is concluded that their success resulted from individual characteristics of the four participants and is not attributable to information they learned during the experiment.

CHAPTER 5

DISCUSSION

5.1 Introduction

It is well-known that part of the second language learner's task is to determine the phoneme contrasts of the target language, and that this is critical to using the second language in all contexts (spoken or written, productive or receptive). While second language learners often do not implement novel phonemic contrasts in exactly the same way as native speakers do, as often evidenced, for example, by a "foreign accent", it is clear that knowledge of the contrasts is critical to second language communication. Experiments 1 and 2 were designed to investigate the types of evidence that second language learners can make use of from their second language input to learn novel sound contrasts. Experiment 3 was designed to further investigate whether or not these types of evidence lead learners to extend their phonetic learning to the lexicon in creating phonemically-distinctive representations. This final chapter serves as a summary of the findings and their implications for a theory of the evidence needed to learn novel sounds in second language acquisition. First, I consider each of the research questions posed in Chapter 1 in light of the findings of the experiments.

5.2 Research questions addressed

5.2.1 How do Lexical-Contrast-Based and Distribution-Based evidence compare in their effects on the perception of second language sounds?

Experiment 1 results indicated that both Lexical-Contrast-Based and Distribution-Based evidence can contribute to learning novel sound contrasts. However, participants trained with evidence indicating that the second language sounds [g] and [k] were distinctive did not exhibit significantly enhanced discrimination of the sounds over untrained participants as was originally expected. There were no significant differences between performance by either the Bimodal group (who received Distribution-Based evidence in support of a contrast) or the Contrast group (who received Lexical-Contrast-Based evidence, and arguably, Distribution-Based evidence as well in support of a contrast) and performance by the No-Training group. Instead, however, the participants who were given Distribution-Based evidence against a distinction (Monomodal training) discriminated the sounds *less* accurately than those who were not trained, and the participants who were given Lexical-Contrast-Based evidence against a distinction (No-Contrast) were nearly-significantly less accurate than the untrained participants. This finding suggests that the impact of training in the Monomodal and No-Contrast groups was actually to suppress an ability to discriminate the sounds that participants appear to have had prior to any training in the language.

Possible explanations for this surprisingly accurate discrimination of a novel sound contrast by the participants who received no training in the new language were discussed in section 2.6.2. There it was concluded that participants in this group were using an ability to detect phonetic differences between the sounds without treating the sounds as phonemically contrastive. In the absence of any knowledge about this language other than the information provided by the test pairs in the discrimination test, participants in the No-Training condition were able to identify this phonetic difference. For the experimental test items in the discrimination task, then, these participants tapped into an ability to detect phonetic differences that are not phonemic in their native language. This does make sense in light of the fact that second language learners can in fact learn to perceive and produce novel sound contrasts in many cases. In order for them to do this, they must at some point have an ability to detect phonetic differences that are not contrastive in their native language.

It was also found that participants trained with Lexical-Contrast-Based evidence in support of a distinction exhibited the most accurate discrimination of the sounds out of all of the training groups. Their discrimination was significantly more accurate than that of the Bimodal group. Although it cannot be argued that Bimodal or Contrast training significantly improved discrimination accuracy over the No-Training condition, it is concluded that Contrast training resulted in more accurate performance. However, the confound in the Distribution training conditions between Lexical-Contrast-Based evidence (pictures indicating word meaning) and Distribution-Based evidence (the inherently "bimodal" distribution of auditory stimuli in this condition) makes this finding difficult to interpret. Drawing conclusions about relative influence of Distribution-Based and Lexical-Contrast-Based evidence in support of a contrast, then, is left for the discussion of Experiment 2 results.

It does, however, appear that the addition of Lexical-Contrast-Based evidence in the Contrast condition provides an advantage to the learner over the auditory evidence alone (participants in the Contrast condition heard only the near-endpoint stimuli along the acoustic continuum). Discrimination by the Contrast group was significantly more accurate than that by participants who heard the same auditory stimuli as the Discrimination group but without pictures to indicate Lexical-Contrasts.

5.2.2 How do these types of evidence interact?

In Experiment 2, which was designed to provide information about how Distribution-Based and Lexical-Contrast-Based evidence interact in the acquisition of second language sound contrasts, participants heard various combinations of Distribution-Based and Lexical-Contrast-Based evidence. In two of the conditions, participants received agreeing evidence in support of or against there being a contrast between [g] and [k]; in the other two, participants received conflicting evidence for the contrast. The only statistically significant comparison for these conditions was between Bimodal+Contrast and Monomodal+No-Contrast, where participants received agreeing Distribution-Based and Lexical-Contrast-Based evidence either for or against there being a contrast. The Bimodal+Contrast group discriminated the sounds significantly better than the Monomodal+No-Contrast group, further supporting the finding from Experiment 1 that Distribution-Based and/or Lexical-Contrast-Based evidence impact second language speech perception. That the other comparisons were not significant prevents conclusions with respect to the relative impact of either kind of evidence on learning novel second language sound contrasts. A final set of tests compared the discrimination performance by the No-Training group from Experiment 1 to the Bimodal+Contrast and Monomodal+No-Contrast groups. There was no difference between No-Training participants' discrimination and that by the Bimodal+Contrast group, but the Monomodal+No-Contrast group exhibited poorer discrimination accuracy than the untrained participants. This provides further support for the Experiment 1 conclusion that participants had sensitivity to the phonetic differences between the sounds prior to participation in the experiments, and the effect of training was only to suppress pre-existing sensitivity to the difference.

The discrimination task used in this experiment appears to some extent to have tapped into participants' ability to discriminate phonetic differences, and whether it further tapped into knowledge of a phonemic contrast by some participants cannot be concluded from the findings of Experiments 1 and 2. Experiment 3, a word learning experiment, was intended to test for knowledge of phonemic distinctions.

5.2.3 Can short-term phonetic training be exploited in subsequent word learning? If so, under what training conditions?

It was argued that a discrimination task does not necessarily demonstrate sensitivity to phonemic contrasts in section 1.5 above. However, the word learning test in Experiment 3 was designed to test the content of second language lexical representations, which are arguably reflective of phonological knowledge.

Participants in this experiment were taught words containing the sounds they had just heard during training in Experiment 1 or 2 along with their meanings, and were then tested on their ability to detect mismatches in word-meaning pairs. The mismatches of interest were when the sounds [g] and [k] were substituted for each other, and participants' ability to detect the mismatches would be assumed to indicate that they had represented [g] and [k] contrastively in their representations of the newly-learned words.

However, no training groups were successful in detecting these mismatches. This null result does not necessarily mean that none of the groups had learned a phonemic contrast, but it does suggest that participants were not able to exploit any knowledge they might have had in learning a set of new second language words. While no group scored well overall on the mismatched test pairs, there were four participants, spread in no clear pattern across training conditions, who exhibited accuracy above 40% in the mismatched test pairs. It is tentatively concluded from this result that some individuals might have been able to represent the contrast lexically for some words. The overall conclusion from this experiment, however, is that the amount of training that participants received in Experiments 1 and 2 was not sufficient to make any comparisons among word learning across training groups. It is however possible to argue that word learning does impact development of sensitivity to novel phonemic contrasts, since participants in the training conditions that received Lexical-Contrast-Based evidence for the trained distinction might have been learning words during training.

5.3 Implications for second language learning and second language pedagogy

The results of the experiments presented here provide several conclusions that may inform second language pedagogy. First, recall that discrimination of [g] and [k] by participants in the Monomodal and the Monomodal+No-Contrast training conditions exhibited poorer performance than that by participants in the No-Training condition. This suggests that second language learners are already sensitive to some important phonetic differences before exposure to the second language. The results of this research, however, suggest that knowledge of word meaning and exposure to minimal pairs is beneficial to second language learners' ability to improve their discrimination of a novel sound contrast. While neither Bimodal nor Contrast training caused significant improvement in participants' discrimination of the new sounds, participants in the Contrast group exhibited the most accurate discrimination overall. This suggests that the availability of Lexical-Contrast-Based evidence results in more accurate discrimination of two new sounds than hearing the sounds without knowing what words mean.

One particularly interesting finding was that the Bimodal+Contrast group did not exhibit more accurate discrimination than the Bimodal group (who received the same auditory information as the Bimodal+Contrast group but without the Lexical-Contrast-Based evidence provided by the pictures). However, as mentioned above, the addition of pictures in the Contrast group—for whom auditory stimuli were in a bimodal distribution without phonetic variability—resulted in significantly better discrimination by that group than by the Near-Endpoint group (who received the same auditory information as the Contrast group but without the Lexical-Contrast-Based evidence provided by the pictures). It might thus be concluded that the discriminability benefit to training using a bimodal distribution of auditory tokens that includes phonetic variability is not increased by the addition of pictures, though there is a benefit of adding pictures when the bimodal distribution of auditory tokens does not include phonetic variability. More research is needed in order to determine how this result might be implemented pedagogically.

5.4 Linguistic-theoretical interpretations

Recent work in the Optimality-Theoretic (OT) framework provides an account of second language speech perceptual development that may prove useful in interpreting the results of the research presented here (Boersma 1999, Hayes 2002a,b, Escudero & Boersma 2001). In these models, the adult second language learner's initial state is the instantiation of her native language perception "grammar", which is a hierarchy of a set of constraints that determines the assignment of acoustic inputs to abstract phonological categories. A primary benefit of these models is that they provide an explicit theory of learning—that is, the learner's perception of second language sounds becomes more target-like as the native hierarchy of constraints is reorganized to reflect the target language's hierarchy. This reorganization is driven by an algorithm that computes the constraints responsible for mismatches between what the speaker intended to say and what the listener perceived (Escudero & Boersma 2001) and the constraints implicated in distributional properties of the acoustic input (Hayes 2002b). While it is beyond the scope of the present work to provide an OT account of the experimental data provided here, the beginnings of a model that integrates Distribution-Based and Lexical-Contrast-Based evidence is presented in Boersma *et al.* (in press).

5.5 Follow-up studies

These experiments have provided information about the relative influence of different types of evidence in the linguistic input on learners' ability to discriminate a novel phonetic sound contrast. These results lead to many more research questions. First, does the effect of training last beyond the experiment duration? Participants in these experiments were tested immediately following their training session, so one cannot determine from this data what the longer-term effects of training are. In follow-up studies, length of time between training and testing could be manipulated in order to test the longer-term effects of training.

Second, how much training is necessary before learners can extend their phonetic learning to the lexicon? The results of Experiment 3 suggest that the training that participants received was not sufficient for them to extend their phonetic learning to lexical representations. Follow-up studies could manipulate duration of training, or number of training sessions. Third, a related question, how much training is necessary to improve discrimination accuracy beyond participants' accuracy in the present research? Can more training positively impact discrimination accuracy beyond that reported in these experiments?

Fourth, why was performance by the No-Training group so accurate? This question was addressed in some detail in section 2.6.2. However, follow-up studies may further explain this result. It was originally assumed that native speakers of English would not discriminate [g] and [k] without having been trained on the contrast because the sounds are not contrastive word-initially in English. If this is the case, then speakers of languages where [g] and [k] are contrastive should exhibit accurate discrimination of the contrast. It should be noted that the particular properties of [g]-[k] contrast used in the present research are somewhat different from the properties of actual [g]-[k] contrasts in languages like Spanish. For example, the duration of prevoicing in these experiments (between zero and nine milliseconds) is shorter than that for Spanish voiced consonants. However, follow-up studies could test native speakers of Spanish or other languages with similarly-constructed, more authentic contrasts in order to compare their discrimination accuracy to that of non-native listeners.

A fifth question, related to the fourth: what is the effect of the phonological relationship of the learned contrast to participants' native language? As discussed briefly above, the [g]-[k] contrast is not entirely new to native speakers of English, as voiceless unaspirated [k] and prevoiced [g] belong to different phoneme categories in English (but not in the same environment). There are other possible relationships to the native language phoneme inventory. First, the novel contrast might be on a completely novel phonetic dimension (e.g. where one or both of the new sounds has a place of articulation not available in the native language). Another possibility is a contrast at a different location on a dimension that is already used by the native language (see work by Best and her colleagues for a typology of these relationships).

A pilot study performed for this dissertation may help to illuminate this issue. In brief, the pilot study employed the training and testing methods from Experiments 1 and

2, but tried to train participants to discriminate a singleton-geminate consonant contrast. (Languages like Japanese and Arabic distinguish singleton and geminate consonants.) The singleton-geminate consonant contrast is an entirely novel contrast for native speakers of English, as consonant length does not contribute to any native English distinctions word-internally. It was found that there were no differences between training groups on the discrimination test, despite several manipulations of the dimensions of the contrast. There is not enough data to draw firm conclusions from these results, but it is tentatively concluded that this entirely novel contrast was not learnable in the duration of the training session.

5.6 Final note on second language acquisition

Second language acquisition is a complex process that involves, among other things, the simultaneous acquisition of a novel sound system and a lexicon. The experiments presented here shed some light on the types of information that second language learners can use in the process of learning a new language, including information that comes from the acoustic properties of the input alone (Distribution-Based evidence) and from knowledge of words and meanings in the language (Lexical-Contrast-Based evidence).

It has traditionally been assumed that students of second languages infer novel second language sound contrasts from their knowledge of minimal pairs that demonstrate the importance of the distinction between the sounds. The results of the present experiments provide evidence for the utility of this type of information in learning a second language's sound contrasts. Although participants who were exposed to minimal pairs during training discriminated the novel sound contrast with greater accuracy than those who received only Distribution-Based evidence, this dissertation additionally provides evidence supporting learners' ability to learn to discriminate a novel sound contrast in the absence of a lexicon.

APPENDIX

These are the acoustic characteristics of the three synthetically-manipulated continua of tokens used in all three experiments.(from Maye 2000, p. 155).

syllable	/ga/							/ka/
token number	1	2	3	4	5	6	7	8
F3 onset (Hz)	2250	2210	2170	2130	2090	2050	2010	1970
F2 onset (Hz)	2225	2150	2075	2000	1925	1850	1775	1700
F1 onset (Hz)	300	325	350	375	400	425	450	475
Prevoicing (msec)	9	6	3	0	0	0	0	0

Table A1. Characteristics of the experimental tokens ranging from /ga/ to /ka/

syllable	/gæ/							/kæ/
token number	1	2	3	4	5	6	7	8
F3 onset (Hz)	2900	2860	2820	2780	2740	2700	2660	2620
F2 onset (Hz)	2900	2825	2750	2675	2600	2525	2450	2375
F1 onset (Hz)	350	375	400	425	450	475	500	525
Prevoicing (msec)	9	6	3	0	0	0	0	0

Table A2. Characteristics of the experimental tokens ranging from /gæ/ to /kæ/

syllable	/gɚ/						/kɚ/	
token number	1	2	3	4	5	6	7	8
F3 onset (Hz)	2450	2325	2275	2200	2125	2075	2000	1925
F2 onset (Hz)	1950	1825	1775	1700	1625	1575	1500	1425
F1 onset (Hz)	400	425	50	475	500	525	550	575
Prevoicing (msec)	9	6	3	0	0	0	0	0

Table A3. Characteristics of the experimental tokens ranging from /gɚ/ to /kɚ/

REFERENCES

Aslin, R.N, J.R. Saffran & E.L. Newport. 1998. Computation of conditional probability statistics by 8-month-old infants. *Psychological Science*, 9, 4, 321-324.

Best, C.T. 1994. The emergence of native-language phonological influences in infants: A perceptual assimilation model. In J.C. Goodman & H.C. Nusbaum, eds. *The Development of Speech Perception: The Transition from Speech Sounds to Spoken Words*. Cambridge: MIT Press. 167-224.

Best, C.T. & W. Strange. 1992. Effects of phonological and phonetic factors on cross-language perception of approximants. *Journal of Phonetics*, 20, 3, 305-330.

Best, C.T., G.W. McRoberts & E. Goodell. 2001. Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system. *Journal of the Acoustical Society of America*, 109, 2, 775-794.

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

Boersma, P. 1999. On the need for a separate perception grammar. Ms, University of Amsterdam.

Boersma, P., P. Escudero & R. Hayes. in press. Learning abstract phonological from auditory phonetic categories: An integrated model of the acquisition of language-specific sound categories. To be published in the *Proceedings of the International Congress of Phonetic Sciences*, Barcelona, August 2003.

Catford, J.C. & D.B. Pisoni. 1970. Auditory vs. articulatory training in exotic sounds. *Modern Language Journal*, 54, 7, 477-481.

Escudero, P. & P. Boersma. 2001. Modelling the perceptual development of phonological contrasts with Optimality Theory and the gradual learning algorithm. *Penn Working Papers in Linguistics, Proceedings of the 25th Penn Linguistics Colloquium.*

Flege, J.E, O.-S. Bohn & S. Jang. 1997. Effects of experience on non-native speakers' production and perception of English vowels. *Journal of Phonetics*, 25, 4, 437-470.

Flege, J.E. 1989. Chinese subjects' perception of the word-final English /t/-/d/ contrast: Performance before and after training. *Journal of the Acoustical Society of America*, 86, 5, 1684-1697.

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. 1995a. Two procedures for training a novel second language phonetic contrast. *Applied Psycholinguistics*, 16, 4, 425-442.

Flege, J.E. 1995b. Second language speech learning: Theory, findings, and problems. In W. Strange, ed. *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*. Baltimore: New York Press. 233-277.

Flege, J.E., M.J. Munro & I.R.A. MacKay. 1995. Effects of age of second-language learning on the production of English consonants. *Speech Communication*, 16, 1, 1-26.

Goto, H. 1971. Auditory perception by normal Japanese adults of the sounds 'L' and 'R'. *Neuropsychologia*, 9, 3, 317-323.

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

Guenther, F.H., F.T. Husain, M.A. Cohen & B.G. Shinn-Cunningham. 1999. Effects of categorization and discrimination training on auditory perceptual space. *Journal of the Acoustical Society of America*, 106, 5, 2900-2912.

Hayes, R.L. 2002a. The perception of Japanese consonant length by non-native listeners. *LabPhon* (Laboratory Phonology) 8, New Haven, June 2002. Poster.

Hayes, R. 2002b. An OT model of L2 speech perception. *Conference on Contrast in Phonology*, Toronto, May 2002. Poster.

Hayes, R. & K. Masuda. in prep. Word Learning and novel second language sounds: How do native speakers of English encode Japanese consonant quantity in memory? Ms, University of Arizona.

Iverson, P., P.K. Kuhl, & R. Akahane-Yamada. 2003. A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition*, 87, 1, B47-B57.

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

Kato, H. & K. Tajima. 2002. Native and non-native perception of phonemic length contrasts in Japanese: A categorization study. *Journal of the Acoustical Society of America* 112, 5, 2387.

Kato, H. K. Tajima & R. Akahane-Yamada. 2001. Native and non-native perception of phonemic length contrasts in Japanese. *Journal of the Acoustical Society of America*, 110, 2686.

Kingston, J. & E. Moreton. 1998. The discovery of natural classes by non-native listeners. *Journal of the Acoustical Society of America*, 103, 5, 2986.

Kingston, J., C. Bartels, J. Rice, D. Moore, R. Thorburn, J. Benkí & N.A. Macmillan. 1996. Effects of training procedures and vowel set on learning non-native vowel categories. *Journal of the Acoustical Society of America*, 100, 4, 2724.

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, 1-2, 125-139.

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

Landahl, K.L. & M.S. Ziolkowski. 1995. Discovering phonetic units: Is a picture worth a thousand words? *Papers from the Regional Meetings, Chicago Linguistic Society* 31, 1, 294-316.

Levy, E.S. & W. Strange. Effects of consonantal context on perception of French rounded vowels by American English adults with and without French language experience. *Journal of the Acoustical Society of America* 111, 5, 2361-2362.

Liberman, P. & S.E. Blumstein. 1988. *Speech Physiology, Speech Perception, and Acoustic Phonetics*. Cambridge, U.K.: Cambridge University Press.

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

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

MacKain, K.S. 1982. Assessing the role of experience on infants' speech discrimination. *Journal of Child Language*, 9, 3, 527-542.

MacKain, K.S., C.T. Best & W. Strange. 1981. Categorical perception of English /r/ and /l/ by Japanese bilinguals. *Applied Psycholinguistics*, 2, 4, 369-390.

Marsh, G. & R.J. Mineo. 1977. Training preschool children to recognize phonemes in words. *Journal of Educational Psychology*, 69, 6, 748-753.

Maye, J. 2000. The Acquisition of Speech Sound Categories on the Basis of Distributional Information. Doctoral Dissertation, University of Arizona, Arizona, USA.

Maye, J., J.F. Werker & L. Gerken. 2002. Infant sensitivity to distributional information can affect phonetic discrimination. *Cognition*, 82, 3, B101-B111.

Mayo, L.H., M. Florentine & S. Buss. 1997. Age of second-language acquisition and perception of speech in noise. *Journal of Speech and Hearing Research*, 40, 3, 686-693.

McClaskey, C.L., D.B. Pisoni & T.D. Carrell. 1983. Transfer of training of a new linguistic contrast in voicing. *Perception & Psychophysics*, 34, 4, 323-330.

Miller, J.L. 1994. On the internal structure of phonetic categories: a progress report. *Cognition* 50, 271-285.

Miller, J.L., T.B. O'Rourke & L.E. Volaitis. 1997. Internal structure of phonetic categories: Effects of speaking rate. *Phonetica*, 54, 21-137.

Miyawaki, K. 1975. An effect of linguistic experience: The discrimination of /r/ and /l/ by native speakers of Japanese and English. *Perception & Psychophysics*, 18, 5, 331-340.

Mochizuki, M. 1981. The identification of /r/ and /l/ in natural and synthesized speech. *Journal of Phonetics*, 9, 3, 283-303.

Mondini, M., P.M. van Alphen & J.L. Miller. 2002. Native-language influence on phonetic perception in Dutch-English bilinguals. *Journal of the Acoustical Society of America* 112, 5, 2357.

Pegg, J.E. & J.F. Werker. 1997. Adult and infant perception of two English phones. *Journal of the Acoustical Society of America*, 102, 6, 3742-3753.

Pisoni, D.B., R.N. Aslin, A. Perey & B. Hennesy. 1982. Some effects of laboratory training on identification and discrimination of voicing contrasts in stop consonants. *Journal of Experimental Psychology: Human Perception and Performance*, 8, 2, 297-314.

Pisoni, D.B., H. Nusbaum & B. Greene. 1985. Perception of synthetic speech generated by rule. *Proceedings of IEEE 73*, 1665-1676.

Pisoni, D.B. & S. Lively (1995) Variability and invariance in speech perception: A new look at some old problems in perceptual learning. In W. Strange, ed. *Speech Perception and Linguistic Experience: Issues in Cross-language Research*. Baltimore: York Press. 433-459.

Polka, L. 1992. Characterizing the influence of native language experience on adult speech perception. *Perception & Psychophysics*, 52, 1, 37-52.

Pruitt, J.S. 1996. The perception of Hindi dental and retroflex stop consonants by native speakers of Japanese and American English. Dissertation, University of South Florida, Florida, USA.

Repp, B. 1984. Categorical perception: Issues, methods and findings. In N.J. Lass, ed. *Speech and Language: Advances in Basic Research and Practice. Volume 10.* New York: Academic Press.

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

Smits, R., J. Sereno & A. Jongman. (submitted) Categorization of sounds. Submitted to *Journal of Experimental Psychology: Human Perception and Performance*.

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

Sundara, M., L. Polka. L. Campisi, F. Genesee & C. Marcoux. 2002. Learning to perceive English /d/ versus /ð/: A comparison of French, English, and English-French bilingual 4-year-olds. *Journal of the Acoustical Society of America* 111, 5, Pt. 2, 2363.

Tajima, K., A. Rothwell & K.G. Munhall. 2002. Native and non-native perception of phonemic length contrasts in Japanese: Effect of identification training and exposure. *Journal of the Acoustical Society of America*, 112, 5, 2387.

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

Wayland, R.P. & S.G. Guion. 2001. Discrimination of Thai tones by naive and experienced native English speakers. *Journal of the Acoustical Society of America*, 110, 5, 2686.

Wayland, S.C., J.L. Miller & L.E. Volaitis. 1994. The influence of sentential speaking rate on the internal structure of phonetic categories. *Journal of the Acoustical Society of America*, 95, 5, 2694-2701.

Werker, J.F. 1994. Cross-language speech perception: Developmental change does not involve loss. J. Goodman & H. Nusbaum, eds. *Speech Perception and Word Recognition*. Cambridge: MIT Press. 93-120.

Werker, J.F. & J.S. Logan. 1985. Cross-language evidence for three factors in speech perception. *Perception & Psychophysics* 37, 1, 35-44.

Werker, J.F. and R.C. Tees. 1984. Phonemic and phonetic factors in adult cross-language speech perception. *Journal of the Acoustical Society of America* 75, 6, 1866-1878.

Werker, J.F., J.H. Gilbert, K. Humphrey & R.C. Tees. 1981. Developmental aspects of cross-language speech perception. *Child Development* 52, 1, 349-355.

Yamada, R.A. 1995. Age and acquisition of second language speech sounds: Perception of American English /r/ and /l/ by native speakers of Japanese. In W. Strange, ed. *Speech Perception and Linguistic Experience*. Baltimore: York Press. 305-320.

Yamada, R.A. & Y. Tohkura. 1992. The effects of experimental variables on the perception of American English /r/ and /l/ by Japanese listeners. *Perception & Psychophysics*, 52, 4, 376-392.